



Accurate and efficient prediction of photonic crystal waveguide bandstructures using neural networks

CASPAR F. SCHWAHN AND SEBASTIAN A. SCHULZ* 

SUPA, School of Physics and Astronomy, University of St Andrews, St Andrews, KY16 9SS, Fife, UK
**sas35@st-andrews.ac.uk*

Abstract: We demonstrate the use of neural networks to predict the optical properties of photonic crystal waveguides (PhCWs) with high accuracy and significantly faster computation times compared to traditional simulation methods. Using 100,000 PhCW designs and their simulated bandstructures, we trained a neural network to achieve a test set relative error of 0.103% in predicting gap guided bands. We use pre-training to improve neural network performance, and numerical differentiation to accurately predict group index curves. Our approach allows for rapid, application-specific tailoring of PhCWs with a runtime of sub-milliseconds per design, a significant improvement over conventional simulation techniques.

Published by Optica Publishing Group under the terms of the [Creative Commons Attribution 4.0 License](#). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

Slow light in photonic crystal waveguides (PhCW) has the potential to enhance optical delay and nonlinear optical interactions, and to reduce the size and power consumption of components such as optical switches and modulators [1,2]. In recent years, PhCWs have been used for applications including beamsteering for LiDAR [3,4], autocorrelators [5], enhancing the sensitivity of optical interferometers [6] and studying photonic transitions [7], enabled by the development of dispersion and loss engineered PhCWs. Here, structural parameters such as the radii or position of individual rows of holes are modified to control the waveguide performance, e.g. the group index and available slow light bandwidth [2,8–10] or the propagation loss [11,12].

Further uptake of slow light PhCWs is limited by the complexity of the device design. Compared to other integrated photonic devices, such as ring resonators or *y*-splitters, PhCWs can have more structural parameters (see Fig. 1) that together determine the device performance. Furthermore, there is no known analytical description linking the device performance to the device structure.

The standard approach to designing PhCWs then is to predict the device performance using numerical simulations. Here, the use of approximate methods, such as the Plane Wave expansion [13] or guided mode expansion methods [14], significantly reduces computation requirements compared to fully vectorial solvers such as finite difference time domain or finite elements methods without sacrificing accuracy.

Computational requirements can be reduced further through 2D approximations of the 3D structure, such as the effective index [15] or the more accurate effective period methods [16], at the cost of yielding approximate results which must be validated by 3D simulations before device fabrication. These advanced computation methods combine accuracy with reduced computation cost, e.g. dropping the time taken for a single simulation to the order of a minute.

Typical PhCW engineering methods shift the position (perpendicular or parallel to the waveguide) or radius of individual rows of holes and have often been restricted to only changing one of these parameters at a time (for typically two or three rows of holes). Wider design

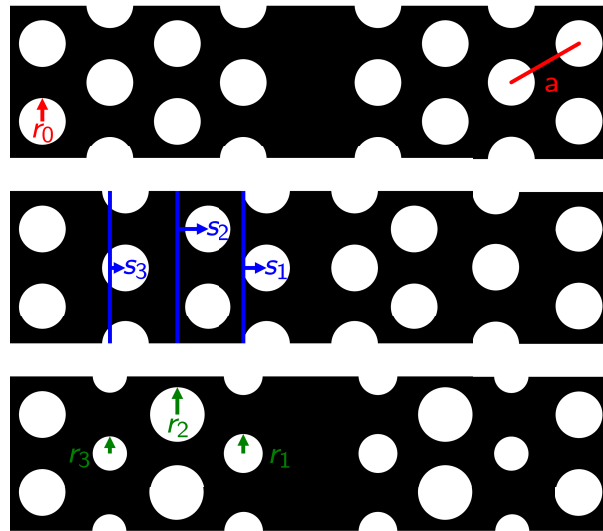


Fig. 1. (top) A W1 row defect PhCW with lattice spacing a and base hole radius r_0 . Common modifications include (middle) perpendicular displacements s_i of the position of rows [9] and (bottom) variations in the radii of holes r_i [8]. PhCWs have reflective symmetry across the y -axis hence modes have odd or even y -parity.

spaces have been opened up by replacing the prohibitively slow manual analysis and assessment of device performance with automated optimization methods, e.g. Refs. [17–19]. Even using these optimization algorithms, only minute subsets of the full parameter space can be explored. Classical optimization algorithms, restricted by the time taken to simulate each design parameter combination, run through 10,000 to 100,000 possible device realizations (using 2D approximations). If we assume a 7-dimensional parameter space (as we will outline later), with at least 100 possible values per parameter we achieve $100^7 = 1 \times 10^{14}$ combinations. To access this full design space a new simulation paradigm is needed, one that achieves the accuracy of 3D simulations, while simultaneously reducing the computation time by several orders of magnitude.

