Journal of Applied Physics ARTICLE pubs.aip.org/aip/jap

Deep learning-based characterization of ion implantation parameters for photo modulated optical reflectance

Cite as: J. Appl. Phys. **136**, 063102 (2024); doi: 10.1063/5.0210816
Submitted: 27 March 2024 · Accepted: 19 July 2024 ·
Published Online: 9 August 2024

Xuesong Wang, Lijun Zhang, Yong Sun, Jin Min, Zhongyu Wang, Shiyuan Liu, Xiuguo Chen, and Zirong Tang[a)]

AFFILIATIONS

School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China

[a)]**Author to whom correspondence should be addressed:** zirong@hust.edu.cn

ABSTRACT

Photo modulated optical reflectance (PMOR) is an ideal ultra-shallow junction area metrology technique for measurement of transistor dopant distribution in integrated circuit fabrication, and the characterization of process parameters such as implant energy, implant angle, and implant dose has a significant impact on the accuracy of the ion implantation process. This study utilized deep learning to analyze various process parameters concurrently and assessed its performance on boron-doped silicon samples, the data were obtained from the power curves measured from Carrier Illumination (CI) experiments in PMOR, a deep learning model with multi-task learning architecture was developed and trained to characterize multiple process parameters, and the PMOR model incorporating a multi-task learning architecture for process parameters demonstrated superior performance in terms of accuracy and speed of characterization. The analyses indicated that applying deep learning methods to CI metrology in PMOR technology is feasible. In particular, compared with the conventional carrier irradiation technique, the ability to obtain the implantation dose and doping profile along with other process parameters such as implantation energy, implantation angle, and implantation type can better assist in the accurate realization of the ion implantation process with acceptable accuracy and time cost.

I. INTRODUCTION

Photo modulated optical reflectance[1,2] (PMOR) is a pump–probe technology where the light intensity serves as the external modulation variable. The strength of the pump laser is varied at a specific frequency, which alters the sample's reflectivity by interacting with light and matter. The probing laser uses a locking-based detection technique to measure the change in reflectance at the same modulation frequency or a higher harmonic. With a very high sensitivity level, measuring reflectivity changes down to 10^{-6}, PMOR initially showed the ability to determine various transport parameters (e.g., thermal conductivity, carrier diffusivity, and carrier lifetime) for bulk materials and thin-film metals, semiconductors, or superconductors,[3–7] with engineering realizations focusing on the monitoring of ion implantation by statistical process control (SPC).[8–16]

Carrier Illumination (CI) is a special implementation of PMOR, described as "devices and methods for determining the distribution of active dopants in semiconductor wafers." The pump laser operates in a quasi-static state (2 kHz) at a fixed pump modulation frequency and is characterized by its pump power (or irradiance) varying over a range. The relationship between the differential probe reflectance and pump power, called as the "power curve,"[17] provides insights into the deep active doping distribution; one of the main challenges in the manufacturing of CMOS devices is the need for increasingly tight control of implantation parameters such as dose, energy, and tilt angle for a given implant, where the dose determines the final concentration in the ion implantation process and where the energy determines the range (i.e., depth) of the ions, and where different implantation angles affect the implanted samples (furrowing and shadowing effects). Therefore, to ensure the accuracy of the ion implantation process, it is

necessary to characterize the process parameters, such as dose and energy, simultaneously. Based on CI theory, methods for extracting wafer implantation dose and active doping profiles from power curves are well established. However, extracting process parameters such as energy, angle, and species is still challenging, so an algorithm is needed to extract the process parameters of ion implantation quickly and accurately.

Deep learning algorithms are a machine learning type trained to recognize complex mappings between inputs and outputs, making them suitable for identifying ion implantation process parameters from power profiles measured by PMOR technology and predicting specific ion implantation parameters based on individual wafer power profiles. Deep learning models have been successfully employed in numerous real-world, large-scale applications due to their robust feature extraction capabilities. Many large recommender systems[18] have used multi-task learning with deep neural network (DNN) models.[19] Loew and Bradley[20] employed deep learning (artificial neural network—ANN) to estimate parameters for patterns resulting from ion bombardment on solid surface development. These studies are instructive for multi-process parameter extraction in CI signals. Migrating deep learning-based approaches to CI metrology is problematic due to the diversity of samples, accuracy requirements, and time consumption in IC manufacture. This study seeks to build a deep learning-based system to predict ion implantation process parameters using measurement data from CI signals. A multi-task learning architecture for process parameters (PPMTL)-based deep learning model was created and utilized. Carrier Illumination experiments were conducted to provide a dataset for training, validation, and testing the model. A logical training approach was devised to instruct this model. The practicality of the deep learning strategy is assessed by analyzing the recognition results in the presence of noise and considering the time consumed.

II. PHOTO MODULATED OPTICAL REFLECTANCE (PMOR) EXPERIMENT

The CI experimental setup in this study is built using pump–probe technology, illustrated in Fig. 1; the pump and probe lasers are semiconductor lasers in the UV and red wavelength bands, respectively, with a beam diameter of 1.2 mm, the pump light is used to excite plasma and heat waves in the sample, and the detector light is used to detect the change of the surface properties of the sample. The pump and probe beams are merged through dichroic mirrors to form the superimposed beam, which is then turned into parallel light through a collimator. Through the collimator into parallel light, the superposition of the beam, through the polarizer, polarization beam splitter and beam sampling plate. The sampling plate directs part of the beam to the detector (silicon PIN photodiode), and the rest of the beam passes through the beam quarter-wave plate and the objective lens, which focuses the beam on the wafer. The cross-polarization method is used in this experimental setup, where the incident and reflected superimposed light passes through the quarter-wave plate twice before and after passing through the objective lens. The superimposed light reflected to the polarizing beam splitter is in orthogonal polarization to the incident light passing through the polarizing beam splitter, and the superimposed beam, which carries the information about the sample, is reflected to the detector as it passes through the polarizing beam splitter.

