Engineering Applications of Artificial Intelligence 139 (2025) 109590

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# Etching process prediction based on cascade recurrent neural network ☆

Zhenjie Yao [a,b], Ziyi Hu [a,b], Panpan Lai [a,b], Fengling Qin [a,b], Wenrui Wang [a,b], Zhicheng Wu [a], Lingfei Wang [a], Hua Shao [a], Yongfu Li [c], Zhiqiang Li [a,b], Zhongming Liu [d], Junjie Li [a,b,*], Rui Chen [a,b,*], Ling Li [a,b]

[a] Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics, Chinese Academy of Sciences, Beijing, China
[b] School of Integrated Circuits, University of Chinese Academy of Sciences, Beijing, China
[c] Department of Micro and Nano Electronics Engineering, Shanghai Jiao Tong University, Shanghai, China
[d] Changxin Memory Technologies Inc., Hefei, Anhui, China

ARTICLE INFO

*Keywords:*
Etching model
Etching profile prediction
Cascade recurrent neural network
Cascade combination

ABSTRACT

Etching is one of the most critical processes in semiconductor manufacturing. Etch models have been developed to reveal the underlying etch mechanisms, which employs rigorous physical and chemical process simulation. Traditional simulation is very time consuming. The data-driven artificial intelligence model provides an alternative modeling approach. In this paper, a Cascade Recurrent Neural Networks (CRNN) is proposed to model and predict etching profiles. The etching profile is represented by polar coordinates and modeled by the recurrent neural networks, the corresponding etching parameters (e.g., pressure, power, temperature, and voltage) are integrated into the network through cascade combination layers. Experimental results on a dataset of 10,000 simulated etching profiles demonstrated the effectiveness of our method: compared with traditional etching simulation methods, CRNN can speedup 21,000× with an average error of less than 0.7 nm for 1 step prediction. Furthermore, compared to simple deep neural networks, the Mean Absolute Errors (MAE) could be reduced from 1.7329 nm to 1.3845 nm for 10 steps prediction. Finally, the effectiveness and accuracy of CRNN etching predictor is validated through fine-tuning on experimental data.

## 1. Introduction

To further increase the density and functionality of large scale integrated circuit, device concepts with increased structural complexity at ultra-scaled dimensions are proposed: as examples of advanced logic technology, Gate-all-around Field Effect Transistor (GAAFET) is firstly proposed as the device architecture of sub-3 nm, as next, the Complementary Field Effect Transistor (CFET) device is expected to further extend the GAAFET family to technology nodes of 1 nm or even below. To support the adoption of both GAAFET and CFET, complex and precise lateral etching technique are needed (Kim et al., 2020; Subramanian et al., 2020). On the other hand, for the memory applications, to continue the miniaturization of Dynamic Random Access Memory (DRAM) technology, the concept of 3D-stacking is adopted, which has to be manufactured in a 3D-stacked way and require precise control of both vertical and lateral etching (Wang et al., 2020). To summarize, accurate etching control and prediction is becoming one of the key knobs of the semiconductor manufacturing process (Chen et al., 2020).

The underlying physical mechanism of the etching process is composed of multiple complex interactions, including adsorption, dissociation, sputtering etc.. Besides the intrinsic physical mechanism, the complexity of etching process is further manifesting itself through strong dependence on the target etching profile, such as undercutting (Fukumoto et al., 2009), tapering (Guo and Sawin, 2009), bowing (Huo et al., 2022), and microtrenching (Hoekstra et al., 1998). All these effects impacts the critical dimension (CD) of features and therefore the device performance. To avoid the complex multi-physics issues, foundries typically rely on actual etching experiments (aka. "learning-cycle") for process calibration, such pragmatic approach could ensure an outcome but is expensive. Instead, etching simulation provides profiles during the entire evolution process of the etching, which allows us to take snapshots of the intermediate states during etching process. By observing the profiles under different process conditions at different

☆ This work is supported by the National Key R&D Program of China (2023YFB3611300), CAS Youth Interdisciplinary Team (JCTD-2022-07), Chinese Academy of Sciences Supporting Technology Talent Project (E2YR01×001). The code and dataset are available at https://github.com/yzjba/AI-Etch.

* Corresponding authors at: Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics, Chinese Academy of Sciences, Beijing, China.
*E-mail addresses:* lijunjie@ime.ac.cn (J. Li), chenrui1@ime.ac.cn (R. Chen).

https://doi.org/10.1016/j.engappai.2024.109590
Received 11 December 2023; Received in revised form 5 August 2024; Accepted 29 October 2024
Available online 8 November 2024

time, engineers gain insights into the etching process, identify the correlation between etching parameters and various effects and optimize the parameters (Kuboi et al., 2016). In summary, the importance of the intermediate etching profiles in semiconductor manufacturing lies in optimizing process design and cost-saving. Various models, such as the String method (Abdollahi-Alibeik et al., 1999; Vyvoda et al., 1999), Level-set method (Sethian et al., 1999; Kokkoris et al., 2004), and Monte Carlo method (Hu et al., 2023; Hoekstra et al., 1998), have been developed for etching simulation. Those models are very time-consuming (Hoekstra and Kushner, 1998). A more efficient etching simulation methods is desired to enable model-driven process development and optimization.

Machine learning, especially deep neural networks, has dramatically improved the performance of various applications, such as computer vision (He et al., 2017), natural language processing (Devlin et al., 2018), healthcare (Yao et al., 2018), and self-driving (Badue et al., 2021), nonlinear system (Song et al., 2023a,b, 2024). Deep neural networks had shown superior performance in the field of semiconductor manufacturing process simulation and Electronic Design Automation (EDA), such as lithography and mask optimization (Chen et al., 2022; Yang et al., 2022), chemical mechanical planarization (Bao and Chen, 2020), etching (Chen et al., 2020; Xiao and Ni, 2021), static timing analysis (Savari and Jahanirad, 2020), and vulnerability analysis (Rahimifar et al., 2024). The successful applications in various fields have inspired us to use deep neural networks for process simulation.

There have been some attempts to use neural networks for etching modeling. Chen et al. proposed a deep neural networks to predict etch bias, which achieves average relative error less than 15% for 2D pattern (Chen et al., 2020). Xiao et al. developed a multiscale modeling and Recurrent Neural Networks (RNN) based optimization framework of a plasma etch process on a three-dimensional substrate (Xiao and Ni, 2021). Kanarik et al. introduced a human–machine collaborative etching process optimization framework based on Bayesian optimization, which could reduce the experiment cost by half for processing recipe determination (Kanarik et al., 2023). However, in these works, neural networks is used to predict key parameters of etching (such as depth, roughness, and critical dimension), rather than complete profiles. Predicting the complete etching profiles offers the potential to partially replace conventional etching simulation. Profiles of one etching process can be regarded as multivariate time series, and there are many successful cases of predicting multivariate time series based on RNN and graph neural networks (Shih et al., 2019; Wu et al., 2020; Yao et al., 2021; Yang et al., 2021; Zerveas et al., 2021). Jeong mentioned a RNN for etching prediction (Jeong et al., 2021). Myung et al. proposed a transformer for etching profile prediction, which adopted inductive bias to avoid results of anti physical laws (Myung et al., 2021).

Most existing works mainly focuses on key parameters of etching morphology and cannot predict profiles. Few studies that can predict profiles have not undergone comprehensive customization for the etching process. There is still room for improvement in prediction performance. Therefore, we customize an end-to-end, purely data-driven cascade recurrent neural network (CRNN) for etching prediction. The Long Short Term Memory (LSTM) is used for extracting spatial and temporal features. By combining the extracted spatial and temporal features from etching profiles, as well as corresponding etching parameters in a cascade manner, better etch prediction performance is achieved. The main contributions of this paper are summarized in :

- **Application innovation.** An unconstrained, fully data-driven deep learning model for etching profiles prediction.
- **Profile data representation and modeling.** Polar coordinate is adopted to effectively represent the etching profiles. The Recurrent Neural Network (RNN) are developed to model the stepwise etching depths, which captures the temporal features of the etching data.
- **Model innovation.** Etching parameters are integrated into the neural networks by novel cascade combination layers, improved accuracy of the etching process prediction is demonstrated.

Furthermore, the effectiveness of the CRNN etching prediction model is validated on both simulated data and experimental data, which outperforms various conventional neural networks. The remaining part of this paper is organized as follows. Section 2 introduces the dataset and data representation. Section 3 formulates the etching profile problem and introduces the CRNN in detail. Section 4 presents the experimental results on both 10 000 profiles from 200 simulated etch processes and experimental data. Finally, Section 5 gives conclusions and discusses future works.

