

ENHANCED OPTICAL CHARACTERIZATION OF MULTILAYER THIN FILMS USING CONVOLUTIONAL NEURAL NETWORKS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Hongrui Zhang

June 2024

© 2024 Hongrui Zhang
ALL RIGHTS RESERVED

ABSTRACT

Autonomous experiments and multi-objective optimization are pivotal in advancing materials science, enabling the efficient discovery and characterization of novel materials. In multi-objective optimization for transparent conducting oxides, accurately determining the optical properties and thickness of these multilayer thin films is essential. However, this task is challenging due to the ill-posed nature of the fitting problem, which can lead to entrapment in local minima and slow processing times. Thus, precise initial guesses are necessary for effective fitting. This study addresses these challenges using Bi_2O_3 thin films as a case study, introducing a method that employs Convolutional Neural Networks (CNNs) to extract these properties from reflectance data of films processed by lateral gradient laser spike annealing (lg-LSA). A simulated reflectance dataset, generated using the Tauc-Lorentz (TL) model and Transfer Matrix Method (TMM), was used to train the CNN to predict TL parameters and thickness. The CNN model achieved a high mean R^2 score of 0.9754 on simulated data. The model's accuracy improved on experimental data from a mean R^2 score of 0.5892 to 0.8408 after subsequent trust region reflective (TRF) fitting. The CNN model demonstrated robustness across the lg-LSA processed stripe and accurately predicted the thickness profile, aligning well with experimental observations. This research underscores the potential of integrating machine learning techniques for high-throughput, autonomous material characterization, providing a solid foundation for future enhancements and applications across diverse material systems.

BIOGRAPHICAL SKETCH

Hongrui Zhang was born in Guangzhou, China. In 2022, he earned his Bachelor of Science degree in Materials Science and Engineering (MSE) from Guangdong Technion—Israel Institute of Technology, where he gained valuable research experience in nanofabrication and molecular electronics. Following his undergraduate studies, he pursued a master's degree in MSE at Cornell University, joining Prof. Michael Thompson's research group.

Driven by a strong interest in autonomous experiments, Hongrui focused his research on applying Convolutional Neural Networks (CNNs) to determine the optical properties of thin films from reflectance data. Throughout his master's studies, he augmented his expertise in materials science with coursework in Computer Science, equipping him with the computational skills necessary for his innovative research. His work aims to improve the accuracy and efficiency of high-throughput, autonomous material characterization, an essential component of multi-objective optimization in materials discovery.

Upon completing his master's degree, Hongrui plans to continue his research in autonomous experiments within materials science as a PhD student in MSE at Boston University. He is eager to further his studies and contribute to advancing the field through pioneering research and collaboration.

This document is dedicated to all Cornell graduate students.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my parents and family for their unwavering support, encouragement, and understanding throughout my two-year academic journey at Cornell. Their belief in me has been a constant source of motivation and strength.

I am profoundly grateful to my advisor, Prof. Michael Thompson, for his steadfast support, guidance, and encouragement throughout my master's research. His insights and expertise have been invaluable to my work and personal growth. I also extend my heartfelt thanks to Prof. Bruce van Dover for his valuable feedback and suggestions that greatly enhanced the quality of my research. Special thanks to my minor committee member, Prof. Carla Gomes, for her insightful comments and encouragement.

I would also like to thank all my colleagues and friends in Prof. Thompson's group. In particular, I sincerely appreciate Ming-Chiang Chang for his detailed guidance and selfless support of my research. His expertise and assistance have been instrumental in my progress. I also wish to thank Aine Connolly, Cameron Gorsak, Katie Gann, and Duncan Sutherland for their invaluable advice on my research and future academic pursuits. My peers Tianhai Luo and Hancheng Wang also deserve high appreciation for their support and companionship.

Lastly, I would like to thank all the faculty, staff, and fellow students in the Department of Materials Science and Engineering at Cornell University for fostering a collaborative and inspiring environment that has contributed significantly to my academic and personal development.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
1 Introduction	1
1.1 Multi-objective Optimization in Autonomous Material Discovery	1
1.2 Exemplary Materials System: Transparent Conducting Oxides (TCOs)	3
1.3 Lateral Gradient Laser Spike Annealing (lg-LSA)	6
1.4 Transparency Characterization	7
1.5 Extracting Optical Properties From Reflectance	9
1.6 Convolutional Neural Network (CNN) Approach	12
2 Methods	16
2.1 TCO Sample and lg-LSA Setup	16
2.2 Reflectance Measurement	16
2.3 Transfer Matrix Method (TMM)	19
2.4 Tauc-Lorentz (TL) Model	21
2.5 Generating Training Data	24
2.5.1 Training Set and Testing Set	27
2.6 Convolutional Neural Network (CNN)	28
2.6.1 Architecture of CNN	29
2.6.2 Training Process	35
2.7 Trust Region Reflective Algorithm	36
2.8 Baseline Model	38
3 Results and Discussion	40
3.1 CNN Accuracy for Simulated Test Set	40
3.1.1 Single Reflectance Spectra	40
3.1.2 Number of Feature Extraction Layers	44
3.1.3 Learning Rate	47
3.1.4 Training Data Size	49
3.1.5 Input Dimension	49
3.1.6 Best Model	51
3.1.7 Parameters Accuracy Comparison	53
3.2 CNN Prediction and TRF Fitting for Experimental Test Set	54
3.2.1 Before TRF Fitting	54

3.2.2	With TRF Fitting	61
4	Conclusion and Future Work	69
4.1	Conclusion	69
4.2	Future Work	71
	Bibliography	74

LIST OF TABLES

2.1	Convolutional filter sizes for different numbers of layers in the CNN architecture.	34
2.2	Max-pooling kernel sizes for different numbers of layers in the CNN architecture.	35
2.3	Explored variable values for determining model accuracy.	38
3.1	Time Comparison of CNN Prediction and TRF Fitting.	64

LIST OF FIGURES

1.1	Schematic of decision and performance spaces in a multi-objective optimization problem. The red-shaded region on the left represents the decision space of experimental parameters, and the grey-shaded region on the right denotes the performance space of material properties (transparency and electrical conductivity). A point in the decision space leads to a feasible solution in the performance space (point A), a point outside of the decision space leads to an infeasible solution (point B), and point C is an optimal solution that lies on the Pareto front.	4
1.2	Illustration of the principle of lg-LSA. (a) Setup of lg-LSA. A scanned laser heats the material with a peak temperature gradient on the material. Lg-LSA conditions can be varied across the wafer. A camera is used to inspect the optical image of the LSA. (b) Bottom: An optical image of a stripe after lg-LSA. Top: Temperature heatmap of the laser and Gaussian profile of the peak temperature gradient.	7
1.3	(a) Structure of the typical multilayer sample. (b) Example of a single reflectance spectrum.	9
1.4	Reflectance measurement and heatmap: (a) Optical image of the LSA stripe. The grey dashed line represents the position of the reflectance measurements. Three representative measurements (in three different phases) are shown in (b). (c) Reflectance heatmap, where the x-axis is the position across the stripe (represented as the data index), and the y-axis is the photon energy E . The color represents the absolute reflectance.	10
1.5	Determining transparency from reflectance measurement: (a) Common multi-layer thin film system; (b) Variation of n and k with photon energy of the incident beam. (c) The theoretical reflectance is calculated using the TMM, and n, k are modified by fitting the theoretical reflectance to the measured reflectance data.	11
1.6	Workflow of the project. The measured reflectance data serves as an input of a CNN model and predicts initial parameters for the reflectance. Then, the initial reflectance goes through further fitting to improve the accuracy. The n, k, d associated with this reflectance is a robust estimate of the true optical properties and thickness of the top layer.	14
2.1	Illustration of the principle of reflectance measurement.	17
2.2	Illustration of the transfer matrix method (TMM) applied to the multilayer thin films in this project.	20

2.3	The dielectric function, $\varepsilon(E)$, of the Tauc-Lorentz model as a function of photon energy. The real part $\varepsilon_{\text{re}}(E)$ (light blue) and the imaginary part $\varepsilon_{\text{im}}(E)$ (red) are shown, with the key parameters of the model annotated.	24
2.4	Illustration of Latin Hypercube Sampling. Each row and column has only one sampled combination of parameter 1 and parameter 2, ensuring the evenness of the sample.	26
2.5	Workflow illustrating the data generation process using the TL model. Parameters are sampled via Latin hypercube sampling (LHS) to produce n, k values for the top layer, which, along with thickness d , are used in the transfer matrix method (TMM) to simulate noisy reflectance data. An example of 10 randomly selected reflectance datasets is shown.	27
2.6	Neural Network Architecture and Activation Function. (a) A three-layer neural network comprising an input layer with two input units x_1 and x_2 , a hidden layer with two neurons h_1 and h_2 , and an output layer with one output unit o_1 . The weights w_1 to w_6 and biases b_1 to b_3 connect the layers as depicted. (b) The ReLU activation function used in the hidden and output layers, defined as $\text{ReLU}(x) = \max(0, x)$, demonstrates its effect on input values ranging from -10 to 10.	30
2.7	Illustration of the convolution, batch normalization, activation, and max pooling layers on a 1D data.	32
2.8	CNN architecture for predicting material optical properties from reflectance spectra. The network consists of feature extraction layers and fully connected layers for parameter prediction. The input reflectance spectrum is processed through 64 filters and n_{layers} number of feature extraction layers. The fully connected layers have dimensions of 3000, 1200, 300, and 6, with a dropout rate 0.5. The final output predicts six parameters $A, E_0, G, E_g, \varepsilon_\infty$, and d	33
3.1	Baseline model prediction for reflectance, n , k , and parameter comparison for a randomly selected single simulated data (Data 0 means it is the first data in the dataset). (a) Radar chart of parameters comparison. (b) True and predicted n values. (c) True and predicted k values. (d) True and predicted reflectance values, with MSE and R^2 score calculated.	43

3.2	Baseline model prediction for reflectance, n , k , and parameter comparison for an outlier (data 404) with the largest MSE in the dataset. (a) Radar chart of parameters comparison. (b) True and predicted n values. (c) True and predicted k values. (d) True and predicted reflectance values, with MSE and R^2 score calculated.	45
3.3	Box plots of metrics for different numbers of feature extraction layers. (a) Mean squared error (MSE) of normalized parameters at different numbers of feature extraction layers. (b) R^2 score of reflectance at different numbers of feature extraction layers.	47
3.4	Box plots of metrics for different learning rates. (a) Mean squared error (MSE) of normalized parameters at different learning rates. (b) R^2 score of reflectance at different learning rates.	48
3.5	Box plots of metrics for different training data sizes. (a) Mean squared error (MSE) of normalized parameters at different training data sizes. (b) R^2 score of reflectance at different training data sizes.	50
3.6	Box plots of metrics for different input dimensions. (a) Mean squared error (MSE) of normalized parameters at different input dimensions. (b) R^2 score of reflectance at different input dimensions.	51
3.7	The best model prediction for reflectance, n , k , and parameter comparison of the same single simulated data used for the baseline model (data 0). (a) Radar chart of parameters comparison. (b) True and predicted n values. (c) True and predicted k values. (d) True and predicted reflectance values, with MSE and R^2 score calculated.	52
3.8	True and predicted values of all parameters. From left to right, top to bottom, subplots of the parameters A , E_0 , G , E_g , ε_∞ , d . The red line shows the ideal $y = x$ condition.	54
3.9	True and predicted values of n , k at photon energy 2.2eV. The blue and green scattering dots have coordinates (true value, predicted value), and the red line has an analytical formula $y = x$	55
3.10	The best CNN model prediction for reflectance, n , k , and parameter comparison for randomly selected experimental data (Data 0 means it is the first data in the dataset). (a) Radar chart of predicted parameters. (b) Predicted n curve. (c) Predicted k curve. (d) Predicted and experimental reflectance curves, with calculated MSE and R^2 scores.	56
3.11	Box plots of R^2 score of reflectance for simulated data (left) and experimental data (right).	57

3.12	(a) Heatmap of the predicted n , (b) Heatmap of the predicted k , (c) Heatmap of the predicted reflectance with Mean MSE and Mean R^2 score calculated, and (d) Experimentally measured reflectance for comparison.	59
3.13	CNN predicted thickness d across the stripe.	60
3.14	The best CNN model prediction with TRF fitting for reflectance, n , k , and parameter comparison for the same single experimental data used in figure 3.10 (data 0). (a) Radar chart of CNN predicted parameters (in orange) and the TRF optimized parameters (in red). (b) Predicted and fitted n values. (c) Predicted and fitted k values. (d) Predicted, fitted, and experimental reflectance, with MSE and R^2 score between the fitted reflectance and the experimental reflectance.	62
3.15	Box plots of R^2 score of reflectance before TRF fitting (left) and after TRF fitting (right).	63
3.16	(a) Heatmap of the fitted n , (b) Heatmap of the fitted k , (c) Heatmap of the fitted reflectance with Mean MSE and Mean R^2 score calculated, and (d) Experimentally measured reflectance for comparison.	65
3.17	Predicted thickness (in blue) and fitted thickness (in orange) across the stripe.	66
3.18	Profilometry characterization of a laser spike annealed stripe rich in bismuth in a Bi-Ti-O composition spread[1].	67

CHAPTER 1

INTRODUCTION

1.1 Multi-objective Optimization in Autonomous Material Discovery

Autonomous experimentation, an approach that utilizes artificial intelligence (AI) to minimize the number of experiments needed to achieve insightful results, has garnered significant attention in the materials science community[2][3][4]. A notable autonomous framework, the Scientific Autonomous Reasoning Agent (SARA)[5], has been developed by our groups at Cornell to accelerate the exploration of new metastable materials. Exploration in this context refers to broadly searching the parameter space to uncover unknown regions of composition and processing that might yield novel or significantly improved materials. This approach is crucial for identifying potential areas of interest where optimal materials might be found. The counterpart concept is exploitation[6], which focuses more narrowly, aiming to deepen the investigation within identified regions to optimize and thoroughly understand material properties.

While this framework has shown promising results in efficiently navigating multidimensional parameter spaces to discover novel material systems, its capabilities in exploitation, which specifically focuses on the investigation within a specific region of the phase boundaries to extract more detailed and refined information, are not yet well-developed. Therefore, it is crucial to expand the

capabilities within exploitation.

Exploitation aims at identifying an optimal material system by focusing on fine-tuning parameters in the most promising regions. Exploitation is often accomplished through multi-objective optimization, where the goal simultaneously optimizes several parameters or properties. A classic example is in transparent conducting oxides (TCOs), where one seeks to optimize both electrical conductivity and transparency across the visible spectrum. These properties are typically at odds, making achieving high transparency and low resistivity challenging under the same experimental conditions.

In such multi-objective optimization tasks, it is useful to consider both the decision space and the performance space. The decision space, which encompasses the range of experimental parameters (e.g., processing temperatures and compositions), represents the set of inputs that can be adjusted to influence the material properties. Meanwhile, the performance space describes the material's output characteristics, such as electrical conductivity and transparency, mapped from the decision space. In this context, feasible solutions lie within the performance space, whereas infeasible solutions fall outside. For multi-objective optimization, the goal is to establish a set of optimal feasible solutions, or Pareto-optimal solutions, where no objective (e.g., transparency) can be improved without worsening another (e.g., electrical conductivity). This set of Pareto-optimal solutions forms the Pareto front or Pareto-optimal front in the performance space, providing a boundary that defines the best trade-offs between conflicting objectives in the performance space.

Figure 1.1 illustrates the relationship between the decision space and the corresponding performance space. Points within the decision space lead to feasible solutions, such as point A, while points outside the decision space, that is, outside the specified range of experimental parameters, result in infeasible solutions like point B. Infeasible solutions typically correspond to materials with insufficient transparency or electrical conductivity to meet acceptable thresholds. Certain combinations of experimental parameters yield Pareto-optimal solutions (point C), where improving transparency would degrade electrical conductivity and vice versa. The primary objective of this multi-objective optimization problem is to identify the set of optimal solutions that constitute the Pareto front, representing the best trade-offs between conflicting material properties.

By finding the Pareto front, exploitation helps in fine-tuning the solutions within the most promising regions identified during the exploration phase and achieving the best possible trade-offs among the multiple objectives. This approach enables identifying an optimal material system more efficiently and effectively.

1.2 Exemplary Materials System: Transparent Conducting Oxides (TCOs)

As suggested previously, an interesting system for multi-objective optimization would be transparent conducting oxides (TCOs), which are widely used in a

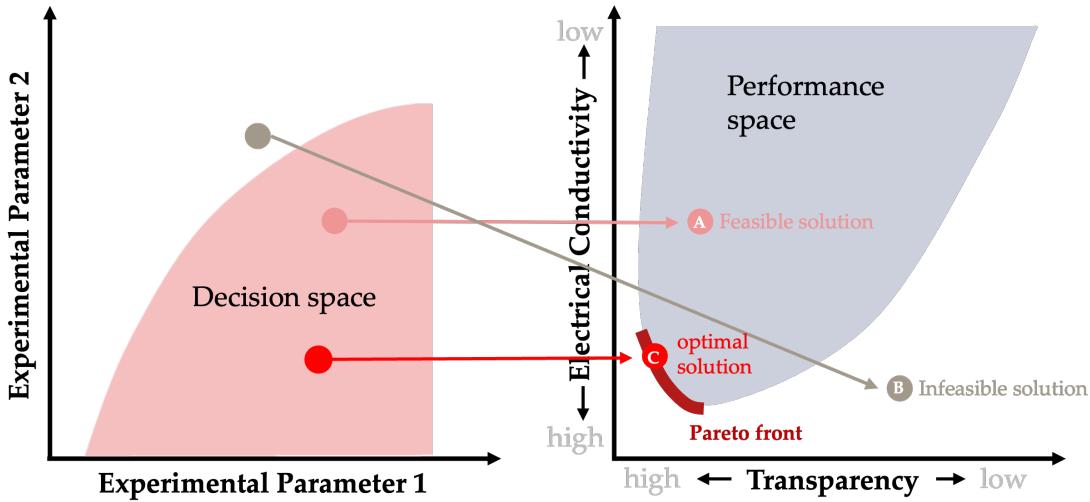


Figure 1.1: Schematic of decision and performance spaces in a multi-objective optimization problem. The red-shaded region on the left represents the decision space of experimental parameters, and the grey-shaded region on the right denotes the performance space of material properties (transparency and electrical conductivity). A point in the decision space leads to a feasible solution in the performance space (point A), a point outside of the decision space leads to an infeasible solution (point B), and point C is an optimal solution that lies on the Pareto front.

broad range of technologies, including cell phones and computer displays[7]. Common TCOs include indium tin oxide (ITO), cadmium tin oxide (CTO), and indium gallium zinc oxide (IGZO)[8]. For a TCO with a complex composition, different phases often present different electrical and optical properties and must be independently investigated.