During the measurements, the modulation amplitude of the pump power was increased from 0 to 22 mW, typically in each of the 22 steps (Figs. 2 and 3), and the signals (dc, first and second harmonics) were recorded at each step. The first harmonic components at each delay time were recorded sequentially using a computer and normalized by dividing by dc to obtain $\Delta R/R$ vs pump laser power (power curve).

In order to obtain the required measurement signals for the study, B was implanted into 300 mm P-type single-sided polished silicon wafers, and CI experiments were performed on them in this study. These samples consisted of silicon wafers implanted with doses ranging from 1×10^{13} to 5×10^{15} ion/cm^2, energies ranging from 2 to 50 keV, and angles ranging from -10° to 10°, with a supplier-guaranteed implant angle accuracy of $\pm 0.5^\circ$. Figure 4 shows the power curves of silicon wafers derived from carrier illumination experiments performed at different ion implantation process parameters. These curves describe the power response of

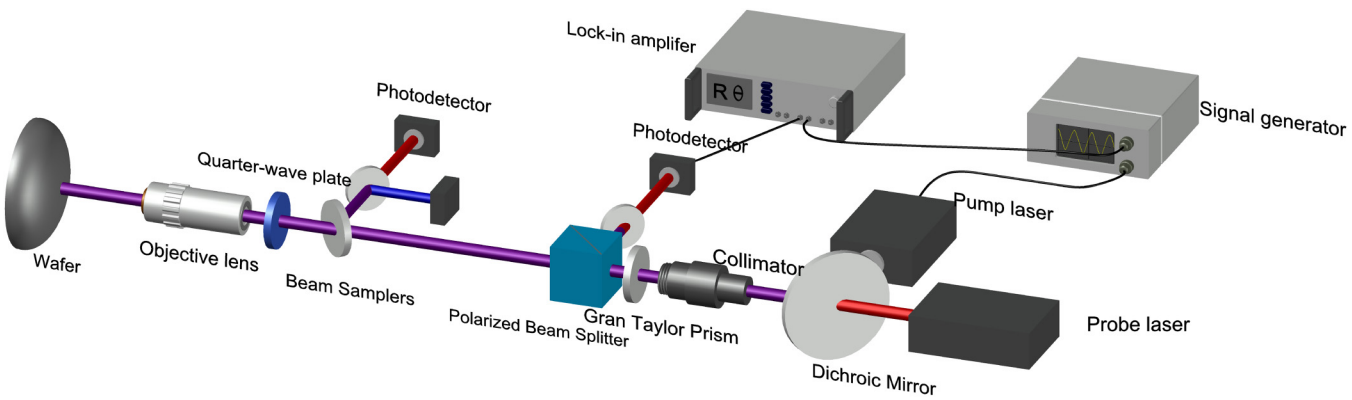

FIG. 1. CI experimental setup.

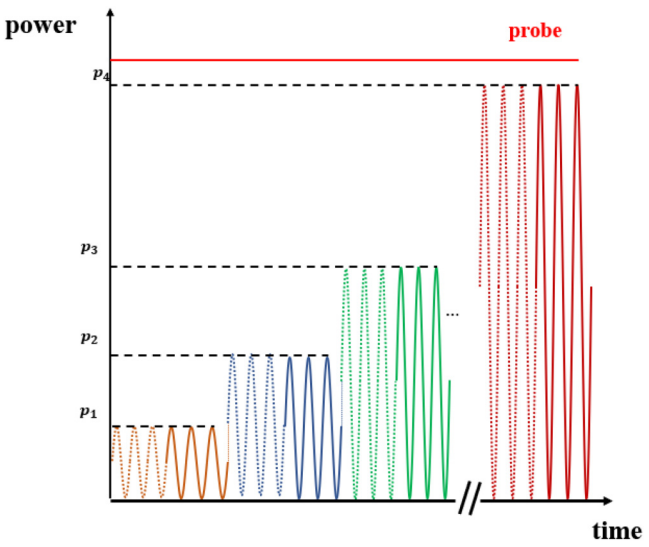

FIG. 2. Time-dependent curve of pump and probe lasers.

silicon wafers to different implantation angles, energies, and doses. At the laser wavelength of this experimental setup, all samples absorbed and converted enough light to produce relatively strong CI signals. From the measured power curves above, a power bulge near 12 mW and saturation at about 20 mW can be observed. These phenomena are mainly due to the slowing down of the CI signal change when the pump laser power reaches 12 mW, as well as the decrease in the linearity of the power regulation as the semiconductor laser approaches its maximum output power (about

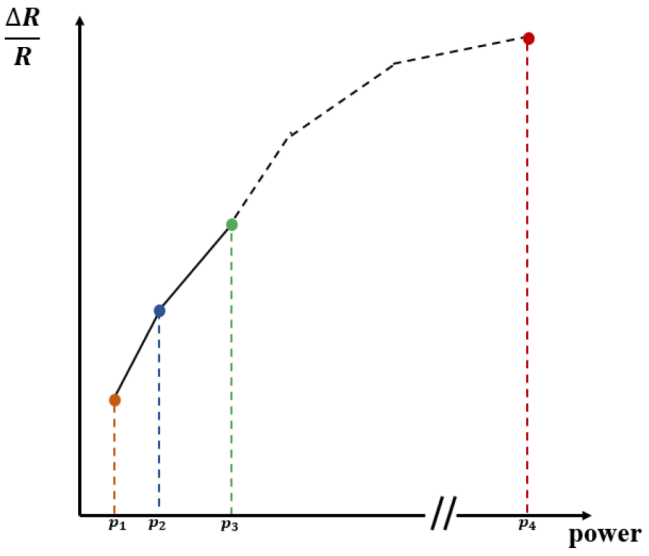

FIG. 3. Principle of CI power curve measurement in PMOR technology.

