Multilayer Perceptron Models for Band Diagram Prediction in bi-dimensional Photonic Crystals

Adriano da Silva Ferreira[1,2], Gilliard Nardel Malheiros Silveira[1,3], and Hugo Enrique Hernández Figueroa[1]

[1]School of Electrical and Computer Engineering (FEEC), University of Campinas (UNICAMP), Campinas, Brazil.
[2]São Paulo Federal Institute of Education, Science and Technology (IFSP), Hortolândia, Brazil.
[3]São Paulo State University (UNESP), São João da Boa Vista, Brazil.
{adrianof, hugo}@decom.fee.unicamp.br, gilliardn@yahoo.com.br

Abstract—We modeled Multilayer Perceptron (MLP) Artificial Neural Network for predicting band diagrams (BD) of bi-dimensional photonic crystals. Datasets for MLP training were created by relating geometric and material properties to BDs of triangular- and square-lattice photonic crystals. We demonstrate that fast-training MLP models are able to estimate accurate BDs and existing photonic band gaps through rapid computations.

Index Terms—photonic crystal; photonic band gap; multilayer perceptron; prediction.

I. INTRODUCTION

PHOTONIC crystals (PhCs) are optical structures composed of dielectric materials periodically distributed in one, two or three dimensions. Such structure can be designed to exhibit photonic band gaps (PBGs): spectral regions in which light cannot propagate in certain directions [1]. PhCs presenting complete PBGs (C-PBGs) are capable of blocking light propagation in all directions for transverse electric (TE) and transverse magnetic (TM) light polarizations. PBG-based PhCs are therefore potential solutions for the problem of light control and manipulation [2].

PBG design generally relies upon design processes which require iterative computations of optical properties. Electromagnetic solvers can effectively perform such computations by means of numerical methods such as Finite Elements [3] and Plane Wave Expansion [4], among others. However, such numerical analyses are often computationally-demanding, leading to time-consuming design processes [6].

In order to provide an efficient procedure for building band diagrams (BDs) of bi-dimensional (2D) PhCs, we modeled Multilayer Perceptron (MLP) Artificial Neural Network (ANN) [5] for speedy computations of PhC BDs by means of fast-training ANN models. The MLP fast computing capability has been reported in the problem of calculating coupling efficiency of 2D photonic couplers [6] and PBG's relative bandwidth of square-lattice 2D PhCs [7], [8]. Furthermore, a study with square-lattice 2D PhCs demonstrated that an MLP model was able to process PhC geometric parameter to predict TE fundamental modes, and another MLP model was able to process geometric and material attributes to estimate PBGs and their central frequencies [9]. Finally, MLP was also employed to compute high order modes of 2D and 3D photonic crystals with different lattices and material backgrounds [10].

Here, we designed an MLP model to process PhC geometric parameters and estimate the TM and TE BDs of a testing triangular-lattice 2D PhC, as well as we modeled an MLP to process PhC geometric and material attributes to build the TM and TE BDs of a square-lattice 2D PhC. Also, existing PBGs and C-PBGs of the case-study PhCs are computed from the MLP predicted modes. Thus, it is demonstrated that unique MLP models are capable of providing BDs and PBGs at once.

This paper is organized as follows. Section 2 presents the case-study PhC types and the creation of MLP training datasets as well as the MLP basic concepts and training methodology. Section 3 presents the MLP's BD and PBG predictions and computing performance regarding MIT Photonic Bands (MPB), an electromagnetic solver which computes BDs and electromagnetic modes of periodic dielectric structures [11]. Finally, section 4 discusses the results and draw conclusions.

II. MATERIALS AND METHODS

In this section, we present the case-study PhC types, the MLP basic concepts and architectures, and the dataset generation for MLP training. All PhCs' BDs simulations were performed with mesh resolution $64/a$ (where a is the PhC constant lattice) and 16 interpolated k points between the Irreducible Brillouin Zone (IBZ) corner points.

A. PhC types

We first present a triangular-lattice 2D PhC type composed of dielectric cylinders with radius $r = 0.2a$ and embedded in air [2] (Fig. 1(a)). The cylinders and air's dielectric constants are $\epsilon = 12$ and $\epsilon = 1$, respectively. We considered the BD's lowest two TM and TE modes for evaluating the ANN approach, since this PhC originally possesses an absolute TM PBG of $\Delta\omega = 0.1708(\omega a/2\pi c)$, at the central frequency of $\bar{\omega} = 0.3598(\omega a/2\pi c)$−lower frequency of $0.2744(\omega a/2\pi c)$ and upper frequency of $0.4452(\omega a/2\pi c)$−with a fractional gap size of $\Delta\omega/\bar{\omega} = 47.47\%$ (Fig. 1(b)).