For the multi-objective optimization problem in TCOs, we need to optimize both the visible band transparency and the electrical conductivity and find the Pareto front within its objective space. Specifically, we want to minimize the electrical resistivity ρ [9] while maximizing the fraction of light within the visible band (400-700nm) that is transmitted by the material, the transmittance T . For

incident light into a material, the intensity $I(z)$ as a function of depth z is given by[10]:

$$I(z) = I_0 e^{-\alpha z} \quad (1.1)$$

where I_0 is the initial intensity and $\alpha(\lambda)$ is the absorption coefficient. This absorption coefficient is related to the complex index of refraction via:

$$\alpha(\lambda) = \frac{4\pi k}{\lambda} \quad (1.2)$$

where k is the imaginary part of the complex refractive index $\tilde{n} = n - ik$. α^{-1} is also known as the extinction coefficient and describes how light is absorbed in the material. Ignoring reflections, the transmittance T at depth z is defined as:

$$T(z) = \frac{I(z)}{I_0} \quad (1.3)$$

and thus

$$T(z) = e^{-\alpha z} = e^{-\frac{4\pi}{\lambda} kz} \quad (1.4)$$

For a material with a fixed thickness, the transmittance is maximized for the smallest k at a fixed wavelength. Therefore, the goal of maximizing T is equivalent to minimizing k .

In summary, the goal of the multi-objective optimization problem is to minimize both ρ and k , both being intrinsic properties of the material. This goal requires investigating different phases of material in a high-throughput manner, which we explore using the lateral gradient laser spike annealing (lg-LSA) method.

1.3 Lateral Gradient Laser Spike Annealing (lg-LSA)

As the objective space is composed of transparency and conductivity, the decision space that determines these properties includes the controlled experimental parameters. One experimental method to investigate this problem is lateral gradient laser spike annealing (lg-LSA), which has been used by the SARA project and has proven successful in high-throughput autonomous experiments exploring the objective space of crystal phases. Lg-LSA was developed by Bell et al. [11] as a technique to anneal thin films with a temperature gradient using a scanned intense laser beam. The result is often the formation of different material phases via rapid heating and cooling. Figure 1.2 illustrates the principle of lg-LSA, with the lg-LSA experimental setup in 1.2 (a). A laser heats the material with a lateral intensity gradient; the stage holding a wafer moves in the vertical direction (the purple arrow) to produce a stripe, whose optical image is shown in the bottom part of figure 1.2 (b). The upper part of figure 1.2 (b) shows the temperature profile across the stripe (the lateral gradient). The central temperature T_{center} can be adjusted by varying the laser power and the heating duration (dwell time) controlled by the stage velocity.

In the context of multi-objective optimization, the decision space of the lg-LSA experiment is composed of T_{center} and dwell time during lg-LSA. In addition, sample wafers can include compositional variations, allowing alloy phases and composition optimization. This system thus provides a useful platform for multi-objective optimization.

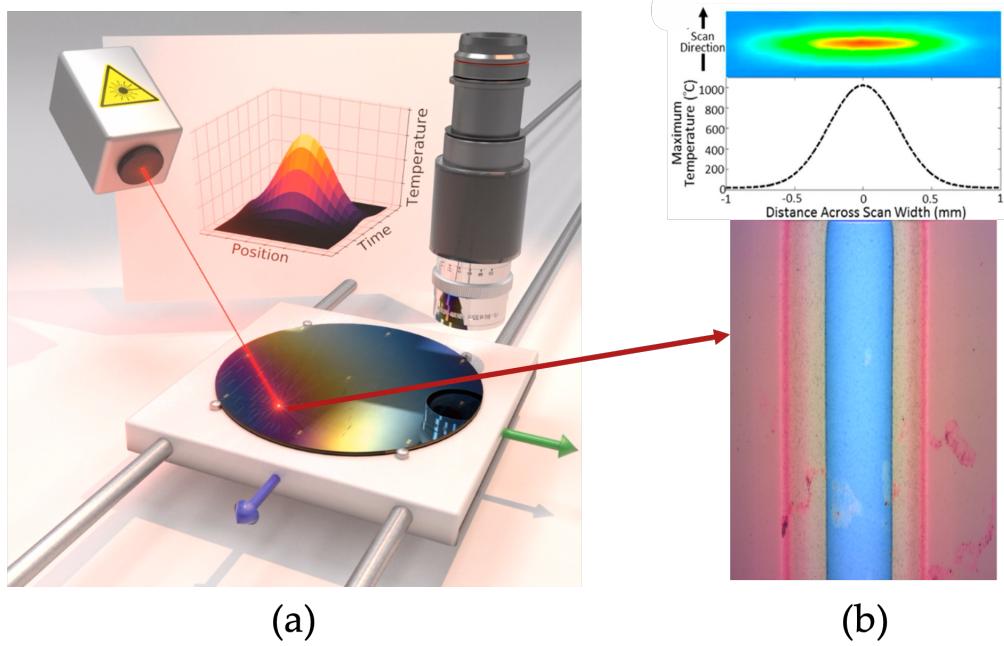


Figure 1.2: Illustration of the principle of lg-LSA. (a) Setup of lg-LSA. A scanned laser heats the material with a peak temperature gradient on the material. Lg-LSA conditions can be varied across the wafer. A camera is used to inspect the optical image of the LSA. (b) Bottom: An optical image of a stripe after lg-LSA. Top: Temperature heatmap of the laser and Gaussian profile of the peak temperature gradient.

1.4 Transparency Characterization

With different phases presented using lg-LSA, the next step is characterizing the electrical and optical properties. This project focuses on the transparency characterization during in-situ anneals. To accurately determine the optical properties of a material, the most definitive method is ellipsometry[12], which measures the change in polarization as an incident light reflects or transmits through a material. The measured data are fitted to theoretical functions to obtain the underlying complex refractive index n and k as a function of wavelength and film

thickness d . While widely used in materials characterization, the technique is incompatible with the lg-LSA high-throughput experiments. First, the size of the light spot on the surface of the thin film is too large, with typical spot sizes of 1 to 10 millimeters. For lg-LSA characterization, the probe must be less than $10\mu m$ to access phases within each stripe. Second, the complex optics of ellipsometry would be challenging to integrate into an lg-LSA setup for full automation.

A more straightforward method is to use a reflectance spectrometer[13]. This device utilizes an illumination fiber to direct light onto the sample, a reflectance fiber to collect the reflected light, and a spectrometer to analyze the data. This approach offers significant advantages over ellipsometry. First, the region of the sample probed can be readily focused to below $10\mu m$, which is sufficiently small to accurately measure each phase within an lg-LSA stripe. Additionally, the fibers' compact nature allows easy integration into an lg-LSA system. These features make the reflectance spectrometer an ideal choice for the precise optical characterization required in this project.

Using a reflectance spectrometer, we must still determine the optical properties n, k as a function of the wavelengths λ , and also the thickness of the films d using models. For convenience, we represent n, k as functions of the photon energies E instead of λ using

$$E = \frac{hc}{\lambda} = \frac{1240[\text{eV} \cdot \text{nm}]}{\lambda} \quad (1.5)$$

where E is the photon energy, h is Planck's constant ($6.626 \times 10^{-34} J \cdot s$), and c is the speed of light ($3.000 \times 10^8 m/s$). In this way, n, k can be represented as $n(E)$, $k(E)$ with the energy typically given in eV (≈ 1.5 to 3.0 eV for visible light).

The structure of the multilayer thin film characterized in this project is illustrated in figure 1.3 (a), where $n(E)$, $k(E)$, and d represent the optical properties of the top layer. The interaction of light with this multilayer structure results in constructive and destructive interference, leading to characteristic features in the reflectance spectra, often observed as 'peaks' and 'valleys'. These features are depicted in figure 1.3 (b).

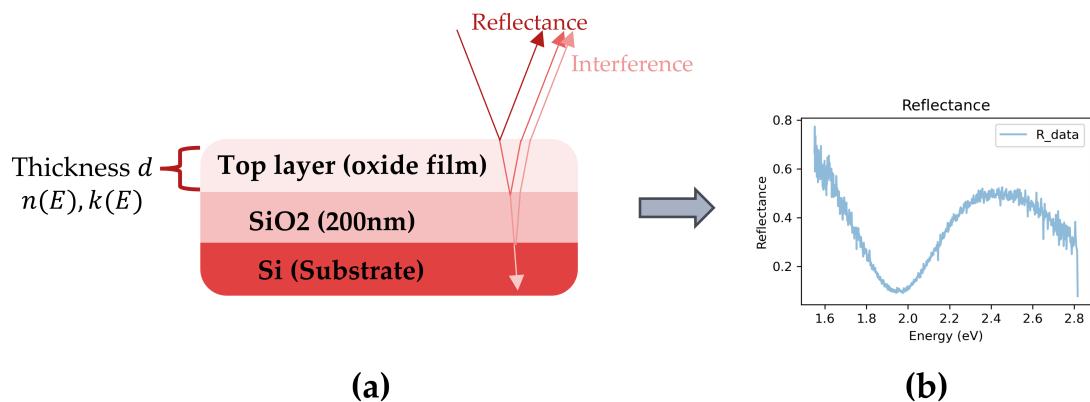


Figure 1.3: (a) Structure of the typical multilayer sample. (b) Example of a single reflectance spectrum.

Figure 1.4 illustrates how the reflectance spectrometer measurement has been integrated into the lg-LSA system. The grey dash line in figure 1.4 (a) represents the scan position when reflectance data was collected, with the measured reflectance values shown in figure 1.4 (b) and figure 1.4 (c).

1.5 Extracting Optical Properties From Reflectance

In reflectance spectroscopy, both $n(E)$, $k(E)$, and thickness d at any location on the thin film must be determined from the reflectance measurement $R(E)$. This

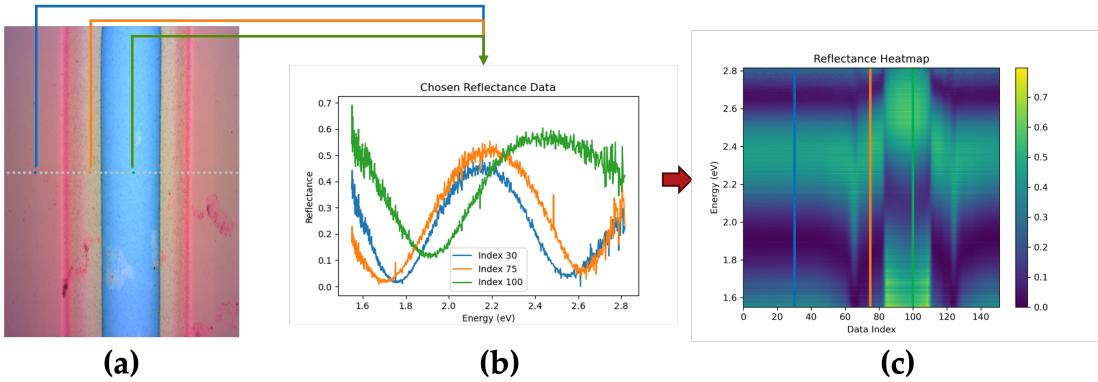


Figure 1.4: Reflectance measurement and heatmap: (a) Optical image of the LSA stripe. The grey dashed line represents the position of the reflectance measurements. Three representative measurements (in three different phases) are shown in (b). (c) Reflectance heatmap, where the x-axis is the position across the stripe (represented as the data index), and the y-axis is the photon energy E . The color represents the absolute reflectance.

determination is done by fitting the calculated reflectance to the experimentally measured reflectance. To calculate the theoretical reflectance spectra for a multilayer thin film system, the transfer matrix method (TMM) is employed[14]. TMM is a computational technique used to model the optical properties of multi-layered structures. This method represents layers and layer boundaries by matrices with entries calculated from the n , k , and d values of each layer, which are then multiplied into a matrix M . The overall reflectance is then determined from the entries of M . The curves representing $n(E)_{fit}$, $k(E)_{fit}$ and d_{fit} of the top layer are adjusted to minimize the difference between $R(E)_{calculated}$ and $R(E)_{experimental}$ using some form of a least-squares algorithm.

Figure 1.5 shows how $n(E)$, $k(E)$, and d are determined from the reflectance measurement. Figure 1.5(a) shows a common multi-layer thin film system with a silicon substrate and a 200 nm SiO_2 layer. The top layer can be a deposited

TCO or other transparent thin film. Figure 1.5(b) depicts the variation of n and k with the photon energy of the incident beam. Theoretical reflectance can be calculated from each layer's n , k , and d values using the TMM. Then, with initial guesses for n , k , and d , one can determine more accurate n , k , and d by fitting the theoretical reflectance to the measured reflectance data, as depicted in Figure 1.5 (c).

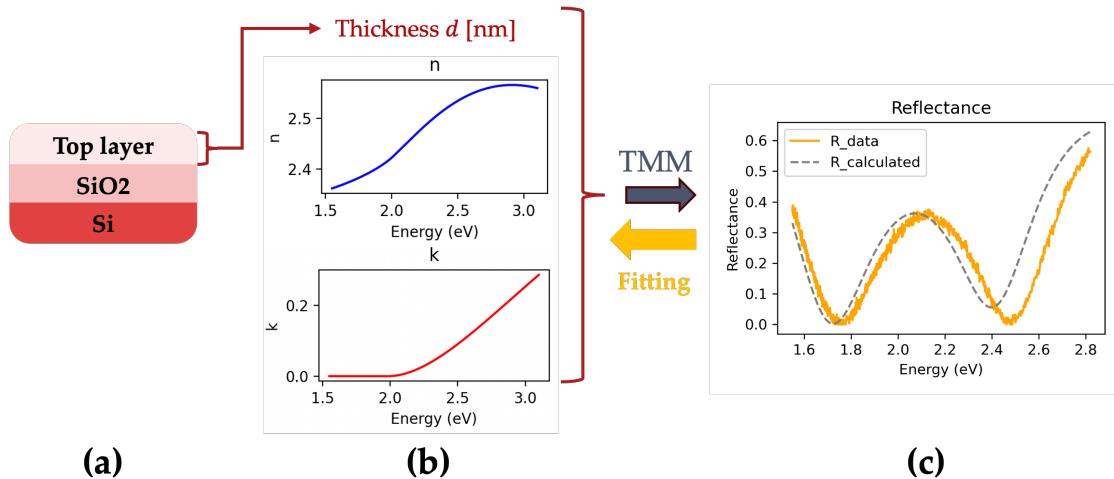


Figure 1.5: Determining transparency from reflectance measurement: (a) Common multi-layer thin film system; (b) Variation of n and k with photon energy of the incident beam. (c) The theoretical reflectance is calculated using the TMM, and n , k are modified by fitting the theoretical reflectance to the measured reflectance data.

However, this fitting approach faces the challenge of being an 'ill-posed problem'[15], which means that there are potentially more than one solution. In the context of this project, it means that there might be multiple combinations of n , k , d that can result in similar R . Moreover, an ill-posed problem typically also implies that small changes in starting conditions may lead to very different outcomes, requiring the initial guesses of n , k , and d to be carefully chosen; if not, the fitting process can easily fall into local minima instead of global minimum

and producing inaccurate results.

Traditionally, initial guesses of n and k are based on the materials before any heat treatment and stored as a database[16]. However, this approach cannot guarantee good initial guesses for phases present after annealing, and obtaining n and k measurements for all material phases is challenging. Therefore, a method for producing good initial guesses for each measurement is crucial.

1.6 Convolutional Neural Network (CNN) Approach

To address these drawbacks and produce good initial guesses, this project explores the application of machine learning, specifically convolutional neural networks (CNNs), known for their effectiveness in pattern recognition within spectra. It has been proven useful in many materials and research applications. For example, Szymanski et al. demonstrated the efficacy of CNN in interpreting the multi-phase X-ray diffraction spectra of solid state materials[17]. Ziatdinov et al. also showed that CNN is helpful in analyzing electron and scanning probe microscopy images[18].

In the area of applying CNN to extract the optical properties, a previous study has demonstrated CNNs' potential to decipher n , k , and d values directly from data[19]. However, there are drawbacks to this study's approach. First, it relies on both reflectance and transmission data, whereas our setup is limited to reflectance measurements. Also, their study generated reflectance data on a confined wavelength range as the training data, which means it can only pre-

dict n, k over this range, limiting its ability to predict n, k outside of this range. Moreover, the n, k predicted by their algorithm does not necessarily satisfy the Kramers–Kronig (KK) relations[20][21], a mathematical relationship between n and k . These drawbacks lead to significant concerns regarding the physicality and reliability of these predictions.

This project proposes an innovative methodology utilizing convolutional neural networks (CNNs) to analyze constrained reflectance data by employing a Tauc-Lorentz (TL) model for optical properties[22], a well-established physical model for amorphous semiconductors at photon energies exceeding their optical bandgap. This methodology addresses the limitations above and enhances measurement accuracy. Initially, a training dataset of the reflectance, R_{true} , is generated by selecting TL parameters and thicknesses d (collectively termed as 'parameters') within specified ranges, with these reflectance curves serving as inputs to predict the corresponding TL parameters and thicknesses d that produced these reflectance curves. Subsequently, the predicted reflectance R_{pred} is computed utilizing these predicted parameters to provide a robust initial guess for subsequent fitting processes. When the CNN prediction attains sufficient accuracy to render a reliable initial guess, a least-squares fitting method, specifically the trust region reflective (TRF) algorithm, is employed to reconcile the calculated reflectance with the experimental reflectance. By refining the parameters of the TL model and d , a refined fit of the reflectance spectra is achieved, thereby enabling precise determination of the refractive index n and extinction coefficient k , essential for determining material transparency. Figure 1.6 depicts the workflow of this methodology.

To summarize, the primary goal of this project is to address the ill-posed reverse problem of extracting the refractive index curve (n), the extinction coefficient curve (k), and the film thickness (d) from reflectance measurements. The objective is to produce accurate initial guesses for these parameters to minimize the discrepancy between calculated reflectance using the transfer matrix method (TMM) and experimental reflectance. The Tauc-Lorentz (TL) model is employed to model the optical properties (n and k) of the top layer in a multi-layer system. Concurrently, a convolutional neural network (CNN) is utilized to predict initial values of n , k , and d . This approach is pivotal in advancing the multi-objective optimization of materials and significantly enhances the capabilities of autonomous experimental studies.

