

Inverse design of 3D integrated high-efficiency grating couplers using deep learning

Yu Wang(王玉), Yue Wang(王越), Guohui Yang(杨国辉), Kuang Zhang(张狂), Xing Yang(杨星), Chunhui Wang(王春晖), and Yu Zhang(张雨)

Citation: Chin. Phys. B, 2026, 35 (**2**): 024101. DOI: 10.1088/1674-1056/adf69c
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Electron beam modeling and analyses of the electric field distribution and space charge effect

Yueling Jiang(蒋越凌) and Quanlin Dong(董全林)
Chin. Phys. B, 2022, 31 (**5**): 054103. DOI: 10.1088/1674-1056/ac3d84

Switchable directional scattering based on spoof core—shell plasmonic structures

Yun-Qiao Yin(殷允桥), Hong-Wei Wu(吴宏伟), Shu-Ling Cheng(程淑玲), and Zong-Qiang Sheng(圣宗强)
Chin. Phys. B, 2022, 31 (**5**): 054101. DOI: 10.1088/1674-1056/ac3503

Surface plasmon polaritons frequency-blue shift in low confinement factor excitation region

Ling-Xi Hu(胡灵犀), Zhi-Qiang He(何志强), Min Hu(胡旻), and Sheng-Gang Liu(刘盛纲)
Chin. Phys. B, 2021, 30 (**8**): 084102. DOI: 10.1088/1674-1056/abe22e

Electromagnetic field of a relativistic electron vortex beam

Changyong Lei(雷长勇), Guangjiong Dong(董光炯)
Chin. Phys. B, 2020, 29 (**8**): 084102. DOI: 10.1088/1674-1056/ab96a9

Inverse design-based metamaterial transparent device and its multilayer realization

Li Ting-Hua (李廷华), Huang Ming (黄铭), Yang Jing-Jing (杨晶晶), Yuan Gang (袁刚), Cai Guang-Hui (蔡光卉)
Chin. Phys. B, 2014, 23 (**5**): 054102. DOI: 10.1088/1674-1056/23/5/054102

Inverse design of 3D integrated high-efficiency grating couplers using deep learning

Yu Wang(王玉)[1,2], Yue Wang(王越)[2], Guohui Yang(杨国辉)[3], Kuang Zhang(张狂)[3], Xing Yang(杨星)[4,5], Chunhui Wang(王春晖)[2,†], and Yu Zhang(张雨)[4,5,‡]

[1] *The College of Computer and Control Engineering, Northeast Forestry University, Harbin* 150040, *China*
[2] *National Key Laboratory of Laser Spatial Information, Harbin Institute of Technology, Harbin* 150001, *China*
[3] *The School of Electronic and Information Engineering, Harbin Institute of Technology, Harbin* 150001, *China*
[4] *The State Key Laboratory of Pulsed Power Laser Technology, College of Electronic Engineering, National University of Defense Technology, Changsha* 410073, *China*
[5] *Jianghuai Advance Technology Center, Hefei* 230037, *China*

(Received 13 April 2025; revised manuscript received 26 June 2025; accepted manuscript online 1 August 2025)

In recent years, the use of deep learning to replace traditional numerical methods for electromagnetic propagation has shown tremendous potential in the rapid design of photonic devices. However, most research on deep learning has focused on single-layer grating couplers, and the accuracy of multi-layer grating couplers has not yet reached a high level. This paper proposes and demonstrates a novel deep learning network-assisted strategy for inverse design. The network model is based on a multi-layer perceptron (MLP) and incorporates convolutional neural networks (CNNs) and transformers. Through the stacking of multiple layers, it achieves a high-precision design for both multi-layer and single-layer raster couplers with various functionalities. The deep learning network exhibits exceptionally high predictive accuracy, with an average absolute error across the full wavelength range of 1300–1700 nm being only 0.17%, and an even lower predictive absolute error below 0.09% at the specific wavelength of 1550 nm. By combining the deep learning network with the genetic algorithm, we can efficiently design grating couplers that perform different functions. Simulation results indicate that the designed single-wavelength grating couplers achieve coupling efficiencies exceeding 80% at central wavelengths of 1550 nm and 1310 nm. The performance of designed dual-wavelength and broadband grating couplers also reaches high industry standards. Furthermore, the network structure and inverse design method are highly scalable and can be applied not only to multi-layer grating couplers but also directly to the prediction and design of single-layer grating couplers, providing a new perspective for the innovative development of photonic devices.

Keywords: deep learning, inverse design, grating couplers, photonic devices

PACS: 41.85.–p, 42.25.–p, 84.71.Mn

DOI: 10.1088/1674-1056/adf69c **CSTR:** 32038.14.CPB.adf69c

1. Introduction

Silicon photonics platforms have rapidly developed due to their high compatibility with existing semiconductor manufacturing processes, low cost, and ability to integrate densely with electronic chips.[1–4] However, silicon as an optical material faces challenges such as high optical losses in certain spectral regions, susceptibility to crosstalk, nonlinear effects under high power, and temperature sensitivity.[5] These issues limit its performance in telecommunications and high-power applications. Overcoming these limitations requires hybrid integration with other materials to enhance the overall performance and functionality of the silicon photonics platform. Silicon nitride (Si_3N_4) has emerged as a highly regarded alternative material due to its low optical losses and broad transparency window. It offers excellent optical transmission properties and supports complex optical functions such as nonlinear optics and efficient optical modulation. Si_3N_4 has a medium to high refractive index and superior passive optical performance compared to silicon, especially at telecommunication wavelengths. Its nonlinear magnetic susceptibility is much lower than silicon, allowing it to handle higher power levels with minimal losses, making it suitable for high-power optical applications such as photonic integrated LiDAR and biosensors.[1,6] Despite the smaller electro-optic and thermo-optic coefficients of Si_3N_4, which limit its use in active photonic devices, hybrid integration with materials having higher electro-optic or thermo-optic coefficients, such as lithium niobate (LNOI),[7,8] indium phosphide (InP) [9] and gallium arsenide (GaAs),[10,11] can enable active functionalities on a silicon nitride-based platform. This hybrid approach leverages the advantages of Si_3N_4 and introduces other materials to achieve functions such as modulation and switching. Studies have shown that Si_3N_4 waveguides can be integrated into SOI, LNOI and other platforms to form 3D integrated structures, where Si_3N_4 is typically used for the transmission and

†Corresponding author. E-mail: wangch_hit@163.com
‡Corresponding author. E-mail: zhangyu24d@nudt.edu.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

reception of high-power light in waveguides, optical fibers and free space. Si and LNOI are used for active modulation of the optical paths.[1,12]

