Determining the layer structure and thickness using model-based white light interferometry

Yian Cheng
University of Michigan
Ann Arbor, Michigan, USA
yianch@umich.edu

Mingqian Ma
University of Michigan
Ann Arbor, Michigan, USA
mamq@umich.edu

Changwen Xu
University of Michigan
Ann Arbor, Michigan, USA
changwex@umich.edu

Abstract

Accurate thin-film thickness measurement is crucial in various high-tech fields, including electronics, optics, and materials science. Traditional methods like Atomic Force Microscopy (AFM) and X-ray Reflectometry (XRR) are reliable but face limitations such as high cost, contact-induced damage, and inefficiencies in multi-layer structures. To overcome these challenges, we propose a model-based approach integrating White-Light Interferometry (WLI) with Spectral Reflectometry (SR), achieving non-destructive, high-accuracy measurements. We propose a deep-learning-based model pipeline that incorporates denoising, structure classification and layers' length regression by a hybrid combination of U-Net, CNN and MLP structures. This research provides a robust, scalable framework for thin-film characterization, meeting the stringent precision demands of modern industrial applications.

Our website can be found HERE.

Keywords

White Light Interferometry, Regression, Encoder-Decoder Architecture, Thin Film Thickness Measurement, Spectral Reflectometry

ACM Reference Format:

Yian Cheng, Mingqian Ma, and Changwen Xu. 2024. Determining the layer structure and thickness using model-based white light interferometry . In . ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Thin-film measurement is a critical process in materials science and engineering, especially in fields like electronics, optics, and coatings, where precise control of film thickness is essential for ensuring proper device functionality and performance. High-tech industries rely heavily on accurate thickness measurements for quality control, particularly when working with films typically less than 1 µm in thickness. Traditional measurement techniques, such as stylus profilometry, Atomic Force Microscopy (AFM) [8], and X-ray Reflectometry (XRR) [4], are well-established; however, they come with notable limitations. Contact methods like stylus profilometry risk damaging delicate films, while methods such as AFM and XRR are time-consuming, costly, and face difficulties in accurately measuring multi-layer structures due to complex interactions with light, including multiple reflections and absorptions.

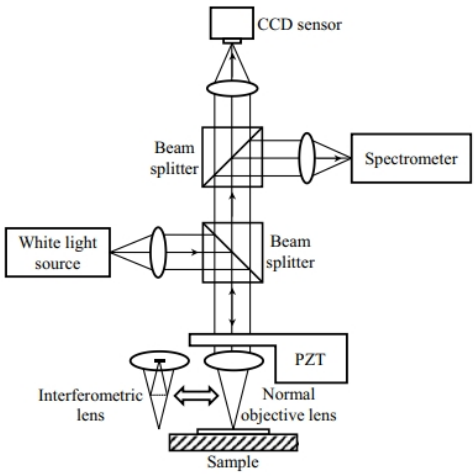

Figure 1: Optical configuration in spectrally resolved white-light interferometry [9]

To overcome these limitations, non-destructive optical methods such as spectral reflectometry (SR) offer a promising alternative. SR, particularly when combined with white-light interferometry (WLI), provides a reliable approach for high-accuracy measurements over large areas without compromising resolution. In this project, we leverage this combined method to measure and analyze the layer structure and thickness of a wafer with a target measurement error of 1–2 nm. Using a commercial WLI instrument, we aim to characterize multiple known structures composed of materials like SiO_2, Si_3N_4, $p-Si$, and $BK7$ across different wafer regions.

In this project, we aim to enhance thin-film measurement accuracy and efficiency by employing model-based analysis techniques. We will specifically address the challenges posed by multi-layer structures and noise interference, which are common issues in thin-film metrology. Our contributions include developing and evaluating a unified model for predicting both layer composition and thickness from spectral data, as well as incorporating noise-resilient processing methods to improve robustness. Through these advancements, this work seeks to provide a more effective and scalable approach to thin-film measurement that meets the precision demands of modern high-tech industries.

