MIT Open Access Articles

This is a supplemental file for an item in DSpace@MIT

**Item title:** Deep Learning for Rapid Analysis of Spectroscopic Ellipsometry Data

**Link back to the item:** https://hdl.handle.net/1721.1/139811
Deep learning for rapid analysis of spectroscopic ellipsometry data
Yifei Li, Yifeng Wu, Heshan Yu, Ichiro Takeuchi, R. Jaramillo*

Yifei Li, Rafael Jaramillo
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
*email: rjaramil@mit.edu

Yifeng Wu
853 Commodore Dr, Apt. 460, San Bruno, CA, USA.

Dr. Heshan Yu, Ichiro Takeuchi
Department of Materials Science and Engineering, University of Maryland, College Park, MD, USA.

Keywords: spectroscopic ellipsometry, deep learning, high-throughput, phase-change materials

Abstract
High-throughput experimental approaches to rapidly develop new materials require high-throughput data analysis methods to match. Spectroscopic ellipsometry is a powerful method of optical properties characterization, but for unknown materials and/or layer structures the data analysis using traditional methods of nonlinear regression is too slow for autonomous, closed-loop, high-throughput experimentation. Here we introduce and study three methods (termed spectral, piecewise, and pointwise) of spectroscopic ellipsometry data analysis based on deep learning. After initial training, the incremental time for inferring optical properties can be a thousand times faster than traditional methods. We present results for multilayer sample structures with optically isotropic materials, appropriate for high-throughput studies of thin films of phase-change materials such as Ge-Sb-Te (GST) alloys. We also present results for studies on highly-birefringent layered materials, exemplified by the transition metal dichalcogenide MoS$_2$. We discuss how the materials under test and the experimental objectives may guide the choice of analysis methods. We demonstrate the utility of our approach by analyzing data measured on a composition spread of Ge-Sb-Te phase-change alloys containing 177 distinct compositions, and identifying the composition with optimal phase-change figure of merit in only 1.4 sec of analysis time.
1. Introduction

High-throughput materials development aided by machine learning and big data resources has been a mainstay of materials science and engineering for nearly a half century, with many commercial successes in the field of structural materials.[1] More recently, efforts have broadened to include materials for electronics, photonics, optoelectronics, and energy conversion and storage.[2,3] As the diversity of materials addressed by high-throughput approaches expands, new bottlenecks must be overcome. On the experimental side, bottlenecks are encountered at multiple stages of materials synthesis and characterization.

Here we focus on optical properties characterization by spectroscopic ellipsometry, relevant for the development of materials for photonics applications, including phase-change materials for strong modulation of optical phase and amplitude in confined geometry.[4] Spectroscopic ellipsometry is increasingly relevant for rapid in-line metrology, due to developments in accelerating data acquisition.[5] We demonstrate the use of a deep learning (DL) algorithm to quickly analyze spectroscopic ellipsometry data measured on a thin film with a composition spread in the Ge-Sb-Te (GST) class of phase-change materials. We use our results to calculate the phase-change figure-of-merit (FoM), which depends on the complex refractive index in both the crystalline and amorphous phases. A map of FoM over the composition spread identifies the composition most promising for application. Our DL algorithm infers film complex refractive index and thickness approximately one thousand times faster than traditional nonlinear regression. This speedup is important for integrating spectroscopic ellipsometry in high-throughput and autonomous experimental materials development, and for related applications in high-speed metrology and inspection. We also present the application of DL to analyze spectroscopic ellipsometry data measured on layered (*i.e.* two-dimensional, 2D) materials, and discuss the limitations imposed by materials with large refractive index and strong birefringence.

Spectroscopic ellipsometry is well-poised for analysis by DL because it is an information-rich experimental method that for practical reasons is often analyzed with highly-simplified models. Our work is an extension of previous work developing neural networks and related methods of experimental data analysis for spectroscopic ellipsometry, begun decades ago.[6] Our work also relates to ongoing work to develop DL for other methods for which the information content in the data is often greater than conventional analysis can infer; examples include X-ray diffraction and computational design of photonic devices.[7,8]
1.1. Introduction to data analysis in spectroscopic ellipsometry

Ellipsometry is an optical characterization method that measures the change of polarized light reflected (or transmitted) from flat samples. We show in Figure 1a a schematic diagram of reflection ellipsometry. As a high-precision and non-destructive measurement, ellipsometry is a mainstay method for measuring material properties (e.g., refractive index) and characterizing device processing (e.g., layer thickness). Ellipsometry is highly-valued in commercial semiconductor wafer processing, due to its accuracy in characterizing thin films such as oxides on semiconductor substrates, and its ability to map large wafers. With proper modelling and calibration, ellipsometry can also characterize materials properties besides optical properties such as composition, crystallinity, doping, and electrical conductivity.