Secondly, we present a square-lattice 2D PhC type composed of dielectric cylinders with radius $r = 0.28a$, connected by dielectric veins with width $w = 0.042a$ [12] (Fig. 2(a)). Cylinders and veins have dielectric constant $\epsilon = 11.56$ and are embedded in air. We considered the lowest four TM modes since there are PBGs between the
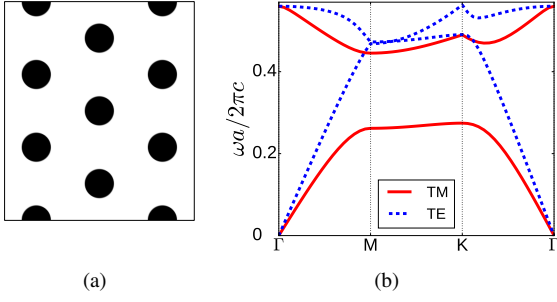
Fig. 1. (a) Triangular-lattice PhC type composed of dielectric cylinders and (b) its TM/TE BD. Γ, M, K, and Γ represent the IBZ's corner points, ω the angular frequency, a the constant lattice and c the speed of light in vacuum.
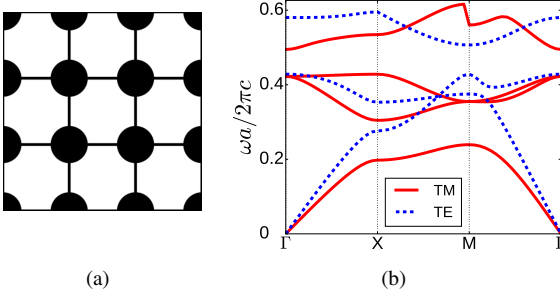

Fig. 2. (a) Square-lattice PhC composed of dielectric cylinders and veins and (b) its TM/TE BD. Γ, X, M, Γ represent the IBZ's corner points, w the angular frequency, a the constant lattice and c the speed of light in vacuum.

1-2 curves (delimited by the lower and upper frequencies of $0.2393(\omega a/2\pi c)$ and $0.3045(\omega a/2\pi c)$, respectively, with $\Delta\omega = 0.0652$, $\bar{\omega} = 0.2719$ and $\Delta\omega/\bar{\omega} = 23.97\%$), and the 2-4 curves (lower frequency of $0.4282(\omega a/2\pi c)$ and upper frequency of $0.4947(\omega a/2\pi c)$, with $\Delta\omega = 0.0665$, $\bar{\omega} = 0.4614$ and $\Delta\omega/\bar{\omega} = 14.41\%$). We also considered the three lowest TE modes since a PBG appears between the 2-3 curves (lower and upper frequencies of $0.4285(\omega a/2\pi c)$ and $0.5066(\omega a/2\pi c)$, respectively, with $\Delta\omega = 0.0781(\omega a/2\pi c)$, $\bar{\omega} = 0.46755(\omega a/2\pi c)$ and $\Delta\omega/\bar{\omega} = 16.70\%$). Moreover, this PhC possesses a C-PBG of $0.0662(\omega a/2\pi c)$ between the 2th TE and 4th TM modes (lower frequency of $0.4285(\omega a/2\pi c)$ and upper frequency of $0.4947(\omega a/2\pi c)$, with $\Delta\omega = 0.0662(\omega a/2\pi c)$, $\bar{\omega} = 0.4616(\omega a/2\pi c)$ and $\Delta\omega/\bar{\omega} = 14.34\%$) (Fig. 2(b)).

B. PhC dataset generation

We have generated different datasets for MLP training. PhCs are characterized by their geometric and material parameters. First, datasets that relate only PhC geometric data and respective TM and TE BDs were created for the case study with the triangular-lattice PhC, allowing us analyzing the MLP capability for computing BDs and PBGs only from PhC geometric data. Secondly, we generated datasets relating PhC geometric and material parameters to TM and TE BDs for the square-lattice PhC.

The triangular-lattice PhC type was geometrically parameterized by the cylinders' radius r. Five variant PhCs were

created for MLP training by linearly varying the original radius. This process produced variant training PhCs with the following radii: $0.12a$, $0.16a$, $0.2a$, $0.24a$, $0.28a$. Two distinct datasets, relating training PhCs' geometric data and respective BDs, were then created for each polarization.