Grating couplers[13,14] and edge couplers[15] are the two most commonly used optical input/output interfaces. Grating couplers are widely used in various photonic architectures due to their high integration, high alignment tolerance, wide bandwidth and cost-effectiveness. Maire *et al.*[16] designed and fabricated a high-efficiency surface grating coupler for Si_3N_4 waveguides. The results showed that at a wavelength of 1310 nm, the coupling efficiency exceeded 60%, and the −3 dB tolerance of the angle and wavelength was 4° and 50 nm, respectively. However, due to the lower refractive index and grating design limitations of Si_3N_4 grating couplers, their coupling efficiency is usually lower than that of silicon grating couplers. To maximize the coupling efficiency of Si_3N_4 grating couplers as much as possible, Pan *et al.*[17] designed a double-layer focusing grating coupler, in which the first layer is a silicon nitride grating and the second layer is a Si_slab layer. The coupling efficiency reaches 78% at 1550 nm and the 1-dB bandwidth reaches 100 nm. Mak *et al.*[18] successfully designed a SiN-on-Si double-layer grating coupler based on a global optimization algorithm for the first time. The measured median insertion loss is −2.1 dB and the median 1-dB bandwidth is 70.8 nm. Most of the methods used in this type of research are traditional forward design and inverse design methods,[19–21] both of which require complex modeling processes and a large number of computing resources. Moreover, it is difficult to quickly design a grating coupler that meets the needs of various complex applications. Deep learning techniques,[22–25] by predicting the optical properties of nontraditional materials and the performance of complex grating structures, have opened new avenues for the design of grating couplers. This approach can uncover innovative solutions that may be overlooked by traditional manual design processes, assisting researchers in developing new types of grating couplers that meet specific application needs.[26–29] However, current research is limited to single-layer grating couplers, with minimal application in multi-layer designs. The challenges in multi-layer designs mainly stem from the interactions between different layers and the unpredictable interlayer effects, which often reduce the accuracy of predictions.

To accurately predict the coupling efficiency of multi-layer grating couplers, this paper proposes a deep learning-assisted inverse design method. The study first takes the three-layer grating coupler as the basic geometric structure and simulates the grating as a 300×3 matrix, where '0' and '1' represent the fully etched area and the non-etched area, respectively. This matrix data are used as input for training in the deep learning network. To enhance prediction accuracy, we integrate the MLP training module with the CNN and transformer architecture as the core component of the deep learning network. The training results show that the prediction error of the network structure ranges from 0%–0.15% in most cases, with an average error of only 0.17%. Particularly at the specific wavelength of 1550 nm, the prediction error consistently remains below 0.09%. In addition, we combine genetic algorithms with the network structure, using the inverse design approach to optimize grating couplers with varying performances. Simulation results indicate that single-wavelength grating couplers designed using this method can achieve up to 82% coupling efficiencies, comparable to metal mirror grating couplers. At the same time, the performance of dual-wavelength and broadband grating couplers also reached a relatively high level in the industry. Notably, the design time for this inverse design method is less than 5 min, demonstrating a significant speed advantage. Importantly, this network structure and inverse design method are not only applicable to multi-layer grating couplers but can also be directly applied to the prediction and design of single-layer grating couplers, demonstrating extreme scalability.

2. A 3D integrated grating coupler structure design

Figure 1 shows a schematic diagram of the structure of a three-layer grating coupler. From bottom to top, the waveguide materials are LNOI, Si_3N_4 and Si, with thicknesses of 900 nm, 400 nm and 220 nm, respectively. Each layer is separated by SiO_2, with thicknesses of 400 nm, 300 nm, and 2 μm. The thickness of the top SiO_2 cladding is set to 2 μm. The grating area of each layer is composed of 300 discrete pixels represented by '0' and '1', where '0' and '1' represent the fully etched area and the non-etched area, respectively, and the pixel unit size is 50 nm, as shown in Fig. 1(b). Based on these parameters, the corresponding grating coupler model is established and simulated using Lumerical FDTD software. After the simulation is completed, the structural distribution of the grating coupler is discretized into a 300×3 matrix, and the matrix values are composed of randomly distributed '0' and '1'. In addition, the coupling efficiency curve is output along with the structural data to establish the database required for the deep learning network. When the database contains a sufficient number of samples, this structure can be used to approximate the simulation of grating couplers with arbitrary distributions, and the pixel combination can be appropriately adjusted according to the actual processing requirements. When all the matrix values of the upper or lower layers are '0', different combinations of double-layer or single layer Si_3N_4 grating couplers can be represented. This shows that by changing parameters the designed grating coupler architecture is highly scalable and can be used to equivalently simulate traditional grating couplers, sub-wavelength grating couplers and multi-layer superimposed grating couplers.

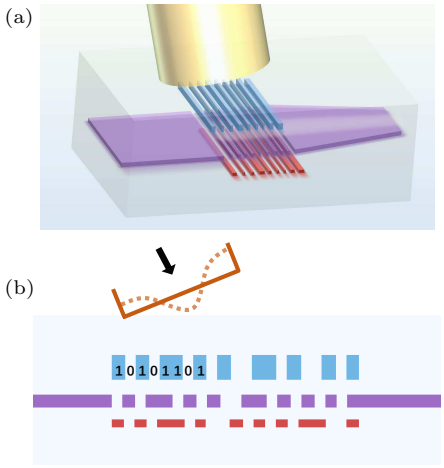


Fig. 1. Schematic structure of three-layer grating coupler. (a) 3D model diagram. (b) Front view.

3. Deep learning network design and inverse design

As shown in Fig. 2, the conventional design process and inverse design process of grating couplers rely on electromagnetic simulation software to scan various parameters and then select the optimal result. Although this method is simple to operate, but its efficiency is very low when there are many design target parameters or the design target is difficult to achieve. This study uses the trained deep learning network structure to replace the traditional electromagnetic simulation steps. This approach significantly reduces the time required for device simulation and can be used in combination with any inverse design algorithm to improve design efficiency, as shown in Fig. 2(c).

