

Artificial Intelligence in Nanophotonics: From Design to Optical Computing

Yongqiang Hu^{1,2}, Jingmin Huang^{1,2}, Junyi Li^{1,2}, Yu Sui^{1,2}, Jiarui Zhang^{1,2}, Tao Zheng^{1,2}, Wenxiao Wang^{1,2}, Chengyi Zhu^{1,2}, Qianyi Guo^{1,2}, Lihua Li^{1,2}, Xilin Zhang³, Nihong Chen³, Juan Chen³, Yitong Chen⁴, Xiaorui Zheng⁵, Zhongmin Yang^{1,2*}, and Jiewei Chen^{1,2*}

¹School of Optoelectronic Science and Engineering, South China Normal University, Guangzhou 510631, China

²Guangdong Provincial Key Laboratory of Nanophotonic Functional Materials and Devices, Guangdong Basic Research Center of Excellence for Structure and Fundamental Interactions of Matter, South China Normal University, Guangzhou 510006, China

³Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, School of Psychology, South China Normal University, Guangzhou 510631, China

⁴Institute of Image Communication and Network Engineering, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

⁵Zhejiang Key Laboratory of 3D Micro/Nano Fabrication and Characterization, School of Engineering, Westlake University, Hangzhou 310030, China

(Received 8 April 2025; accepted manuscript online 30 June 2025)

The bidirectional convergence of artificial intelligence and nanophotonics drives photonic technologies toward unprecedented levels of intelligence and efficiency, fundamentally reshaping their design paradigms and application boundaries. With its powerful data-driven and nonlinear optimization capabilities, artificial intelligence has become a powerful tool for optical design, enabling the inverse design of nanophotonics devices while accelerating the forward computation of electromagnetic responses. Conversely, nanophotonics provides a wave-based computational platform, giving rise to novel optical neural networks that achieve high-speed parallel computing and efficient information processing. This paper reviews the latest progress in the bidirectional field of artificial intelligence and nanophotonics, analyzes the basic principles of various applications from a universal perspective, comprehensively evaluates the advantages and limitations of different research methods, and makes a forward-looking outlook on the bidirectional integration of artificial intelligence and nanophotonics, focusing on analyzing future development trends, potential applications, and challenges. The deep integration of artificial intelligence and nanophotonics is ushering in a new era for photonic technologies, offering unparalleled opportunities for fundamental research and industrial applications.

DOI: 10.1088/0256-307X/42/8/080802

CSTR: 32039.14.0256-307X.42.8.080802

1. Introduction. Since the advent of artificial intelligence (AI), researchers have been committed to giving machines human-like cognitive abilities through AI.^[1,2] Although artificial general intelligence has many capabilities, such as abstract reasoning, decision-making, learning from experience, creativity, and social interaction, replacing human thinking with artificial general intelligence and breaking through human thinking to achieve artificial superintelligence is still very challenging.^[3] In contrast, artificial narrow intelligence is a technical means to perform specific tasks under specific constraints.^[4] It has performed well in autonomous driving, speech recognition, image classification, and recommendation systems and has been widely used. In addition, the deep integration of AI with disciplines such as physics, materials science, and engineering continues to generate innovations, paving the way for its expansion into more emerging fields.^[5–9]

Nanophotonics is a discipline that studies the interaction between light and matter and focuses on explor-

ing and breaking through the optical diffraction limit.^[10] Changing the material's optical properties and achieving unnatural control of light waves is possible by selecting specific materials and designing structures with special functions at the wavelength and subwavelength scales.^[11] Related designs include negative index materials,^[12–14] chiral materials,^[15–17] metalens,^[18–20] metasurfaces,^[21–24] and meta-waveguides,^[25] which respectively control optical signals in on-chip waveguides and free space. Traditionally, the design of nanophotonic materials and devices mainly relies on numerical simulation methods, such as the finite-difference time-domain method,^[26] finite element method,^[27] rigorous coupled wave analysis,^[28] and transfer matrix method.^[29] Although these methods have good performance in theory and engineering applications, they are computationally intensive, time-consuming, heavily dependent on prior expertise; it is often difficult to obtain the global optimal design solution.^[30,31] Fortunately, the introduction of AI provides strong support for design-

*Corresponding authors. Email: yangzm@scut.edu.cn; jwchen@m.scnu.edu.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

ing and simulating nanophotonics.^[32–37] In the forward numerical simulation of nanophotonic materials and devices, AI can be used as a high-speed search tool or as an alternative to physical solvers, breaking away from the traditional complex computational process based on Maxwell's equations and directly establishing an implicit mapping relationship between material or structural parameters and target responses, thereby significantly improving computational efficiency.^[30,38–41] In structural design and optimization, AI can also achieve an efficient inverse design of structural or material parameters based on the required optical response, revolutionizing the traditional design paradigm.^[42,43] This integration of AI and nanophotonics, often classified as “AI for Nanophotonics”, represents a transformative photonic design approach that effectively bridges the gap between computational efficiency and performance optimization.^[30,44]

While AI has powerful computing capabilities, its deployment and operation depend highly on platforms based on silicon-based electronic architectures, such as central processing units, graphics processing units, and field-programmable gate arrays.^[45–50] These traditional architectures rely on charge transmission in metal wires. They are inevitably subject to problems such as ohmic loss and bandwidth limitations caused by the von Neumann

bottleneck.^[51–54] Light has many natural advantages as an information transmission medium. The interaction mechanism between it and nanophotonic structures is similar to data transmission in electronic neural networks to a certain extent.^[55–59] Notably, the inherent parallelism of optical systems, combined with the ultrafast propagation speed of photons ($\sim 10^8$ m/s), enables information processing delays close to zero and low power consumption.^[60,61] These advantages make nanophotonic devices for optical computing a potential alternative to traditional von Neumann electronic architectures, especially in large-scale data transmission and computing tasks.^[58,62,63] In recent years, with the continuous maturity of micro-nano optical processing technology, achieving precise control of structural dimensions and ultra-high resolution patterning is no longer a technical bottleneck, providing a solid foundation for the materialization, miniaturization, efficiency, and functional diversity of optical neural networks (ONNs).^[64–67] This type of research direction that uses nanophotonics technology to address challenges in the field of AI is usually summarized as “Nanophotonics for AI”.^[68–73]

This review delves into the interdisciplinary integration of nanophotonics and AI, exploring how AI intersects with nanophotonics, integrated optics, and guided wave optics. The research focuses on how AI-driven methods

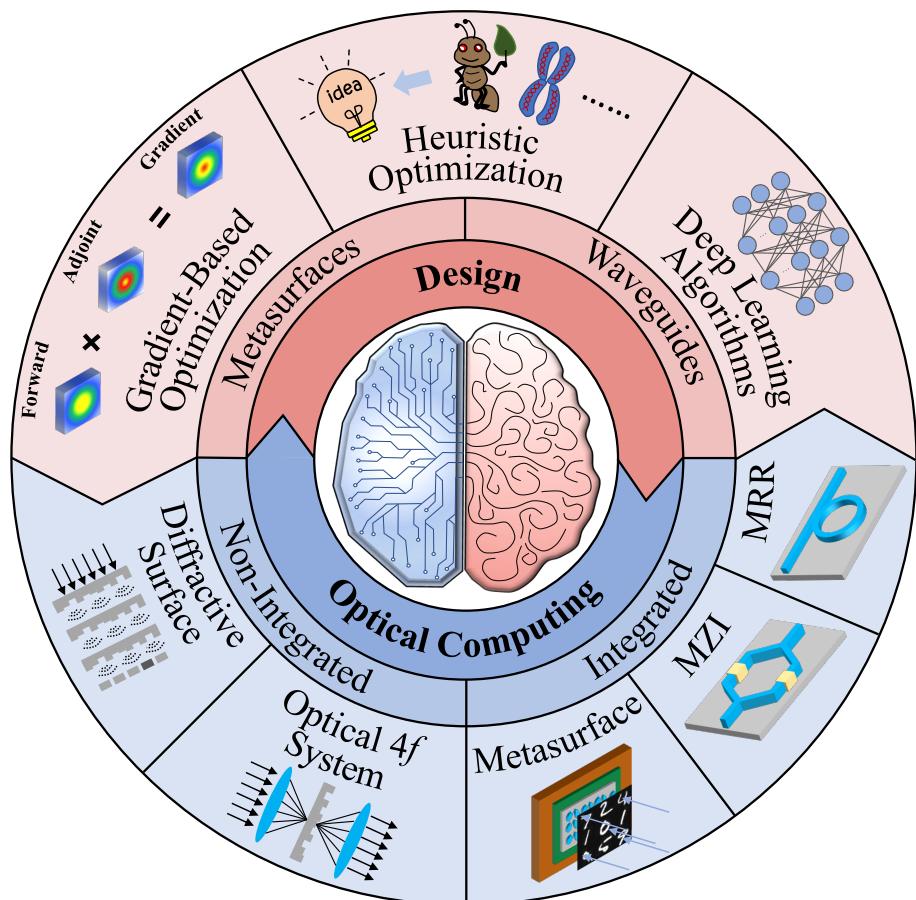


Fig. 1. Overview of the two parts in our work: “AI for Nanophotonics” and “Nanophotonics for AI.”

can revolutionize photonic device design and how nanophotonic platforms can provide new approaches to advanced computing. As shown in Fig. 1, this article consists of two core parts: “AI for Nanophotonics” and “Nanophotonics for AI.” First, the “AI for Nanophotonics” section introduces the basic principles of gradient optimization, heuristic optimization, and deep learning inverse design algorithms. It summarizes their typical research results in the two main fields of meta-waveguides and metasurfaces. These algorithms’ global optimization, speed, and size optimization capabilities are discussed, providing practical solutions to common problems in the current design of nanophotonic devices. Secondly, the “Nanophotonics for AI” section explains the physical mechanisms and latest advances of integrated and non-integrated ONNs. It explores their application potential in fully connected computing, convolution computing, and nonlinear processing. Furthermore, we summarize the emerging challenges of inverse design algorithms including computational efficiency and manufacturability; we also outline future directions for intelligent photonic computing in integrated photonic circuits as well as multimodal optical sensing.

2. AI for Nanophotonics. Nanophotonics has emerged as a leading interdisciplinary field, merging principles from optics, materials science, and nanotechnology, and attracting considerable attention in both scientific research and practical applications in recent years. The field encompasses diverse research areas, including optical imaging,^[12,74–77] display,^[78,79] photonic crystals,^[80] metamaterials,^[11,81–84] meta waveguides, and plasmonic nanostructures.^[85,86] The design of these structures plays a central role in their functionality. Traditionally, photonic design relies on analytical models such as multilayer interference theory and Mie scattering solutions,^[87,88] or brute-force numerical techniques like finite-difference time-domain and finite element methods.^[26,27] These approaches offer precise solutions for well-defined structures but struggle with complex geometries requiring extensive parameter sweeps and iterative refinements. The AI-driven inverse design provides a new path for this paradigm which can efficiently search for the optimal structure in a large-scale parameter space achieving freedom and accuracy far exceeding traditional design methods. For instance, AI-driven inverse design algorithms have been employed in optical communication to optimize devices such as grating couplers and multimode waveguides.^[89–91] In materials science, these algorithms have facilitated innovative designs of plasmonic materials and metasurfaces.^[92–94] Furthermore, within the domains of sensing and photovoltaics AI-based inverse design techniques have optimized key components including plasmon-phonon couplers and solar cells.^[95,96]

The inverse design of nanophotonic devices seeks to identify the global minimum of an objective function, $f(m, g)$, within the material space M and geometric space G , thereby optimizing device performance: $\min f(m, g)$, $m \in M$, $g \in G$, where $f(\cdot)$ represents the objective function related to the figure of merit (FOM) of the device. In recent years, the rapid advancement of computational

power has opened up new avenues for designing complex nanostructures, particularly with AI-driven approaches. Based on their underlying principles, these inverse design algorithms fall broadly into heuristic optimization algorithms, gradient-based optimization algorithms, and deep learning algorithms. The selection of an appropriate inverse design method is critical for specific applications and physical models. This section will systematically elucidate the fundamental principles of various inverse design algorithms and delve into their applications in the design of nanophotonic devices.

2.1. Heuristic Optimization Algorithms. Heuristic algorithms, a class of computational methods guided by empirical rules and heuristic strategies, have emerged as essential tools for addressing complex optimization problems. While they do not inherently guarantee convergence to a global optimum, their ability to efficiently identify near-optimal solutions within constrained computational resources makes them highly valuable across diverse scientific and engineering domains. These algorithms widely assist in optimizing structures with intricate geometries, such as spacecraft antenna design in the early 1990s, where evolutionary algorithms played a pivotal role.^[97–99] Similarly, heuristic approaches have widespread applications in optical beam shaping.^[100,101] Although heuristic algorithms do not necessarily outperform deep learning in the inverse design of nanophotonic devices, their global optimization capability makes them suitable for optimizing large-scale array-based nanophotonic devices.^[102,103] By leveraging appropriate objective functions and constraints, these methods enable structural parameter optimization, enhance design flexibility and, in some instances, reduce fabrication complexity.^[104] Among the various heuristic optimization techniques, genetic algorithms (GAs), simulated annealing, ant colony optimization, and particle swarm optimization (PSO), are among the most widely adopted. The following section will focus on two heuristic algorithms that have demonstrated significant efficacy in nanophotonics device design: GA and PSO.