We built datasets with geometric and material properties by linearly modifying geometric values and using two distinct dielectric constants. The cylinders' radius and veins' width of the square-lattice PhC type were multiplied by the proportions [0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3]. By combining the seven different radii and widths and then merging these forty-nine geometric configurations with two dielectric constants $\epsilon_1 = 8.9$ and $\epsilon_2 = 11.56$, ninety eight variant PhCs were created for MLP training. After computing their TE and TM BDs with MPB, one dataset was created for each polarization.

C. Multilayer Perceptron: basic aspects

MLP [5] is a classical multilayer feed-forward ANN whose architecture consists of an input layer, one or more hidden layers, and an output layer. Adjacent layers are fully interconnected. An input signal (input layer) is sent through−and processed by−hidden neurons in a layer-by-layer basis towards the output layer. The neural model is the perceptron, which consists of weights, bias, a summation, and a linear or non-linear activation function (Fig. (3)).

A perceptron of a given layer processes an m-length input signal $\{x_j\}_{j=1}^m$ as

$$y = f\left(\sum_{j=1}^m w_j x_j + b\right),$$

where w_j is the weight of the input signal's element x_j, b is the bias, and $y = f(.)$ is the activation function, which can be linear or non-linear, characterizing the perceptron as non-linear structure and MLP as a generic tool for non-linear function mappings [5]. Such mappings can be performed in a supervised manner by training algorithms that adjust the MLP weights in order to map the relation between input-output data.

D. MLP training method and designed architectures

Given a finite labeled dataset $\left\{(s_i, d_i)\right\}_{i=1}^N$ which relates N pairs of input patterns $\{s_i\}_{i=1}^N$ and target outputs $\{d_i\}_{i=1}^N$, the MLP training process is accomplished by iterative adjustments to the MLP weights aiming to minimize the different between the MLP answers $\{y_i\}_{i=1}^N$ and $\{d_i\}_{i=1}^N$ when processing $\{s_i\}_{i=1}^N$. An error-based metric of performance indicates the proximity between MLP answers and target outputs.

When MLP processes the input pattern s_i, a squared error is generated at the output of the network as

$$E(i) = \sum_{z=1}^q (d_{iz} - y_{iz})^2, \tag{1}$$

where q is the number of neurons in the output layer and d_{iz} and y_{iz} are the z-th elements of the target output and the MLP output, respectively. An average performance regarding

the whole dataset $\left\{(s_i, d_i)\right\}_{i=1}^N$ is obtained by means of a mean squared error (MSE) as

$$MSE = J(W) = \frac{1}{N} \sum_{i=1}^N E(i), \qquad (2)$$

where $E(i)$ is given by (1), W is the whole set of MLP weights and $J(.)$ is the cost function. The training goal is to find a W^* set that minimizes (2). We shall organize the W set through a number of hidden layers and hidden neurons in each layer, until finding an MLP architecture that provides a suitable dimension to obtain the W^* set and therefore to define a good approximation model.
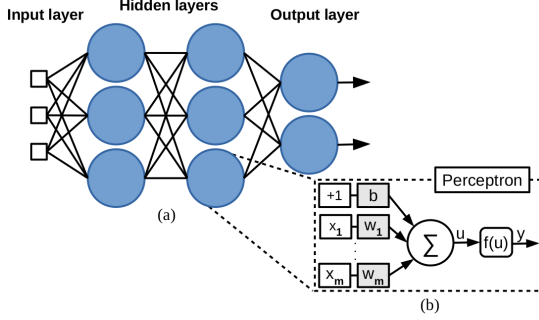

Fig. 3. Schematic of an (a) ANN with two hidden layers and (b) a perceptron with input $\{x_j\}_{j=1}^m$, weights $\{w_j\}_{j=1}^m$, bias b and activation function $f(u)$.