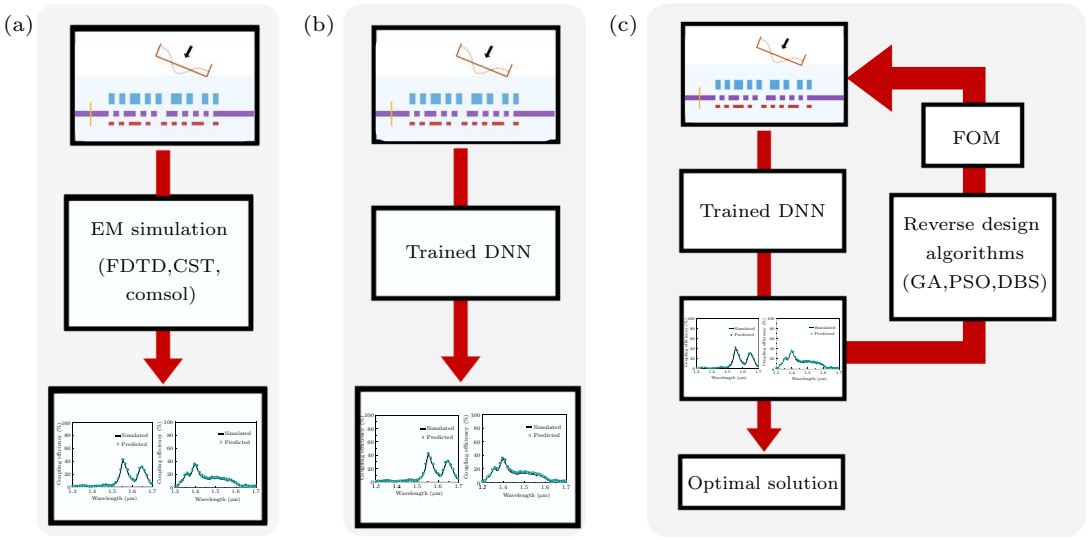


Fig. 2. (a) Prediction flowchart for traditional electromagnetic simulation. (b) Prediction flowchart based on deep learning network structure. (c) Inverse design flowchart incorporating the deep learning network.

The first step in designing the deep learning network is establishing a database. During the simulation process, the pixel combination of the grating couplers is set to a 300×3 matrix with an arbitrary distribution. During the simulation process using Lumerical FDTD, the light source uses a Gaussian beam with the wavelength range of 1300–1700 nm. The position of the light source is centered in the grating area, with a mode field radius of 5.2 μm and an incidence angle of 5°. The boundary conditions around the simulation domain are set to perfectly matched layer (PML), and the mesh grid resolution is established at 5 nm. Based on these simulation parameters, a dataset of 40000 grating couplers with various distributions was simulated for the neural network. To achieve accurate predictions for the majority of structures with minimal data, half of the structures were set to completely random, while the other half were configured as special structures, including single-layer, double-layer, uniform distribution, and spaced values. During the actual training process, the ratio of random structures to special structures was 1:1. The training, validation, and test sets were randomly selected from both the random and special structure databases, with 70% of the data used for the training set, 15% for the test set and 15% for the validation set, ensuring that each dataset contained sufficient numbers of both completely random and special structures. Upon completion of the simulations, we recorded the coupling efficiency, denoted as T, for each grating coupler. These data, along with the discrete 300×3 matrices, were compiled into a dataset. The dataset comprises 201 sampling points, each representing the coupling efficiency values within the wavelength range of 1300–1700 nm, to be used for subsequent analysis and deep learning network training.

The second step involves constructing the deep learning network structure. The structure has been further optimized based on the network architecture previously proposed,[30] consisting of a three-layer neural network structure: MLP, CNN and transformer. Each layer transforms the input data into higher-level feature representations through nonlinear activation functions. Compared to individual neural network lay-

ers (e.g., CNN, GAN, MLP), this design offers advantages such as enhanced feature extraction capability, improved expressive power and greater predictive accuracy. It effectively handles nonlinear problems, making it suitable for predicting the coupling efficiency of grating couplers across multiple angles and a wide wavelength range. Figure 3(a) shows the flowchart of the deep learning network structure prediction. The main steps to achieve prediction using this network structure include:

(i) Generation of a 300×3 matrix corresponding to the output of the arbitrarily distributed multi-layer grating coupler, along with the coupling efficiency curves at different angles.

(ii) Integration of the input matrix data with coupling efficiency using the single pixel information aggregation (SPIA) component, which also incorporates single-pixel positional information.

(iii) Extraction of local structural information using the local information acquisition (LIA) component, which consists of eight residual blocks, each comprising a 1×1 convolution layer and a 1×3 convolution layer.

(iv) Simultaneous embedding of position information and wavelength information into the designed network structure via transpose attention modules, as shown in Fig. 3(b). The output of the wavelength encoding module is concatenated with the input to the transpose attention module and then fed into the first MLP module. The output of the first module, after transposition and concatenation with the output from the position encoding module, is fed into the second MLP module.

(v) Conversion of latent space features into corresponding coupling efficiency through a fully connected module with non-shared parameters, outputting the prediction results.