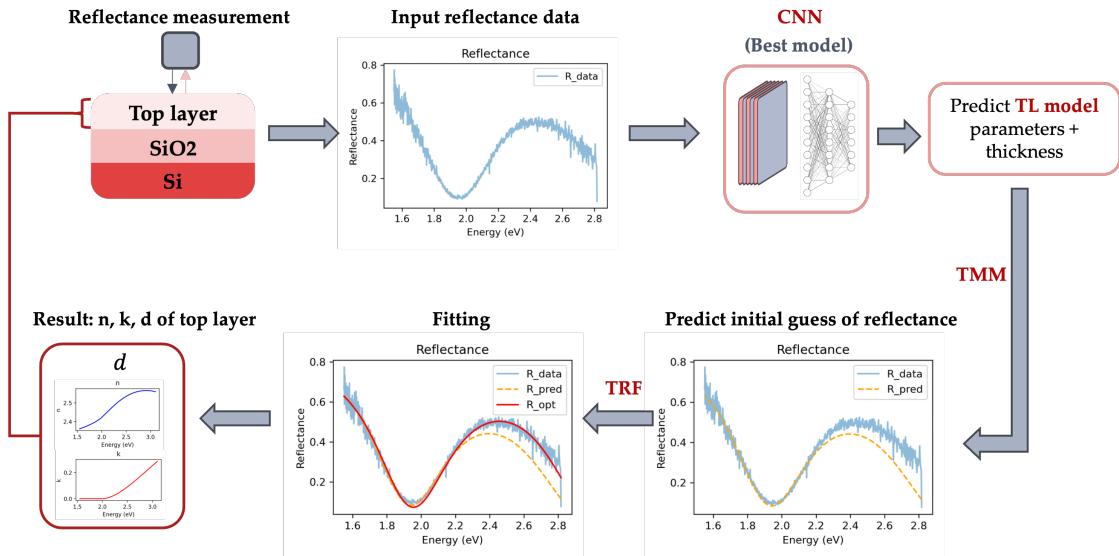


Figure 1.6: Workflow of the project. The measured reflectance data serves as an input of a CNN model and predicts initial parameters for the reflectance. Then, the initial reflectance goes through further fitting to improve the accuracy. The n, k, d associated with this reflectance is a robust estimate of the true optical properties and thickness of the top layer.

In the next chapter, several methods in this workflow, including reflectance

measurement, CNN, TL model, TMM, and TRF algorithm, will be introduced.

CHAPTER 2

METHODS

2.1 TCO Sample and lg-LSA Setup

While the ultimate goal of this work is optimizing the properties of transparent conducting oxides (TCOs), the specific focus is on determining the optical properties, particularly absorption in the visible band, through measurements of the reflectance on an absorbing Si substrate. As an initial test film, we chose a system extensively studied in our lab, which exhibits multiple phase transitions under the lateral gradient laser spike annealing (lg-LSA) conditions. The samples for the initial study were Bi_2O_3 thin films with thicknesses of about 100 nm. The Bi_2O_3 was deposited on silicon wafers with a 200 nm SiO_2 barrier layer, creating a three-layer multilayer thin film system with Si as the substrate, SiO_2 as a buffer, and Bi_2O_3 as the top layer. For initial explorations, the lg-LSA was set to a center temperature of 1080°C with a dwell time of 2500 ms. Figure 1.2 (b) in the introduction section shows the optical image of this stripe, exhibiting distinct boundaries between the phases, which are also clearly evident in the reflectance analysis.

2.2 Reflectance Measurement

Reflectance measurements were obtained using a reflectance spectrometer, the operational principle of which is illustrated in Figure 2.1. White light was di-

rected onto the sample surface through an illumination 1mm diameter fiber and a 10x or 20x objective lens. Incident light reflected off the sample was imaged into a reflectance fiber with a spatial resolution of $\approx 10\mu m$. The collected light was then analyzed by a spectrometer covering the 300nm to 850nm range. The range of incident angles was relatively small and was approximated as normal incidence. This approximation significantly simplifies the modeling of the problem by eliminating the need to account for angular variations in reflectance, thereby facilitating a more straightforward interpretation and analysis of the optical properties.

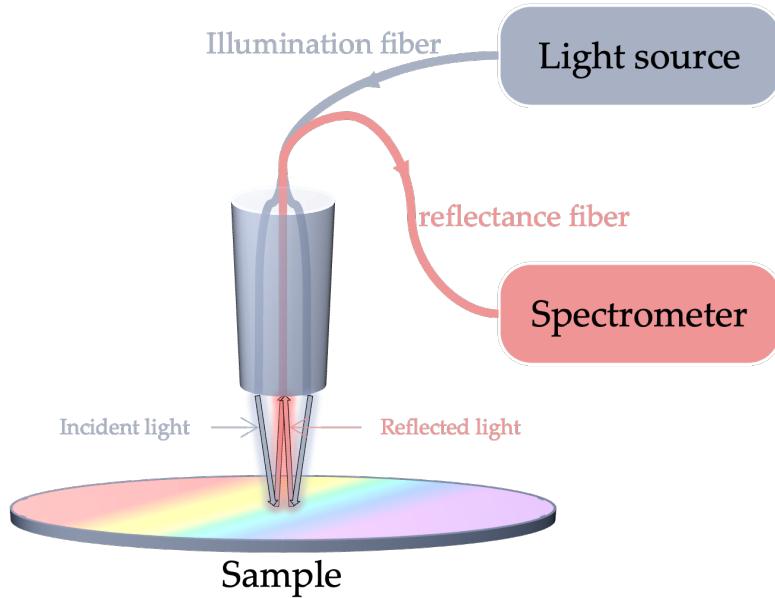


Figure 2.1: Illustration of the principle of reflectance measurement.

To obtain absolute reflectance values, we first measured the intensity of the reflected light as a function of wavelength λ without a sample in the path. This measurement, denoted as $I(\lambda)_{\text{blank}}$, accounts for light reflected from optical elements within the system. Next, we measured the spectrum of a bare silicon

wafer, known for its well-characterized reflectance $R_{\text{Si}}(\lambda)$, to serve as a calibration reference. This data is noted as $I(\lambda)_{\text{Si}}$. The absolute reflectance $R(\lambda)$ of the sample was then calculated using equation 2.1:

$$R(\lambda) = \frac{I(\lambda)_{\text{measured}} - I(\lambda)_{\text{blank}}}{I(\lambda)_{\text{Si}} - I(\lambda)_{\text{blank}}} \cdot R_{\text{Si}}(\lambda) \quad (2.1)$$

where $I(\lambda)_{\text{measured}}$ is the intensity of the reflected light with the sample in the path.

In this project, reflectance measurements were conducted within the wavelength range of $\lambda = 440 \text{ nm}$ to $\lambda = 800 \text{ nm}$, corresponding to photon energies of 1.55 eV to 2.81 eV. Initially, the data range was set between 400 and 800 nm (1.55 eV to 3.1 eV), encompassing the visible spectrum. However, it was observed that the experimental data within the 400 to 440 nm (2.81 eV to 3.1 eV) was excessively noisy and thus unsuitable for reliable data analysis. Consequently, the data range was cropped to exclude this noisy portion. Despite this adjustment, the selected range from 440 nm to 800 nm still adequately covers the visible spectrum from blue light to red light, making it particularly relevant for investigating the optical properties of thin films. This refined range ensures the acquisition of high-quality data necessary for accurate analysis and characterization of the materials.

151 measurements were conducted across the stripe, with a spacing of $10 \mu\text{m}$ between each measurement point. This dense sampling ensures a detailed and accurate representation of the reflectance properties across the stripe.

2.3 Transfer Matrix Method (TMM)

The transfer matrix method (TMM) is commonly utilized to analyze the optical properties of multilayer thin films[14]. In this approach, \vec{F}_1^+ and \vec{F}_1^- denote the electric field vectors of the forward and backward traveling waves at the first interface, while \vec{F}_{m+1}^+ and \vec{F}_{m+1}^- represent these electric field vectors at the last interface (m being the total number of layers). The following matrix equation describes the transformation through the multilayer stack:

$$\begin{pmatrix} \vec{F}_1^+ \\ \vec{F}_1^- \end{pmatrix} = \mathbf{M} \begin{pmatrix} \vec{F}_{m+1}^+ \\ \vec{F}_{m+1}^- \end{pmatrix} = \mathbf{B}_{0,1} \cdot \prod_{j=1}^m \mathbf{P}_j \mathbf{B}_{j,j+1} \begin{pmatrix} \vec{F}_{m+1}^+ \\ \vec{F}_{m+1}^- \end{pmatrix} \quad (2.2)$$

where \mathbf{M} is the total transfer matrix. $\mathbf{B}_{0,1}$ accounts for the boundary conditions at the interface between layer 0 (air) and layer 1 (the top layer), and $\mathbf{B}_{j,j+1}$ accounts for the conditions at the interfaces between consecutive layers j and $j+1$. \mathbf{P}_j is the propagation matrix for the j -th layer, encapsulating the effects of the layer's thickness and material properties on the phase and amplitude of the vector fields traveling through it.

Figure 2.2 illustrates the application of the TMM method to the multilayer thin films in this project. Note that $m = 2$ here since only two layers (the top layer of Bi_2O_3 and the intermediate SiO_2 layer) are considered for the light propagation calculation.

For a given layer, the propagation matrix \mathbf{P} for light at normal incidence is

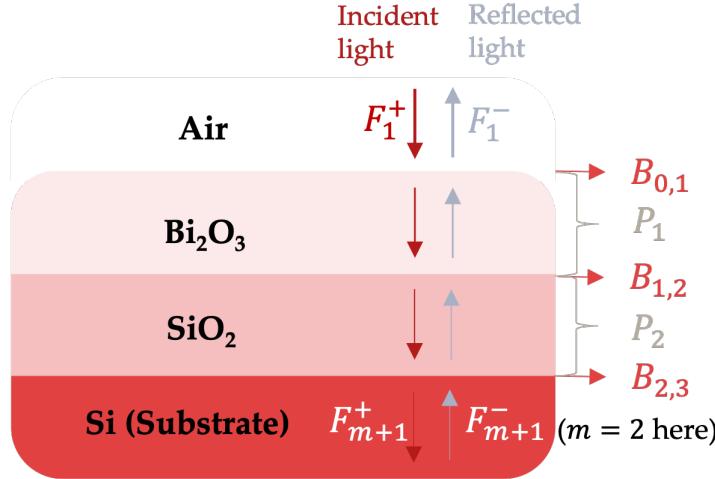


Figure 2.2: Illustration of the transfer matrix method (TMM) applied to the multilayer thin films in this project.

given by:

$$\mathbf{P} = \begin{pmatrix} e^{-i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix} \quad (2.3)$$

where $\phi = \frac{2\pi\tilde{n}d}{\lambda}$ is the phase thickness of the layer, \tilde{n} is the complex refractive index of the material ($\tilde{n} = n - ik$), d is the thickness of the layer, and λ is the wavelength of light in the medium.

The boundary matrix \mathbf{B} between two adjacent layers (say layer j and layer $j + 1$) with refractive indices \tilde{n}_j and \tilde{n}_{j+1} , respectively, is given by:

$$\mathbf{B} = \frac{1}{2\tilde{n}_j} \begin{pmatrix} \tilde{n}_j + \tilde{n}_{j+1} & \tilde{n}_j - \tilde{n}_{j+1} \\ \tilde{n}_j - \tilde{n}_{j+1} & \tilde{n}_j + \tilde{n}_{j+1} \end{pmatrix} \quad (2.4)$$

After computing \mathbf{P} and \mathbf{B} matrices, we can multiply them together to get \mathbf{M} . To calculate reflectance, we noticed that since equation 2.2 can also be written as:

$$\vec{F}_1^+ = M_{1,1}\vec{F}_{m+1}^+ + M_{1,2}\vec{F}_{m+1}^- \quad (2.5)$$

$$\vec{F}_1^- = M_{2,1}\vec{F}_{m+1}^+ + M_{2,2}\vec{F}_{m+1}^- \quad (2.6)$$

Considering that after crossing the last layer, the wave does not undergo any further reflection ($\vec{F}_{m+1}^- = 0$), the reflection coefficient for waves traveling from the air to the substrate can be derived from the above relations:

$$\tilde{r}^+ = \frac{\vec{F}_1^-}{\vec{F}_1^+} = \frac{M_{2,1}}{M_{1,1}} \quad (2.7)$$

The absolute reflectance R of the multilayer system is the modulus squared of the reflection coefficient given by:

$$R = |\tilde{r}^+|^2 \quad (2.8)$$

Knowing n, k and d of each layer allows the absolute reflectance R to be calculated.

2.4 Tauc-Lorentz (TL) Model

The Tauc-Lorentz (TL) model was developed as a mathematical formula to model the refractive index (n) and extinction coefficient (k) of amorphous semiconductors at photon energies more significant than their optical band gap[22]. The model has five adjustable parameters of an 'oscillator', each of which plays a specific role in characterizing the material's optical properties:

- **Amplitude (A):** This parameter determines the strength of the oscillator, representing the peak value of the imaginary part of the dielectric function, and hence reflects the intensity of the absorption.
- **Resonance Energy (E_0):** The resonance energy of the electronic transition and it is the photon energy at which the peak absorption occurs.

- **Broadening Parameter (G):** Also known as the damping factor, this parameter describes the width of the absorption peak, accounting for the broadening effects due to various scattering mechanisms and the lifetime of excited states.
- **Bandgap Energy (E_g):** This is the optical band gap energy, below which the material is transparent to photons. It defines the energy threshold above which significant absorption begins.
- **High-Frequency Dielectric Constant (ϵ_∞):** The high-frequency dielectric constant represents the contribution to the dielectric function from electronic polarizability at photon energies well above the band gap. It is a measure of the material's response to high-frequency electric fields.

These parameters are used to calculate the real and imaginary parts of the dielectric function, denoted as $\epsilon_{\text{re}}(E)$ and $\epsilon_{\text{im}}(E)$ respectively. Then, the real and imaginary parts of the refractive index, denoted as $n(E)$ and $k(E)$, can be calculated from $\epsilon_{\text{re}}(E)$ and $\epsilon_{\text{im}}(E)$. The real part of the dielectric function is given by

$$\begin{aligned}\epsilon_{\text{re}}(E) = \epsilon_\infty + & \frac{AGa_{\ln}(E)}{2\pi\kappa(E)^4\alpha E_0} \ln\left(\frac{E_0^2 + E_g^2 + \alpha E_g}{E_0^2 + E_g^2 - \alpha E_g}\right) \\ & - \frac{Aa_{\text{atan}}(E)}{\pi\kappa(E)^4 E_0} \left(\pi - \arctan\left(\frac{2E_g + \alpha}{G}\right) + \arctan\left(\frac{\alpha - 2E_g}{G}\right) \right) \\ & + \frac{4AE_0 E_g (E^2 - \gamma^2)}{\pi\kappa(E)^4 \alpha} \left(\arctan\left(\frac{\alpha + 2E_g}{G}\right) + \arctan\left(\frac{\alpha - 2E_g}{G}\right) \right) \\ & - \frac{AE_0 G (E^2 + E_g^2)}{\pi\kappa(E)^4 E} \ln\left(\frac{|E - E_g|}{E + E_g}\right) \\ & + \frac{2AE_0 G E_g}{\pi\kappa(E)^4} \ln\left(\frac{|E - E_g|(E + E_g)}{\sqrt{(E_0^2 - E_g^2)^2 + E_g^2 G^2}}\right),\end{aligned}$$

and the imaginary part is defined by

$$\varepsilon_{\text{im}}(E) = \begin{cases} \frac{AE_0G(E-E_g)^2}{E((E^2-E_0^2)^2+G^2E^2)}, & \text{if } E > E_g \\ 0, & \text{otherwise.} \end{cases}$$

The real and imaginary parts of the refractive index are given by

$$n(E) = \sqrt{0.5 \left(\sqrt{\varepsilon_{\text{re}}(E)^2 + \varepsilon_{\text{im}}(E)^2} + \varepsilon_{\text{re}}(E) \right)}, \quad (2.9)$$

and

$$k(E) = \sqrt{0.5 \left(\sqrt{\varepsilon_{\text{re}}(E)^2 + \varepsilon_{\text{im}}(E)^2} - \varepsilon_{\text{re}}(E) \right)}, \quad (2.10)$$

respectively. Auxiliary functions include:

$$a_{\ln}(E) = (E_g^2 - E_0^2)E^2 + E_g^2G^2 - E_0^2(E_0^2 + 3E_g^2)$$

$$a_{\text{atan}}(E) = (E^2 - E_0^2)(E_0^2 + E_g^2) + E_g^2G^2$$

$$\kappa(E) = \sqrt[4]{((E^2 - \gamma^2)^2 + 0.25\alpha^2G^2)}$$

where

$$\gamma = \sqrt{\max(E_0^2 - 0.5G^2, 0)}$$

$$\alpha = \sqrt{\max(4E_0^2 - G^2, 0)}$$

Figure 2.3 illustrates how the parameters of the Tauc-Lorentz model influence the dielectric function. The real part of the dielectric function, $\varepsilon_{\text{re}}(E)$, and the imaginary part, $\varepsilon_{\text{im}}(E)$, are plotted in light blue and red, respectively. The values of TL parameters are $A = 100\text{eV}$, $E_0 = 4.5\text{eV}$, $G = 1\text{eV}$, $E_g = 3.5\text{eV}$, and $\varepsilon_\infty = 3.5$.

As depicted in the figure, A is proportional to the peak height of $\varepsilon_{\text{im}}(E)$, E_0 is the central peak position, and G represents the width of the peak. The optical

bandgap E_g marks the energy at which $\varepsilon_{\text{im}}(E)$ starts to rise. On the right, the value of ε_∞ is shown, representing the high-energy limit of the real part of the dielectric function.

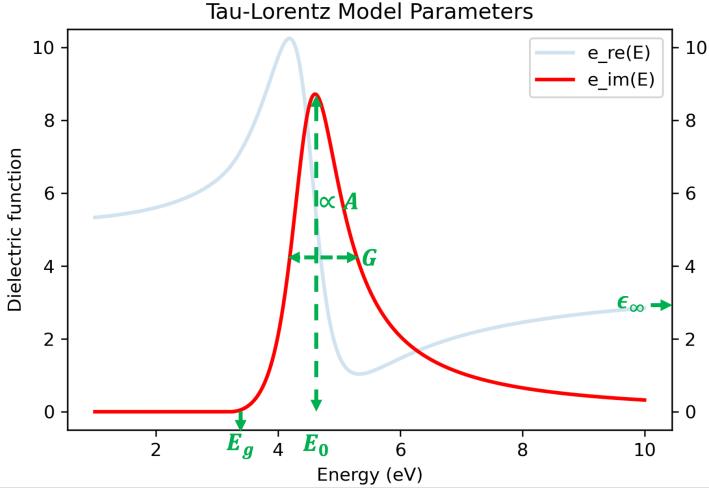


Figure 2.3: The dielectric function, $\varepsilon(E)$, of the Tauc-Lorentz model as a function of photon energy. The real part $\varepsilon_{\text{re}}(E)$ (light blue) and the imaginary part $\varepsilon_{\text{im}}(E)$ (red) are shown, with the key parameters of the model annotated.

2.5 Generating Training Data

With the Transfer Matrix Method (TMM) and Tauc-Lorentz (TL) model established, we can generate a simulated reflectance dataset for training the Convolutional Neural Network (CNN) model. The training data is generated by selecting TL parameters and thickness d within specific ranges. The ranges for the TL parameters were determined as follows:

- $A \in [1 \times 10^{-6}, 100]$ and $\varepsilon_\infty \in [1, 10]$: These ranges are chosen based on empirical data and previous research findings[23].

- $E_g \in [0.5, 3.2]$: This range is selected as it represents the typical band gap range of TCOs and includes the expected band gap range of Bi_2O_3 [24].
- $E_0 \in [3.2, 7]$: The upper limit is informed by previous research[23], while the lower limit is set for two reasons. First, the TL model dictates that E_0 must be larger than E_g . Second, it was observed that if both E_0 and E_g fall within the energy range of 1.55 to 3.1 eV (before excluding the noisy data in the 2.81 eV to 3.1 eV range), peaks appear in the n and k curves with values up to 90, causing exponential calculations in equation 2.3 to overflow. Hence, to ensure E_0 is greater than E_g and outside the problematic energy range, E_0 is set to start at 3.2 eV.
- $G \in [0, \sqrt{2}E_0]$: This range is based on the TL model, which requires G to be greater than 0 and less than $\sqrt{2}E_0$.