Analyzing ellipsometry data amounts to solving an inverse problem. For a specific optical model including sample geometry and incidence plane, there exists an analytic forward function that maps material properties to experimental data: ellipsometric angles ($\Psi$ and $\Delta$) as a function of wavelength ($\lambda$) and angle of incidence (AoI). To interpret materials properties from experimental data, the inverse function (i.e., the backward function) needs to be solved, explicitly or implicitly. However, the forward function is complex and nonlinear, which makes inversion difficult. For well-understood materials and a limited number of unknowns (e.g. oxide thickness), ellipsometry data is straightforward to analyze. However, in more complex scenarios the analysis is often inaccurate, highly time-consuming, or both. Such can be the case when material properties or the sample structure are unknown.

There are several established approaches used to derive optical properties from spectroscopic ellipsometry data by inverting the forward function. The most common approach is nonlinear regression, combined with models that represent optical materials properties using a small number of parameters, such as Drude and Tauc-Lorentz models. A related approach is spline analysis.\cite{9,10} In this approach, the imaginary part of the dielectric function is parametrized by a piecewise polynomial, and the real part is calculated by the Kramers-Kronig (KK) relation. In this case, the nonlinear regression acts on the spline parameterization, rather than on the underlying spectral model. The spline approach avoids assuming particular models for the material and can improve accuracy, but comes with an added computational cost. A third approach may be termed wavelength-by-wavelength.\cite{11,12} At a given wavelength, the forward function depends only on the
sample properties, not on the evaluation at other wavelengths. Hence, at a given wavelength, the input and output tensors of the forward function have relatively low dimensionality (e.g., less than 10), which makes it possible to analytically approximate the inverse function. However, this approach is valid only when the forward function is one-to-one, and there is no simple criterion to determine whether this is the case. Also, since the function is evaluated wavelength-by-wavelength, this approach doesn’t guarantee the continuity, smoothness, and KK-consistency of the inferred spectrum. Each of these established approaches can be time-consuming, inaccurate, and may involve strong assumptions.

1.2. Introduction to deep learning

The three methods introduced above can also be considered (classified) as a type of conventional machine learning (ML). These traditional ML algorithms train a mathematical model based on experimental data, and the training parameters are the material properties of interest, such as the refractive index. The spectroscopic ellipsometry inverse problem involves a large number of parameters, on the same order as the number of wavelengths measured. Training time is proportional to the parameter space volume, which grows exponentially with the number of parameters. Conventional machine learning based on nonlinear regression may converge reasonably well, but only when the model parameters are initialized well, and this often requires much experience.

These challenges are well addressed by DL. DL refers to a large class of ML approaches based on artificial neural networks (ANNs), and is extensively used in many aspects of modern society, from natural language processing to advertisement bidding.\textsuperscript{13} DL is also increasingly useful for materials science.\textsuperscript{7,14} Like conventional ML, DL involves training a mathematical model based on data. However, the model is a function that maps experimental data to materials properties, and the model parameters do not themselves have immediate physical significance. Compared to conventional ML, the accuracy and analysis speed of DL increases exponentially with the size of the training data set. DL can also learn the inverse function on a lower-dimensional manifold instead of over the entire parameter space. These advantages counter the exponential complexity arising from geometrically-large parameter spaces, making DL feasible in many cases for which conventional ML is challenging. Further, ANN evaluation involves only matrix multiplication and function evaluation, which are orders-of-magnitude faster than nonlinear regression. ANN training emulates the process by which an experienced expert may initialize model parameters for nonlinear
regression. Therefore, although training ANNs may take hours, DL can subsequently process large data sets much faster than conventional ML. This is an important advantage for high-throughput analysis of spectroscopic ellipsometry data to assist rapid materials development.

As with any method of data analysis, the accuracy of the results inferred from experimental data cannot be quantified if the ground truth (i.e. sample properties in their entirety) remains unknown. This challenge affects conventional ML and DL alike. Below we address this challenge by quantifying accuracy relative to the ground truth for the case of simulated data, and relative to ellipsometry data for the case of experimental data.