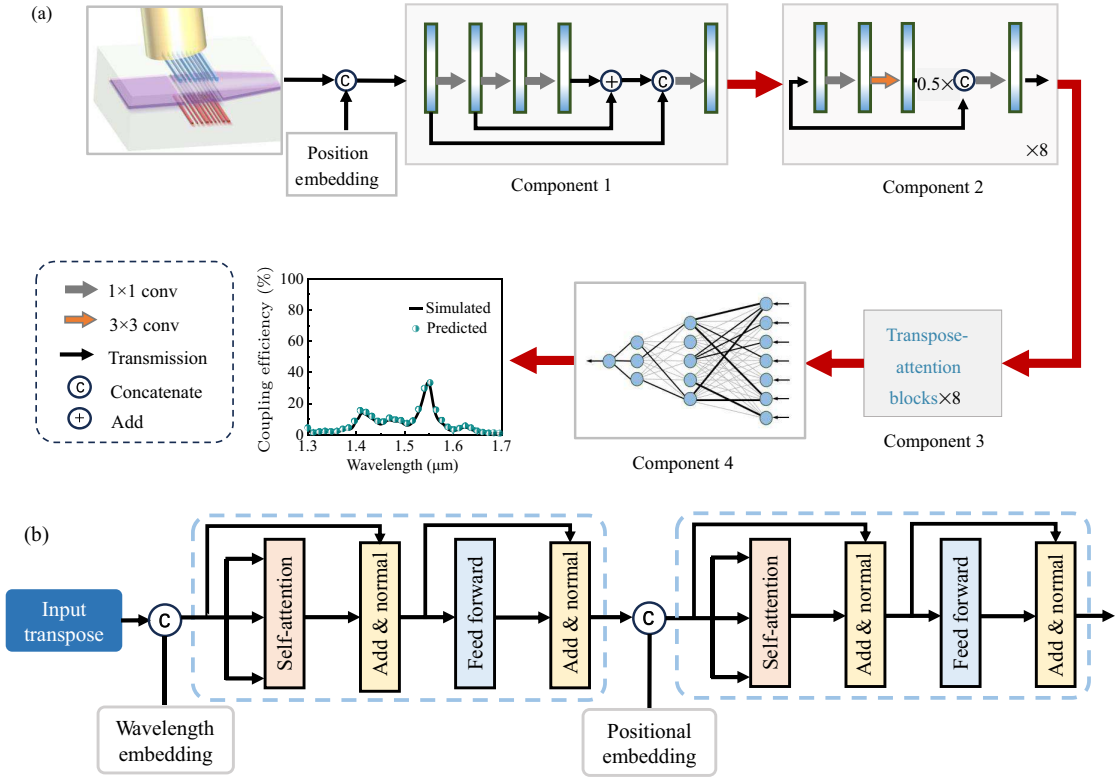


Fig. 3. (a) Prediction flowchart based on the deep learning network structure. (b) Structure of the transposed attention block.

This design aims to precisely predict the performance of grating couplers under various configurations through a highly specialized network structure, thereby facilitating efficient processing and accurate modeling of complex multidimensional data. In the overall training process of the deep learning network, we used the adaptive moment estimation (Adam) algorithm as the optimization method. This algorithm is based on adaptive estimation of first-order gradients using low-order moments. During this phase, the learning rate of the model was set to 0.001, with 80 training epochs and a batch size of 1800. The loss function was set as follows:

$$\mathrm{LOSS} = |\ln Y - \ln T|, \tag{1}$$

where T represents the simulated coupling efficiency at different wavelengths and Y indicates the predicted coupling efficiency from the deep learning network at various wavelengths. The entire training of the neural network and the subsequent reverse design process were conducted on a platform comprising an Intel (R) Core (TM) i7-13700K and two NVIDIA RTX 4090 GPUs. The training time required was approximately 72 hours. As shown in Fig. 4(a), the convergence curve of the trained deep learning network's prediction accuracy for

the coupling efficiency of grating couplers with respect to the iteration number is displayed. With the introduction of the deep learning network, the prediction absolute error (the average error between the simulated coupling efficiency and the predicted coupling efficiency over a wide wavelength range) gradually decreased from an initial 67.1% to approximately 0.29% as the number of iterations increased. After around 40 iterations, the error began to stabilize, with both prediction accuracy and the loss function showing minimal changes. At this stage, the mean squared error (MSE) is 2.18×10^{-5}. These results indicate that the designed model can achieve a stable state with only a few iterations, and the predictions for unknown data are almost consistent with the simulated values. Figure 4(b) presents the average error statistics of the model on the test set, where the majority of the prediction absolute errors range from 0% to 0.15%, with an overall average absolute error of only 0.17%. Notably, at specific wavelengths, such as at 1550 nm, the prediction absolute errors are all less than 0.09%. Figures 4(c) and 4(d) show the comparison between the predicted curve and the actual simulation curve under the lowest and highest prediction accuracy, respectively, with the maximum absolute error being only 2%. In the case of poor prediction results, most errors occur when the coupling efficiency is less than 10%. If the prediction is made only for data with coupling efficiencies higher than 20%, the maximum absolute error is less than 0.3% in all cases. These results indicate that our model predictions are very close to the actual values, and their accuracy remains at a high level.[31–34] In addition, the prediction results from the neural network were trained and optimized based on the materials and thicknesses shown in Fig. 1. The selection of these thicknesses and materials was primarily designed based on current trends in photonic integrated substrates. If different material substrates and layer thicknesses are used, the neural network can still perform adequately. It simply requires collecting relevant datasets by treating the different material properties and thickness features as variables during the training process. However, as the number of feature parameters increases, the required dataset size will become significantly larger. Finally, to clearly illustrate the performance of the neural network designed in this study, we compared it with models using standalone MLP and CNN networks, as well as similar methods from other studies, as shown in Table 1. It is evident that the use of a multilayer network structure in this study significantly improved the prediction accuracy of the grating coupler across a wide wavelength range.

Table 1. Comparison with other similar neural network schemes for a grating coupler.

Reference	Network architecture	Types of GC	Wavelength (nm)	Prediction error
[31]	MLP	subwavelength GC	1400–1700	6.8%
[32]	CNN	nonuniform GC	1200–1700	$< 1\%$
[33]	MLP	inclined/hybrid GC	532	0.05%
[34]	MLP	chirped GC	1400–1650	$< 1\%$
This paper	MLP	discrete matrix GC	1300–1700	$< 0.44\%$
This paper	CNN	discrete matrix GC	1300–1700	$< 0.49\%$
This paper	CNN+MLP+transformer	discrete matrix GC	1300–1700	0–0.15%

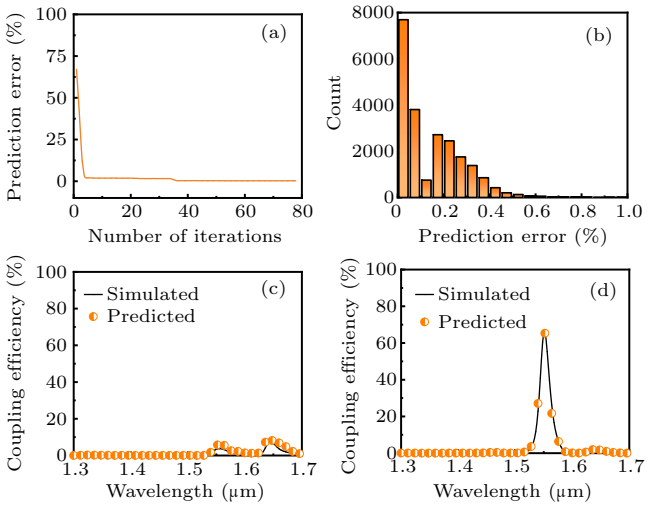


Fig. 4. (a) Convergence curve of coupling efficiency predicted by deep learning network. (b) Statistics of average prediction error of deep learning network on test set. (c) Comparison of the prediction curve and actual simulation curve in the case of lowest prediction accuracy. (d) Comparison of the prediction curve and actual simulation curve in the case of highest prediction accuracy.