Deep Learning for the design of photonic devices in general rapidly gained popularity over the past few years [20–22]. Example applications include designing integrated power splitters [23] and computing the band structures of 2D photonic crystals [24–26]. In fact, neural networks have recently been applied to the forward problem of predicting the normalised delay bandwidth products of PhCWs [27] from various design parameters. Here, we focus on the more general task of predicting 6 full gap guided bands from a 7-dimensional design parameter space and achieve excellent accuracy - a mean test error of 0.103%. At the cost of a minimal reduction in accuracy compared to the plane-wave-expansion-based MIT Photonics Band solver (MPB), the NNs are 5 orders of magnitude faster than 2D simulations.

Furthermore, this task is ideally suited to approaches similar to transfer learning [28] which is a machine learning technique where knowledge of one problem is used to simplify another related problem. Transfer learning has previously been applied to photonic problems [28,29], e.g. to train a NN to predict the transmission spectra of 8-layer photonic film using a NN pre-trained on 10-layer films. Here we use a similar approach, first pre-training a NN on a large 2D MPB dataset before continuing training using a small 3D dataset, to reach a test set relative error of 0.296%. Compared to directly training a NN on the 3D dataset (relative error 1.22%), pre-training led to a 75% reduction in error. Thus we can achieve accurate prediction of the 3D-bandstructure (which represents the true device bandstructure, and not an approximation as produced by 2D

simulations), without the need for a large 3D-dataset which would be computationally expensive to produce.

At inference time, the NN is 7 orders of magnitude faster with a runtime of 8.4×10^{-5} seconds per design on a desktop computer compared to several hours for 3D MPB simulations.

2. Methods

2.1. Design space and band structures

Photonic crystal waveguides are created by introducing a line defect in a photonic crystal slab consisting of air holes in a silicon membrane (Fig. 1(top)). We parameterize the PhCW designs under consideration as a 7D vector

$$\mathbf{d} = (r_0, s_1, s_2, s_3, r_1, r_2, r_3) \quad (1)$$

where r_i and s_i refer to the radii (Fig. 1(bottom)) and perpendicular shifts (Fig. 1(middle)) of holes of the inner three rows, respectively, and r_0 denotes the radii of all other holes. This vector represents a commonly used design space, with several groups highlighting the importance of the first three rows for slow light PhCW design [10,11,30,31]. To ensure sufficient coverage of the design space given limited computational resources, we restrict the design space to 7 dimensions by freezing out parallel shifts as these often lead to designs with very high propagation losses [30]. We impose the following design constraints to account for fabrication limitations such as minimum feature sizes or the need for a connected membrane [9,11,32]:

- minimum hole radius of $r_i = 0.2 a$
- minimum hole separation of $0.1 a$

where a is the lattice constant, here fixed at 400 nm. The design space \mathbf{d} was uniformly sampled subject to the above constraints to avoid biasing the NN. (The distribution of design parameters is shown in [Supplement 1](#).)

For PhCW applications, we are typically interested in the guided modes beneath the light line (Fig. 2), as these are intrinsically lossless. We also need to know if these bands are single mode at our desired operation frequency (which is application dependent) and especially if there are leaky modes – those above the light line [33]. Hence, we chose as targets for our neural network several gap guided bands ω_{n,k_x} over the entire Brillouin zone, $0 \leq k_x a/2 \leq 0.5$.

2.2. Dataset generation

We simulated each design using the MIT Photonic Bands (MPB) eigensolver [13] to compute the lowest 30 frequencies ω at 101 evenly spaced values of k_x . (A copy of the MPB control file, together with other underlying data is available [34].) The 2D effective period approximation [16] was used to reduce the simulation time by approximately two orders of magnitude. Since the PhCWs have reflective symmetry, bands can be classed by even and odd y-parity where the y-direction is in the plane of the PhCW but perpendicular to the light propagation. When bands of the same parity approach each other they anti-cross, a key factor in controlling the slow light behaviour of the bands [10], while bands of opposite parity cross. The parity of each ω, k_x pair was obtained using an inbuilt MPB function so that each band structure could be split into even and odd bands. Bands of the same parity were sorted by frequency so that they can be uniquely labelled by $\omega_{n,k}$. We selected even bands $\in [10, 11, 12]$ and odd bands $\in [11, 12, 13]$ as these typically lie in the region of interest – in the band gap and below the light line $\omega = c k_x$ (see Fig. 2). In total, we simulate 100,000 examples which, using 2D approximations, took 5000 cpu core hours (~ 3 cpu core minutes/design) on a high-performance computing cluster. A second dataset was generated using 1000 3D simulations over 6500 cpu core hours (~ 6.5 cpu core hours/design).