CSE598, Nov. 2024, Michigan, USA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

2 Problem Description

Thin-film thickness measurement is a vital process in many high-technology fields, including electronics, optics, and coatings, where precise control over material dimensions impacts device functionality and performance. Traditional methods like stylus profilometry, Atomic Force Microscopy (AFM), and X-ray Reflectometry (XRR) have proven accurate but are limited by several factors: contact-based approaches can damage fragile films, while other methods are time-consuming, costly, and less effective when dealing with multi-layer structures due to complex light reflections and absorptions. These challenges have driven the exploration of non-destructive optical approaches, particularly Spectral Reflectometry (SR) in combination with White-Light Interferometry (WLI) as shown in the Fig 1 which offers high accuracy over large surface areas without sacrificing resolution.

SR is an optical measurement technique that determines thin-film thickness by analyzing interference patterns in light reflected from the film's interfaces. The combination of SR and WLI enhances this approach, as WLI enables differentiation between film thickness and surface profile information by modulating the optical path difference (OPD) between a reference mirror and the sample. Low-frequency domains in the resulting spectra represent thickness information, while high-frequency domains capture surface profiles. This combined method allows simultaneous measurement of both film thickness and surface structure.

In this project, we aim to determine the structure and thickness of each layer within a 4"-sized wafer sample, which contains multiple known material configurations (e.g., SiO_2, Si_3N_4, p-Si, BK7). Our objective is to achieve a target error margin of 5–10 nm in thickness measurement. This measurement precision is essential for quality control in manufacturing settings where material consistency directly affects product reliability.

To tackle this problem effectively, we frame it as an **inverse problem** where the goal is to infer physical layer structures from observed spectral data. Unlike traditional forward problems, where the properties of a given structure are calculated, an inverse problem requires determining the underlying structure from the observed spectrum. This inverse approach is challenging due to the non-unique mapping between spectra and layer configurations, especially in the presence of noise.

The ML component of this project addresses this inverse problem using a model-based analysis with machine learning. We aim to develop a predictive model that learns to map spectral data inputs (the reflectance spectra collected by the WLI system) to target outputs representing both layer composition and thickness $T = \{t_1, t_2, ..., t_n\}$, where t_i denotes the thickness of the layer i. Our data will consist of generated spectra for various layer configurations, allowing the model to learn patterns in spectral variation corresponding to different material compositions and thicknesses.

This project will implement a unified model that performs both classification and regression. The model first classifies the spectral data to determine the combination of materials present, then predicts each layer's thickness in the detected configuration. By leveraging this ML approach, we expect to handle the inverse design problem more efficiently, overcoming the limitations of traditional methods and improving the robustness and scalability of thin-film measurements for high-tech applications.

3 Related Works

A computationally efficient theoretical model for low-coherence interferometric profilers has been raised by [5], where our data generator is built upon. This model integrates both geometric and spectral effects through an incoherent superposition of ray bundles spanning a range of wavelengths, incident angles, and pupil plane coordinates. It provides a physical model that gives insight into the key characteristics of interferometric signals, especially in surface and thin-film metrology. Another work utilizing model-based algorithms is raised by ZYGO Corporation [6] to measure the thickness of thin films using their commercial white-light interferometric instruments. This model separates the calibration into two parts with pupil and field calibration and it enables the full-field film surface metrology over a full range of interferometric magnification. While this is quite related and directly put forward by the company, it also finds out that "poor knowledge of material properties will limit performance and may even prevent a meaningful measurement." And it fails to measure the film that is thinner than $50nm$. Those model-based analyses pose limitations to the thickness determination. The interference spectra, especially when materials have absorption properties, can distort the measurements, necessitating more advanced models.

