BraggNet: Complex Photonic Integrated Circuit Reconstruction Using Deep Learning

Jonathan Cauchon, Jean-Michel Vallée, Jonathan St-Yves and Wei Shi, *Member, IEEE*

Abstract—We propose the use of a deep learning model to reconstruct the physical design of complex coupled-cavity systems from their spectral response. The model is demonstrated using silicon photonic grating-assisted, contra-directional couplers consisting of thousands of Bragg periods. The contra-directional couplers are modeled as coupled optical cavities, for which a transfer matrix model is used to generate a synthetic dataset comprising a strategic design parameter space. A modular approach is adopted to build a deep learning model made of 6 sub-models trained to make high-accuracy prior and posterior predictions on specific physical quantities. The free-form, architecture-independent model allows the user to include any geometries to the design parameter space simply by including them in the training dataset. Upon proper training, the model achieves 1.4% mean absolute error on device reconstruction and hence proves suitable for inverse design applications. To show the potential of the approach, a second dataset is generated to emulate the fabrication conditions of a nominal design hindered by fabrication imperfections. The model is then trained on this dataset and used to reconstruct fabricated devices from experimental measurements, showing promise for on-chip fabrication diagnosis.

I. INTRODUCTION

Silicon photonics offers small-footprint and low-cost operation of telecommunication, bio-sensing and quantum computing tasks, to name a few. This technology has attracted massive research attention in recent years, leading to significant progress in integrated photonic devices and systems. While design of single components often relies on rigorous electromagnetic simulations, circuit-level design scales rapidly in complexity. In the realm of linear optics, simulation paradigms such as transfer matrix models based on scattering matrices of optical components have been proposed and used as a powerful frequency-domain simulation tool [1].

Allowing for device encapsulation and versatility, such models rely on linear algebra operations that are computationally efficient and can be accelerated by modern computing hardware like graphics processing units (GPUs). One of their fundamental limitations lies in the fact that linear operations performed in the frequency domain do not allow for direct inverse operations. In other words, the knowledge of a complete system's scattering matrix encapsulates the system in a black box that does not allow to retrieve the individual sub-components easily. Reconstructing a coupled photonic system from its frequency response can be very useful for inverse design and fabrication metrology purposes. In particular, the current state of the art for reconstruction of complex Bragg gratings is the layer peeling (LP) algorithm based on time-domain feature extraction [2]. While this model is based on analytical equations and hence physically accurate, it lacks practicality because it is computationally expensive. It is

sensitive to noise in the input data and becomes unstable when used with strongly-reflecting gratings. In addition, it does not allow for targeted solution sub-spaces. For those reasons, it is hardly usable in a data-driven experimental context.

Sophisticated optimization algorithms make use of parameter-space navigation in order to find local or global extrema [3]. While such algorithms are effective, they require to navigate the parameter space every time a design is queried. This often requires the model to be run many times, which can be computationally-intensive and time-consuming. The ML paradigm for inverse design, on the other hand, relies on a first step of model training and optimization using a set of provided training data. While the dataset creation and training processes can be just as time-consuming, they yield a trained model that can handle any query in a matter of microseconds. This is highly desirable for inverse design and fabrication diagnostics purposes. Artificial intelligence (AI) has proven to offer highly-effective photonics design tools [4]. This task has generally been tackled in one of three ways.

Firstly, inverse design has been demonstrated using generative adverserial networks (GANs). This unsupervised learning technique aims to train a model to generate artificial input data that displays statistical properties similar to the training dataset's. Notable examples of this approach include the synthetic generation of realistic human portraits [5]. In photonics, such an approach has been demonstrated to design high-dimensionality devices such as free-form metasurfaces [6], [7].

Secondly, parameter-based regression models offer forward, inverse or bidirectional design of photonic components. Parameter-based models are built on the assumption that the device's physical properties and/or its spectral response follow a certain shape, usually based on theoretical models. The presumed shape is usually a function fully defined by a handful of parameters. This creates a low-dimensionality input/output parameter space that can be learned by a ML model using multivariate regression. For instance, a forward DL-based model has been demonstrated to predict the Lorentzian oscillator-shaped optical response of a nanoslit array [8]. Inverse design demonstrations include core-shell nanoparticles [9], materials targeting certain properties [10], and topological photonics [11]. Bidirectional modeling has also been demonstrated by training separate models for forward and inverse design like integrated Bragg gratings [12]. Some demonstrations also include coupled bidirectional models that use an iterative, back-and-forth training process [13] [14].

The third approach lies in the use of deep learning for free-form parameter space design. A good example of this

is the design of integrated power splitters targeting specific splitting ratios [15]. A pixel grid of 2.6 X 2.6 μm^2 is defined as parameter space where each pixel is independent and can either contain silicon or not. While parameter-based methods assume a certain shape to which it seeks to fit parameters, a free-form approach allows for an infinite design parameter space. It then relies on the provided dataset's parameter space contents to learn possible output shapes. This approach is more versatile because the training dataset can be supplemented at any time with further design parameter shapes. In this sense, it constitutes an architecture-independent approach.

In this work, we propose BraggNet, a deep learning model, to reconstruct non-uniform contra-directional coupler (CDC) filters from their frequency response. CDC profiles are treated as pixels and synthetic datasets are used to limit the model to a desired parameter space. We demonstrate how a broad parameter space can allow versatile inverse design. To our knowledge, very little effort has been put into using machine learning for fabrication diagnosis applications. As opposed to the layer peeling algorithm, which uses extensive computational power at every query, a deep learning model uses most of the computing resources during the training process, prior to actual use. As a result, once such a model is trained and put into production, querying the model is almost instantaneous, which allows it to scale to chip- and wafer-scale analyses. We hence propose and demonstrate how training BraggNet on an experimental-emulating dataset allows to reconstruct devices from experimental measurements, paving the way for wafer-scale fabrication quality assessment.