## 2. Dataset and representation

In this work, a plasma etching tool named iProcess-ETCH (Hu et al., 2023) is adopted to simulate the feature profile evolution in the process of etching Si substrate with Cl2 under different process conditions. Cellular representation (Rezvankhah et al., 2013; Than and Büttgenbach, 1994) is adopted to represent an etched trench structure. Monte Carlo method (Mahorowala and Sawin, 2002; Mahorowala et al., 2002) is used to simulate particle incident trajectories and surface reaction processes. The model includes five kinds of reactions: adsorption, spontaneous reaction dissociation, physical sputtering, high-energy particle removal and reactive etch. While the ion enhancement etching indicates that high-energy ions are directly etched, and low-energy ions have a certain probability of reflection (Winters and Plumb, 1991).

The dataset are generated by simulating the etching Si substrate with Cl2 under different process conditions. During each simulation, we simulate the etch process on a 2D region of 700 nm × 200 nm and recorded the results on 700 × 200 cells with 1 nm resolution. The etching process was simulated in the total reaction time of 5 s and the etching profile was recorded each 0.1 s, therefore, each simulation process results in 50 profiles.

By changing the four etching process conditions (chamber pressure, ion energy, temperature and chamber power, other input parameters, such as pattern density and hard mask topography, keep unchanged), we obtained 200 different etching process records composed of 10,000 profiles. In our experiments, 150 etching records (7500 profiles) are randomly selected for training, and the left 50 (2500 profiles) etch records format the test dataset.

One recorded etching process is shown in Fig. 1. As shown in the figure, dark gray indicates the hard mask, light gray is the silicon substrate, and the six black dashed lines indicate the etching profiles at 0s, 1s, 2s, 3s, 4s, and 5s, respectively. In this work, we focus on the profiles of the silicon substrate without considering the hard mask part. Due to the etching range exceeding the range without mask, the entire etching profile is bowl shaped. There may be two points on the profile on one vertical line, so the entire profile cannot be represented by depth directly. As an alternative, we adopt polar coordinates, which is capable of representing bowl shaped curves, to represent the entire etching profile. As shown in Fig. 1, each point $p$ on the profile can be represented by an angle $\theta$ and a distance $d$ from the center of the uncovered silicon substrate $o$. By uniformly sampling the distance on each angle from the etched profile, each profile can be discretized into a multi-dimensional vector.

## 3. Cascade recurrent neural networks for etching profile prediction

### 3.1. Problem formulation

The etched profiles spanning multiple time steps form a multivariate time series. Therefore, the etching profile prediction problem can be formulated as a multivariate time-series prediction task. The objective is to forecast the future etching profile based on historical profiles and

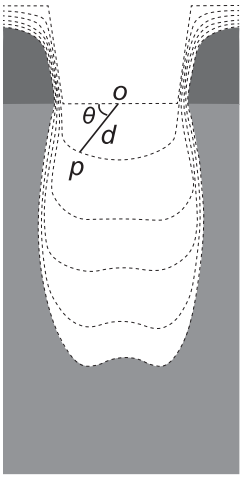



**Fig. 1.** Etching profiles.

**Table 1**
Notation description.

| Notation | Description |
|---|---|
| $N$ | Number of samples of each profile. |
| $x_{it}$ | Distance on the direction of $i$ degree at the $t_{th}$ time step. |
| $X$, $X_h$, $X_p$ | Profiles, Profiles in history window, and Profiles in prediction window. |
| $T$, $T_h$, $T_p$ | Number of time steps in $X$, $X_h$, and $X_p$. |
| $X_e$ | Etching parameters. |
| $e$ | Number of etching parameters |
| $\vec{x}(t)$ | Profile of time step t. |
| $\vec{y}(t)$ | Predicted profile of time step t. |
| $f_t$ | Forget gate of LSTM cell at $t_{th}$ time step. |
| $i_t$ | Input gate of LSTM cell at $t_{th}$ time step. |
| $h_t$ | Hidden state of LSTM cell at $t_{th}$ time step. |
| $o_t$ | Output gate of LSTM cell at $t_{th}$ time step. |
| $K_t$ | Intermediate state of the LSTM cell at $t_{th}$ time step. |
| $C_t$ | State of the LSTM cell at $t_{th}$ time step. |
| $\theta$ | Model parameters. |
| $\alpha$ | Learning rate. |
| $h^{(l)}$ | Features in $l$th hidden layer. |
| $W_*$ | Different weights in neural networks. |
| $b_*$ | Different biases in neural networks |

etch parameters. Let $X \in \mathbb{R}^{N\times T}$ denotes the profiles of an etching process, where $N$ represents the number of samples for each profile, and $T$ denotes the number of time steps. Here, $x_{it}$ represents the distance in the direction of $i$ degrees at the $t$th time step. Specifically, the aim is to predict future profiles $X_p \in \mathbb{R}^{N\times T_p}$ using historical profiles $X_h \in \mathbb{R}^{N\times T_h}$ and etching parameters $X_e \in \mathbb{R}^{e\times 1}$ as inputs. Mathematically, this entails finding the following mapping:

$$f(X_h, X_e) \to X_p. \tag{1}$$

The mathematical notations are listed in Table 1.

In this paper, we sample 0 to 180 degrees at intervals of 1 degree, a profile result in a vector of 181 unique distance on each angle, which means $N = 181$.

In our experiments, we record the etching profile every 0.1 s, which result in 50 profiles in 5 s.

### 3.2. Proposed cascade recurrent neural networks

To comprehensively consider etching profile records and etching parameters, we designed a cascade neural network architecture based on RNN, named CRNN. The architecture of CRNN is illustrated in Fig. 2. CRNN is utilized to partially replace traditional physical simulation methods, offering a speedup of $21,000\times$ while maintaining accuracy. As depicted in the figure, CRNN requires two types of features: historical etching profiles and etching parameters. An LSTM is employed to model the historical etching records, while the etching parameters are integrated into the network architecture through multiple cascades. We will now proceed to introduce each part in detail.

**Table 2**
Statistical characteristics of features.

| Data | min | max | mean | std |
|---|---|---|---|---|
| Etching depth (nm) | 1 | 372 | 69.18 | 37.61 |
| Pressure (MTorr) | 40 | 160 | 112.40 | 25.63 |
| Power (W) | 20 | 260 | 117.80 | 60.04 |
| Temperature (K) | 300 | 900 | 603.00 | 144.06 |
| Voltage (V) | 8 | 20 | 11.55 | 3.16 |

#### 3.2.1. Input and preprocessing

As shown in Fig. 2, the input of the network consists of two parts: one part is the historical profiles, and the other part is the etching parameters. In our experimental setting, the input includes 48 etching depths. If there are less than 48 depths, the previous etching depths is supplemented with 0. So the input of LSTM unit is of size $181 \times 48$, the etching parameters are of size 4.

Table 2 describes the statistical characteristics of all the etching features. The etching profile is composed of a series of depth data, measured in nanometers. There are four etching parameters: chamber pressure, chamber power, temperature, and ion energy (determined by voltage). The range of values for features varies significantly and normalization is necessary. All the features are normalized by

$$\tilde{x} = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{2}$$

where $x_{max}$ and $x_{min}$ indicate the maximum and minimum value of the corresponding feature. All the historical etching profiles share the same maximum and minimum value.

Furthermore, through the differential operation of cumulative etching depth, we can obtain the etching depth of each time step, which is fed into the proposed neural networks.

#### 3.2.2. LSTM for etching depth data

As mentioned above, etching profile prediction is essentially a multivariate time series prediction problem. RNN is an effective modeling method for multivariate time series. The LSTM model is essentially a specific form of RNN, which addresses the long and short-term memory issues of RNN by incorporating gates into the model. The structure of an LSTM cell is illustrated in Fig. 3. Three logic control units (Forget Gate, Input Gate, Output Gate) are integrated into the RNN, allowing it to effectively capture long and short-term information in the multivariate time series. These gates control the input and output of information flow and the state of the cell by adjusting the weights on the connections between the memory unit of the neural network.

For the forget gate, we have

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \tag{3}$$

where $\sigma(\cdot)$ is sigmoid activation, $[\cdot,\cdot]$ indicates concatenation of two vectors.

