PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Evaluating regression-based techniques for modelling fabrication variations in silicon photonic waveguides

James, Aneek, Wang, Alexander, Wang, Songli, Bergman, Keren

Aneek E. James, Alexander Wang, Songli Wang, Keren Bergman, "Evaluating regression-based techniques for modelling fabrication variations in silicon photonic waveguides," Proc. SPIE 11843, Applications of Machine Learning 2021, 1184305 (1 August 2021); doi: 10.1117/12.2594255

SPIE.

Event: SPIE Optical Engineering + Applications, 2021, San Diego, California, United States

Evaluating regression-based techniques for modelling fabrication variations in silicon photonic waveguides

Aneek E. James*[a], Alexander Wang[a], Songli Wang[a], Keren Bergman [a]
[a]Lightwave Research Laboratory, Department of Electrical Engineering, Columbia University, New York, NY 10027 USA

ABSTRACT

For integrated silicon photonics to mature as an industry platform, robust methods for measuring and extracting the geometry of fabricated waveguides are needed. Due to the cost and time needed for SEM or AFM imaging, a method of extracting waveguide variability though optical measurements is often preferred. Here, we present a study of regression-based machine learning (ML) techniques that enable such variability extraction while maintaining compatibility with wafer-scale optical measurements. We first explicitly investigate the issue of non-unique effective and group index pairs that can affect the accuracy of regression-based techniques. Training data is then generated by simulating several geometries of wire waveguides in Lumerical's MODE solver to simulate defects due to process variances. Finally, a representative set of ML regression techniques are tested for their ability to accurately estimate the geometries of said simulated waveguides. To the best of the authors' knowledge, this work represents the first attempt in the literature to i.) explicitly study the effects of non-uniqueness in optical measurement-based metrology and ii) present a model that potentially overcomes said non-uniqueness. This work represents an important step towards the maturing models for process variations in silicon photonic platforms.

Keywords: Silicon photonics, process variations, metrology, machine learning, linear regression

1. INTRODUCTION

The growth of the field in silicon photonics is an indication of the massive potential of the field to revolutionize several technologies needed for the future. Data-intensive applications such as machine learning and artificial intelligence have pushed data bandwidths to their limits, demanding advanced optical interconnect technologies such as dense wavelength division multiplexing (DWDM) systems. Other fields such as microwave photonics have looked to the technology to potentially provide a platform for easily reconfigurable analog front ends to RF and microwave transmitters and receivers. The growth of these big application areas has placed an increased demand on the reliability and robustness of silicon photonics as a technology, particularly with regards to devices that hold the key to these advancements, such as microring resonators (MRRs) and maximally flat interleaver filters. To control costs and enable rapid development and prototyping, measurement of the width and heights of fabricated waveguides would aid in PDK development. Scanning Electron Microscopy (SEM) and atomic force microscopy (AFM) measurements are extremely accurate options, but unsuitable for wafer-scale deployment. A more scalable solution would be to use optical measurements of some measurement devices to infer the geometry of the fabricated waveguides.

Several efforts have been made toward this end. Liu et al. used a MRR to get this information, and decomposed the variation seen in the devices across the wafer[1]. Hammond et al. worked on this by using machine learning techniques to infer the width and thickness variations of several measured photonic devices[2]. To do this, they simulated many devices in a photonic device simulator and fed the output of that as training data into the machine learning algorithm. They used a deep learning model and generated over 100,000 simulations to extract this data. Xing et al. went for a simpler route, using linear regression to fit the variations of effective and group indices measured by Mach-Zehnder Interferometers (MZIs)[3].

In this paper, we demonstrate an introductory comparison of different machine learning models for extracting process variability in commercial SOI foundries. We first explore the design space for different silicon photonic waveguides using the effective index method to address the potential issue of the non-uniqueness of group and effective indices.

