ACS Photonics

pubs.acs.org/journal/apchd5

Perspective

Algorithm-Driven Paradigms for Freeform Optical Engineering

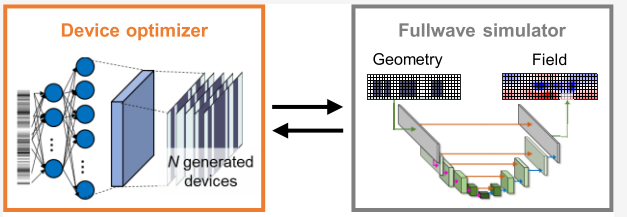
Mingkun Chen, Jiaqi Jiang, and Jonathan A. Fan*

Cite This: *ACS Photonics* 2022, 9, 2860−2871

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Advances in modern manufacturing have enabled the multiscalar patterning of dielectric media with nearly arbitrary layouts, presenting unique opportunities to revolutionize the design and fabrication pipeline for photonic technologies. In this Perspective, we discuss how algorithms based on classical optimization and deep learning are establishing a new conceptual framework for freeform optical engineering. These tools can specify suitable design parameters for a desired objective, automate the high-speed optimization of freeform devices, and augment manufacturing processes to mitigate challenges set by freeform fabrication. A central feature of many of these algorithms is their utilization of data and physics to model and exploit high-dimensional relationships between geometric structure and electromagnetic response within the constraints of Maxwell's equations. We anticipate that these algorithm-driven methods will streamline optical systems design at the physical limits of structured media and become standard academic and industry tools for scientists and engineers.



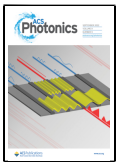
KEYWORDS: *Optical engineering, Performance bounds, Freeform optimization, Maxwell simulator, Machine learning, Fabrication*

The field of optical engineering is based on understanding and exploiting the relationship between geometric shape and electromagnetic response to enable the experimental implementation of optical systems. The concepts span a multitude of length scales, from much larger than the wavelength where waves can be treated as rays to much less than the wavelength where light−matter interactions require a full vectorial wave treatment. Traditionally, the relationship between geometry and response has been principally built from insights based on physical laws, where a combination of simplifications, approximations, and constraints to physical systems have led to intuitive models and design strategies. A simple example is the lensmaker's equation, which was developed for designing imaging systems and for which spherical surfaces were considered due to their utility in light focusing and practical constraints posed by manufacturing. Analogous concepts at much smaller length scales have been established for the customized manipulation of electromagnetic waves in integrated and free space photonics, where relatively simple geometric shapes described by a handful of free parameters serve as the typical basis for device design.

Recent developments in high-precision manufacturing have pushed the experimental patterning of geometric features to new limits, presenting opportunities to redefine the optical engineering process. Advanced machining can now be used to fabricate refractive optical components with nearly arbitrary surface profiles,[1,2] lithography can support the routine patterning of thin film materials into nanoscale freeform shapes,[3−5] and additive manufacturing can produce nearly arbitrary three-dimensional structures with feature sizes spanning the macro- to submicrometer scale.[6,7] These fabrication methods present an immense geometric feature space for optical design where, in principle, every voxel in the patterning process is a free design parameter. Concepts that take full advantage of this feature space can push system capabilities to new levels and meet the challenges posed by the next generation of quantum, communication, sensing, imaging, and display systems, which include the following: (1) operation in conditions that involve low photon counts, intolerance to stray light, or limited energy footprints, thereby requiring light manipulation with exceptionally high efficiencies,[8−10] (2) multiplexing of optical responses that are a function of incident angle, wavelength, and polarization,[11−13] (3) ultracompact form factors that maximize device density in a system,[5,14,15] and (4) unusual curvilinear or freeform layouts that enhance optical system parameters such as field of view.[16,17]

In this Perspective, we discuss emergent directions in optical engineering that leverage new advances in algorithms to design and fabricate freeform optical systems with unprecedented capabilities. This new breed of optical engineering has been made possible due to the convergence of multiple develop-

Received: April 29, 2022
Revised: July 26, 2022
Accepted: July 26, 2022
Published: August 9, 2022



https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

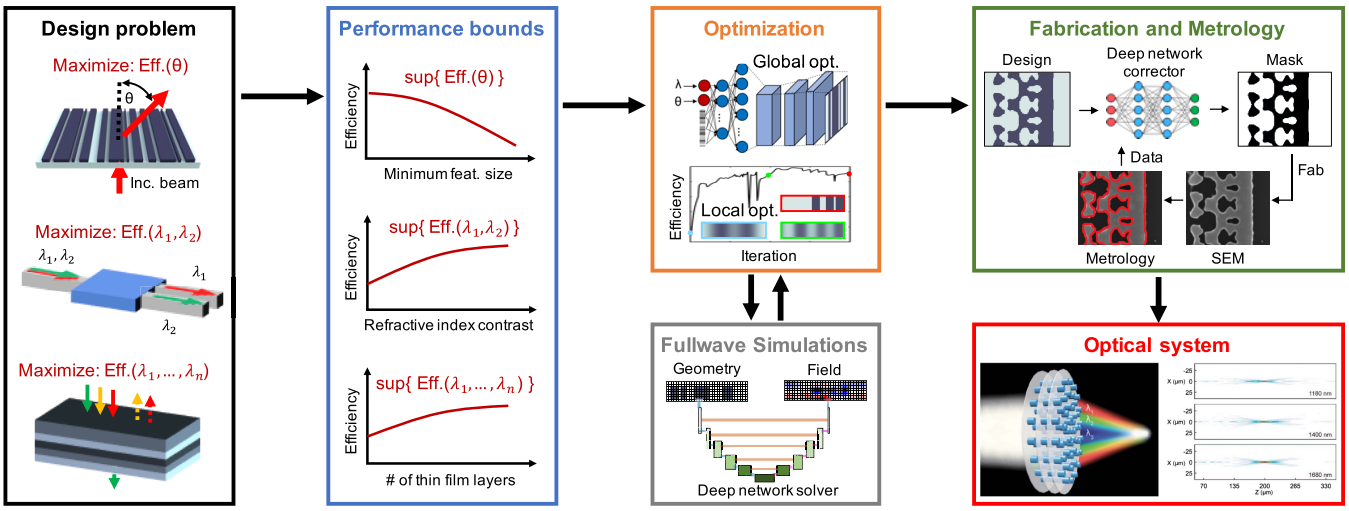


Figure 1. Overview of the algorithm-driven pipeline for optical engineering. Optical design objective is cast as an optimization problem, and performance bounds calculations ensure the presence of high-performing devices in the design space. Iterative freeform optimization algorithms coupled with high-speed fullwave electromagnetic solvers are then used to identify candidate devices. Fabrication processes based on lithography or additive manufacturing are augmented with artificial intelligence algorithms to reliably produce experimental devices. Material used for this figure is adapted with permission from refs 18, 19, 20, and 5. Copyright 2019, 2020, 2022, and 2018 American Chemical Society, The Optical Society, arXiv https://creativecommons.org/licenses/by/4.0/, and American Chemical Society.

ments over the last two decades, including the explosive growth of machine learning, advances in optimization theory, proliferation of new computing hardware, and continued development and refinement of experimental fabrication tools. Optimization methods are at the heart of this algorithm-driven pedagogy, where they aim to capitalize on the full geometric feature space made accessible by current state-of-the-art fabrication methods. However, optimization methods are only part of this growing ecosystem of algorithms, as meaningful optical design requires full accounting of experimental manufacturing capabilities and the physical limits of light−matter interactions with structured media.

Our framework for algorithm-driven optical engineering, with the goal of designing and prototyping high-performing experimental devices in an accelerated and automated manner, is outlined in Figure 1. Performance bound calculations guide the selection of the device footprint and material parameters to ensure that the geometric feature space can support designs with a desired performance metric. Iterative freeform optimization algorithms, operating within constraints set by experimental fabrication, effectively search for optima within the full geometric feature space, and they are augmented by ultrafast electromagnetic solvers that can dramatically accelerate the computing of Maxwell's equations. To reliably fabricate these devices, machine-learning algorithms are used to handle and correct manufacturing errors, which are particularly nontrivial in freeform designs. In this Perspective, we will focus our discussion on wavelength- and subwavelength-scale photonic devices, though the presented concepts apply to optical systems at all length scales, and we note the exciting algorithmic developments in ray-based optics emerging from the computer graphics and computer vision communities. We anticipate that these approaches to optical engineering will empower practitioners to move away from asking *how* to realize an optical system and focus on *what* system functionality is useful for an application.