The third step involves integrating genetic algorithms with deep learning to optimize the design process. Genetic algorithms are a heuristic search technique primarily used to solve complex optimization problems. Compared to gradient descent and particle swarm optimization algorithms,[35] genetic algorithms exhibit stronger global search capabilities, adaptability and parallelism. In addition, the grating coupler designed in this paper is based on a discrete distribution structure, allowing its representation through the conversion of 0 and 1. This characteristic aligns perfectly with the nature of genetic algorithms. Therefore, we choose to combine genetic algorithms with the designed neural network for subsequent design optimization.

The core mechanism of this algorithm includes operations, such as selection, crossover and mutation, through which iterative searches are performed in the solution space, with the aim of finding optimal or near-optimal solutions. However, these operations are also the main reason the algo-

rithm takes longer to run. By replacing the traditional electromagnetic simulation step with the deep learning model introduced in this article, the time required for optimization can be significantly reduced. The entire optimization process is illustrated in Fig. 5.

(i) According to the above grating coupler equivalent matrix method, 200 300×3 matrices are randomly generated as 200 initial populations.

(ii) A specific figure of merit (FOM) is defined based on the target requirements. Use the trained deep learning model to evaluate the FOM of each population.

(iii) Use the roulette method to select the next generation population based on the proportion of FOM values and determine which populations will be used for crossover or mutation. The specific process is shown in the flowchart of Fig. 5.

(iv) The process is terminated if the maximum number of iterations is reached or if there is no significant improvement in the FOM over 50 consecutive generations.

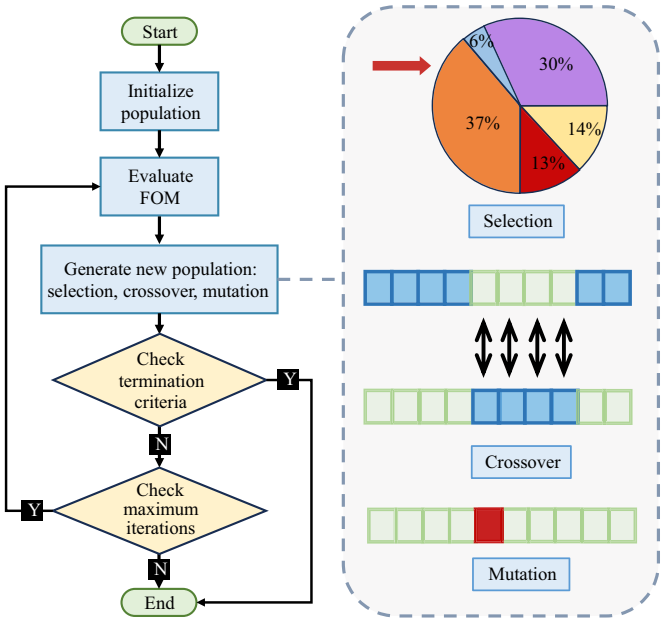


Fig. 5. Flowchart of inverse design combining genetic algorithms and deep learning. Inset illustrates the processes of selection, crossover and mutation.

Through the method of combining genetic algorithms and deep learning, the optimization process can be effectively accelerated while maintaining or improving the quality of the solution. Compared with traditional inverse design algorithms, our method requires only 0.06 s on average to calculate the FOM for each individual. Through testing, it was found that the deep learning model was able, in only 5 min, to identify the optimal solution that met the design objectives. Under the same settings — including identical design targets, number of iterations, and population size — the FDTD-based electromagnetic simulation software required approximately 10 hours to achieve a similar result. Moreover, the optimized results obtained from deep learning were largely consistent with those from the FDTD simulation. This finding demonstrates that the proposed deep learning network possesses both high prediction accuracy and extremely fast optimization speed. Compared to traditional electromagnetic simulation, the optimization process is accelerated by approximately 120 times. It is worth noting that the observed acceleration in the optimization task is significantly lower than the acceleration observed for predicting a single structure. This discrepancy is primarily attributed to hardware limitations, which introduce uncontrollable delays during the iterative operations of the genetic algorithm. In addition, this approach utilizes a discrete grating coupler architecture, allowing the deep learning model to be applied to various types of grating coupler designs, as well as to grating couplers intended for multiple functionalities.

4. Optimization and performance analysis of grating couplers of inverse design

4.1. Design of multi-layer single-wavelength grating couplers

The wavelength grating coupler is the most common type of grating coupler design, primarily aimed at maximizing the coupling efficiency between fibers and photonic integrated platforms at the target wavelengths. Such high coupling efficiency significantly reduces signal loss and enhances the overall transmission performance of the system, which is crucial in fields such as LiDAR and optical communications. For 3D integrated photonic platforms, employing a multilayer grating coupler design is considered the best strategy to improve coupling efficiency. We use an inverse design method to optimize the commonly used operational wavelengths of 1310 nm and 1550 nm. We set the incident angle to 5°, and the objective function is defined as

$$\mathrm{FOM} = T_\lambda, \tag{2}$$

where λ is the target wavelength and T_λ is the coupling efficiency at the target wavelength.

The FOM is imported into the aforementioned inverse design model for optimization. As shown in Fig. 6(a), the grating coupler designed by this model shows relatively smooth convergence in coupling efficiency. After 100 iterations, the coupling efficiencies reach high levels of 82% and 85% at two wavelengths, surpassing the results reported in the vast majority of existing literature, as shown in Figs. 6(b) and 6(c). From the figure, it can be observed that both grating couplers achieve their maximum values at their respective center wavelengths, with 3-dB bandwidths of 30 nm and 32 nm. Due to the specific configuration of the three-layer grating coupler, irregular fluctuation effects occur in regions outside the center wavelength. However, this does not affect the high coupling efficiency near the center wavelength. In addition, compared to traditional methods combining electromagnetic simulation with inverse

design, our model covers more possibilities and can prevent, to the greatest extent, the inverse design from falling into the local optimal solution. At the same time, the speed advantage of the model is also obvious. The design of the grating coupler for two working wavelengths can be completed within 5 min, while the traditional method takes more than 10 hours. In order to further explore the principle of the ultra-high coupling efficiency of the multi-layer grating coupler, we also simulated its cross-sectional electric field distribution, and the results are shown in Fig. 6(d). As shown in Fig. 6(d), the top layer of LNOI significantly enhances the mode matching between the Si_3N_4 layer grating coupler and the single-mode fiber through grating modulation. At the same time, the bottom Si layer effectively reduces the leakage of coupled light to the substrate, thereby significantly improving the overall efficiency of the grating coupler.