*aej2149@columbia.edu

Applications of Machine Learning 2021, edited by Michael E. Zelinski, Tarek M. Taha
Jonathan Howe, Proc. of SPIE Vol. 11843, 1184305 · © 2021 SPIE
CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2594255

Next, we review the most common machine learning models that can be employed to solve this problem, as well as our criteria for comparison. We then use our training data to feed the different models and give a hierarchy of appropriate solutions for different devices.

2. EXPLORING WAVEGUIDE PARAMETER NON-UNIQUENESS

A key requirement for robust optical measurement metrology is the addressing the potential for non-unique solutions. Specifically, addressing whether two different geometries of waveguides can produce the same effective index. To explore this potential issue, the effective index of the silicon wire waveguide is explored analytically using the effective index method (EIM). The operation of the effective index is described in full in several previous works[4,5]. Though there are some errors introduced in this way, one can see that it can behaviorally track the effective index value of a waveguide for different geometries.

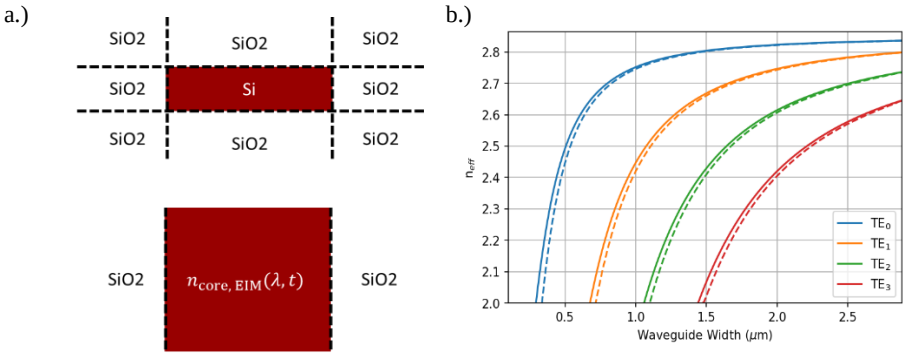

Fig 1. a) Graphical explanation of the EIM method. The 3D waveguide structure is reduced into a 2D slab structure to allow for simple analysis of effective and group indices. b) comparison of accuracy of the EIM method (solid) vs simulations in Lumerical MODE (dashed) for the TE0 – TE3 modes.

Using this analytical framework, we can explore the design space for both a silicon channel waveguide. Fig 2. shows a 3D plot of the width and height as functions of effective and group indices for a silicon wire waveguide. The waveguide width and height are varied based on ranges that are of interest to silicon photonic designs and 3σ values that are typical of worst-case sweeps across different foundry processes, respectively. There are roughly two forms of non-uniqueness that can be encountered. The first is a truly non-unique region, denoted by regions where a given surface bends in on itself. This can only be avoided in one of two ways. The first is by using more sophisticated models for the surface than simple polynomial regression. The second is by designing waveguides such that you avoid the possibility of ambiguity in the first place. While this is a fine solution, it would be preferable to develop a model that could work for a wide range of waveguide designs. Addressing this is the focus of this paper.

The authors would also like to note a second source of ambiguity where—though every effective and group index pair yields a unique geometry—several combinations of waveguides can yield similar results. This is a consequence of larger mode confinement of wider waveguides; as the waveguide geometries increase, the sensitivity of the effective and group indices on the geometry decreases. In this scenario, the uniqueness of any given solution depends primarily on the precision to which the extracted group and effective indices can be trusted and is beyond the scope of this work.

3. LINEAR REGRESSION TECHNIQUES FOR MODELING PROCESS VARIATONS

To overcome this non-uniqueness, regression-based ML techniques are evaluated for their potential efficacy. This section includes the review of the different ML techniques that were applied during the study. For each technique, we provide a brief review of the machine learning method, the motivations, and drawbacks behind the consideration of the method, and a summary of any notable mathematical properties. Finally, we discuss the model we fitted our data to