The thickness range for the samples is set to $d \in [10, 150]$ nm, reflecting the potential thickness range of our test samples.

Latin Hypercube Sampling (LHS)[25] was used to sample parameters within these ranges. Latin Hypercube Sampling is a statistical method used to generate a distribution of plausible collections of parameter values from a multidimensional distribution. This method ensures that each parameter is uniformly sampled over its range by dividing the range of each parameter into N intervals, where N is the total number of generated data, and then sampling once from each interval. The multidimensional sampling result is a random combination of the parameters sampled in each dimension. Compared to simple random sampling, LHS enhances the coverage of the sample space compared to random sampling, thereby improving the accuracy and reliability of the results.

Figure 2.4 illustrates how LHS works. Each row and column in the grid represents an interval for a parameter, and only one sample is taken from each row and each column. This sampling method ensures an even distribution of the samples across the parameter space.

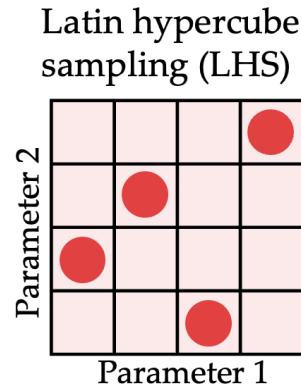


Figure 2.4: Illustration of Latin Hypercube Sampling. Each row and column has only one sampled combination of parameter 1 and parameter 2, ensuring the evenness of the sample.

After sampling N sets of TL parameters, the same number of n and k curves are also generated, and with the sampled d , N simulated reflectance can be calculated using transfer matrix method. To mimic experimental reflectance curves, Gaussian noise is added to the reflectance:

$$\tilde{R} = R + \epsilon \quad (2.11)$$

where \tilde{R} is the noisy reflectance data, R is the original reflectance data, and ϵ is the Gaussian noise. We can adjust the Gaussian noise's standard deviation σ as the noise level, which is set to be 0.01 here.

Figure 2.5 illustrates the workflow of the data generation.

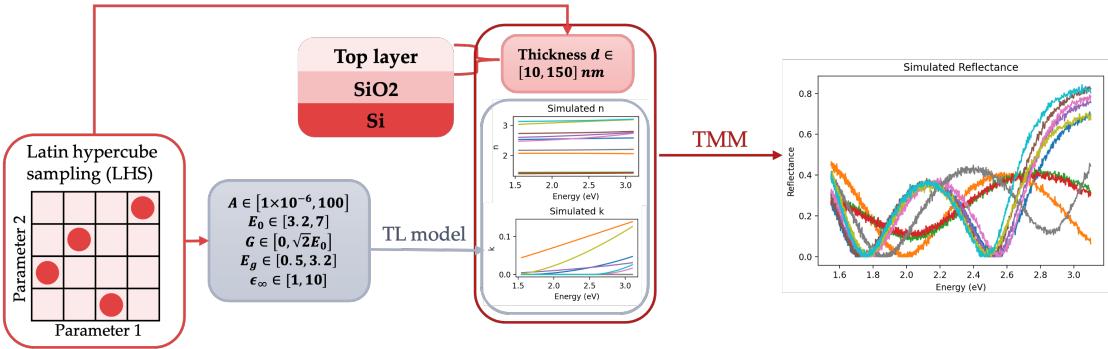


Figure 2.5: Workflow illustrating the data generation process using the TL model. Parameters are sampled via Latin hypercube sampling (LHS) to produce n, k values for the top layer, which, along with thickness d , are used in the transfer matrix method (TMM) to simulate noisy reflectance data. An example of 10 randomly selected reflectance datasets is shown.

2.5.1 Training Set and Testing Set

The data generation workflow creates both the training and testing sets. The reflectance values serve as the input values X , while the parameters $A, E_0, G, E_g, \epsilon_\infty$, and d that generated these reflectance values are the label values y . Training sets with 10,000, 20,000, and 50,000 samples were generated to evaluate the impact of dataset size on model accuracy. Additionally, training sets with input dimensions of 1,000, 500, and 200 points were created to investigate the effect of down-sampling on model accuracy. For experimentally measured reflectance, the input dimension here refers to the number of points sampled from the original data. The original reflectance data are first linearly interpolated. Then, different numbers of points within the wavelength range [1.55, 2.81] eV are linearly sampled (this energy range was selected by observing the wavelength range of experimentally measured data where the noise is less significant). This approach ensures that the corresponding wavelengths are

consistent across different datasets.

For the testing set, a dataset with 2,000 samples was generated within the same parameter ranges as the training set. Different models were evaluated on this consistent testing set to ensure fair comparisons. It is important to note that all label values are normalized to the range [0, 1] to improve convergence speed and prevent numerical instability during training:

$$y_{\text{norm}} = \frac{y - y_{\min}}{y_{\max} - y_{\min} + \epsilon} \quad (2.12)$$

where y represents the original label values, and y_{\min}, y_{\max} are the minimum and maximum values of the labels, respectively. A small constant $\epsilon = 10^{-8}$ was added to the denominator to prevent division by zero.

The CNN model was initially tested on the simulated testing set to evaluate its performance. Once the results were satisfactory, the model was subsequently applied to an experimental dataset to assess its real-world applicability and robustness. By first validating the model on simulated data, we could identify and address any potential issues in a controlled environment before moving to experimental data, thus enhancing the overall reliability and validity of the model's predictions.

2.6 Convolutional Neural Network (CNN)

A Neural network is a machine learning technique that mimics the structure of the human brain to process complex data inputs[26]. It consists of interconnected layers of nodes, or “neurons,” where each neuron in one layer is con-

nected to several neurons in the next layer. The connections between these neurons have associated weights w and biases b that are learned during training by a technique called backpropagation. By passing data through these layers, a neural network can learn the weights and biases without explicit programming, enabling quick and accurate prediction based on input data.

Figure 2.6 illustrates a simple neural network with two neurons in the input layer, two in the hidden layer, and one in the output layer. In the figure, each input x goes through some computation, including the linear computation $wx + b$, and non-linear computation $\sigma(wx + b)$, where σ is the activation function that introduces non-linearity and allows the model to learn non-linear functions. A commonly used activation function is called **ReLU**, or denoted as σ in equation 2.13:

$$\sigma = \max(0, x) \quad (2.13)$$

Convolutional Neural Network (CNN) is a specialized type of neural network that is particularly powerful for processing data with a grid-like structure, such as images or, in our case, spectral data. It can automatically detect important features without human supervision.

2.6.1 Architecture of CNN

A CNN consists of two main components: feature extraction layers and fully connected layers. The feature extraction layers extract the features of the input

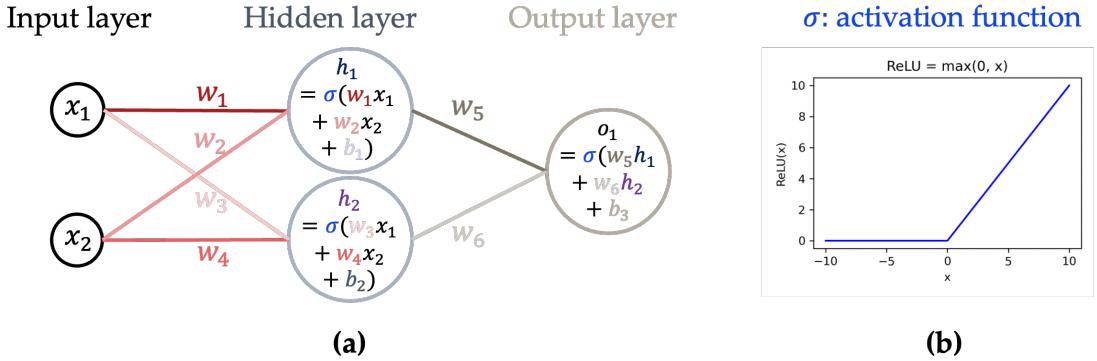


Figure 2.6: Neural Network Architecture and Activation Function. (a) A three-layer neural network comprising an input layer with two input units x_1 and x_2 , a hidden layer with two neurons h_1 and h_2 , and an output layer with one output unit o_1 . The weights w_1 to w_6 and biases b_1 to b_3 connect the layers as depicted. (b) The ReLU activation function used in the hidden and output layers, defined as $\text{ReLU}(x) = \max(0, x)$, demonstrates its effect on input values ranging from -10 to 10.

data and produce a feature map. This feature map is then fed into the fully connected layers to learn the representations within it and perform the final prediction.

The **feature extraction layers** consist of four parts:

- **Convolution:** Convolutional layers apply convolutional filters to the input reflectance spectra, creating feature maps. A certain number of randomized filters with a filter size are used to convolute with the input reflectance spectra to extract their features, which are the peaks and valleys in the spectra. The result is a feature map set that highlights different aspects of the reflectance spectra.
- **Batch Normalization:** Batch normalization is used to normalize the output of the convolutional layers using its mean and variance to improve the

training speed and stability[26].

- **Activation:** Activation functions introduce non-linearity into the network, enabling it to learn complex patterns. The activation function used in this project is the most commonly used Rectified Linear Unit (ReLU), as defined in equation 2.13
- **Pooling:** The pooling layers are used to reduce the spatial dimensions of the feature maps. Max pooling, which selects the maximum value from a patch of the feature map, is used in this project.

Figure 2.7 gives a concrete example of how these layers work. In the figure, 1D data of five elements first does convolution with a randomized filter, $[1, -1]$, to do feature extraction. The extracted feature then goes through batch normalization using the mean μ and the squared standard deviation σ^2 . Then, the activation layer uses the activation function ReLU to introduce nonlinearity. The max pooling layer chooses the max values within each pool to reduce the dimension, which results in a feature map.

After the feature extraction layers, the fully connected layers take the feature maps and perform regression. The fully connected layers consist of three parts:

- **Flatten:** This layer converts the 2D feature maps into a 1D vector since the dense layers expect a 1D input. Flattening preserves the spatial information of reflectance and prepares it for the fully connected layers.
- **Dense layers:** These are the hidden layers of the fully connected layers. In the dense layers, a technique called 'dropout' is applied to prevent overfitting, a phenomenon that the model is accurate on the training data but

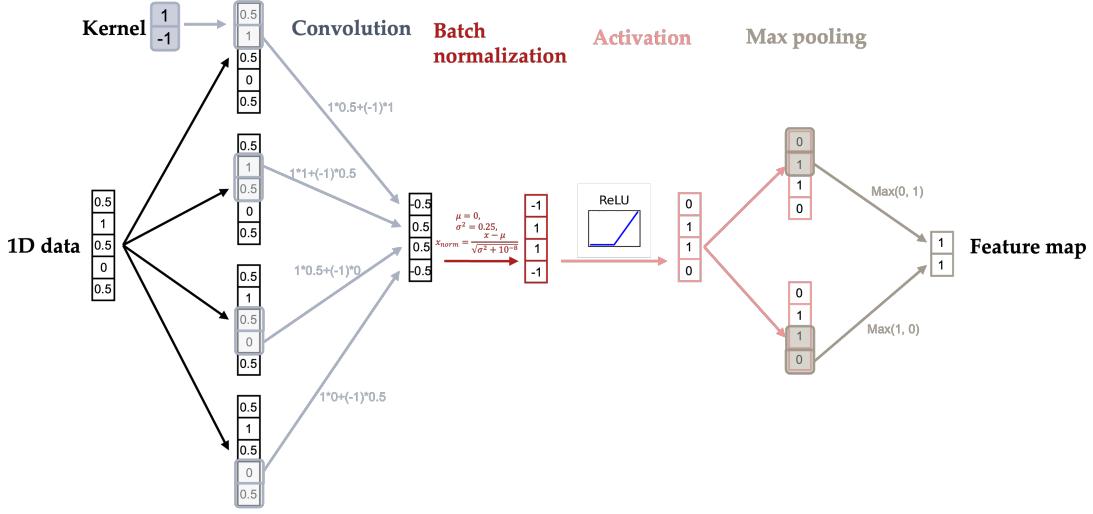


Figure 2.7: Illustration of the convolution, batch normalization, activation, and max pooling layers on a 1D data.

performs poorly on the testing data[26]. Dropout randomly sets a fraction of the input units to zero with probability p (also named the dropout rate) at each update during training, which helps to prevent the network from becoming too reliant on specific neurons and improves generalization. In this project, $p = 0.5$.

- **Output:** The output layer produces the final prediction of the network, which are the normalized TL parameters $A, E_0, G, E_g, \varepsilon_\infty$ and thickness d .

Figure 2.8 illustrates the complete architecture of the CNN used in this project. The architecture consists of feature extraction layers followed by fully connected layers. Each 1D input reflectance spectrum is initially passed through a set of 64 convolutional filters, which are randomly initialized. These filters capture different features from the input, converting the 1D input into a 2D feature map. The size of each filter varies across the feature extraction layers. The input is processed through n_{layers} feature extraction layers, each applying

a series of convolution operations followed by batch normalization, activation functions, and pooling, ultimately producing a refined feature map.

In the fully connected layers, the resulting 2D feature map is flattened into a 1D vector of length 3000 and then passed through dense layers with dimensions of 1200 and 300, with dropout (rate = 0.5) applied to mitigate overfitting. The final output layer predicts the six parameters $A, E_0, G, E_g, \varepsilon_\infty$, and d .

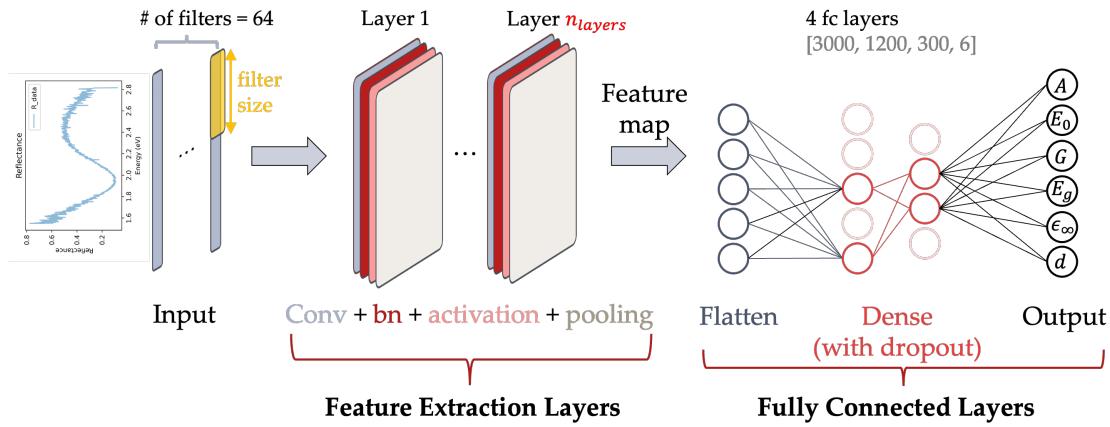


Figure 2.8: CNN architecture for predicting material optical properties from reflectance spectra. The network consists of feature extraction layers and fully connected layers for parameter prediction. The input reflectance spectrum is processed through 64 filters and n_{layers} number of feature extraction layers. The fully connected layers have dimensions of 3000, 1200, 300, and 6, with a dropout rate 0.5. The final output predicts six parameters $A, E_0, G, E_g, \varepsilon_\infty$, and d .

The feature extraction layers are critical in determining the performance of a CNN. Generally, more layers can capture more complex and detailed features from the input data, enhancing the model's ability to learn intricate patterns. However, increasing the number of layers also comes with challenges. Too many layers can lead to overfitting, where the model performs well on training data but poorly on unseen data. Additionally, deeper networks are more challenging to train due to issues like vanishing gradients and increased com-

putational complexity. Thus, finding the optimal number of feature extraction layers is crucial to balancing the model’s performance and training efficiency. In this project, we investigate the effects of using 4, 5, 6, and 7 feature extraction layers to determine the most effective architecture.

In this project, the convolutional filter sizes and max-pooling kernel sizes vary at each layer, as shown in Tables 2.1 and 2.2. The convolutional filter sizes decrease progressively across layers, allowing the model to capture finer features of the reflectance spectra. For example, the first layer uses a filter size of 150, which is about the width of a peak, while the subsequent layers use smaller filters, down to as small as 3 in the seventh layer, to detect more localized features.

Similarly, the max-pooling kernel sizes also vary between layers. The initial layers use larger pooling sizes (e.g., 3 in Layer 1 and Layer 2), which reduce the feature map dimensions significantly, while later layers utilize smaller pooling sizes (e.g., 1 in the deeper layers), maintaining finer spatial information as the network goes deeper. These progressive reductions in filter size and pooling size ensure that the CNN effectively captures global and local features, leading to better predictive performance on the reflectance data.

n_{layers}	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7
4	150	100	75	50	-	-	-
5	150	100	75	50	15	-	-
6	150	100	75	50	15	5	-
7	150	100	75	50	15	5	3

Table 2.1: Convolutional filter sizes for different numbers of layers in the CNN architecture.

<i>n_{layers}</i>	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7
4	3	3	2	1	-	-	-
5	3	3	2	1	1	-	-
6	3	3	2	1	1	1	-
7	3	3	2	1	1	1	1

Table 2.2: Max-pooling kernel sizes for different numbers of layers in the CNN architecture.

2.6.2 Training Process

During training, CNNs use **backpropagation**[26] to update the weights and biases of the network. Backpropagation involves calculating the gradient of the loss function with respect to each weight by the chain rule, iteratively updating the weights in the direction that minimizes the loss. The loss function used in this project is the mean squared error (MSE) of the normalized parameters:

$$L = \frac{1}{N} \sum_{i=1}^N [(A_i - \hat{A}_i)^2 + (E_{0i} - \hat{E}_{0i})^2 + (G_i - \hat{G}_i)^2 + (E_{gi} - \hat{E}_{gi})^2 + (\epsilon_{\infty i} - \hat{\epsilon}_{\infty i})^2 + (d_i - \hat{d}_i)^2] \quad (2.14)$$

where $A_i, E_{0i}, G_i, E_{gi}, \epsilon_{\infty i}$, and d_i are the true parameters, and $\hat{A}_i, \hat{E}_{0i}, \hat{G}_i, \hat{E}_{gi}, \hat{\epsilon}_{\infty i}$, and \hat{d}_i are the predicted parameters for the i -th sample, and N is the total number of samples. During the training process, all parameters are normalized to between [0, 1] using equation 2.12, and the predicted parameters are denormalized back during the inference stage. A smaller MSE value indicates a higher accuracy in the regression task of this project.

