An Open-Source Artificial Neural Network Model for Polarization-Insensitive Silicon-on-Insulator Subwavelength Grating Couplers

Dusan Gostimirovic, *Member, IEEE*, and Winnie N. Ye, *Senior Member, IEEE*

(Invited Paper)

***Abstract*—We present an open-source deep artificial neural network (ANN) model for the accelerated design of polarization-insensitive subwavelength grating (SWG) couplers on the silicon-on-insulator platform. Our model can optimize SWG-based grating couplers for a single fundamental-order polarization, or both, by splitting them counter-directionally at the grating level. Alternating, SWG sections are adopted to reduce the reflections (loss) of standard, single-etch devices—further accelerating the design time by eliminating the need to process a second etch. The model of this device is trained by a dense uniform dataset of finite-difference time-domain (FDTD) optical simulations. Our approach requires the FDTD simulations to be made up front, where the resulting ANN model is made openly available for the rapid, software-free design of future standard photonic devices, which may require slightly different design parameters (e.g., fiber angle, center wavelength, and polarization) for their specific application. By transforming the nonlinear input–output relationship of the device into a matrix of learned weights, a set of simple linear algebraic and nonlinear activation calculations can be made to predict the device outputs 1830 times faster than numerical simulations, within 93.2% accuracy of the simulations.**

***Index Terms*—Silicon photonics, subwavelength devices, polarization insensitivity, grating couplers, machine learning, artificial neural networks.**

I. Introduction

THE large refractive index contrast of the silicon-on-insulator (SOI) platform (1.45 to 3.48) allows for ultracompact light confinement in densely integrated photonic circuits; however, it also causes lossy coupling of light into and out of the circuits, as there is a large modal mismatch between the 9 μm optical fiber and the standard single-mode SOI waveguide. A grating coupler offers an efficient solution to this issue by capturing the entire optical mode on its surface and diffracting it into the direction of the waveguide. These devices can be placed anywhere on the photonic circuit and can be quickly aligned to for the rapid prototyping of multiple devices on a single chip. Grating couplers suffer, however, from being highly dependent on the wavelength, polarization, and incident angle of the coupled light. As feature requirements change between different applications—or even between different parts of the same chip—new designs must be carried out to accommodate for the specific changes. This uses considerable amounts of computational resources. Broadband designs that reduce wavelength dependence have been reported [1], [2]; however, their peak performance generally falls short of devices designed to maximize transmission for a specific wavelength.

Designing photonic circuits requires numerical electromagnetic solvers that can take minutes to hours to characterize a single device. Tens or hundreds of simulations are often needed to optimize the device for a specific figure of merit. Although this is unavoidable for most novel device-level designs, the development of standard secondary components of the photonic circuit further delays the time to fabrication and testing. Like the electronics industry, libraries of compact models have become more available for system-level simulations [3] but are not flexible to the specific needs of different applications (e.g., working at a different polarization or center wavelength). We address this problem by developing flexible compact models of essential secondary components for the acceleration of a wider range of future photonic designs [4]. In this work, we use machine learning to model and design 1D polarization-insensitive subwavelength grating (SWG) couplers. SWG sections reduce backreflections in single-etch couplers [5], [6], with recent designs achieving sub-decibel coupling efficiencies [7], [8]. The 1D grating style can also efficiently couple both fundamental-order, orthogonal polarizations by splitting them in opposite directions at the grating level [9].