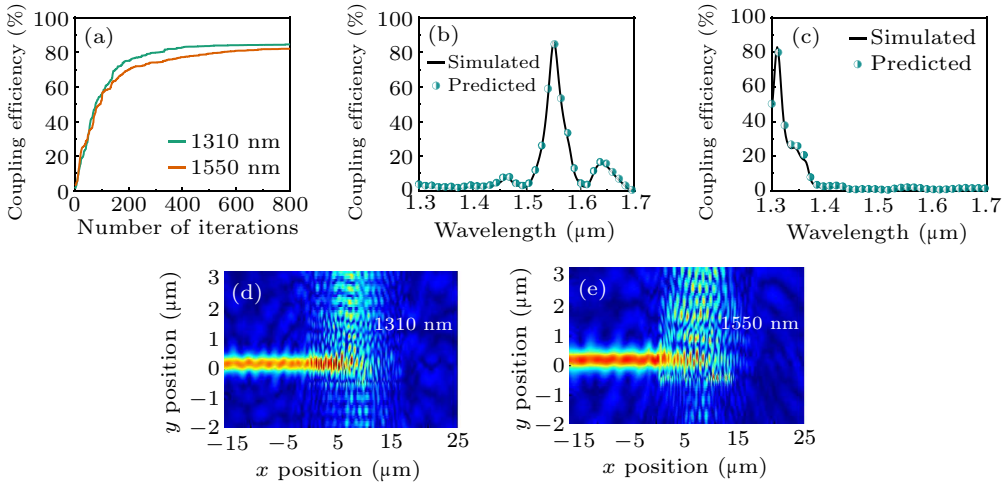


Fig. 6. (a) Coupling efficiency convergence curve of the inverse-designed single-wavelength grating coupler. (b) Predicted and simulated coupling efficiency curves for the grating coupler at a working wavelength of 1310 nm. (c) Predicted and simulated coupling efficiency curves for the grating coupler at a working wavelength of 1550 nm. (d) Cross-sectional electric field distribution of grating couplers at both wavelengths.

4.2. Design of multi-layer dual-wavelength grating couplers

Multi-wavelength grating couplers are pivotal components in photonic integrated circuits (PICs), designed to efficiently couple light between optical fibers and photonic chips over multiple wavelengths. These devices leverage the diffraction properties of grating structures to facilitate the coupling process, which is essential for applications requiring the use of different wavelengths, such as in telecommunications, sensing and biomedical fields. To achieve high coupling efficiency at both 1550 nm and 1310 nm wavelengths simultaneously, the optimization objective function in our design is set as follows:

$$\mathrm{FOM} = \sum_{i=1}^{K} T_{\lambda_i} - \eta \sum_{i=1}^{K} \left| T_{\lambda_i} - \bar{T}_{\lambda_i} \right|, \tag{3}$$

where K represents the number of target working wavelengths, which is set to 2 here, T_{λ_i} represents the coupling efficiency at the ith wavelength, η represents the weight, and $\bar{T}_{\lambda_i}$ represents the average coupling efficiency across multiple wavelengths.

Through multiple experiments, we have determined that setting η to 1 enables our dual-wavelength grating couplers to achieve the highest coupling efficiency, as shown in Figs. 7(a) and 7(b). The efficiencies of this dual-wavelength coupler reached 50% and 60%, respectively. Although these values are lower than the performance of single-wavelength grating couplers, they exceed the results reported in most of the existing literature. Similarly, the design time for the dual-wavelength grating coupler is also within 6 min, further confirming the speed advantage of our design method. In addition, the electric field distributions illustrated clearly demonstrate that at both target wavelengths, the beams can be stably coupled into the waveguide layer and stably transmitted in the fundamental mode.

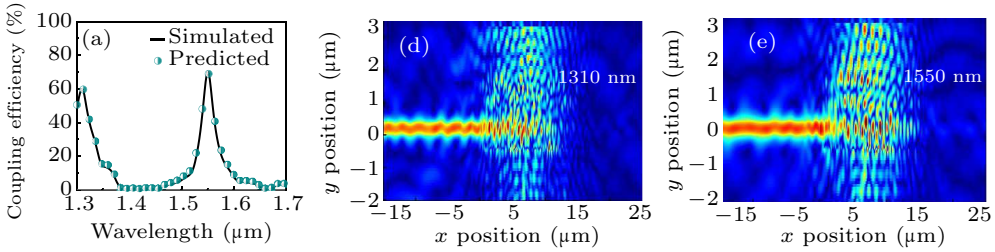


Fig. 7. (a) Predicted and simulated coupling efficiency curves for dual-wavelength grating couplers. (b) Cross-sectional electric field distribution of the grating couplers at the two operating wavelengths.

4.3. Design of multi-layer broad wavelength grating couplers

Broad wavelength grating couplers are specialized optical components used in photonics to couple light efficiently across a wide range of wavelengths. These devices are crucial for applications that require transmitting and processing light over broad spectral ranges, such as in telecommunications, datacom and various sensing technologies. In our design, the corresponding optimization objective function is set as follows:

$$\mathrm{FOM} = \sum_{i=n_1}^{i=n_2} T_{\lambda_i} - \eta \sum_{i=n_1}^{i=n_2} \left| T_{\lambda_i} - \bar{T}_{\lambda_i} \right|, \tag{4}$$

where n_1 and n_2 represent the sequence numbers corresponding to the starting wavelengths of the broad wavelength coupling range. We conducted inverse design for two broadband grating couplers, with their operating wavelength ranges set at 1520 nm–1570 nm and 1400 nm–1600 nm, respectively, with η also set to 0.5. The results, as shown in Figs. 8(a) and (b), indicate that both grating couplers exhibited relatively flat coupling efficiencies within their respective target wavelength ranges, with coupling efficiencies exceeding 68% and 22%, respectively. These values have achieved a high level comparable to existing broad wavelength grating couplers. As shown in Fig. 8, the coupling efficiency of the broad wavelength grating coupler exhibits a noticeable decrease compared to the single-wavelength grating coupler. The primary reason for this decline is that the design was optimized with broadband performance as the main objective. The results demonstrate that this design achieves a flat coupling efficiency curve over a 200 nm bandwidth, which is highly advantageous for broadband photonic devices. For applications where the requirement for a broad wavelength range is relatively lower, the coupling efficiency can be adjusted directly by controlling the value of η. Finally, the designed grating coupler is compared with other grating couplers with similar architectures, as shown in Table 2. The performance of the designed grating coupler is superior to other architectures.