In backpropagation, the weights in the neural network are updated by **gradient descent** algorithm, as stated in equation 2.15:

$$w_{t+1} = w_t - \alpha \nabla_w L(w_t) \quad (2.15)$$

Where $\nabla_w L(w_t)$ denotes the gradient of the loss function with respect to the weights at that iteration t , and α represents the learning rate. This learning rate α is pivotal as it dictates the magnitude of steps taken toward the optimal solution. A small learning rate may lead to slow convergence, whereas a large learning rate risks overshooting the global minimum.

To enhance the efficiency of this process, the **Adam** optimizer[27], an advanced form of gradient descent that can dynamically adjust the learning rate, was employed. This optimizer sped up convergence and enhanced the model's performance on complex tasks. The impact of various initial values of α (e.g., 10^{-3} , 10^{-4} , 10^{-5}) was specifically investigated to determine the most effective learning rate that ensures both accuracy and optimal convergence.

Another training technique, **early stopping**[26], was deployed to prevent overfitting. Early stopping stops the model from training when the validation loss stops improving for a certain number of epochs, namely 'patience'. The patience parameter for early stopping was set to 10, meaning training stopped if the validation performance did not improve for 10 consecutive epochs. This approach helps to ensure that the model does not overfit the training data and can generalize well to new, unseen data.

2.7 Trust Region Reflective Algorithm

After the CNN prediction provided an initial guess of TL parameters and d , the Trust Region Reflective (TRF) algorithm was used to optimize the TL param-

eters and d to minimize the difference between calculated reflectance and experimental reflectance. This optimization technique is instrumental in solving nonlinear least-squares problems subject to bounds on the variables.

The TRF algorithm is developed based on the Levenberg–Marquardt (LM) algorithm[28][29]. The LM algorithm combines the concepts of gradient descent and the Gauss–Newton method to minimize the sum of squared residuals iteratively. The update rule for the LM algorithm can be written as:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{J}^\top \mathbf{J} + \lambda \mathbf{I})^{-1} \mathbf{J}^\top \mathbf{r} \quad (2.16)$$

Where \mathbf{x}_k is the current estimate of the parameter vector, \mathbf{J} is the Jacobian matrix of partial derivatives of the residuals, λ is the damping factor, \mathbf{I} is the identity matrix, and \mathbf{r} is the vector of residuals.

The LM algorithm is effective for solving unconstrained nonlinear least-squares problems. Still, it does not handle variable bounds, so it cannot enforce constraints to ensure that the solution remains within specified ranges. On the other hand, the TRF algorithm extends the LM algorithm by incorporating trust regions and allowing variable bounds. In the TRF approach, the optimization problem is solved iteratively within a trust region around the current estimate, and the size of this region is adjusted based on the agreement between the predicted and actual reductions in the residuals. The trust region subproblem can be formulated as follows:

$$\min_{\mathbf{p}} \|\mathbf{J}\mathbf{p} + \mathbf{r}\|^2 \quad \text{subject to} \quad \|\mathbf{p}\| \leq \Delta \quad (2.17)$$

where \mathbf{p} is the step within the trust region, and Δ is the radius of the trust region.

The TRF algorithm adjusts Δ dynamically and ensures that it is always within the variable bounds, specified as $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$. This way, the parameter vectors are ensured to be within the variable bounds and eliminate the non-physical solutions. In the context of this project, the variable bounds were the bounds for TL parameters and d when generating the simulated data.

In summary, the TRF algorithm effectively ensures physical solutions when optimizing the parameters, making the optimization results reliable.

2.8 Baseline Model

To summarize, four variables are explored to determine their effects on model accuracy: the number of feature extraction layers n_{layers} (CNN architecture parameter), the learning rate α (CNN training parameter), the size of the training data N , and the input dimension of the training data dim . The values of the variables are listed in Table 2.3:

Variable	Values
Number of feature extraction layers n_{layers}	4, 5, 6, 7
Learning rate α	$10^{-3}, 10^{-4}, 10^{-5}$
Size of the training data N	10,000, 20,000, 50,000
Input dimension of the training data dim	200, 500, 1000

Table 2.3: Explored variable values for determining model accuracy.

The model with $n_{\text{layers}} = 4$, $\alpha = 10^{-3}$, $N = 10,000$, and $dim = 1000$ is designated as the baseline model. This configuration served as the starting point for further experimentation and comparison. By systematically varying each

of these parameters, we sought to understand their individual and combined effects on the accuracy and performance of the CNN model.

CHAPTER 3

RESULTS AND DISCUSSION

In this chapter, the validity and accuracy of the CNN are evaluated using synthetic data, where training and testing data sets are generated as outlined in Chapter 2. Key CNN model parameters (n_{layers} , (α) , (N) , and dim) are systematically assessed for their impact on the model performance. The analysis begins with synthetic data to establish the best CNN model with the highest accuracy, followed by applying the trained model to experimental data, providing insight into its real-world applicability.

3.1 CNN Accuracy for Simulated Test Set

First, the baseline model is tested on the simulated testing dataset. The evaluation results on single reflectance spectra and on the whole dataset are investigated.

3.1.1 Single Reflectance Spectra

Figure 3.1 shows the CNN-determined parameters for a single simulated test reflectance spectra. In Figure 3.1 (a), the predicted parameters and the true parameters are represented in a radar chart, where the true parameters (in blue) are all normalized to the same value (0.5) to illustrate how the predicted parameters (in orange) differ from the true parameters. The real values of true

and predicted parameters are attached in the figure. From figure 3.1 (a), we can see that predicted A (oscillator strength) deviates from true A the most. In contrast, the predicted d (thickness) deviates from true d the least, indicating that the CNN model performs well in predicting thickness, but performs poorly in predicting the strength of the TL oscillator.

Next, figure 3.1 (b) and figure 3.1 (c) show the predicted n, k curves calculated using the predicted parameters (in yellow) and the true n, k calculated using the true parameters (in blue). Δ is the percentage of the largest absolute error between the predicted value and the true value with respect to the maximum of true and predicted values. It serves as an evaluation of how the predicted n and k shifts away from the true n and k . The Δ value is 6.9% for n and 38.7% for k , indicating a more accurate prediction for n value. The accuracy of n and k vary considerably with energy, deviating most strongly at short wavelengths.

Lastly, figure 3.1 (d) shows the predicted reflectance (in yellow) calculated from the predicted n, k , and d values alongside the true reflectance (in blue) calculated from the true n, k , and d values. This subplot visually demonstrates how the predicted reflectance deviates from the true reflectance, providing a direct method for evaluating the accuracy of the CNN model. Alongside this visual representation, the mean squared error (MSE) and R-squared (R^2) values (coefficient of determination) offer more quantitative metrics to indicate the accuracy of the CNN model. These metrics for the reflectance data are calculated using the following equations:

$$\text{MSE} = \frac{1}{\text{dim}} \sum_{i=1}^{\text{dim}} (R_{\text{pred},i} - R_{\text{true},i})^2 \quad (3.1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{\text{dim}} (R_{\text{pred},i} - R_{\text{true},i})^2}{\sum_{i=1}^{\text{dim}} (R_{\text{true},i} - \bar{R}_{\text{true}})^2} \quad (3.2)$$

Where $R_{\text{pred},i}$ denotes the predicted reflectance values by the CNN model, $R_{\text{true},i}$ represents the true reflectance values, \bar{R}_{true} is the average reflectance over the full energy range, and dim is the input dimension (the number of observations). A smaller MSE value implies a higher accuracy of the predictions, while an R^2 score closer to 1 indicates a model that explains a larger proportion of the variance in the reflectance data around its mean.

It is worth noticing that although R^2 is traditionally used for evaluating linear regression models[30] while reflectance spectra are non-linear, it remains a useful metric because it provides a measure of how well the model captures the overall variance in the data. The mean reflectance provides a single scalar reference point that allows for a standardized way to measure the variance, and using this mean value ensures that we have a common basis for calculating the unexplained variance, regardless of the specific shape of the reflectance data.

In figure 3.1 (d), the MSE value is 0.0068, illustrating a high level of precision in the reflectance predictions by the model. However, the R^2 value of 0.64 suggests that the model has only moderate effectiveness in capturing the variability in the reflectance data. Ongoing experimentation and model refinement, including adjustments to the number of feature extraction layers, learning rate, size of training data, and input dimensionality, are necessary to optimize model

performance and enhance predictive accuracy.

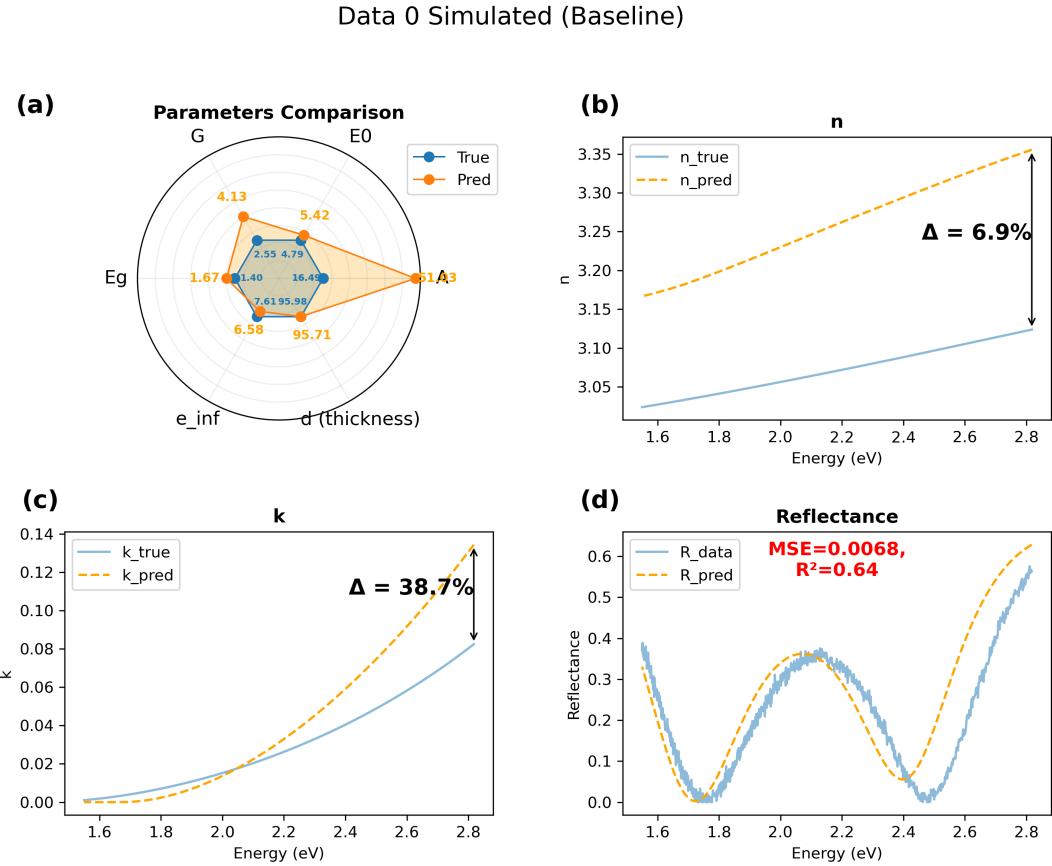


Figure 3.1: Baseline model prediction for reflectance, n , k , and parameter comparison for a randomly selected single simulated data (Data 0 means it is the first data in the dataset). (a) Radar chart of parameters comparison. (b) True and predicted n values. (c) True and predicted k values. (d) True and predicted reflectance values, with MSE and R^2 score calculated.

Before exploring the variables of the CNN model, it is important to address several unusual outcomes observed in the simulated test set. Specifically, a few test data yield negative R^2 values, prominently illustrated in figure 3.2 (d), where the mean squared error (MSE) reaches a notably high level of 0.0266, and the R^2 score goes to an extremely negative value of -97.07. Such adverse R^2 scores indicate that the model's predictions deviate dramatically from the ac-

tual data. This significant discrepancy arises because the sum of squared residuals (SS_{res}), the numerator in equation 3.2, substantially exceeds the total sum of squares (SS_{tot}), the denominator in equation 3.2 that measures the variance around the mean of the actual data.

Moreover, the prediction errors for the parameters and n, k are substantial, as depicted in figures 3.2 (a) to (c). For example, the large error in A in figure 3.2 (a) results in notable inaccuracies in the subsequent predictions of n and k shown in figures 3.2 (b) and (c), where the deviations are 51.4% and 99.6%, respectively.

These significant errors likely stem from the model's inability to accurately capture the intricate patterns within the reflectance data, particularly the non-obvious 'peaks' and 'valleys' highlighted in figure 3.2 (d). Data with such large errors are critical outliers disproportionately affecting the overall model performance. Enhancing the model to reduce the occurrence of these outliers is essential for improving its predictive accuracy and reliability.

In the following subsections, the effects of CNN variables, including the number of feature extraction layers (n_{layers}), the learning rate (α), the size of training data (N), and the input dimension (dim) on the overall model performance are discussed.

3.1.2 Number of Feature Extraction Layers

For the simulated test set with 2,000 data, the effect of the number of layers (n_{layers}) is investigated first. Figure 3.3 illustrates the box plots of the mean

Data 404 Simulated (Baseline)

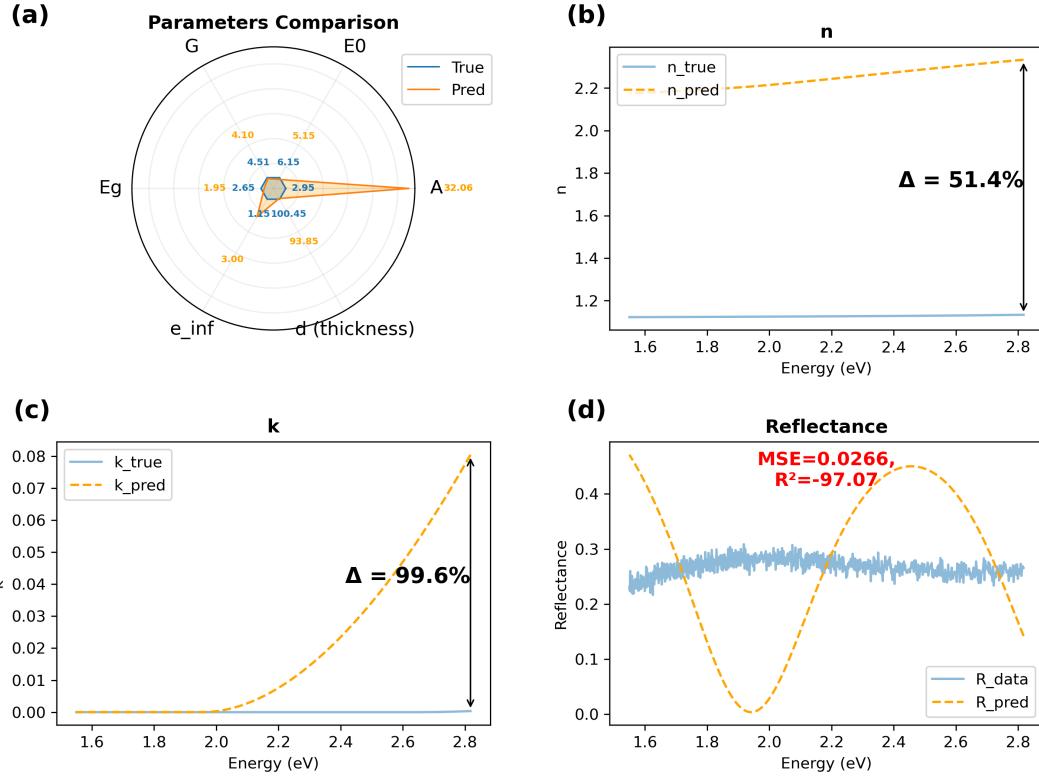


Figure 3.2: Baseline model prediction for reflectance, n , k , and parameter comparison for an outlier (data 404) with the largest MSE in the dataset. (a) Radar chart of parameters comparison. (b) True and predicted n values. (c) True and predicted k values. (d) True and predicted reflectance values, with MSE and R^2 score calculated.

squared error (MSE) values of predicted parameters, which are normalized (the testing loss) and the R^2 scores reflectance. A box plot is a graphical representation that summarizes data through its quartiles[31]. In a box plot, the box's central line represents the data's median. The edges of the box represent the interquartile range (IQR), which contains the middle 50% of the data. The whiskers extend to the smallest and largest values within 1.5 times the IQR from the lower and upper quartiles, respectively, indicating the variance of the

data: a longer whisker represents greater variance. Points outside this range are considered outliers and are plotted individually. The box plot offers important statistical insights into a dataset.

From figure 3.3 (a), we can see that as the number of feature extraction layers increases, the MSE decreases. This observation suggests that deeper models can better capture the complex relationships in the data. The median MSE decreases from 0.0385 with 4 layers to 0.0223 with 6 layers, indicating improved accuracy. However, there is a slight increase to 0.0242 when using 7 layers, suggesting potential overfitting. In figure 3.3 (b), the trend in the R^2 score of reflectance also indicates a similar conclusion. As n_{layers} increases, the variance decreases (the length of the whisker decreases). The median R^2 score increases significantly from 0.2222 with 4 layers to 0.9394 with 6 layers but slightly decreases at 7 layers, suggesting that the best n_{layers} is 6. Notice that many outliers are present in the box plots of 3.3 (b), suggesting that the model does not generalize well across all data points.

Overall, these results indicate that increasing the number of convolutional layers improves the model's accuracy in predicting normalized parameters and reflectance. However, there is a point beyond which adding more layers may lead to overfitting. Therefore, choosing 6 layers as the CNN architecture parameter appears to be an optimal choice that balances high accuracy and prevents overfitting.

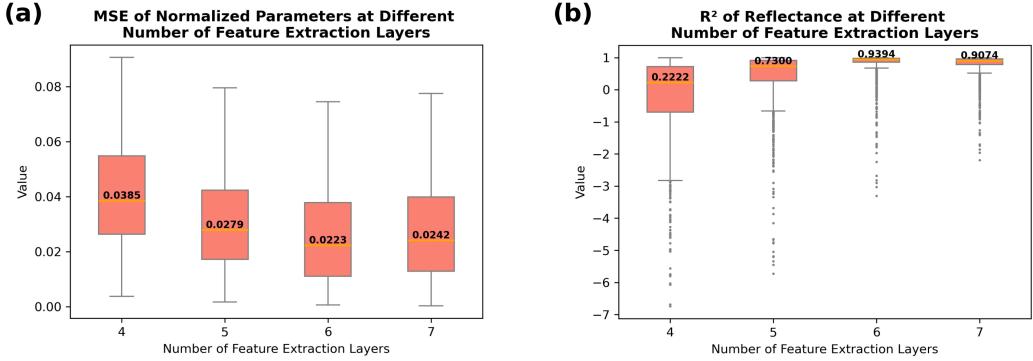


Figure 3.3: Box plots of metrics for different numbers of feature extraction layers. (a) Mean squared error (MSE) of normalized parameters at different numbers of feature extraction layers. (b) R^2 score of reflectance at different numbers of feature extraction layers.