PERFORMANCE BOUNDS

Prior to performing the detailed optimization of an optical device, the first step is specifying the scope of its geometric feature space. This scope takes different forms depending on the physical system and includes the choice of the areal footprint in a chip-based device, number of layers in a multilayer metasurface, and range of refractive index values used in any device. Ideally, the feature space is specified to be sufficiently large as to ensure the presence of high-performing devices but not so large that it adds unnecessary device size, weight, or fabrication steps. To this end, performance bounds calculations have been developed that specify the physical limits of the best devices within a geometric feature space, thereby providing direct insights on the scope of a feature space without directly needing to identify the devices themselves. They are particularly important for electromagnetics problems because the design landscapes are high dimensional and nonconvex, meaning there are many local optima of varying quality and it is very difficult to find the global optimum.

Bounds calculation methods can be roughly classified into two categories. The first is physical bounds, which are derived from fundamental physical laws and are independent of a particular device geometry or domain of optimization. These bounds are often in the form of concise analytical expressions and typically provide physical intuition as to their origins. In an early example, it was shown that the near-field heat transfer rate between two closely spaced bodies, independent of shape, is bounded by $|\chi|^2/\mathrm{Im}\chi$, where χ is the susceptibility of each body (Figure 2a).[21] These bounds were derived from conservation of energy and reciprocity laws, and they have successfully served to simplify material selection for practical near-field heat transfer systems. In another example, material-dependent upper bounds to the extinction cross section of nanoparticles with fixed volume were derived from restrictive sum rules involving the quasistatic eigenmodes of the particles (Figure 2b).[22] More recently, a framework based on local conservation laws arising from the complex Poynting theorem

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

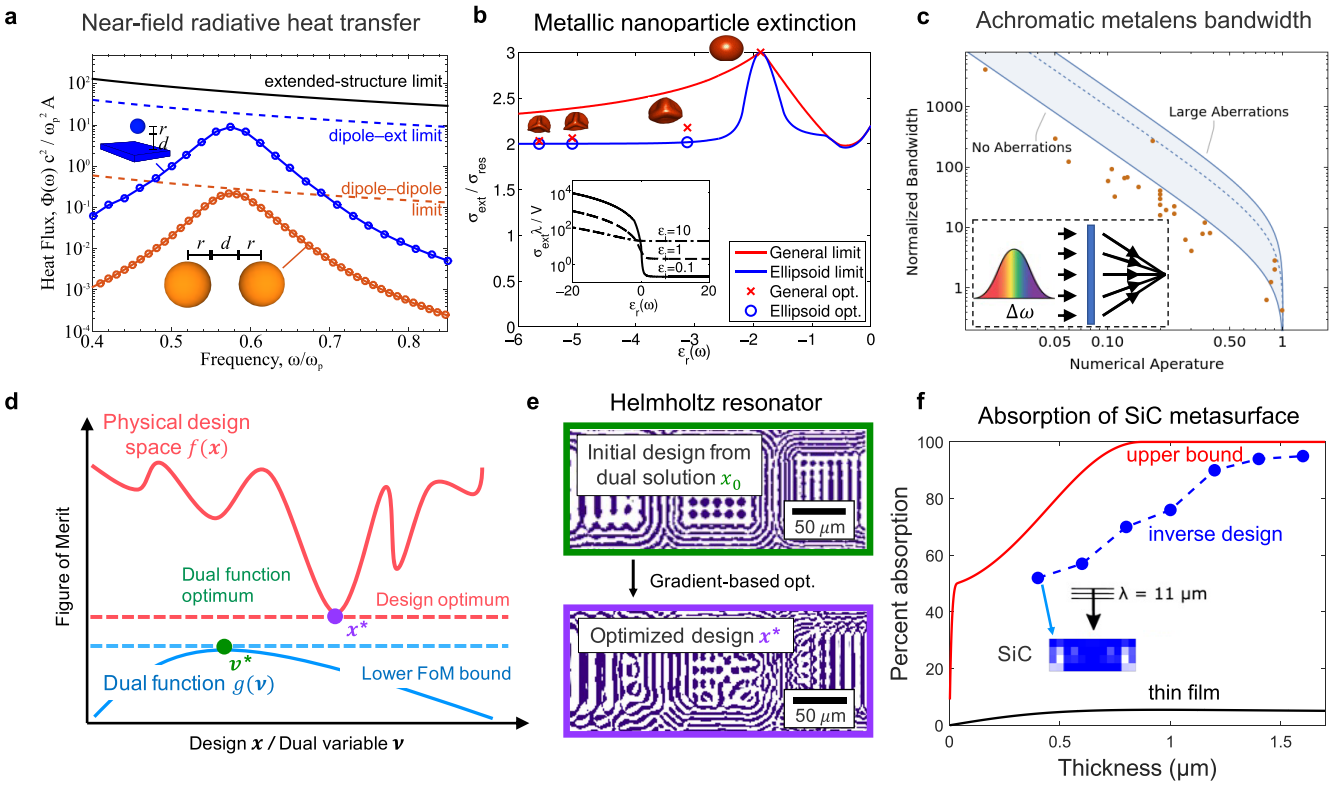


Figure 2. Performance bound calculations for photonic systems. (a–c) Tight physical bounds, derived from physical laws, specify performance limits for (a) near-field radiative transfer between two objects, (b) extinction cross section of metal nanoparticles, and (c) achromatic metalens bandwidth. (d–f) Computational bounds are calculated using principles in convex optimization. (d) Lagrange dual functions can provide tight bounds to nonconvex physical design spaces. (e) Solution of a Lagrange dual function can suggest a good starting point for gradient-based optimization, as shown in the case of a Helmholtz resonator. (f) Lagrange dual methods provide computational bounds for absorption in a silicon carbide metasurface. (a) Adapted with permission from ref 21. Copyright 2015 American Physical Society. (b) Adapted with permission from ref 22. Copyright 2014 American Physical Society. (c) Adapted with permission from ref 23. Copyright 2020 The Optical Society. (e) Adapted with permission from ref 24. Copyright 2019 American Chemical Society. (f) Adapted with permission from ref 25. Copyright 2020 The Optical Society.

was developed to compute bounds for various electromagnetic scattering phenomena, including the minimum size of a scatterer required to support a desired S matrix response, maximal bandwidth-averaged extinction from a scatterer given a bandwidth, and absorption bounds for a scatterer given its complex dielectric constant and dimension.[26]

Approximations and simplifications to physical laws can also be used to derive physical bounds. In the calculation of the upper bounds for light trapping in a nanophotonic structure, light trapping was modeled as the coupling of an incident plane wave into guided modes and bounds to the trapping enhancement factor were computed using statistical couple-mode theory.[27] Bandwidth limits of achromatic metalenses were derived by modeling thin broadband metalenses as a set of local delay lines, which use group delay to compensate for the time difference between light traveling from the metalens center to the focal spot and light traveling from the metalens edge to the focal spot (Figure 2c).[23] With this approximation, a bandwidth limit can be computed that depends on the metalens numerical aperture, delay line length, and metasurface dielectric constant.

The second category of bounds calculations is computational bounds, which frame a specific optical design problem as a physically constrained optimization problem and uses Lagrange duality to mathematically calculate the bounds.[24,28] While the physical design space $f(x)$ is nonconvex and the search for the optimal device x^* is intractable, the Lagrange dual-function $g(\nu)$ is convex and its optimum ν^* can be readily solved to provide the bound (Figure 2d).[24] As these bounds are derived from optimization theory, they are more problem specific and are often tighter but tend to lack physical intuition. To be more specific on the formalism, an optical design problem can be generally expressed as a constrained optimization problem, for example

$$\min_{\epsilon,E} \quad \|E - \hat{E}\|_2^2$$
$$\text{s. t.} \quad \nabla^2 E + \epsilon\mu\omega^2 E = 0 \tag{1}$$

with design variable ϵ, field E, and target field $\hat{E}$: Our objective is to optimize the dielectric constant ϵ according to the desired target field $\hat{E}$. The physical constraint is the wave equation derived from Maxwell's equation. To incorporate this constraint into the optimization objective, the Lagrangian is defined as the objective function augmented with a weighted constraint function

$$L(\epsilon, E, \nu) = \|E - \hat{E}\|_2^2 + \nu^T(\nabla^2 E + \epsilon\mu\omega^2 E) \tag{2}$$

The weight variable ν is referred to as the Lagrange multiplier or dual variable. The minimum value of $L(\epsilon,E,\nu)$, upon optimization over the design variable ϵ and field E, will vary as a function of ν.