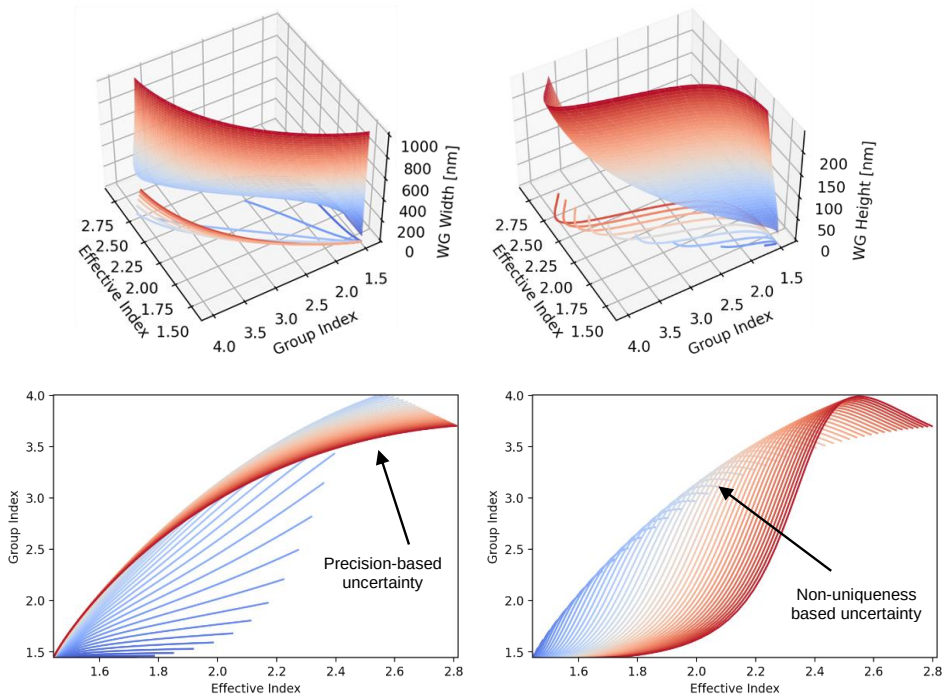

Fig 2. 3D Surface and contour plots of waveguide width and height as a function of effective and group indices. Looking at the plots, the main roadblocks in modeling waveguide width and height arise from uncertainty in the group and effective index measurements and non-unique regions of the surfaces. The non-unique regions are denoted by where the surfaces fold in on themselves, while regions that are sensitive to measurement accuracy can be identified by a bunching up of the contour lines. This second source of ambiguity is more of an issue for determining waveguide width than it is for height.

3.1 Linear Regression

Also known as ordinary least-squares (OLS) regression and linear least squares, this is the most straightforward machine learning technique and will also serve as our baseline during the study. Specifically, given several observations, (x_i, y_i), $i = 1,..,n$, we assume the model: $y = x_i^T \beta + \epsilon_i$. OLS will yield a β that minimizes the sum of squared residuals $\Sigma_i = 1^n (y_i - x_i^T \beta)^2$ where β is the desired vector of regression coefficients and ϵ is an error vector such that $E(\epsilon_i | x_i) = 0$. However, it is commonly accepted that a major drawback of this technique is that it is highly susceptible to outliers and multicollinearity[6]. Multicollinearity is when the independent variables within our model are correlated with one another. Given the relationship between group and effective indices, we suspect that our regression terms are highly susceptible to multicollinearity. Furthermore, OLS is prone to overfitting, which is when the model begins to predict the random error within the data rather than relationships between the variables. For these reasons, we wish to explore more sophisticated techniques for linear regression problems[7].

3.2 Ridge Linear Regression

As described in the previous section, OLS runs into issues when severe multicollinearity and overfitting are present. The reason for this is that OLS provides a zero-bias estimate, but often this introduces such large variance that the estimate ends up unusable anyways. Alternatively, ridge regression introduces a slight amount of bias into the estimate to reduce much of the variance. This provides more useful coefficient estimates in such cases of severe multicollinearity. The standard model for multivariate linear regression is $Y = \beta X + \epsilon$. The OLS estimator for β is $\hat{\beta} = (\mathrm{X'X})^{-1}\,\mathrm{X^T y}$. The ridge regression estimator is instead $\hat{\beta} = \left(X^T X + kI_p\right)^{-1} X^T y$ [8,9].