Figure 1: Schematics of experimental method and deep learning model. (a) Schematic diagram of an ellipsometry experiment; $\theta$ = angle of incidence (AoI). (b) Schematic diagram of an artificial neural network with one convolutional layer and a multi-layer perceptron with two layers and residual connection

2. Models and Methods

2.1. Neural network design choices

We choose a multi-layer perceptron (MLP) with the rectified linear unit (ReLU) activation function. ReLU is defined as

$$f(x) = \max(0, x)$$

Eqn. 2.1
In back-propagation, ReLU gives either constant gradient (positive input) or zero gradient (negative input). Hence, compared to other widely-used activation functions (e.g. sigmoid and tanh), ReLU has fewer vanishing-gradient problem and higher training efficiency. Also, ReLU has the lowest computational cost, making it a good match for high-throughput analysis. Another advantage of ReLU in real-value inference is that ReLU is unbounded. Therefore, an ANN using ReLU can infer values outside of the range of training data.

For several of our analyses (model for isotropic thin film, and spectral and piecewise models for birefringent samples, defined below) we use a convolutional neural network (CNN) layer to improve the training efficiency. The convolutional operation is defined as

\[ S(i, j) = \sum_n \sum_m I(i + m, j + n)K(m, n) \]  

Eqn. 2.2

\( S \) is the output tensor or feature matrix. \( I \) is the input tensor (with size \( N_\lambda \times \text{number of AoIs in our case} \)), and \( K \) is kernel to be trained.\(^{[13]}\) Since both refractive index and the forward functions are smooth, the measured ellipsometric angles should be close for nearby wavelengths and AoIs. CNNs exploit such relationships and improve the smoothness of the inference results by using a limited number of parameters in the kernel, and by sharing parameters through convolution operations.

Since the forward function has a considerable linear component, we add residual connection to the MLP. Therefore, the output of each MLP layer is

\[ x^{l+1} = x^l + \text{ReLU}(W^l x^l) \]  

Eqn. 2.3

where \( x^l \) and \( W^l \) are the output tensor and weight matrix of layer \( l \), respectively. Our numerical experiments verify that the MLP has better training efficiency and accuracy with residual connection than without.

We use L1 and ridge regularization to avoid overfitting. For each model, we optimize the neural network parameters (layer number, layer size, and regularization coefficient) by direct search.

2.2. Experimental model and neural network for isotropic materials

We first consider a three-layer sample with isotropic materials, presented in Figure 2a. The top layer is a thin film of GST, the middle layer is silicon oxide, and the bottom layer is a semi-infinite Si wafer. The GST and oxide layer thicknesses are typically 30 and 2000 nm, respectively; the actual values are not known \textit{a priori}. The refractive indices of Si and SiO\(_2\) are known. Therefore, we require that the neural network infer the refractive index and thickness of the GST
film and the thickness of the SiO$_2$ layer from ellipsometric angles ($\Psi$ and $\Delta$) as a function of wavelength and AoI.

In the corresponding experiments, a sample response is measured at five values of AoI (50, 55, 60, 65, and 70°) from the ultraviolet (UV) to the near-infrared (NIR). Therefore, the input tensor has dimension $N_\lambda$ (number of wavelengths) $\times$ 5 (number of AoIs) $\times$ 2 ($\Psi$ and $\Delta$), and the output tensor has dimension $(N_\lambda \times 2 (n$ and $k)) + 2$ (layer thicknesses). As $N_\lambda$ increases, the inference error decreases but the training data set size, neural network size, and training time all increase. We choose $N_\lambda = 100$ to balance these trade-offs. The neural network consists of two layers of a 3-by-3 CNN and five layers of MLP with residual connection.$^{[15]}$

We generate a training data set by simulating spectroscopic ellipsometry data for a range of material properties and sample structures. We vary the GST film thickness in the range 25 – 35 nm, and the oxide layer thickness in the range 1900 – 2100 nm. We choose a smaller fractional variation of the oxide thickness because oxidized silicon wafers are made commercially by well-controlled processes, and the nominal oxide thickness is reliable. The generated refractive index spectra of the film should fulfill two criteria: $n$ and $k$ are to be in a reasonable range (e.g. 0 – 10 and $|n + ik| > 1$), and are to be KK-consistent. We fulfill these criteria by calculating Drude and Tauc-Lorentz models with bounded random parameters and combining the generated spectra in linear combinations. The training and the test data sets are free from random error.

We train the neural network using the optimizer ADAM.$^{[16]}$ After about 1,000 epochs, the loss function of the training data and the loss function of the test data converge to the same minimum value, which indicates that the network training has completed without overfitting. Generating training data takes approximately 3.5 CPU-hours on an Intel i7-9700K processor. Neural network training takes approximately 1 GPU-hours on a Nvidia GeForce RTX 2060 Super video card.