Processing spectral data using machine learning models has gained significant traction in thin film analysis. Traditionally, white light interferometry spectra have been interpreted using physical models and simulations to extract material properties such as composition and layer thickness. However, these methods can be computationally intensive and may not fully capture the complexities of real-world data. As a result, machine learning approaches have been employed to address these challenges and provide more efficient and accurate data-driven solutions.·

Convolutional Neural Networks (CNNs) [10] have been particularly effective in processing spectral data due to their ability to capture local patterns and hierarchical features within the spectra. CNNs can effectively identify subtle variations associated with different material compositions and layer structures, making them suitable for classification tasks in spectral analysis. For instance, CNNs have been employed for classification of Raman Spectra [12], classification of pharmaceutical tablets using VIS-NIR spectroscopy [2], and predicting drug content in tablets from near infrared (NIR) spectra [3].

In recent years, Transformer models [15] have emerged as a powerful tool for processing spectral data due to their ability to model long-range dependencies and complex interactions within input sequences [7]. Originally developed for natural language processing tasks, Transformers are particularly well-suited for spectra analysis because they use self-attention mechanisms to assign varying levels of importance to different regions of the input spectra. In the recent work [16], researchers demonstrate that transformer-based models significantly outperform 1D-CNNs and recurrent neural networks (RNN) when trained on spectral data, and they can even achieve equivalent performance as 2D-CNNs and 3D-CNNs which use much more than spectral information on image tasks.

Accurate thin-film measurements using Spectral Reflectometry (SR) combined with White-Light Interferometry (WLI) are often compromised by noise from environmental disturbances and instrument limitations. Traditional denoising techniques, such as Fourier Transform Filtering and Wavelet Transform, have been employed to mitigate these issues. Fourier Transform Filtering separates signal components based on frequency, allowing for the attenuation of high-frequency noise, but may not perform well with non-stationary signals like those in SR and WLI. Wavelet Transform decomposes signals into different frequency bands, enabling localized noise reduction, and has been widely applied in image and signal processing [13].

Advancements in deep learning have introduced data-driven approaches to signal denoising, offering several advantages. For example, CNN-based autoencoders have been employed to denoise lidar signals, demonstrating superior performance over traditional methods by effectively capturing complex noise characteristics [11]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are adept at modeling temporal dependencies in sequential data and have been applied to denoise electrocardiogram (ECG) signals, effectively preserving temporal structures while reducing noise [1]. Generative Adversarial Networks (GANs) consist of a generator and a discriminator network, trained adversarially to produce denoised signals that are indistinguishable from clean signals, and have been applied to denoise ECG signals, demonstrating improved performance over traditional methods [14]. Deep learning methods offer the advantage of learning complex noise patterns directly from data, reducing the need for manual feature extraction and parameter tuning. However, they require large datasets for training and significant computational resources.

4 Dataset Preparation

To prepare our dataset, we generate spectra for multilayer thin-film structures to address the inverse design problem, where the goal is to predict a material's physical structure based on its observed optical spectrum. While simulating the spectrum response of an arbitrary structure can be done quickly using physical simulators, performing inverse design on spectra lacks a fast, direct approach. To train a large model for this purpose, we rely on a synthesized dataset to provide sufficient data. By generating synthetic spectra for known multilayer structures, we can use these inverted data pairs to train machine-learning models that later infer material configurations from real-world spectral measurements.

The dataset, referred to as "I4D," was generated through a multilayer optical interferometry simulation, capturing interferometric signals across multiple parameters and capturing data in four dimensions including the time, spatial, spectral, and polarization status. This preparation involved several stages to ensure that the data would reflect realistic conditions for thin-film measurements and that it was comprehensive enough for model training. A total of eight structure datasets were generated to represent common scenarios in a particular application, with each dataset corresponding to a specific layer structure and varying layer thicknesses. For initial network testing, we focus specifically on a three-layer structure composed of SiO_2, Si_3N_4 and SiO_2. The thickness of each layer was varied within a specified range (0.1 to 0.5 μm) to provide a broad coverage of layer combinations, which was discretized to yield a resolution of 200 data points per layer, resulting in a dataset size of 200^3 points. To reduce the dataset size, we randomly selected 0.5% of these points, producing the "I4D Large" dataset, making the dataset to be of dimension $40,000 \times 279$ with ground truth, where 279 is the length per data. To facilitate efficient training and comparison, we also generated a smaller version, "I4D Small," with a reduced resolution of 100 data points per layer, and the dataset size reduces to $5,000 \times 279$.