Inspired by natural selection, GA has emerged as a powerful optimization tool across various scientific disciplines. Pioneeringly introduced by Holland,^[108] GA effectively solves complex design problems by simulating the process of biological evolution. In nanophotonics, the application of GAs is particularly noteworthy. At the heart of GA lies the encoding of design parameters into “chromosomes”, each representing a potential device design. These parameters can be crucial elements constituting micro- and nanostructures, such as the unit cells’ size, position, or material properties. By cleverly defining a fitness function—often directly related to the optical performance of the device, such as spectral response or phase distribution—GA can iteratively improve the design. Figure 2(a) shows that the typical GA process includes the following key steps: First, an initial population containing many random chromosomes is generated. Subsequently, each individual is evaluated based on the fitness function, and individuals with high fitness are selected for reproduction. Crossover and mutation operations simulate gene recombination and

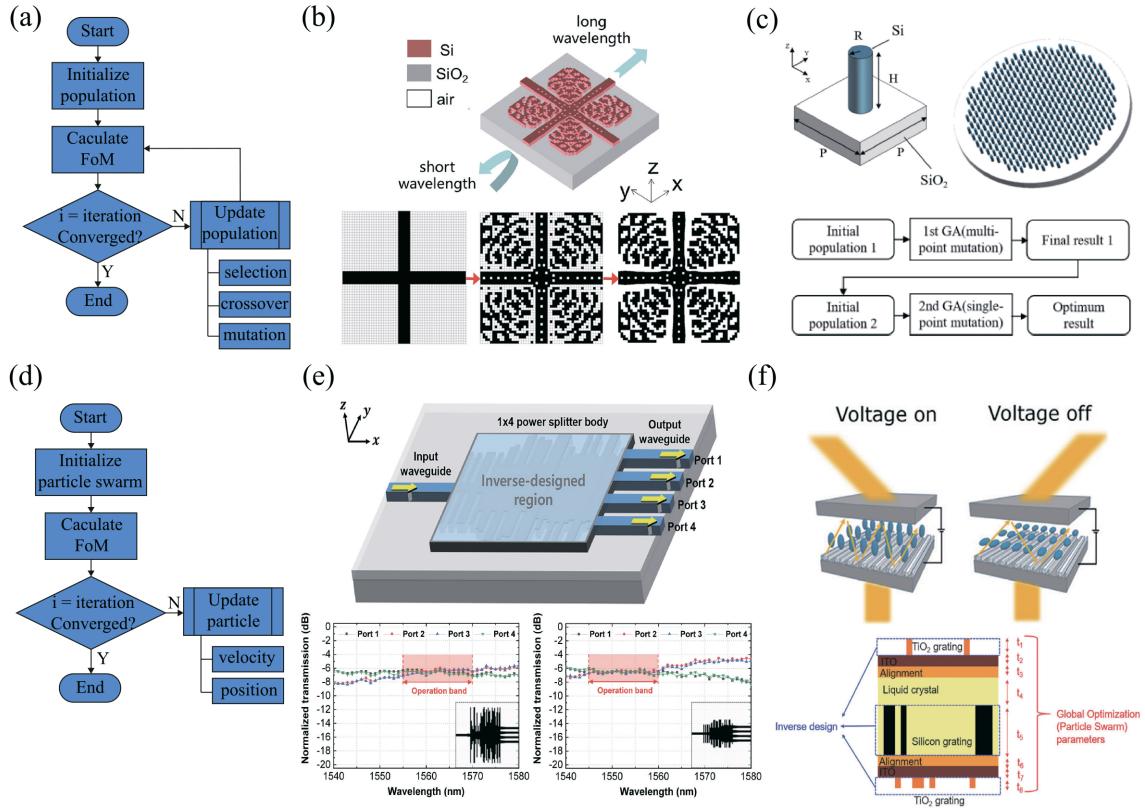


Fig. 2. GA and PSO for the inverse design of nanophotonic devices. (a) The flowchart of GA. (b) Schematic of the metamaterial-based long-pass filter and the design process using GA.^[105] (c) Schematic of the 2D metasurface and the design process using dual-process genetic algorithm.^[106] (d) The flowchart of PSO. (e) Schematic of the 1×4 optical power splitter inversely designed using the PSO algorithm and the corresponding measured transmission spectrum.^[107] (f) Tunable metasurface via inverse design combining the PSO algorithm and the adjoint optimization method.^[103]

mutation, generating new offspring chromosomes. The process repeats until the fitness function converges or reaches a preset optimization target. In this way, GA can effectively explore the design space, overcome the limitations of local optima, and ultimately find design solutions close to global optimum. GA has emerged as a powerful tool for inverse design in nanophotonic devices.

Figure 2(b) illustrates that a GA-driven optimization strategy designs a metamaterial-based integrated photonic filter. In this approach, the optimized binary pixel pattern (black: silicon, white: air) allows for selective transmission of long wavelengths while reflecting shorter ones.^[105] In the domain of metasurfaces, a dual-process genetic algorithm enhances structural optimization for near-infrared metalenses.^[106] Figure 2(c) shows that this method refines the initial structural parameters by applying different mutation operators across two sequential GA stages before further optimizing complex design variables, resulting in more extensive and intricate metalenses with superior optical performance.

As a favored optimization strategy, PSO draws inspiration from collective behaviors in nature and is particularly well-suited for the design of nanophotonic devices. Proposed by Kennedy and Eberhart in the early 1990s, PSO simulates the patterns of bird flocking and fish schooling, providing a unique approach to solving complex optimization problems.^[109] Unlike GAs, the fundamental unit of

PSO is a “particle” rather than a chromosome, as shown in Fig. 2(d). In the design of nanophotonic devices, the position of each particle represents a potential device design scheme, such as a specific combination of parameters for the device structure. The position of a particle, denoted as X_t , evolves dynamically based on its previous position X_{t-1} and velocity V_t , following the update rule,

$$X_t = X_{t-1} + V_t, \quad (1)$$

where the velocity update is given by

$$V_t = wV_{t-1} + c_1r_1(p_{\text{best}} - X_{t-1}) + c_2r_2(g_{\text{best}} - X_{t-1}), \quad (2)$$

where w represents the inertia weight that determines how past velocities influence the current movement, while c_1 and c_2 serve as learning factors to balance exploration and exploitation. The terms r_1 and r_2 are random numbers within the range [0, 1], ensuring stochasticity in the search process. The variable p_{best} records the best-known position of an individual particle, whereas g_{best} tracks the globally optimal solution identified within the swarm. The search mechanism of PSO primarily relies on individual experience and global optima, enabling the identification of optimal solutions within the complex parameter space of nanophotonic device design.^[110] A PSO approach has been applied to design a series of 1×4 optical power splitters, as illustrated in Fig. 2(e). By parameterizing the widths and

lengths of silicon bars in the input, output, and coupling regions, this optimization method iteratively adjusted the device architecture, ultimately yielding a variety of optical power splitters and demonstrating its effectiveness in photonic circuit design.^[107]

Compared to GA, PSO generally exhibits lower computational complexity and faster convergence speed. However, this rapid convergence can also lead to the risk of being trapped in local optima, necessitating careful parameter tuning in practical applications.^[111] To address the limitations of conventional techniques, hybrid optimization strategies that integrate the strengths of multiple methods have gained increasing attention.^[112] One such approach involves a hybrid evolutionary PSO algorithm, which has been applied to designing complex dual-channel multilayer metagratings.^[113] This method creatively replaces the selection operator used in GA with the velocity update mechanism of PSO, effectively mitigating the risk of local optima and significantly improving global search capabilities. Another optimization framework combines PSO with an adjoint method to enhance the performance of a liquid crystal-based electrically tunable beam-switching metasurface [Fig. 2(f)]. In this strategy, PSO is responsible for optimizing the thicknesses of the multilayer grating while the adjoint method fine-tunes the surface structural parameters, improving controllability and tuning precision.^[103]

2.2. Gradient-Based Optimization Algorithms.

Gradient-based optimization algorithms solve optimization problems by leveraging the gradient of the objective function. While heuristic algorithms perform well for relatively small design spaces, their efficiency diminishes as the problem complexity and parameter dimensionality increase.^[114] Unlike heuristic approaches, gradient-based methods assume the objective function is continuously differentiable, significantly reducing computational cost. In addition, when the design parameters vary in a continuous range, such as width, thickness, refractive index, and when the gradient can be calculated, the gradient-based topology optimization method can achieve a fast and stable search process and have mathematically verifiable convergence near the local minimum of the objective function.^[115] For those problems that are essentially non-differentiable, such as when discrete selection is required from multiple candidate materials with different characteristics, heuristic optimization methods are more suitable as the preferred solution.^[116] Despite its computational efficiency compared to heuristic methods, traditional topology optimization still faces significant computational overheads mainly due to iterative gradient evaluations required in methods like finite-difference-based optimization.

Adjoint method has become a powerful gradient-driven approach in the inverse design of nanophotonic devices.^[121,122] The mathematical foundation of this approach stems from Pontryagin's maximum principle in control theory, and it has achieved remarkable success in fields such as computational fluid dynamics. The application of adjoint methods to electromagnetic design optimization was first introduced in 2013, revolutionizing the rapid design of silicon photonic Y-splitters. This approach demonstrated exceptional convergence speed, even within large

design spaces.^[31] Figure 3(a) illustrates that the adjoint method typically involves one forward and one adjoint simulation. By calculating the interaction between the forward and adjoint fields within the nanophotonic device, gradient information concerning the design parameters is obtained, which is then used for iterative optimization until convergence criteria are met. For instance, in the design of a silicon photonic device as shown in Fig. 3(b), if the objective is to maximize or minimize the electric field at a specific location x_0 , and the optimization region Ω is a rectangular area composed of $M \times N$ pixels each filled with a different dielectric constant ϵ , then the FOM is

$$\text{FOM} = |\mathbf{E}(x_0)|^2. \quad (3)$$

Within the optimization region Ω , a dielectric perturbation $\Delta\epsilon_r$ within a volume element at position x induces a corresponding field perturbation $\Delta\mathbf{E}(x_0)$ at the target location x_0 . Consequently, the variation in the FOM can be expressed as follows:

$$\Delta\text{FOM} = \text{Re}[\overline{\mathbf{E}^{\text{old}}(x_0)} \cdot \Delta\mathbf{E}(x_0)], \quad (4)$$

where $\mathbf{E}^{\text{old}}(x_0)$ is the value of the original electric field, the electric dipole moment induced by a dielectric perturbation $\Delta\epsilon_r$ within a volume element ΔV at position x is given by

$$\mathbf{p} = \epsilon_0 \Delta\epsilon_r(x) \Delta V \mathbf{E}^{\text{old}}(x). \quad (5)$$

A key consideration is optimizing $\Delta\mathbf{E}(x_0)$ to obtain a better FOM, where $\Delta\mathbf{E}(x_0)$ can be expressed as

$$\begin{aligned} \Delta\mathbf{E}(x_0) &= \overline{\overline{G^{\text{EP}}}}(x_0, x) \mathbf{p} \\ &= \epsilon_0 \Delta\epsilon_r(x) \Delta V \overline{\overline{G^{\text{EP}}}}(x_0, x) \mathbf{E}^{\text{old}}(x), \end{aligned} \quad (6)$$

here, $\overline{\overline{G^{\text{EP}}}}(x_0, x)$ represents Green's function, which characterizes the influence of the electric dipole moment at x on the electric field at x_0 . Consequently, the gradient of the FOM function concerning the optimization parameter $\epsilon_r(x)$ can be derived as

$$\frac{\partial \text{FOM}}{\partial \epsilon_r(x)} = \epsilon_0 \Delta V \text{Re}[\overline{\mathbf{E}^{\text{old}}(x_0)} \cdot \overline{\overline{G^{\text{EP}}}}(x_0, x) \mathbf{E}^{\text{old}}(x)], \quad (7)$$

according to the reciprocity theorem, $G(x_0, x) = G(x, x_0)$, thus Eq. (7) can be rewritten as

$$\begin{aligned} \frac{\partial \text{FOM}}{\partial \epsilon_r(x)} &= \text{Re}[\mathbf{E}^{\text{adj}}(x) \cdot \mathbf{E}^{\text{old}}(x)], \\ \mathbf{E}^{\text{adj}}(x) &= \epsilon_0 \Delta V \overline{\mathbf{E}^{\text{old}}(x_0)} \cdot \overline{\overline{G^{\text{EP}}}}(x_0, x), \end{aligned} \quad (8)$$

where $\mathbf{E}^{\text{adj}}(x)$ represents the adjoint field induced by an electric dipole with amplitude $\epsilon_0 \Delta V \overline{\mathbf{E}^{\text{old}}(x_0)}$ at position x_0 [as illustrated in Fig. 3(b)]; thus, a single forward and adjoint simulation can efficiently obtain shape derivatives across the entire design space with multiple degrees of freedom. The optimization iteratively converges toward an optimal design by computing the FOM gradient and applying geometry modifications proportional to this gradient.^[123]

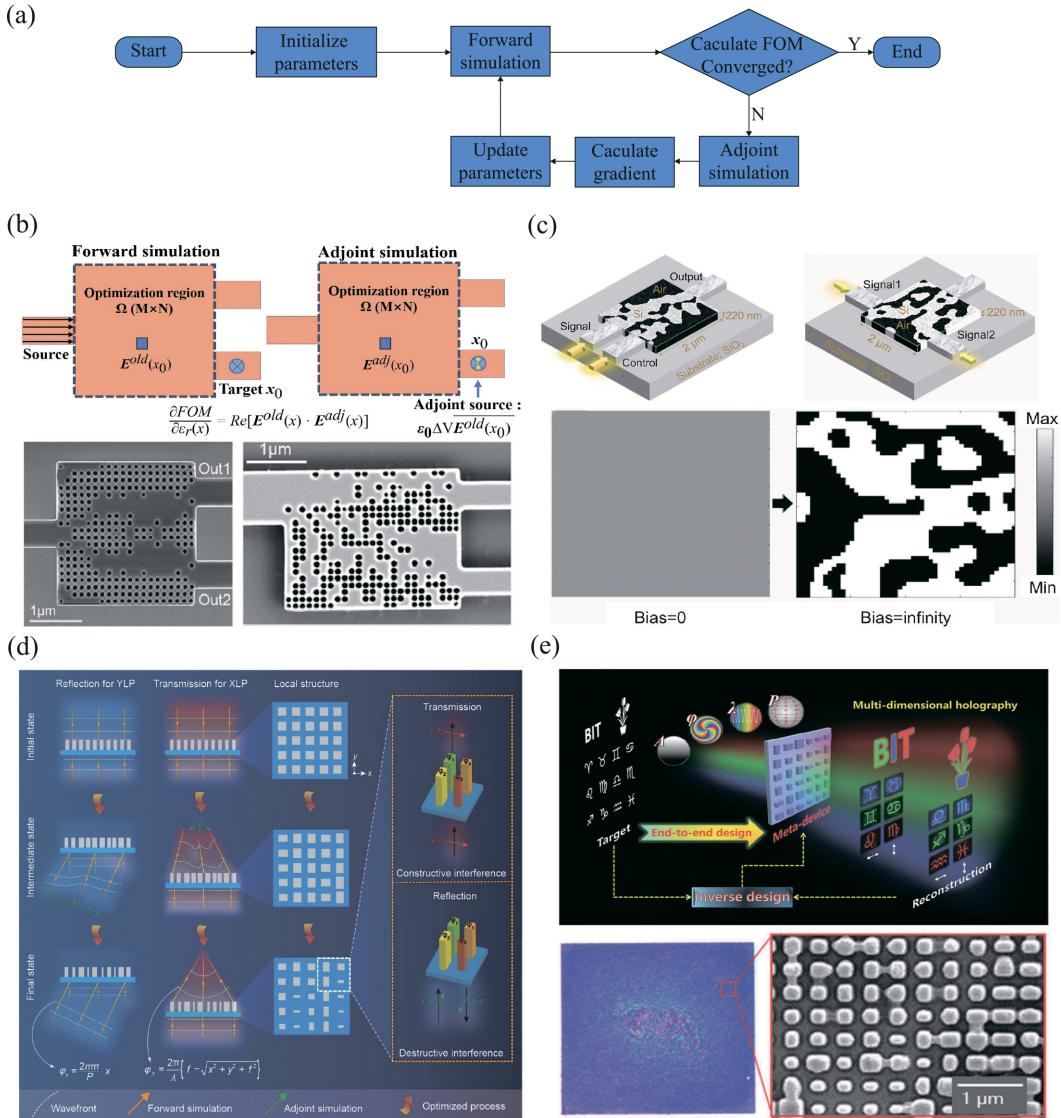


Fig. 3. Adjoint method for nanophotonic device optimization. (a) Adjoint method flowchart. (b) The divider and mode demultiplexer are optimized via the adjoint method, with the orange region representing the schematic of the adjoint field that computes gradient. [117] (c) Design of integrated photonic circuit unit via the adjoint method. [118] (d) Schematic of the full-space metalens optimization via the adjoint method, optimizing the reflected wavefront in the y -polarization channel and the transmitted wavefront in the x -polarization channel. [119] (e) The end-to-end inverse design framework is based on the adjoint method for multidimensional metasurface holography and includes a scanning electron microscope image of a sample. [120]

In the field of silicon photonics, adjoint optimization has been effectively utilized to achieve the rapid inverse design of a single-mode 3 dB power splitter and a dual-mode demultiplexer, as illustrated in Fig. 3(b). These devices feature remarkably compact footprints of $2.6 \mu\text{m} \times 2.6 \mu\text{m}$ and $2.6 \mu\text{m} \times 3 \mu\text{m}$, respectively. [117] This optimization strategy has shown a fivefold increase in speed compared to conventional deep brain stimulation algorithms. Building on this advancement, an improved adjoint method was later introduced and experimentally validated, facilitating the development of ultrafast, ultra-compact, and ultra-low-power photonic integrated circuits on a Silicon-on-Insulator platform. [118] Figure 3(c) shows that they first designed an all-optical XOR logic gate and two optical switches separately. Then they integrated them into an all-optical computing circuit capable of recognizing two-

bit logic signals.