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860–2871

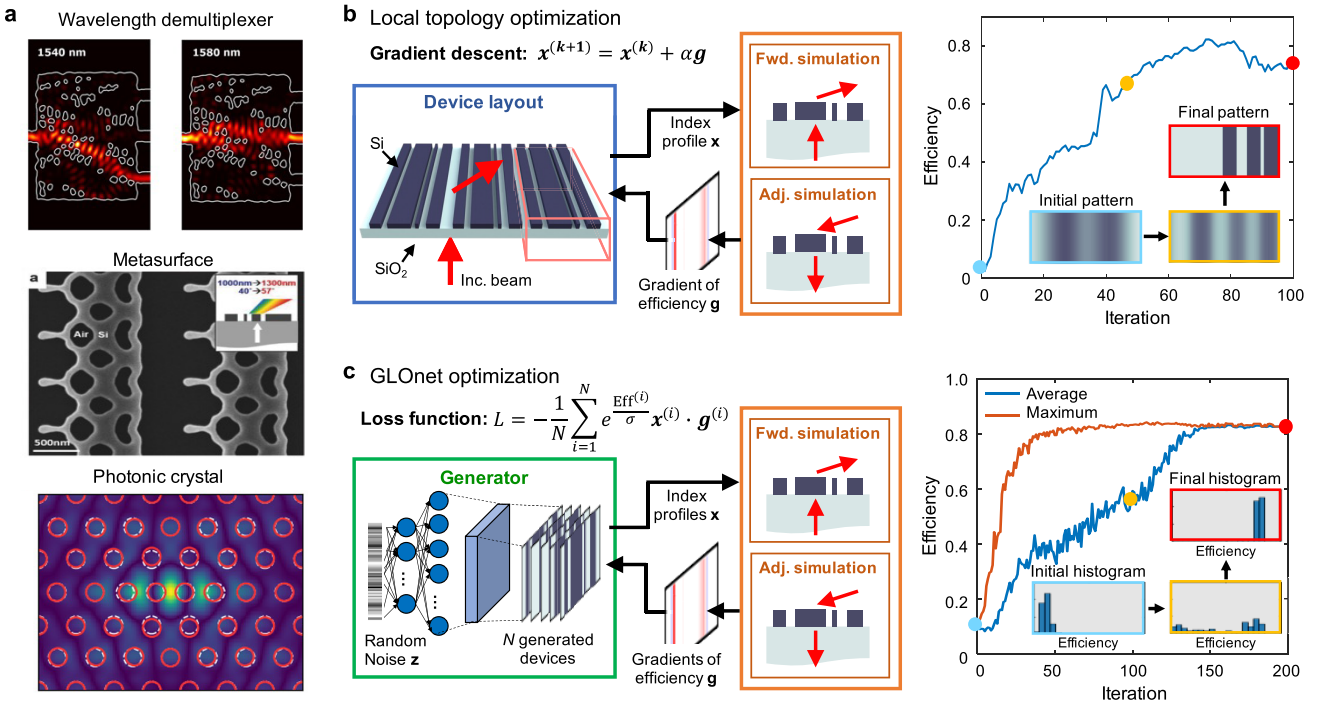


Figure 3. Freeform optimization of photonic devices. (a) Examples of freeform-optimized devices, including a chip-based wavelength demultiplexer, free space metagrating, and chip-based photonic crystal. (b) Schematic of a local freeform optimizer based on the adjoint variables method in which a dielectric distribution is iteratively updated using gradients computed from forward and adjoint simulations. FoM increases over the course of optimization, and constraints, such as the enforcement of binary dielectric values, are gradually incorporated. (c) Schematic of the global freeform optimization algorithm GLOnet, which frames the population-based search for the global optimum as the training of a generative deep network. Over the course of optimization, the figures of merit featured in the generated device distributions increase and converge to high values. (a) Adapted with permission from refs 31, 11, and 32. Copyright 2018, 2017, and 2020 American Chemical Society, Wiley Online Library, and American Chemical Society. (b) Adapted with permission from ref 19. Copyright 2020 The Optical Society. (c) Adapted with permission from ref 18. Copyright 2019 American Chemical Society.

The Lagrange dual function is defined as the infimum (i.e., greatest lower bound) of $L(\epsilon, E, \nu)$ as a function of ν

$$g(\nu) = \inf_{\epsilon,E} L(\epsilon, E, \nu) \tag{3}$$

$g(\nu)$ is a concave function even if the original problem is nonconvex. In addition, by definition

$$g(\nu) = \inf_{\epsilon,E} L(\epsilon, E, \nu) \leq \inf_{\nabla^2 E + \epsilon\mu\omega^2 E = 0} L(\epsilon, E, \nu)$$
$$= \inf_{\nabla^2 E + \epsilon\mu\omega^2 E = 0} \|E - \hat{E}\|_2^2 \tag{4}$$

and it therefore yields a lower bound on the optimal value of the original problem. Nontrivial bounds can therefore be calculated by solving the following convex optimization problem

$$\max_{\nu} g(\nu) \leq \min_{\nabla^2 E + \epsilon\mu\omega^2 E = 0} \|E - \hat{E}\|_2^2 \tag{5}$$

As a convex problem, optimization of the Lagrange dual function is tractable and straightforward using convex optimization algorithms. These computational bounds have been applied to a variety of wave-based systems, including Helmholtz resonators (Figure 2f)[24] and metasurface absorbers (Figure 2e).[25] We additionally note that the dual-function optimum ν^* can be used to find a good starting point $\epsilon_0 = \arg\min_\epsilon L(\epsilon, E, \nu^*)$ for optimization in the physical design space, as demonstrated for Helmholtz resonators.[24]

Research into electromagnetics performance bounds calculations is relatively nascent, and to date, nontrivial bounds have been derived for only a handful of photonic systems. These formalisms have nonetheless proven impactful in cases where the calculated bounds are tight. For example, bandwidth bounds for local achromatic metasurfaces have been important in discerning the physical limitations of broadband metalenses and their potential application in full color imaging systems. Looking ahead, there are many open research opportunities to develop new methods for computing tight bounds as these formalisms are problem specific and require detailed tailoring to ensure that the bounds are nontrivial. In the example of achromatic metasurfaces, new bounds formalisms are required for emergent classes of nonlocal volumetric devices, which have much greater capacity for large-aperture broadband operation and cannot be described with the local delay line approximation. A particular long-term challenge for all bounds formalisms involves the incorporation of manufacturing constraints, such as minimum feature size and sensitivity to random imperfections, which is critical when using these algorithms to guide the design of devices for experimental implementation.

FREEFORM OPTIMIZATION ALGORITHMS

Optical design is an optimization task in which the goal is to maximize a figure of merit (FoM) describing the desired optical response. It is particularly challenging because the feature space for photonics is large and nonconvex, and it can

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

be difficult to know where in the landscape to search. For simple low-dimensional problems, solutions can be identified through the brute force sweeping of candidate geometries or be optimized using one of several heuristic gradient-free optimization methods.[29,30] However, for systems where the dimensionality of the feature space is large and the FoM is complex and multiobjective, conventional approaches can become ineffective and computationally intractable. Fortunately for electromagnetics systems, computationally efficient methods have been developed to calculate gradients that specify how device parameters can be perturbed in a manner that improves the FoM, and these gradients can be used in an optimizer to efficiently search within a feature space. While gradient-based optimization cannot guarantee the discovery of exceptional devices in a nonconvex landscape, its ability to identify high-performing freeform devices with relatively few simulations makes them a clear solution for freeform device design.