For the input gate, we have

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i). \tag{4}$$

The intermediate state of the cell is defined by

$$K_t = tanh(W_K[h_{t-1}, x_t] + b_K). \tag{5}$$

The state of the cell is determined by old and intermediate state of the cell together.

$$C_t = f_t \odot C_{t-1} + i_t \odot K_t, \tag{6}$$

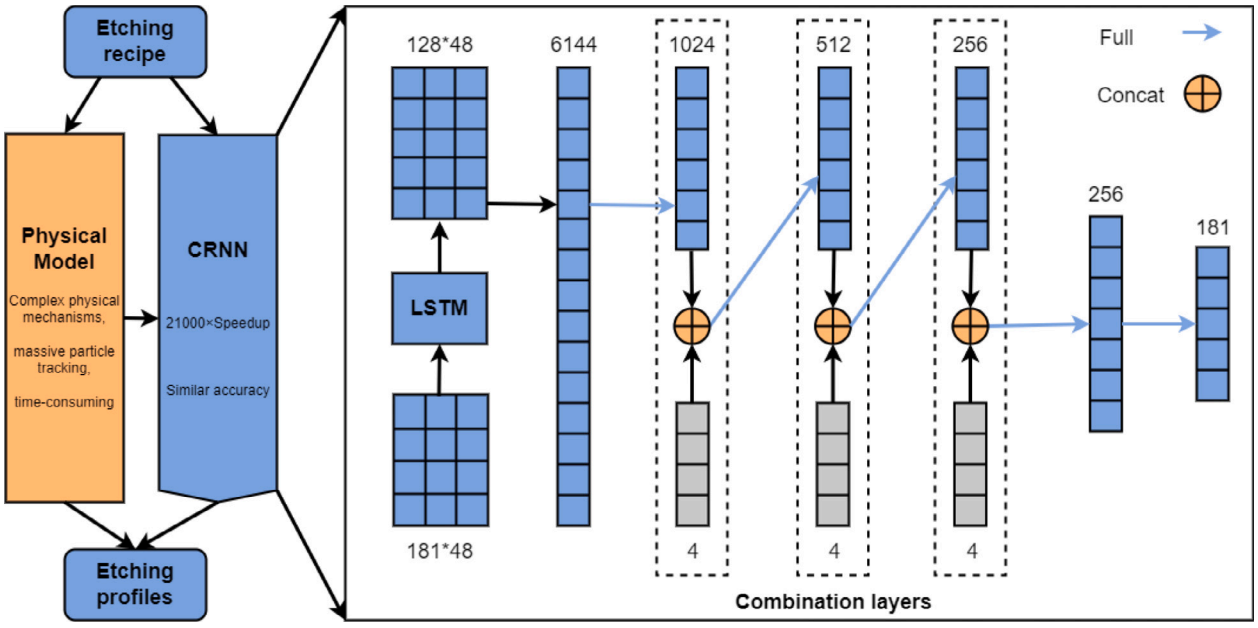



**Fig. 2.** The architecture of CRNN based etching simulation.

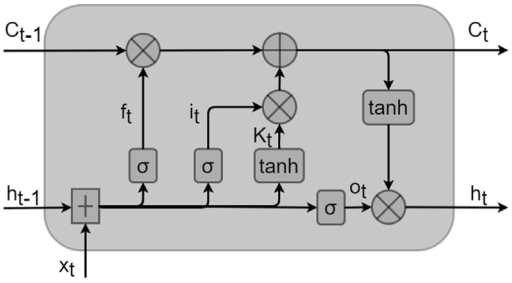



**Fig. 3.** The structure of LSTM.

in which $\odot$ indicates point-wise multiplication. $i_t$ and $f_t$ control the combination by weighting.

As for the output gate, we have

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o). \tag{7}$$

The hidden state is determined by output weighting as

$$h_t = o_t \odot tanh(C_t). \tag{8}$$

In our setting, the input of LSTM is a multivariate time series with 48 time steps and 181 variables as inputs, the output is with 48 time steps and 128 features. The output $128 \times 48$ matrix is expanded into a vector of 6144 elements, the vector of 6144 features are converted into a vector of 1024 features through a fully connected layer, which are the final features extracted from the etching profiles. Mathematically, that is

$$h^{(l)} = \sigma(W^{(l)} h^{(l-1)} + b^{(l)}), \tag{9}$$

where $h^{(l)}$ is the features in $l$th hidden layer, $W^{(l)}$ and $b^{(l)}$ is weight and bias, the activation function $\sigma$ defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{10}$$

LSTM is capable of capturing long-term and short-term dependencies in sequential data, effectively modeling the historical etching profiles, and making accurate predictions for subsequent etching profiles.

### 3.2.3. Cascade combination layers

The features extracted from etching profiles by LSTM are combined with etching parameters in a cascade manner, through the fully connected layer. Specifically,

$$h^{(l)} = \sigma(W^{(l)} concat(h^{(l-1)}, X_e) + b^{(l)}), \tag{11}$$

where $concat(h^{(l-1)}, X_e)$ is the concatenation of $h^{(l-1)}$ and $X_e$.

Then the vector combined with 4 etching parameter features in the first combination layer, in which it is concatenated with 4 etching parameter features, and transformed into vector of 512 features through the fully connected layer. In the second combination layer, the vector of 512 features is concatenated with 4 etching parameter features again, and transformed into a vector of 256 features. In the third combination layer, the vector of 256 features is concatenated with 4 etching parameter features again, and transformed into a vector of 256 features, which is the final feature vector. The final feature layer is transformed to the output layer, which is the final etching depth of the next time step.

The 4 etching parameters are integrated into the backbone neural networks by combination layer in a cascade manner. The information contained in the etching profiles and etching parameters is fully modeled, allowing for more accurate prediction of future etching depths.

Although cascade combinations may increase the number of fully connected layers, these layers are the most computationally efficient connections in neural networks. Consider that CRNN is 21,000 times faster than traditional methods, the slight increase in complexity from the combined layers can be ignored.

Compared to conventional direct concatenation, the proposed cascade combination layer allows multiple joint learning between profile features and etching parameter features. After each concatenation, interactive feature extraction is performed through fully connected layers, resulting in higher level features. Through the cascade "concatenation-combination" process, full integration of profile features and etching parameters is achieved, thereby enhancing the accuracy of subsequent etching profile predictions.

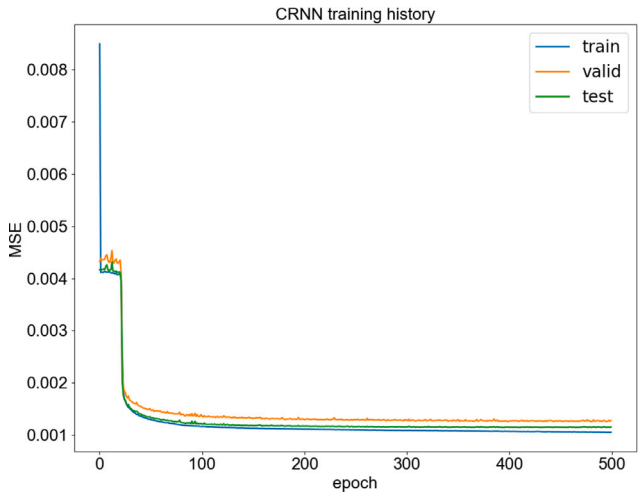



**Fig. 4.** Learning curves.

#### 3.2.4. Training of CRNN

During training, The proposed CRNN are optimized using techniques such as backpropagation and gradient descent. As shown in Algorithm 1, the parameters of the network, including the weights and biases of all the layers and gates, are adjusted iteratively to minimize the Mean Squared Error (MSE) loss function, through the optimization algorithms Adam (Kingma and Ba, 2014). The MSE loss is defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{12}$$

The parameters are updated through the gradients obtained.

It is worth noting that the LSTM part utilizes Backpropagation Through Time (BPTT) to adjust its parameters. BPTT is a variant of the backpropagation algorithm specifically designed for RNN, which extends the traditional backpropagation algorithm to sequential data by unfolding the network across time steps. This allows the network to learn from past experiences and adjust its parameters accordingly.