II. PROBLEM DEFINITION

The proposed demonstration of complex system reconstruction uses contra-directional couplers, a Bragg filtering device that finds use in telecommunications [16] and quantum computing [17]. The basic operation principle of the CDC consists in designing an asymmetric waveguide coupler (i.e., a pair of optical waveguides with different propagation constants) whilst creating Bragg gratings on the waveguides to achieve effective coupling between two confined lateral modes propagating in opposite directions. It is essentially a four-port device where selected pass-band wavelengths are dropped while the remaining spectral content propagates through the input waveguide [18]. While uniform gratings are their simplest form, CDCs often incorporate non-uniform spatial profiles such as apodization and chirp in order to obtain specific performances [16]. In that case, the device is spatially segmented in sub-devices that exhibit their own spectral characteristics. A transfer matrix model (TMM) then can be used to cascade the coupled response of each individual segment along the grating to compute the overall complex electric field spectral response. A non-uniform CDC device is split into N_{seg} uniform grating segments. Fig. 1 shows the geometrical and optical characteristics of a single segment as well as the complete system.

The mode coupling coefficients κ_{mn} are given by overlap integrals dictated by coupled-mode theory [19],

$$\kappa_{mn} = \frac{\omega}{4} \iint \mathbf{E}_m^*(x,y)\Delta\varepsilon(x,y)\mathbf{E}_n(x,y) \; dydx, \quad (1)$$
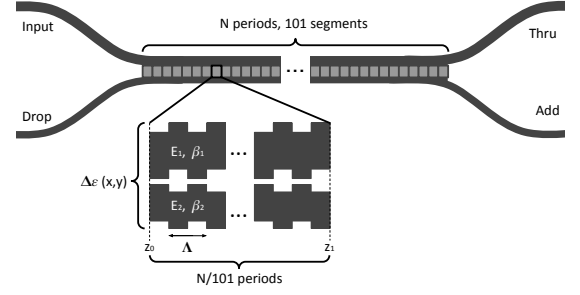

Figure 1: Schematic of a non-uniform CDC device and one of its segments.

where $\mathbf{E}_m$ and $\mathbf{E}_n$ are the complex electric fields of the optical modes guided in waveguides m and n, respectively. The phase-matching condition is $\beta_B = \pi/\Lambda$. Hence, the waveguides' supported modes are detuned from the phase matching condition by

$$\begin{aligned}\Delta\beta_1 &= \beta_1 - \pi/\Lambda - j\alpha \\ \Delta\beta_2 &= \beta_2 - \pi/\Lambda - j\alpha,\end{aligned} \quad (2)$$

where α is the propagation loss parameter. For each uniform segment, interface matrix S_1 and propagation matrix S_2 are obtained as follows:

$$S_1 = \begin{bmatrix} j\Delta\beta_1 & 0 & 0 & 0 \\ 0 & j\Delta\beta_2 & 0 & 0 \\ 0 & 0 & -j\Delta\beta_1 & 0 \\ 0 & 0 & 0 & -j\Delta\beta_2 \end{bmatrix},$$

$$S_2 =$$
$$\begin{bmatrix} -j\Delta\beta_1 & 0 \\ 0 & -j\Delta\beta_2 \\ j\kappa_{11}^*e^{-2j\Delta\beta_1 z_1} & j\kappa_{12}^*e^{-j(\Delta\beta_1+\Delta\beta_2)z_1} \\ j\kappa_{12}^*e^{-j(\Delta\beta_1+\Delta\beta_2)z_1} & j\kappa_{22}^*e^{-2j\Delta\beta_2 z_1} \end{bmatrix}$$
$$\left.\begin{matrix} -j\kappa_{11}e^{2j\Delta\beta_1 z_1} & -j\kappa_{12}e^{j(\Delta\beta_1+\Delta\beta_2)z_1} \\ -j\kappa_{12}e^{j(\Delta\beta_1+\Delta\beta_2)z_1} & -j\kappa_{22}e^{2j\Delta\beta_2 z_1} \\ j\Delta\beta_1 & 0 \\ 0 & j\Delta\beta_2 \end{matrix}\right].$$
$$(3)$$

The transfer matrix of the nth segment is then given by

$$M_n = e^{(S_{1,n}+S_{2,n})(z_{1,n}-z_{0,n})}, \quad (4)$$

where the matrix exponential is obtained by a Padé approximation of order 13 [20]. Provided that the CDC is separated in N_{seg} segments, the total transfer matrix of the system is obtained by backward multiplication of the individual segments' transfer matrices:

$$M = \prod_{n=1}^{N_{seg}} M_{(N_{seg}-n+1)}. \quad (5)$$

The drop and thru electric fields can then be obtained by basic operations on M [18]. Eq. 5 shows that the system's spectral response is given by N_{seg} cascaded 4 X 4 matrix multiplications, which is why it is uneasy to extract individual segment matrices from the total transfer matrix.