Beyond silicon photonics, adjoint optimization has revolutionized metasurface design. [124–129] Conventional approaches rely on local periodic approximation (LPA) conditions, optimizing unit-cell responses before arranging them according to predefined rules. However, metasurfaces often deviate from LPA at high phase gradients, reducing efficiency and increasing design complexity. An adjoint-optimization-based inverse design methodology was introduced to address this limitation, allowing independent control over full-space wavefronts, both transmitted and reflected, in metasurfaces. [119] This approach leverages polarization-selective local interference between adjacent meta-atoms in a monolayer dielectric metasurface, effectively suppressing aperiodic electromagnetic crosstalk at large phase gradients. Figure 3(d) shows that this

strategy provides a new avenue for optimizing high-performance, multifunctional metasurfaces. Meanwhile, researchers developed an end-to-end adjoint-based design framework to realize a multidimensional holographic metasurface, as illustrated in Fig. 3(e). This approach enabled 12-channel holographic multiplexing across polarization, wavelength, and image planes.^[120] This method eliminates constraints imposed by LPA, establishing a new paradigm for metasurface design.

2.3. Deep Learning Algorithms. The origins of deep learning can be traced back to the 1940s when multilayer perceptrons were first designed to mimic human cognitive processes.^[130] Over the past decades, the rapid evolution of deep learning has revolutionized numerous scientific domains, including autonomous driving, speech recognition, and medical diagnostics.^[131] This unprecedented progress is due to the significant improvement in hardware computing power, big data technology, and continuous breakthroughs in optimization algorithms, which have gradually made it the core driving force of AI technology and demonstrated strong data modeling and feature extraction capabilities in multiple cutting-edge disciplines. Deep learning is a type of data-driven algorithm. Its core principle is to perform layer-by-layer feature extraction, nonlinear transformation, and optimization solutions on input data through multilayer neural networks. In this process, there is no need to rely on the explicit expression of traditional physical models. Instead, through large-scale calculations and iterations, complex high-dimensional mapping relationships are approximated in a “brute force” manner, achieving efficient data processing and decision-making capabilities. In the design of nanophotonic devices, deep learning makes designing and optimizing complex optical devices possible with its efficient computing power. Its core goal is to build a bidirectional mapping relationship

between two physical spaces.^[132–134] As shown in Fig. 4, the first space, namely g-space, represents the working conditions including the incident field of the device and structural parameters used to describe it; while the second space, namely s-space, represents the electromagnetic response or output field of the device under corresponding g-space conditions. Deep learning methods can be used to efficiently predict s-space responses when g-space is known by replacing traditional forward numerical simulation; they can also inversely solve for required g-space parameters when target responses from s-space are known to achieve rapid design optimization. In this section, we review main research progress using deep learning technology for solving inverse design problems related to nanophotonic devices. As shown in Fig. 5, these methods can be roughly summarized into three models: tandem models, generative adversarial models, and hybrid models reflecting strategies and advantages of deep learning across different design scenarios.

As mentioned above, deep learning is a powerful data-driven approach for forward prediction and inverse design of nanophotonics devices. In forward modeling, known geometric and material parameters are fed into a pre-trained neural network to obtain the corresponding electromagnetic response directly.^[137] This process is straightforward as it follows a one-to-one mapping. However, in inverse design, the problem becomes one-to-many, meaning that multiple structural configurations can produce identical optical responses. This inherent ill-posedness significantly hampers the convergence and stability of neural networks in inverse design. To address this non-uniqueness issue, the concept of a tandem neural network was employed, as shown in Fig. 5(a).^[138,139] This approach first trains a forward network to predict optical responses accurately. Once trained, the forward network is fixed, and its input

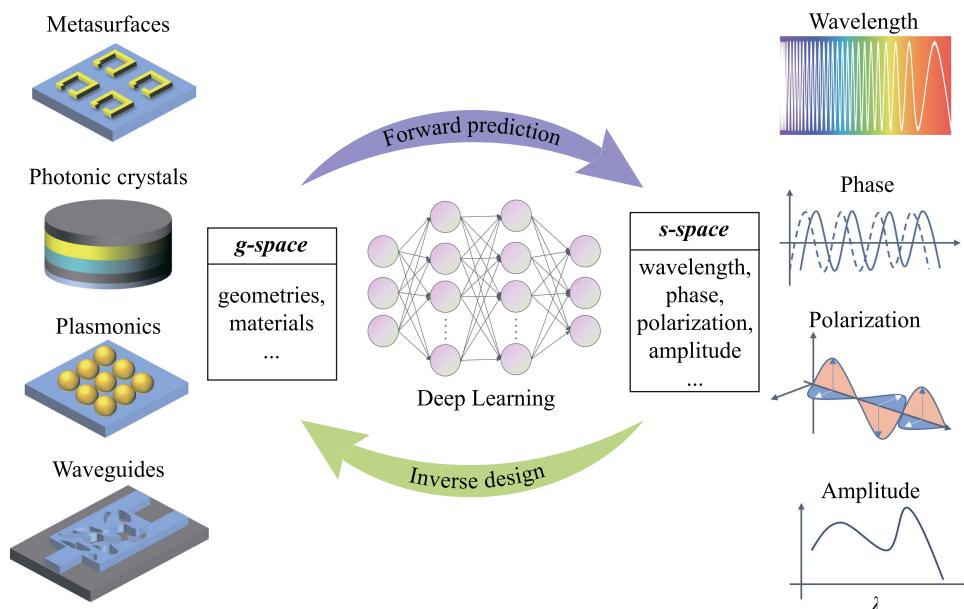


Fig. 4. Overview of deep learning algorithms for nanophotonics design, where deep learning algorithms enable the mutual mapping between g-space and s-space.

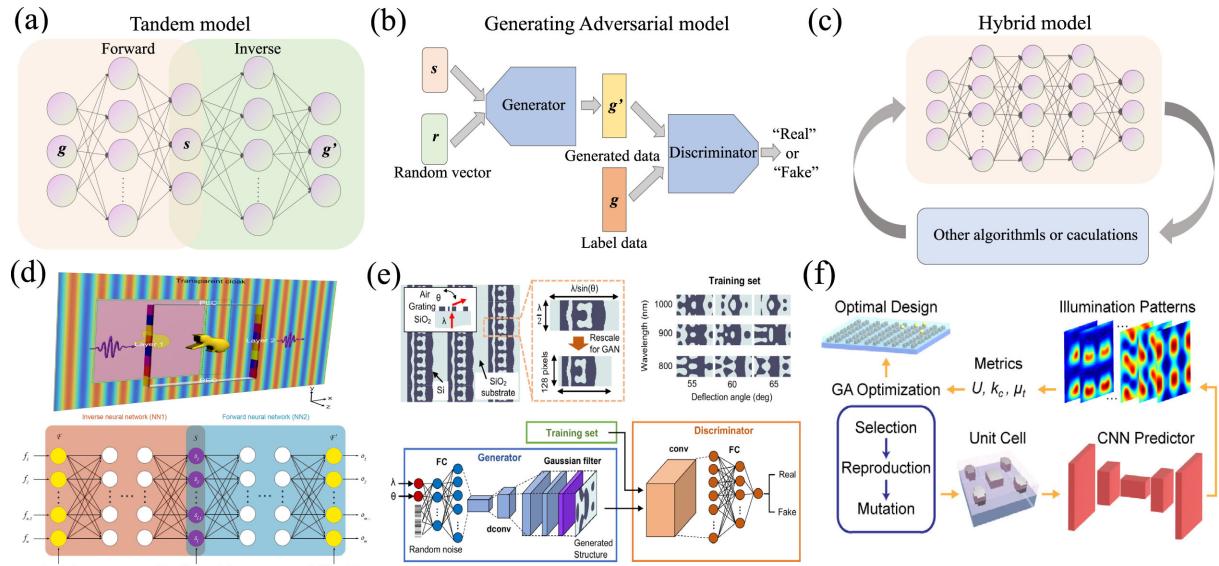


Fig. 5. Three mainstream deep learning models for nanophotonics device design. (a) Schematics of the Tandem model. (b) Schematics of the generative adversarial model. (c) Schematics of the Hybrid model. (d) Design of a transmitted metasurface cloak via the Tandem model.^[132] (e) Design of a free-form diffractive metagrating via the generative adversarial model.^[135] (f) Design of a nanoantenna array for localized plasmonic structured illumination microscopy via the Hybrid model.^[136]

is directly connected to the output of the inverse network, guiding the optimization path toward physically meaningful solutions. By incorporating learned priors from the forward model, this strategy significantly improves both the stability and accuracy of inverse design. The introduction of the tandem model has provided a crucial foundation for deep learning-driven optical device design and has rapidly expanded to various domains. For instance, figure 5(d) illustrates that a transmitted metasurface cloak was designed using a tandem neural network.^[132] Based on this approach, a dual-channel tandem model was later introduced to achieve polarization-multiplexed structural color coding in metasurfaces. By integrating polarization information into the network architecture, this approach enhances the design flexibility of metamaterials and demonstrates the potential of deep learning in designing multi-channel nanophotonic devices.^[38]

Generative adversarial networks are a type of deep learning model commonly used in the inverse design of nanophotonics devices. The core idea is to simulate the data distribution in the training set through adversarial training between the generator and the discriminator.^[140–144] As shown in Fig. 5(b), the generator receives a random variable r as input and tries to generate structural data g' similar to the distribution of real samples g in the training set; the discriminator is responsible for judging whether the input data is a real sample or a generated sample.^[135] When the structure generated by the generator is very close to that of a real sample, the discriminator is more likely to judge it as real data. Otherwise, it is regarded as forged, thereby jointly improving model performance in this continuous game. Generative adversarial networks initially found application in fields such as image style transfer and image synthesis.^[145] Due to their powerful image generation ability, they have grad-

ually been introduced into designing micro-nano photonic structures to address complex structure design problems and irregular distributions. This method can directly use two-dimensional images of photonic structures as generation targets to achieve efficient inverse design. For instance, as Fig. 5(e) depicts, a conditional generative adversarial network was trained to design a two-dimensional metasurface beam deflector. The network can generate a metasurface two-dimensional pattern with high deflection efficiency according to input working wavelength and deflection angle.^[135] In addition, thanks to the high sensitivity of the discriminator during authenticity determination processes, generative adversarial networks have an ability to alleviate non-uniqueness problems in inverse design to some extent, further improving their applicability and effectiveness in actual design tasks.^[146]

Despite their advantages in computational efficiency and high-dimensional nonlinear fitting, deep learning-based inverse design methods remain fundamentally data-driven. Before training, large-scale datasets must be obtained through simulations or experiments, which is a time-intensive process. Furthermore, most existing studies rely on the assumption of LPA, where deep learning is employed primarily to optimize unit cell structures. However, this assumption overlooks inter-unit coupling effects, posing limitations on applying deep learning in large-scale complex photonic device design. To address these challenges, integrating traditional heuristic optimization algorithms with advanced deep learning approaches has emerged as a promising solution. This hybrid strategy not only leverages the strengths of both methods, reducing computational demands but also enables multi-solution search mechanisms, mitigating the one-to-many mapping issue inherent in inverse problems.^[147–149] Deep learning is often a high-speed surrogate for electromagnetic numerical solvers within such hybrid optimization frameworks.

As illustrated in Figs. 5(c) and 5(f), a trained forward neural network was combined with a GA, enabling the rapid inverse design of plasmonic nanoantenna arrays for localized plasmonic structured illumination microscopy.^[136] Moreover, integrating physics-informed priors or gradient-based optimization techniques can alleviate conventional deep learning approaches' computational and geometric constraints.^[150–154] For instance, multi-objective adjoint optimization combined with deep learning to optimize the design of large-aperture metalenses.^[155] This approach accelerated the design process and effectively tackled the challenges associated with large-aperture metasurface engineering.

3. Nanophotonics for AI. AI algorithms are reshaping the design of photonic devices and driving innovations in AI computing architectures. Currently, the implementation and deployment of AI heavily rely on large-scale datasets and electronic computing platforms based on the Von Neumann architecture. Although the rapid advancement of semiconductor technology over the past two decades has fueled the progress of deep learning, the limitations of the Von Neumann architecture in terms of parallel computing efficiency and energy consumption have become major bottlenecks for scaling AI to more complex and real-time applications. Since matrix operations inherent in ordinary digital logic computation as well as neural network computation are analogous to the interaction of light with matter and the propagation properties of light, it is possible to realize an ultra-low-power, ultra-high-speed, and ultra-wide-bandwidth computing and transmission platform.^[69,156–159] ONNs, a novel computational paradigm utilizing photonic devices as computational platforms, harnessing light propagation to transfer information and perform computations.^[160–164] ONNs are widely regarded as a potential solution to computational bottlenecks in future development of AI, with their advantages stemming not only from low latency, low power consumption, large bandwidth, and parallel processing inherent in light propagation but also from nanophotonics devices' ability to regulate phenomena such as diffraction, scattering, and interference that effectively mimic matrix operations and nonlinear activations in artificial neural networks.^[156,165,166] Thus, ONNs offer significant computational potential capable of overcoming efficiency limitations faced by traditional computing architectures.

In neural networks, whether it is convolution calculation or fully connected calculation, its essence is to realize the connection between neurons through matrix-vector multiplication (MVM). ONNs also follow this mathematical principle, performing MVM through optical means to complete the calculation process of neurons in each layer.^[156,167] Its mathematical expression can be $\mathbf{Y} = \mathbf{A} \cdot \mathbf{X}$, where \mathbf{X} is the input vector, \mathbf{A} is the weight matrix, and \mathbf{Y} is the output vector. At present, the mainstream implementation schemes of optical MVM mainly fall into three categories: matrix calculation based on plane light conversion (PLC), matrix calculation based on on-chip Mach-Zehnder interferometer (MZI), and matrix calculation based on wavelength-division multiplexing (WDM). These three types of schemes are manifested in two forms

at the device level: integrable and non-integrable, providing multiple feasible paths for developing optical computing. In this section, we focus on "Nanophotonics for AI", exploring and presenting the current leading ONN architectures from the perspectives of integrated and non-integrated physical structures.