There are two primary ways to calculate gradients for photonic systems. The most widely used method is the adjoint variables method,[13,33,34] in which gradients for every geometric feature in a device are calculated with a computational cost that is independent of the number of features. The concept was introduced to photonics by Ole Sigmund[35,36] and Eli Yablonovitch[37] in the early 2000s and has since been applied to a wide range of free space and chip-based photonic systems (Figure 3a).[11,31,32] To conceptualize the adjoint variables method, we can frame the effect of a geometric feature perturbation on the FoM as the excitation of polarization currents at the feature position and the measurement of the electric field at the FoM point of evaluation (i.e., spatial position in the near or far field corresponding to the FoM objective). With Lorentz reciprocity, this problem is equivalent to placing a current at the FoM point of evaluation and measuring the electric field at the feature position. In this so-called adjoint simulation, the electric field can be evaluated at all locations and be used to help compute gradients for all features in the system. The second method for calculating gradients is to use an electromagnetic simulator that can perform autodifferentiation.[38] The idea is that the relationship between device layout and electromagnetic fields can be described by an analytical and differentiable function, and changes in the field distribution can link with changes in the device layout by use of the chain rule. With autodifferentiation serving as a tool for neural network training in machine learning, there now are accessible and standardized software packages to adapting algorithms, including Maxwell solvers, to this framework.

Gradient calculations can be readily incorporated into gradient descent algorithms to perform local freeform optimization (Figure 3b).[9,19,39] An initial device layout with either randomly initialized geometric features or specified features based on an educated guess is used as a starting point,[40] and gradient calculations are used to iteratively perturb the layout in a manner that improves the FoM. Various modifications can be made to the algorithm to further shape the optimization process. Adaptive-learning rates can ensure that the gradient magnitudes are suitable and effective throughout the iterative optimization process.[41,42] In addition, the inclusion of momentum in the gradient descent algorithm can help the optimization trajectory escape regions of the feature space containing vanishing gradients.[43,44] While local gradient methods have benefits of straightforward implementation, they are ultimately local optimizers that search within limited regions of the feature space near the initial device layout.

Gradients can also be used to perform population-based searches for the globally optimal device in a design landscape. In a recent development, global topology optimization networks (GLOnet) was proposed in which the search for the global optimum is mediated by the training of a generative neural network.[18,45,46] A schematic of the algorithm is shown in Figure 3c. First, the network is initialized so that the input latent random variable maps to all possible outputted photonic devices. Upon sampling of the outputted device distribution, the device figures of merit and local gradients are calculated using a Maxwell solver, inputted into a special loss function, and backpropagated into the network. In performing this procedure in an iterative manner, the outputted device distribution gets pushed toward the global optimum during GLOnet training, and it ultimately collapses ideally at the global optimum upon completion of network training. GLOnet has been adapted for various photonics systems including metasurfaces and thin film stacks, and while it does not guarantee the global optimal device, it can yield exceptionally high-performing devices with capabilities beyond those designed using other methods.

A detailed discussion of how GLOnet works and why it is so effective is beyond the scope of this Perspective, but we summarize a few key points. First, as a population-based optimizer that reframes the optimization of physical device parameters to the optimization of weights in a high-dimensional neural network space, GLOnet smoothens the design landscape so that the distribution of outputted devices can effectively evolve without easily getting trapped in local optima. Second, GLOnet can readily scale to high-dimensional problems without a corresponding increase in computational resources because the batch size does not scale with problem dimensions. As such, GLOnet can be effectively and efficiently applied to problems featuring hundreds to thousands of dimensions. Third, there is a strong relationship between network architecture and algorithm capability, and proper exploitation of this relationship provides GLOnet the best opportunity to search for high-performance freeform devices. The presence of such a strong relationship also indicates that experience and hyperparameter tuning are required to properly maximize the potential of the algorithm.

For both local and global gradient-based optimizers, experimental material and device constraints can be directly incorporated to ensure that the final device designs are practically manufacturable. For final devices comprising discrete dielectric materials, level set formalisms can be used or regularization penalty terms can be added to the FoM that push continuous grayscale dielectric values to discrete values over the course of optimization.[18,39,47,48] Devices can be made robust to geometric variations by evaluating multiple devices with different perturbed boundary configurations for each iteration and using their collective gradients to update the design.[49,50] Minimum feature size constraints can be incorporated into the optimization algorithm using reparameterization, in which devices are represented in a continuous latent space and analytical mathematical transformations are used to transform those representations to constrained physical devices.[51] Gradients calculated on physical devices map back to their latent space representation using the chain rule. By explicitly baking in hard constraints to the device description

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

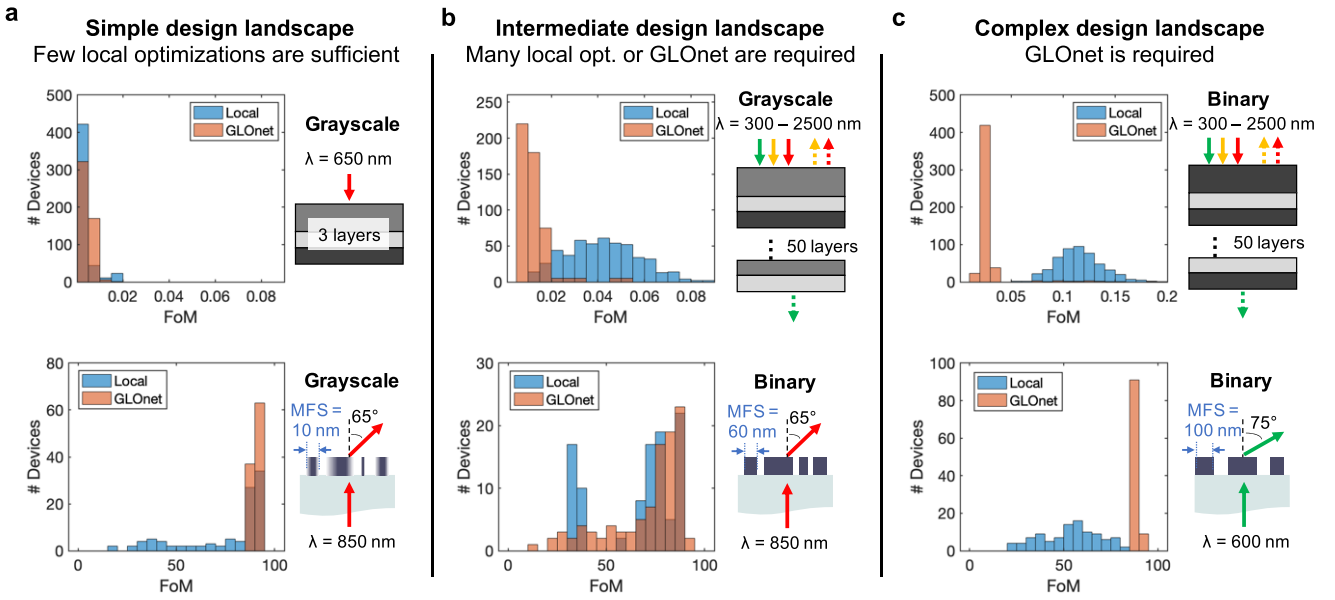


Figure 4. Benchmark of local and global freeform optimization algorithms for different design landscapes based on thin film stack filter (FoM to be minimized) and dielectric metagrating deflector (FoM to be maximized) model systems. (a) For simple design landscapes, including grayscale devices designed for simple objective functions, local freeform optimization is sufficient. (b) For intermediate design landscapes featuring more complex objective functions and device constraints, many local optimization attempts or global optimization is required. (c) For highly complex design landscapes, global freeform optimization is the only practical method for identifying high-performing devices. MFS: Minimum feature size.

itself, the optimization algorithms can more effectively search within a feature space comprising only suitable candidate devices.

The existence of a wide variety of optimization methods begs the following question: which method is the most effective and computationally efficient for a given problem? There is no general answer, and the choice and suitability of an optimization algorithm are highly dependent on the specific optimization objective, problem constraints, device form factor, and device material composition.[52] Some trends can be garnered from optimization histograms for representative thin film stack and metasurface systems, as shown in Figure 4. For the thin film stack system, the FoM is defined as the deviation between the actual and the desired device spectral profile and is to be minimized with a lower bound of zero, while the FoM for the metasurfaces is the device efficiency and is to be maximized to an upper bound of 100%. Grayscale thin film and metasurface systems with simple objectives appear to feature design landscapes that are particularly smooth with many good local optima, such that one or a few local gradient-based optimization runs are sufficient (Figure 4a). Thin film systems and high contrast metasurfaces featuring more complex objective functions have design landscapes that feature more variation in local optima performance, thereby requiring either GLOnet or a large ensemble of local optimization runs with different initializations (Figure 4b). High-dimensional, multiobjective problems have complex design landscapes featuring many poor local optima, in which case local gradient-based optimization is no longer effective and global optimization is required (Figure 4c).