3.1.3 Learning Rate

Next, the effect of learning rate (α) was investigated. Figure 3.4 (a) illustrates the box plots of the MSE of normalized parameters, and figure 3.4 (b) illustrates the R^2 score of reflectance. In figure 3.4 (a), the median of the MSE of normalized parameters increases as the learning rate decreases, and the variance increases slightly. In figure 3.4 (b), the median R^2 score of reflectance decreases slightly from 0.9394 at $\alpha = 10^{-3}$ to 0.9173 at $\alpha = 10^{-4}$, and then drops significantly to 0.7940 at $\alpha = 10^{-5}$, with a considerable increase in the variance. A reduction in the learning rate beyond this point tends to degrade model performance, potentially due to two primary factors: early stopping and the propensity of the optimization algorithm to become trapped in local minima.

- **Early stopping:** In the training, early-stopping was deployed such that if the validation loss did not improve for more than 10 epochs, the training

stopped. If the learning rate is too small, the optimization may stop too early because minimal changes in loss may be misinterpreted as convergence, preventing the model from reaching a more optimal solution.

- **Trapped in local minimum:** A smaller learning rate increases the risk of the optimization process getting stuck in local minima. This problem occurs because smaller steps are more likely to settle in shallow valleys of the loss function, which might not be the global minimum. As a result, the model may exhibit suboptimal performance on new unseen data outside the training set.

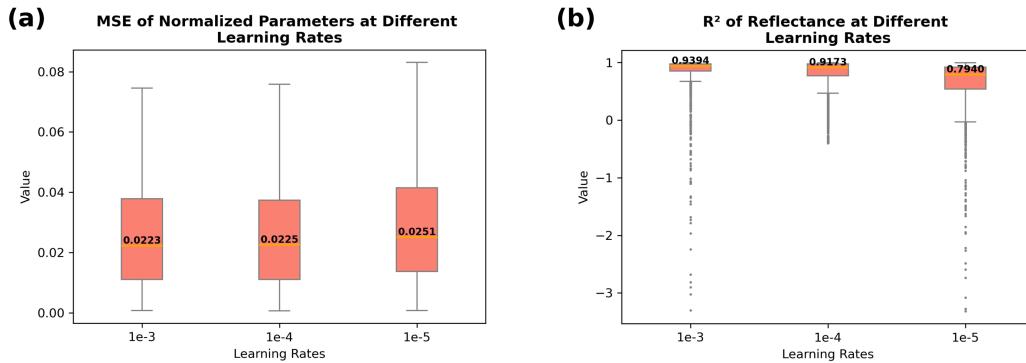


Figure 3.4: Box plots of metrics for different learning rates. (a) Mean squared error (MSE) of normalized parameters at different learning rates. (b) R^2 score of reflectance at different learning rates.

These results suggest the best learning rate is $\alpha = 10^{-3}$, giving the lowest median MSE value with the highest median R^2 score.

3.1.4 Training Data Size

As described by Goodfellow et al. (2016)[26], the scaling law in deep learning highlights that increasing training data size generally enhances model generalization. This principle is evidenced in figure 3.5, which presents the box plots of previously discussed metrics for different training data sizes. Specifically, figure 3.5 (a) demonstrates a reduction in both the median MSE value and its variance as the training data size increases from 10,000 to 50,000. More compellingly, figure 3.5 (b) illustrates the benefits of expanding the training dataset more straightforwardly, with the median R^2 score improving significantly from 0.9394 at $N = 10,000$ to 0.9754 at $N = 50,000$, approaching the upper limit of R^2 score of 1. Simultaneously, the variance in the R^2 scores decreases markedly, with no outliers below 0. These observations suggest that $N = 50,000$ is the optimal size for the training dataset. While further increases in N could enhance performance, the computational cost at $N = 50,000$ is already considerable, and the benefits would not likely justify the additional resources.

3.1.5 Input Dimension

The dimensionality of input data profoundly influences model performance by affecting its ability to capture input features[26]. While decreasing the input dimension can reduce the computational cost, overly low dimensions may impair the model’s capacity to accurately capture the patterns in reflectance spectra, such as peaks and valleys. Therefore, it is worth investigating whether lower-dimensional inputs can achieve performance comparable to higher-dimensional

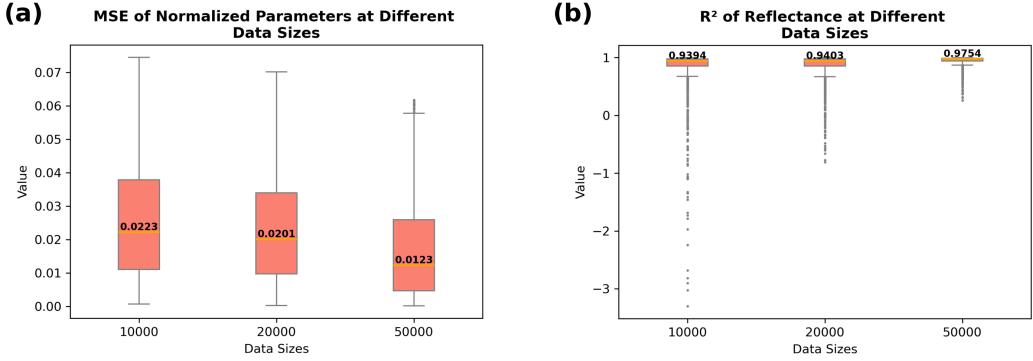


Figure 3.5: Box plots of metrics for different training data sizes. (a) Mean squared error (MSE) of normalized parameters at different training data sizes. (b) R^2 score of reflectance at different training data sizes.

settings. As depicted in figure 3.6, the median MSE and its variance remain consistent across different input dimensions, with $dim = 500$ exhibiting marginally superior results. Similarly, the R^2 scores are nearly the same across dimensions, with high values close to 1, as illustrated in figure 3.6 (b). These results suggest that the chosen input dimensions are sufficiently robust to capture the essential features of the reflectance spectra, allowing the CNN to learn effectively, irrespective of the input dimension size.

Since $dim = 500$ performs slightly better with the lowest median MSE of 0.0113 and highest median R^2 of 0.9808 among the models with different input dimensions, it is selected as the optimal input dimension for the best model.

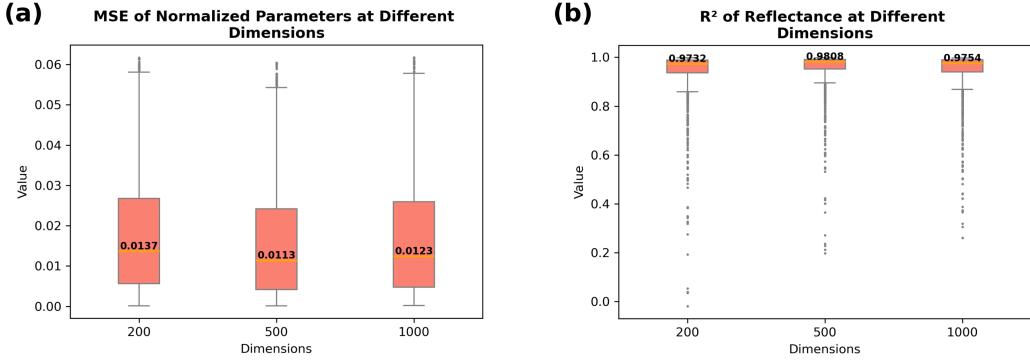


Figure 3.6: Box plots of metrics for different input dimensions. (a) Mean squared error (MSE) of normalized parameters at different input dimensions. (b) R^2 score of reflectance at different input dimensions.

3.1.6 Best Model

Based on these results, the best training model has the variable combination of: $n_{\text{layers}} = 6$, $\alpha = 10^{-3}$, $N = 50,000$, and $dim = 500$. Figure 3.7 shows the "best" CNN result for the same single reflectance spectra used for the baseline model (data 0). Compared to figure 3.1, the difference between the predicted parameters and the true parameters all decreases, with a parameter A decreasing significantly, as depicted in figure 3.7 (a). In figure 3.7 (b), Δ of n decreases from 6.9% of the baseline model to 2.1% of the best model, and in figure 3.7 (c), Δ of k decreases from 38.7% of the baseline model to 19.7% of the best model, representing a considerable improvement in n, k prediction. In figure 3.7 (d), the predicted reflectance almost overlaps the true reflectance extremely well, and the MSE value of reflectance decreases from 0.0068 of the baseline model to 0.0002. More importantly, the R^2 score rises dramatically from 0.64 for the baseline model to 0.99 for the best model, indicating an almost perfect prediction. These results

imply that the best model can extract n, k accurately on the simulated test data, potentially producing an initial guess close to the global minimum in the fitting process and solving the ill-posed problem. However, tests must be evaluated to establish the more general validity of the experimental data.

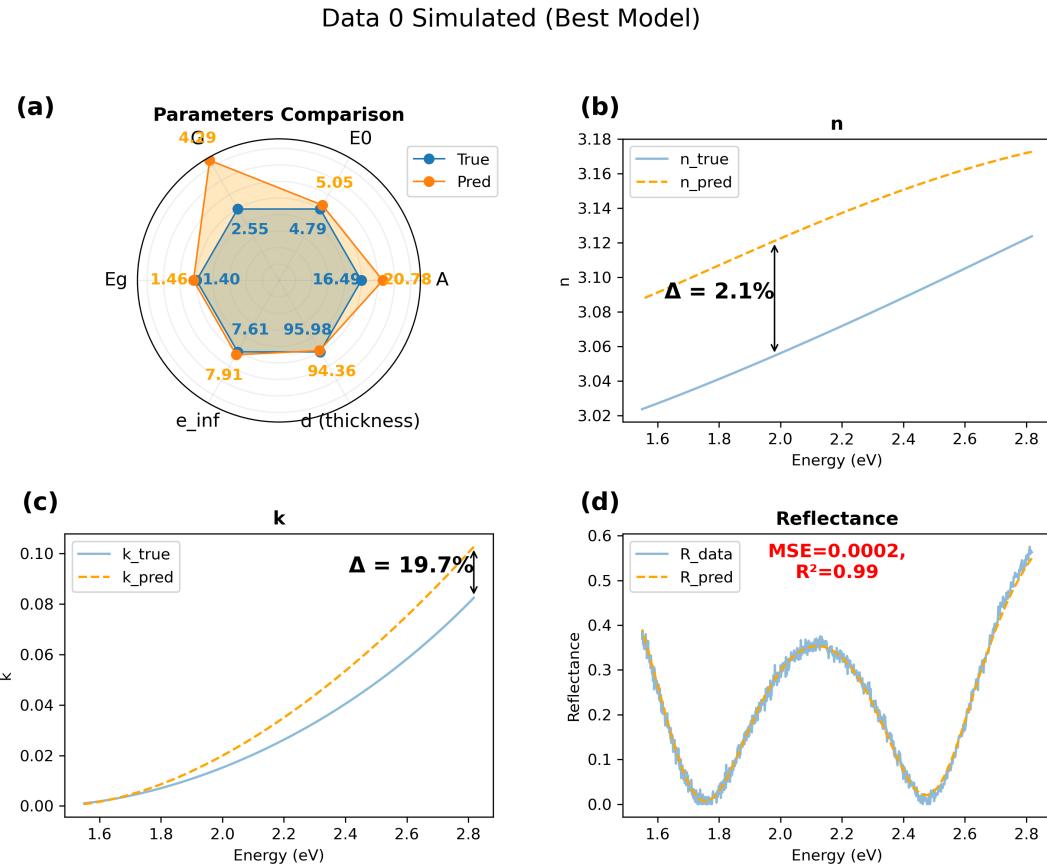


Figure 3.7: The best model prediction for reflectance, n , k , and parameter comparison of the same single simulated data used for the baseline model (data 0). (a) Radar chart of parameters comparison. (b) True and predicted n values. (c) True and predicted k values. (d) True and predicted reflectance values, with MSE and R^2 score calculated.

3.1.7 Parameters Accuracy Comparison

Before leaving the CNN predictions, we want to explore which parameters (out of $A, E_0, G, E_g, \varepsilon_\infty, d$) are predicted the most accurately. Figure 3.8 shows 6 subplots, and each subplot contains scattered yellow dots for the true value and predicted value. Points closer to the red line ($y = x$) indicate better predictions. It can be noticed that the thickness d , a parameter not part of the TL model, has the best prediction accuracy, with an R^2 score of 0.99. On the other hand, parameters E_0, G , and A all have poor accuracy, with R^2 score of 0.47, 0.47, 0.68 respectively. This observation indicates that the TL model might be too complicated and that its parameters are relatively hard to learn by a CNN. Above all, while the model can learn the thickness of the top layer well, it performs relatively poorly when learning the representation of the TL model.

To further assess the prediction accuracy of n and k at 2.2 eV, Figure 3.9 compares the true n, k values—calculated using the true parameters—with the predicted n, k values—derived from the predicted parameters. The figure demonstrates that the n, k predictions align closely with the true values. Interestingly, even though the predictions for the TL model parameters are not highly accurate, the predicted n, k values remain relatively precise. This observation suggests that the model can still capture the optical properties effectively despite the parameter prediction discrepancies.

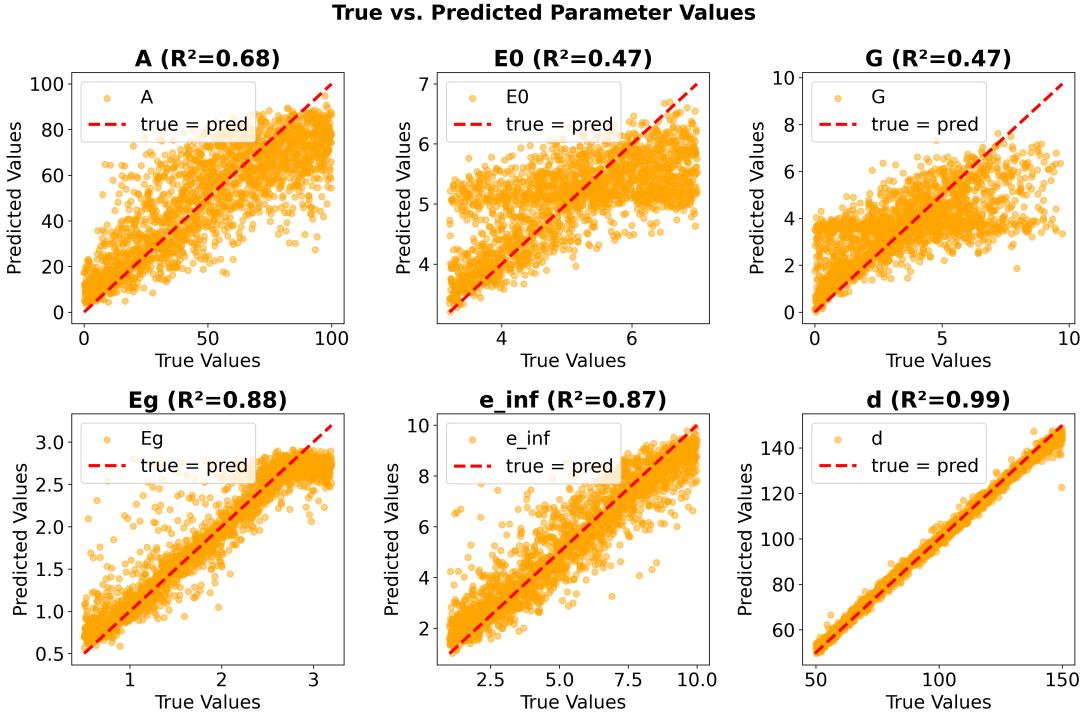


Figure 3.8: True and predicted values of all parameters. From left to right, top to bottom, subplots of the parameters $A, E_0, G, E_g, \varepsilon_\infty, d$. The red line shows the ideal $y = x$ condition.

3.2 CNN Prediction and TRF Fitting for Experimental Test Set

3.2.1 Before TRF Fitting

After identifying the best parameters to train the model, the model was deployed to examine its performance on a set of 151 experimentally measured reflectance spectra across an lg-LSA processed stripe. The evaluation results were examined to see if they could serve as useful initial guesses for the TRF fitting.

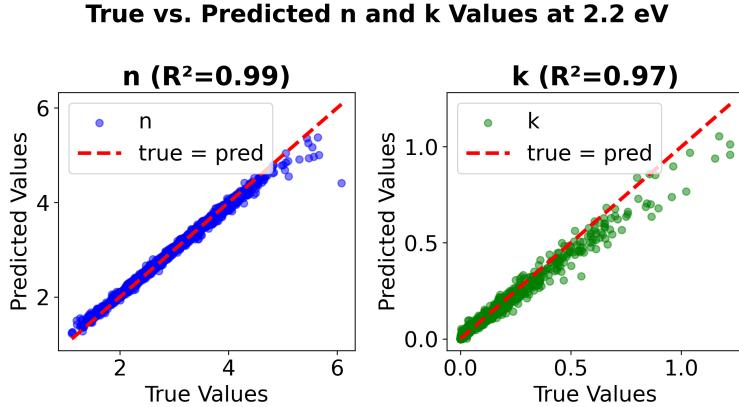


Figure 3.9: True and predicted values of n, k at photon energy 2.2eV. The blue and green scattering dots have coordinates (true value, predicted value), and the red line has an analytical formula $y = x$.

Single Reflectance Spectra

Similar to Section 3.1.1, the model was first tested on a single reflectance spectrum. Figure 3.10 shows the evaluation result of a single experimental reflectance spectrum. Figure 3.10 (a) shows the predicted parameters in a single radar chart, and figure 3.10 (b) and (c) show the predicted n and k . Notice that only predicted values are shown since the true values are unknown.