Table 1: Datasets generated with eight specific structures

Dataset	Layers	Substrate
0	$SiO_2/Si_3N_4/SiO_2$	BK7
1	$SiO_2/Si_3N_4/SiO_2/Si_3N_4$	Glass
2	$SiO_2/Si_3N_4/SiO_2/p-Si/SiO_2$	BK7
3	$SiO_2/Si_3N_4/SiO_2/p-Si/Si_3N_4$	Glass
4	$SiO_2/Si_3N_4/SiO_2$	Si
5	$SiO_2/Si_3N_4/SiO_2/Si_3N_4$	Si
6	$SiO_2/Si_3N_4/SiO_2/p-Si/SiO_2$	Si
7	$SiO_2/Si_3N_4/SiO_2/p-Si/Si_3N_4$	Si

Key parameters in the simulation included the numerical aperture (NA) and sample height of the interferometric setup. The nominal NA was set to 0.3, with an error level randomly sampled within a ±10% range to simulate slight variations, reflecting potential inconsistencies in real-world setups. For simplicity, the sample height, which influences the overall shape of the generated data, was set to zero in this initial stage.

In real-world applications, optical measurements are rarely noise-free. To create a dataset that generalizes well to real-world scenarios, we introduced random noise into the generated spectra. Both Gaussian and Poisson noise were added to the interferograms, simulating realistic noise conditions and enhancing the model's robustness in handling noisy data. Different noise levels were applied to achieve signal-to-noise ratios (SNRs) of 30, 40, 50 dB, with those chosen to align with typical real-world data. Observations with added noise show significant deviation, ensuring the model learns to infer structures accurately even under challenging conditions.

With these two main datasets generated, we further preprocessed the data by applying traditional signal processing methods for denoising. Afterward, the data were transformed into the frequency domain, and specific key segments were cropped, reducing the data length from 279 to 50 points. This transformation and cropping help retain essential information while optimizing the dataset for model input. These final datasets will be described further in Section 5.1.1.

As the "I4D Large" dataset gives the best result, we generated eight datasets containing this 1. The eight datasets correspond to eight unique structures that are vital in a specific application. To avoid data explosion, we only double the generated quantities of the generated data onto the four- and five-layer structure, making the data inside each of those datasets with more than three layers 80,000.

5 Method

5.1 Data Denoising

As the real experimental value is always noisy, we apply various methods to denoise the model.

5.1.1 Traditional Denoising. An alternative way to perform inverse design from noisy data is to train a denoising model and apply it before the inverse design model. We here also explore this possibility. As the real-world noise distribution may be complicated, we explore both the possibilities of traditional denoising methods from signal processing as well as new deep-learning denoising methods.

In traditional signal processing, denoising signals, particularly in the frequency domain, is essential for improving data quality, especially when working with phase data. Several effective signal-processing techniques are available to help reduce noise. Research indicates that noise in White-Light Interferometry (WLI) primarily affects the magnitude component after applying a Fourier Transform. The method operates as follows: it first transforms the signal to the frequency domain, then identifies the spectral magnitude data, and finally crops the useful phase data based on the meaningful magnitude. These refined phase data are then stored as a dataset for efficient computation and comparison.

5.1.2 Deep-Learning Based Denoising. We utilize a deep-learning-based denoising approach to enhance the quality of the spectral data and improve the accuracy of thin-film thickness measurements. Specifically, we employ a U-Net architecture, which has demonstrated exceptional performance in signal and image denoising tasks due to its encoder-decoder structure with skip connections. The encoder extracts multi-scale features from the noisy input spectra, compressing it into a latent representation that captures the essential characteristics of the signal. The decoder then reconstructs the denoised spectra by upsampling this representation, with the skip connections ensuring the preservation of fine-grained details from the original input.