For this demonstration, a dataset was created by using devices with 101 grating segments ($N_{seg} = 101$). A given segment is completely defined by three physical quantities: its number of grating periods, its coupling power κ and its phase-matching condition λ_B. Therefore, a complete CDC device is completely defined by its total number of grating periods N, its coupling power profile along the propagation direction $\kappa(z)$ (101 values total), and its phase-matching condition profile along the propagation direction $\lambda_B(z)$ (101 values total). With the proposed TMM model, the knowledge of these 203 numerical values allows to fully define a given CDC and simulate its spectral response. For the proof of concept, pseudo-random physical parameters were employed to simulate CDCs to create a dataset. It was decided that N would vary between 300 and 3000 periods, κ would have a Gaussian shape comprising maximum coupling values between 10 and 50 mm^{-1} and Gaussian indices between 0 and 10, hence creating a variety of apodization profile shapes. The effective indices were set to fixed values 2.4 and 2.6, while the grating pitch was set to be between 305 and 315 nm, resulting in phase-matching conditions λ_B between 1525 and 1575 nm. All $\lambda_B(z)$ profiles where set to be linear and centered at 1550 nm, with pseudo-random slopes. Technically speaking, the two pseudo-randomly generated grating profiles used floats, which provides a virtually infinite number of design combinations (limited by computer float precision).

A total of 18 262 such pseudo-random devices were generated and simulated, creating a synthetic dataset. The spectral response was simulated for 1001 wavelength values between 1500 and 1600 nm. For the DL reconstruction task, the 203 values defining the physical grating are the targets, and the drop amplitude and group delay spectra are the inputs. The dataset hence exhibits size reduction $2002 \rightarrow 203$. Since the data was randomly generated, a given sample is not necessarily a good design. It is just a random one generated from the parameter space defined above. In order to emulate experimental conditions, noise was added to the dataset spectra. Noisy transformations on the dataset were applied to add a Gaussian noise to the spectra. Standard deviations of 5 db and 1 ps were applied on the drop amplitude and group delay spectra, respectively. Those figures were obtained from a commonly-used optical spectrum analyzer's specification sheet [21]. A noise floor of -50 dB was added to the data to increase realism. Fig. 2 shows a few random samples of the dataset.

III. DEEP LEARNING MODEL AND RESULTS

Inverse modeling non-uniform CDC devices consists in finding a model f with parameters θ such that

$$N, \ \kappa(z), \ \lambda_B(z) = f\left(\begin{pmatrix} |r(\lambda)|^2 \\ \tau_g(\lambda) \end{pmatrix}, \ \theta \right), \tag{6}$$

which constitutes a high-dimensionality multivariate regression problem. As photonics designers, we can use prior knowledge of the system to choose a good model architecture and provide knowledge transfer. The CDC task suffers from its high dimensionality because the outputs can show inter-dependence. For instance, two different combinations of design parameters can yield very similar spectral responses.
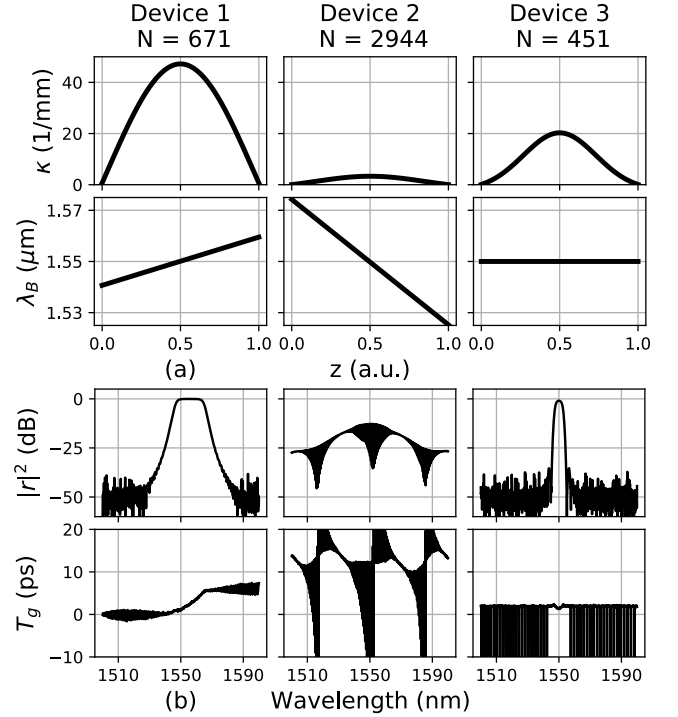

Figure 2: Three samples from the created dataset. (a) Pseudo-randomly-generated grating physical parameters including the number of grating periods (N), the Gaussian apodization profile along the grating ($\lambda_B(z)$), and the linear coupling power profile along the grating ($\kappa(z)$). (b) The simulated spectral responses of the CDC devices with added noise, including the drop response ($|r|^2$) and the drop group delay (T_g).
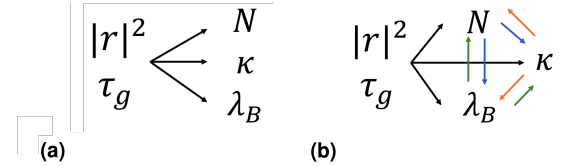

Figure 3: (a) A naive approach to the problem, in which all outputs are decoupled. (b) A more complete approach, in which outputs are considered interdependent. The direction of the arrows indicates the flow of relevant information that can be used to reach an accurate solution.

For this reason, the prior knowledge of certain targets can go a long way in predicting other targets of the system. For instance, filter bandwidth depends on both the number of grating periods and the coupling power of the device. Prior knowledge of one of those values can thus be useful in inferring the other one. A naive approach, as displayed in Fig. 3a, would consist in decoupling the three output variables and training a unidirectional model to predict the design from the CDC's spectra alone. However, in order to increase the coherence between the outputs and acknowledge their coupled nature, a more complete model is proposed, displayed in Fig. 3b. In this model, each output is dependent not only upon the model's inputs, but also upon the other outputs. Prior knowledge is useful in predicting the inverse design.

As highlighted by Bayes' Theorem and Bayesian network theory [22], estimators widely benefit from the combined use of priors and posteriors because it restricts the parameter space dimensionality.