3.1. Non-Integrated ONNs.

3.1.1. Diffractive Surfaces. ONNs using diffractive surfaces essentially achieve optical computing by manipulating light fields, that is, completing matrix multiplication based on the form of PLC.^[73] This framework is characterized by the modulation of various degrees of freedom of light at dimensions on the order of the wavelength ($\geq \lambda/2$). Each diffractive surface acts as a two-dimensional pixelated structure, where individual elements function as artificial neurons. These neurons interconnect through the Rayleigh-Sommerfeld diffraction mechanism, which is equivalent to fully connected networks in conventional electronic neural systems, as shown in Fig. 6(a), and this process facilitates the efficient transfer of optical information.^[164] The modulation function $t(x, y)$ associated with each pixel on the diffractive surface can be analogously regarded as a weight term in artificial neural networks, serving as a learnable parameter that is iteratively refined during training. Consequently, when the input light field $U_{\text{in}}(x, y)$ is introduced, the output light field $U_{\text{out}}(x, y)$ can be expressed as $U_{\text{out}}(x, y) = t(x, y)U_{\text{in}}(x, y)$.

During the training process of diffractive ONNs, light modulation is primarily achieved by adjusting the phase through individual diffractive units. Each unit's structural characteristics (height h) or material properties (refractive index n_{medium}) induce a corresponding phase shift $\Delta\varphi = \frac{2\pi h}{\lambda}(n_{\text{medium}} - n)$, where n represents the refractive index of air, conventionally set to 1. When the refractive index n_{me} of the diffractive surface is fixed, network training can be accomplished by optimizing the height h of each diffractive unit.^[164,168,169] This approach is commonly employed, and the corresponding modulation function can be expressed as $t(x, y) = \exp[\frac{j2\pi h(x, y)}{\lambda}(n_{\text{medium}} - n)]$. Alternatively, the height h can be held constant while phase optimization can be achieved by varying the refractive index of each unit.^[170,171] In this case, the modulation function is given as $t(x, y) = \exp\{\frac{j2\pi h}{\lambda}[n_{\text{medium}}(x, y) - n]\}$.

In 2018, researchers pioneered a diffractive surface-based ONN capable of classifying and imaging handwritten digits (MNIST) and fashion products (Fashion-MNIST) in the terahertz band (0.4 THz), as illustrated in Fig. 6(b).^[164] During training, input data was processed via optical diffraction calculations, with the network architecture and neuron coefficients optimized based on output errors. Post-training, the Diffractive Deep Neural Network (D²NN) design finalizes, enabling fabrication through 3D printing or photolithography. The diffractive layers, composed of VeroBlackPlus RGD 875, achieved a classification accuracy of 93.39% with seven layers. In 2022, this team innovatively proposed a polarization-multiplexed all-optical diffraction computational architecture.^[172] By embedding linear polarizer arrays between the P trainable

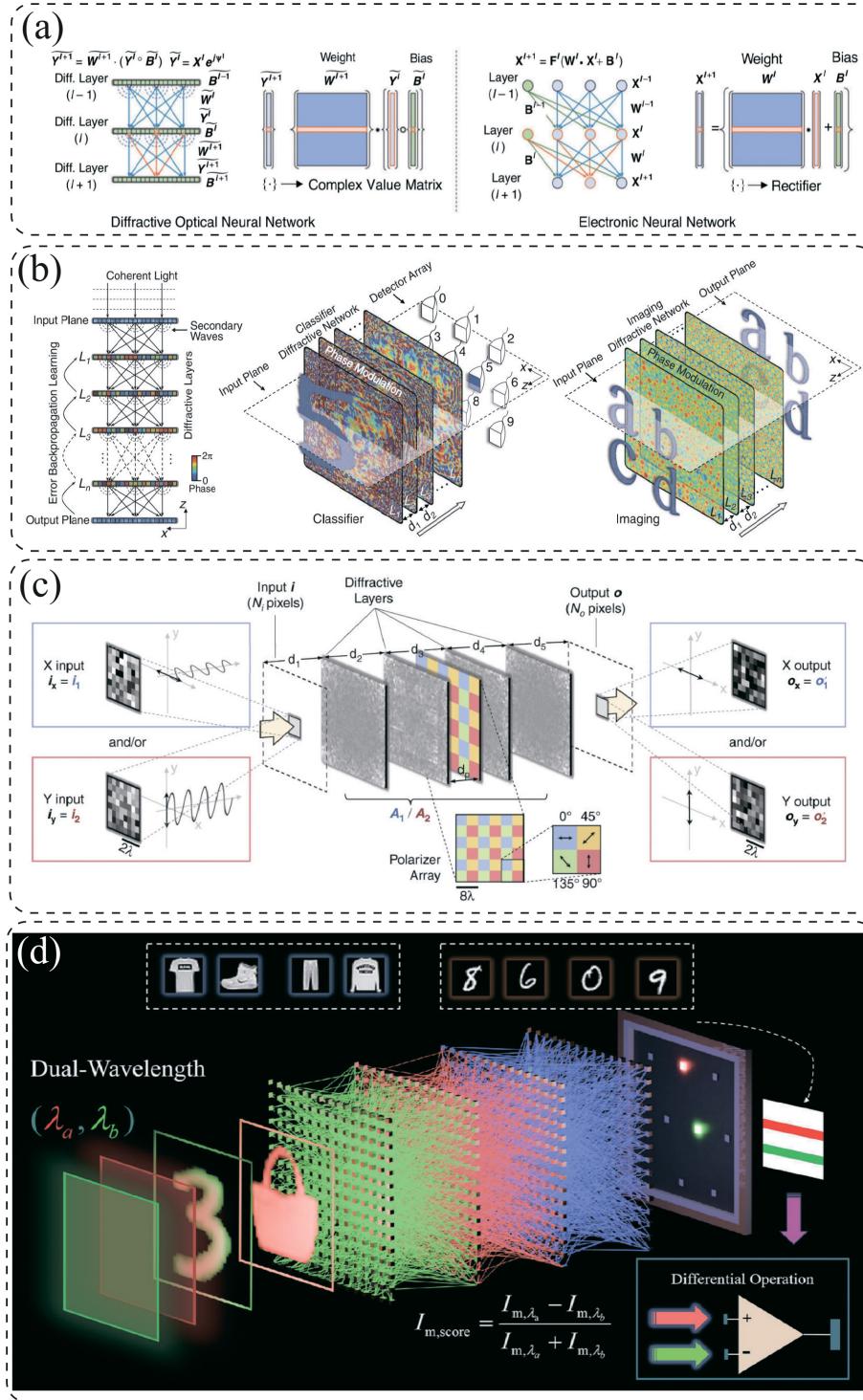


Fig. 6. Non-integrated optical neural network based on diffractive surfaces. (a) Comparison of the computational principles between diffractive optical neural networks and traditional electronic neural networks.^[164] (b) Schematic of the physical process of D²NN and its implementation for classification and imaging.^[164] (c) Schematic of polarization-multiplexed all-optical diffractive computing.^[172] (d) Schematic of DW-D²NN for optical target recognition.^[173]

isotropic diffraction layers [Fig. 6(b)], different complex linear transformations are mapped to specific combinations of polarization states, and parallel optical calculations of multiple sets of transformations are realized. Compared to traditional scalar optical field designs, this work exploits the vector properties of light while maintaining isotropy in the diffraction layers and achieving

two operational modes, sequential (SeqPA) and simultaneous (SimPA), through “polarization seeds”. This breakthrough provides a new perspective for the development of multifunctional optical processors. In 2025, some researchers designed a dual-wavelength differential diffractive neural network (DW-D²NN) at the wavelength dimension level.^[173] As shown in Fig. 6(d), the network handles

different classification tasks at different operating wavelengths by utilizing the complementary optical responses of two wavelengths (532 nm and 640 nm). The proposed structure achieves classification accuracies of 98.7% on MNIST and 90.1% on Fashion-MNIST datasets respectively; these accuracies are higher than those obtained using traditional single-wavelength methods. These works

show that diffractive ONNs have great potential for parallel computing. In addition, many research efforts indicate that diffractive ONNs are also developing towards integration and portability.^[168,169,174–176]

3.1.2. Optical 4f System. ONNs based on the 4f system are a unique type of non-integrated optical neural computing architecture.^[177–184] Unlike ONNs based on

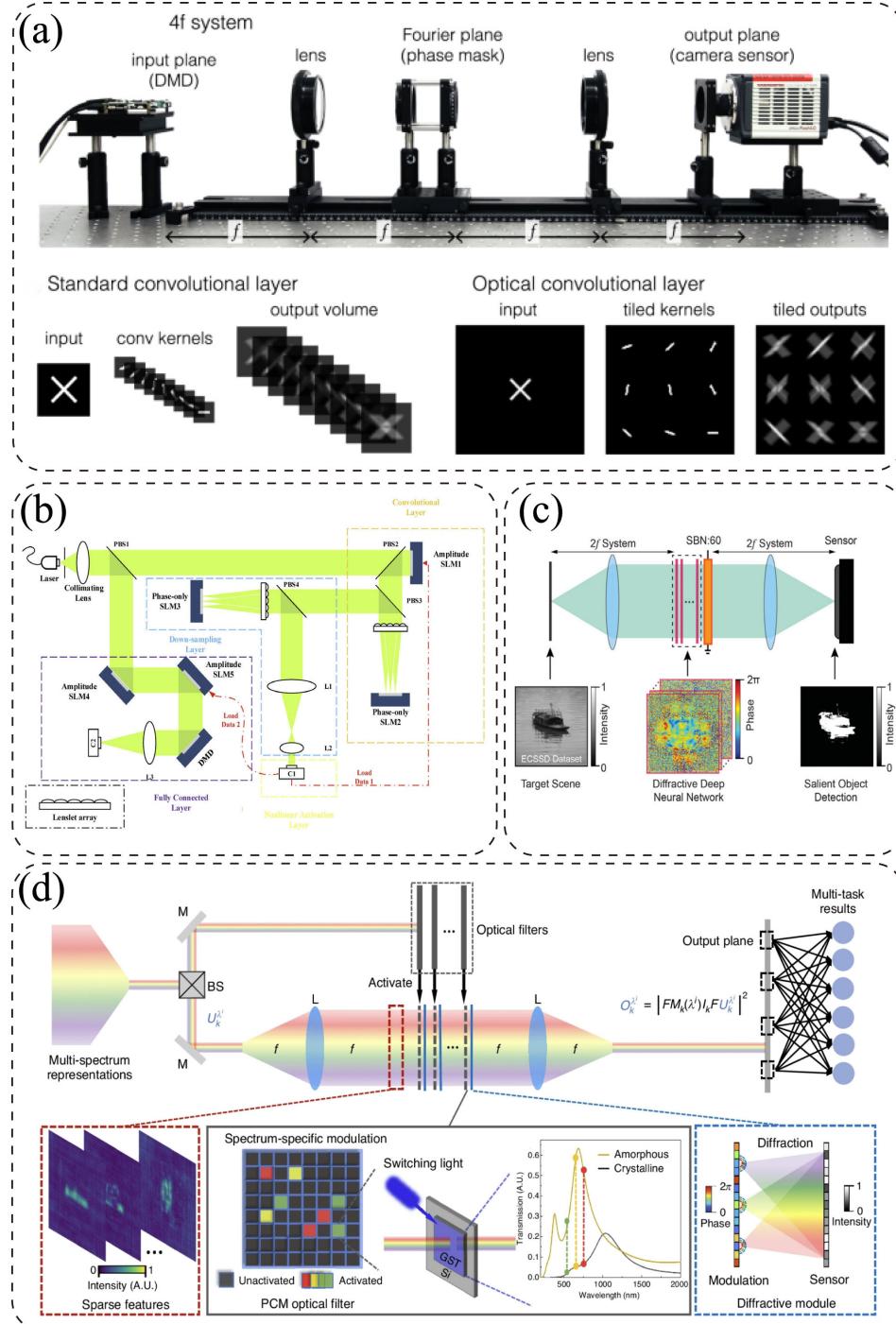


Fig. 7. Non-integrated optical neural network based on the optical 4f system. (a) Experimental setup of the 4f optical system for convolution operation and a comparison of the standard convolutional layer component with the optical convolutional layer component.^[177] (b) Schematic of the implementation of optronic convolutional neural network, where the yellow dotted box section represents the convolution module of the network.^[183] (c) Schematic of the implementation of F-D²NN.^[184] (d) Schematic of the implementation of L²ONN based on the 4f optical system, in which at the Fourier plane of the 4f optical system is a multi-layered sparse optical layer (optical filter) composed of phase change material.^[179]

diffractive surfaces, this system performs computations in the frequency domain by utilizing the principle of Fourier transform, although its essence can still be classified as a PLC-based principle. According to the convolution theorem, convolution operations in the space-time domain are equivalent to multiplication in the frequency domain. In the $4f$ system, the two-dimensional light field is first mapped to the frequency domain through lenses to realize the Fourier transform. Subsequently, specific modulation or modification operations are introduced in the Fourier plane. Finally, the modulated frequency domain signal is inversely transformed back into the spatial domain through a second lens to complete the equivalent convolution operation. In addition to the convolution function, the $4f$ system can combine nonlinear optical materials or diffractive optical structures in the Fourier plane to introduce a nonlinear activation mechanism, giving it stronger information extraction and nonlinear fitting capabilities.^[178,184] This makes ONNs based on the $4f$ system have greater potential in terms of expressiveness and functionality.

In 2018, researchers introduced a hybrid optoelectronic convolutional neural network. They used optimized diffractive optics to enhance performance.^[177] A pre-optimized phase mask was placed in the Fourier plane of a $4f$ system, as shown in Fig. 7(a). This setup enabled optical convolution. The network achieved higher image classification accuracy by integrating an electronic nonlinear activation function and a fully connected layer than traditional fully connected neural networks. It also reduced training costs. In 2021, an optronic convolutional neural network based on the optical $4f$ system was proposed. This structure is different from the traditional $4f$ system composed of two lenses, and its convolution module adopts the combination design of a lens array and spatial light modulator, which can realize multi-channel convolution calculation.^[183] As shown in the yellow wireframe in Fig. 7(b), the operations of multiple convolutional kernels are assigned to different lenses for processing, significantly improving the efficiency of parallel computing. In addition, the network also integrates a nonlinear activation function module and a fully connected computing module after the convolution module to form a complete classification network architecture. Experimental results show that the performance of the optronic convolutional neural network in image classification tasks is better than that of some ONNs based on electronic computing, and its recognition accuracy is higher.

In addition to convolutional calculations, placing a trainable D²NN in the Fourier plane of a $4f$ system can significantly improve the performance of ONNs, called Fourier-space diffractive deep neural network (F-D²NN).^[173] As shown in Fig. 7(c), this design directly modulates the optical field spectrum in the frequency domain and works with the nonlinear activation layer of the SBN:60 ferroelectric thin film to improve MNIST classification accuracy to 97.0% and enhance the system's robustness. In 2024, a reconfigurable ONN based on the $4f$ system was developed, named the lifelong-learning optical neural network (L²ONN).^[179] They introduced phase

change materials placed in the Fourier plane of a $4f$ system to dynamically control the connections of optical neurons to achieve efficient multitask learning [Fig. 7(e)]. Unlike previous models, their system allows for adjusting the activation state of the optical filter according to tasks instead of remaining unchanged. This adaptability greatly enhances the reusability and flexibility of ONNs and expands their bandwidth.