.

3.3 Lasso Regression

Lasso (least absolute shrinkage and selection operator) selects variables in a way to increase prediction accuracy by identifying a simpler model. This type of regression is useful for problems with high multicollinearity, and, as the resulting model is sparse, leads to easier interpretation in many cases. It is like ridge regression in purpose and methodology except it includes predictor variable selection. It does this by performing L1 regularization, which adds a penalty equal to the magnitude of the coefficients, thus favoring simpler models[10].

3.4 Polynomial Regression Model

An important decision in our study was our choice of model. The presence of non-unique regions in the design space implies that a simple polynomial model may not sufficiently describe the surfaces plotted in Figure 2. We now describe our alternative model in full. Consider all possible monomials based on the algebraic composition of the set of prediction terms $\{n_{\text{eff}}, n_{\text{g}}, \lambda, y\}$, where n_{eff} is the effective index of the waveguide, n_{g} is the group index, λ is the wavelength the two measurements are taken at, and y is the geometric attribute of the channel waveguide we wish to predict (width or thickness). Let (i, j, k) represent the powers that $n_{\text{eff}}, n_{\text{g}}, \lambda$ are respectively raised to for a particular monomial. We ensure a closed-form solution by limiting the power of y term in the regression to be ≤ 1 (i.e., y^0 or y^1 for each monomial). In this way, the resulting model for waveguide width and height is given as the N-th order Padé approximant.

$$y(n_{\text{eff}}, n_{\text{g}}, \lambda) = \sum_{\{i+j+k \leq N\}}^{i,j,k} \left[a_{ijk} \cdot n_{\text{eff}}^{i} \cdot n_{\text{g}}^{j} \cdot \lambda^{k}\right] \Big/ \left(1 + \sum_{\{i+j+k \leq N\}}^{i,j,k} \left[b_{ijk} \cdot n_{\text{eff}}^{i} \cdot n_{\text{g}}^{j} \cdot \lambda^{k}\right]\right), \quad (1)$$

where a_{ijk} denotes the regression terms where y was raised to a power of 0 and b_{ijk} denotes the regression terms where y was raised to a power of 1. Though originally expected to be of lower importance than our choice of regression technique, fitting the data to this model had important consequences on our study that will be described in Section 5.

4. PROCESS VARAITION SIMULATION

To get the training data for the regression algorithms, we used the MODE solver to calculate the group and effective indices for a range of waveguide geometries. An example of the simulation environment and results are shown in Figure 4. The waveguide width was swept from 400-1000 nm, while the height was swept from 200-240 nm, with the group and effective indices evaluated over a wavelength span from 1450 – 1650 nm. In total, 1050 unique data points were generated for use in the model discovery process.

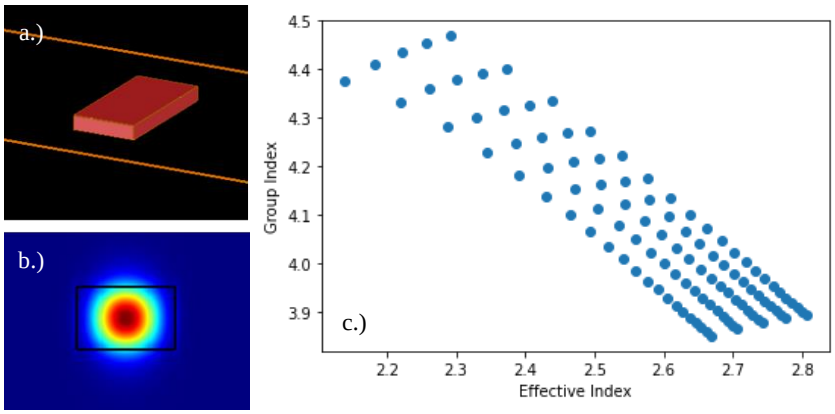

Fig 3. Example set of simulated data used in the regression fit. a) 3D CAD model of waveguide. b.) Example Mode profile. c.) Group of group and effective index pairs for several different geometries. Evaluated at λ=1550nm.