In order to obtain a complete dynamic model as displayed in Fig. 3b, as well as to provide encapsulation and scalability, the problem is separated into 6 sub-problems to which 6 sub-models are sought. For each output, one naive (prior) model and one complete (posterior, depending on other outputs) model are defined. These 6 sub-models are defined by functions f_1 through f_6, given in Eqs. 7. Upon obtaining such functions, the 6 models can interact with each other dynamically in order to create a complete neural network (NN) model.

$$
\begin{array}{ll}
\textit{Prior} & \textit{Posterior} \\[4pt]
\tilde{N}_{naive} = f_1\left(\frac{|r|^2}{\tau_g}\right), & \tilde{N}_{complete} = f_2\left(\frac{|r|^2}{\tau_g}\,\middle|\,\kappa, \lambda_B\right) \\[12pt]
\tilde{\kappa}_{naive} = f_3\left(\frac{|r|^2}{\tau_g}\right), & \tilde{\kappa}_{complete} = f_4\left(\frac{|r|^2}{\tau_g}\,\middle|\,N, \lambda_B\right) \\[12pt]
\tilde{\lambda}_{B\,naive} = f_5\left(\frac{|r|^2}{\tau_g}\right), & \tilde{\lambda}_{B\,complete} = f_6\left(\frac{|r|^2}{\tau_g}\,\middle|\,N, \kappa\right)
\end{array} \tag{7}
$$

A. Inverse Design

The dataset was split in a 80-20 ratio between train and test data. The test data was set aside and never trained upon, in order to provide accurate performance assessment and avoid overfitting. Each individual sub-model is trained at performing its own simple task. This makes the learning process easier because the solution space is limited and models can converge relatively easily. This approach is commonly referred as transfer learning [23], since it uses the input as well as previously-acquired knowledge to generate a final output. Each sub-model consists in a 3-layer fully-connected NN (one hidden layer) and various layers sizes. During the training process, techniques such as the Adam optimizing algorithm [24] and batch size increase [25] were used to minimize the loss criterion on the test dataset. Poutyne, a convenient training framework for PyTorch, was used to conduct most training routines [26]. A custom loss criterion was defined as the mean absolute error (MAE), averaged on the three physical quantities of interest (N, $\kappa(z)$ and $\lambda_B(z)$). Since the three quantities have different orders of magnitude, this allows to optimize the regression model uniformly.

Table I: Summary of all trained models' number of parameters and performance on the test dataset.

Model	Num. Params (10^3)	Mean Abs. Error (%)
N - naive	20.3	1.57
N - complete	23.9	0.70
κ - naive	200.8	2.78
κ - complete	351.9	2.02
λ_B - naive	200.8	0.71
λ_B - complete	262.0	0.48
BraggNet	1069.8	1.40

Upon proper training of the 6 sub-models, they were cascaded as shown in Figure 4. Since all sub-models were pretrained to perform their specific task, further training of the
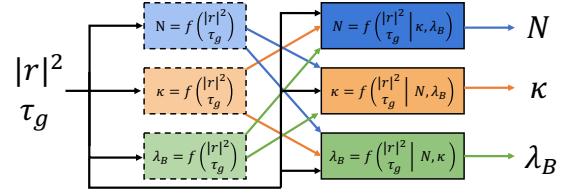

Figure 4: General architecture of the BraggNet model. 6 sub-models are trained to achieve either high prior or posterior performance. The drop amplitude and group delay spectra first enter the naive (prior) models in order to obtain zeroth-order estimations. Those estimations are then fed into the complete (posterior) models to obtain higher-accuracy estimations.

whole BraggNet allowed to reach the best performances. Like in a sport team, each player is good at their own position. However, team trainings are necessary in order to optimize the team's performance.

Table I contains a summary of all trained models in terms of number of parameters and performance on the test data, while Fig. 5 shows its performance of three sample predictions. All sub-models show mean absolute errors below 3% on their predictions. Once the 6 pre-trained models are put together to form BraggNet and trained further, the complete trained model shows a mean absolute error of 1.40% on its device reconstructions. It is worth noting the effect of the use of a cascaded model as opposed to a naive one. The naive approach would consist in using the three naive models for predictions, which would result in a mean absolute error of 1.69% on all physical quantities. By using a complete, posterior approach, the performances are significantly higher. It is also worth noting that the average of the complete sub-models' performances (1.07%) is significantly lower than BraggNet's performance. This is because the complete sub-models, trained using actual target values, are used in BraggNet with estimations provided by the naive models. Therefore, it is reasonable to consider that the accuracy of the predicted priors used have an impact on the accuracy of BraggNet.

B. Noise Analysis

To use BraggNet on experimental data for fabrication diagnosis, a proper noise analysis is needed to ensure the usability of the model and asses what performances to expect. Since it features average pooling layers at its inputs, the model is mostly insensitive to the Gaussian noise applied on the transmission and group delay spectra because the averaging kernel operations act as filters. However, relative the noise floor present (relative to the maximum reflection) in a given drop spectrum plays a major role in the device reconstruction task.