3.2. Integrated ONNs. On-chip integration has emerged as a key approach for miniaturizing ONNs, attracting significant attention from academia and industry.^[185,186] Conventional free-space ONNs rely on bulky optical components, which hinder their practical integration.^[156] In contrast, integrated ONNs achieve compact designs and can be roughly divided into two categories. The first type exploits silicon photonics, where light propagates through waveguides and implements MVM via coherent interference or optical resonance principles.^[187] The second type is based on integrated metasurfaces such as subwavelength structures.^[188] Like D²NN, it implements MVM-based PLC, but these systems are integrated. This section provides a comprehensive overview of representative research advances in on-chip integrated ONNs.

3.2.1. MZI. The MZI is a fundamental optical device widely employed in optical measurement, communication, sensing, modulation, and computing. Named after Ludwig Mach and Ludwig Zehnder, its operation relies on the coherent interference of light waves.^[189] By integrating beam splitters and phase shifters, the MZI enables precise control of both the amplitude and phase of light, providing a critical foundation for high-precision optical computation. In 1994, a triangular MZI array was introduced and constructed using cascaded beam splitters and phase shifters, demonstrating its capability for unitary matrix operations in optical systems.^[190] Later, in 2016, this architecture was optimized, leading to the development of a matrix-based MZI array.^[191] These two configurations have become the dominant frameworks for MZI array design. As illustrated in Fig. 8(a), when an input optical field \mathbf{E}_{in} propagates through an MZI array with $N = 9$ channels, the transformation between \mathbf{E}_{in} and the output field \mathbf{E}_{out} can be described by an $N \times N$ unitary matrix \mathbf{U} , expressed as $\mathbf{E}_{\text{out}} = \mathbf{U} \cdot \mathbf{E}_{\text{in}}$. In both frameworks, constructing a unitary matrix \mathbf{U} for an N -channel system requires a network of $\frac{N(N-1)}{2}$ MZIs. The transfer matrix of an individual MZI responsible for linking input channels m and n is given by

$$\mathbf{T}_{m,n}(\theta, \varphi) = \begin{bmatrix} 1 & 0 & & \cdots & & & 0 \\ 0 & 1 & & & & & \\ & & \ddots & & & & \\ \vdots & & & e^{i\varphi} \cos \theta & -\sin \theta & & \\ & & & e^{i\varphi} \sin \theta & \cos \theta & & \\ \vdots & & & & & \ddots & \\ 0 & & \cdots & & & & 1 & 0 \\ 0 & & \cdots & & & & 0 & 1 \end{bmatrix}. \quad (9)$$

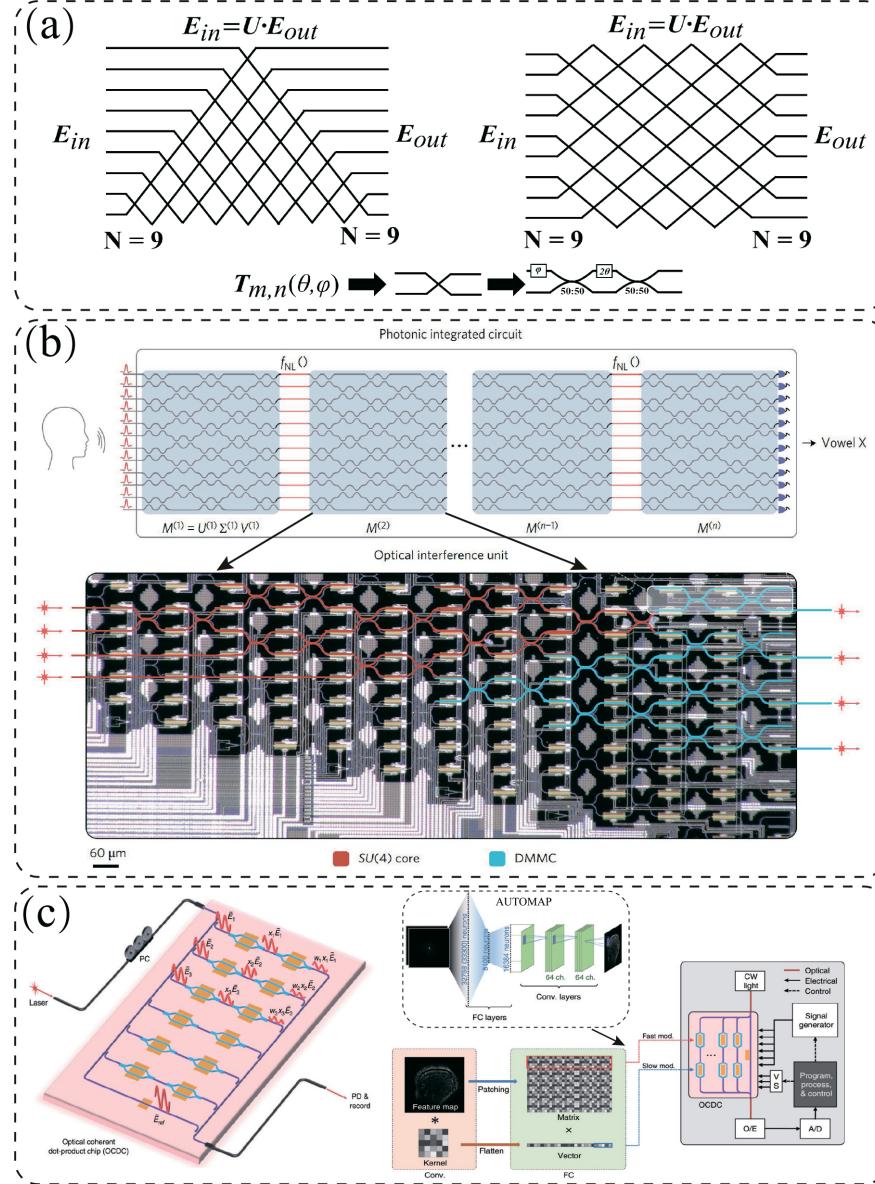


Fig. 8. Integrated optical neural network based on MZI. (a) Two classic unitary transform MZI mesh architectures: rectangular architecture and triangular architecture. (b) A photonic integrated circuit diagram is based on an MZI array and an optical micrograph of neurons in one layer.^[161] (c) The conceptual schematic of the optical coherent dot-product chip and its implementation in the AUTOMAP network.^[193]

The unitary matrix U of the entire grid can be expressed as $U = \prod_{(m,n) \in S} T_{m,n}(\theta, \varphi)$. Any weight matrix W can be decomposed via singular value decomposition into the form $W = U \Sigma V^*$, where Σ is a diagonal matrix, and U and V^* are unitary matrices.^[192] By applying distinct phase modulations to each MZI unit and adjusting the interconnections between units, arbitrary transmission operators can be designed, including fully connected computations, convolutional computations, and nonlinear activation functions commonly used in deep learning algorithms. In 2017, researchers experimentally demonstrated a programmable cascaded MZI ONN based on the singular value decomposition method, achieving fully connected computation for the first time.^[161] Figure 8(b) illustrates that the optical interference unit (OIU) module designed

in this work consists of 56 programmable MZIs. The red section on the left represents the unitary matrix (U and V^*), while the blue section on the right corresponds to the diagonal matrix Σ . Arbitration weight matrix operators W can be configured by precisely controlling optical beam splitters, phase shifters, and attenuators. In 2021, researchers introduced a silicon-based optical coherent dot-product chip, capable of executing deep learning regression tasks.^[193] Leveraging this chip, they successfully implemented the AUTOMAP network architecture, which integrates fully connected and convolutional computations as illustrated in Fig. 8(c). Their approach demonstrated image reconstruction performance on par with a 32-bit digital computer, highlighting the potential of ONNs in tackling complex regression problems.

Optical MVM based on MZI arrays relies on the coherence of a single wavelength. As computational complexity increases, cascading MZI structures expand system dimensions and design complexity while limiting parallel

computing efficiency.

3.2.2. Micro-Ring Resonators. WDM offers an incoherent approach where distinct wavelengths serve as optical carriers, each mapped to a specific matrix

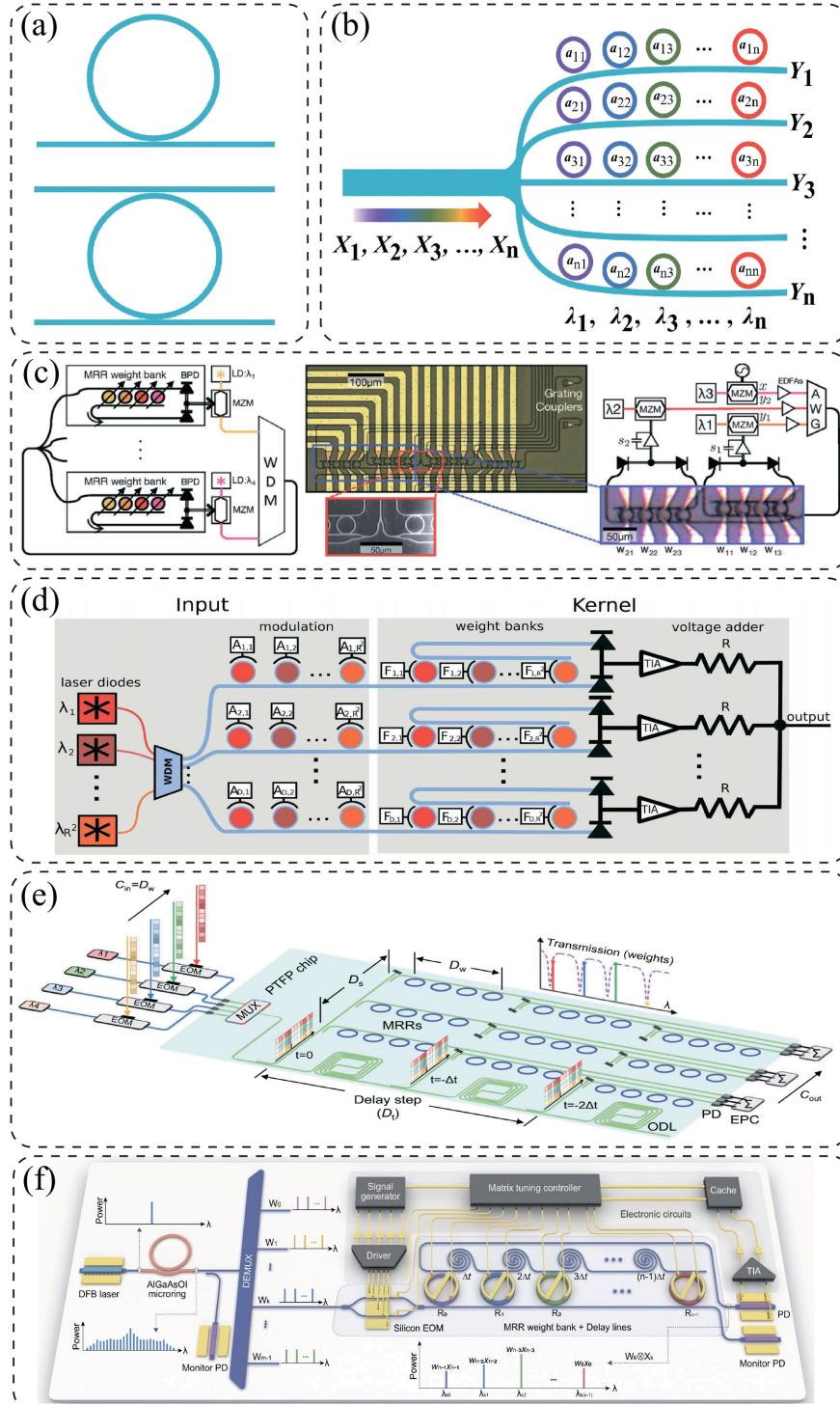


Fig. 9. Integrated optical neural network based on MRR. (a) Two classic MRRs: all-pass MRR configuration and add-drop type MRR configuration. (b) Schematic of MVM implemented by an MRR array based on WDM. (c) Concept of a broadcast-and-weighted network with MRR used as neurons, along with a corresponding experimental setup and sample micrograph.^[194] (d) Schematic of an MRR-array-based digital electronic-analog photonic for single convolved pixel operations.^[195] (e) The concept of the photonic tensor flow processor chip is one in which convolution operations are implemented through hybrid manipulation of optical wavelength (corresponding to D_w), spatial dimension (corresponding to D_s), and time delay steps (corresponding to D_t).^[196] (f) The concept of the PPU chip is that the convolution unit consists of optical delay lines alternately arranged with a tunable MRR.^[197]

element.^[194–202] Micro-ring resonators (MRRs) are fundamental to WDM-based architectures, enabling selective filtering and enhancement of optical signals that satisfy specific resonance conditions. Figure 9(a) shows that MRRs typically exist in all-pass and add-drop structures, each providing distinct mechanisms for manipulating optical signals.^[156] Figure 9(b) illustrates the principle of optical MVM in neural networks using an array of MRRs combined with WDM. In this scheme, input vector $\mathbf{X} = [X_1, X_2, \dots, X_n]$ undergoes encoding into optical signals at distinct wavelengths $[\lambda_1, \lambda_2, \dots, \lambda_n]$, where the optical power at each wavelength represents the magnitude of the corresponding vector element. Each row of n micro-rings resonates at these specific input wavelengths, with individual rings dynamically implementing the weight coefficients a_{ij} through tuning. The final weighted summation is performed by a photodetector, yielding the output \mathbf{Y} , mathematically described as

$$\begin{aligned}\mathbf{Y} &= \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \mathbf{AX} \\ &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}. \quad (10)\end{aligned}$$

This broadcast-weighted computational framework was first proposed and experimentally validated by the research team behind the development.^[194,198] As illustrated in Fig. 9(c), the approach enables precise modulation of matrix weight elements a_{ij} through continuous tuning of MRRs, thereby facilitating the implementation of ONNs. In 2020, the same group extended this concept by developing a hybrid digital electronic-analog photonic architecture leveraging micro-ring weight banks, enabling optical convolutional neural network computations.^[195] Figure 9(d) shows that performing a single convolution kernel operation within this framework requires n^2 wavelengths and n^2 MRRs for an $n \times n$ kernel.

Leveraging multiplexing across additional dimensions has proven to be a practical approach to further enhance the integration density and parallel computing capability of ONNs based on micro-ring weight banks. In 2022, an integrated photonic tensor flow processor was introduced, combining optical delay lines with micro-ring weight banks [Fig. 9(e)].^[196] This architecture enables convolutional neural networks for video recognition by leveraging hybrid control across three dimensions: wavelength, spatial channels, and temporal delay steps, where the spatial dimension represents different convolution kernel groups, and the wavelength dimension corresponds to the number of channels. In 2023, a photonic process-

ing unit (PPU) based on a microcell-driven chip was proposed, as shown in Fig. 9(f).^[197] This architecture implements a time-wavelength plane-stretched optical convolution method in which convolution units are structured as alternating optical delay lines and micro-ring weight banks. The system enhances computational throughput and parallel processing capability by leveraging an integrated optical frequency comb to drive the micro-ring weight banks.

3.2.3. Integrated Metasurfaces.

Integrated metasurface-based ONNs exhibit functional parallels to diffractive ONNs. Especially, some subwavelength-scale unit structures (dimensions $< \lambda/2$) enable unique light-matter interactions governed by coupled scalar and vector diffraction theories.^[203] These systems primarily manifest in two-dimensional and one-dimensional integrated metasurfaces. Owing to their ultrathin profiles, exceptional light-field modulation efficiency, and tailorabile material properties, subwavelength metasurfaces deliver superior performance in the visible spectrum while enabling high integration density and serving as versatile platforms for multichannel multiplexing.