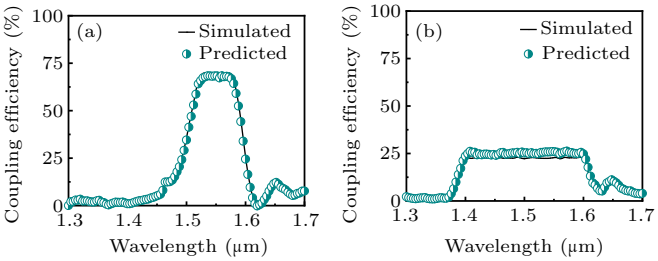


Fig. 8. Predicted and simulated coupling efficiency curves for grating couplers with central wavelength ranges of (a) 1520 nm–1570 nm and (b) 1400 nm–1600 nm.

Table 2. Comparison with other similar neural network schemes for grating coupler comparison of performance with other similar multi-layer grating couplers.

Reference	GC structure	Wavelength (nm)	Coupling efficiency
[18]	Si_3N_4–Si double-layer structure	1550 nm	61%
[19]	height-variable structure	1503 nm–1587 nm	57.5%
[36]	dual-etched structure	1306 nm &1527 nm	54.5%&45.3%
[37]	multilayer topologically optimized structure	1550 nm	50%
[38]	dual-layer Si structure	1530 nm–1565 nm	50%–70%
This paper	LNOI–Si_3N_4–Si multilayer structure	1310 nm/1550 nm	82%/85%
This paper	LNOI–Si_3N_4–Si multilayer structure	1310 nm & 1550 nm	50%&60%
This paper	LNOI–Si_3N_4–Si multilayer structure	1520 nm–1570 nm/1400 nm–1600 nm	68%/22%

4.4. Design of single-layer single-wavelength grating couplers

The results above have demonstrated that the inverse design algorithm combined with deep learning can effectively design multi-layer grating couplers with various functionalities. In addition, we can also design single-layer grating couplers of silicon nitride by using discrete matrices with specific design inputs, such as setting the matrix values of LNOI layers to 0, and setting the matrix values of one of the silicon layers and Si_3N_4 layers to 0. In this design approach, since each layer of the multi-layer grating coupler is fully etched, the designed single-layer grating couplers are also fully etched. According to Eq. (1), we designed a single-wavelength grating coupler for the 1550 nm wavelength, which is presented in Fig. 9(a), with maximum efficiency reaching 34%. Although these efficiency values are significantly decreased compared to multi-layer grating couplers, they are still higher than fully etched grating couplers with equivalent feature sizes. As shown in Fig. 9(b), the significant decrease in coupling efficiency is mainly since most of the light beam is directly incident on the substrate layer and absorbed by it. This design case demonstrates the scalability of our design algorithm. This means that our algorithm is applicable not only to the design of multi-layer grating couplers, but also to the design of single-layer grating couplers. Moreover, although our discussion primarily focuses on a single angle of incidence, the design can be adapted to accommodate multiple angles of incidence based on specific requirements, fully meeting the needs of various application scenarios. This combination of flexibility and efficiency ensures a high degree of applicability and reliability of the design solution in practical applications.

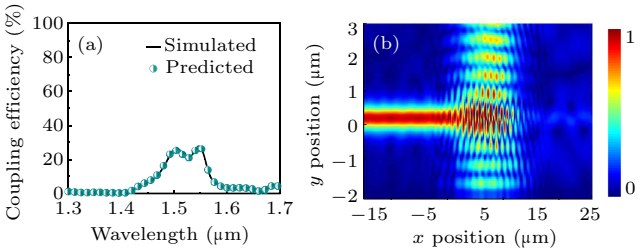


Fig. 9. (a) Predicted and simulated coupling efficiency curves for a single-layer grating coupler. (b) Cross-sectional electric field distribution of the grating coupler at a working wave-length of 1550 nm.

5. Experiments and testing

To validate the performance of the designed grating coupler, a combination of electron beam lithography and plasma etching was employed to fabricate the single-layer silicon nitride grating coupler shown in Fig. 9. For the convenience of subsequent testing, two identical grating couplers were placed in parallel with a spacing of 127 μm. The mode conversion region of the grating coupler was designed with a 300 μm long tapered waveguide to ensure that the insertion loss of the mode converter is negligible. Straight waveguides and Euler bend waveguides with a bending radius of 15 μm were used in the intermediate section, both contributing negligible insertion loss. As shown in Fig. 10(a), optical microscope images and corresponding SEM detailed images of the fabricated grating coupler are provided. It can be observed from the images that, due to the high precision of electron beam lithography, no significant process defects were observed in the grating region, and the grating lines remained intact.

The grating coupler was characterized using a fiber array (FA). During the testing process, the upper port was set as the input port, and the lower port was set as the output port, enabling the measurement of the transmission loss of the two grating couplers. The transmission loss curve of a single grating coupler was then obtained by dividing the measured result by two. Figure 10(b) presents a comparison between the transmission loss predicted by deep learning and the actually measured transmission loss. Due to limitations of the testing fixture, the testing angle of the FA was constrained to 8°, which caused a shift in the transmission loss curve in terms of wavelength. The central wavelength shifted to 1.535 μm, and the corresponding transmission loss was −6.2 dB. It can be seen that while the actual transmission loss curve exhibited an overall decrease compared to the computed results, the overall trend remained consistent. Periodic fluctuations were also observed in the transmission loss curve, primarily due to multiple reflections of light between the optical fiber, the surface of the grating coupler, and the substrate, as well as unavoidable fiber jitter. Overall, the experimental results are in good agreement with the predictions of deep learning, demonstrating the reliability of deep learning in predicting the performance of grating couplers. For multilayer grating couplers, similar experimental validation has not yet been conducted due to fabrication limitations. Future work will explore the verification of multilayer grating couplers through multi-step overlay and bonding techniques.

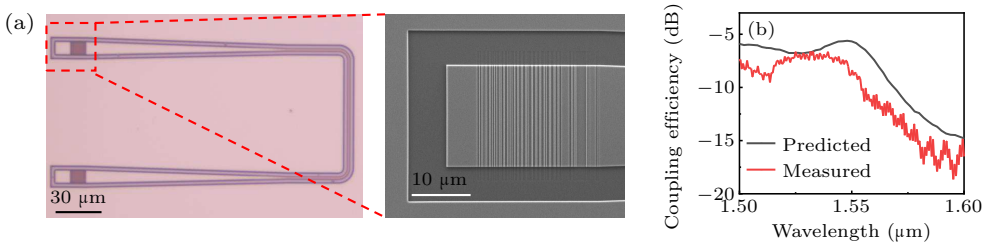


Fig. 10. Experimental test results of a single-layer grating coupler. (a) Experimental fabrication microscope and SEM images. (b) Measured coupling efficiency curve.