2.3. Experimental model and neural network for birefringent materials

In previous work, we performed spectroscopic ellipsometry measurements on transition metal dichalcogenide (TMD) crystals.$^{[17,18]}$ TMDs are highly anisotropic due to their layered crystal structures. We measured crystals of 2H-MoS$_2$, 1T-TiS$_2$, 1T-ZrS$_2$, 2H-MoTe$_2$, and 1T-MoTe$_2$, all of which are birefringent with the optical axis perpendicular to the crystal surface. We label the complex refractive index for polarization perpendicular and parallel to the optical axis as ordinary $(n_o, k_o)$ and extraordinary $(n_e, k_e)$, respectively.
Since the refractive index of TMDs is large, incident light is strongly refracted for practical values of the AoI. Nearly all optical measurements of TMDs are made on crystals with the layered face exposed. Therefore, within the material light propagates nearly-parallel to the optic axis, and as a result, ellipsometry is far more sensitive to the ordinary than to the extraordinary refractive index. Under these conditions, for measurements on semi-infinite TMD crystals, the ordinary refractive index is well-approximated by the so-called effective refractive index ($\bar{n}_{\text{eff}}$) (referred to as the pseudo refractive index in some literature), which is an analytic solution of the inverse model for the simple case of a semi-infinite isotropic sample:

$$\bar{n}_{\text{eff}} = \sin(\vartheta) \sqrt{1 + \tan^2(\vartheta) \left(\frac{1-\rho}{1+\rho}\right)^2}$$  \hspace{1cm} \text{Eqn. 2.4}

$\vartheta$ is the AoI and $\rho$ is defined as

$$\rho = \tan\Psi \exp(i\Delta)$$  \hspace{1cm} \text{Eqn. 2.5}

We note in passing that the effective refractive index can be useful in determining film thickness. If the refractive index of the material is known, then for any given AoI the anisotropic film can be modeled as an isotropic film, and the thickness can be efficiently determined by nonlinear fitting.

We consider three approaches to analyze spectroscopic ellipsometry data for highly-anisotropic materials: pointwise ($N_\lambda = 1$), piecewise ($N_\lambda = 10$), and spectral learning ($N_\lambda = 100$). The input tensor size is $N_\lambda \times 9$ (AoI) $\times 2$ ($\Psi$, $\Delta$), and the output tensor size is $N_\lambda \times 4$ ($n_o$, $k_o$, $n_e$, $k_e$). For spectral learning and piecewise learning, we use a neural network consisting of four layers of one-dimensional CNN and four layers of MLP with residual connection. For pointwise learning, we use seven layers of MLP without residual connection. In all cases we use ReLU as the activation function.

For spectral learning, the process of generating random training and testing data is similar to the case described in Sec. 2.2 of a multilayer sample with isotropic materials, but in this case the ordinary and extraordinary refractive index spectra are generated separately. In piecewise learning, the generated refractive index spectra have at most two monotone segments with limited slope, and lie within the range 0 - 10. Like spectral learning, piecewise learning provides continuous and smooth inference, but drops the KK-relation. In pointwise learning, $n_o$, $k_o$, $n_e$, and $k_e$ are generated by uniformly-distributed random variables ranging from 0 - 10.

Training uses the optimizer ADAM and takes approximately 1,000 epochs to converge. To generate training data takes approximately 30 min for spectral learning, 6 hrs for piecewise
learning, and 2 min for pointwise learning, on an Intel i7-9700K CPU. The training processes take approximately 30 min for spectral learning, 1 hr for piecewise learning, and 12 min for pointwise learning, on an Nvidia GeForce RTX 2060 Super graphics card.

Figure 2: Sample models used in this study. (a) Model of three-layer sample with isotropic materials. (b) Model of a semi-infinite, birefringent sample with optical axes normal to the exposed face.

3. Results and discussion

In this section we present the performance of our DL methods on simulated and experimental data. In Sec. 3.1 we evaluate the performance of our methods on simulated data, for which we can easily quantify the inference error. We measure inference accuracy by considering the absolute difference between the inferred $n \& k$ and the original model, and the percentage error in the inferred layer thickness. We calculate the mean absolute error (MAE) and the standard deviation
of the refractive index error, averaging the errors for \( n \) & \( k \), and of the thickness error. To demonstrate the advantage of DL, we also compare ANN-inferred results with the results from traditional nonlinear regression, performed using CompleteEASE software. In Sec. 3.2 we present analysis of experimental data measured on birefringent MoS\(_2\) crystals, and in Sec. 3.3 we present analysis of experimental data measured on a GST thin film with a composition spread.