For experimental data, the only approach to assess the prediction accuracy is by comparing the predicted reflectance to the experimental reflectance, as shown in figure 3.10 (d). The accuracy of this experimental data is not as high as that of the simulated data used to train and test the CNN model. Specifically, the MSE is 0.0095 compared to the values of 0.0002 for simulated data, and the R^2 score is only 0.54. Given the limitation of the TL model, this is not entirely unexpected. Based on the simulated data, the CNN model's generalized ability does not completely imitate the experimental data. Even with Gaussian noise

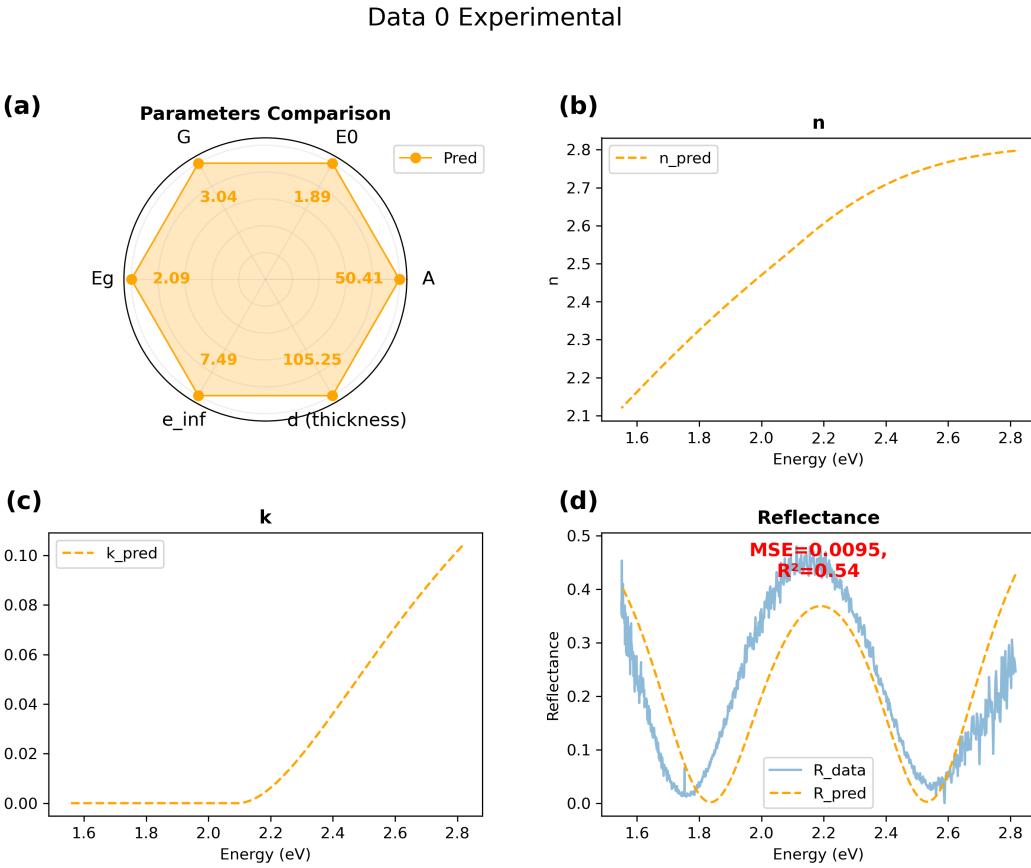


Figure 3.10: The best CNN model prediction for reflectance, n , k , and parameter comparison for randomly selected experimental data (Data 0 means it is the first data in the dataset). (a) Radar chart of predicted parameters. (b) Predicted n curve. (c) Predicted k curve. (d) Predicted and experimental reflectance curves, with calculated MSE and R^2 scores.

and the range of parameters used for data generation, the model cannot perfectly produce simulated data representing real-world data.

All Reflectance Spectrum

The decrease in accuracy is further apparent when the evaluation is repeated for all 151 experimental reflectance spectra across the stripe. Figure 3.11 shows

the box plot of R^2 scores for (i) the simulated reflectance test spectra and (ii) the experimental reflectance spectra.

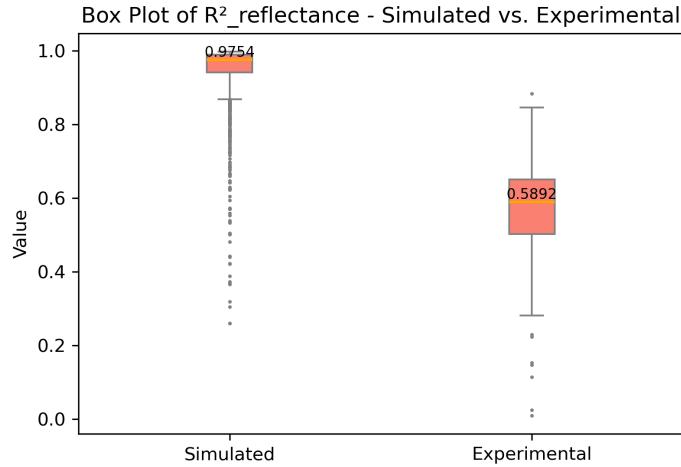


Figure 3.11: Box plots of R^2 score of reflectance for simulated data (left) and experimental data (right).

It can be noticed that the median of R^2 score drops from 0.9754 for test data to 0.5892 for experimental data, with a larger variance for the experimental data. Also, the CNN model accuracy drops dramatically but likely provides sufficiently robust initial values for subsequent TRF fitting.

Reflectance Spectrum Across the Stripe

Beyond the box plot which does not include spatial information of the measured spectrum, it is also important to look at the predicted n, k , and reflectance across the stripe to assess the accuracy of CNN predictions. The predicted n, k and reflectance should be very similar for spectra drawn from within the same phase of the LSA stripe. Figure 3.12 shows a heatmap of predicted n (figure 3.12 (a)),

a heatmap of predicted k (figure 3.12 (b)), a heatmap of predicted reflectance ((figure 3.12 (c)) across the stripe, and experimentally measured reflectance in figure 3.12 (d). Key observations include:

- The x-position (represented by the data index) where transitions occur shows common points for n , k and reflectance. Wherever transitions occur in the experimental reflectance heatmap, there is a corresponding transition in the predicted n heatmap, predicted k heatmap, and the predicted reflectance heatmap. This correspondence indicates that the CNN model is robust across the stripe.
- In figure 3.12 (c), some noises exist in the predicted reflectance, and it deviates from the experimental reflectance heatmap in figure 3.12 (d). The peaks and valleys of each predicted reflectance spectra oscillate to different energies across the reflectance heatmap. For the predicted reflectance, the mean MSE is 0.0114, and the mean R^2 score is only 0.50, representing only a moderate prediction accuracy. However, the major features of the phase transformations are captured in the baseline CNN results before any refinement with the TRF algorithm.
- In the heatmap of predicted k , at high energy range (2.6 to 2.8 eV, a range of blue to purple light), the k value is close to 0 between data indices range of 82 to 110 (the center stripe where annealing takes place) comparing to other regions, implying that Bi_2O_3 becomes more transparent after annealing. The annealing condition changes the phase and increases the band gap energy, making the material more transparent at the high energy side.

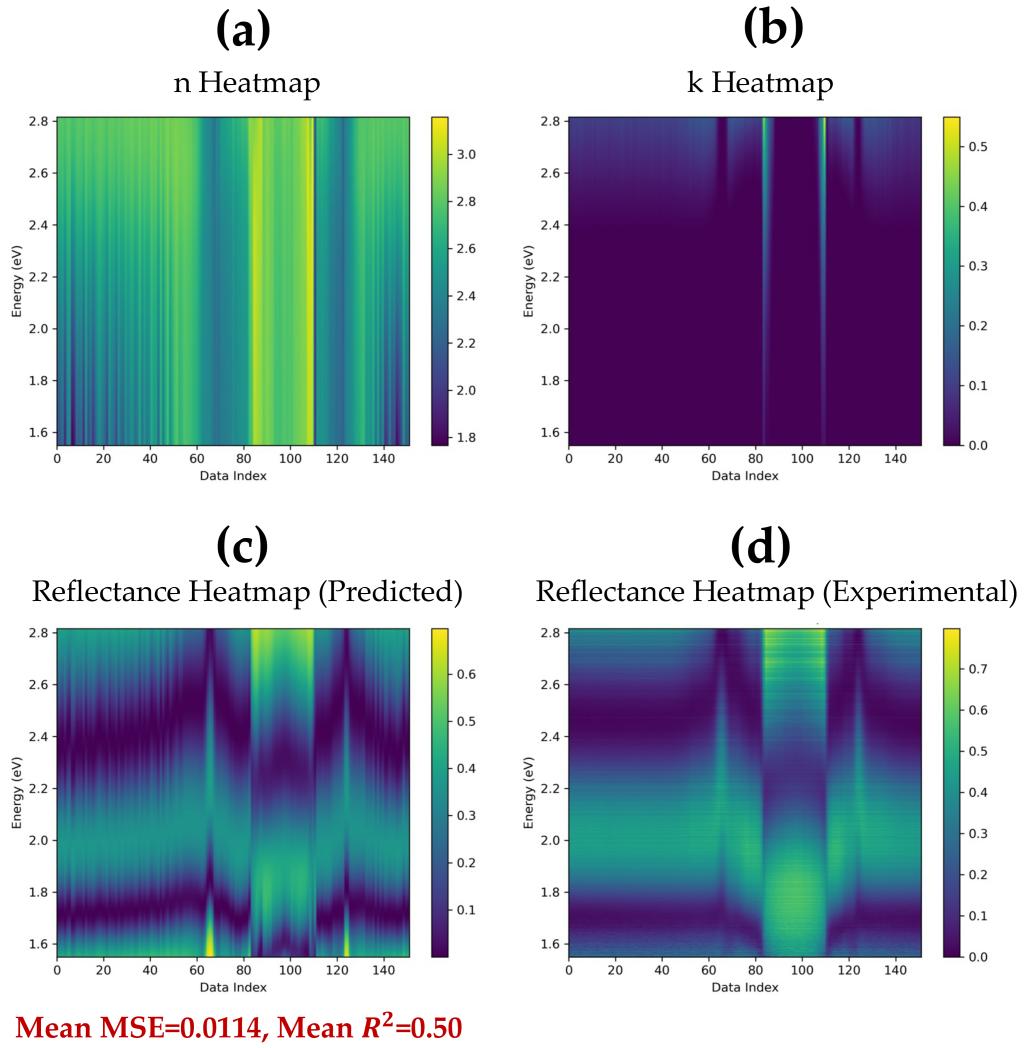


Figure 3.12: (a) Heatmap of the predicted n , (b) Heatmap of the predicted k , (c) Heatmap of the predicted reflectance with Mean MSE and Mean R^2 score calculated, and (d) Experimentally measured reflectance for comparison.

Thickness Profile Across the Stripe

Figure 3.13 shows the predicted thickness d profile across the stripe (in blue). Before annealing (regions of data indices range from 0 to 60 and 130 to 150), the thickness remains constant at about 105nm, as expected. A thickness reduc-

tion of about 40nm occurs at the center annealed location (data indices range from 82 to 110), while the thickness increases by about 40nm on either side of the annealed location. This change might be because, upon annealing, the laser pushes surface materials at the annealed position to the sides, creating a thickness change.

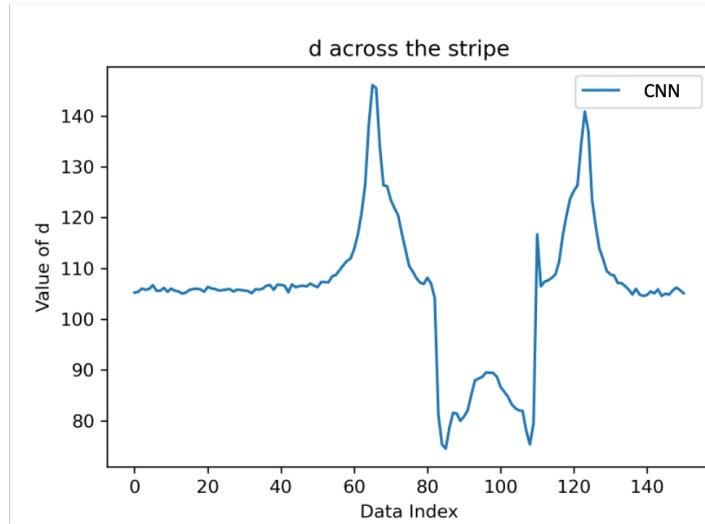


Figure 3.13: CNN predicted thickness d across the stripe.

To summarize, although the CNN model accuracy decreases in the experimental test set compared to the simulated test set, it maintains a certain level of robustness by being consistent within the same phase across the LSA stripe. TRF fitting is expected to decrease the residuals between predicted and experimental reflectance.

3.2.2 With TRF Fitting

In this subsection, the initial results returned by the CNN model are further refined using TRF fitting.

Single Reflectance Spectra

Figure 3.14 shows the evaluation result of the same single experimental testing reflectance spectra (data 0), after TRF fitting conducted. In Figure 3.14 (a), the predicted parameters and the fitted parameters are represented in a radar chart, where the initial CNN predicted parameters (in yellow) are all normalized to the same value (0.5) to illustrate how the TRF fitted parameters (in red) differ from the predicted parameters. The parameters E_0, A, G change the most dramatically. In figure 3.14 (b) and figure 3.14 (c), the fitted n and k also change significantly from the origin CNN predicted values. In figure 3.14 (d), comparing to the predicted reflectance, the TRF fitted reflectance curve more closely approximates the experimental reflectance, with a decreased MSE of 0.0023 (compared with 0.0095 of the CNN predicted reflectance) and an improved R^2 score of 0.88 (compared with 0.54 of the CNN predicted reflectance). The improved accuracy can be expected as TRF fitting seeks to minimize the residual between predicted and experimental reflectance. Unsurprisingly, the fitted reflectance still cannot completely fit the experimental reflectance as the n and k are still based on a simple one-oscillator TL model. As a mathematical formula that models the optical properties of amorphous semiconductors at photon energies greater than their band gap, the TL model fails to fully model the crystallized phase and n, k

below the band gap accurately. More sophisticated models[32][33] are needed to model the full structure of the experimental reflectance more accurately.

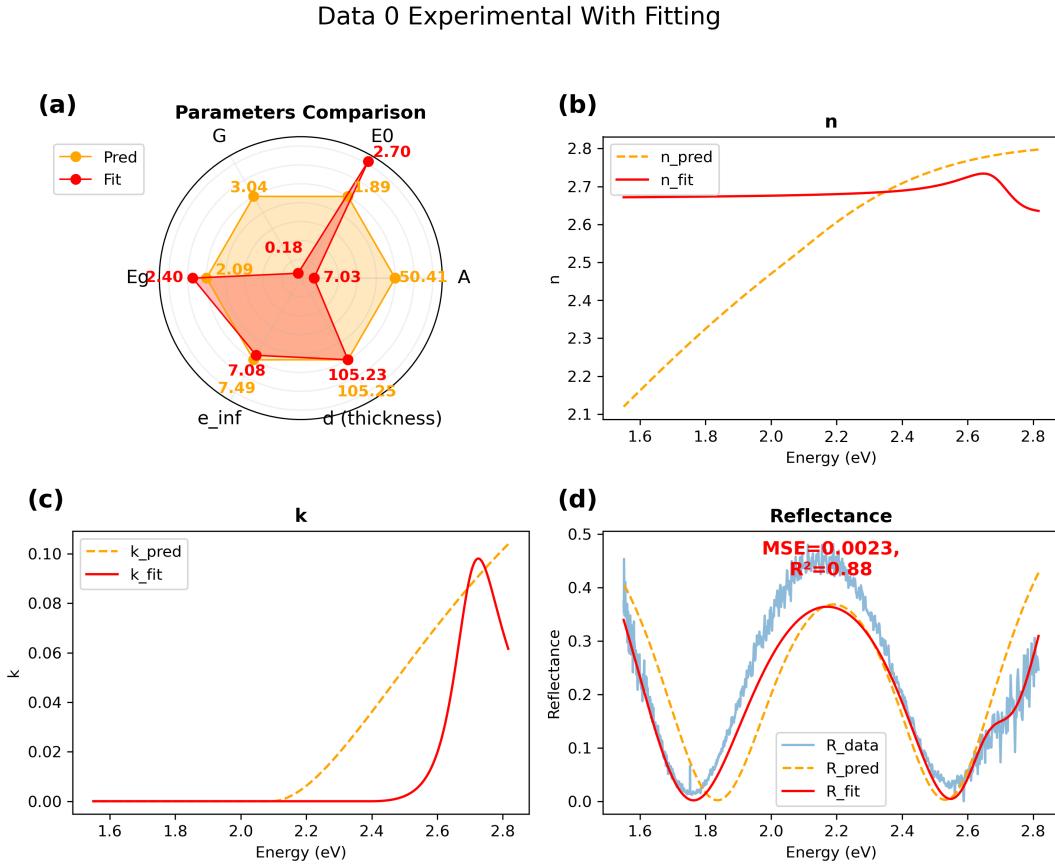


Figure 3.14: The best CNN model prediction with TRF fitting for reflectance, n , k , and parameter comparison for the same single experimental data used in figure 3.10 (data 0). (a) Radar chart of CNN predicted parameters (in orange) and the TRF optimized parameters (in red). (b) Predicted and fitted n values. (c) Predicted and fitted k values. (d) Predicted, fitted, and experimental reflectance, with MSE and R^2 score between the fitted reflectance and the experimental reflectance.

All Reflectance Spectrum

The TRF fitting was repeated for all 151 experimental reflectance spectra across the stripe. Figure 3.15 shows the box plot of R^2 score of reflectance before applying TRF fitting and after applying TRF fitting. As expected, the median of the R^2 score improves from 0.5892 to 0.8408, and the variance of the R^2 score decreases. The fact that the median of the R^2 score of experimental data fails to reach above 0.98 as the simulated data confirms the limitation of the TL model in modeling real-world experimental reflectance spectrum.

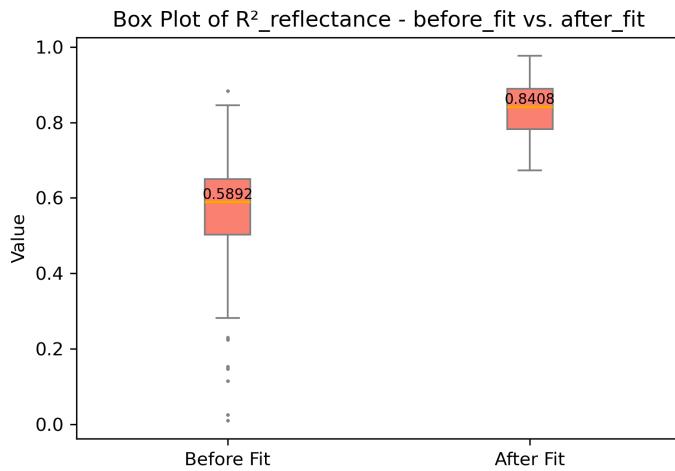


Figure 3.15: Box plots of R^2 score of reflectance before TRF fitting (left) and after TRF fitting (right).

It is important to note that while the accuracy of the TRF fitting method improves, the computational cost is significantly higher than the initial CNN prediction. Table 3.1 compares the time required for each method. Even on a simple laptop processor, the CNN prediction requires only 0.014 seconds on average per evaluation, totaling 2.21 seconds for the entire dataset. In contrast, TRF fitting is approximately 60 times slower. Given that the accuracy improve-

ment factor is only about 1.43 (as indicated by comparing the median of the R^2 scores before and after fitting), it raises an essential consideration: the trade-off between efficiency and enhanced accuracy. Whether this trade-off justifies the increased computational expense is a critical question. The two methods can be used synergistically in an experimental flow. The CNN is sufficiently fast that it can be incorporated within an experimental workflow to identify and explore promising regions of the parameter space. Later, the TRF refinement can be used to assess the results more accurately.

	CNN Prediction	TRF Fitting
Avg. Time of 1 Data	0.014 seconds	0.82 seconds
Total Time of 151 Data	2.21 seconds	124.08 seconds = 2.07 mins

Table 3.1: Time Comparison of CNN Prediction and TRF Fitting.

Reflectance Spectrum Across the Stripe

The heatmap of n , k , and reflectance after fitting are again investigated to examine the consistency within each phase. The heatmap of fitted n , fitted k , and fitted reflectance are illustrated accordingly in figure 3.16 (a), (b), and (c), with the experimentally measured heatmap in figure 3.12 (d) for comparison. It can be observed that after fitting, n, k are very different from those before fitting. Still, the position where transitions occur remains the same, indicating the robustness of both the CNN and TRF fitting. In figure 3.16 (b), the high energy k values all become very low, implying that various lg-LSA conditions exist that can all result in high transparency. In figure 3.16 (c), the predicted reflectance heatmap looks more similar to the experimental reflectance heatmap after fitting, with the mean MSE values decreasing from 0.0114 before fitting to 0.0004

after fitting, and Mean R^2 score improving from 0.50 to 0.83. Above all, the TRF fitting method effectively improves the model accuracy.