We add a few additional thoughts on the optimization of freeform photonic devices. First, the design landscapes for photonic systems are not intuitive, and experience is the best way to deduce the suitability of a method for freeform optimization. Experience includes the running of local and global optimizers on representative model systems, with proper hyperparameter tuning, to provide benchmarks on the quality of local optima within the design landscape. It also includes quantitative comparisons of the performance of optimized devices with tight bounds calculations. Second, when appropriate and feasible, it always helps to relax the system parameters (i.e., dielectric value constraints) or add more design degrees of freedom (i.e., increase the size of the device) to modify the design landscape in a manner that improves the outcomes from gradient-based optimization. Third, while we anticipate that gradient-based optimization will be the basis for most problems in photonics, there are problems, such as the topology optimization of metal-based systems, where population-based gradient-free methods are a more viable solution. These problems are generally characterized as having high dimensionality, gradients that are vanishing or that provide limited information, landscapes that are highly irregular and discontinuous, and computationally expensive hyperparameter tuning.

■ ACCELERATION OF ELECTROMAGNETIC SOLVERS

A practical bottleneck of all optimization methods is the time and computational resources required to perform fullwave simulations. While many accurate time domain (i.e., FDTD and DGTD) and frequency domain (i.e., FDFD and FEM) solvers have been developed by the computational electromagnetics community, they all operate on the same premise: the computational problem is first set up from scratch, and then a series of calculations is used to produce fullwave solutions that are consistent with Maxwell's equations. For time domain simulations performed on standard computing processing units (CPUs), field updates are computed in each voxel for each time step and the total compute time scales linearly with the number of voxels.[53] For frequency domain simulations, Maxwell's equations are solved by inverting a large matrix and computing time scales as approximately the square

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

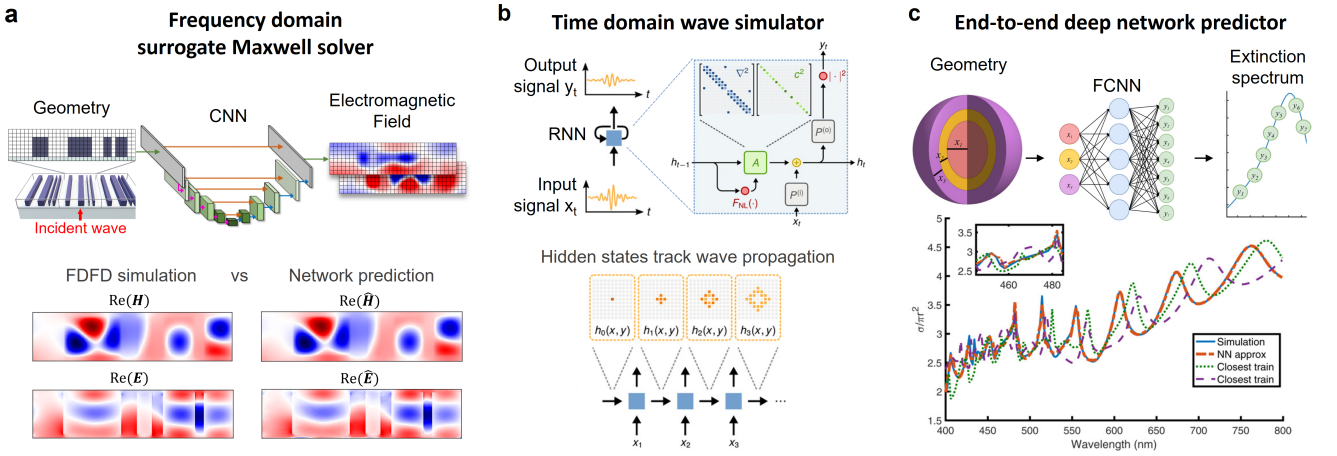


Figure 5. Accelerating electromagnetic wave solvers with deep neural networks. (a) Convolutional neural networks (CNNs) can serve as surrogate frequency domain solvers in which the input is the simulated geometry and the output is the electromagnetic field response. Predicted fields match well with those simulated with a fullwave solver. (b) Time domain wave phenomena can be modeled using a recurrent neural network (RNN) where the physical system operates on an input signal x and produces an output signal y. Hidden states $h(x, y)$ track wave propagation phenomena within the system. (c) Deep neural networks can be trained end-to-end to predict more general electromagnetic phenomena given an input geometry; in this example, the extinction spectra of concentric shell scatterers. (a) Adapted with permission from ref 20. Copyright 2022 ARXIV https://creativecommons.org/licenses/by/4.0/. (b) Adapted with permission from ref 60. Copyright 2019 American Association for the Advancement of Science https://creativecommons.org/licenses/by/4.0/. (c) Adapted with permission from ref 61. Copyright 2018 American Association for the Advancement of Science https://creativecommons.org/licenses/by/4.0/.

of the number of voxels.[54] These Maxwell solvers can be serially called hundreds to thousands of times in an iterative optimization algorithm, thereby taking hours to days to perform a single optimization run. Fortunately, innovations in computing hardware and software have led to new classes of Maxwell solvers with orders of magnitude faster speeds.

On the hardware front, the proliferation of graphics processing units (GPUs) has enabled new ways of parallelizing calculations within an electromagnetics simulation. GPUs are specialized computing hardware that were initially developed for computer graphics applications and have become mainstream hardware for machine-learning tasks. For FDTD algorithms, the field update at each voxel position is a local calculation involving only neighboring voxels, and the large-scale parallelization of these calculations with GPUs has enabled a 20× speedup in computing time.[55] For FDFD and RCWA frequency domain solvers, which solve Maxwell's equations through matrix computations, GPU-supported mathematical libraries and programming models[56,57] have vastly improved the speed and efficiency of matrix operations particularly for large matrices. In another recent development, broad access to servers and cloud computing resources have been game changing, as it now allows anyone with an Internet connection to have immediate access to thousands of CPUs. Access to these resources allows for the local optimization of multiple devices or for the evaluation of a batch of devices within a global optimization iteration to be performed in parallel. Individual simulations of large domains can also be accelerated using domain decomposition techniques in which the evaluation of the domain is performed by iteratively solving smaller subdomain problems in a parallelizable manner.[58,59]

On the software front, concepts in machine learning have emerged as new algorithmic approaches for accelerating the evaluation of Maxwell's equations. A key insight with these methods is that most electromagnetics problems are not arbitrary but are regularized: practical devices are made from a small selection of materials and have limited form factors, such as patterned thin films on a substrate. As the electromagnetic field solutions from these devices possess related features, structure−field relations from a subset of these devices can help inform the solution of a related unsolved problem. Importantly, training data representing these structure−field relations can be readily and accurately curated with fullwave solvers.

Surrogate Maxwell solvers in the frequency domain have been realized using convolutional neural networks (CNNs), which are deep networks developed in the computer vision community that specialize in image processing tasks (Figure 5a). In an initial demonstration, CNNs were implemented in which the inputs were images of the photonic devices and the outputs were maps of the full electromagnetic field responses.[62] The networks trained exclusively using simulated training data, and they could produce accurate field solutions for a variety of freeform nanoantenna devices. Physics-augmented CNNs have since been implemented as surrogate solvers in which the networks were trained with two loss terms: data loss specified by training data and physical loss in which the outputted fields were explicitly constrained to obey the wave equation.[20] These physical constraints were rigorously enforced with a finite-difference Maxwell formalism that specified the device dielectric distribution and electromagnetic fields on the Yee grid. The incorporation of physics in these networks led to smoother and more accurate field solutions, in part because the finite difference representation of the wave equation explicitly enforced physical relationships between neighboring voxels on a discrete field map. These surrogate simulators were trained to model dielectric metasurfaces and used in conjunction with the adjoint variables method and GLOnet to perform gradient-based optimization, where high-efficiency freeform devices were designed to be 3−4 orders magnitude faster than with conventional algorithms. The best devices from the optimization process exceeded the perform-