In 2022, Hu *et al.* demonstrated a multiplexed diffractive neural network based on on-chip metasurfaces, as shown in Fig. 10(a).^[203] The architecture integrates with a complementary metal-oxide-semiconductor imaging sensor, and the designed diffractive neural network builds on subwavelength-scale metasurfaces and introduces the concept of polarization multiplexing, which enables the network to achieve chip-level multispectral sensing in the visible light band and improve the areal density of neurons. In 2023, Chen *et al.* demonstrated an all-analog photo-electronic computing chip ACCEL, integrating a metasurface-based diffractive neural network and analog electronic computing in one chip, as illustrated in Fig. 10(b).^[204] By fusing large-scale photonic neurons and nonlinear analog electronic computing without time-consuming analog-to-digital converters, ACCEL exceeded end-to-end computing speed and energy efficiency in intelligent vision tasks such as ImageNet classification. A parallel effort proposed a diffractive optical neural network (DONN) based on a one-dimensional integrated diffractive surface, as shown in Fig. 10(c).^[205] Its core is a one-dimensional dielectric metasurface with silicon grooves filled with silicon dioxide. The length of the silicon grooves and their spacing jointly affect the effective refractive index of the structure, thereby controlling the phase delay. In this study, the team constructed a three-layer DONN model to classify iris data sets, achieving theoretical classification accuracy up to 90%. In subsequent research, the group further designed an optical convolutional neural network, as shown in Fig. 10(d), based on one-dimensional subwavelength metasurfaces and proposed an optical convolution unit (OCU) module.^[206] The study shows that by introducing structural reparameterization ideas, the module can adapt to any real-valued convolution kernel and significantly improve computational throughput.

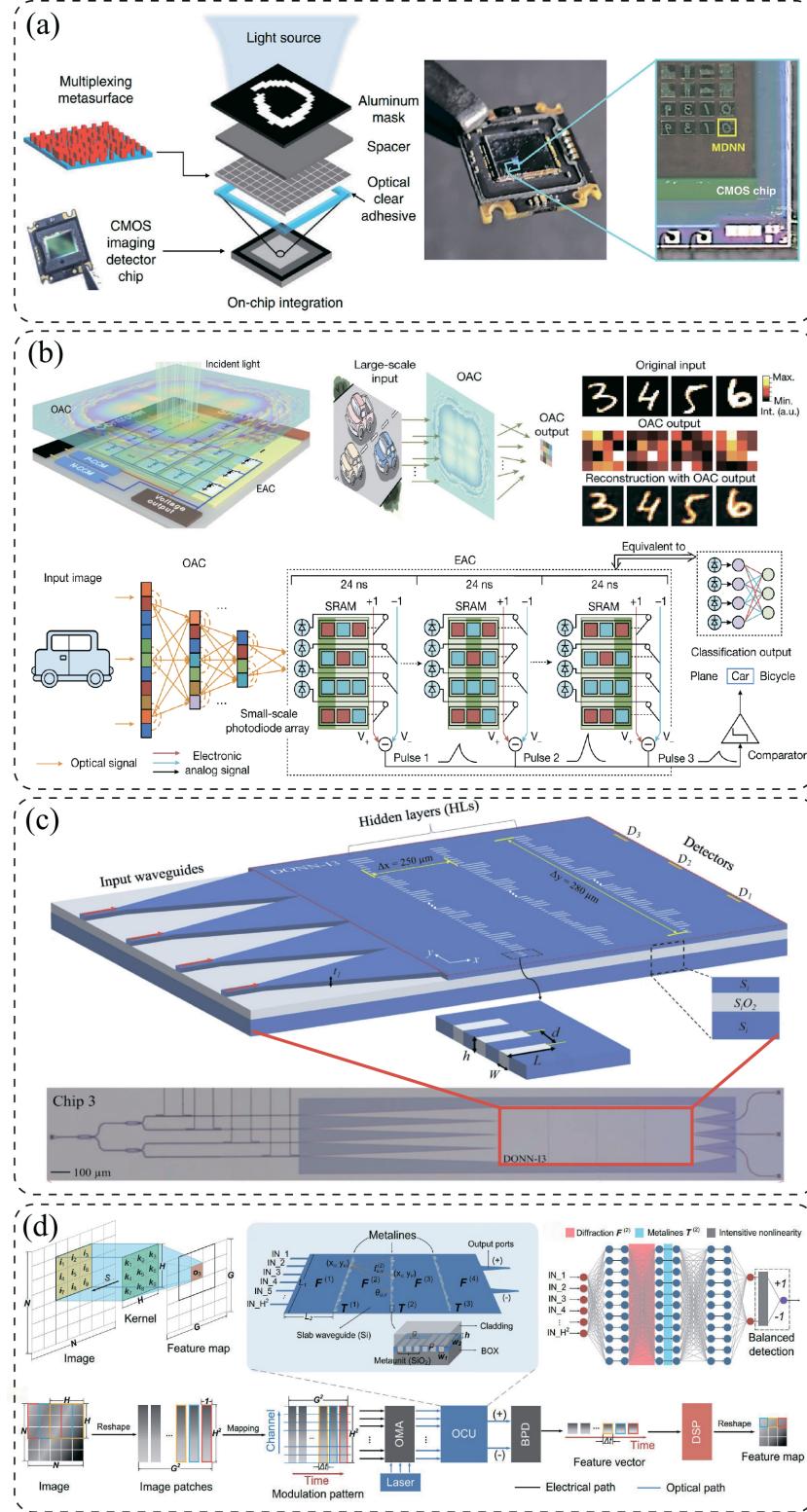


Fig. 10. Integrated optical neural network based on subwavelength metasurfaces. (a) Experimental setup of the on-chip multiplexed diffractive neural network.^[203] (b) The architecture of the on-chip ACCEL.^[204] (c) Schematic of the on-chip DONN-I3 structure and its sample micrograph.^[205] (d) Schematic of optical image convolution processing based on the OCU module.^[206]

4. Discussion and Outlook. Integrating nanophotonics and AI constitutes an emerging interdisciplinary discipline that promotes the development of nanophotonics and important breakthroughs in AI. AI has significantly

improved nanophotonic devices' numerical simulation and inverse design capabilities, enabling original complex and computationally intensive problems to be solved more efficiently. On the other hand, nanophotonic materials and

devices provide a new computing architecture for AI computing, showing unique advantages over traditional electronic computing architectures in terms of energy consumption, computing speed, and bandwidth. The two are symbiotic and mutually reinforcing; for example, the construction of optical neural networks also relies on inverse design ideas. Taking the diffractive optics neural network [shown in Fig. 6(a)] as an example, which is based on the Rayleigh–Sommerfeld diffraction mechanism and is trained on task-specific data labels to optimize the phase distribution of each layer for efficient computation.^[164] Regarding integrated optical neural networks, designs based on adjoint methods have also emerged in an endless stream.^[160,163,207–209] In addition, nanophotonics-based computing platforms can perform everyday tasks such as image classification and speech recognition and can also be used for inverse design.^[210,211] Silicon-on-insulator-based photonic simulators have been developed to design optical multiple-input–multiple-output descramblers, optical matrix computations, and tunable wavelength selective switches.^[210] Although this type of application is not inverse design in a broad sense, it still brings breakthroughs and development opportunities in this field.

The rapid advancement of nanophotonics technology

faces fundamental challenges—computational paradigms and manufacturing processes—that hinder its transition from laboratory prototypes to engineering applications. This section discusses these core issues and proposes potential breakthrough directions for future development.

4.1. Synergistic Strategies for Overcoming Technical Bottlenecks

4.1.1. Design Methods. The inverse design method is driven by electromagnetic simulation. As shown in Table 1, whether it is the early data acquisition of deep learning algorithms or the calculation of FOM by heuristic algorithms, simulation calculations are indispensable and generally face the problem of “dimensionality disaster”. For instance, the multiparameter optimization process for broadband achromatic metalenses requires handling over 10^6 discontinuous simulation data points, with traditional finite-difference time-domain methods generating redundant computations through grid discretization. Current research focuses on constructing neural operator networks embedded with Maxwell’s equations. By introducing physical constraints to reduce data dimensionality—such as employing Fourier neural operators to capture intrinsic electromagnetic field modes—the training data demand can be reduced by two orders of magnitude.

Table 1. Performance comparison of three types of optimization algorithms.

Performance	Types		
	Heuristic optimization	Gradient-based optimization	Deep learning
Computational complexity	Depends on the calculation of FOM.	Low	High during training
Convergence speed	Low	High	High
Data requirement	Moderate	Moderate	High
Application scenarios	Discrete parameter optimization	Continuous parameter optimization	Complex parameter optimization
Tolerance to fabrication error	Moderate	Moderate	Moderate

4.1.2. Computational Efficiency. Enhancing optimization requires transcending the limitations of conventional algorithm architectures. While hybrid algorithms combining genetic and adjoint methods achieve $3\text{--}5\times$ acceleration, centimeter-scale devices (more than 10^8 meta-atoms) still demand weeks of computation. Recent advances in quantum tensor networks offer novel solutions: utilizing matrix product states to characterize meta-atom topological configurations can compress parameter spaces to polynomial complexity. Experimental validation shows this approach achieves a $10^3\times$ improvement in convergence speed for silicon-based metasurface design.

4.1.3. Manufacturing-Design Mismatch. As shown in Table 1, systemic solutions are urgently needed for nanoscale fabrication errors, particularly in multilayer heterostructures where efficiency losses exceed 30% compared to simulations. The fault-tolerant co-optimization algorithm has achieved significant progress: by constructing a Markov decision process for fabrication errors through deep reinforcement learning, researchers can design meta-

atom units that maintain functionality under $\pm 15\text{ nm}$ deformations.

4.2. Evolutionary Directions for Intelligent Photonic Systems.

4.2.1. Photonic Computing Integration. The fusion of optical computing and nanophotonics catalyzes revolutionary optoelectronic architectures that promise to redefine the boundaries of information processing. By harnessing subwavelength-engineered meta-atoms’ unique wavefront manipulation capabilities, researchers are pioneering ultra-compact photonic circuits capable of encoding, transmitting, and processing data at light-speed scales. The breakthrough lies in exploiting the extreme field confinement within precisely designed meta-structures to induce nonlinear optical phenomena, effectively creating optical analogs of electronic transistors that bypass the thermal limitations of conventional silicon-based systems.^[212,213] This paradigm shift enables optical neural networks to perform parallel matrix operations through coherent light interference patterns, achieving computational through-

puts exceeding petascale operations per second while consuming orders of magnitude less energy than state-of-the-art electronic processors. Looking ahead, such terahertz-frequency photonic accelerators may soon power real-time brain-scale AI models, ultra-low-latency quantum-classical hybrid systems, and distributed edge computing nodes that seamlessly integrate sensing, computation, and communication into unified photonic frameworks.

4.2.2. Multimodal Optical Sensing. The frontier of metasurface technology and meta-waveguides in the field of nanophotonics is undergoing a paradigm shift, transcending traditional limitations of single-parameter optical detection and ushering in an era of multiphysics sensing systems capable of simultaneously decoding complex interactions across light, matter, and energy. Next-generation nanophotonics devices are poised to integrate multidimensional data acquisition—spanning spectral, polarization, phase, and spatial-resolved information—into ultra-compact platforms by harnessing the subwavelength-scale precision of meta-atoms.^[214,215] This evolution will unlock real-time, cross-domain sensing capabilities, enabling dynamic mapping of thermal gradients, mechanical stresses, chemical compositions, and electromagnetic fields within a single device. Future systems may merge adaptive metasurface architectures with machine learning algorithms to autonomously correlate multiphysical signatures, revolutionizing applications from biomedical diagnostics (e.g., label-free cellular metabolism tracking) to environmental intelligence networks (e.g., distributed pollution monitoring). The convergence of nanophotonic design freedom and multimodal data fusion could further birth intelligent optical skins for robotics, offering tactile-visual-auditory synesthesia or ultra-precise light detection and ranging (LiDAR) systems for autonomous vehicles that simultaneously detect objects, analyze material properties, and monitor atmospheric conditions. As metaphotonics matures toward self-calibrating, energy-autonomous sensing ecosystems, they will redefine the boundaries of photonics by bridging quantum-scale phenomena and macroscopic physical worlds through seamless physics-rich optical interfaces.

Nanophotonics technology is shifting from “design-to-function” to “learning-to-evolve.” We anticipate intelligent optical systems capable of environmental adaptation by integrating AI-driven optimization, atomic-scale fabrication, and neuromorphic photonic architectures. These “optical lifeforms” will not only manipulate complex light fields but also autonomously evolve their functional structures, potentially redefining the interface between photonic devices and biological systems.

Acknowledgements. This work was supported by the National Key R&D Program of China (Grant No. 2024YFB3614600), the National Natural Science Foundation of China (Grant No. 52402185), Guangzhou Basic and Applied Basic Research Foundation (Grant No. 2025A1515011800), and Shenzhen Science and Technology Program (Grant No. JCYJ20241202123712017).