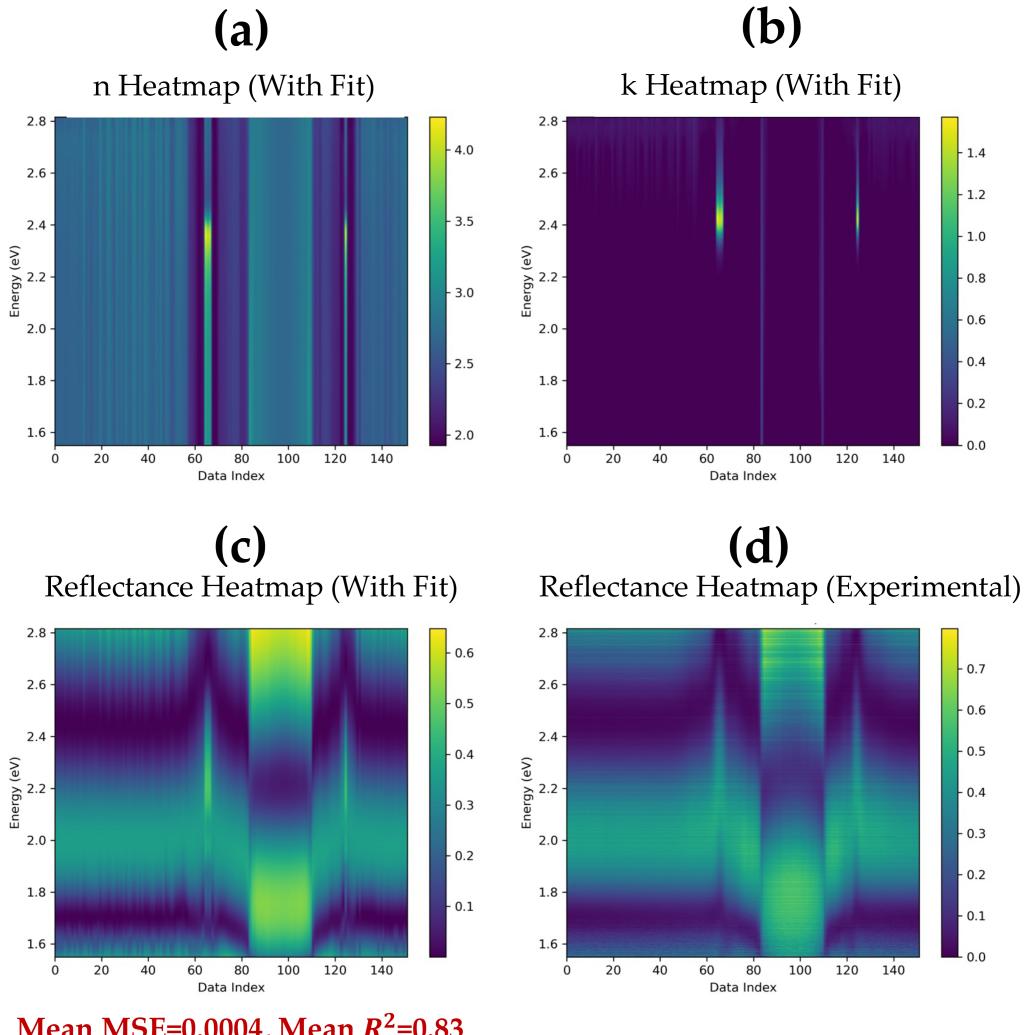


Figure 3.16: (a) Heatmap of the fitted n , (b) Heatmap of the fitted k , (c) Heatmap of the fitted reflectance with Mean MSE and Mean R^2 score calculated, and (d) Experimentally measured reflectance for comparison.

Thickness Profile Across the Stripe

Lastly, the effect of TRF fitting on the thickness d is investigated. Figure 3.17 shows CNN predicted thickness (in blue) and TRF fitted thickness (in orange) across the stripe. The shape of the TRF optimized thickness profile is similar to the shape of the CNN predicted thickness profile, but it is much noisier and, at some point, reaches the boundary set for the TRF algorithm (for instance, at data index = 60, the value of d reaches the maximum of 150nm). This drawback indicates the limitation of TRF fitting, which cannot maintain consistency across the stripe. More advanced methods, such as incorporating a cost function that accounts for the difference between adjacent points (reflecting physical continuity requirements), may offer effective solutions.

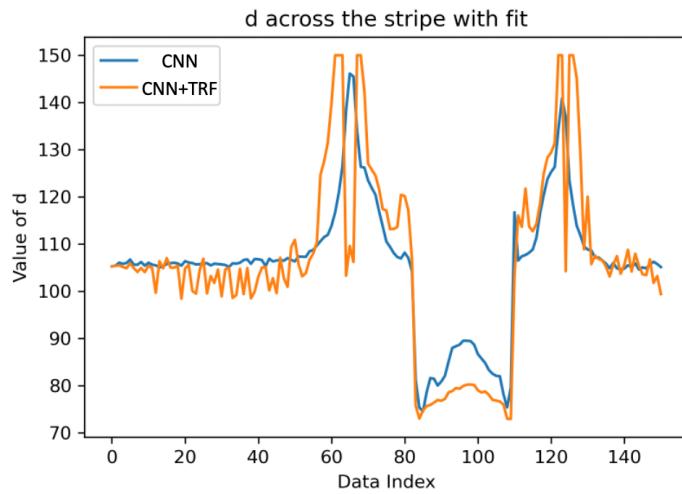


Figure 3.17: Predicted thickness (in blue) and fitted thickness (in orange) across the stripe.

As discussed in Section 3.1.7, the CNN model demonstrates high accuracy in predicting the thickness d of materials. Specifically, it accurately predicts the thickness profile of materials displaced from the annealed area, suggesting

a high likelihood of this prediction being correct. Significant support for this assertion is provided by an experimental thickness profile measured in a previous thesis from our group[1], as shown in Figure 3.18. This figure depicts the profilometry characterization of a laser spike annealed stripe enriched with bismuth within a Bi-Ti-O composition spread. Notably, the thickness increases as the x-position approaches the center of the annealed area and then decreases sharply within the annealed area itself. Despite differences in material composition and thickness, the similarities between these profiles underscore the physical validity of using both CNN prediction and TRF fitting for thickness evaluation.

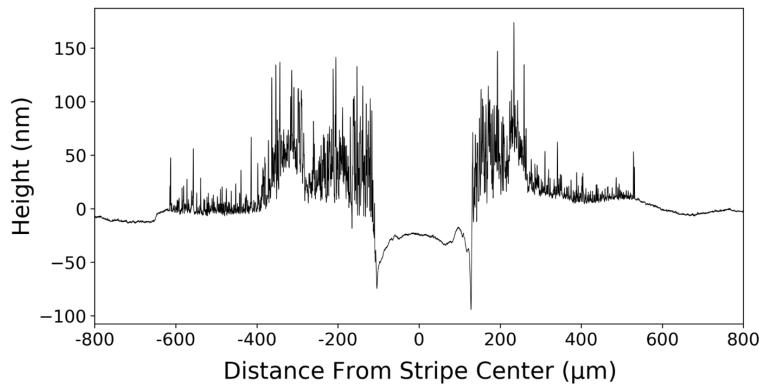


Figure 3.18: Profilometry characterization of a laser spike annealed stripe rich in bismuth in a Bi-Ti-O composition spread[1].

In summary, while the TRF fitting method significantly enhances model accuracy, it is not sufficiently complex to fully model the experimental data, an inherent limitation of the TL model. Nonetheless, with a mean R^2 score of 0.83, the method provides a reliable characterization of the optical properties n , k , and thickness d , particularly d . By analyzing the heatmap of k , we can clearly identify the LSA-processed regions characterized by high transparency, effec-

tively solving the ill-posed problem and boosting the multi-objective optimization studies.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 Conclusion

This study presents a novel approach to address the ill-posed reverse problem of determining the optical properties n, k and thickness of optical thin films from reflectance measurements. These are crucial for applications in multi-objective optimization in autonomous experiments. By applying Convolutional Neural Networks (CNNs) to produce accurate initial guesses for the Trust Region Reflective fitting algorithm (TRF), we developed a method to extract these properties from reflectance data of multilayer films processed by lateral gradient laser spike annealing (lg-LSA). The CNNs are readily trained using simulated reflectance datasets generated using the Tauc-Lorentz (TL) model and the Transfer Matrix Method (TMM).

Through extensive experimentation with various CNN variables, including the number of feature extraction layers (n_{layers}), learning rate (α), training data size (N), and input dimension (dim), a robust CNN training model was identified with these insights:

- The optimal n_{layers} value is 6, balancing model complexity and performance. Increasing the number of layers beyond this appears to lead to overfitting, where the model performs well on training data but poorly on unseen data.

- A learning rate (α) of 10^{-3} achieves the best convergence speed and accuracy. Higher learning rates may cause the model to overshoot minima, while lower rates may slow convergence and result in the CNN trapped in local minima.
- A training data size (N) of 50,000 samples ensures sufficient parameter space coverage, significantly improving the model's generalization ability. Increasing the dataset size further would continue to improve performance but at a diminishing rate and higher computational cost.
- An input dimension (dim) of 500 maintains a good balance between computational efficiency and model accuracy. While reducing the input dimension can lower computational costs, it may also result in the loss of critical information, affecting accuracy.

This model achieved a high mean R^2 score of 0.9754 on the simulated dataset, showcasing its effectiveness. Moreover, the model predicted the thickness effectively, with an R^2 score of 0.99.

However, some limitations were observed. First, the CNN model could not correctly learn from reflectance spectra with non-significant peaks and valleys, indicating the limitation of this model. Second, when applying the model to experimental data, the model's accuracy was lower than on the simulated data, with a mean R^2 score of 0.5892. After applying TRF fitting, the mean R^2 score improved to 0.8408, which, although better, still did not reach a highly accurate fit ($R^2 > 0.9$). The TL model used in the simulated data cannot accurately represent real-world experimental data. The TL model primarily models amorphous

semiconductors at photon energies higher than their optical band gap, while experimental curves cover photon energy ranges above and below possible Bi_2O_3 band gaps.

Nevertheless, the CNN prediction demonstrated robustness across the stripe. The CNN accuracy was consistent within the same phase for data across the stripe, indicating the model's robustness in predicting optical properties across different regions of the LSA-processed thin film.

Moreover, the shape of the predicted and fitted thickness profile aligns with experimentally measured data, showing that the LSA-processed thin film will have a lower thickness at the center and higher thickness on the sides because the laser pushes materials to the sides at the annealed location. This observation further proves the effectiveness of CNN prediction in thickness characterization.

These findings highlight the potential of combining machine learning and optimization techniques for high-throughput, autonomous material characterization, providing a robust foundation for further advancements.

4.2 Future Work

Building on the findings of this study, several enhancements and new directions are proposed to improve the accuracy and broaden the applicability of the developed methods:

- **TL Model Improvement:** The current study uses a TL model with a single oscillator to generate data. Future work could explore using TL models with multiple oscillators to capture more complex optical behaviors, although this will exponentially increase the problem's dimensionality, posing significant computational challenges. Additionally, employing advanced TL models, such as TL models that incorporate Urbach tails[33] could improve the accuracy of modeling n and k values below the band gap. Experimenting with a broader range of TL parameters will also help to refine the model and better represent the diverse properties of the materials studied.
- **Data Generation Improvement:** Adjusting the level of Gaussian noise to see how it affects model accuracy is another potential improvement. Moreover, introducing higher noise levels at low and high photon energy ranges with a moderate noise level at the central range can better mimic real-world data, providing more robust training for the CNN model. This approach will help to bridge the gap between simulated and experimental data, leading to more accurate predictions.
- **TRF Fitting Improvement:** In the current study, TRF fitting improves the accuracy of predictions but could benefit from further enhancements. Incorporating the minimization of differences between adjacent reflectance data into the cost function of the fitting process could ensure consistency within the same phase. This adjustment would address the issue of noise and variation within phases, leading to smoother and more reliable predictions.
- **Validation of Prediction:** One significant limitation of this study is the

validation of the n and k predictions, that we cannot obtain information of the true n and k values. To ensure the predicted values are close to the true optical properties, transmission measurements could be employed to get transmittance directly, and there get k via equation 1.4. n can be further obtained via the Kramers–Kronig (KK) relations[20][21]. This approach would provide a more comprehensive validation of the optical properties. Developing such an experimental setup is still under conception but will be crucial for future work. Also, experimental thickness characterization, including profilometry, can be used to validate the accuracy of thickness prediction.

- **Model Evaluation and Alternative Techniques:** Conducting extensive cross-validation using diverse experimental datasets will be essential to refine model accuracy and enhance generalization capabilities. Additionally, exploring alternative machine learning techniques, such as ensemble methods or transformer models, may offer potential improvements in performance and robustness.

These proposed future directions aim to address the current limitations and expand the applicability of the developed methods, paving the way for a more accurate, efficient, and comprehensive material characterization. With these improvements, the methods developed in this study could be applied to a wider range of materials and thin films with complex composition spreads, which will eventually help to achieve the ultimate goal of integrating these refined techniques into autonomous experiments, thereby significantly advancing the field of materials science research.

BIBLIOGRAPHY

- [1] Sofia Chen. Structure and dielectric properties of Bi-Ti-O thin films, 2023. Unpublished senior thesis, Department of Materials Science and Engineering.
- [2] Nathan J Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E Kumar, Tanjin He, David Milsted, Matthew J McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature*, 624(7990):86–91, 2023.
- [3] Gary Tom, Stefan P Schmid, Sterling G Baird, Yang Cao, Kourosh Darvish, Han Hao, Stanley Lo, Sergio Pablo-García, Ella M Rajaonson, Marta Skreta, et al. Self-driving laboratories for chemistry and materials science. *Chemical Reviews*, 2024.
- [4] Sathya R Chitturi, Akash Ramdas, Yue Wu, Brian Rohr, Stefano Ermon, Jennifer Dionne, Felipe H da Jornada, Mike Dunne, Christopher Tassone, Willie Neiswanger, et al. Targeted materials discovery using bayesian algorithm execution. *npj Computational Materials*, 10(1):156, 2024.
- [5] Sebastian Ament, Maximilian Amsler, Duncan R Sutherland, Ming-Chiang Chang, Dan Guevarra, Aine B Connolly, John M Gregoire, Michael O Thompson, Carla P Gomes, and R Bruce van Dover. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. *Science Advances*, 7(51):eabg4930, 2021.
- [6] Richard S Sutton. Reinforcement learning: An introduction. *A Bradford Book*, 2018.
- [7] Rakesh A Afre, Nallin Sharma, Maheshwar Sharon, and Madhuri Sharon. Transparent conducting oxide films for various applications: A review. *Reviews on advanced materials science*, 53(1):79–89, 2018.
- [8] Claes G Granqvist. Transparent conductors as solar energy materials: A panoramic review. *Solar energy materials and solar cells*, 91(17):1529–1598, 2007.
- [9] William Lowrie and Andreas Fichtner. *Fundamentals of Geophysics*. Cambridge University Press, 2020.

- [10] Eugene Hecht. *Optics*. Pearson Education India, 2012.
- [11] Robert T Bell, Alan G Jacobs, Victoria C Sorg, Byungki Jung, Megan O Hill, Benjamin E Treml, and Michael O Thompson. Lateral temperature-gradient method for high-throughput characterization of material processing by millisecond laser annealing. *ACS Combinatorial Science*, 18(9):548–558, 2016.
- [12] Harland G. Tompkins and Eugene A. Irene. *Handbook of Ellipsometry*. Springer Berlin, Heidelberg, 2005.
- [13] David W. Ball. Modern Spectroscopy, 2004. Review of *Modern Spectroscopy* by J. Michael Hollas, New York: John Wiley & Sons, Inc., 2004, 428 pp., paperback. ISBN 0-470-84416-7.
- [14] Leandro N. Acquaroli. Matrix method for thin film optics. *arXiv preprint arXiv:1809.07708*, 2018.
- [15] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.
- [16] RefractiveIndex.INFO - Refractive index database. <https://refractiveindex.info/>, 2024.
- [17] Nathan J. Szymanski, Christopher J. Bartel, Yan Zeng, Qingsong Tu, and Gerbrand Ceder. Probabilistic deep learning approach to automate the interpretation of multi-phase diffraction spectra. *Chemistry of Materials*, 33(11):4204–4215, 2021.
- [18] Maxim Ziatdinov, Ayana Ghosh, Chun Yin Wong, and Sergei V. Kalinin. Atomai framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nature Machine Intelligence*, 4(12):1101–1112, 2022.
- [19] Siyu Isaac Parker Tian, Zekun Ren, Selvaraj Venkataraj, Yuanhang Cheng, Daniil Bash, Felipe Oviedo, J Senthilnath, Vijila Chellappan, Yee-Fun Lim, Armin G Aberle, et al. Tackling data scarcity with transfer learning: a case study of thickness characterization from optical spectra of perovskite thin films. *Digital Discovery*, 2(5):1334–1346, 2023.

- [20] R. de L. Kronig. On the theory of dispersion of x-rays. *JOSA*, 12(6):547–557, 1926.
- [21] Hendrik Anthony Kramers. *La diffusion de la lumière par les atomes*. E. J. Brill, Leiden, Netherlands, 1928.
- [22] G. E. Jellison Jr and F. A. Modine. Parameterization of the optical functions of amorphous materials in the interband region. *Applied Physics Letters*, 69(3):371–373, 1996.
- [23] Rajdeep Dutta, Siyu Isaac Parker Tian, Zhe Liu, Madhavkrishnan Lakshminarayanan, Selvaraj Venkataraj, Yuanhang Cheng, Daniil Bash, Vijila Chellappan, Tonio Buonassisi, and Senthilnath Jayavelu. Extracting film thickness and optical constants from spectrophotometric data by evolutionary optimization. *PLOS ONE*, 17(11):e0276555, 2022.
- [24] Materials Project. Bi₂O₃ (bismuth oxide) - Materials Data, 2023. Accessed: June 26, 2023.
- [25] Boxin Tang. Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397, 1993.
- [26] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [27] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [28] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.
- [29] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [30] NR Draper. *Applied regression analysis*. McGraw-Hill. Inc, 1998.
- [31] John Wilder Tukey. *Exploratory Data Analysis*, volume 2. Springer, 1977.

- [32] Dmitriy V. Likhachev, Natalia Malkova, and Leonid Poslavsky. Modified tauc-lorentz dispersion model leading to a more accurate representation of absorption features below the bandgap. *Thin Solid Films*, 589:844–851, 2015.
- [33] Luis V. Rodríguez-de Marcos and Juan I. Larruquert. Analytic optical-constant model derived from tauc-lorentz and urbach tail. *Optics Express*, 24(25):28561–28572, 2016.