For this project, the U-Net model is trained using paired datasets of noisy and clean spectra generated from our simulation process. The loss function combines Mean Squared Error (MSE) to minimize reconstruction errors and a perceptual loss term to ensure the model learns spectral features critical for thin-film measurements.

5.2 Two-Stage Classification-Regression Model

Our initial approach involves a two-stage predictive pipeline for both regression and classification. The first stage is a classification model trained to categorize the input spectra into one of the eight possible layer combinations. Each combination represents a unique sequence of material compositions in the thin film structure. We utilize a neural network encoder such as a Convolutional Neural Network (CNN) to capture the intricate patterns in the spectra associated with different layer sequences. The CNN processes the spectral data to extract relevant features that distinguish between the different combinations using a final classification head.

Following classification, the second stage of our model consists of eight separate regression models, each corresponding to one of the classified combinations. These regression models are separately trained to predict the thicknesses of each layer within the model's specific combination. By isolating the regression tasks, each model can focus on the relationships between the spectra and the layer thicknesses for its assigned sequence. The regression models employ multi-layer perception (MLP) prediction heads to map the extracted features to precise thickness predictions of each layer.

We apply this two stage regression-classification model due to the inhomogeneity of the problems. A unified regression-classification model cannot learn the implicit mapping of the combined problem. We deduce the underlying reason is because that the structures are very diverse so a simple regression problem cannot handle well the regression task even with the help of classification head to identify the proper structure. We'll demonstrate more details of the implementation and the comparison in the experiment section.

6 Experiments

6.1 Deep Learning for Signal Denoising

We employ a deep learning approach to denoise the signal data using a U-Net architecture. This model captures complex patterns in the noisy data, effectively learning to reconstruct the underlying clean signal. The model was trained with mean squared error (MSE) as the loss function, optimizing it to minimize the reconstruction error between the denoised output and the ground truth.

The results demonstrate promising denoising capabilities. Specifically, the average Mean Squared Error (MSE) between the denoised output and the ground truth was observed to be 0.0006, indicating high accuracy in reconstructing the original signal. Additionally, we achieved an average Peak Signal-to-Noise Ratio (PSNR) of 32.58 dB, highlighting the model's effectiveness in reducing noise while preserving signal quality.

Figure 3 presents qualitative results from the denoising model, showcasing the noisy input, the denoised output, and the ground truth signal for a subset of test cases (We show 2 for simplicity; please refer to our slides for more). As shown, the model successfully attenuates noise while closely aligning the output with the ground truth signal.

Figure 4 is the frequency domain plot from left data point shown in Figure 3. we can see that the U-Net can effectively suppress the error level in the magnitude plot of the frequency domain, but not very well in the phase plot. This is normal as the phase noise is usually hard to remove.

6.2 Classification and Regression

We trained Convolutional Neural Network (CNN) as the encoder for representation learning of white light spectrum signals. The CNN model utilized 1D convolutional layers with the input data formatted as a single channel. The architecture progressively increased the number of channels while decreasing the dimensionality within each channel through max pooling. Each convolutional layer was followed by a BatchNorm, an activation function, and a max pooling layer to reduce dimensionality and capture essential features effectively. To justify our choice of CNN as the encoder, we also trained a Multi-Layer Perceptron (MLP) whose architecture follows a pyramidal structure, initially increasing the dimensionality of the hidden layers before gradually decreasing it.

We first evaluated our model on one of the eight datasets so that the thin film architecture is fixed and we only need to predict the thin film thickness for each composition. We used the first dataset