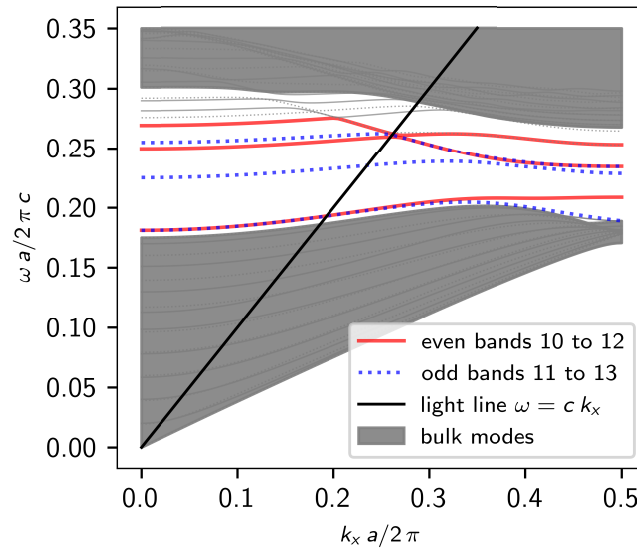


Fig. 2. Band diagram of a PhCW showing the bands under consideration, which are centred on the band-gap, labelled by symmetry.

The recommended min-max normalisation scheme [35] was used to rescale all features to the interval $[0, 1]$ using the known parameter ranges over which the designs were generated.

We split both the 2D and 3D datasets into a standard training, validation and test set in the ratio 70:15:15.

2.3. Neural networks

The NN's task is to learn the forward problem: predicting guided bands from the design parameters of a PhCW. A fully connected network is appropriate for the input vector since there is no spatial ordering of the design parameters which fully describe a PhC waveguide. The architecture of the best-performing model is a stack of 5 dense layers of 500 nodes followed by two fully connected branches of a further 5 layers with 500 nodes each: one branch for the even parity bands and one for the odd parity bands, as outlined in Fig. 3. We use the LeakyReLU activation function (with default $\alpha = 0.3$) because it is fast to evaluate and outperforms the common Rectified Linear Unit activation (ReLU) function [36]. Each branch is finally connected to 303 output nodes which represent the 3 flattened bands of 101 k_x points for each parity (see Fig. 3). The total number of trainable parameters is 3,814,606. We implemented the NN using the popular *Keras* machine learning library [37] with a *TensorFlow* backend [38].

We chose the mean squared error (MSE) for the loss function and the Nadam optimizer [39] with default hyperparameters except for the initial learning rate which was tuned to 1×10^{-4} . To minimize overfitting and avoid tuning the number of epochs, early stopping [40] terminates training when the validation loss does not improve over 300 epochs and restores the NN's weights to the lowest validation loss state reached during training. Additionally, we reduce the learning rate on plateau [35] meaning the learning rate is halved when the validation loss does not improve for 200 epochs. We find that L2 regularisation [41] of 1×10^{-9} is necessary to mitigate overfitting.

Training the final NN for 4000 epochs took 27 hours on a tesla v100 graphics card and the NN converged to a minimum validation loss of 9.43×10^{-7} and training loss 8.31×10^{-7} . The NN is available at [34].

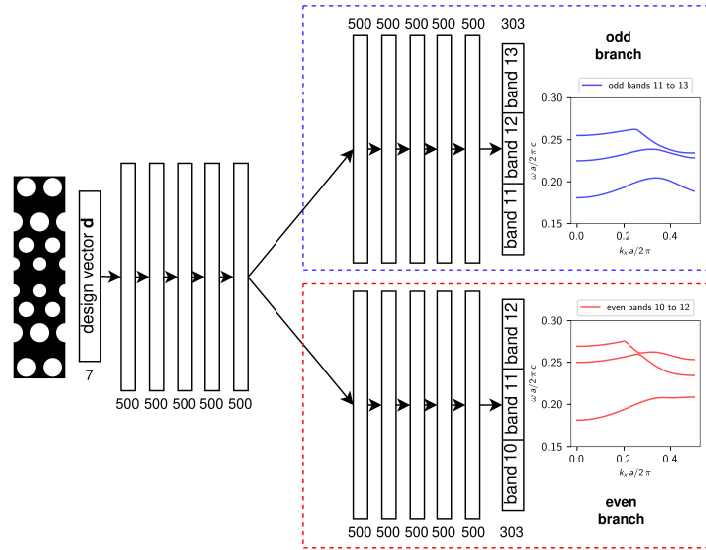


Fig. 3. The input layer takes a 7D design vector \mathbf{d} and the NN outputs two 303 dimensional vectors – one for each parity branch – composed of 3 bands of 101 k_x points each. Arrows show dense connections all of which (except the last) are followed by LeakyReLU activations.

2.4. Direct learning and pretraining

We trained a NN on the 3D dataset with randomly initialised weights (‘direct learning’). Other than the L2 regularisation value which was tuned to 1×10^{-8} , the same hyperparameters were used as for the final 2D NN described in the previous section.

We used the NN from section 2.3 that was trained on the large and relatively inexpensive 2D dataset, as the pretrained NN for the small 3D dataset. Concretely, we gradually unfroze layers for training with a low learning rate and L2 regularisation (following method from Ref. [35]). First, the weights of the two highest layers of each branch were unfrozen and trained with a learning rate of 1×10^{-6} and L2 of 1×10^{-7} until convergence at 5000 epochs. This was repeated for the four highest layers and then all layers but the first 4 dense layers with L2 of 1×10^{-8} . Finally, we fine-tuned the entire model with a low learning rate of 1×10^{-8} . Higher learning rates quickly led to overfitting especially when more layers were trainable, and we arrived at the method outlined here by trial and error.