Artificial neural networks (ANNs) are universal approximators that can generalize the input–output relationship of any complex nonlinear function [10]. Although ANNs are most widely used in voice and image recognition applications, they can be used to accurately model RF and microwave circuits [11]–[13]. Their use in photonics is limited, but recent works have presented models that accurately approximate light scattering spectra [14], [15]. Given a large dataset of inputs and out-

Manuscript received August 1, 2018; revised October 30, 2018; accepted December 3, 2018. Date of publication December 6, 2018; date of current version January 10, 2019. This work was supported by the Natural Sciences and Engineering Research Council of Canada. *(Corresponding author: Winnie N. Ye.)*

The authors are with the Department of Electronics, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: DusanGostimirovic@cmail.carleton.ca; winnie_ye@carleton.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTQE.2018.2885486

puts (e.g., the properties and performance metrics of a photonic device), we use the ANN—a matrix of weighted values—to learn the input–output relationship of the polarization-insensitive SOI SWG coupler. With the trained ANN model, a set of simple linear algebraic and nonlinear activation calculations are made to predict the output of the device 1,830 times faster than the numerical simulations, without the need for empirical approximations. The ANN model uses a significant amount of computational resources up front in order to accelerate and automate future designs. Other frameworks have been developed for the automated design of photonic grating couplers [16]; however, every design requires the full, resource-heavy numerical process. We demonstrate the time-saving potential of the machine learning process by running a brute-force, parametric sweep solver on the grating coupler model, which finds an optimal design 61 times faster than a commercial, numerically solved particle swarm optimizer (PSO) [17] often used in the design of photonic devices [2], [18]. The trained model and the design framework are both made openly available on our GitHub repository [19].

II. Modelling

Designing photonic devices with our machine learning framework is split into three main sections: data acquisition from numerical simulations, the construction and training of a deep ANN model, and device optimization using inference of the trained model. This process does not change for different device types; thus, we have generalized and automated the process for the accelerated development of future flexible, compact, and open models of standard photonic components. We use the MATLAB numerical computing environment to handle this process, with Lumerical FDTD as a back-end finite-difference time-domain numerical solver.

A. Data Acquisition

The data needed to train and validate the model can be acquired experimentally, directly from real devices; however, training the model to a high accuracy requires hundreds of time- and resource-consuming measurements. Machine-learning-based design is thus more suitable when data are acquired numerically, as the measurements can be taken quickly and automatically, with a lower chance of external factors (e.g., measurement and fabrication error) being learned by the model. Given a reliable source of numerically solved data, the final accuracy of the trained model can be taken as the main source of device validation.

A simulation file is first created (manually) for the photonic device. Ideally, every parameter of the device is fully variable and learnable—to train a robust model useful for any design case—but compute resources are still limited. It is important to choose the parameters that are likely to cause a meaningful change in the device performance and to define a range of acceptable values for them. These may be continuous values, such as the height of a waveguide, or discrete, like specific materials or polarization states. The variable parameters are known as the *features* of the ANN model. Likewise, the outputs of interest

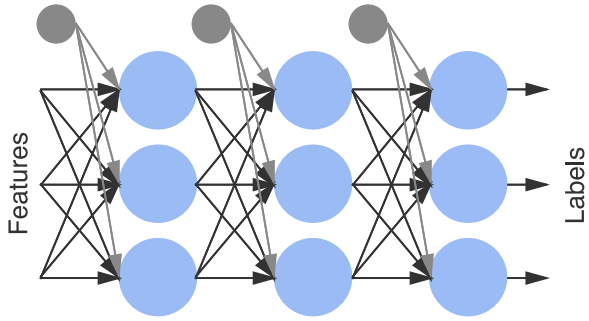

Fig. 1. An example structure of a 3:3:3:3 fully interconnected multilayer perceptron artificial neural network.

are extracted into the *labels* of the model. Labels can be a single value, like the center wavelength, or vectorized into multiple labels, like for the transmission spectrum at a device port.

Our data acquisition algorithm generates a set of randomized feature values within the ranges specified, numerically solves for the specified labels, and stores them together in a data object as an *example*. This is repeated for a user-specified number of examples. A more suitable approach is also available, which uniformly sweeps through the range of features. Here, a specified resolution of the uniform sweep determines the number of examples that are acquired. A uniform order, however, introduces an external source of bias into the ANN model, negatively affecting its ability to generalize the photonic device. After the data are acquired, we automatically randomize the order to eliminate the bias. We also automatically split the example set into two separate groups: training and validation, with an 85:15 ratio. We train the model with the training set and evaluate the performance of the model with the validation set. Each feature and label are also normalized between zero and one, as the parameters of a photonic device may have largely different orders of magnitude—otherwise negatively affecting the training rate and accuracy.