22 mW). Despite these nonlinear features, the deep learning model could identify and exploit subtle differences in the power profile through the training process, thereby extracting multiple independent ion implantation process parameters and ensuring the accuracy and reliability of the analysis results. In general, the traditional extraction of ion implantation process parameters in PMOR measurement uses the controlled variable method, i.e., controlling a single variable to test a group of wafers and distinguishing them according to the strength of the signal value, which is not only time-consuming and laborious but also requires a group of wafers for each process parameter, making it impossible to extract multiple process parameters for a single wafer at the same time. The traditional method is also easily affected by noise signals, and the accuracy cannot be guaranteed. Therefore, this paper proposes a deep learning-based process parameter extraction method, which can quickly and accurately extract multiple process parameters of ion implantation from the power curve simultaneously.

Aside from the fundamental construction mentioned, the experimental setup incorporates a lock-in amplification approach and a focusing module (not visible in the optical channel) to achieve a relatively high signal-to-noise ratio. The experimental setup contains two detector modules, each consisting of two photodetectors and a dichroic mirror, which separates the pump light from the detector light and a filter to filter stray light. A signal generator modulates the pump beam at 2 kHz. The signal generator modifies the pump beam at a frequency of 2 kHz. The lock-in amplifier isolates and boosts the 2 kHz signal from the photodetector output that corresponds to the modulation frequency.

III. DEEP LEARNING MODELS FOR CHARACTERIZATION OF MULTIPLE PROCESS PARAMETERS

The goal of the research is to create a unified model capable of learning numerous objectives and tasks simultaneously. The power curve measured by the CI is essentially a sequential sequence, so RNN models commonly used to process sequential data can be employed. We achieve multitask learning by appending multiple fully connected layers on top of the RNN output, with each fully connected layer corresponding to a specific multitask. Multitask learning models have been demonstrated to improve predictions for all tasks through the use of regularization and transfer learning.[21] Nevertheless, commonly used multitask models like DNN and RNN are often affected by the inter-task relationships,[22,23] leading to conflicts that may hinder predictions, especially when model parameters are widely distributed across tasks. This study utilizes the PPMTL model [24], which explicitly represents task connections. The model adjusts the parameterization by utilizing modulation and gating networks to balance shared information and task-specific information.

The PPMTL model has been shown to perform well under different task relationships;[24] the model takes as input a one-dimensional array created from the power curves obtained by the CI, and the input dimension is set to 22 since each power curve measures 22 points, in this study, the extracted parameters for the ion implantation are the implant energy, implant dose, and implant angle. Therefore, the model output dimension should be 3. It is worth mentioning that the method of extracting the

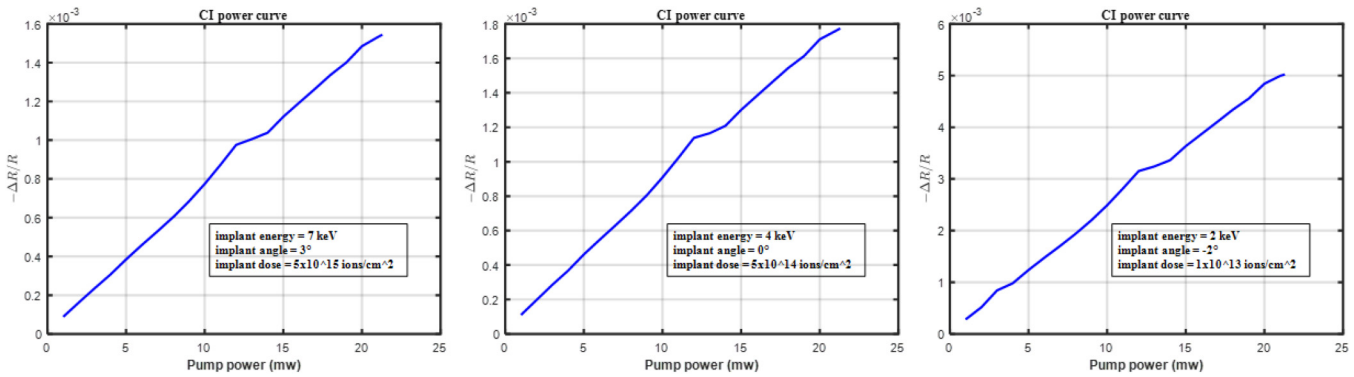

FIG. 4. Experimentally measured power profiles of silicon wafers with different ion implantation process parameters.

implantation dose from the power curve has been very well developed, but it is still included in this study. First, it can be cross-referenced with the traditional method. Furthermore, in future studies of simultaneous extraction of multiple ion implantation process parameters, it is essential to omit the traditional dose extraction step to ensure the integrity of the extraction process. While it is theoretically possible to jointly investigate the implant ion species, the absence of samples featuring ion implants other than boron precludes their inclusion in this study.