3. Results

3.1. Prediction of band structure

Our aim is to use the NN to predict the band structure of our PhC waveguides. Thus for an assessment of the accuracy of these predictions, we calculate the relative errors in the predicted frequency $\hat{\omega}_{s,b,k}$ for each band b at each k -value for every sample s in the test set. As shown in Fig. 4 most errors are small with 95% of errors falling in the range -0.322% to $+0.326\%$. To evaluate the accuracy of a single band structure prediction we average over all k_x points (N_k) and all odd and even bands (N_b) for a mean absolute percentage error (MAPE_s) defined as:

$$\text{MAPE}_s = \frac{1}{N_b N_k} \sum_{b,k} \left| \frac{\omega_{s,b,k} - \hat{\omega}_{s,b,k}}{\omega_{s,b,k}} \right| \quad (2)$$

the distribution of which is plotted in Fig. 5. While the MSE was used for the loss function, the MAPE_s is a suitable metric for the evaluation of a band structure prediction as the deviation of all 606 data points is weighted equally.

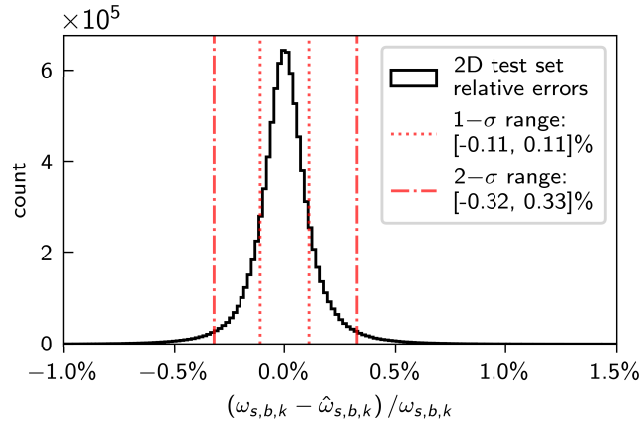


Fig. 4. Distribution of percentage errors (PEs) of individual neurons' outputs across the entire test set showing that 95% of PEs lie in the interval $[-0.32\%, 0.33\%]$

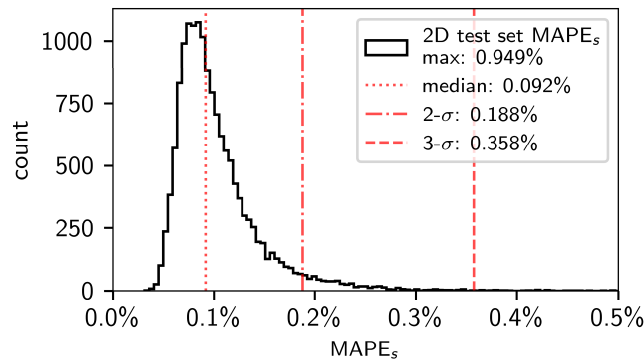


Fig. 5. Distribution of the mean relative error per band structure in the test set. 95% of errors are below 0.188% and the worst prediction had $\text{MAPE}_s = 0.949\%$.

For a qualitative comparison, four predicted and actual band structures are plotted in Fig. 6 corresponding to the median, 95th and 99.7th percentiles as well as the prediction with the highest MAPE_s in the test set.

The MAPE over all 15,000 samples in the 2D test set is 0.103% indicating the NN is highly accurate and can generalize very well to unseen samples.

Once trained, it takes the NN $8.4 \cdot 10^{-4}$ s per design per cpu-core to predict band structures. This is a factor of 214,000 times faster than 2D simulations on the same hardware.

3.2. Group index calculation

Since PhCWs are typically used for their slow light enhancement of light-matter interactions, the key figures of merit not only include the bandstructure (i.e. operating frequency), but also the group index and group index distribution, typically quantified through the group index bandwidth product (GBP) [10]. Therefore once both the group index and frequency are known as a function