**Algorithm 1** Training the CRNN

Initialize model parameters $\theta$ ($W_f$, $W_i$, $W_K$, $W_o$, $b_i$, $b_f$, $b_o$, $b_c$...)
Set optimizer Adam
**for** each epoch **do**
  Initialize gradients
  **for** each batch of data **do**
    Get current batch of data and labels, $(x^{(b)}, y^{(b)})$
    Forward pass: $\hat{y}^{(b)} = CRNN(x^{(b)})$
    Compute loss: $\mathcal{L}^{(b)} = \mathrm{MSE}((y^{(b)}, \hat{y}^{(b)})$
    Backward pass: $\nabla^{(b)} = \frac{\partial L^{(b)}}{\partial \theta}$
    Update parameters: $\theta_{t+1} = \theta_t - \alpha \cdot \nabla^{(b)}$
  **end for**
**end for**

Fig. 4 shows the learning curves of CRNN on training, validation, and testing dataset. From the figure, we find that the MSE curve rapidly decrease at about 20 epochs, then gradually stabilizes. After training more than 100 epochs, it essentially converges, with only a slight continuous decrease in MSE on the training set, while the MSE on the validation and test set remain stable. Therefore, we opt to train CRNN for 200 epochs in experiments.

### 3.3. Multiple step prediction

Prediction for more than 1 time step is achieved through recursive 1 step prediction, which can be summarized as Eqs. (13) and (14), in which $\vec{x}(t)$ is the etching depth on $t$th time step, $\vec{y}(t+1)$ is the predicted etching depth on $t+1$ time step, and $\vec{y}(t+s)$ is the predicted etching depth on $t+s$ time step. $f(\cdot)$ is the 1 step prediction function, which is a CRNN. Prediction on multiple time steps could be obtained by recursive 1 time step prediction, which replacing the unknown true etching depth with the previous single step prediction result.

$$\vec{y}(t+1) = f([\vec{x}(t-T_h+1), \vec{x}(t-T_h+2), \ldots, \vec{x}(t-1), \vec{x}(t)], x_e). \tag{13}$$

$$\vec{y}(t+s) = f([\vec{x}(t-T_h+s), \vec{x}(t-T_h+s+1), \ldots, \vec{x}(t), \vec{y}(t+1), \vec{y}(t+2), \ldots, \vec{y}(t+s-1)], x_e). \tag{14}$$

### 3.4. Fine-tuning method

In practice, obtaining the etching profile involves observing the cross-section of a trench with a Scanning Electron Microscope (SEM). It is impractical to record all sequential steps in cross-sectional profiles during the etching process. Consequently, acquiring etching profiles in practice is challenging, which hinders the training of neural networks. To address this issue, we employed a fine-tuning method to transfer the trained model based on simulation data to practical data. The architecture is shown in Fig. 5. As illustrated in the figure, a large simulation dataset is utilized to pre-train the base neural network. Subsequently, fine-tuning the base neural network using a small dataset comprising experimental etching profiles enables the derivation of a feasible neural network prediction model, which can be deployed to predict actual etching profiles.

## 4. Experimental results

### 4.1. Experimental setting

In the experiments, we set $T_h = 48$ and $T_p$ range from 1 to 10. The number of neurons in each layer is set as Fig. 2. The CRNN models are trained by minimizing the mean square error using Adam optimizer (Kingma and Ba, 2014) with initial learning rate 0.002 for 200 epochs. The batch size is 64. All experiments in this study were conducted using Python and implemented with the PyTorch framework. These general parameters were selected from a small range around the optimal values identified in previous work. The specific key parameters, such as the number of cascaded combination layers, have been thoroughly tested and are presented in the subsequent experiments.

### 4.2. Comparison

Mean Absolute Errors (MAE) is adopted to evaluate and measure the prediction performance of different methods, which is defined as

$$\mathrm{MAE}(y, \hat{y}) = \mathrm{E}\left(\left|y_i - \hat{y}_i\right|\right) = \frac{\sum_{i=0}^{N}\left|y_i - \hat{y}_i\right|}{N}. \tag{15}$$

where $|\cdot|$ denotes taking absolute values, $N$ is the number of predicted points. Smaller MAE indicates better prediction performance.

We compare the proposed CRNN with conventional DNN (Deep Neural Network), Transformer, CNN (Convolutional neural network) and RNN. In this paper, DNN is implemented by multilayer perceptron. Combining different neural networks with different etching parameter integration manner, we obtain the following baseline models for performance comparison in Table 3.

The prediction performance of all the 5 models are listed in Table 4. The table provides a list of prediction error of different time steps, including 1 step, 3 steps, 5 steps, 7 steps and 10 steps (One time step corresponds to 0.1 s etching). To quantitatively represent the ratio of error reduction, we have provided the ratio of MAE to that of DNN for each model, named Ratio2DNN (the smaller the better). By comparing the prediction results of different models and time steps, we have the following observations:

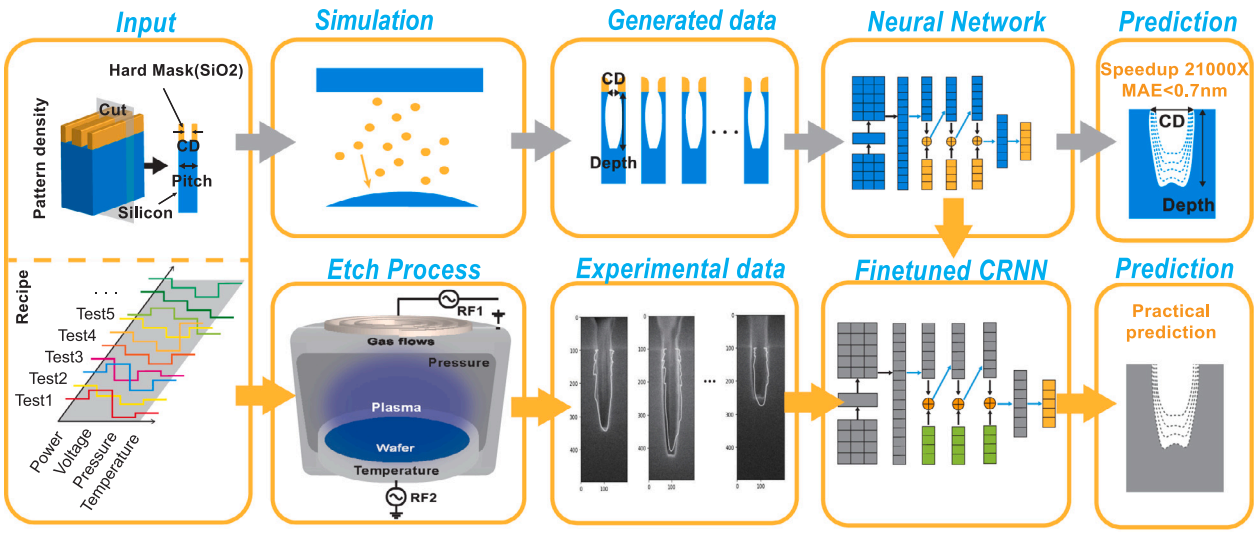



**Fig. 5.** Workflow of fine-tuning. CRNN is trained with simulated data and can be used for acceleration. Through the fine-tuning method, a small amount of experimental data can be used to calibrate the pre-trained CRNN to a more practical one.

**Table 3**
Various neural networks for comparison.

| Model | Note |
|---|---|
| DNN | Predicting the future etch depth based on latest cumulative etching depth and last step etching depth, and etching parameters with DNN. |
| TRANS | Predicting the future etch depth with historical one-step etching depths with Transformer, and integrate the etching parameters once only. |
| RNN | Predicting the future etch depth with historical one-step etching depths with RNN, and integrate the etching parameters once only. |
| CNN | Predicting the future etch depth with historical one-step etching depth with CNN, and integrate the etching parameters once only. |
| CCNN | Predicting the future etch depth with historical one-step etching depth with CNN, and integrate the etching parameters in a cascade manner. |
| CRNN | Predicting the future etch depth with historical one-step etching depth with RNN, and integrate the etching parameters in a cascade manner. |