Side-lobe suppression ratio (SLSR) is mainly dependent on the apodization and chirp of the CDC device. Therefore, the noise floor level inevitably influences the accuracy of the possible reconstruction. In order to quantify this effect, a standard BraggNet training procedure was defined: 100 epochs at a learning rate of 0.01 followed by 50 epochs at a learning rate of 0.001. For different noise floor values, BraggNet was
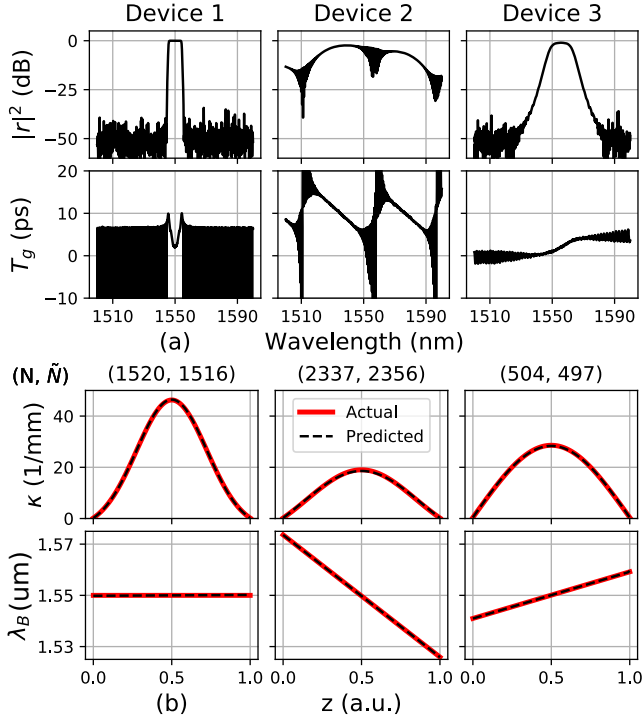
Figure 5: Random examples of BraggNet's reconstructions (predictions) of test data and comparison with the actual design values. N and $\tilde{N}$ are the design number of periods and reconstructed number of periods, respectively.
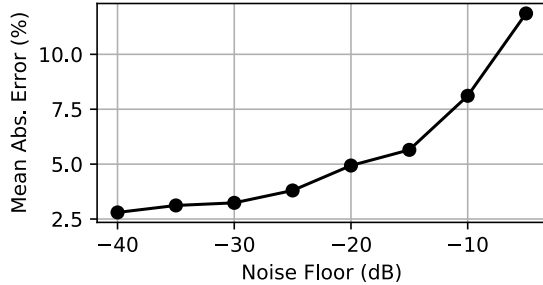

Figure 6: Effect of the measuring instrument's noise floor on the reconstruction performance of BraggNet. The dashed line shows the best-obtained performance without any noise applied to the dataset.

trained from scratch using the standard procedure to determine what reconstruction performance it could achieve with the given noise floor value. Figure 6 shows the mean absolute errors on the test data obtained during this experimentation. For high noise floor values, the error on reconstruction is very important and goes beyond 10% at the noise level of -8 dB. Therefore, the model is ineffective in presence of a high noise level due to the lack of information obtainable from the side lobes. The noise floor is referred to is relative, For experimental measurements of CDC devices, the relative noise floor is typically around -25 dB beneath the maximum reflection, which coincides with typical single-stage device SLSR values [27]. At this floor level, the training procedure yields a MAE of 3.8% on test data.

C. Fabrication Diagnosis

In order to demonstrate the potential of the deep learning model for practical applications, it was further used to reconstruct fabricated CDC devices from their measured responses. In the case of inverse design, a wide parameter space was covered in the dataset in order to enable flexible and versatile device design. For fabrication diagnosis, a much more narrowed-down dataset was created to emulate on-chip fabrication variation and non-uniformity. Namely, a nominal design was chosen and stochastic variations were applied to the design to emulate fabrication imperfection. The chosen nominal design is a broadband apodized CDC with no chirp as shown in [16].

The modifications to the nominal design were all made randomly, following a normal distribution centered at the nominal value and with a determined standard deviation. The maximum and minimum coupling powers κ_{max} and κ_{min} of the apodization profile were offset from the nominal values. For the chirp profile, five points were randomly generated along the grating, with variations from the nominal values following a normal distribution. From those five points, a third-degree polynomial fit was obtained and from this fit, the chirp profile of the sample was determined. By choosing five points semi-randomly separated from each other, we ensured that the dataset comprised different frequency contents in the chirp variation (i.e. some sample have very low-frequency chirp while some other ones have higher-frequency chirp). These chirp profiles hence emulate on-chip index variations along the grating [28]. The distorted profiles were fed into the transfer matrix model that yielded the corresponding filter responses. A noise floor value of -25 dB on the drop spectrum and a gaussian noise with an amplitude of 0.1 ps were applied to once again emulate typical experimental conditions observed with our instruments [21].

Additionally, other considerations were taken to make the dataset compliant with experimental data. Firstly, instead of defining a phase-matching conditions centered at 1550 nm on all samples, the phase matching condition is normalized in terms of relative detuning from λ_B. Since the phase matching condition is linear in terms of wavelength ($\lambda_B = 2\bar{n}_{eff}\Lambda$), this allows to apply the model on a measurement with any center wavelength. Secondly, instruments measuring group delays do not eliminate the effects of other components of the optical circuit such as routing waveguides and grating couplers. For this reason, there is an arbitrary offset between measured and simulated group delays. To normalize this situation, we set the group delay value to be zero at the phase-matching wavelength on all samples. Thirdly, experimental group delay is often very noisy outside the reflection band. Since there is negligible optical power reflected at these wavelengths, the value of the simulated/measured group delay has no physical meaning. By weighting the group delay with respect to drop-port optical power (in linear scale), we drastically suppress the noisy, meaningless group delay values. The weighted group delay is given by:

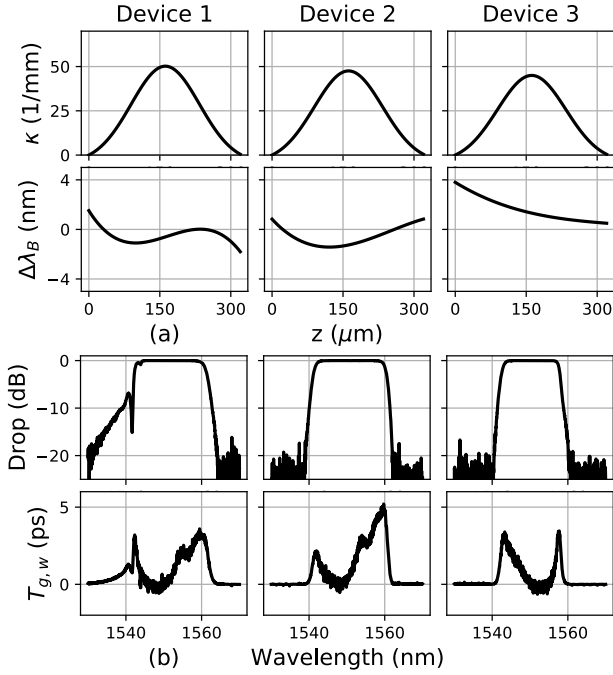$$\tau_{g,w} = \tau_g \ 10^{\frac{drop \ [dB]}{10}}. \tag{8}$$

Figure 7: Samples of the experimental-oriented dataset showing (a) the coupling and phase-matching condition profiles of the device along the propagation direction, and (b) the amplitude and group delay spectral responses of the fabrication-distorted devices.
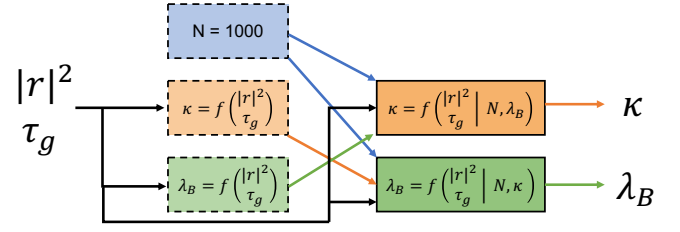

Figure 8: Adaptation of BraggNet for applications on experimental data. The number of grating period remains unhindered by fabrication, which makes it known from the design. N-models can hence be replaced by the known value of 1000.

An experimental-emulating dataset comprising 50 000 data samples was generated this way. Fig. 7 shows three samples of the dataset. It can be seen that the chirp profile has a significant impact on the quality of the filter's response, which can be distorted by wafer non-uniformity and fabrication errors.

BraggNet itself was slightly simplified to comply with experimental operation. While fabrication errors affect the yielded apodization and chirp profiles through smoothing effects and silicon height variations, the device's number of periods remains unhindered by fabrication and thus can be deemed exact. Consequently, there is no need for N-models to make predictions on N. Instead, a simple bypass is used on the model, as displayed in Fig. 8. The known value of N (1000) is fed directly to the posterior sub-models. This also contributes to increase accuracy because posterior models will use an exact N value as opposed to the usual estimated N value. BraggNet was re-trained on this new dataset, following the same training routine as before. Performances similar to those shown in Fig. 6 were reached (with an MAE of 3.2%).

To demonstrate the fabrication diagnosis capabilities of the experimental-oriented model, some on-chip non-uniformities were manually created along the fabricated device. To do so, the device was fabricated with a segmented micro-heater 2 μm above it, allowing to create virtually-arbitrary temperature profiles that change the local effective index and therefore the local phase-matching condition. This technique is used in fiber Bragg grating to generate chirp for delay compensation applications [29]. The device was then measured using different thermal chirps. The drop and group delay response were fed into BraggNet, as shown on Fig. 9. The grating profile predicted by BraggNet were then fed into the transfer matrix model for comparison with the experimental measurements.

Fig. 9 shows three measurement samples displaying different levels of chirp. A passive (not chirped), a lightly-chirp and a heavily-chirped device are each shown. In all three cases, it can be seen that the predicted coupling coefficient κ is smaller than the designed value. This is expected since lithography smoothing typically leads to bigger-than-designed inter-waveguide gaps and hence weaker coupling. Interestingly, the predicted κ_{max} on all three devices are similar despite the remarkable difference in bandwidth and filter shape. The model hence understood that this spectral widening comes from chirp instead of stronger coupling. Looking at the chirp reconstruction on Fig. 9b, it is evident that the model caught the chirping effect. As smoothed-out temperature profiles were applied along the device, BraggNet detected that the maximum phase mismatch increases with the bandwidth and shifts towards longer wavelength. Looking at the lefter-most chirp profile on Fig. 9, namely the passive-response chirp, it is seen that the reconstructed and predicted chirps are similar but not quite the same. This deviation is attributed to the on-chip fabrication non-uniformity, that typically cause effective index variations of $\Delta n \approx 1 \times 10^{-3}$ along the silicon waveguides [30].

Fig. 9c shows the output of the transfer matrix model, where the BraggNet-predicted grating profiles are fed and noise is applied. It can be seen that the reconstruction is very accurate at the drop and group delay level. That is, BraggNet picked up on the features underlying the chirp and apodization. Comparing the BraggNet-reconstructed responses to the experimental results, we can see that the overall behavior show an excellent agreement, while the small features (e.g., location of ripples) are slightly off. The weighted group delays also match well in the profile with mismatch in fine features.