When analyzing computed data, the ground truth is known, and therefore inference accuracy is simple to evaluate by comparing ground truth and inferred sample properties, as we do in Sec. 3.1. When analyzing experimental data, the ground truth is never fully known – in that no physical sample is ever fully and accurately characterized – and the inference accuracy has less clear meaning. In Sec. 3.2-3.3 we use the error between the experimental and reconstructed ellipsometry data to quantify inference accuracy.

### 3.1. Isotropic materials

In Table 1 we show the error statistics for the training run. The MAE and \( \sigma \) for the inferred refractive index are well under 0.1, which suggests that the method will be reliable for high-throughput materials characterization. The error percentage of the SiO\(_2\) layer thickness is significantly smaller than that of the film because we have more reliable prior knowledge of the SiO\(_2\) thickness.

| Inferred values      | MAE   | \( \sigma \) |
|----------------------|-------|-------------|
| film refractive index | 0.0329| 0.0578      |
| film thickness (%)   | 3.55  | 3.90        |
| SiO\(_2\) thickness (%) | 0.12  | 0.21        |

**Table 1:** The accuracy of the inferred film refractive index, film thickness, and SiO\(_2\) layer thickness as measured by the mean absolute error (MAE) and standard deviation (\( \sigma \)).

In Figure 3a we demonstrate an example of inferring sample properties from randomly-generated spectroscopic ellipsometry data. The MAE of the inferred refractive index is 0.0384. Using the inferred sample properties to reconstruct the forward model, we find that the MAE of the ellipsometric angles is 0.731° (average for both \( \Psi \) and \( \Delta \)) across the full spectrum. In Figure 3b we show that there is little-to-no correlation between the inferred film refractive index and film thickness, which is often a concern when analyzing ellipsometry data. In the inset of Figure 3b we show that the refractive index MAE histogram is clustered below 0.1.
In Figure 3c and 3d we infer the materials properties of a GST film from experimental data by two approaches: using our trained ANN, and using nonlinear regression. In nonlinear regression, we use a Tauc-Lorentz model with a single peak (as is common in analyzing GST data), and we manually adjust the initial parameters to approximate the data. We compare these approaches by visualizing reconstructed ellipsometry data, $\Psi$ at AoI = 55°. The ANN-inferred spectrum is within $\pm 0.5^\circ$ (orange shadow region) of the experimental value. The result of nonlinear regression is less accurate.

To further compare our trained ANN to nonlinear regression, we analyze 177 experimental ellipsometry spectra by both approaches. The error distributions (Figure 3e) demonstrate that the ANN produces more accurate results. The MAE between experimental spectra and reconstructed spectra from ANN-inferred results have an average of 0.911° across all 177 samples. Nonlinear regression produces an average MAE of 5.048°, five times larger than the DL result. This error is not completely unexpected given that nonlinear fitting was limited to a single-Tauc-Lorentz-peak model. DL is also much faster. With a preselected optical model, nonlinear regression takes approximately 30 minutes to analyze all data sets, whereas the ANN takes 700 ms. This thousand-fold speedup in solving the inverse problem of ellipsometry data analysis reflects the difference between linear computations and nonlinear regression. We note that the ReLU function is nonlinear, making ANN inference not completely linear, but it is nevertheless computationally very inexpensive.
Figure 3: Results of using our ANN to infer film refractive index and thickness for a multilayer sample with isotropic materials, as in Figure 2a. Panels (a-b) show results for randomly-generated data, and panels (c-d) show results for experimental data. The refractive index and thickness data correspond to the “GST film” layer as in Fig. 2a. (a) Comparing the inferred refractive index to the original, randomly-generated data. (b) The distribution of the refractive index MAE vs. film thickness error percentage. (Insert) Distribution of the refractive index MAE. (c) Experimentally-measured (at AoI 55°) ellipsometry angle $\Psi$ (presented as an orange band to indicate the ±0.5° error range), compared to the spectra reconstructed by our ANN, and by nonlinear regression (i.e.
manual fitting). (d) The refractive index determined by ANN inference and by manual fitting regression. (e) Error distributions of the ellipsometry angle from ANN inference and manual fitting.

To understand the sources of inaccuracy, in Figure 4 we show two examples drawn from generated data that are representative of (outlying) cases with significant inference error. In Figure 4a we show a case with a sharply-peaked refractive index, typical of a material with an absorption resonance. The inferred model reproduces the peak position, but there is substantial error near the peak maximum. Figure 4b shows that the corresponding reconstructed ellipsometric angle spectrum is nevertheless very close (MAE = 3.219°) to the generated one.