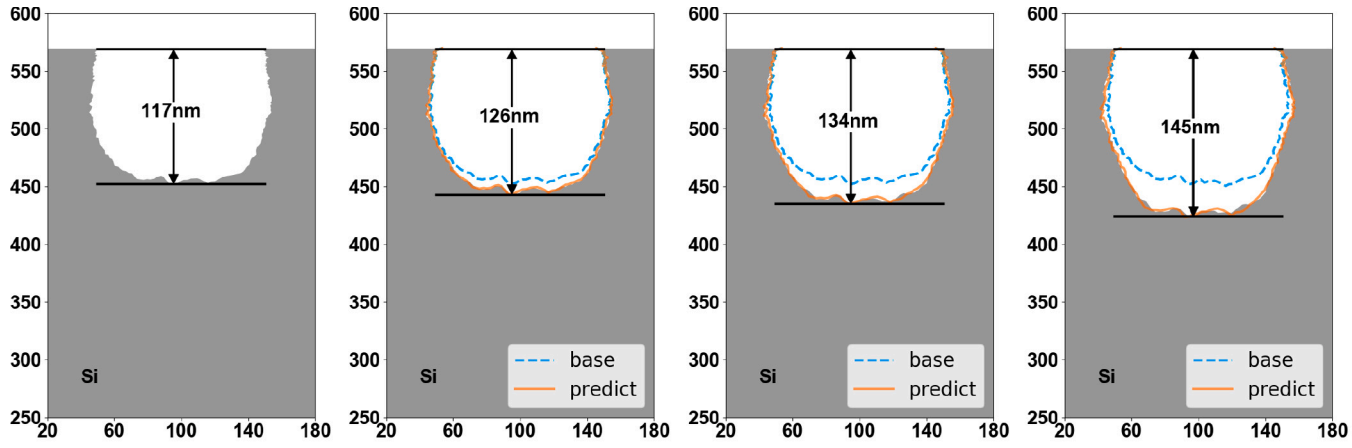
- **Transformer is not as good as other neural networks.** The poor performance of the Transformer model may be attributed to the insufficient amount of data. For example, vision Transformer (Dosovitskiy et al., 2020) requires pre-trained on more than 100 million images to surpass CNN. Our dataset contains only 10,000 profiles, which is insufficient to fully leverage the modeling capabilities of the Transformer. In such cases of limited sample size, Transformer may not be a good choice.
- **RNN and CNN works better than DNN.** Compare the 4 RNN/CNN models and the DNN model, we can find that all the RNN/CNN models outperform the DNN model significantly. Taking RNN as an example, the average MAE of RNN for 10 time steps is 1.5190 nm, while that of DNN is 1.7329 nm, the MAE reduced by 12.3%. The comparison between CNN, as well as the CCNN and CRNN, and DNN, all support the same inference. Due to their excellent sequence modeling capability, CNN and RNN are more suitable for modeling etching profiles than DNN.
- **Cascade combination of etching parameters did improve the prediction performance.** CRNN integrates etching parameters in a cascade manner, which significantly improving prediction performance, the MAE for 10 time steps reduced from 1.5190 nm to 1.3845 nm. Compared with CNN, the MAE of CCNN for 10 time steps reduced from 1.4819 nm to 1.4225 nm. Cascade combination of etching parameters is effective for both RNN and CNN.

- **Recursive multi-step prediction can provide acceptable prediction performance for the etching task.** For all the five models, the MAE increases as the time step increases, due to the inevitable accumulation of errors. However, the multi-step prediction error does not increase explosively, but maintains a stable and controllable growth trend. Moreover, the performance margin of the CRNN model increases with the increase of time steps. For 1 time step prediction, CRNN reduced MAE by 2.7%; for 10 time steps prediction, it reduced by 20.1%.
- **The proposed CRNN is the optimal etching predictor.** CRNN achieves better performance than all the other models, including CCNN. For 10 time steps prediction, CRNN reduced MAE by 20.1%, while CCNN, which is the second best model, reduced MAE by 17.9%.CRNN models the stepwise etching depth with RNN, integrates etching parameters in a cascade manner, leading to the best prediction performance for etching profile prediction.

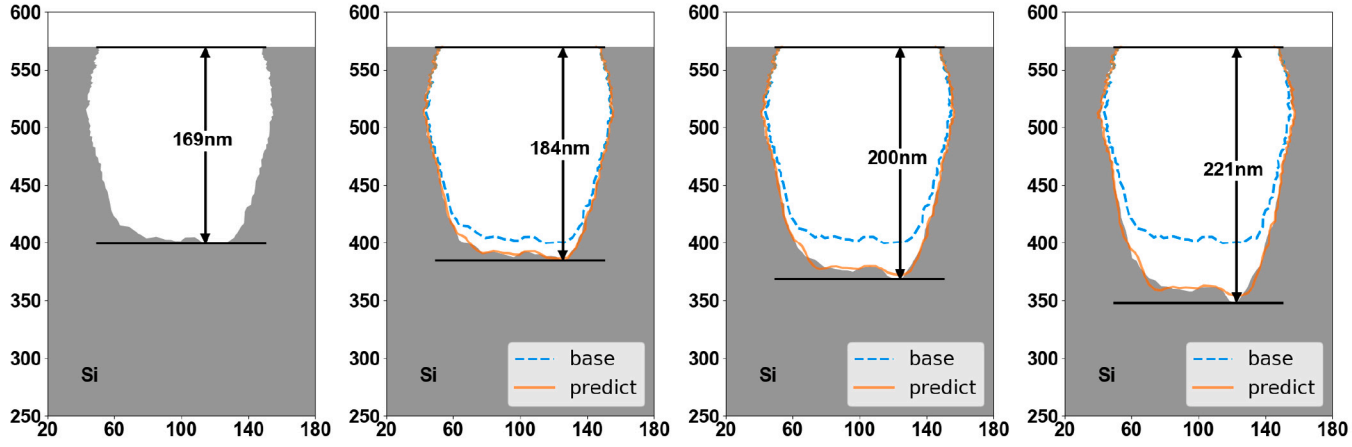
### 4.3. Prediction examples

Fig. 6 presents two groups of examples of multi-step prediction results for etching profiles with CRNN. Figs. 6(b), 6(c), 6(d) are prediction results of 1, 3 and 10 time steps with the same base etch profile (Fig. 6(a)), respectively. Figs. 6(f), 6(g), 6(h) are prediction results of 1, 3 and 10 time steps with the same base etch profile (Fig. 6(e)), respectively. The dashed lines in all the figures are its base profile, the orange lines are predicted profile of CRNN, while the actual etching profile could be figure out by the edge of Si. It is easy to find that, no matter how many prediction steps are involved, the difference between the predicted profiles and the ground truth is negligible. There are only a few gaps between the predicted profile and its ground truth. The prediction profiles not only fit the overall profile, but also capture many local details on the profile. The proposed CRNN did achieve accurate etching profile prediction.

Comparing the two groups of figures, it can be observed that as the number of prediction steps increases, the error gradually accumulates, distinguishable gaps between the predicted profile and the true profile increases slightly.




(a) base profile (b) 3 steps prediction (c) 6 steps prediction (d) 10 steps prediction




(e) base profile (f) 3 steps prediction (g) 6 steps prediction (h) 10 steps prediction

**Fig. 6.** Predicted etching profiles.

**Table 4**
Etching depth MAE comparison of different approaches (nm).

| Model | | 1 step (0.1 s) | 3 steps (0.3 s) | 5 steps (0.5 s) | 7 steps (0.7 s) | 10 steps (1.0 s) |
|---|---|---|---|---|---|---|
| DNN | MAE | 0.6939 | 1.0136 | 1.2356 | 1.4435 | 1.7329 |
| | Ratio2DNN | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| TRANS (Myung et al., 2021) | MAE | 0.6819 | 1.0258 | 1.2836 | 1.5229 | 1.8636 |
| | Ratio2DNN | 0.9827 | 1.0120 | 1.0388 | 1.0550 | 1.0754 |
| RNN (Jeong et al., 2021) | MAE | 0.6755 | 0.9670 | 1.1538 | 1.3132 | 1.5190 |
| | Ratio2DNN | 0.9735 | 0.9540 | 0.9338 | 0.9097 | 0.8766 |
| CNN | MAE | 0.6796 | 0.9749 | 1.1604 | 1.3095 | 1.4819 |
| | Ratio2DNN | 0.9794 | 0.9618 | 0.9391 | 0.9072 | 0.8552 |
| CCNN | MAE | 0.6763 | 0.9587 | 1.1256 | 1.2614 | 1.4225 |
| | Ratio2DNN | 0.9746 | 0.9458 | 0.9110 | 0.8738 | 0.8209 |
| CRNN | MAE | **0.6752** | **0.9544** | **1.1173** | **1.2407** | **1.3845** |
| | Ratio2DNN | **0.9731** | **0.9416** | **0.9043** | **0.8595** | **0.7989** |