We proposed the following MLP scheme for computing PhC BDs: considering that 2D PhC's BDs consist of TM and TE modes that are formed by the set of all normalized frequencies given at every wave vector element (k point), an MLP structure can be designed to map the relation between geometric and material attributes (inputs) and the set of normalized frequencies (outputs) at a specific k point. Therefore, the MLP input layer can be formed by an l-length set of geometric attributes $(g_1, ..., g_l)$ along with the dielectric constant ϵ and the k point and magnitude $(k_x, k_y, k_z, k_{mag}/2\pi)$ as

$$s = [(g_1, g_2, ..., g_l), \epsilon, (k_x, k_y, k_z, k_{mag}/2\pi)], \qquad (3)$$

(ϵ is omitted for datasets with only geometric data). The MLP output layer can be formed by q output neurons which estimate the q modes' normalized frequencies at a given k point as

$$d = [f_1, ..., f_n, ..., f_q]. \qquad (4)$$

We trained one MLP for each created dataset, which resulted in distinct ANN architectures. By incrementing the number of hidden layers and hidden neurons, we trained different ANN structures until finding suitable MLP models. MLP has been trained with the second order, limited-memory Broyden-Fletcher-Goldfarb-Shanno-B (L-BFGS-B) algorithm (a complete description of L-BFGS-B can be found in [13]). Before training MLP, the input patterns were normalized by removing the mean and scaling to unit variance.

Table I presents the obtained ANN architectures for each PhC dataset. Architectures are given in tuples where the number of elements corresponds to the number of MLP layers

TABLE I
MLP ARCHITECTURES PER PhC DATASET.

Dataset	architecture	transfer function*
triangular-lattice PhC:TM	(5, 8, 8, 8, 2)	hyperbolic tangent
triangular-lattice PhC:TE	(5, 8, 8, 8, 2)	hyperbolic tangent
square-lattice PhC:TM	(7, 15, 15, 15, 4)	hyperbolic tangent
square-lattice PhC:TE	(7, 14, 14, 14, 3)	hyperbolic tangent

* output neurons consist of linear transfer functions in all MLP models.

TABLE II
ANN TRAINING PERFORMANCE PARAMETERS.

Dataset	time(s)	α*	iterations	MSE
triangular-lattice PhC:TM	0.17	5e−4	163	4.01e−6
triangular-lattice PhC:TE	0.36	5e−4	391	1.01e−5
square-lattice PhC:TM	8.9	1e−4	1123	1.63e−5
square-lattice PhC:TE	7.63	9e−3	988	9.76e−6

* α is the MLP learning rate and has been fixed along the iterations.

and each element value represent the number of nodes per layer. Table II presents the training parameters performances including the training mean elapsed real times (an average of 5 calls for each dataset).

III. NUMERICAL RESULTS

In this section, the MLP's BDs predicting and computing performances are assessed regarding MPB. We show the MLP and MPB's TM and TE BDs computations for a testing PhC of each type. Then, we compare the run-times of both approaches for each test PhC's TM and TE DB.

A. Triangular-lattice 2D PhC

This testing PhC has cylinders with radius $r = 0.3a$ and dielectric constant $\epsilon = 12$, and presents PBGs between both the 1-2th TM and TE modes. The absolute TM PBG is $\Delta\omega = 0.1708(\omega a/2\pi c)$−lower frequency of $0.2745(\omega a/2\pi c)$ and upper frequency of $0.4452(\omega a/2\pi c)$ −at the central frequency $\bar{\omega} = 0.3599(\omega a/2\pi c)$ and with fractional gap size $\Delta\omega/\bar{\omega} = 47.45\%$. The absolute TE PBG is $\Delta\omega = 0.0369(\omega a/2\pi c)$−lower and upper frequencies of $0.3476(\omega a/2\pi c)$ and $0.3845(\omega a/2\pi c)$, respectively−at $\bar{\omega} = 0.3660(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 10.08\%$.

The BDs and PBGs computed by MLP agreed very well with the MPB results. MLP obtained a TM PBG between the 1-2 modes, with $\Delta\omega = 0.1698(\omega a/2\pi c)$−lower frequency of $0.2754(\omega a/2\pi c)$ and upper frequency of $0.4452(\omega a/2\pi c)$−at $\bar{\omega} = 0.3603(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 47.13\%$ (Fig 5(a)). Also, MLP computed a TE PBG between the 1-2 modes with $\Delta\omega = 0.0366(\omega a/2\pi c)$−lower frequency of $0.3485(\omega a/2\pi c)$ and upper frequency of $0.3851(\omega a/2\pi c)$−at $\bar{\omega} = 0.3668(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 9.97\%$ (Fig. 5(b)). The MLP MSE for predicting the test PhC's TM and TE BDs were 5.15e−6 and 4.72e−6, respectively.