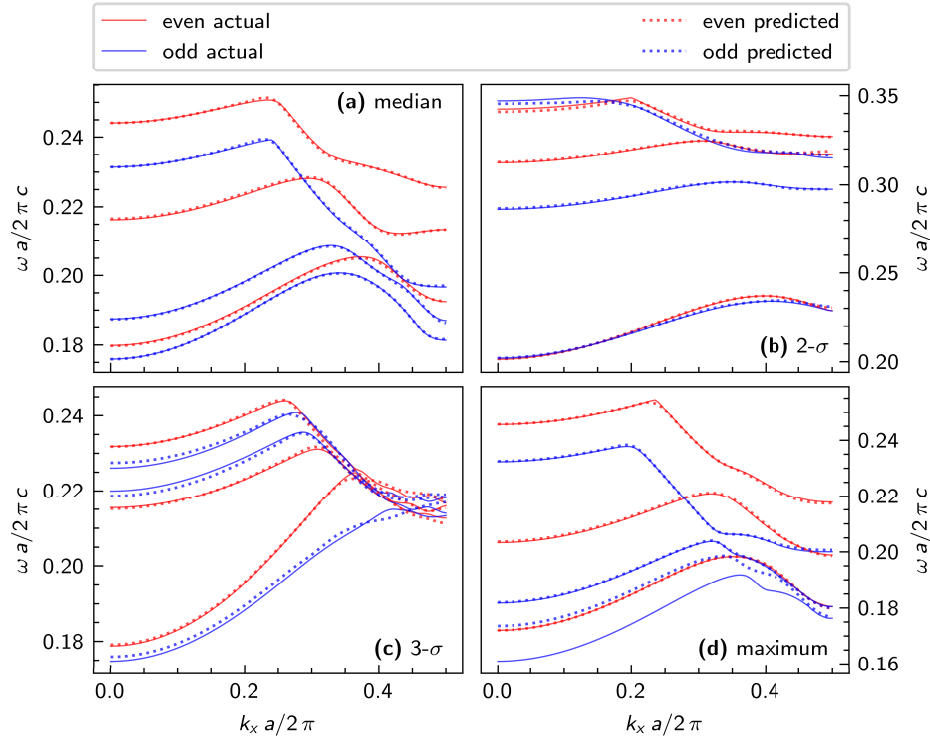


Fig. 6. Predicted (dotted lines) and actual (solid lines) band structures from the selected boundaries shown in Fig. 5.

of k_x , we have the key elements required from PhC waveguide simulations. The group index, n_g is related to the gradient of the bandstructure via:

$$n_g(k_x) = \frac{dk}{d\omega(k_x)}, \quad (3)$$

and we can obtain the group index distribution through numerical differentiation of each band.

An example group index plot is shown in Fig. 7 for the band structure with the greatest MAPE_s of the entire test set. Even in this largest error case, the two sets of group index curves are visually in good agreement, indicating that we can use the output from our neural network to predict the group index of PhCWs.

Due to the divergence of the group index where the bands are flat, the MAPE is not a meaningful metric, as small changes in the exact divergence point can quickly lead to extremely large errors, localized in a very small number of k -points. A similar effect occurs at stationary points (e.g. inflection points) within the band. Additional exemplary group index plots are shown in the Supplement 1.

3.3. 3D dataset: direct learning and pretraining

We trained a NN on the 3D dataset with randomly initialised weights ('direct learning') which achieved a MAPE of $\sim 1.22\%$ on the 3D test set containing 150 elements.

The pretrained NN reached a 3D test set MAPE of $\sim 0.296\%$. Pretraining therefore reduced the test error by 75% compared to direct learning, indicating highly accurate predictions of PhC waveguide bands using a NN, with an accuracy matching 3D simulations. To assess the impact of training set size on the error, we trained 5 NN on subsets of the 2D training and validation sets.

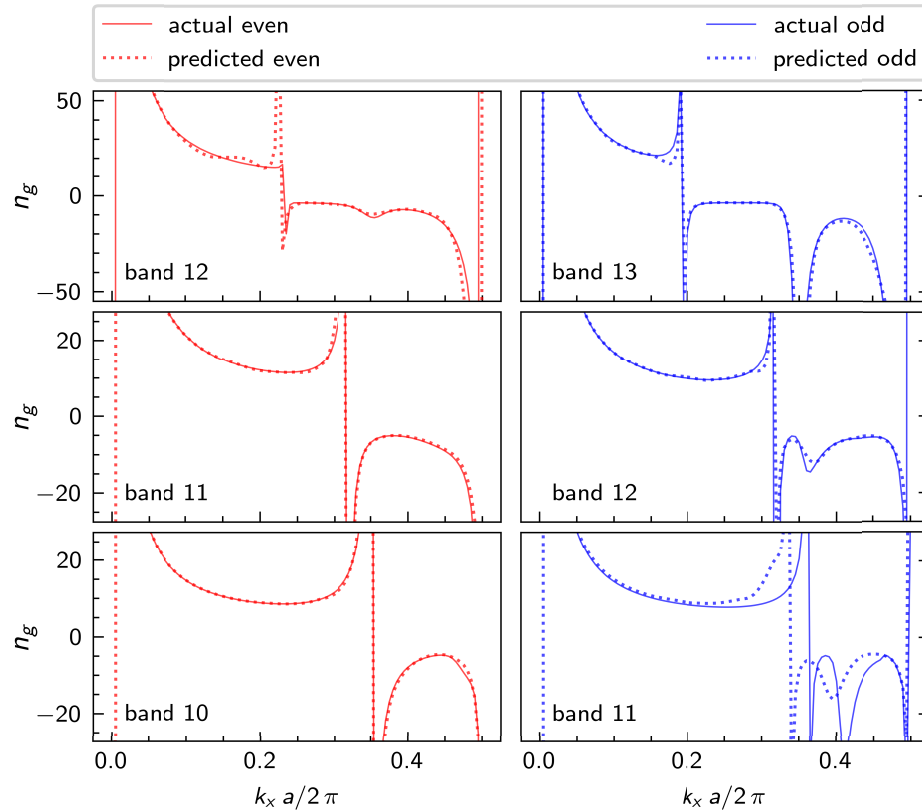


Fig. 7. The group index as a function of k_x (Eq. (3)) for the worst predicted band structure in the test set (Fig. 6(d)). The actual (solid lines) and predicted (dotted lines) bands are plotted qualitatively showing good agreement in the slow light region.