In Figure 4c we show a case with substantial dispersion on the short-wavelength side of the spectrum, but without an extremum. In this case, both $n$ and $k$ become badly-inferred near the end of the range. Figure 4d shows that the ellipsometric angle reconstructed from the inferred spectrum is accurate over most of the wavelength range, but becomes badly behaved on the short-wavelength side.

For the cases considered in Figure 4 the error distribution histograms are sharply-peaked around zero, but have long tails on the high-error side, see the insets to Figure 4a and 4b. For instance, the error distribution for $n$ shown in Figure 4a (inset) has MAE and $\sigma$ of 0.194 and 0.457, respectively; for the case shown in Figure 4c (inset), the numbers are 0.132 and 0.456. These are an order of magnitude higher than for cases without sharp extreme or strong dispersion at the edge of the spectral range. Nevertheless, even in such cases the DL approach is valuable for high-throughput material evaluation, as long as the limitations are borne in mind.
Figure 4: Examples of inference with significant error. (a-b) show a case with a significant extremum in the spectrum. (a) The inferred refractive index and the original, generated data. (insert) Histogram of the error distribution in $n$. Most cases have small errors (corresponding to the sharp peak around 0). Cases with large error as shown in the main panel are rare but do occur (corresponding to the long tail in the histogram). (b) The reconstructed and original ellipsometric angle $\Psi$ with AoI 60°. (c-d) show a case with strong dispersion at one edge of the spectral range. (c) The inferred and original refractive index; inset is a histogram of error distribution, as in panel (a). (d) The reconstructed and original ellipsometric angle $\Psi$ with AoI = 60°.

3.2. Birefringent materials
To evaluate the performance of our DL method for birefringent materials (illustrated in Figure 2b), we analyze experimental data measured on a crystal of 2H-MoS$_2$.\textsuperscript{[17]} In Figure 5a we show the in-plane ordinary refractive index $n_o$ inferred from the experimental data using the piecewise method, compared to that calculated directly using the effective approximation (Eqn. 2.1), and to that calculated by density functional theory (DFT).\textsuperscript{[17]} The DL-inferred result and the effective approximation are in close correspondence. The DFT calculations capture key features in the spectrum, but suffer from known inaccuracies.

In Figure 5b we compare the in-plane refractive index $n_o$ inferred by our three methods: pointwise, piecewise, and spectral. The pointwise and piecewise methods produce similar results. The spectral method reproduces broad features but misses the sharp resonances. The usefulness and applicability of the spectral approach depend on the material under investigation and the analysis goals. For the case of isotropic, amorphous GST, it is fair to assume that the optical properties vary only weakly within the spectral range. In this case, spectral learning not only preserves KK-consistency, continuity, and smoothness, but also improves the accuracy at each individual wavelength. However, birefringent MoS$_2$ has a complicated optical spectrum including sharp absorption resonances. In this case, spectral learning smears out sharp features. The degree of smearing depends on the number of wavelength points chosen for analysis; this is an important input to the inference process, that can be chosen based on the problem at hand. In brief, spectral learning with relatively sparse wavelength point density works well for smooth features, whereas piecewise and pointwise learning are more suitable for material with complicated spectra.

For the case of birefringent MoS$_2$, DL inference is accurate for the in-plane refractive index, but not so for out-of-plane. In Table 2 we show the MAE and $\sigma$ for the in-plane ($n_o$, $k_o$) and out-of-plane ($n_e$, $k_e$) refractive index inferred by our three methods: The in-plane inference error is acceptably low, but the out-of-plane error is so large as to make the inference useless.

| Method    | $\text{MAE of } n_o, k_o$ | $\text{MAE of } n_e, k_e$ | $\sigma$ of $n_o, k_o$ | $\sigma$ of $n_e, k_e$ |
|-----------|---------------------------|---------------------------|-------------------------|-------------------------|
| Pointwise | 0.124                     | 2.467                     | 0.394                   | 2.862                   |
| Piecewise | 0.153                     | 2.142                     | 0.292                   | 2.635                   |
| Spectral  | 0.172                     | 0.715                     | 0.360                   | 1.100                   |