A. PPMTL-based model

The PPMTL architecture is mostly based on the widely known Collaborative Bottom multi-task deep neural network architecture.[25] The Collaborative Bottom model structure has several bottom layers that are shared among all tasks after the input layer. Then, each task has a separate network "tower" above the bottom representation. The PPMTL approach does not utilize a single bottom network for all tasks, as depicted in Fig. 5. Instead, it

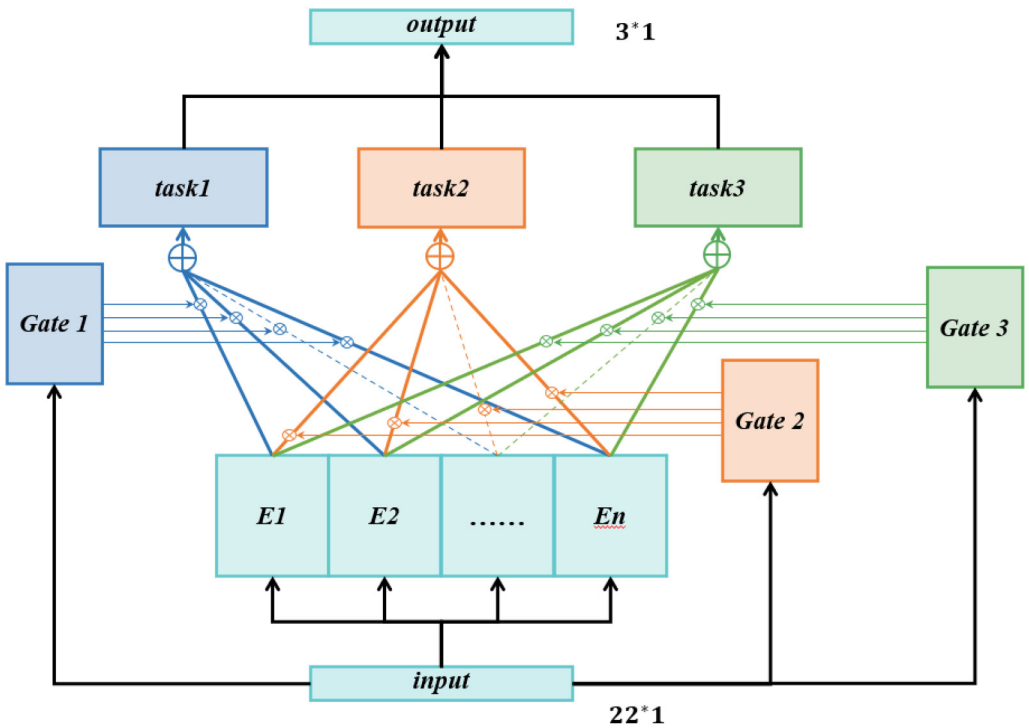

FIG. 5. PPMTL model architecture.

employs a collection of bottom networks, each referred to as an expert. Every expert in our document is a feed-forward network. Next, we will establish a gated network for each task. The gated network uses input characteristics and output softmax gates to integrate experts with different weights, allowing various tasks to utilize the experts in unique ways. The conclusions from the assembled experts are then sent to a tower network created for this particular purpose. Gated networks for various tasks can learn distinct mixing patterns of expert combinations to capture task relationships and, more precisely, the output of task k,

$$y_k = h^k\left(f^k(x)\right), \tag{1}$$

$$f^k(x) = \sum_{i=1}^{n} g_i^k(x) f_i(x), \tag{2}$$

$$g^k(x) = \text{softmax}(W_{gk}x), \tag{3}$$

where f_i, $i = 1, \ldots, n$, are n expert networks, h^k, $k = 1, 2, 3$, correspond to each task (i.e., three ion implantation parameters), and the model is implemented using the same multilayer perceptron with ReLU activation. The gating network involves a linear transformation of the input through a softmax layer, with the trainable matrix $W_{gk} \in \mathbb{R}^{(n\times 22)}$.

IV. MODEL TESTING

After finalizing the design of the deep learning model architecture, it is essential to create a dataset with diverse data samples and an effective training approach for model training. Training will determine all adjustable parameters in the model, enabling it to derive ion implantation process parameters from CI signals.

A. Key functions for training the model

When developing the training approach, it is crucial to focus on the loss and evaluation index functions. The backpropagation technique updates all trainable parameters in the model during the training process. The loss function, which quantifies the discrepancy between the model's output and the real data, is fundamental to the backpropagation method. The choice of this function determines how the model parameters are updated. On the other hand, evaluating the indicator function helps to select the optimal combination of model performance parameters.

Since the identification of ion implantation process parameters in this paper is a regression problem, the loss function commonly used in regression scenarios is the mean square error (MSE). The value is calculated by averaging the sum of the squares of the discrepancies between the estimated values and the real values using the following expression:

$$\text{Loss} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{4}$$

where n is the number of tasks to be studied, y denotes the process parameter's actual value, and $\hat{y}$ is the process parameter's predicted value.

Considering that the same algorithmic model is used in this work to solve different problems, the root mean square error (RMSE) and the coefficient of determination (R^2) are used as evaluation metrics. R^2 is also evaluated as the best measure of the linear regression method with the expression

$$R^2 = 1 - \frac{\sum_i \left(\hat{y}^{(i)} - y^{(i)}\right)^2}{\sum_i \left(\bar{y} - y^{(i)}\right)^2}, \tag{5}$$

where $\hat{y}$ is the prediction result, y is the actual value, and $\bar{y}$ is the mean of the prediction result. The denominator represents the error in the baseline (mean), and the numerator represents the error generated by the model's prediction results.