Before doing any training, it is important to manually inspect the data to make sure the seeding simulations produced reasonable results, as the training algorithm does not have any knowledge of the underlying physics: only the relationship between the inputs and outputs. Because the data are acquired numerically, we save time here by not having to inspect each example. If the simulation was set up correctly, we can expect every example to be meaningful, even if the results are undesired. Unlike conventional, optimization-based design approaches, poor results are not discarded; rather, they are used to train the model to find optimized designs across a broad range of applications.

B. Model Training

The ANN structure (model) we use is the multilayer perceptron, as shown in Fig. 1. The structure consists of at least three layers of fully interconnected neurons. A deep model simply adds more hidden (middle) layers, which generally produces more accurate results for highly nonlinear functions [20]. The connections between the neurons each hold a value (weight)

that ultimately represents a working part of the modelled photonic device. We randomly initialize the weights with a mean of zero and a standard deviation of one. At each hidden layer, we have the choice of using sigmoidal, hyperbolic tangent, and rectified linear (ReLu) activation functions. In general, the ReLu function produces higher training accuracies. Because we are predicting a set of continuous values at the output, we omit the activation function at the output layer. We evaluate the performance of the model by calculating the squared error cost function. The error is minimized by full-batch gradient descent, which adjusts the weights of the model after every example is evaluated by the cost function. We find that stochastic and mini-batch gradient descent do not offer better performance for our current models. The rate of gradient descent is decided by a set of hyperparameters (learning rate and momentum) that can cause major performance variations; it takes several iterations of trial and error to find the best starting values. Towards the end of training, the learning rate may be lowered to descend the model towards the minimum of the gradient function with less noise—ultimately finding a lower minimum.

C. Device Design With Model Inference

We can run inference on a well-trained ANN model to accurately predict the output of a device for any set of features within the range of simulated data. Inference is the process of forward propagation through the network to get a prediction, but with no backward propagation and weight adjustment to follow. Given the weights and activation functions used in the network, via our open-source model, inference can be run regardless of the simulation and training software used to develop the model.

To design for a specific device, we have the choice to use any optimization algorithm and consider the model as a black box that takes feature inputs and produces predictions. An evolutionary algorithm, such as particle swarm optimization, can be used; however, given the ultrafast prediction speed of an ANN model, a brute-force parametric sweep is more than capable of producing a rapid optimization. Our open-sourced device development framework uses a simple parametric sweep based on user-defined conditions and constraints. A condition is what the framework solves for (e.g., minimize loss at $\lambda_0 = 1.55 \pm 0.01$ μm), and a constant is a feature that is not included in the sweep (e.g., keep etch depth at 0.22 μm). Given a sweep resolution, the solver finds the best fits by running inference on each set of features in the sweep.

Given the difficulty of achieving 100% training accuracy, it may be necessary to cross reference the solved device in the original numerical solver and perform a quick, narrow optimization to find a truer optimum. If results vary significantly, it may be necessary to retrain the network or collect more training data.

III. Results

The generalized ANN modelling process is used to create a compact, flexible model of a polarization-insensitive SOI subwavelength grating coupler, as shown in Fig. 2. Due to the wide usage of this type of device for coupling light into and out of photonic circuits, each application may require slightly different performance metrics (e.g., center wavelength or polarization state). We use the trained model to show the accelerated design of the device, for an arbitrary set of metrics, in comparison with a conventional, numerically solved approach.