References

- [1] Goertzel B 2014 *J. Artif. Gen. Intell.* **5** 1
- [2] Lake B M, Ullman T D, Tenenbaum J B, and Gershman S J 2017 *Behav. Brain Sci.* **40** e253
- [3] Chaudhary S 2025 *Int. J. Inf. Technol. Manag. Inf. Syst.* **16** 1131
- [4] LeCun Y, Bengio Y, and Hinton G 2015 *Nature* **521** 436
- [5] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
- [6] Butler K T, Davies D W, Cartwright H, Isayev O, and Walsh A 2018 *Nature* **559** 547
- [7] Jordan M I and Mitchell T M 2015 *Science* **349** 255
- [8] Hu J, Peng D, Chen J M, Huete A R, Yu L, Lou Z, Cheng E, Yang X, and Zhang B 2025 *The Innovation* **6** 100868
- [9] Wang K, Liang J, Li Z, Zhou H, Nie C, Deng J, Zhao X, Peng X, Chen Z, Peng Z, Huang D, Jang H S, Kong J, and Zou Y 2024 *FlexMat* **1** 234
- [10] Zheludev N I and Kivshar Y S 2012 *Nat. Mater.* **11** 917
- [11] Yu N and Capasso F 2014 *Nat. Mater.* **13** 139
- [12] Pendry J B 2000 *Phys. Rev. Lett.* **85** 3966
- [13] Shelby R A, Smith D R, and Schultz S 2001 *Science* **292** 77
- [14] Soukoulis C M, Linden S, and Wegener M 2007 *Science* **315** 47
- [15] Li W, Coppens Z J, Besteiro L V, Wang W, Govorov A O, and Valentine J 2015 *Nat. Commun.* **6** 8379
- [16] Bai J and Yao Y 2021 *ACS Nano* **15** 14263
- [17] Ashalley E, Acheampong K, Besteiro L V, Yu P, Neogi A, Govorov A O, and Wang Z M 2020 *Photon. Res.* **8** 1213
- [18] Chen W T, Zhu A Y, Sanjeev V, Khorasaninejad M, Shi Z, Lee E, and Capasso F 2018 *Nat. Nanotech.* **13** 220
- [19] Wang S, Wu P C, Su V C, Lai Y C, Chen M K, Kuo H Y, Chen B H, Chen Y H, Huang T T, Wang J H, Lin R M, Kuan C H, Li T, Wang Z, Zhu S, and Tsai D P 2018 *Nat. Nanotech.* **13** 227
- [20] Zheng X, Jia B, Lin H, Qiu L, Li D, and Gu M 2015 *Nat. Commun.* **6** 8433
- [21] Qian C, Tian L, and Chen H 2025 *Light Sci. Appl.* **14** 93
- [22] Li T, Xu H, Panmai M, Shao T, Gao G, Xu F, Hu G, Wang S, Wang Z, and Zhu S 2024 *Ultrafast Sci.* **4** 0074
- [23] Shi T, Deng Z L, Geng G, Zeng X, Zeng Y, Hu G, Overvig A, Li J, Qiu C W, Alí A, Kivshar Y S, and Li X 2022 *Nat. Commun.* **13** 4111
- [24] Li J, Ni Y, Zhao X, Wang B, Xue C, Bi Z, Zhang C, Dong Y, Tong Y, Tang Q, and Liu Y 2024 *Light Sci. Appl.* **13** 177
- [25] Xie C, Zou X, Zou F, Yan L, Pan W, and Zhang Y 2021 *Chin. Phys. B* **30** 120703
- [26] Taflove A, Hagness S C, and Piket-May M 2005 *The Electrical Engineering Handbook* (Elsevier) p. 629
- [27] Szabó B and Babuska I 2021 *Finite Element Analysis: Method, Verification and Validation* (Newark: John Wiley & Sons, Incorporated)
- [28] Moharam M G, Gaylord T K, Grann E B, and Pommet D A 1995 *J. Opt. Soc. Am. A* **12** 1068
- [29] Zhan T, Shi X, Dai Y, Liu X, and Zi J 2013 *J. Phys. Condens. Matter* **25** 215301
- [30] Ma W, Liu Z, Kudyshev Z A, Boltasseva A, Cai W, and Liu Y 2021 *Nat. Photon.* **15** 77
- [31] Lalau-Keraly C M, Bhargava S, Miller O D, and Yablonovitch E 2013 *Opt. Express* **21** 21693
- [32] Goh G B, Hodas N O, and Vishnu A 2017 *J. Comput. Chem.* **38** 1291
- [33] Chen C, Zuo Y, Ye W, Li X, Deng Z, and Ong S P 2020 *Adv. Energy Mater.* **10** 1903242
- [34] Shen P X, Jiang W J, Li W K, Lu Z D, and Deng D L 2021 *Acta Phys. Sin.* **70** 140302 (in Chinese)

- [35] Chen J, Zhou Z, Kim B J, Zhou Y, Wang Z, Wan T, Yan J, Kang J, Ahn J H, and Chai Y 2023 *Nat. Nanotech.* **18** 882
- [36] Chen J, Zhou Y, Yan J, Liu J, Xu L, Wang J, Wan T, He Y, Zhang W, and Chai Y 2022 *Nat. Commun.* **13** 7758
- [37] Chen J, Zhang T, Wang J, Xu L, Lin Z, Liu J, Wang C, Zhang N, Lau S P, Zhang W, Chhowalla M, and Chai Y 2022 *Sci. Adv.* **8** eabn3837
- [38] Hu Y, Zhang W, Chen Y, Zuo H, Tian M, Tang M, Li L, Xie Z, and Huang Y 2024 *Results Phys.* **61** 107805
- [39] Hu Y, Tang Z, Hu J, Lu X, Zhang W, Xie Z, Zuo H, Li L, and Huang Y 2023 *Opt. Commun.* **540** 129488
- [40] Tang Z, Li L, Zhang H, Yang J, Hu J, Lu X, Hu Y, Qi S, Liu K, Tian M, Jin J, Zhang Z, Lin H, and Huang Y 2022 *Mater. Des.* **223** 111264
- [41] Lu X, Li W, Zhu Z, Hu Y, Tang Z, Zhang W, Liu K, Su Y, Zheng J, Chen W, Tang M, Xie Z, Huang Y, and Li L 2022 *Adv. Theory Simul.* **5** 2100338
- [42] Qie J, Khoram E, Liu D, Zhou M, and Gao L 2021 *Photonics Res.* **9** B104
- [43] Xu Y, Zhang X, Fu Y, and Liu Y 2021 *Photon. Res.* **9** B135
- [44] Wiecha P R, Arbouet A, Girard C, and Muskens O L 2021 *Photon. Res.* **9** B182
- [45] Liu Y, Qian K, Wang K, and He L 2022 *IEEE Syst. J.* **16** 1424
- [46] Misra J and Saha I 2010 *Neurocomputing* **74** 239
- [47] Poon C S and Zhou K 2011 *Front. Neurosci.* **5** 108
- [48] Silver D, Huang A, Maddison C J, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, and Hassabis D 2016 *Nature* **529** 484
- [49] Graves A, Wayne G, Reynolds M, Harley T, Danihelka I, Grabska-Barwińska A, Colmenarejo S G, Grefenstette E, Ramalho T, Agapiou J, Badia A P, Hermann K M, Zwols Y, Ostrovski G, Cain A, King H, Summerfield C, Blunsom P, Kavukcuoglu K, and Hassabis D 2016 *Nature* **538** 471
- [50] Moons B and Verhelst M 2017 *IEEE J. Solid-State Circuits* **52** 903
- [51] Waldrop M M 2016 *Nature* **530** 144
- [52] Guo X, Xiang J, Zhang Y, and Su Y 2021 *Adv. Photon. Res.* **2** 2170019
- [53] Zhou C, Li Q, Li C, Yu J, Liu Y, Wang G, Zhang K, Ji C, Yan Q, He L, Peng H, Li J, Wu J, Liu Z, Xie P, Xiong C, Pei J, Yu P S, and Sun L 2023 arXiv:2302.09419 [cs.AI]
- [54] Hines M L and Carnevale N T 1997 *Neural Comput.* **9** 1179
- [55] Caulfield H J, Kinser J, and Rogers S K 1989 *Proc. IEEE* **77** 1573
- [56] Fisher A D, Lippincott W L, and Lee J N 1987 *Appl. Opt.* **26** 5039
- [57] Farhat N H, Psaltis D, Prata A, and Paek E 1985 *Appl. Opt.* **24** 1469
- [58] Goodman J W, Dias A R, and Woody L M 1978 *Opt. Lett.* **2** 1
- [59] Lugt A V 1964 *IEEE Trans. Inf. Theory* **10** 139
- [60] Athale R A, Psaltis D, and Wagner K 1988 *Appl. Opt.* **27** 1641
- [61] Guo R, Lang Q, Zhang Z, Hu H, Liu T, Wang J, and Cheng Z 2024 *Chip* **3** 100104
- [62] Engheta N 2007 *Science* **317** 1698
- [63] Caulfield H J and Dolev S 2010 *Nat. Photon.* **4** 261
- [64] Sayed S and Selvaganapathy P R 2022 *Microsyst. Nanoeng.* **8** 20
- [65] Lugli P, Harrer S, Strobel S, Brunetti F, Scarpa G, Tornow M, and Abstreiter G 2007 *7th IEEE Conference on Nanotechnology (IEEE NANO)*, Hong Kong, China, 2007, p. 1179
- [66] Stokes K, Clark K, Odetade D, Hardy M, and Oppenheimer P G 2023 *Discov. Nano* **18** 153
- [67] Wang S, Zhou Z, Li B, Wang C, and Liu Q 2021 *Mater. Today Nano* **16** 100142
- [68] Wu J, Lin X, Guo Y, Liu J, Fang L, Jiao S, and Dai Q 2022 *Engineering* **10** 133
- [69] Fu T, Zhang J, Sun R, Huang Y, Xu W, Yang S, Zhu Z, and Chen H 2024 *Light Sci. Appl.* **13** 263
- [70] Ciacci G, D'Andrea C, Matteini P, Alessi E M, Di Ruzza S, Pascali M A, Colantonio S, Giannetti A and Barucci A 2024 *Italian Conference on Optics and Photonics (ICOP)*, Firenze, Italy, 2024, p. 1
- [71] Ma Q, Liu C, Xiao Q, Gu Z, Gao X, Li L, and Cui T J 2022 *Photon. Insights* **1** R01
- [72] Yang Y, Xin H, Liu Y, Cheng H, Jin Y, Li C, Lu J, Fang B, Hong Z, and Jing X 2024 *Chin. J. Phys.* **89** 991
- [73] Hu J, Mengu D, Tzarouchis D C, Edwards B, Engheta N, and Ozcan A 2024 *Nat. Commun.* **15** 1525
- [74] Jacob Z, Alekseyev L V, and Narimanov E 2006 *Opt. Express* **14** 8247
- [75] Liu Z, Lee H, Xiong Y, Sun C, and Zhang X 2007 *Science* **315** 1686
- [76] Chi H, Hu Y, Ou X, Jiang Y, Yu D, Lou S, Wang Q, Xie Q, Qiu C, and Duan H 2025 *Adv. Mater.* **37** 2419621
- [77] Wang D, Wang X, Pan Y, Bian J, Liu K, Guo J, Lin J, Sun Z, Gou S, Sheng C, Dong X, Su H, Zhu Y, Sun Q, Xu Z, Guo A, Shao L, Chen H, and Bao W 2024 *Adv. Opt. Mater.* **12** 2400547
- [78] Zhang H, Jia S, Yin D, and Feng J 2024 *Laser Photon. Rev.* **18** 2400358
- [79] Xuan Z, Li J, Liu Q, Yi F, Wang S, and Lu W 2021 *The Innovation* **2** 100081
- [80] Li T, Liu G, Kong H, Yang G, Wei G, and Zhou X 2023 *Coord. Chem. Rev.* **475** 214909
- [81] Smith D R, Padilla W J, Vier D C, Nemat-Nasser S C, and Schultz S 2000 *Phys. Rev. Lett.* **84** 4184
- [82] Cai W, Chettiar U K, Kildishev A V, and Shalaev V M 2007 *Nat. Photon.* **1** 224
- [83] Schurig D, Mock J J, Justice B J, Cummer S A, Pendry J B, Starr A F, and Smith D R 2006 *Science* **314** 977
- [84] Kildishev A V, Boltasseva A, and Shalaev V M 2013 *Science* **339** 1232009
- [85] Meng Y, Chen Y, Lu L, Ding Y, Cusano A, Fan J A, Hu Q, Wang K, Xie Z, Liu Z, Yang Y, Liu Q, Gong M, Xiao Q, Sun S, Zhang M, Yuan X, and Ni X 2021 *Light Sci. Appl.* **10** 235
- [86] Li S, Miao P, Zhang Y, Wu J, Zhang B, Du Y, Han X, Sun J, and Xu P 2021 *Adv. Mater.* **33** 2000086
- [87] Byrnes S J 2016 arXiv:1603.02720 [physics.comp-ph]
- [88] Bohren C F and Huffman D R 2004 *Absorption and scattering of light by small particles* (New York: Wiley-CH Verlag GmbH & Co KGaA)
- [89] Hooton S, Beausoleil R G, and Van Vaerenbergh T 2021 arXiv:2107.00088 [physics.comp-ph]
- [90] Radford T W, Wiecha P R, Politi A, Zeimpekis I, and Muskens O L 2025 *ACS Photon.* **12** 1480
- [91] Sun P, Van Vaerenbergh T, Fiorentino M, and Beausoleil R 2020 *Opt. Express* **28** 3756
- [92] Zhang T, Wang J, Liu Q, Zhou J, Dai J, Han X, Zhou Y, and Xu K 2019 *Photon. Res.* **7** 368
- [93] Qiu T, Shi X, Wang J, Li Y, Qu S, Cheng Q, Cui T, and Sui S 2019 *Adv. Sci.* **6** 1900128
- [94] Zhang Q, Liu C, Wan X, Zhang L, Liu S, Yang Y, and Cui T J 2019 *Adv. Theory Simul.* **2** 1800132