B. Model training implementation

The dataset acquired from the experimental procedure outlined in Sec. II comprises 360 samples for training, validation, and testing the model. Out of the total, 226 samples were allocated to the training set, while the remaining samples were evenly divided between the validation and testing sets. In addition, a design was made to ensure that the distribution of samples was the same for all three sets. The construction and development of the descriptive model were carried out in Python 3.6 utilizing TensorFlow 2.5. The backpropagation technique was implemented using the Adam optimizer with an initial learning rate of 0.001. After 300 epochs, the model reached a steady state regarding the evaluation metrics on the test set, and the optimal combination of model parameters was trained.

During the training process, RMSE and R^2 stabilized at 300 epochs for all three subtasks, and the model did not exhibit significant overfitting when comparing the evaluation metrics of the training and validation sets. Using R^2 in conjunction with RMSE helps fully assess the model's accuracy and generalization capabilities. In contrast, R^2 assesses overall explanatory power, while RMSE focuses on the magnitude of specific prediction errors. Combining these two metrics provides a more complete picture of model performance in different areas, helping to make more accurate model selection and optimization decisions.

V. RESULT AND DISCUSSION

This work evaluates the parameter extraction accuracy and recognition speed of the PPMTL model and a typical single-task model recurrent neural network (RNN) in order to determine the utility of the proposed deep learning-based technique after obtaining the training model. The overall performance of the two training models is also evaluated and compared, followed by a detailed discussion on how the accuracy of the models can be improved, thus comprehensively analyzing the utility of the proposed deep learning-based strategy.

A. Model comparison

This paper first estimates the process parameters from the CI power curve data, including three regression variables (energy,

angle, and dose). Figure 6 shows the performance of the proposed PPMTL method, compared with using various single-task models. A recurrent neural network (RNN) is a classical neural network structure with a recursive structure that performs well when processing sequential data. Long short-term memory (LSTM) is a special type of recurrent neural network (RNN) architecture that typically outperforms RNNs in dealing with long-term dependencies and time series prediction tasks. In the PPMTL model, the implantation dose is the primary job, while the implantation angle and implantation energy are secondary tasks. For hyperparameter tuning, the root mean square error (RMSE) of the principal task on the validation set serves as the target metric. We performed thousands of experiments for each method using the hyperparameter tuner to find the optimal hyperparameter settings.

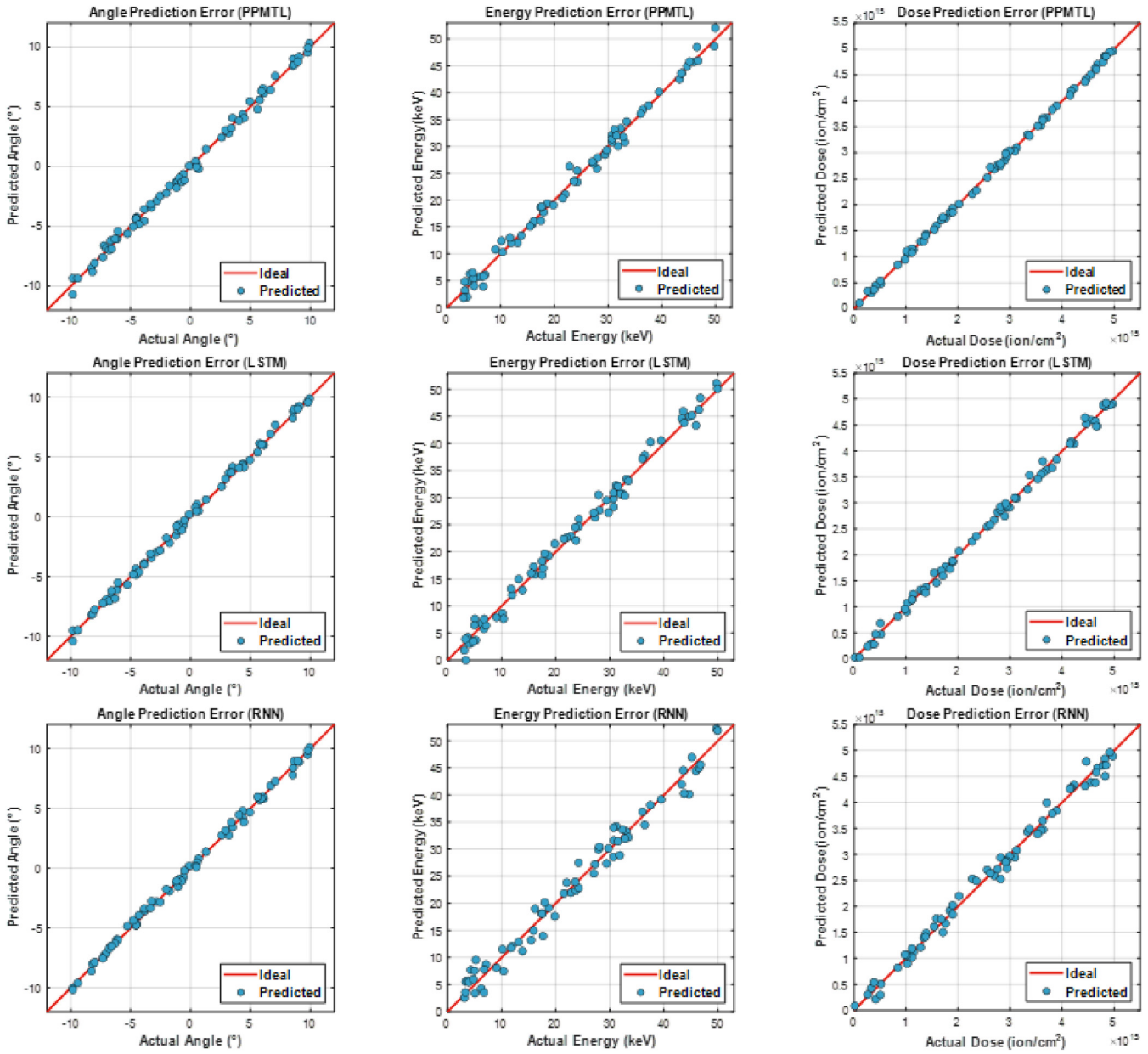

FIG. 6. Actual and predicted values of the PPMTL model, the RNN model, and LSTM for the three process parameters, with blue representing the actual points and red representing the perfectly consistent curve.