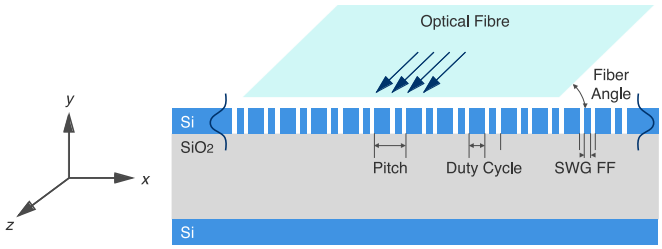

Fig. 2. A 2D cross-section of a straight SOI subwavelength grating coupler. The labeled dimensions are the features (variable parameters) defined in our ANN model.

A. Training the Grating Coupler Model

A 2D Lumerical FDTD simulation file is created for a straight (1D) grating coupler with alternating SWGs. Only a cross-sectional, 2D (x, y) FDTD simulation is needed to accurately characterize this device, as it extends uniformly in the z direction past the extent of the incident optical mode. For a focusing-style grating coupler [21], which reduces the required footprint, a longer-duration, 3D FDTD simulation would be needed. For simplicity, we demonstrate the modelling of the straight grating, noting that the process would remain identical for the focused grating. We define the features (variable parameters) of the model to be the fiber angle, polarization, grating pitch, the duty cycle of the large grating, and the fill factor (FF) of the subwavelength grating. The range of variable values are 5° to 20°, 0 (TE_0) and 1 (TM_0), 0.5 to 1.5 μm, 0.4 to 0.8, and 0.2 to 0.6, respectively. We create a uniform distribution of the variable parameters, producing 9,190 examples, and we simulate for each set of features—each taking 23.8 seconds to complete on an 18-core, 3.4 GHz, Intel Core i9-7980XE CPU. For each simulation, we extract the 200-point transmission spectrum at each end of the grating coupler for wavelengths between 1.3 and 1.7 μm. The spectra are then automatically discretized into their maximum transmission values, with the corresponding λ_0 values. Along with the automated data engineering steps the framework takes, we remove the examples that have maxima on the edges of the spectrum, as the true maximum may lie outside of it. We are then left with 4,831 training examples and 852 validation examples.

Given the sizes of the feature and label vectors, we automatically start with a 5-neuron input layer and a 4-neuron output layer—both without activation functions. Through trial and error, we achieve the highest model accuracy from a structure with three ultrawide (much larger than the input/output layers) hidden layers of 100, 50, and 50 neurons. Each neuron is equipped with a ReLu activation function. The learning rate was originally set to $\alpha =$ 1E-4 and manually decreased when the error curve would flatten or start to experience noise. The final learning rate was $\alpha =$ 2E-6. At each iteration, a mean percentage error was

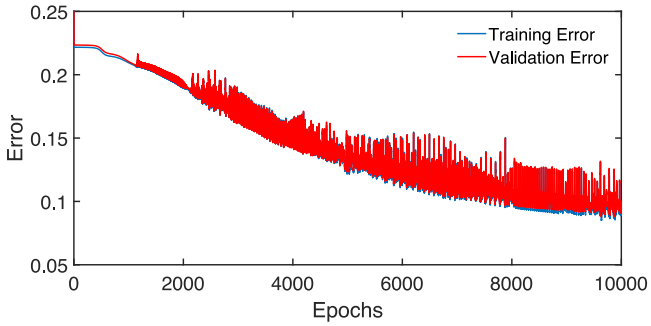

Fig. 3. Training and validation error for the first 10,000 iterations of the ANN training process. The error at the first iteration is 0.31.

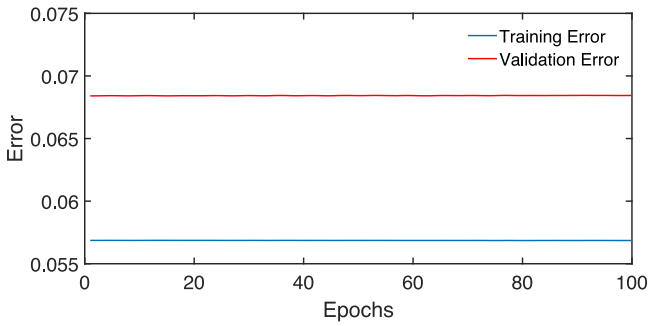

Fig. 4. Training and validation error for the last 100 iterations of the ANN training process, showing the stable minimum of the trained function.