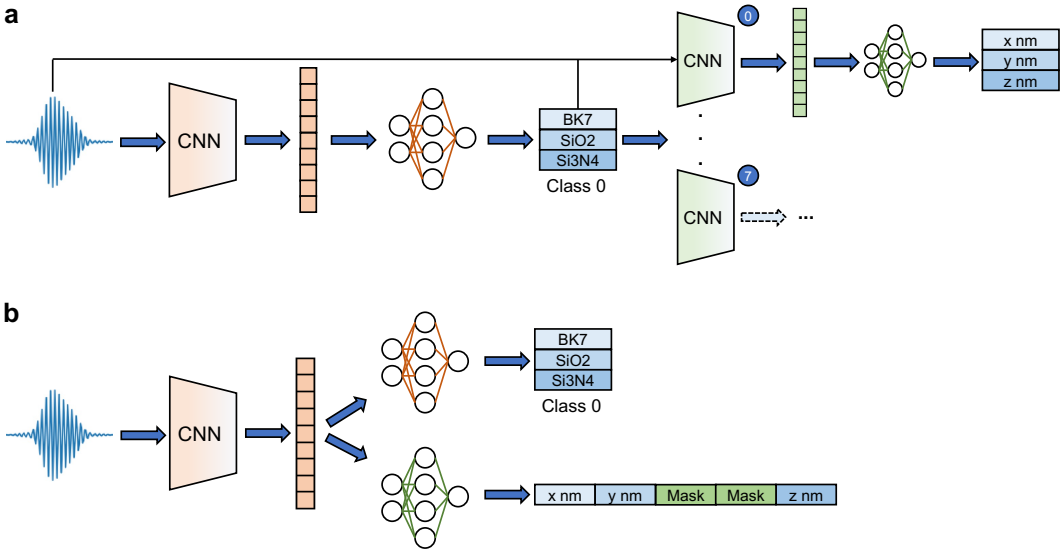

Figure 2: Illustration of (a) training separate classification and regression models (b) training a single model which integrates classification and regression.

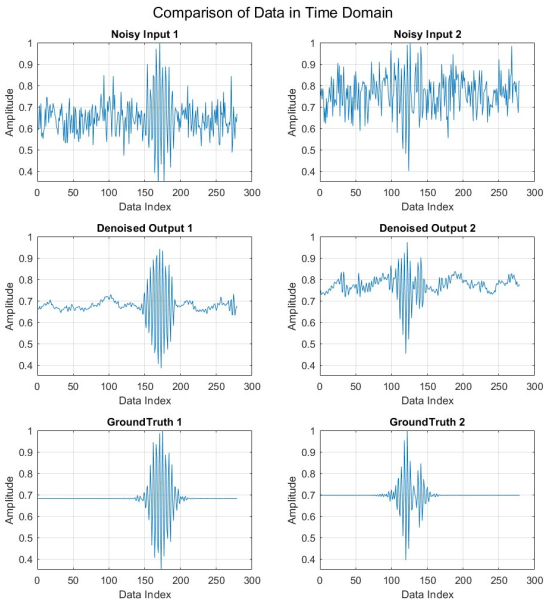

Figure 3: Examples of Signal Denoising using UNet

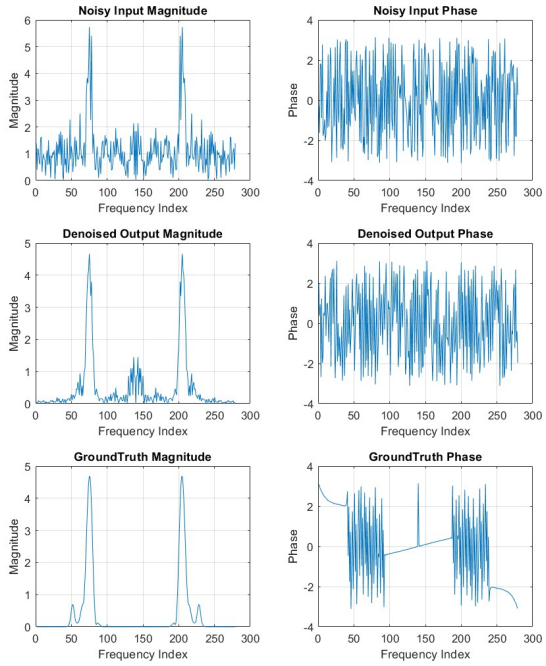

Figure 4: Frequency Plot of One Data Point

for this experiment. For this dataset, the thin film architecture consists of three layers: Layer 1 (SiO_2), Layer 2 (Si_3N_4), and Layer 3 (SiO_2), and we generate four versions: I4D Small, I4D Large, I4D Cropped Small, and I4D Cropped Large, as mentioned in the Data Preparation section. Each dataset was evaluated both with and without the introduction of noise. An 8:1:1 split was employed for the training, validation, and testing datasets, respectively. Overall,

Table 2: Mean Squared Error (MSE) of MLP and CNN on the test set of the four datasets we generated. MLP is used as the baseline model.