TABLE I. RMSE of three trained models on different subtasks.

	Angle (deg)	Energy (keV)	Dose (ions/cm^2)
PPMTL	0.127	0.767	1.713×10^{13}
LSTM	0.108	1.106	3.12×10^{13}
RNN	0.118	1.284	4.602×10^{13}

After the hyperparameter tuner identifies the optimal hyperparameters for each approach, we perform 300 training iterations on the training dataset for each method using random parameter initialization. The results are assessed on the test dataset using the same experimental conditions to compare the two models.

The overall recognition effect of the trained models is reflected in the performance of the parameter extraction ability on the dataset; Fig. 6 shows the predicted parameter values vs the actual parameter values for the PPMTL model, the RNN model, and the LSTM model after training and tuning using the 67 data points in the test set. For convenience, the root mean square errors of the three ion implantation process parameters are compiled in Table I. As seen from Table I and Fig. 6, for each of the three process parameters, all three models agree with the red 1:1 line. However, the PPMTL model consistently outperforms both the RNN and LSTM models, particularly in predicting ion implantation process parameters. The PPMTL model achieves an implantation energy RMSE of 0.767 keV, an implantation angle RMSE of 0.127°, and an implantation dose RMSE of 1.7×10^{13} ions/cm^2, representing only 1.59%, 0.63%, and 0.34% of the training range, respectively. In contrast, the RNN model shows RMSE values of 1.284 keV for energy, 0.118° for angle, and 4.6×10^{13} ions/cm^2 for dose, which are 2.6%, 0.59%, and 0.92% of the training range. Similarly, the LSTM model demonstrates RMSE values of 1.106 keV for energy, 0.108° for angle, and 3.12×10^{13} ions/cm^2 for dose, corresponding to 2.2%, 0.54%, and 0.62% of the training range, respectively. It can be seen that the PPMTL model performs better in terms of the accuracy of other ion implantation parameters, except for the implantation angle error, which is slightly higher than that of the single-task model. The multi-task model is not optimized for the implantation angle task on the validation set, while the single-task model is optimized. The single-task model excels in the auxiliary task, leading to the conclusion that PPMTL performs exceptionally well and surpasses the single-task model in overall performance. It accurately predicts the three parameters in the CI power curve within a 1.6% margin of the training range. Another metric commonly evaluated in machine learning is R^2. We utilized this metric to evaluate our model's performance. Figure 7 illustrates the comparison between the multi-task model and the single-task model's performance on each subtask. This supports the effectiveness of our model on the test set and its precise prediction of the parameters in the CI power curve.

B. Inference time

PMOR is usually widely used in online measurement scenarios that require high speed. Therefore, when selecting an ion implantation parameter extraction algorithm for CI metrology, it is crucial to take into account the algorithm's time cost. The time cost of the deep learning-based approach in this study is mostly driven by the inference time of the model, which is the time needed for the model to calculate the inputs to the appropriate outputs.

In practical uses of CI metrology, parameter extraction algorithms typically analyze individual signals rather than multiple signals at once. Thus, the focus is on the processing time of a single signal. Figure 8 displays a comparison of the processing times of the two models, with computations performed on an AMD Ryzen 7 4800H CPU.

According to the results in Fig. 8 , the single-task model has a shorter inference time than the PPMTL model. The single-task model requires fewer operations to perform a single inference compared to the PPMTL model due to its lower complexity. However, it is essential to note that the sum of the reasoning times for the single-task model is much more significant than the reasoning time for the PPMTL model. This further demonstrates the advantage of the PPMTL model for parameter extraction in CI metrology. The single-task model is also a good choice if only a single implant parameter is of interest. However, in online monitoring of the implant level of ion implantation equipment, accurate

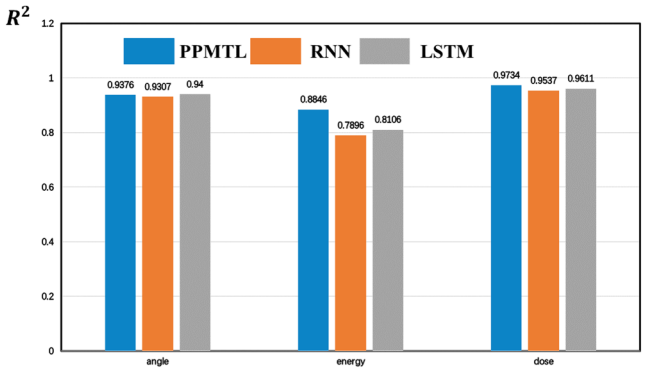

FIG. 7. Comparison of the three model assessment metrics R^2.