The dataset sizes and corresponding relative errors are plotted in Fig. 8. To achieve the same accuracy with direct learning a prohibitively large 3D dataset would have been required.

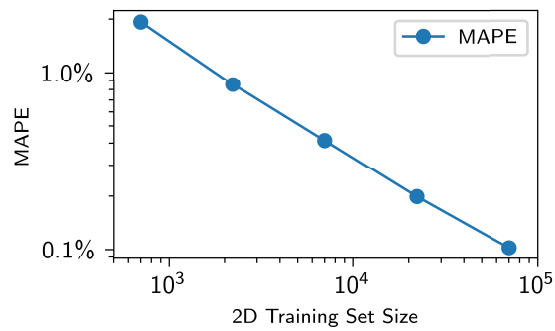


Fig. 8. Learning curve showing the 2D test set relative error of NNs trained on subsets of the 2D training set of varying size.

4. Discussion

Our initial NN was trained on a large dataset generated using 2D simulations and the MAPE used to evaluate the predictive accuracy was calculated over a test dataset of 15,000 elements. The mean relative error (0.103%), together with a small spread of the MAPE_s shown in Fig. 5 (max (MAPE_s) = 0.949%) indicates that the NN generalises well to unseen designs. A percentage error of 0.1% corresponds to a 1.55 nm error in the operating wavelength for a waveguide operated near $\lambda = 1550$ nm. A similar shift is associated with a 0.5 nm variation in the hole diameter and slab thickness [42], thus the predictions from our NN fall well within the variation of state-of-the-art fabrication processes.

Although our 2D NN is capable of accurately predicting the band structure of PhC waveguides, 2D simulations produce biased band structures and the associated group index curves are typically 10 to 20% out [16] compared to those obtained using 3D simulations, which generally match experimental results [42]. The creation of a 100,000 element 3D training set is prohibitively computationally expensive. As a pilot, we generated a 1000 element 3D dataset to directly train a NN. The resulting 3D test set predictions had an unsatisfactory MAPE of 1.22%. Motivated by recent applications of transfer learning in photonics [28], we used the pre-trained 2D NN as a starting point for training on the 3D dataset. The resulting error on the 3D dataset was 0.296% which was 75% lower than by direct learning highlighting the utility of this method for problems where only small datasets are computationally feasible.

Our NN models the forward problem of predicting 6 bands near the area of the expected band gap given a 7D vector of structural PhCW parameters. NNs to predict a single Figure of Merit (FoM), the Normalised Delay Bandwidth Product, were used in an alternative machine learning approach to evaluate PhCW performance [27]. In comparison, the key drawback of our approach is the increased complexity of predicting 606 outputs, which necessitates a larger training set and a NN with more parameters than for the direct prediction of a single FoM. We believe this is outweighed by the additional knowledge that can be gained from observing the full bands, which can answer important questions. For example: is the waveguide single mode or multi-mode at a given frequency? Do these additional modes lie below or above the light line? These factors strongly impact propagation losses [33] and should be considered when deciding which designs can be used for a given application.

The NN can also be used to gain insight into the behaviour of PhCWs, e.g. by plotting the evolution of bands as design parameters are varied.

For the group index prediction, we note that the NN predicted curves are ‘noisy’ compared to those calculated from MPB simulations (Fig. 7). The accuracy of band predictions (and further numerical processing) could be improved by smoothing bands for example by applying a moving average over the bands and/or group index curves, similar to methods used to remove noise from experimental group index measurements [43]. Alternatively, the smoothing function itself can be learned by appending trainable convolutional layers to the NN [44].

We note there are design parameter combinations for which all or most bands are in fact bulk modes, however, this is not a feature of the NN. Instead, it is a consequence of our training data generation pipeline, where we selected specific bands by band index, rather than identifying a bandgap and then selecting the bands that are within the bandgap.

5. Conclusion

We present a neural network that offers a fast and highly accurate alternative to conventional MPB simulations of photonic crystal waveguides. Our NN achieves an improvement in simulation time of 5 orders of magnitude compared to 2D simulations, and of 7 orders compared to 3D simulations, with a sub-millisecond CPU time per design.

We compare the NN band structure predictions with a large test set of 2D-MPB simulations and show that our NN has a mean absolute percentage error $\sim 0.1\%$ and also predicts the group index behaviour of the designs under consideration. We use this NN as a pretrained NN for a small dataset of full 3D simulations. Comparison with a 3D test set yields a MAPE $\sim 0.3\%$, however, this time with respect to 3D simulations that closely match experimental results [42]. This indicates that after pretraining our NN is in fact more accurate than 2D simulations.