Dataset	Feature	MLP	CNN
I4D Small	No noise	0.0126	0.0070
	Noise	0.0134	0.0091
I4D Large	No noise	0.0128	0.0014
	Noise	0.0134	0.0077
I4D Cropped Small	No noise	0.0128	0.0125
	Noise	0.0129	0.0127
I4D Cropped Large	No noise	0.0124	0.0107
	Noise	0.0131	0.0126

the CNN model outperformed the MLP in all the cases, as shown in Table 2. The superior performance of the CNN can be attributed to its ability to capture local patterns and hierarchical features in the spectral data through convolutional layers. The introduction of noise into the datasets resulted in a consistent degradation of model performance for both MLP and CNN architectures. For instance, in the I4D Small dataset, the MLP's MSE increased from 0.0126 (no noise) to 0.0134 (with noise), while the CNN's MSE increased from 0.0077 to 0.0097 under the same conditions. The sensitivity to noise makes it challenging to apply the models to experimental data which are more noisy than simulation data.

When eight different thin film architectures are presented, we first trained a CNN model to classify the thin films based on the signals. The CNN encoder is concatenated with an output head which predicts the probability that the data point belongs to each class and the data point is assigned to the class with the highest probability as the predicted label. During training, the cross entropy is used as the loss function, and the model is evaluated according to the prediction accuracy. The model achieves 93.14% on the test set. In comparison, MLP only achieves 33.86% accuracy on the test set. Such experimental results demonstrate that CNN has an edge over MLP for signal processing and representation learning.

Then we trained separate models for the eight thin film architecture datasets. Similar to previous experimental settings, we employed an 8/1/1 data split. Table 3 summarizes the MSE values of CNN on the eight thin film architectures. The CNN model achieves low MSE on all eight datasets when the features without noise are used as input. When there are noises introduced to the white light interferometry spectrum, the MSE values significantly increase. This again highlights the sensitivity of the model performance to noise in the input data.

Given the superior performance on classification and regression of CNN on non-noisy input, it's natural to integrate the classification and regression model into a single one. Consequently, the model can efficiently infer the compositions in the thin film and predict their thicknesses. Meanwhile, the single encoder is expected to benefit from gradients backpropagated from both the classification and regressor head so that it can learn richer information than training classification and regression models separately.

Therefore, we train the as shown in Figure 2(b). We adopted the same training strategy as previous experiments. Although the

Table 3: Mean Squared Error (MSE) of CNN on the test set of the eight datasets we generated.

Dataset	Input feature	
	No noise	With Noise
0	0.0014	0.0077
1	0.0003	0.0051
2	0.0036	0.0056
3	0.0008	0.0057
4	0.0001	0.0035
5	0.0006	0.0054
6	0.0010	0.0055
7	0.0006	0.0049

Table 4: Model performance of integrating UNet and CNN.

Input feature	Classification accuracy	Regression MSE
No noise	93.14%	0.0014
Noise	72.92%	0.0077
UNet denoised	71.72%	0.0075

classification accuracy on the test set achieves 92.89% which is comparable to the separate classification model, the regression MSE is only 0.2471 which is quite poor, even though we use non-noisy signals. Therefore, more efforts are required to improve the performance of the integrated model.

In order to deal with the poor model performance when trained on noisy data, we incorporate the pretraiend denoising UNet model so that the CNN encoder will take the denoised signals as input. We trained the model in this way and tested the model performance on classification on the eight architectures and regression on the first architecture. As shown in Table 4, although the UNet effectively denoised the noisy spectrum, the model performance is very close to the one without UNet preprocessing and significantly falls behind the one trained on non-noisy features.