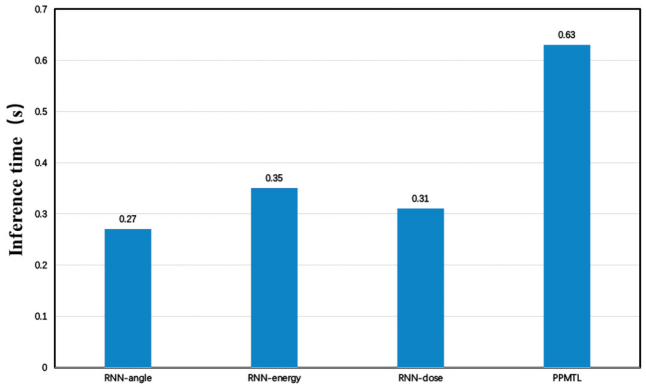

FIG. 8. Comparison of inference times for two models for a single power curve.

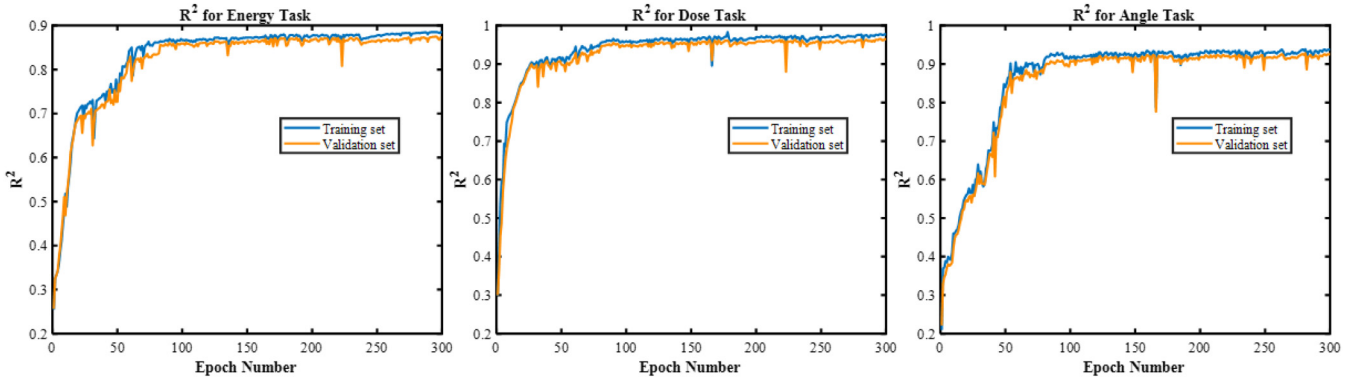

FIG. 9. The fractional R^2 vs epochs for the three subtasks on validation and training data.

characterization of multiple parameters can better reflect the changes in the implant level. Therefore, simultaneous multi-task characterization is of practical interest in this context.

C. Discussion of model accuracy improvement

One may wonder if the predicted accuracy of PPMTL may be enhanced by increasing dataset training, incorporating more data (such as power curves and corresponding parameters), or expanding the network's number of layers. We first assess whether more training could enhance the model's performance. Figure 9 shows the trend of the R^2 evaluation metrics with epochs for the three subtasks. The blue line represents the training set, and the yellow line represents the validation set. The graphs indicate a notable improvement in R^2 at the start of training. By 300 epochs, further training reaches a point of diminishing returns, with no substantial enhancement in the overall R^2.

Second, we will explore the option of augmenting the dataset size utilized for training PPMTL. Increasing the amount of data can effectively decrease the root mean square error (RMSE) on the validation set. In any analysis, having more data is consistently beneficial. The network can accurately estimate parameters within the training parameter range with an accuracy of up to 1.6%. We rate the model performance positively; however, in cases where higher accuracy is required, expanding the dataset by conducting experiments is a viable way. As the study progresses, we will continue to expand the dataset by testing more and more samples. This has significant implications for engineering applications, allowing us to meet the needs of our customers for large-scale test samples and to generate richer datasets.

Finally, consider whether increasing the depth of the network is beneficial. Overall, PPMTL needs help fully understand the data patterns at the beginning of training, i.e., it is in an underfitting state.[26] Through ongoing training, the network learns patterns and decreases the root mean square error (RMSE) on both the training and validation datasets. Eventually, the model may overfit the unique characteristics of the training set, leading to overfitting. Overfitting is challenging since it causes a reduction in RMSE on the training set while increasing RMSE on the validation set. Additional layers aid in more accurately recognizing patterns and characteristics within the data, contributing to the success of deep learning. Deeper networks may not always be superior as increasing the number of layers can result in overfitting, causing the model to underperform on the test data.[27] In this study, we used ResNet 25 because when comparing the results of ResNet 25 (25 layers of neurons) and ResNet 50 (50 layers), it was found that the more extensive network showed more overfitting.

VI. CONCLUSION

We investigated a deep learning-based approach for extracting multiple parameters from CI signals. We successfully applied deep learning techniques to the field of CI metrology. Specifically, we conducted a series of Carrier Illumination experiments for different implantation parameters to obtain the data required for deep learning. Based on this, we trained and compared two deep-learning models and found that the PPMTL model demonstrated excellent overall performance in training accuracy and recognition speed. Thus, deep learning provides an effective solution for fast and accurate extraction of ion implantation process parameters in CI metrology. Although our study mainly focuses on CI metrology, we also found that this approach is potentially feasible in another special implementation of PMOR technology, Therma-Probe (TP), which extracts ion implant process parameters from offset curves. The PPMTL model is anticipated to be beneficial for PMOR metrology in manufacturing intricate structured integrated circuits as more data samples are collected.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NNSFC) (Grant No. 52375541) and the National Key Research and Development Programme of China (No. 2022YFF0709104).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Xuesong Wang: Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (equal). **Lijun Zhang:** Investigation (equal); Software (equal). **Yong Sun:** Investigation (equal). **Jin Min:** Investigation (equal). **Zhongyu Wang:** Investigation (equal). **Shiyuan Liu:** Funding acquisition (equal); Investigation (equal). **Xiuguo Chen:** Funding acquisition (equal); Investigation (equal). **Zirong Tang:** Funding acquisition (equal); Investigation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]J. Bogdanowicz, *Photomodulated Optical Reflectance: A Fundamental Study Aimed at Non-Destructive Carrier Profiling in Silicon* (Springer Science & Business Media, 2012).