B. Square-lattice PhC

This testing PhC has cylinders' radius $r = 0.34$ and veins' width $w = 0.045$, and dielectric constant $\epsilon = 8.9$. Two TM
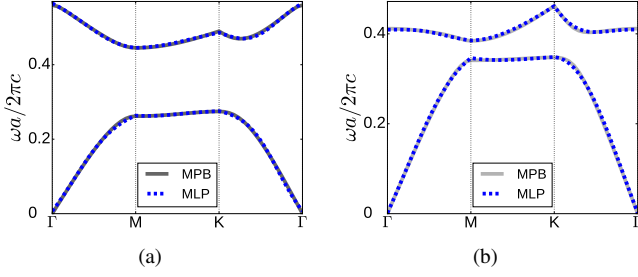
Fig. 4. MPB and MLP computations for the first two TM and TE modes of the testing triangular-lattice PhC: (a) TM and (b) TE BDs.

PBGs appear in this structure: the first one is between the 1-2th TM modes, with $\Delta\omega = 0.0352(\omega a/2\pi c)$−lower frequency of $0.2539(\omega a/2\pi c)$ and upper frequency of $0.2892(\omega a/2\pi c)$−at $\bar{\omega} = 0.2715(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 12.98\%$; and the second is between the 3-4th TM modes, with $\Delta\omega = 0.0783(\omega a/2\pi c)$−lower frequency of $0.4317(\omega a/2\pi c)$ and upper frequency of $0.5100(\omega a/2\pi c)$−at $\bar{\omega} = 0.4708(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 16.62\%$. Furthermore, this PhC has a TE PBG between the 2-3 modes, with $\Delta\omega = 0.054(\omega a/2\pi c)$ −lower frequency of $0.4254(\omega a/2\pi c)$ and upper frequency of $0.4794(\omega a/2\pi c)$−at $\bar{\omega} = 0.4524(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 11.94\%$. Finally, a C-PBG appears between the 3th TM and the 3th TE modes, with $\Delta\omega = 0.0477(\omega a/2\pi c)$ −lower and upper frequencies of $0.4317(\omega a/2\pi c)$ and $0.4794(\omega a/2\pi c)$, respectively−at $\bar{\omega} = 0.4555(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 10.47(\omega a/2\pi c)$.

MLP predicted with good accuracy the two TM PBGs: the first one is between the 1-2 modes, with $\Delta\omega = 0.0300(\omega a/2\pi c)$−lower frequency of $0.2550(\omega a/2\pi c)$ and upper frequency of $0.2850(\omega a/2\pi c)$−at $\bar{\omega} = 0.2700(\omega a/2\pi c)$ and $\Delta\omega/\bar{\omega} = 11.1\%$; and the second one is between the 3-4 modes, with $\Delta\omega = 0.0716(\omega a/2\pi c)$−lower frequency of $0.4339(\omega a/2\pi c)$ and upper frequency of $0.5055(\omega a/2\pi c)$ −at $\bar{\omega} = 0.4697(\omega a/2\pi c)$ and $\Delta\omega/\bar{\omega} = 15.24\%$ (Fig. 5(a)). Also, MLP reasonably predicted the TE PBG between the 2-3th modes, obtaining $\Delta\omega = 0.0496(\omega a/2\pi c)$ −lower frequency of $0.4292(\omega a/2\pi c)$ and upper frequency of $0.4788(\omega a/2\pi c)$−at $\bar{\omega} = 0.454(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 10.92\%$ (Fig. 5(b)). Finally, a C-PBG between both the 3th TM and TE modes was obtained from MLP: $\Delta\omega = 0.0449(\omega a/2\pi c)$−lower and upper frequencies of $0.4339(\omega a/2\pi c)$ and $0.4788(\omega a/2\pi c)$, respectively−at $\bar{\omega} = 0.4563(\omega a/2\pi c)$ and with $\Delta\omega/\bar{\omega} = 9.83\%(\omega a/2\pi c)$. The MLP MSE for estimating the test PhC's TM and TE BDs were $1.79e{-}5$ and $4.06e{-}6$, respectively.

C. MLP and MPB computing performances

Table III lists the MLP and MPB's mean elapsed real times for calculating the testing PhCs. The measured times correspond to a mean of 5 calls of each serial process running on a laptop with Intel Core i-5 2410M (2.3 GHz, 3 MB L3 cache), DDR3 RAM (6 GB, 1333MHz), and Linux Ubuntu. MLP was coded in Python. The Linux wall-clock time function was used for time statistics.

TABLE III
MLP AND MPB MEAN ELAPSED REAL TIMES (S).