5. REGRESSION TECHNIQUE EVALUATION AND DISCUSSION

For comparing the three different methods, we used a randomized training test split of the retrieved Lumerical MODE simulation data. The ratio of train to test samples was 100:1, where the points included in each set were randomly chosen. A run of each regression method was done 1000 times, and the result for each figure of merit was averaged over the 1000 trials.

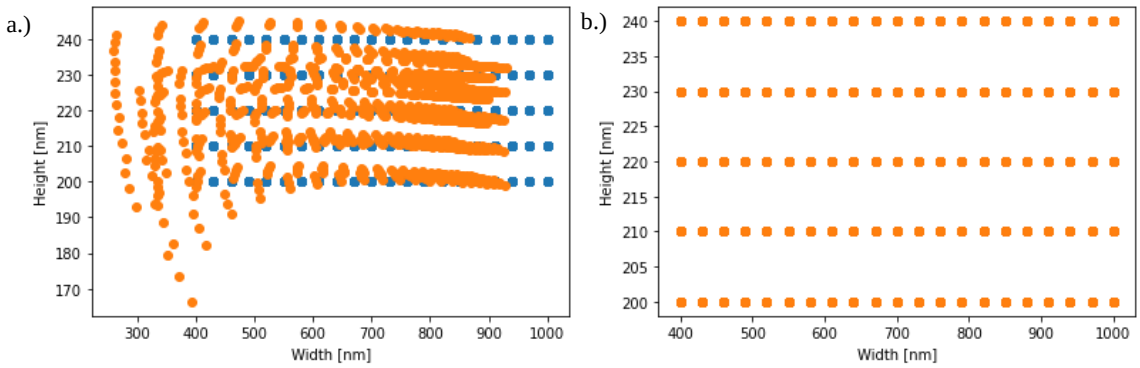

Fig 4. Comparison of the accuracy of the model predictions (orange) and the actual waveguide geometry (blue) for a a.) first order model and b.) a second order full linear regression model. More overlap between the orange and blue points represents a more accurate model.

An unexpected result of this regression is shown in Figure 4. The performance of OLS regression – the most straightforward technique – was surprisingly accurate, achieving a root mean squared error (RMSE) of 0.53 nm and 0.04 nm for width and height respectively with a second order model. For all intents and purposes, this is an accurate enough extraction of width and height for most requirements of PDK development. This is despite the expected complications of including wavelength as a prediction variable and the lower model order in comparison to some other reported methods[3]. Interestingly, use of the more advanced machine learning algorithms increased the error of the model, contrary to prior expectations. The results are described in detail in Table 1. The authors suspect that with hyperparameter tuning this relative increase in error for the advanced methods could be mitigated. Given the already accurate OLS results, however, the usefulness of any increased accuracy would be minimal in comparison to the required effort.

Table 1. Comparison of model accuracy between ordinary least squares (OLS), Ridge, and LASSO regression for a second-order Padé approximant.

Figure of Merit	OLS	Ridge	LASSO
RSME (width, height)	(0.535 nm, 0.036 nm)	(2.115 nm, 0.148 nm)	(4.714 nm, 1.575 nm)
R^2 (width, height)	(0.99999, 0.99999)	(0.99982, 0.99985)	(0.99907, 0.98271)

These results imply that model that the data is fitted to is far more important than the regression technique used. We confirm this by also performing the regression accuracy test for a normal polynomial model (i.e. the model described in Xing et al.[3]). The resulting drop in model accuracy is shown in Figure 5. Without the terms that enable the Padé approximant form, the model far less accurate, likely due to its unsuitability to handle the non-unique regions described in Section 2.