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

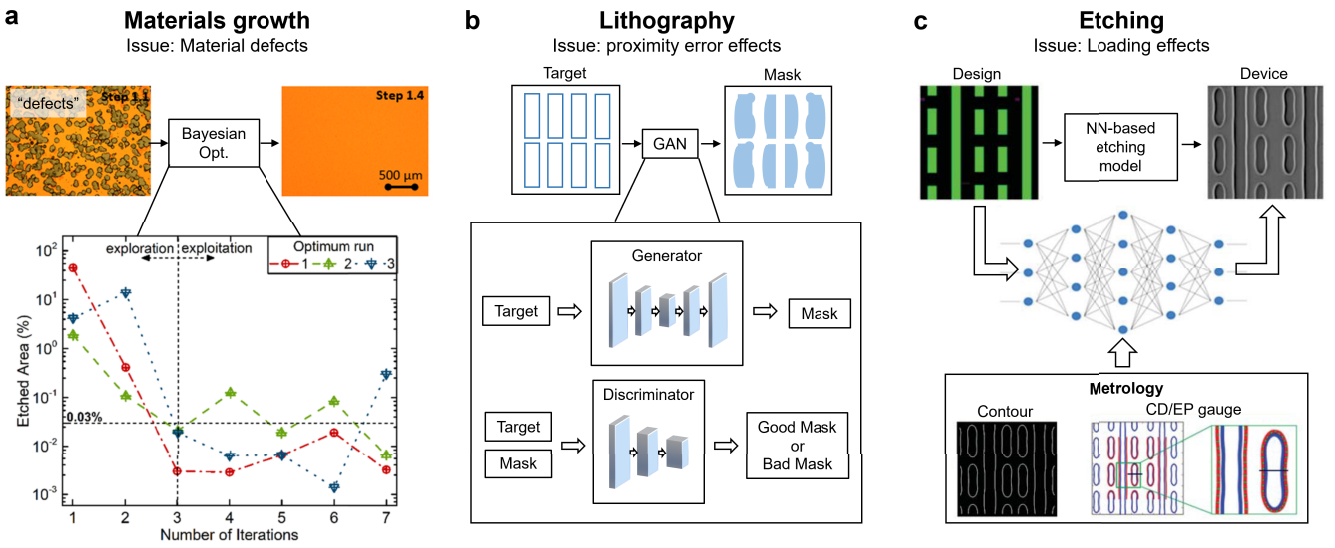


Figure 6. Algorithm-enhanced fabrication of freeform photonic devices. (a) Bayesian optimization can fine tune materials growth recipes to significantly reduce material defects. (b) Convolutional neural networks can be trained to reduce proximity error in the lithographic patterning of freeform shapes. (c) Deep networks can learn and mitigate loading effects during etching processes. (a) Adapted with permission from ref 75. Copyright 2021 American Chemical Society. (c) Adapted with permission from ref 76. Copyright 2020 International Society for Optics and Photonics.

ance of the best devices in the training set, indicating the generalization capabilities of these algorithms.

Recent research additionally indicates the potential of deep neural networks to accurately model the time domain electromagnetic behavior. In one study, it was shown that recurrent neural networks (RNNs) have the same algorithmic form as finite-difference wave solvers. With the RNN framework, the input signal is an incident source, the output signal is the system wave response at a point in space, and the hidden states embody the wave fields within the system (Figure 5b). These concepts were used to model and design physical systems supporting nonlinear scalar wave behavior.[60] Graphical neural networks (GNNs), which are an emergent class of neural network architecture that specialize in the processing of graphical data, have also been shown to effectively model wave phenomena in the time domain. They are a natural network architecture for wave modeling because the process of wave propagation, in which wave energy passes between neighboring voxels, can be readily modeled as energy flow between interconnecting nodes of a graph. Initial demonstrations of GNNs trained with simulation data indicate that these networks could serve as accurate surrogate time domain simulators for fluid mechanics systems and operate nearly an order of magnitude faster than conventional solvers.[63] More advanced algorithms utilizing transformers, in which correlations in time within data are learned, exhibit even more accurate wave simulation capabilities.[64] We anticipate that the application of these concepts to photonic systems and their incorporation of physics augmentation will enable accurate and accelerated electromagnetic time domain simulations.

Significant efforts have also been made to train deep networks end to end, where the input to the network is a device geometry and the output is a performance metric such as scattering spectrum, far-field profile, or efficiency[61,65,66] (Figure 5c). In these networks, the physical relationships between layout and optical performance are implicitly learned through the training process. A properly trained network can be evaluated with millisecond speeds and can be directly used in an optimization procedure for a desired performance metric using backpropagation, which is a gradient-based method in which the device geometry is iteratively modified to minimize error between the network output and the desired performance metric.[67,68] End-to-end networks have been used to successfully model a variety of nanostructure scatterers, wavelength demultiplexers, metamaterial unit cells, and thin film stack systems, and they have had general success in modeling systems described with a handful of free parameters.[69] However, it has proven challenging to apply these concepts to high-dimensional systems without the use of a tremendous amount of training data, in part because these networks attempt to capture more abstracted input−output relationships that cannot be explicitly regularized with physical constraints such as Maxwell's equations. These concepts are therefore well suited for low-dimensional problems, whereas the combination of surrogate Maxwell solvers with auxiliary physics-based algorithms, such as the adjoint variables method or near-to-far-field transformations, have more potential for deep network generalization to high-dimensional systems.

FABRICATION OF FREEFORM DEVICES

The final step to experimental device prototyping is fabrication, which can be performed using thin film processing methods based on CMOS chip fabrication or additive manufacturing. Thin film processing involves a sequence of steps in which the material is deposited, lithographically patterned, and etched, and multilayer devices can be achieved by repeating these steps together with planarization procedures. Microscale and nanoscale additive manufacturing typically use light to pattern a photosensitive material, and they include stereolithography[70] and two-photon polymerization[71] platforms. A principal challenge in fabricating freeform photonic devices with either of these methods is that the designs often demand the patterning of structures with disparate length scales or

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

curvilinear shapes, leading to particularly nontrivial manufacturing errors. In the case of thin film processing, manufacturing errors can arise in the form of defects in material deposition,[72] lithographic proximity error,[73] and loading effects in etching.[74] For all of these errors, the processing of an individual feature is sensitive to the presence and processing of disparate neighboring features. For additive manufacturing, proximity errors from patterning and mechanical stresses arising from patterning and postprocessing steps can lead to shape deformations in the final structure.

Algorithms that can accurately model and correct fabrication errors are therefore indispensable to ensuring the success of device manufacturing and minimizing the requirement of intensive process calibration for each device. Among the many efforts aimed at this task, deep neural networks have emerged as promising algorithms due to their ability to capture nonlinear relationships between desired and experimentally realized geometric features in a device. Importantly, deep networks can implicitly learn these relationships through experimental data and account for factors of unknown physical or chemical origin that can be impossible to explicitly capture through purely physical modeling. While many of these techniques were initially developed for high-throughput tool sets used in the semiconductor and integrated circuit industry, their versatility makes them readily adaptable to a broad range of fabrication tool sets.

In thin film processing, machine-learning methods have been successfully employed to improve materials growth, lithography, and etching steps. For materials growth, use of the data sciences to improve the quality of grown materials has become a field in itself.[77] To cite a representative example, Bayesian optimization can be used to identify material growth parameters in just a few trials that yield optimal films with virtually no porosity[75] (Figure 6a). Deep neural networks have become a standard tool to perform optical proximity correction (OPC) in the lithography process, in which lithography mask patterns are adjusted to improve the fidelity of the corresponding output patterns. In early examples, multilayer perceptron networks (MLPs)[78] and CNNs[79] were used to predict optimal mask layouts given the desired exposure pattern. Generative neural networks serving as an inverse design tool for lithography have also been trained in a generative adversarial network framework to produce candidate quasi-optimal masks for given target patterns[80] (Figure 6b). Concepts from OPC have been readily extended to the modeling of etching errors, where CNNs have been used to accurately model the etching masks for given target patte profiles of nanoscale features[76] (Figure 6c).