This leads to understanding the limitations of the model. Since it is based on 101 grating segments, it treats each segment as a lumped entity. This bring granularity when used with experimental measurements. In our case, the devices have 1000 grating periods, which means that each grating segment comprises roughly 10 periods. Any fabrication variations occurring within the 10-period scale cannot be resolved by the model in training or in application. Therefore, the current model is more suitable for extraction of low-frequency features of CDC devices.
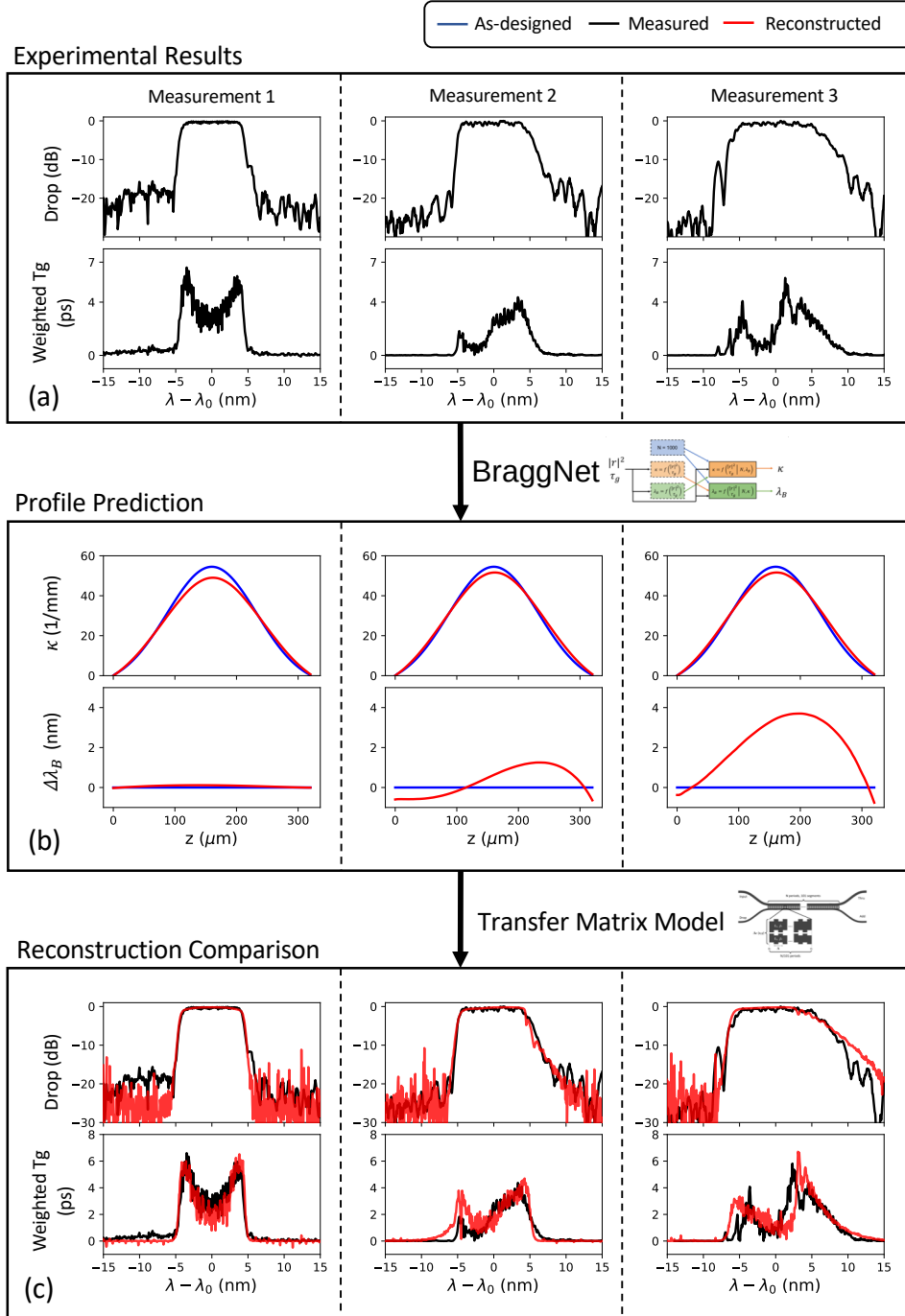
Figure 9: Demonstration of BraggNet applied on experimental data. (a) three conditioned measurement of a CDC device under different thermal chirp conditions, including the measured drop and weighted group delay spectra. (b) BraggNet prediction of the apodization and chirp profile of the experimental data. (c) Comparison between the original measurements and the output of the transfer matrix model fed with BraggNet's profile prediction for the three measurements, including the drop and weighted group delay spectra. The central wavelength was offset manually to match the experimental and reconstructed spectra.

IV. CONCLUSION

We have demonstrated that a free-form deep learning approach can be used for efficient reconstruction (inverse design) of non-uniform contra-directional couplers comprising a large number of grating segments, which constitutes a complex coupled-cavity optical system. Such a free-form approach is architecture-independent because the model can learn any design parameter space included in the training dataset. The approach is also modular because it comprises 6 sub-models trained specifically for particular tasks. These sub-models can be trained further together to form BraggNet. Our model has shown efficient and accurate inverse-design capabilities in the realm of Gaussianly-apodized, linearly-chirped contra-directional couplers, reaching a mean absolute error of 1.4% on validation data. A noise analysis was performed considering a standard training procedure. This showed that BraggNet is able to offer sub-5% mean absolute error on validation data so long as the amplitude noise floor was below 20 dB.

To show the potential and versatility of BraggNet, a second dataset emulating fabrication variations and comprising realistic noise figures was created. Upon proper re-training and simple architecture simplifications, BraggNet achieved a 3.2% mean absolute error on validation data. It was then used to reconstruct as-fabricated CDC devices with manually-induced thermal chirp profiles. Once fed back into the transfer-matrix model, these reconstructions proved to be accurate and capable of picking up on most low-frequency features such as bandwidth, band-edge angle and group delay profile. While the demonstrated reconstructions show great potential, their are still limited by the model's granularity, which prevents the extraction of higher-frequency variations along the grating.

Based on BraggNet's performances for inverse design and comprehensive fabrication diagnosis, we expect to achieve eventually robust data-driven fabrication assessment and diagnosis.