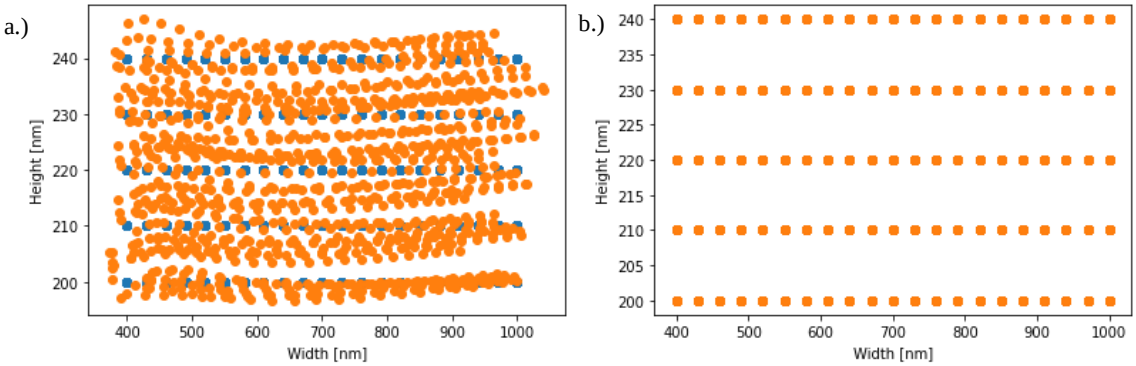

Fig 5. Comparison of the accuracy of the model predictions (orange) and the actual waveguide geometry (blue) for a a.) third-order polynomial model and b.) a third-order Padé approximant model.

6. CONCLUSION

In conclusion, we have investigated several methods of data collection for silicon photonic waveguides' effective and group indices using linear regression algorithms. The validity of such an approach is explored by mapping the design space of wire waveguides using the EIM to explore the regions of non-uniqueness. The accuracy of representative sample of linear regression techniques were compared to data simulated in Lumerical MODE. This comparison showed that a sub-nanometer accurate model of waveguide width and height as a function of wavelength, effective, and group indices can be achieved without complex machine learning routines. This is primarily achieved by the authors' novel use of a Padé approximant-based model instead of a normal polynomial model. It is presently unclear why the Padé approximant is more mathematically capable of modeling width and height variations, and further investigation should be performed to understand these results. Nonetheless the key conclusion is that optical measurement metrology is primarily dependent on the accuracy of group and effective index measurements. This study represents a significant step towards cheaply connecting waveguide geometry with optical measurements and, by extension, accurately representing process variations in photonic integrated circuit design to aid in statistical characterization of silicon photonic devices.

REFERENCES

[1] Lu, Z., Jhoja, J., Klein, J., Wang, X., Liu, A., Fluekiger, J., Pond, and J., Chrostowski, L., "Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability," Opt. Express 25(9), 9712-9733 (2017).

[2] Hammond, A., Potokar, E., and Camacho, R. M., "Accelerating silicon photonic parameter extraction using artificial neural networks," OSA Continuum 2(6), 1964-1973 (2019).

[3] Xing, Y., Dong, J., Dwivedi, S., Khan, U., and Bogaerts, W., "Accurate extraction of fabricated geometry using optical measurement," Photon. Res. 6(11), 1008-1020 (2018).

[4] Knox, R. M., and Toulios, P. P., "Integrated circuits for the millimeter through optical frequency range," Proc. MRI Symp. Submillimeter Waves, 497-516 (1970).

[5] Ming-Liu, J., [Photonic Devices], Cambridge University Press, 107-110 (2005).

[6] Farrar, D. E., and Glauber, R. R., "Multicollinearity in regression analysis: the problem revisited," The Review of Economic and Statistics 49(1), 92-107 (1967).

[7] Yu, C. and Yao, W., "Robust linear regression: A review and comparison," Communications in Statistics—Simulation and Computation, 46(8), 6261-6282 (2017).

[8] Hoerl, A. E., and Kennard, R. W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, 12(1), 55-67 (1970).

[9] McDonald, G. C., "Ridge Regression," Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 93-100 (2009).

[10] Hesterberg, T., Choi, N. H., Meier, L., and Fraley, C., "Least angle and ℓ_1 penalized regression: A review," Statist. Surv., 2, 61-93 (2008).