Table 2: The accuracy of the pointwise, piecewise, and spectral learning methods for birefringent materials, as in Figure 2b. The MAE and $\sigma$ values are the averages for the real and imaginary indices.
To understand the poor accuracy of inferred \((n_e, k_e)\), we show in Figure 5c-d further analysis of the results of the simplest method, pointwise inference, working with generated data. The inferred \(n_e\) are tightly distributed around 5, which is the average of generated \(n_e\). The results suggest that the model doesn’t learn enough information from \(n_e\) & \(k_e\), and treats them as random variables, whose expectation value is the optimal inference. Therefore, we conclude that the pointwise forward function is not numerically one-to-one. It is not possible to more accurately infer \(\tilde{n}_e\) without additional measurements or assumptions. The fundamental explanation for this lies in the physical details of refraction at the air-sample interface, as explained in our earlier work.\(^{[17]}\) Fortunately, the inference error of \(\tilde{n}_o\) is about constant for sufficiently large \(\tilde{n}_e\). The observation suggests that \(\tilde{n}_o\) can be reliably inferred even without accurate knowledge or inference of \(\tilde{n}_e\). However, this means that the NN in effect simply learns Eqn. 2.1, which can be calculated directly. Therefore, in this case, DL offers no advantage in accuracy or speed.
Figure 5: Inferring the ordinary and extraordinary refractive indices of a birefringent material, as in Fig 2b. (a-b) Analyzing experimental data measured on MoS\textsubscript{2}. (a) Real part of the ordinary refractive index determined by three methods: DFT-calculated, effective index calculated directly from the experimental data as in Eqn. 2.10, and inferred by piecewise DL. (b) $n_o$ inferred by the three DL methods: pointwise, piecewise, and spectral. (c-d) Analyzing the error in inferred $n_e$ and $n_o$ using the piecewise method and generated data. (c) Inferred $n_e$ vs. the generated $|n_e|$. (d) Error of inferred $n_o$ vs. the generated $|n_e|$.

3.3. High-throughput discovery of phase-change materials for photonics

To demonstrate the use of DL-based spectroscopic ellipsometry data analysis in high-throughput material discovery, we analyze data collected for 177 different compositions in the
GST system. The composition spread is prepared by co-sputtering, as described elsewhere\cite{4}. The resulting sample is a GST thin film on an oxidized Si wafer, with a composition gradient across the wafer. The composition is measured by wavelength dispersive spectroscopy in an electron microprobe system (JEOL JXA-8900)\cite{4}. The ellipsometry data is measured for material in both the amorphous and crystallized states. The difference in optical properties across the amorphous-crystalline transformation defines the phase-change figure of merit (FoM):

$$\text{FoM} = \frac{|n_{\text{crystal}} - n_{\text{amorphous}}|}{k_{\text{crystal}} + k_{\text{amorphous}}}. \quad \text{Eqn. 3.1}$$

This choice of FoM emphasizes the importance of a large change in the real refractive index ($n$) and low optical loss ($k$), appropriate for refractive applications such as optical phase modulators\cite{19,20}. In Figure 6a we show $\lambda$-dependent FoM data for several different compositions. The optimal materials composition is wavelength-dependent. In Figure 6b we represent the FoM at $\lambda = 800$ nm as a colormap on the ternary phase diagram. The highest FoM of 0.872 is achieved for composition Ge$_{0.30}$Sb$_{0.27}$Te$_{0.43}$. The entire data analysis takes 1.4 sec, which includes inferring the refractive index spectra of both amorphous and crystalline phases across 177 different compositions and 100 distinct wavelengths. In Figure 6c we show the distribution of inferred thickness of the GST films; the thickness varies smoothly with composition, as expected. In Figure 6d-e we show the experimental data $\Psi$ at an AoI of 70°, and the inferred refractive index ($n$, $k$) for the composition with the highest FoM at 800 nm.
Figure 6: High-throughput discovery of phase-change materials by co-sputtering, ellipsometry measurements, and DL-based analysis. (a) $\lambda$-dependent FoM for compositions $\text{Ge}_{0.09}\text{Sb}_{0.17}\text{Te}_{0.74}$, $\text{Ge}_{0.11}\text{Sb}_{0.30}\text{Te}_{0.50}$, $\text{Ge}_{0.30}\text{Sb}_{0.27}\text{Te}_{0.43}$, and $\text{Ge}_{0.33}\text{Sb}_{0.39}\text{Te}_{0.28}$. (b) FoM at $\lambda = 800$ nm as a function of composition. The peak FoM is found for $\text{Ge}_{0.30}\text{Sb}_{0.27}\text{Te}_{0.43}$. (c) GST film thickness as a function of composition (data measured in the amorphous phase). (d-e) Experimental data $\Psi$ at AoI of 70° (d) and inferred refractive
index (e) for the composition with highest FoM at 800 nm, in both the amorphous and crystalline phases.