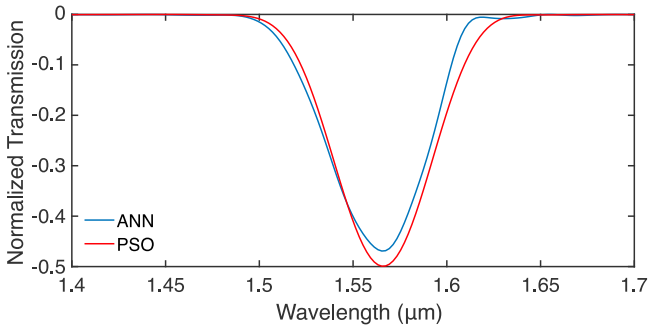

Fig. 5. Normalized transmission spectra for the ANN- and PSO-designed single-polarization SOI grating couplers.

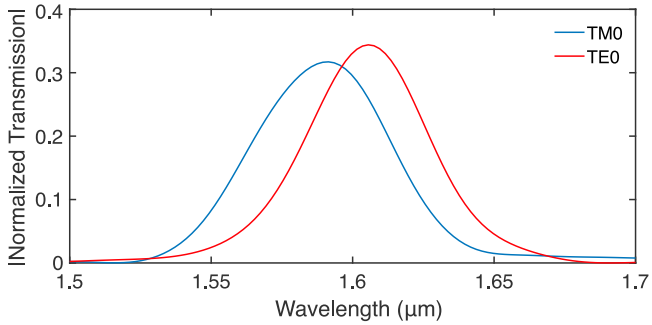

Fig. 6. Absolute transmission spectra of the forward (TM_0) and backward (TE_0) directions of the polarization-insensitive grating coupler.

calculated based on the training data and the validation data. The error plot for these two datasets is shown in Fig. 3 for the first 10,000 iterations (epochs), which covered the steepest section of the gradient descent. The noise is a result of the model bouncing around local minima as it descends down the error gradient. We stopped training when the validation error reached a minimum value: 6.8%. Although further training will continue to reduce training error (past 5.7%), validation error will start to increase, implying an overfit function. Because we are generalizing the function for any unseen set of features, basing the error on already seen data would be misleading. Fig. 4 shows the last 100 iterations of training, where the error values have stabilized. Note that with more data, and further tuning of the structure of the ANN, the validation error can be further reduced.

Inference of the trained model takes 0.013 seconds to complete, which is 1,830 times faster than the FDTD simulation. This is a relatively low speedup factor, as this particular simulation runs quickly. For many potential models (e.g., a ring resonator that simulates in 25 minutes [22]), we can expect to see speedup factors above 100,000.

B. Designing the Grating Coupler

To demonstrate and test the design process using the open-source ANN model and parametric-sweep solver, we create an SOI subwavelength grating coupler optimized at $\lambda_0 = 1.55$ μm for the fundamental TE mode. Using a sweep step resolution of nine, in 8.1 seconds, the solver finds the features of the device as: fiber angle $= 5°$, pitch $= 0.875$ μm, duty cycle $= 0.65$, and SWG fill factor $= 0.6$. The labels were found as: $T_{TE0,max} = -0.5$ and $\lambda_0 = 1.55$ μm. We run this device back into the FDTD solver to cross-check the results. Fig. 5 shows the FDTD transmission spectrum ($T_{TE0,max} = -0.47$ at $\lambda_0 = 1.56$ μm), which is within the range of expected error. To compare, we create the same device with the particle swarm optimizer in Lumerical FDTD. We use a generation size of five and stop the process when the maximum transmission begins to converge (five generations). In 492 seconds, the optimizer finds the device as: fiber angle $= 8.5°$, pitch $= 0.886$ μm, duty cycle $= 0.8$, and SWG fill factor $= 0.6$ to achieve: $T_{TE0,max} = -0.5$ at $\lambda_0 = 1.56$ μm. Despite using a less sophisticated optimization method, the ANN model arrives at a similar result and accelerates the design process by a factor of 61.