[2]F. Dortu, "Low frequency modulated optical reflectance for the one-dimensional characterization of ultra-shallow junctions," Doctoral thesis (Katholieke Universiteit Leuven, 2009).

[3]I. Barbereau, B. Forget, and D. Fournier, "Characterization of electronic transport properties in semiconductors by scanning photothermal microscopy," Opt. Express **6**, S479–S482 (1996).

[4]B. C. Forget and D. Fournier, "Electronic transport properties characterization of silicon wafers by modulated photoreflectance," J. Phys. IV France **1**(6), 6277–6282 (1991).

[5]B. C. Forget, I. Barbereau, D. Fournier, S. Tuli, and A. B. Battacharyya, "Electronic diffusivity measurement in silicon by photothermal microscopy," Appl. Phys. Lett. **69**, 1107–1109 (1996).

[6]D. Fournier and A. C. Boccara, "Thermal wave probing of the optical electronic and thermal properties of semiconductors," Mater. Sci. Eng., B **5**, 83–88 (1990).

[7]B. Li, L. Pottier, J. P. Roger, D. Fournier, and E. Welsch, "Thermal characterization of film-on-substrate systems with modulated thermoreflectance microscopy," Rev. Sci. Instrum. **71**, 2154–2160 (2000).

[8]L. Nicolaides, A. Salnick, and J. Opsal, "Study of low energy implants for ultra-shallow junctions using thermal wave and optical techniques," Rev. Sci. Instrum. **74**, 563–565 (2003).

[9]L. Nicolaides, A. Salnick, and J. Opsal, "Nondestructive analysis of ultrashallow junctions using thermal wave technology," Rev. Sci. Instrum. **74**, 586–588 (2003).

[10]A. Salnick and J. Opsal, "Quantitative photothermal characterization of ion-implanted layers in si," J. Appl. Phys. **91**, 2874–2882 (2002).

[11]G. Smets, E. Rosseel, G. Sterckx, J. Bogdanowicz, W. Vandervorst, and D. Shaughnessy, "Transfer from rs-based to pmor-based ion implantation process monitoring," AIP Conf. Proc. **1321**, 426–431 (2011).

[12]P. Borden, C. Ferguson, D. Sing, L. Larson, L. Bechtler, K. Jones, and P. Gable, "In-line characterization of preamorphous implants (PAI)," in *2000 International Conference on Ion Implantation Technology Proceedings (Cat. No. 00EX432)* (IEEE, 2000), pp. 635–638.

[13]C. Christofides, I. A. Vitkin, and A. Mandelis, "Photothermal reflectance investigation of processed silicon. I. Room-temperature study of the induced damage and of the annealing kinetics of defects in ion-implanted wafers," J. Appl. Phys. **67**, 2815–2821 (1990).

[14]B. C. Forget and D. Fournier, "Characterization of implanted silicon wafers by the non-linear photoreflectance technique," Mater. Sci. Eng., B **24**, 199–202 (1994).

[15]B. C. Forget, D. Fournier, and V. E. Gusev, "Non linear recombination processes: Application to quantitative implantation characterization," J. Phys. IV France **4**(7), 155–158 (1994).

[16]W. Vandervorst, T. Clarysse, B. Brijs, R. Loo, Y. Peytier, B. J. Pawlak, E. Budiarto, and P. Borden, "Carrier illumination as a tool to probe implant dose and electrical activation," AIP Conf. Proc. **683**, 758–763 (2003).

[17]F. Dortu, T. Clarysse, R. Loo, and W. Vandervorst, "Extracting active dopant profile information from carrier illumination power curves," J. Vac. Sci. Technol. B **24**, 375–380 (2006).

[18]E. S. Paul Covington and J. Adams, "Deep neural networks for youtube recommendations" in *RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems* (Association for Computing Machinery, 2016), pp. 191–198.

[19]T. Bansal, D. Belanger, and A. McCallum, "Ask the GRU: Multi-task learning for deep text recommendations" in *RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems* (Association for Computing Machinery, 2016), pp. 107–114.

[20]K. M. Loew and R. M. Bradley, "Parameter estimation for pattern formation induced by ion bombardment of solid surfaces using deep learning," J. Phys.: Condens. Matter **33**, 025901 (2021).

[21]R. Caruana, "Multitask learning," Learn. Learn **28**, 95–133 (1998).

[22]L. Duong, T. Cohn, S. Bird, and P. Cook, "Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Association for Computational Linguistics, 2015), Vol. 2: short papers, pp. 845–850.

[23]Y. Yang and T. M. Hospedales, "Deep multi-task representation learning: A tensor factorisation approach," arXiv:abs/1605.06391 (2016).

[24]J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts" in *KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Association for Computing Machinery, 2018), pp. 1930–1939.

[25]R. Caruana, "Multitask learning," Mach. Learn. **28**, 41–75 (1997).

[26]A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow* (O'Reilly Media Inc, 2022).

[27]Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, Berlin, 2012), pp. 437–478.