**Table 5**
Ablation study of CRNN.

| Model | Factors | | MAE(nm) | | | | |
|---|---|---|---|---|---|---|---|
| | RNN | Number of CLs | 1 step (0.1 s) | 3 steps (0.3 s) | 5 steps (0.5 s) | 7 steps (0.7 s) | 10 steps (1.0 s) |
| DNN | | | 0.6939 | 1.0136 | 1.2356 | 1.4435 | 1.7329 |
| RNN | ✓ | 1 | 0.6755 | 0.9670 | 1.1538 | 1.3132 | 1.5190 |
| CRNN2 | ✓ | 2 | 0.6755 | 0.9635 | 1.1383 | 1.2797 | 1.4517 |
| CRNN3 | ✓ | 3 | **0.6752** | **0.9544** | **1.1173** | **1.2407** | **1.3845** |
| CRNN4 | ✓ | 4 | 0.6774 | 0.9713 | 1.1456 | 1.2834 | 1.4470 |

**Table 6**
Runtime comparison.

| Simulation | Monte Carlo | CRNN(CPU) | CRNN(GTX960M) | CRNN(A100) |
|---|---|---|---|---|
| Time (s) | 946.9643 | 0.2111 | 0.1121 | 0.0440 |
| Speedup | – | 4486× | 8447× | 21522× |

### 4.4. Ablation study

Ablation study is conducted on the same dataset to check the effectiveness of each component technique of CRNN, whose results are shown in Table 5. There are five models, in which all carefully designed technical components of CRNN are enabled one by one. MAE of 5 different time steps are listed.

DNN is to predict the etch depth with DNN based on latest one-step profile, etching depth and etching parameters. RNN is to predict the etch depth with simple RNN based on all the historical profiles and concatenated etching parameters. Compare the results of RNN and DNN for 10 time steps prediction, MAE of RNN reduced from 1.7329 nm to 1.5190 nm, which validates the advantage of RNN.

Cascading combination is the last customized technology in CRNN, the number of combination layers is a key issue. We have tried four scenarios, namely RNN with one combination layer, CRNN2 with two combination layers, CRNN3 with three combination layers and CRNN4 with four combination layers. It is easy to find adding a combination layer can significantly reduce MAE. However, more combination layers may not necessarily result in better results. The optimal number of combination layers is three, CRNN3 gives the least MAE.

Overall, each technical component of CRNN structure can improve prediction performance, and the optimal number of combination layers is 3. The CRNN mentioned in other parts of the paper has 3 combination layers.

### 4.5. Error distribution

In order to further analyze the causes of prediction errors, we conducted a more in-depth statistical analysis of the prediction errors. Fig. 7 shows the average MAE over different angles and time steps.

Fig. 7(a) presents the average MAE from different angles. We can find that the error in the side and center parts of the etching area is low. Specifically, the side parts mean the section less than 30 degrees and greater than 150 degrees, and the central part is between 85 degrees and 95 degrees. The average MAE in the side and central part is less than 0.6 nm. The reason for the low error in the side part is that there are few changes, and only a small number of particles will collide or react with the side region. The central region is the most concentrated region of particle flow, which usually kept flat in the etching process with low reflection effects, result in stable etching intensity and strong predictability. The morphology of other areas undergoes significant changes during the etching process, resulting in significant differences in etching depth at each step and poor predictability.

Fig. 7(b) displays the average MAE at different time steps. In the initial stage of etching, due to the limited available information, the prediction error is large. As the etching process runs, more and more information is involved, and the prediction error gradually decreases. The average MAE decreases from over 1 nm in 1 step to 0.5 nm in 20 steps. However, the addition of more information cannot further reduce the prediction error, the mean MAE keep steady at around 0.5 nm.

**Table 7**
Fine-tuning results.

| | Base NN | Finetuned NN |
|---|---|---|
| MAE (nm) | 14.1813 | 7.9028 |

**Table 8**
Fine-tuning results.

| | 1 step | 2 steps | 3 steps |
|---|---|---|---|
| MAE (nm) | 9.6038 | 6.7288 | 7.3757 |

### 4.6. Runtime comparison

Table 6 lists the runtime of conventional Monte Carlo model and CRNN, in term of simulating a etch process of 5 s (50 steps). We can see that the trained CRNN takes 0.2111 s to inference a etching process of 5 s on CPU (Intel(R) Xeon(R) 6246, 3.30 GHz) while conventional Monte Carlo model need more than 946 s for the same simulation. If calculated on GPUs (Nvidia GTX960M and A100), the acceleration can reach 8,447 and 21,000 times, respectively. CRNN based simulation models can accelerate up to $21,000\times$, with an average error of less than 0.7 nm.

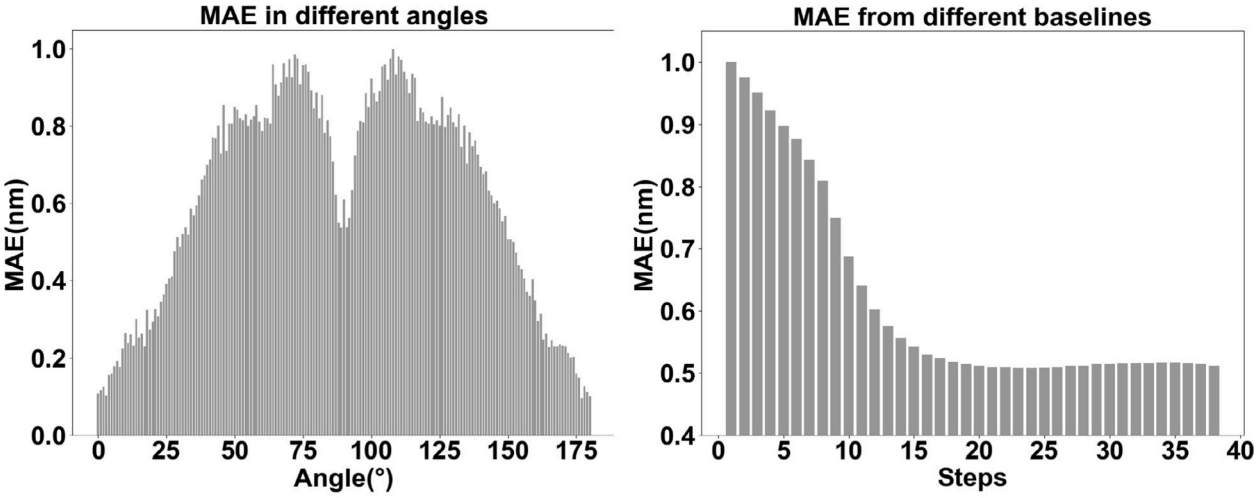
### 4.7. Fine-tuning method

As mentioned in Section 3.4, it is impossible to record the profiles during the etching process. We validated the practicality of this scheme by etching experimental data at different times using the same recipe. Etching on 5 different pitches yields 5 different proxy profiles. The 5 profiles take turns to be test data, while the left 4 for network fine-tuning. Fig. 8 shows the prediction results of 2 profiles with different CDs. The predicted profiles are basically consistent with the actual etching profiles, which shows the effectiveness of the model after fine-tuning. Table 7 lists the average MAE before and after fine-tuning, we can find that the MAE decreases significantly after fine-tuning. Table 8 lists the average MAE at different time steps. Note that the prediction error in the first step is greater than that in the other steps, which is consistent with the results shown in Figs. 7(b) and 8.

In this section, we conducted extensive testing of the proposed CRNN algorithm. We used both simulated and real data, including quantitative comparisons and qualitative observations. The comparison results with state-of-the-art algorithms fully demonstrate the effectiveness of the novel algorithm.