Because the ANN model was trained for both fundamental-order polarizations, at both ends of the grating coupler, we can design for a polarization-insensitive device as well. We sweep to solve for maximum TM_0 transmission in the forward direction of the grating, and TE_0 in the backward direction. In 8.4 seconds, the solver finds the features of the device as: fiber angle $= 14.375°$, pitch $= 0.75$ μm, duty cycle $= 0.65$, and SWG fill factor $= 0.25$. The FDTD results were found as: $T_{TE0,max} = 0.34$ and $T_{TM0,max} = -0.32$ at $\lambda_{0,TE0} = 1.61$ μm and $\lambda_{0,TM0} = 1.591$ μm. Note that the opposite signs in transmission represent the counter-directional propagation of the two modes. The wavelength of polarization insensitivity is $\lambda_{0,TEM} = 1.596$ μm, where the matched transmission is $|T_{TEM,max}| = 0.31$ (-5.1 dB). This is within the range of previously reported polarization-insensitive SOI grating couplers (-3 dB, -6.5 dB) [9], [23]. Fig. 6 shows the FDTD transmission spectra of the device.

IV. Conclusion

We present an openly available compact, generalized model of a polarization-insensitive SOI subwavelength grating coupler, featuring accelerated design of silicon photonic integrated circuits. The model is a deep artificial neural network trained by a dense set of finite-difference time-domain optical data in our machine learning framework. The model predicts an output for the device 1,830 times faster than the corresponding numerical simulation, at 93.2% accuracy of the simulation. We believe that the open-sourcing of standard photonic neural network models is beneficial for the rapid development of new photonic devices, given the speed and flexibility of the model, as well as the abstraction of the underlying physics and simulation into a simple, universally solvable package. Using the same automated modelling process, we plan to grow our open-source model library on GitHub with more standard components.

References

[1] D. Taillaert, P. Bienstman, and R. Baets, "Compact efficient broadband grating coupler for silicon-on-insulator waveguides," *Opt. Lett.*, vol. 29, no. 23, pp. 2749–2751, 2004.

[2] Y. Wang *et al.*, "Design of broadband sub-wavelength grating couplers with low back reflection," *Opt. Lett.*, vol. 40, no. 20, pp. 4647–4650, 2015.

[3] Lumerical Inc., "Photonic integrated circuit simulation," 2018. [Online]. Available: https://www.lumerical.com/products/interconnect/, Accessed on: Jan. 3, 2019.

[4] D. Gostimirovic and W. Ye, "Automating photonic design with machine learning," *in Proc. IEEE 15th Int. Conf. Group IV Photon.*, Cancun, Mexico, 2018.

[5] R. Halir *et al.*, "Waveguide grating coupler with subwavelength microstructures," *Opt. Lett.*, vol. 34, no. 9, pp. 1408–1410, 2009.

[6] P. Cheben *et al.*, "Subwavelength integrated photonics," *Nature*, vol. 560, pp. 565–572, 2018.

[7] R. Marchetti *et al.*, "High-efficiency grating-couplers: demonstration of a new design strategy," *Sci. Rep.*, vol. 7, 2017, Art. no. 16670.

[8] D. Benedikovic *et al.*, "Subwavelength index engineered surface grating coupler with sub-decibel efficiency for 220-nm silicon-on-insulator waveguides," *Opt. Express*, vol. 23, no. 17, pp. 22628–22635, 2015.