- [95] He M, Nolen J R, Nordlander J, Cleri A, Lu G, Arnaud T, McIlwaine N S, Diaz-Granados K, Janzen E, Folland T G, Edgar J H, Maria J, and Caldwell J D 2023 *Adv. Mater.* **35** 2209909
- [96] Hussein M, Mahmoud K R, Hameed M F O, and Obayya S S A 2017 *J. Photon. Energy* **8** 022502
- [97] Haupt R L 1995 *IEEE Antennas Propag. Mag.* **37** 7
- [98] Michielssen E, Sajer J M, Ranjithan S, and Mittra R 1993 *IEEE Trans. Microw. Theory Tech.* **41** 1024
- [99] Rahmat-Samii Y and Michielssen E 1999 *Electromagnetic optimization by genetic algorithms* (New York: J. Wiley)
- [100] Yang Z, Huangfu H, Zhao Y, Duan M, Wang D, and Zuo H 2022 *Opt. Lett.* **47** 5929
- [101] Yang Z, Fang L, Zhang X, and Zuo H 2021 *Opt. Lasers Eng.* **144** 106646
- [102] Jafar-Zanjani S, Inampudi S, and Mosallaei H 2018 *Sci. Rep.* **8** 11040
- [103] Chung H and Miller O D 2020 *ACS Photon.* **7** 2236
- [104] Schneider P I, Garcia Santiago X, Soltwisch V, Hammerschmidt M, Burger S, and Rockstuhl C 2019 *ACS Photon.* **6** 2726
- [105] Xu P, Zhang Y, Zhang S, Chen Y, and Yu S 2020 *Opt. Lett.* **45** 4072
- [106] Xiao S, Zhao F, Wang D, Weng J, Wang Y, He X, Chen H, Zhang Z, Yu Y, Zhang Z, Zhang Z, and Yang J 2023 *Opt. Express* **31** 8668
- [107] Kim J, Kim J Y, Yoon J, Yoon H, Park H H, and Kurt H 2022 *Nanophotonics* **11** 4581
- [108] Holland J H 1992 *Sci. Am.* **267** 66
- [109] Kennedy J and Eberhart R 1995 *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, 1995, p. 1942
- [110] Banks A, Vincent J, and Anyakoha C 2007 *Nat. Comput.* **6** 467
- [111] Peer E S, Van Den Bergh F and Engelbrecht A P 2003 *IEEE Swarm Intelligence Symposium. SIS'03* (Cat. No. 03EX706), Indianapolis, IN, USA, 2003, p. 235
- [112] Frankle R T 1976 *Am. J. Clin. Nutr.* **29** 105
- [113] Xu W, Hu L, Shao K, Liang H, He T, Dong S, Zhu J, Wei Z, Wang Z, and Cheng X 2023 *Opt. Express* **31** 41339
- [114] Chi H, Hu Y, Ou X, Jiang Y, Yu D, Lou S, Wang Q, Xie Q, Qiu C, and Duan H 2025 *Adv. Mater.* **37** 2419621
- [115] Jensen J S and Sigmund O 2011 *Laser Photon. Rev.* **5** 308
- [116] Wu Y 2021 *Swarm Evol. Comput.* **62** 100844
- [117] Wang K, Ren X, Chang W, Lu L, Liu D, and Zhang M 2020 *Photon. Res.* **8** 528
- [118] Qi H, Du Z, Hu X, Yang J, Chu S, and Gong Q 2022 *Opto-Electron. Adv.* **5** 210061
- [119] Liu Y, Luo Y, Zhang F, Pu M, Bao H, Xu M, Guo Y, Li L, Li X, Ma X, and Luo X 2024 *Sci. China Phys. Mech. Astron.* **67** 274213
- [120] Yin Y, Jiang Q, Wang H, Liu J, Xie Y, Wang Q, Wang Y, and Huang L 2024 *Adv. Mater.* **36** 2312303
- [121] Giles M B and Pierce N A 2000 *Flow Turbul. Combust.* **65** 393
- [122] Hughes T W, Minkov M, Williamson I A D, and Fan S 2018 *ACS Photon.* **5** 4781
- [123] Li Z, Zhou Z, Qiu C, Chen Y, Liang B, Wang Y, Liang L, Lei Y, Song Y, Jia P, Zeng Y, Qin L, Ning Y, and Wang L 2024 *Adv. Opt. Mater.* **12** 2301337
- [124] Pestourie R, Pérez-Arcibbia C, Lin Z, Shin W, Capasso F, and Johnson S G 2018 *Opt. Express* **26** 33732
- [125] Christiansen R E, Lin Z, Roques-Carmes C, Salamin Y, Kooi S E, Joannopoulos J D, Soljačić M, and Johnson S G 2020 *Opt. Express* **28** 33854
- [126] Lin Z and Johnson S G 2019 *Opt. Express* **27** 32445
- [127] Wang Z Q, Li F J, Deng Q M, Wan Z, Li X, and Deng Z L 2024 *Chin. Opt. Lett.* **22** 023601
- [128] Dainese P, Marra L, Cassara D, Portes A, Oh J, Yang J, Palmieri A, Rodrigues J R, Dorrah A H, and Capasso F 2024 *Light Sci. Appl.* **13** 300
- [129] Wang S, Wen S, Deng Z L, Li X, and Yang Y 2023 *Phys. Rev. Lett.* **130** 123801
- [130] McCulloch W S and Pitts W 1943 *Bull. Math. Biophys.* **5** 115
- [131] Wang X, Zhao Y, and Pourpanah F 2020 *Int. J. Mach. Learn. Cybern.* **11** 747
- [132] Zhen Z, Qian C, Jia Y, Fan Z, Hao R, Cai T, Zheng B, Chen H, and Li E 2021 *Photon. Res.* **9** B229
- [133] Baxter J, Calá Lesina A, Guay J M, Weck A, Berini P, and Ramunno L 2019 *Sci. Rep.* **9** 8074
- [134] Wright L G, Onodera T, Stein M M, Wang T, Schachter D T, Hu Z, and McMahon P L 2022 *Nature* **601** 549
- [135] Jiang J, Sell D, Hoyer S, Hickey J, Yang J, and Fan J A 2019 *ACS Nano* **13** 8872
- [136] Wu Q, Xu Y, Zhao J, Liu Y, and Liu Z 2024 *Nano Lett.* **24** 11581
- [137] Ma W, Cheng F, Xu Y, Wen Q, and Liu Y 2019 *Adv. Mater.* **31** 1901111
- [138] Liu D, Tan Y, Khoram E, and Yu Z 2018 *ACS Photon.* **5** 1365
- [139] Gao L, Li X, Liu D, Wang L, and Yu Z 2019 *Adv. Mater.* **31** 1905467
- [140] Anon 2017 *J. Jpn. Soc. Fuzzy Theory Intell. Inform.* **29** 177
- [141] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, and Courville A 2017 arXiv:1704.00028 [cs.LG]
- [142] Radford A, Metz L, and Chintala S 2015 arXiv:1511.06434 [cs.LG]
- [143] Brock A, Donahue J, and Simonyan K 2018 arXiv:1809.11096 [cs.LG]
- [144] Liu Z, Zhu D, Rodrigues S P, Lee K T, and Cai W 2018 *Nano Lett.* **18** 6570
- [145] Zhu J Y, Park T, Isola P, and Efros A A 2020 arXiv:1703.10593 [cs.CV]
- [146] You H, Du L, Xu X, and Zhao J 2024 *Opt. Commun.* **554** 130122
- [147] Zhang J, Qian C, Fan Z, Chen J, Li E, Jin J, and Chen H 2022 *Adv. Opt. Mater.* **10** 2200748
- [148] Ren Y, Zhang L, Wang W, Wang X, Lei Y, Xue Y, Sun X, and Zhang W 2021 *Photonics Res.* **9** B247
- [149] Hegde R S 2020 *IEEE J. Sel. Top. Quantum Electron.* **26** 1
- [150] Jiang J and Fan J A 2019 *Nano Lett.* **19** 5366
- [151] Peng Y, Fan L, Jin W, Ye Y, Huang Z, Zhai S, Luo X, Ma Y, Tang J, Zhou J, Greenburg L C, Majumdar A, Fan S, and Cui Y 2021 *Nat. Sustain.* **5** 339
- [152] Zhou J, Chen T G, Tsurimaki Y, Hajj-Ahmad A, Fan L, Peng Y, Xu R, Wu Y, Assawaworrarit S, Fan S, Cutkosky M R, and Cui Y 2023 *Joule* **7** 2830
- [153] Peng Y, Lai J C, Xiao X, Jin W, Zhou J, Yang Y, Gao X, Tang J, Fan L, Fan S, Bao Z, and Cui Y 2023 *Proc. Natl. Acad. Sci.* **120** e2300856120
- [154] Gershner E, Chen M, Mao C, Wang E W, Lalanne P, and Fan J A 2022 *ACS Photon.* **10** 815
- [155] Ha Y, Luo Y, Pu M, Zhang F, He Q, Jin J, Xu M, Guo Y, Li X, Li X, Ma X, and Luo X 2023 *Opto-Electron. Adv.* **6** 230133
- [156] Zhou H, Dong J, Cheng J, Dong W, Huang C, Shen Y, Zhang Q, Gu M, Qian C, Chen H, Ruan Z, and Zhang X 2022 *Light Sci. Appl.* **11** 30
- [157] Liu J, Wu Q, Sui X, Chen Q, Gu G, Wang L, and Li S 2021 *Photonix* **2** 5
- [158] Wu B, Zhang W, Zhou H, Dong J, Huang D, Wai P K A, and Zhang X 2023 *Photonix* **4** 37
- [159] Liu W, Xu C, Xie W, Ye J, Zhu Z, Lin Z, and Chen H 2025 *FlexMat* 1

- [160] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, and Yu Z 2019 *Photon. Res.* **7** 823
- [161] Shen Y, Harris N C, Skirla S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Laroche H, Englund D, and Soljačić M 2017 *Nat. Photon.* **11** 441
- [162] Feldmann J, Youngblood N, Wright C D, Bhaskaran H, and Pernice W H P 2019 *Nature* **569** 208
- [163] Hughes T W, Minkov M, Shi Y, and Fan S 2018 *Optica* **5** 864
- [164] Lin X, Rivenson Y, Yardimci N T, Veli M, Luo Y, Jarrahi M, and Ozcan A 2018 *Science* **361** 1004
- [165] Qian C, Kaminer I, and Chen H 2025 *Nat. Commun.* **16** 1154
- [166] Tu H, Liu H, Pan T, Xie W, Ma Z, Zhang F, Xu P, Wu L, Xu O, Xu Y, and Qin Y 2025 *Nat. Commun.* **16** 1369
- [167] Cheng J, Zhou H, and Dong J 2021 *Nanomaterials* **11** 1683
- [168] Lu L, Zeng Z, Zhu L, Zhang Q, Zhu B, Yao Q, Yu M, Niu H, Dong M, and Zhong G 2019 *IEEE Photon. Technol. Lett.* **31** 1952
- [169] Chen H, Feng J, Jiang M, Wang Y, Lin J, Tan J, and Jin P 2021 *Engineering* **7** 1483
- [170] Krueger N A, Holsteen A L, Kang S K, Ocier C R, Zhou W, Mensing G, Rogers J A, Brongersma M L, and Braun P V 2016 *Nano Lett.* **16** 7402
- [171] Ocier C R, Richards C A, Bacon-Brown D A, Ding Q, Kumar R, Garcia T J, Van De Groep J, Song J H, Cyphersmith A J, Rhode A, Perry A N, Littlefield A J, Zhu J, Xie D, Gao H, Messinger J F, Brongersma M L, Tousaint K C, Goddard L L, and Braun P V 2020 *Light Sci. Appl.* **9** 196
- [172] Li J, Hung Y C, Kulce O, Mengu D, and Ozcan A 2022 *Light Sci. Appl.* **11** 153
- [173] Fu T, Lin X, Li J, Luo G, Zheng H, Ji Z, Wang D, Yuan X, Li Y, and Geng Y 2025 *Laser Photon. Rev.* e00433
- [174] Yu H, Huang Z, Lamon S, Wang B, Ding H, Lin J, Wang Q, Luan H, Gu M, and Zhang Q 2025 *Nat. Photon.* **19** 486
- [175] Mengu D, Tabassum A, Jarrahi M, and Ozcan A 2023 *Light Sci. Appl.* **12** 86
- [176] Cheong Y Z, Thekkakara L, Bhaskaran M, Del Rosal B, and Sriram S 2024 *Adv. Photon. Res.* **5** 2300310
- [177] Chang J, Sitzmann V, Dun X, Heidrich W, and Wetzstein G 2018 *Sci. Rep.* **8** 12324
- [178] Zuo Y, Li B, Zhao Y, Jiang Y, Chen Y C, Chen P, Jo G B, Liu J, and Du S 2019 *Optica* **6** 1132
- [179] Cheng Y, Zhang J, Zhou T, Wang Y, Xu Z, Yuan X, and Fang L 2024 *Light Sci. Appl.* **13** 56
- [180] Momeni A, Rahmani B, Malléjac M, Del Hougne P, and Fleury R 2023 *Science* **382** 1297
- [181] Fu W, Zhao D, Li Z, Liu S, Tian C, and Huang K 2022 *Light Sci. Appl.* **11** 62
- [182] Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, Fang L, Lin X, and Dai Q 2019 *Phys. Rev. Lett.* **123** 023901
- [183] Gu Z, Gao Y, and Liu X 2021 *Opt. Express* **29** 5877
- [184] Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, Fang L, Lin X, and Dai Q 2019 *Phys. Rev. Lett.* **123** 023901
- [185] Shastri B J, Tait A N, Ferreira De Lima T, Pernice W H P, Bhaskaran H, Wright C D, and Prucnal P R 2021 *Nat. Photon.* **15** 102
- [186] Lu C C, Yuan H Y, Zhang H Y, Zhao W, Zhang N E, Zheng Y J, Elshahat S, and Liu Y C 2022 *Chip* **1** 100025
- [187] Park T, Mondal S, and Cai W 2025 *Laser Photon. Rev.* **19** 2401520
- [188] Wu Z, Zhou M, Khoram E, Liu B, and Yu Z 2020 *Photon. Res.* **8** 46
- [189] Saghaei H, Elyasi P, and Karimzadeh R 2019 *Photon. Nanostruct. Fundam. Appl.* **37** 100733
- [190] Reck M, Zeilinger A, Bernstein H J, and Bertani P 1994 *Phys. Rev. Lett.* **73** 58
- [191] Clements W R, Humphreys P C, Metcalf B J, Kolthammer W S, and Walsmley I A 2016 *Optica* **3** 1460
- [192] Lawson C L and Hanson R J 1995 *Solving least squares problems* (Philadelphia: SIAM)
- [193] Xu S, Wang J, Shu H, Zhang Z, Yi S, Bai B, Wang X, Liu J, and Zou W 2021 *Light Sci. Appl.* **10** 221
- [194] Tait A N, De Lima T F, Zhou E, Wu A X, Nahmias M A, Shastri B J, and Prucnal P R 2017 *Sci. Rep.* **7** 7430
- [195] Bangari V, Marquez B A, Miller H, Tait A N, Nahmias M A, De Lima T F, Peng H T, Prucnal P R, and Shastri B J 2020 *IEEE J. Sel. Top. Quantum Electron.* **26** 1
- [196] Xu S, Wang J, Yi S, and Zou W H 2022 *Nat. Commun.* **13** 7970
- [197] Bai B, Yang Q, Shu H, Chang L, Yang F, Shen B, Tao Z, Wang J, Xu S, Xie W, Zou W, Hu W, Bowers J E, and Wang X 2023 *Nat. Commun.* **14** 66
- [198] Tait A N, Jayatilleka H, De Lima T F, Ma P Y, Nahmias M A, Shastri B J, Shekhar S, Chrostowski L, and Prucnal P R 2018 *Opt. Express* **26** 26422
- [199] Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M, Le Gallo M, Fu X, Lukashchuk A, Raja A S, Liu J, Wright C D, Sebastian A, Kippenberg T J, Pernice W H P, and Bhaskaran H 2021 *Nature* **589** 52
- [200] Zhang W, Huang C, Peng H T, Bilodeau S, Jha A, Blow E, De Lima T F, Shastri B J, and Prucnal P 2022 *Optica* **9** 579
- [201] Yin R, Xiao H, Jiang Y, Han X, Zhang P, Chen L, Zhou X, Yuan M, Ren G, Mitchell A, and Tian Y 2023 *Optica* **10** 1709
- [202] Cheng J, Xie Y, Liu Y, Song J, Liu X, He Z, Zhang W, Han X, Zhou H, Zhou K, Zhou H, Dong J, and Zhang X 2023 *Nanophotonics* **12** 3883
- [203] Luo X, Hu Y, Ou X, Li X, Lai J, Liu N, Cheng X, Pan A, and Duan H 2022 *Light Sci. Appl.* **11** 158
- [204] Chen Y, Nazhamaiti M, Xu H, Meng Y, Zhou T, Li G, Fan J, Wei Q, Wu J, Qiao F, Fang L, and Dai Q 2023 *Nature* **623** 48
- [205] Fu T, Zang Y, Huang Y, Du Z, Huang H, Hu C, Chen M, Yang S, and Chen H 2023 *Nat. Commun.* **14** 70
- [206] Huang Y, Fu T, Huang H, Yang S, and Chen H 2023 *Photon. Res.* **11** 1125
- [207] Qu Y, Zhu H, Shen Y, Zhang J, Tao C, Ghosh P, and Qiu M 2020 *Sci. Bull.* **65** 1177
- [208] Nikkhah V, Pirmoradi A, Ashtiani F, Edwards B, Aflatouni F, and Engheta N 2024 *Nat. Photon.* **18** 501
- [209] Yang Z, Zhang T, Dai J, and Xu K 2024 *Opt. Express* **32** 46633
- [210] Cheng J, Zhang W, Gu W, Zhou H, Dong J, and Zhang X 2023 *ACS Photon.* **10** 2173
- [211] Gostimirovic D and Soref R 2023 *Sensors* **23** 626
- [212] Li X, Wang J, Yu F, Chen J, Chen X, Lu W, and Li G 2025 *Light Sci. Appl.* **14** 47
- [213] Wang Q, Zhang X, Xu Q, Feng X, Niu L, Chen X, Lu Y, Feng J, Fang M, Zhang X, Zhang W, and Han J 2025 *Adv. Mater.* **37** 2500392
- [214] Fan Y, Huang W, Zhu F, Liu X, Jin C, Guo C, An Y, Kivshar Y, Qiu C W, and Li W 2024 *Nature* **630** 77
- [215] Jiang H, Chen Y, Guo W, Zhang Y, Zhou R, Gu M, Zhong F, Ni Z, Lu J, Qiu C W, and Gao W 2024 *Nat. Commun.* **15** 8347