Testing PhC	MLP	MPB
triangular-lattice PhC:TM	3.46e−4	1.85
triangular-lattice PhC:TE	2.95e−4	1.98
square-lattice PhC:TM	4.92e−4	4.32
square-lattice PhC:TE	5.26e−4	3.36

Altogether, the MLP models, which were trained within few seconds (Table II), hugely outperformed MPB in terms of runtime, computing PhC BDs with four orders of magnitude faster than MPB.
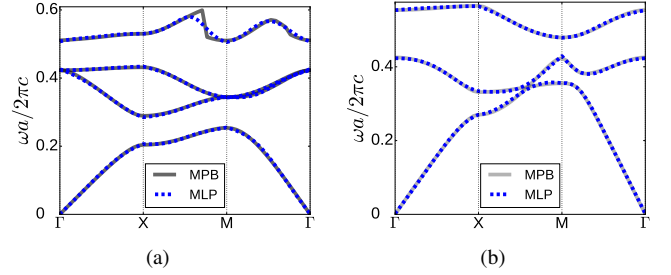

Fig. 5. MPB and MLP computations for the first four TM and first three TE modes of the testing square-lattice PhC: (a) TM and (b) TE BDs.

IV. DISCUSSIONS AND CONCLUSIONS

MLP demonstrated a satisfactorily predicting performance when computing the testing PhCs' BDs. Such results were performed by simple ANN schemes, which only estimate frequencies at a specific k points. It resulted in small ANN models and fast ANN training (supported by efficient training algorithms). The MLP models also computed simultaneously the lowest modes, enabling the calculations of the existing PBGs. The remarkable advantage of MLP was the computing performance, which substantially reduced the runtime for processing PhC BDs in regard to MPB. Combined with the fast training, the proposed approach might be an efficient procedure for supporting electromagnetic solvers, which would provide a fast alternative on the BDs and PBGs computations during PhC design.

REFERENCES

[1] E. Yablonovitch, "Inhibited Spontaneous Emission in Solid-State Physics and Electronics," Physical Review Letters, 58(20), 2059–2062, 1987.

[2] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals, Molding the Flow of Light*, Princeton, NJ: Princeton Univ. Press, 1995.

[3] J. M. Jin, *The Finite Element Method in Electromagnetics*, 2nd ed. New York: Wiley, 2002.

[4] S. Guo and S. Albin, "Simple plane wave implementation for photonic crystal calculations," Opt. Express, vol. 11, no. 2, pp. 167175, 2003.

[5] S. S. Haykin, *Neural networks: a comprehensive foundation*, 2nd Edition, Prentice Hall, Upper Saddle River, NJ, USA 1998.

[6] A. da Silva Ferreira, C. H. da Silva Santos, M. S. Gonçalves, and H. E. Hernández Figueroa, "Towards an integrated evolutionary strategy and artificial neural network computational tool for designing photonic coupler devices," Applied Soft Computing 65, 1–11, 2018.

[7] A. da Silva Ferreira, G. N. Malheiros-Silveira, and H. E. Hernandez-Figueroa, "Complete band-gap prediction of 2D photonic crystals by using multilayer perceptron," in: 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTER-CON), 1-3, 2017.

[8] A. da Silva Ferreira, G. N. Malheiros-Silveira, and H. E. Hernandez-Figueroa, "Predicting complete band-gaps of 2D photonic crystals by using artificial neural networks," in: 2017 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC), 1-5, 2017.

[9] G. N. Malheiros-Silveira and H. E. Hernandez-Figueroa, "Prediction of Dispersion Relation and PBGs in 2-D PCs by Using Artificial Neural Networks," IEEE Photonics Technology Letters, 24(20), 1799–1801, 2012.

[10] A. da Silva Ferreira, G. N. Malheiros-Silveira and H. E. Hernandez-Figueroa, "Computing Optical Properties of Photonic Crystals by Using Multilayer Perceptron and Extreme Learning Machine," Journal of Lightwave Technology, 36(18), 4066-4073, 2018.

[11] S. G. Johnson, J. D. Joannopoulus, "Block-iterative frequency-domain methods for Maxwell's equations in a plane-wave basis," Optics Express, 8(3), 173-190, 2001.

[12] Yuan-Fong Chau, Fong-Lin Wu, Zheng-Hong Jiang, Huang-Yi Li, "Evolution of the complete photonic bandgap of two-dimensional photonic crystal," Optics Express, 19(6), 4862-4867, 2001.

[13] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, Ciyou Zhu, "A Limited Memory Algorithm for Bound Constrained Optimization", SIAM Journal on Scientific Computing, 16(5), 1190-1208, 1995.