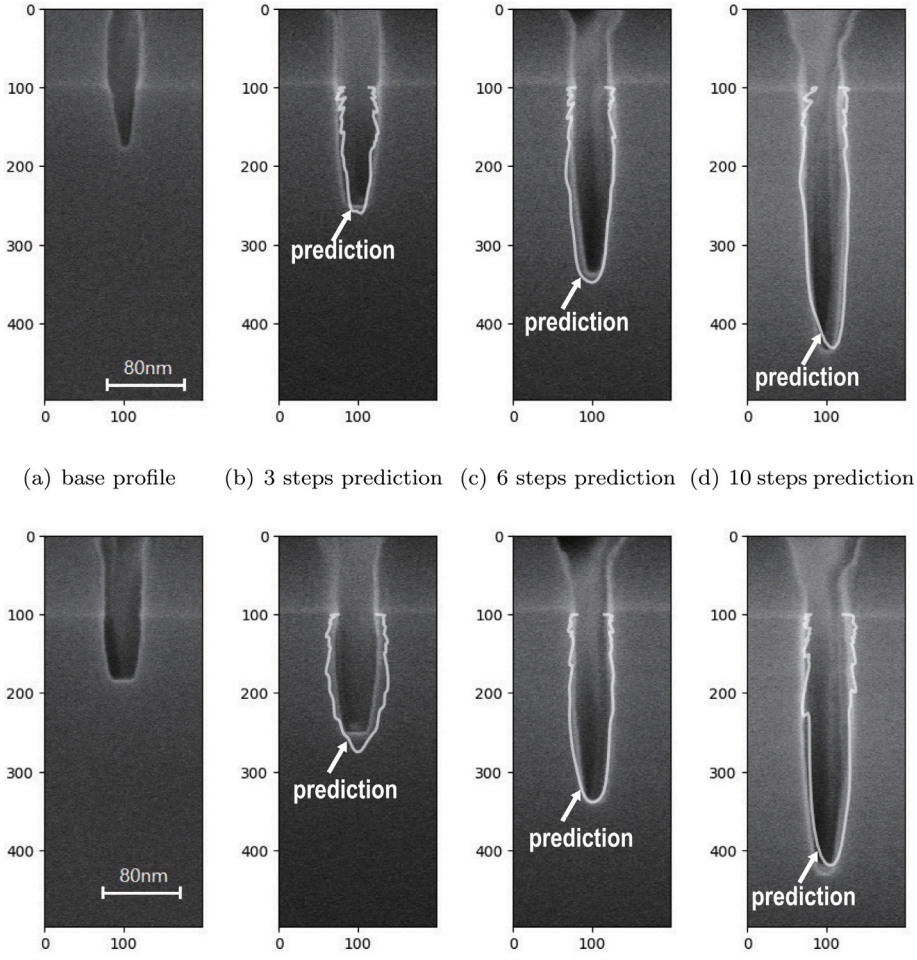
## 5. Conclusion and future works

In this paper, CRNN is proposed for predicting etching profiles. Initially, the etching profile data is converted into a more efficient polar coordinate representation and modeled by LSTM. Etching parameters are combined into the network through a newly proposed cascade combination layers, which leads to comprehensive integration of the profile features and etching parameters. Experimental results conducted on a dataset consisting of 200 etching processes demonstrate the effectiveness of our proposed method. In comparison to traditional physical etching simulation methods, CRNN exhibits a 21,000 times




(a) Mean MAE over different angles (b) Mean MAE over different baselines

**Fig. 7.** Error distribution.




(a) base profile (b) 3 steps prediction (c) 6 steps prediction (d) 10 steps prediction

(e) base profile (f) 3 steps prediction (g) 6 steps prediction (h) 10 steps prediction

**Fig. 8.** Predicted etching profiles.

acceleration in 1-step prediction with an average error of less than 0.7 nm. Furthermore, compared to existing RNN-based etching predictions, the MAE is reduced from 1.5190 nm to 1.3845 nm for a 10-step prediction.

Furthermore, the model, initially trained on a large simulation dataset, undergoes fine-tuning and validation on a smaller experimental dataset. This process illustrates the potential applicability of data-driven AI models in real-world semiconductor manufacturing scenarios. To our knowledge, this is the first purely data-driven etching profile prediction algorithm validated on both simulated and real-world datasets.

However, the proposed model is trained on simulation data, which can only predict profiles for uniformly sampled profiles. But most real-world data is often non-uniformly sampled. In future works, we will focus on improvement from two aspects. On one hand, enhancing the etching process to accumulate more available experimental data; on the other hand, establishing more flexible models that can be applied to real-world data with non-uniformly sampling.

Moreover, the proposed cascaded combination strategy is not only applicable to feature combination in etching modeling but can also be extended to more applications involving heterogeneous feature combinations. Its suitability for multi-modal data fusion is a topic worthy of further exploration.

## CRediT authorship contribution statement

**Zhenjie Yao:** Writing – original draft, Visualization, Software, Methodology. **Ziyi Hu:** Writing – original draft, Visualization, Data curation. **Panpan Lai:** Validation, Data curation. **Fengling Qin:** Writing – review & editing, Conceptualization. **Wenrui Wang:** Software, Data curation. **Zhicheng Wu:** Writing – review & editing, Investigation. **Lingfei Wang:** Writing – review & editing, Methodology. **Hua Shao:** Data curation. **Yongfu Li:** Writing – review & editing, Supervision, Methodology. **Zhiqiang Li:** Supervision, Resources, Funding acquisition. **Zhongming Liu:** Resources. **Junjie Li:** Supervision, Data curation. **Rui Chen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Ling Li:** Supervision, Resources, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References


Abdollahi-Alibeik, S., McVittie, J.P., Saraswat, K.C., Sukharev, V., Schoenborn, P., 1999. Analytical modeling of silicon etch process in high density plasma. J. Vac. Sci. Technol. A 17 (5), 2485–2491. http://dx.doi.org/10.1116/1.581986.

Badue, C., Guidolini, R., Carneiro, R.V., Azevedo, P., Cardoso, V.B., Forechi, A., Jesus, L., Berriel, R., Paixão, T.M., Mutz, F., de Paula Veronese, L., Oliveira-Santos, T., De Souza, A.F., 2021. Self-driving cars: A survey. Expert Syst. Appl. 165, 113816. http://dx.doi.org/10.1016/j.eswa.2020.113816.

Bao, H., Chen, L., 2020. A CNN-based CMP planarization model considering LDE effect. IEEE Trans. Compon. Packag. Manuf. Technol. 10 (4), 723–729. http://dx.doi.org/10.1109/TCPMT.2020.2979472.

Chen, G., Chen, W., Sun, Q., Ma, Y., Yang, H., Yu, B., 2022. DAMO: Deep agile mask optimization for full-chip scale. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 41 (9), 3118–3131. http://dx.doi.org/10.1109/TCAD.2021.3116511.

Chen, R., Hu, H., Li, X., Chen, Y., Xiaojing, S., Dong, L., Qu, L., Li, C., Yan, J., Wei, Y., 2020. ETCH model based on machine learning. In: 2020 China Semiconductor Technology International Conference. CSTIC, IEEE, pp. 1–4. http://dx.doi.org/10.1109/CSTIC49141.2020.9282462.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929.

Fukumoto, H., Eriguchi, K., Ono, K., 2009. Effects of mask pattern geometry on plasma etching profiles. Japan. J. Appl. Phys. 48 (9R), 096001. http://dx.doi.org/10.1143/JJAP.48.096001.

Guo, W., Sawin, H.H., 2009. Review of profile and roughening simulation in microelectronics plasma etching. J. Phys. D: Appl. Phys. 42 (19), 194014. http://dx.doi.org/10.1088/0022-3727/42/19/194014.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2980–2988. http://dx.doi.org/10.1109/ICCV.2017.322.

Hoekstra, R.J., Kushner, M.J., 1998. Comparison of two-dimensional and three-dimensional models for profile simulation of poly-Si etching of finite length trenches. J. Vac. Sci. Technol. A 16 (6), 3274–3280. http://dx.doi.org/10.1116/1.581533.

Hoekstra, R.J., Kushner, M.J., Sukharev, V., Schoenborn, P., 1998. Microtrenching resulting from specular reflection during chlorine etching of silicon. J. Vac. Sci. Technol. B 16 (4), 2102–2104. http://dx.doi.org/10.1116/1.590135.

Hu, Z., Shao, H., Li, J., Lai, P., Wang, W., Li, C., Yan, Q., He, X., Li, J., Yang, T., Chen, R., Wei, Y., 2023. Modeling of microtrenching and bowing effects in nanoscale Si inductively coupled plasma etching process. J. Vac. Sci. Technol. A 41 (6), 063113. http://dx.doi.org/10.1116/6.0003032.

Huo, Z., Cheng, W., Yang, S., 2022. Unleash scaling potential of 3D NAND with innovative Xtacking® architecture. In: 2022 IEEE Symposium on VLSI Technology and Circuits. VLSI Technology and Circuits, IEEE, pp. 254–255. http://dx.doi.org/10.1109/VLSITechnologyandCir46769.2022.9830285.