[9] Y. Tang, D. Dai, and S. He, "Proposal for a grating waveguide serving as both a polarization splitter and an efficient coupler for silicon-on-insulator nanophotonic circuits," *IEEE Photon. Technol. Lett.*, vol. 21, no. 4, pp. 242–244, Feb. 2009.

[10] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[11] Q.-J. Zhang, K. Gupta, and V. Devabhaktuni, "Artificial neural networks for RF and microwave design—From theory to practice," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 4, pp. 1339–1350, Apr. 2003.

[12] A. Hafid Zaabab, Q.-J. Zhang, and M. Nakhla, "Device and circuit-level modeling using neural networks with faster training based on network sparsity," *IEEE Trans. Microw. Theory Techn.*, vol. 45, no. 10, pp. 1696–1704, Oct. 1997.

[13] H. Kabir, Y. Wang, M. Yu, and Q.-J. Zhang, "Neural network inverse modeling and applications to microwave filter design," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 4, pp. 867–879, Apr. 2008.

[14] J. Peurifoy *et al.*, "Nanophotonic particle simulation and inverse design using artificial neural networks," *Sci. Adv.*, vol. 4, no. 6, 2018, Art. no. eaar4206.

[15] D. Liu *et al.*, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photon.*, vol. 5, no. 4, pp. 1365–1369, 2018.

[16] L. Su *et al.*, "Fully-automated optimization of grating couplers," *Opt. Express*, vol. 26, no. 4, pp. 4023–4034, 2018.

[17] Lumerical, Inc., "Particle swarm optimization." [Online]. Available: https://kb.lumerical.com/en/sweeps_optimization_particle_swarm.html, Accessed on: Jul. 18, 2018.

[18] J. Mutitu, "Thin film silicon solar cell design based on photonic crystal and diffractive grating structures," *Opt. Express*, vol. 16, no. 19, pp. 15238–15248, 2008.

[19] D. Gostimirovic, "photonmind-PIC," 2018. [Online] Available: https://github.com/Dusandinho/photonmind-PIC

[20] H. Geoffrey, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[21] F. Van Laere *et al.*, "Compact focusing grating couplers for silicon-on-insulator integrated circuits," *IEEE Photon. Technol. Lett.*, vol. 19, no. 23, pp. 1919–1921, Dec. 2007.

[22] Lumerical, Inc., "Lumerical's 2.5D FDTD propagation method." [Online]. Available: https://www.lumerical.com/support/whitepaper/2.5d_fdtd_propagation_method.html, Accessed on: Jul. 18, 2018.

[23] Z. Cheng and H.-K. Tsang, "Experimental demonstration of polarization-insensitive air-cladding grating couplers for silicon-on-insulator waveguides," *Opt. Lett.*, vol. 39, no. 7, pp. 2206–2209, 2014.

Dusan Gostimirovic (M'18) received the B.Eng. (with high distinction) degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2013. He is currently working toward the Ph.D. degree in electrical and computer engineering at Carleton University, with a thesis focused on silicon nanophotonics. His research interests include the design and fabrication of passive and active silicon-photonic devices for optical communication and computing. He is a member of OSA and SPIE.

Winnie N. Ye (SM'12) received the B.Eng. degree in electrical engineering from Carleton University, Ottawa, ON, Canada. She then studied Photonics and received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Toronto,Toronto, ON, Canada, and Carleton University, respectively. She is a Canada Research Chair (Tier II) in Nano-scale IC Design for Reliable Opto-Electronics and Sensors. She is currently an Associate Professor in the Department of Electronics, Carleton University. Her expertise is in silicon photonics and its applications in biophotonics, telecommunications, and renewable energy. She was the recipient of MRI's Early Researcher Award in 2012 and the Research Achievement Award from Carleton University in 2013. Recently, she won the 2018 IEEE Women in Engineering Inspiring Member Award, and the 2018 Engineering Medal for Research and Development from the Ontario Professional Engineers (PEO), as well as the PEO Ottawa Chapter's 2018 Engineering Excellence Award.