4. Conclusion

We demonstrate the use of DL to analyze ellipsometry data and infer the optical properties of materials. The results demonstrate that DL could be a fast, accurate, and useful method in cases when data analysis speed is at a premium, as in high-throughput materials development, and when the models are complex and the conventional non-linear regression is difficult. The workflow demonstrated here includes optical model construction, training data generation, and DL model setup and training, which take between 1 – 4 hrs on an Intel i7-9700K CPU and an Nvidia GeForce RTX 2060 Super graphics card. Inference from experimental data is then very fast, on the order of milliseconds per data set, a thousand-fold speedup compared to nonlinear regression. This approach is compatible with newly-emerging approaches high-throughput and closed-loop autonomous experimentation and may help to accelerate the discovery and development of new materials for photonic integrated circuits. We demonstrate the utility of our approach by analyzing data measured on a composition spread of Ge-Sb-Te phase-change alloys, zeroing in on the composition with optimal phase-change FoM with only 1.4 s of analysis time.

Appendix A: Detailed neural network structure

In Figure A1, we show detailed neural network structure as block diagram. Each block represents a layer with type on top and output dimension on bottom.
Figure A1: Schema of neural networks used in this work for (a) isotropic film model, (b) spectral learning, (c) piecewise learning, and (c) pointwise learning for birefringent materials.

Acknowledgements

This work was supported by an Office of Naval Research MURI through grant #N00014-17-1-2661. We acknowledge helpful discussions with Wujie Wang and Rafael Gomez-Bombarelli (MIT). The simulations presented in this paper were performed using the Massachusetts Green High Performance Computing Center, with support from the US Department of Energy under grant #DE-FG02-91-ER54109.

References

[1] P. J. Spencer, *Calphad* **2008**, 32, 1.

[2] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, 1, 011002.
[3] J. Holdren, *Materials Genome Initiative Strategic Plan*, National Science And Technology Council, 2014.

[4] A. G. Kusne, H. Yu, C. Wu, H. Zhang, J. Hattrick-Simpers, B. DeCost, S. Sarker, C. Oses, C. Toher, S. Curtarolo, A. V. Davydov, R. Agarwal, L. A. Bendersky, M. Li, A. Mehta, I. Takeuchi, *Nat. Commun.* 2020, 11, 5966.

[5] A. Nazarov, M. Ney, I. Abdulhalim, *Opt. Express* 2020, 28, 9288.

[6] M. Fried, P. Masa, *J. Appl. Phys.* 1994, 75, 2194.

[7] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N. T. P. Hartono, S. Ramasamy, B. L. DeCost, S. I. P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, *Npj Comput. Mater.* 2019, 5, 60.

[8] S. Chugh, S. Ghosh, A. Gulistan, B. M. A. Rahman, *J. Light. Technol.* 2019, 37, 6080.

[9] J. W. Weber, T. A. R. Hansen, M. C. M. van de Sanden, R. Engeln, *J. Appl. Phys.* 2009, 106, 123503.

[10] M. Gilliot, *Appl. Opt.* 2017, 56, 1173.

[11] M. Gilliot, *Appl. Opt.* 2019, 58, 800.

[12] M. Gilliot, *Appl. Opt.* 2019, 58, 8153.

[13] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.

[14] Y. Dong, C. Wu, C. Zhang, Y. Liu, J. Cheng, J. Lin, *Npj Comput. Mater.* 2019, 5, 26.

[15] V. Nair, G. E. Hinton, in *ICML*, 2010, pp. 807–814.

[16] D. P. Kingma, J. Ba, *ArXiv14126980 Cs* 2017.

[17] A. Singh, Y. Li, B. Fodor, L. Makai, J. Zhou, H. Xu, A. Akey, J. Li, R. Jaramillo, *Appl. Phys. Lett.* 2019, 115, 161902.
[18] Y. Li, A. Singh, S. Krylyuk, A. Davydov, R. Jaramillo, in *Low-Dimens. Mater. Devices 2019*, International Society For Optics And Photonics, 2019, p. 110850T.

[19] C. Wu, H. Yu, H. Li, X. Zhang, I. Takeuchi, M. Li, *ACS Photonics* 2019, 6, 87.

[20] A. Singh, S. S. Jo, Y. Li, C. Wu, M. Li, R. Jaramillo, *ACS Photonics* 2020, DOI: 10.1021/acsphotonics.0c00915.

[21] H. Fujiwara, *Spectroscopic Ellipsometry: Principles and Applications*, John Wiley & Sons, Chichester, England; Hoboken, NJ, 2007.
Spectroscopic ellipsometry is a powerful and data-rich metrology to characterize materials, devices, and manufacturing processes. However, traditional data analysis methods tend to be too slow for rapid feedback. We develop deep learning methods for quick & accurate analysis. We demonstrate the efficacy of our methods using data from high-throughput synthesis of phase-change materials for integrated photonics applications.