6. Conclusion

In this study, a novel deep learning prediction network based on MLP, CNN and transformer models was developed, enabling highly accurate predictions for multi-layer grating couplers. The network demonstrated excellent prediction accuracy, with errors ranging between 0% and 0.15% in most cases, and an overall average error of just 0.17%, with a particularly low error of less than 0.09% at the 1550 nm wavelength. To improve design efficiency, the deep learning model was combined with genetic algorithms to form an efficient inverse design strategy, significantly reducing optimization time, with all simulations completed in less than 6 min. To validate the feasibility of this deep learning-based design method, simulation verifications were conducted for single-wavelength, dual-wavelength and broadband multi-layer grating couplers, as well as single-wavelength single-layer grating couplers, all of which achieved outstanding performance. Furthermore, experimental validation was performed on a single-layer silicon nitride grating coupler using the EBL process, confirming the accuracy of both the deep learning predictions and the design methodology. This study highlights the immense potential of deep learning in grating coupler design and provides valuable insights for applying these methods to silicon photonic device design.

Acknowledgements

This study was sponsored by the National Key Scientific Instrument and Equipment Development Projects of China (Grant No. 62027823) and the National Natural Science Foundation of China (Grant No. 61775048).

References

[1] Blumenthal D J, Heideman R, Geuzebroek D, Leinse A and Roeloffzen C 2018 *Proc. IEEE* **106** 2209
[2] Rahim A, Ryckeboer E, Subramanian A Z, *et al.* 2017 *J. Light. Technol.* **35** 639
[3] Shi Y, Zhang Y, Wan Y, Yu Y, Zhang Y, Hu X, Xiao X, Xu H, Zhang L and Pan B 2022 *Photonics Res.* **10** A106
[4] Siew S Y, Li B, Gao F, *et al.* 2021 *J. Light. Technol.* **39** 4374
[5] Sun C, Yang L, Li B, Shi W, Wang H, Chen Z, Nie X, Deng S, Ding N and Zhang A 2021 *Opt. Lett.* **46** 5699
[6] Xiang C, Jin W and Bowers J E 2022 *Photonics Res.* **10** A82
[7] Boes A, Corcoran B, Chang L, Bowers J and Mitchell A 2018 *Photonics Rev.* **12** 1700256
[8] Churaev M, Wang R N, Riedhauser A, *et al.* 2023 *Nat. Commun.* **14** 3499
[9] Xiang C, Guo J, Jin W, *et al.* 2021 *Nat. Commun.* **12** 6650
[10] Goyvaerts J, Kumari S, Uvin S, Zhang J, Baets R, Gocalinska A, Pelucchi E, Corbett B and Roelkens G 2020 *Opt. Express* **28** 21275
[11] Jacob B, Camarneiro F, Borme J, Bondarchuk O, Nieder J B and Romeira B 2022 *ACS Appl. Electron. Mater.* **4** 3399
[12] Subramanian A Z, Ryckeboer E, Dhakal A, *et al.* 2015 *Photonics Res.* **3** B47
[13] Scheerlinck S, Schrauwen J, Laere F V, Taillaert D, Thourhout D V and Baets R 2007 *Opt. Express* **15** 9625
[14] Zhang J, Yang J, Lu H, Wu W, Huang J and Chang S 2015 *Chin. Opt. Lett.* **13** 091301
[15] Papes M, Cheben P, Benedikovic D, *et al.* 2016 *Opt. Express* **24** 5026
[16] Maire G, Vivien L, Sattler G, *et al.* 2008 *Opt. Express* **16** 328
[17] Wang P, Luo G, Xu Y, *et al.* 2020 *Photonics Res.* **8** 912
[18] Mak J C, Wilmart Q, Olivier S, Menezo S and Poon J K 2018 *Opt. Express* **26** 13656
[19] Yoon J, Kim J Y, Kim J, Yoon H, Neeli B, Park H H and Kurt H 2023 *Photonics Res.* **11** 897
[20] Cheng L, Mao S, Li Z, Han Y and Fu H 2020 *Micromachines* **11** 666
[21] Xie Y, Nie M and Huang S W 2024 *Appl. Phys. Lett.* **124** 051108
[22] Ma W, Liu Z, Kudyshev Z A, Boltasseva A, Cai W and Liu Y 2021 *Nat. Photonics* **15** 77
[23] Mourgias-Alexandris G, Moralis-Pegios M, Tsakyridis A, *et al.* 2022 *Nat. Commun.* **13** 5572
[24] Hong L, Liu Y, Xu M and Deng W 2025 *Chin. Phys. B* **34** 018705
[25] Cui J, Li Y, Zhao C and Zheng W 2023 *Chin. Phys. B* **32** 096101
[26] Tu X, Xie W, Chen Z, Ge M F, Huang T, Song C and Fu H 2021 *J. Light. Technol.* **39** 2790
[27] Adibnia E, Ghadrdan M and Mansouri-Birjandi M A 2025 *Opt. Laser Technol.* **181** 111766
[28] Hooten S, Beausoleil R G and Van Vaerenbergh T 2021 *Nanophotonics* **10** 3843
[29] Mengu D, Sakib Rahman M S, Luo Y, Li J, Kulce O and Ozcan A 2022 *Adv. Opt. Photonics* **14** 209
[30] Li Y, Zhang Y, Wang Y, *et al.* 2024 *Adv. Opt. Mater.* **12** 2302657
[31] Gostimirovic D and Winnie N Y 2018 *IEEE J. Sel. Top. Quantum Electron.* **25** 1
[32] Zeng Z, Huang Q and He S 2023 *Prog. Electromagn. Res. M* **119** 143
[33] Chen X, Lin D, Zhang T, Zhao Y, Liu H, Cui Y, Hou C, He J and Liang S 2023 *Appl. Opt.* **62** 2924
[34] Kang S, Gao F, Yu X, Bo F, Zhang G and Xu J 2023 *JOSA B* **40** D21
[35] Augenstein Y, Repan T and Rockstuhl C 2023 *ACS Photonics* **10** 1547
[36] Cheng L, Mao S, Tu X and Fu H Y 2021 *J. Light. Technol.* **39** 5902
[37] Hammond A M, Slaby J B, Probst M J and Ralph S E 2022 *Opt Express* **30** 31058
[38] Dai M, Ma L, Xu Y, Lu M, Liu X and Chen Y 2015 *Opt Express* **23** 1691