Jeong, C., Myung, S., Huh, I., Choi, B., Kim, J., Jang, H., Lee, H., Park, D., Lee, K., Jang, W., Ryu, J., Cha, M.-H., Choe, J.M., Shim, M., Kim, D.S., 2021. Bridging TCAD and AI: Its application to semiconductor design. IEEE Trans. Electron Devices 68 (11), 5364–5371. http://dx.doi.org/10.1109/TED.2021.3093844.

Kanarik, K.J., Osowiecki, W.T., Lu, Y., Talukder, D., Roschewsky, N., Park, S.N., Kamon, M., Fried, D.M., Gottscho, R.A., 2023. Human–machine collaboration for improving semiconductor process development. Nature 1–5. http://dx.doi.org/10.1038/s41586-023-05773-7.

Kim, S., Kim, M., Ryu, D., Lee, K., Kim, S., Lee, J., Lee, R., Kim, S., Lee, J.-H., Park, B.-G., 2020. Investigation of electrical characteristic behavior induced by channel-release process in stacked nanosheet gate-all-around MOSFETs. IEEE Trans. Electron Devices 67 (6), 2648–2652. http://dx.doi.org/10.1109/TED.2020.2989416.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv:1412.6980.

Kokkoris, G., Tserepi, A., Boudouvis, A., Gogolides, E., 2004. Simulation of SiO 2 and Si feature etching for microelectronics and microelectromechanical systems fabrication: A combined simulator coupling modules of surface etching, local flux calculation, and profile evolution. J. Vac. Sci. Technol. A 22 (4), 1896–1902. http://dx.doi.org/10.1116/1.1738660.

Kuboi, N., Fukasawa, M., Tatsumi, T., 2016. Advanced simulation technology for etching process design for CMOS device applications. Japan. J. Appl. Phys. 55 (7S2), 07LA02. http://dx.doi.org/10.7567/JJAP.55.07LA02.

Mahorowala, A.P., Sawin, H.H., 2002. Etching of polysilicon in inductively coupled Cl 2 and HBr discharges. II. Simulation of profile evolution using cellular representation of feature composition and Monte Carlo computation of flux and surface kinetics. J. Vac. Sci. Technol. B 20 (3), 1064–1076. http://dx.doi.org/10.1116/1.1481867.

Mahorowala, A.P., Sawin, H.H., Jones, R., Labun, A.H., 2002. Etching of polysilicon in inductively coupled Cl 2 and HBr discharges. I. Experimental characterization of polysilicon profiles. J. Vac. Sci. Technol. B 20 (3), 1055–1063. http://dx.doi.org/10.1116/1.1481866.

Myung, S., Jang, H., Choi, B., Ryu, J., Kim, H., Park, S.W., Jeong, C., Kim, D.S., 2021. A novel approach for semiconductor etching process with inductive biases. arXiv:2104.02468.

Rahimifar, M., Jahanirad, H., Fathi, M., 2024. Deep transfer learning approach for digital circuits vulnerability analysis. Expert Syst. Appl. 237, 121757. http://dx.doi.org/10.1016/j.eswa.2023.121757.

Rezvankhah, M.A., Shayan, M., Merati, A.R., Pahlevani, M., 2013. Step flow model in continuous cellular automata method for simulation of anisotropic etching of silicon. J. Micro/Nanolithogr. MEMS MOEMS 12 (2), 023004. http://dx.doi.org/10.1117/1.JMM.12.2.023004.

Savari, M.A., Jahanirad, H., 2020. NN-SSTA: A deep neural network approach for statistical static timing analysis. Expert Syst. Appl. 149, 113309. http://dx.doi.org/10.1016/j.eswa.2020.113309.

Sethian, J.A., et al., 1999. Level Set Methods and Fast Marching Methods, vol. 98, Cambridge University Press.

Shih, S.-Y., Sun, F.-K., Lee, H.-y., 2019. Temporal pattern attention for multivariate time series forecasting. Mach. Learn. 108, 1421–1441. http://dx.doi.org/10.1007/s10994-019-05815-0.

Song, X., Peng, Z., Song, S., Stojanovic, V., 2024. Anti-disturbance state estimation for PDT-switched RDNNs utilizing time-sampling and space-splitting measurements. Commun. Nonlinear Sci. Numer. Simul. 132, 107945. http://dx.doi.org/10.1016/j.cnsns.2024.107945.

Song, X., Sun, P., Song, S., Stojanovic, V., 2023a. Quantized neural adaptive finite-time preassigned performance control for interconnected nonlinear systems. Neural Comput. Appl. 35 (21), 15429–15446. http://dx.doi.org/10.1007/s00521-023-08361-y.

Song, X., Wu, N., Song, S., Zhang, Y., Stojanovic, V., 2023b. Bipartite synchronization for cooperative-competitive neural networks with reaction–diffusion terms via dual event-triggered mechanism. Neurocomputing 550, 126498. http://dx.doi.org/10.1016/j.neucom.2023.126498.

Subramanian, S., Hosseini, M., Chiarella, T., Sarkar, S., Schuddinck, P., Chan, B.T., Radisic, D., Mannaert, G., Hikavyy, A., Rosseel, E., Sebaai, F., Peter, A., Hopf, T., Morin, P., Wang, S., Devriendt, K., Batuk, D., Martinez, G.T., Veloso, A., Litta, E.D., Baudot, S., Siew, Y.K., Zhou, X., Briggs, B., Capogreco, E., Hung, J., Koret, R., Spessot, A., Ryckaert, J., Demuynck, S., Horiguchi, N., Boemmels, J., 2020. First monolithic integration of 3D complementary FET (CFET) on 300mm wafers. In: 2020 IEEE Symposium on VLSI Technology. pp. 1–2. http://dx.doi.org/10.1109/VLSITechnology18217.2020.9265073.

Than, O., Büttgenbach, S., 1994. Simulation of anisotropic chemical etching of crystalline silicon using a cellular automata model. Sensors Actuators A 45 (1), 85–89. http://dx.doi.org/10.1016/0924-4247(94)00820-5.

Vyvoda, M., Li, M., Graves, D., 1999. Hardmask charging during Cl 2 plasma etching of silicon. J. Vac. Sci. Technol. A 17 (6), 3293–3307. http://dx.doi.org/10.1116/1.582056.

Wang, Q., De Chen, Y., Huang, J., Joseph, E., 2020. A study of wiggling AA modeling and its impact on the device performance in advanced DRAM. In: 2020 International Conference on Simulation of Semiconductor Processes and Devices. SISPAD, pp. 101–104. http://dx.doi.org/10.23919/SISPAD49475.2020.9241640.

Winters, H.F., Plumb, I.C., 1991. Etching reactions for silicon with F atoms: Product distributions and ion enhancement mechanisms. J. Vac. Sci. Technol. B 9 (2), 197–207. http://dx.doi.org/10.1116/1.585593.

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C., 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 753–763. http://dx.doi.org/10.1145/3394486.3403118.

Xiao, T., Ni, D., 2021. Multiscale modeling and recurrent neural network based optimization of a plasma etch process. Processes 9 (1), 151. http://dx.doi.org/10.3390/pr9010151.

Yang, Z., Chen, L., Zhang, H., Yao, Z., 2021. Residual connection based TPA-LSTM networks for cluster node CPU load prediction. In: 2021 IEEE International Conference on Big Data. Big Data, IEEE, pp. 5311–5316. http://dx.doi.org/10.1109/BigData52589.2021.9671699.

Yang, H., Li, Z., Sastry, K., Mukhopadhyay, S., Kilgard, M., Anandkumar, A., Khailany, B., Singh, V., Ren, H., 2022. Generic lithography modeling with dual-band optics-inspired neural networks. In: Proceedings of the 59th ACM/IEEE Design Automation Conference. DAC '22, Association for Computing Machinery, New York, NY, USA, pp. 973–978. http://dx.doi.org/10.1145/3489517.3530580.

Yao, Z.-J., Bi, J., Chen, Y.-X., 2018. Applying deep learning to individual and community health monitoring data: A survey. Int. J. Autom. Comput. 15 (6), 643–655. http://dx.doi.org/10.1007/s11633-018-1136-9.

Yao, Z., Xu, Q., Chen, Y., Tu, Y., Zhang, H., Chen, Y., 2021. Internet traffic forecasting using temporal-topological graph convolutional networks. In: 2021 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8. http://dx.doi.org/10.1109/IJCNN52387.2021.9534096.

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C., 2021. A transformer-based framework for multivariate time series representation learning. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, pp. 2114–2124. http://dx.doi.org/10.1145/3447548.3467401.