Machine-learning methods have also been developed for additive manufacturing processes to reduce geometric error in the final printed devices.[81] In a basic example, a MLP was trained to model thermal stress within the manufactured structure, from which modifications to the original design were made to compensate for thermal-driven geometrical deformations.[82] CNNs have also been shown to be capable of accurately predicting the shape of printed features as a function of their location, orientation, and geometric design.[83]

Looking ahead, there are many opportunities to further streamline the application of the data sciences to freeform device fabrication. One opportunity is to devise strategies for end-to-end algorithms that can account and correct for errors accumulating from multiple manufacturing steps. These algorithms not only are easier to implement than a suite of algorithms tailored to each step but also have the potential to produce better results because they can mitigate the accumulation of errors over the full serial fabrication flow. Another opportunity is to incorporate data-efficient machine-learning techniques, such as active learning and transfer learning, to limit the amount of data used for algorithm training.[84,85] The reality is that the nature of fabrication errors varies with changes to equipment, operating conditions, and sample type, and it would be impractical to procure a large data set of theoretical designs and actual layouts with each change. Active and transfer learning can allow relatively small data sets of theoretical designs and actual layouts to be used together with a precomputed data-intensive model to accelerate the adaptation of these algorithms to customized equipment and fabrication processes. We anticipate that new practices for integrating fabrication and metrology tools can accelerate and even automate the training of error correction algorithms in a manner that ensures the reliable and rapid prototyping of freeform photonic devices.

OUTLOOK

Algorithmic tools for optical engineering are changing the conceptual foundations for how optical system design is approached. Unlike traditional concepts based on intuitive physical models devised by people, algorithmic-based methods are able to identify new, nonintuitive structure−function solutions by treating the relationship between the device and the optical response as a physics-constrained data relationship. Data science algorithms are playing a large role in this evolution, as they are in many fields of research. However, they have proven to be particularly disruptive in photonics in part because the quantitative physical underpinnings of the field are well developed: fullwave Maxwell solvers effectively serve as ground truth oracles, and they can be used to compute structure−function solutions to expediently produce training data or rigorously calculate gradients. Data science algorithms that combine physics with data are particularly adept at learning and utilizing nonintuitive structure−function relationships while remaining grounded in reality with their explicit incorporation of physics.

Looking ahead, we anticipate many opportunities for algorithms to further enable new paradigms in optical engineering. One is the development of true multiscalar optical design and fabrication platforms that combine refractive, scalar diffractive, and fullwave optics. Optical systems at each length scale feature unique capabilities, and the combination of these different modalities can be synergistic and produce systems with exceptional bandwidth, aberration correction, and functional multiplexing properties. They also provide new avenues to bridging free space and chip-based optical systems. These concepts are complex to implement in part because of difficulties posed by the modeling and fabrication of optical systems featuring orders-of-magnitude different length scales and due to the siloing of optical technical expertise of a given length scale into distinct academic disciplines. Nonetheless, end-to-end analysis of true multiscalar systems is now possible on the algorithms front, where differential ray-tracing algorithms developed for refractive optics feature a compatible mathematical framework with fullwave simulation and design software based on autodifferentiation.[86,87] Recent work on the simulation and codesign of hybrid refractive and metasurface optical systems,[88−90] including those that support aberration-corrected

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

imaging functionality,[91] further highlights the promise of multiscalar optical design.

There are also substantial opportunities to connect optical engineering to physical domains outside of electromagnetics and even to software, which will lead to optical systems with completely new functionality. As algorithm-based optical design involves the maximization of a FoM, it is relatively straightforward to incorporate other physical metrics into the FoM involving a solid mechanics, fluid mechanics, electronic, chemical, or quantum-based objective to produce truly multifunctional systems designed for both optical and nonoptical capabilities. The evaluation and optimization of such figures of merit is further streamlined by the ability of gradient-based calculations, via the adjoint method or autodifferentiation, to be applied to these different physical domains. Software, specifically deep neural networks, can also be incorporated into systems to produce optical hardware−software systems with new capabilities[92,93] such as enhanced imaging properties, fast data processing, and low energy consumption. As both optical devices and neural networks can be individually designed using gradient-based optimization, it is straightforward to perform gradient-based end-to-end design of both the optical device and the neural network simultaneously.

There are still many innovations that will be required for these tools to be truly accepted and broadly used by the photonics community. First, the methods need to be scalable to large domains of arbitrary dimension without a loss in accuracy. This is a challenge for many algorithms, particularly data science algorithms where the domain dimensions are typically fixed and relatively small. Second, the interface between electromagnetics and the data sciences, including the relationship between scientific computing operations and neural network architecture, needs to continue to be refined. A recent example of such a development is the advent of Fourier neural operators,[94] which perform nonlinear processing of data in the Fourier space and are particularly well suited for processing wave-type data forms. Third, these algorithms need to be developed and packaged in a manner where they are openly available,[19] robust, and easy to use. Currently, many of these algorithms require the user to have highly specialized knowledge in the data sciences, optimization, or physics. Mass proliferation of these concepts and their adaptation to user-specific problems will ultimately require algorithms to run with automated and efficient hyperparameter tuning, neural network architecture specification, and active-learning modalities. Fourth, while many of these concepts offer black box-like solutions to device design and implementation, more foundational advances in optical engineering will be enabled by interpretable algorithmic approaches that help to connect nonintuitive structure−function solutions to broader physical insights.

AUTHOR INFORMATION

Corresponding Author

Jonathan A. Fan − *Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States;* orcid.org/0000-0001-9816-9979; Email: jonfan@stanford.edu

Authors

Mingkun Chen − *Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States;* orcid.org/0000-0001-9796-0021

Jiaqi Jiang − *Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States;* orcid.org/0000-0001-7502-0872

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphotonics.2c00612

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are thankful for support from the National Science Foundation under award no. 2103301, ARPA-E under award no. DE-AR0001212, the National Aeronautics and Space Administration under award no. 80NSSC21K0220, and the Office of Naval Research under Award Number N00014-20-1-2105. The authors acknowledge L. Gan for editorial assistance.

REFERENCES

(1) Deng, Q.; Yang, Y.; Gao, H.; Zhou, Y.; He, Y.; Hu, S. Fabrication of micro-optics elements with arbitrary surface profiles based on one-step maskless grayscale lithography. *Micromachines* **2017**, *8*, 314.

(2) Su, L.; Chen, Y.; Yi, A. Y.; Klocke, F.; Pongs, G. Refractive index variation in compression molding of precision glass optical components. *Appl. Opt.* **2008**, *47*, 1662−1667.

(3) Fourkas, J. T. Nanoscale photolithography with visible light. *J. Phys. Chem. Lett.* **2010**, *1*, 1221−1227.

(4) Farahani, R. D.; Chizari, K.; Therriault, D. Three-dimensional printing of freeform helical microstructures: a review. *Nanoscale* **2014**, *6*, 10470−10485.

(5) Zhou, Y.; Kravchenko, I. I.; Wang, H.; Nolen, J. R.; Gu, G.; Valentine, J. Multilayer noninteracting dielectric metasurfaces for multiwavelength metaoptics. *Nano Lett.* **2018**, *18*, 7529−7537.

(6) Wong, K. V.; Hernandez, A. A review of additive manufacturing. *Int. Scholarly Res. Not.* **2012**, *2012*, 208760.

(7) Frazier, W. E. Metal additive manufacturing: a review. *Journal of Materials Engineering and performance* **2014**, *23*, 1917−1928.

(8) Shah, Y. D.; Connolly, P. W.; Grant, J. P.; Hao, D.; Accarino, C.; Ren, X.; Kenney, M.; Annese, V.; Rew, K. G.; Greener, Z. M.; et al. Ultralow-light-level color image reconstruction using high-efficiency plasmonic metasurface mosaic filters. *Optica* **2020**, *7*, 632−639.

(9) Phan, T.; Sell, D.; Wang, E. W.; Doshay, S.; Edee, K.; Yang, J.; Fan, J. A. High-efficiency, large-area, topology-optimized metasurfaces. *Light Sci. Appl.* **2019**, *8*, 48.

(10) Devlin, R. C.; Khorasaninejad, M.; Chen, W. T.; Oh, J.; Capasso, F. Broadband high-efficiency dielectric metasurfaces for the visible spectrum. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 10473−10478.

(11) Sell, D.; Yang, J.; Doshay, S.; Fan, J. A. Periodic Dielectric Metasurfaces with High-Efficiency, Multiwavelength Functionalities. Advanced. *Opt. Mater.* **2017**, *5*, 1700645.

(12) Piggott, A. Y.; Lu, J.; Lagoudakis, K. G.; Petykiewicz, J.; Babinec, T. M.; Vučković, J. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nat. Photonics* **2015**, *9*, 374−377.