This represents a new paradigm in PhC waveguide design. The reduction in computation time means that optimization algorithms or other machine learning-based selection mechanisms can now be used to search a vastly larger range of the full PhC waveguide parameter space, without increase in the computation cost. Similarly, the existence of a forward NN allows the use of machine-learning based approaches to inverse design [45]; our NN could act as a pre-trained part of an inverse NN architecture [46]. This opens the door for application-specific PhCW designs and hence the realization of the full potential of slow light in PhCWs.

Funding. Engineering and Physical Sciences Research Council (EP/V029975/1).

Acknowledgments. The authors thank Andrea Di Falco for useful discussions.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [34].

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. T. F. Krauss, "Why do we need slow light?" *Nat. Photonics* **2**(8), 448–450 (2008).
2. T. Baba, "Slow light in photonic crystals," *Nat. Photonics* **2**(8), 465–473 (2008).
3. D. Vercauteren, N. V. Sampaio, K. Y. Yang, and J. Vučković, "Inverse-designed photonic crystal circuits for optical beam steering," *ACS Photonics* **8**(10), 3085–3093 (2021).
4. T. Baba, H. Ito, H. Abe, T. Tamanuki, Y. Hinakura, R. Tetsuya, J. Maeda, M. Kamata, R. Kurahashi, and R. Shiratori, "Si pic based on photonic crystal for lidar applications," in *Optical Fiber Communication Conference (OFC) 2020*, (Optica Publishing Group, 2020), p. M4H.1.
5. K. Kondo and T. Baba, "On-chip autocorrelator using counter-propagating slow light in a photonic crystal with two-photon absorption photodiodes," *Optica* **4**(9), 1109–1112 (2017).
6. O. S. Magaña-Loaiza, B. Gao, S. A. Schulz, K. M. Awan, J. Upham, K. Dolgaleva, and R. W. Boyd, "Enhanced spectral sensitivity of a chip-scale photonic-crystal slow-light interferometer," *Opt. Lett.* **41**(7), 1431–1434 (2016).
7. M. A. Gaafar, D. Jalas, L. O'Faolain, J. Li, T. F. Krauss, A. Y. Petrov, and M. Eich, "Reflection from a free carrier front via an intraband indirect photonic transition," *Nat. Commun.* **9**(1), 1447 (2018).
8. L. H. Frandsen, A. V. Lavrinenko, J. Fage-Pedersen, and P. I. Borel, "Photonic crystal waveguides with semi-slow light and tailored dispersion properties," *Opt. Express* **14**(20), 9444–9450 (2006).
9. J. Li, T. P. White, L. O'Faolain, A. Gomez-Iglesias, and T. F. Krauss, "Systematic design of flat band slow light in photonic crystal waveguides," *Opt. Express* **16**(9), 6227–6232 (2008).
10. S. A. Schulz, L. O'Faolain, D. M. Beggs, T. P. White, A. Melloni, and T. F. Krauss, "Dispersion engineered slow light in photonic crystals: a comparison," *J. Opt.* **12**(10), 104004 (2010).
11. L. O'Faolain, S. A. Schulz, D. M. Beggs, T. P. White, M. Spasenović, L. Kuipers, F. Morichetti, A. Melloni, S. Mazoyer, J. P. Hugonin, P. Lalanne, and T. F. Krauss, "Loss engineered slow light waveguides," *Opt. Express* **18**(26), 27627–27638 (2010).
12. J. Li, L. O'Faolain, S. A. Schulz, and T. F. Krauss, "Low loss propagation in slow light photonic crystal waveguides at group indices up to 60," *Photonics Nanostructures - Fundam. Appl.* **10**(4), 589–593 (2012).
13. S. G. Johnson and J. D. Joannopoulos, "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis," *Opt. Express* **8**(3), 173–190 (2001).
14. L. C. Andreani and D. Gerace, "Light-matter interaction in photonic crystal slabs," *Phys. Status Solidi B* **244**(10), 3528–3539 (2007).
15. M. Qiu, "Effective index method for heterostructure-slab-waveguide-based two-dimensional photonic crystals," *Appl. Phys. Lett.* **81**(7), 1163–1165 (2002).
16. S. Schulz, A. Park, I. De Leon, J. Upham, and R. Boyd, "Beyond the effective index method: improved accuracy for 2d simulations of photonic crystal waveguides," *J. Opt.* **17**(7), 075006 (2015).
17. M. Minkov and V. Savona, "Wide-band slow light in compact photonic crystal coupled-cavity waveguides," *Optica* **2**(7), 631–634 (2015).
18. S. M. Mirjalili, S. Mirjalili, A. Lewis, and K. Abedi, "A tri-objective particle swarm optimizer for designing line defect photonic crystal waveguides," *Photonics Nanostructures-Fundamentals Appl.* **12**(2), 152–163 (2014).