7 Conclusion

In this work, we propose a general framework for inversely predicting the structure type and the material thickness for each layer from the input white light interferometry data with noise. We propose a combination of UNet for denoising, CNN for classification and MLP for layer thickness regression to address the problem.

We want to achieve the target error of around 1-2nm to enable the real-world applications; now we can only get 5nm. This, however, is quite promising in this field as many local optimization algorithms can fine-tune the thickness prediction once the error is not that far. We could target on getting smaller errors and then apply those algorithms.

8 Future Work

To address the limitations of our current models and enhance their applicability, our future work will focus on expanding the model's capabilities.

- In future work, we will further explore the possibility of applying a uniform regression-classification that can possibly learn the overall distribution of the white light interferometry with the structures and thickness uniformly.
- We plan to collect more real data to better train the denoising model to make it more robust to real-world application and fill in the distribution gap between simulated noise and real-world observation bias.

References

[1] Karol Antczak. 2018. Deep Recurrent Neural Networks for ECG Signal Denoising. *arXiv preprint arXiv:1807.11551* (2018).

[2] Kyung-Jin Baik, Jae Hyung Lee, Youngsik Kim, and Byung-Jun Jang. 2017. Pharmaceutical tablet classification using a portable spectrometer with combinations of visible and near-infrared spectra. In *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 1011–1014.

[3] Esben Jannik Bjerrum, Mads Glahder, and Thomas Skov. 2017. Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics. *arXiv preprint arXiv:1710.01927* (2017).

[4] E Chason and TM Mayer. 1997. Thin film and surface characterization by specular X-ray reflectivity. *Critical Reviews in Solid State and Material Sciences* 22, 1 (1997), 1–67.

[5] Peter de Groot and Xavier Colonna de Lega. 2004. Signal modeling for low-coherence height-scanning interference microscopy. *Applied optics* 43, 25 (2004), 4821–4830.

[6] Martin F Fay and Thomas Dresel. 2017. Applications of model-based transparent surface films analysis using coherence-scanning interferometry. *Optical engineering* 56, 11 (2017), 111709–111709.

[7] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. 2021. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), 1–15.

[8] S Jena, R Tokas, S Thakur, and NK Sahoo. 2015. Characterization of optical thin films by spectrophotometry and atomic force microscopy. *SMC Bulletin* 6, 1 (2015), 1–9.

[9] Taeyong Jo, KwangRak Kim, SeongRyong Kim, and HeuiJae Pahk. 2014. Thickness and surface measurement of transparent thin-film layers using white light scanning interferometry combined with reflectometry. *Journal of the Optical Society of Korea* 18, 3 (2014), 236–243.

[10] Yann LeCun, Yoshua Bengio, et al. 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361, 10 (1995), 1995.

[11] Xinqi Li, Yijian Zhang, Yue Sun, Yu Wang, Shijie Li, Kai Zhong, Zhaoai Yan, Degang Xu, and Jianquan Yao. 2022. New Denoising Method for Lidar Signal by the WT-VMD Joint Algorithm. *Sensors* 22, 16 (2022), 5978.

[12] Jinchao Liu, Margarita Osadchy, Lorna Ashton, Michael Foster, Christopher J Solomon, and Stuart J Gibson. 2017. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. *Analyst* 142, 21 (2017), 4067–4074.

[13] Stephane Mallat. 1999. A wavelet tour of signal processing. *Academic press* (1999).

[14] Chandresh Pravin and Varun Ojha. 2022. Denoising ECG signals and their analysis using Hybrid Deep learning model. *IEEE Access* 10 (2022), 10150811.

[15] A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).

[16] Felipe Viel, Renato Cotrim Maciel, Laio Oriel Seman, Cesar Albenes Zeferino, Eduardo Augusto Bezerra, and Valderi Reis Quietinho Leithardt. 2023. Hyperspectral image classification: An analysis employing CNN, LSTM, transformer, and attention mechanism. *IEEE access* 11 (2023), 24835–24850.