REFERENCES

[1] L. Chrostowski, H. Shoman, M. Hammood, H. Yun, J. Jhoja, E. Luan, S. Lin, A. Mistry, D. Witt, N. A. Jaeger, *et al.*, "Silicon photonic circuit design using rapid prototyping foundry process design kits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–26, 2019.

[2] R. Feced, M. N. Zervas, and M. A. Muriel, "An efficient inverse scattering algorithm for the design of nonuniform fiber bragg gratings," *IEEE Journal of Quantum Electronics*, vol. 35, no. 8, pp. 1105–1115, 1999.

[3] P.-I. Schneider, X. Garcia Santiago, V. Soltwisch, M. Hammerschmidt, S. Burger, and C. Rockstuhl, "Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction," *ACS Photonics*, vol. 6, no. 11, pp. 2726–2733, 2019.

[4] K. Yao, R. Unni, and Y. Zheng, "Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale," *Nanophotonics*, vol. 8, no. 3, pp. 339–366, 2019.

[5] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, 2019.

[6] J. Jiang, D. Sell, S. Hoyer, J. Hickey, J. Yang, and J. A. Fan, "Free-form diffractive metagrating design based on generative adversarial networks," *ACS nano*, vol. 13, no. 8, pp. 8872–8878, 2019.

[7] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, and W. Cai, "Generative model for the inverse design of metasurfaces," *Nano letters*, vol. 18, no. 10, pp. 6570–6576, 2018.

[8] A.-P. Blanchard-Dionne and O. J. Martin, "Teaching optics to a machine learning network," *Optics Letters*, vol. 45, no. 10, pp. 2922–2925, 2020.

[9] S. So, J. Mun, and J. Rho, "Simultaneous inverse design of materials and structures via deep learning: Demonstration of dipole resonance engineering using core–shell nanoparticles," *ACS applied materials & interfaces*, vol. 11, no. 27, pp. 24264–24268, 2019.

[10] A. Zunger, "Inverse design in search of materials with target functionalities," *Nature Reviews Chemistry*, vol. 2, no. 4, pp. 1–16, 2018.

[11] L. Pilozzi, F. A. Farrelly, G. Marcucci, and C. Conti, "Machine learning inverse problem for topological photonics," *Communications Physics*, vol. 1, no. 1, pp. 1–7, 2018.

[12] A. M. Hammond and R. M. Camacho, "Designing integrated photonic devices using artificial neural networks," *Optics express*, vol. 27, no. 21, pp. 29620–29638, 2019.

[13] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, "Plasmonic nanostructure design and characterization via deep learning," *Light: Science & Applications*, vol. 7, no. 1, pp. 1–8, 2018.

[14] W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," *ACS nano*, vol. 12, no. 6, pp. 6326–6334, 2018.

[15] M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin, and K. Parsons, "Deep neural network inverse design of integrated nanophotonic devices," *arXiv preprint arXiv:1809.03555*, 2018.

[16] J. St-Yves, H. Bahrami, P. Jean, S. LaRochelle, and W. Shi, "Widely bandwidth-tunable silicon filter with an unlimited free-spectral range," *Optics letters*, vol. 40, no. 23, pp. 5471–5474, 2015.

[17] A. Afifi, L. Chrostowski, M. Hammood, N. A. Jaeger, S. Shekhar, and J. F. Young, "Contra-directional couplers as pump rejection and recycling filters for on-chip photon-pair sources," in *2019 IEEE 16th International Conference on Group IV Photonics (GFP)*, vol. 1949, pp. 1–2, IEEE, 2019.

[18] W. Shi, X. Wang, C. Lin, H. Yun, Y. Liu, T. Baehr-Jones, M. Hochberg, N. A. Jaeger, and L. Chrostowski, "Silicon photonic grating-assisted, contra-directional couplers," *Optics express*, vol. 21, no. 3, pp. 3633–3650, 2013.

[19] A. Yariv and P. Yeh, *Photonics: optical electronics in modern communications*. Oxford Univ., 2006.

[20] A. H. Al-Mohy and N. J. Higham, "A new scaling and squaring algorithm for the matrix exponential," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 970–989, 2010.

[21] Luna Innovations, *OVA 5000 Optical Vector Analyzer*, 7 2020.

[22] F. V. Jensen *et al.*, *An introduction to Bayesian networks*, vol. 210. UCL press London, 1996.

[23] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264, IGI global, 2010.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," *arXiv preprint arXiv:1711.00489*, 2017.

[26] F. Paradis, "Poutyne: Keras-like framework for py-torch," 2018.

[27] M. Hammood, A. Mistry, M. Ma, H. Yun, L. Chrostowski, and N. A. Jaeger, "Compact, silicon-on-insulator, series-cascaded, contradirectional-coupling-based filters with¿ 50 db adjacent channel isolation," *Optics Letters*, vol. 44, no. 2, pp. 439–442, 2019.

[28] A. D. Simard, G. Beaudin, V. Aimez, Y. Painchaud, and S. LaRochelle, "Characterization and reduction of spectral distortions in silicon-on-insulator integrated bragg gratings," *Optics express*, vol. 21, no. 20, pp. 23145–23159, 2013.

[29] Y. Painchaud, C. Paquet, and M. Guy, "Optical tunable dispersion compensators based on thermally tuned fiber bragg gratings," *Optics and Photonics News*, vol. 18, no. 9, pp. 48–53, 2007.

[30] Z. Lu, J. Jhoja, J. Klein, X. Wang, A. Liu, J. Flueckiger, J. Pond, and L. Chrostowski, "Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability," *Optics express*, vol. 25, no. 9, pp. 9712–9733, 2017.