19. S. Billings, S. A. Schulz, J. Upham, and R. W. Boyd, "Application-tailored optimisation of photonic crystal waveguides," *J. Opt.* **18**(11), 115005 (2016).
20. O. Khatib, S. Ren, J. Malof, and W. J. Padilla, "Deep learning the electromagnetic properties of metamaterials—a comprehensive review," *Adv. Funct. Mater.* **31**(31), 2101748 (2021).
21. J. Jiang, M. Chen, and J. A. Fan, "Deep neural networks for the evaluation and design of photonic devices," *Nat. Rev. Mater.* **6**(8), 679–700 (2020).
22. W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, "Deep learning for the design of photonic structures," *Nat. Photonics* **15**(2), 77–90 (2021).
23. M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin, and K. Parsons, "Deep neural network inverse design of integrated photonic power splitters," *Sci. Rep.* **9**(1), 1368 (2019).
24. T. Christensen, C. Loh, S. Picek, D. Jakobovic, L. Jing, S. Fisher, V. Ceperic, J. D. Joannopoulos, and M. Soljagic, "Predictive and generative machine learning models for photonic crystals," *Nanophotonics* **9**(13), 4183–4192 (2020).
25. A. Nikulin, I. Zisman, M. Eich, A. Y. Petrov, and A. Itin, "Machine learning models for photonic crystal band diagram prediction and gap optimisation," *Photonics Nanostructures-Fundamentals Appl.* **52**, 101076 (2022).
26. F.-L. Hsiao, H.-F. Lee, S.-C. Wang, Y.-M. Weng, and Y.-P. Tsai, "Artificial neural network for photonic crystal band structure prediction in different geometric parameters and refractive indexes," *Electronics* **12**(8), 1777 (2023).
27. K. Hirotani, R. Shiratori, and T. Baba, "Si photonic crystal slow-light waveguides optimized through informatics technology," *Opt. Lett.* **46**(17), 4422–4425 (2021).
28. Y. Qu, L. Jing, Y. Shen, M. Qiu, and M. Soljagic, "Migrating knowledge between physical scenarios based on artificial neural networks," *ACS Photonics* **6**(5), 1168–1174 (2019).
29. J. Zhang, C. Qian, Z. Fan, J. Chen, E. Li, J. Jin, and H. Chen, "Heterogeneous transfer-learning-enabled diverse metasurface design," *Adv. Opt. Mater.* **10**(17), 2200748 (2022).
30. S. Schulz, C. Hu, J. Upham, R. W. Boyd, and L. O'Faolain, "Controllable low-loss slow light in photonic crystals," in *Steep Dispersion Engineering and Opto-Atomic Precision Metrology XI*, S. M. Shahriar and J. Scheuer, eds. (SPIE, 2018).
31. T. Tamura, K. Kondo, Y. Terada, Y. Hinakura, N. Ishikura, and T. Baba, "Silica-clad silicon photonic crystal waveguides for wideband dispersion-free slow light," *J. Lightwave Technol.* **33**(14), 3034–3040 (2015).
32. Y. Zhao, Y.-N. Zhang, Q. Wang, and H. Hu, "Review on the optimization methods of slow light in photonic crystal waveguide," *IEEE Trans. Nanotechnol.* **14**(3), 407–426 (2015).
33. E. Kuramochi, M. Notomi, S. Hughes, A. Shinya, T. Watanabe, and L. Ramunno, "Disorder-induced scattering loss of line-defect waveguides in photonic crystal slabs," *Phys. Rev. B* **72**(16), 161318 (2005).
34. C. Schwahn and S. A. Schulz, "Accurate and efficient prediction of photonic crystal waveguide bandstructures using neural networks - dataset." *University of St Andrews Research Portal* (2023).
35. A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (O'Reilly Media, Inc., 2019), 2nd ed.
36. B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," (2015).
37. F. Chollet, "Keras," <https://keras.io> (2015).
38. M. Abadi, A. Agarwal, and P. Barham, *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," (2015). Software available from tensorflow.org.
39. T. Dozat, "Incorporating Nesterov Momentum into Adam," in *Proceedings of the 4th International Conference on Learning Representations*, (2016), pp. 1–4.
40. L. Prechelt, *Early Stopping — But When?* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012), pp. 53–67.
41. J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks* **61**, 85–117 (2015).
42. K. M. Awan, S. A. Schulz, D. X. Liu, K. Dolgaleva, J. Upham, and R. W. Boyd, "Post-process wavelength tuning of silicon photonic crystal slow-light waveguides," *Opt. Lett.* **40**(9), 1952–1955 (2015).
43. A. Gomez-Iglesias, D. O'Brien, L. O'Faolain, A. Miller, and T. F. Krauss, "Direct measurement of the group index of photonic crystal waveguides via fourier transform spectral interferometry," *Appl. Phys. Lett.* **90**(26), 261107 (2007).
44. C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," *Opt. Express* **27**(20), 27523–27535 (2019).
45. D. Midtvedt, V. Mylnikov, A. Stilgoe, M. Käll, H. Rubinsztein-Dunlop, and G. Volpe, "Deep learning in light–matter interactions," *Nanophotonics* **11**(14), 3189–3214 (2022).
46. R. Singh, A. Agarwal, and B. W. Anthony, "Mapping the design space of photonic topological states via deep learning," *Opt. Express* **28**(19), 27893–27902 (2020).