(13) Hughes, T. W.; Minkov, M.; Williamson, I. A.; Fan, S. Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photonics* **2018**, *5*, 4781−4787.

(14) Faraji-Dana, M.; Arbabi, E.; Arbabi, A.; Kamali, S. M.; Kwon, H.; Faraon, A. Compact folded metasurface spectrometer. *Nat. Commun.* **2018**, *9*, 4196.

(15) Wang, H.; Zhang, Y.; He, Y.; Zhu, Q.; Sun, L.; Su, Y. Compact silicon waveguide mode converter employing dielectric metasurface structure. *Adv. Opt. Mater.* **2019**, *7*, 1801191.
(16) Ma, T.; Yu, J.; Liang, P.; Wang, C. Design of a freeform varifocal panoramic optical system with specified annular center of field of view. *Opt. Express* **2011**, *19*, 3843−3853.
(17) Lee, G.-Y.; Hong, J.-Y.; Hwang, S.; Moon, S.; Kang, H.; Jeon, S.; Kim, H.; Jeong, J.-H.; Lee, B. Metasurface eyepiece for augmented reality. *Nat. Commun.* **2018**, *9*, 4562.
(18) Jiang, J.; Fan, J. A. Global optimization of dielectric metasurfaces using a physics-driven neural network. *Nano Lett.* **2019**, *19*, 5366−5372.
(19) Jiang, J.; Lupoiu, R.; Wang, E. W.; Sell, D.; Hugonin, J. P.; Lalanne, P.; Fan, J. A. MetaNet: a new paradigm for data sharing in photonics research. *Opt. Express* **2020**, *28*, 13670−13681.
(20) Chen, M.; Lupoiu, R.; Mao, C.; Huang, D.-H.; Jiang, J.; Lalanne, P.; Fan, J. A. WaveY-Net: Physics-augmented deep learning for high-speed electromagnetic simulation and optimization. *arXiv* **2022**, 2203.01248.
(21) Miller, O. D.; Johnson, S. G.; Rodriguez, A. W. Shape-independent limits to near-field radiative heat transfer. *Physical review letters* **2015**, *115*, 204302.
(22) Miller, O. D.; Hsu, C. W.; Reid, M. H.; Qiu, W.; DeLacy, B. G.; Joannopoulos, J. D.; Soljačić, M.; Johnson, S. G. Fundamental limits to extinction by metallic nanoparticles. *Physical review letters* **2014**, *112*, 123903.
(23) Presutti, F.; Monticone, F. Focusing on bandwidth: achromatic metalens limits. *Optica* **2020**, *7*, 624−631.
(24) Angeris, G.; Vuckovic, J.; Boyd, S. P. Computational bounds for photonic design. *ACS Photonics* **2019**, *6*, 1232−1239.
(25) Kuang, Z.; Zhang, L.; Miller, O. D. Maximal single-frequency electromagnetic response. *Optica* **2020**, *7*, 1746−1757.
(26) Kuang, Z.; Miller, O. D. Computational bounds to light-matter interactions via local conservation laws. *Phys. Rev. Lett.* **2020**, *125*, 263607.
(27) Yu, Z.; Raman, A.; Fan, S. Fundamental limit of nanophotonic light trapping in solar cells. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17491−17496.
(28) Angeris, G.; Vučković, J.; Boyd, S. Heuristic methods and performance bounds for photonic design. *Opt. Express* **2021**, *29*, 2827−2854.
(29) Jafar-Zanjani, S.; Inampudi, S.; Mosallaei, H. Adaptive genetic algorithm for optical metasurfaces design. *Sci. Rep.* **2018**, *8*, 11040.
(30) Shi, Y.; Li, W.; Raman, A.; Fan, S. Optimization of multilayer optical films with a memetic algorithm and mixed integer programming. *Acs Photonics* **2018**, *5*, 684−691.
(31) Su, L.; Piggott, A. Y.; Sapra, N. V.; Petykiewicz, J.; Vuckovic, J. Inverse design and demonstration of a compact on-chip narrowband three-channel wavelength demultiplexer. *Acs Photonics* **2018**, *5*, 301−305.
(32) Minkov, M.; Williamson, I. A.; Andreani, L. C.; Gerace, D.; Lou, B.; Song, A. Y.; Hughes, T. W.; Fan, S. Inverse design of photonic crystals through automatic differentiation. *Acs Photonics* **2020**, *7*, 1729−1741.
(33) Fan, J. A. Freeform metasurface design based on topology optimization. *MRS Bull.* **2020**, *45*, 196−201.
(34) Miller, O. D. *Photonic design: From fundamental solar cell physics to computational inverse design*; University of California: Berkeley, 2012.
(35) Borel, P. I.; Harpøth, A.; Frandsen, L. H.; Kristensen, M.; Shi, P.; Jensen, J. S.; Sigmund, O. Topology optimization and fabrication of photonic crystal structures. *Opt. Express* **2004**, *12*, 1996−2001.
(36) Jensen, J. S.; Sigmund, O. Topology optimization for nano-photonics. *Laser & Photonics Reviews* **2011**, *5*, 308−321.
(37) Burger, M.; Osher, S. J.; Yablonovitch, E. Inverse problem techniques for the design of photonic crystals. *IEICE Trans. Electron.* **2004**, *87*, 258−265.
(38) Hughes, T. W.; Williamson, I. A.; Minkov, M.; Fan, S. Forward-mode differentiation of Maxwell's equations. *ACS Photonics* **2019**, *6*, 3010−3016.
(39) Sell, D.; Yang, J.; Doshay, S.; Yang, R.; Fan, J. A. Large-angle, multifunctional metagratings based on freeform multimode geometries. *Nano Lett.* **2017**, *17*, 3752−3757.
(40) Yang, J.; Fan, J. A. Topology-optimized metasurfaces: impact of initial geometric layout. *Opt. Lett.* **2017**, *42*, 3161−3164.
(41) Zeiler, M. D. Adadelta: an adaptive learning rate method. *arXiv* **2012**, 1212.5701.
(42) Liu, L.; Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Han, J. On the variance of the adaptive learning rate and beyond. *arXiv* **2019**, 1908.03265.
(43) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, 1412.6980.
(44) Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning*; 2013; pp 1139−1147.
(45) Jiang, J.; Fan, J. A. Simulator-based training of generative neural networks for the inverse design of metasurfaces. *Nanophotonics* **2019**, *9*, 1059−1069.
(46) Jiang, J.; Fan, J. A. Multiobjective and categorical global optimization of photonic structures based on ResNet generative neural networks. *Nanophotonics* **2020**, *10*, 361−369.
(47) Doshay, S.; Sell, D.; Yang, J.; Yang, R.; Fan, J. A. High-performance axicon lenses based on high-contrast, multilayer gratings. *APL Photonics* **2018**, *3*, 011302.
(48) Sell, D.; Yang, J.; Wang, E. W.; Phan, T.; Doshay, S.; Fan, J. A. Ultra-High-Efficiency Anomalous Refraction with Dielectric Metasurfaces. *ACS Photonics* **2018**, *5*, 2402−2407.
(49) Wang, E. W.; Sell, D.; Phan, T.; Fan, J. A. Robust design of topology-optimized metasurfaces. *Optical Materials Express* **2019**, *9*, 469−482.
(50) Wang, F.; Jensen, J. S.; Sigmund, O. Robust topology optimization of photonic crystal waveguides with tailored dispersion properties. *JOSA B* **2011**, *28*, 387−397.
(51) Chen, M.; Jiang, J.; Fan, J. A. Design space reparameterization enforces hard geometric constraints in inverse-designed nanophotonic devices. *ACS Photonics* **2020**, *7*, 3141−3151.
(52) Yang, J.; Fan, J. A. Analysis of material selection on dielectric metasurface performance. *Opt. Express* **2017**, *25*, 23899−23909.
(53) Taflove, A.; Hagness, S. C. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*; Artech House, 2005.
(54) Jin, J.-M. *The finite element method in electromagnetics*; John Wiley & Sons, 2015.
(55) Bo, Z.; Zheng-hui, X.; Wu, R.; Wei-ming, L.; Xin-qing, S. Accelerating FDTD algorithm using GPU computing. *2011 IEEE International Conference on Microwave Technology & Computational Electromagnetics* **2011**, 410−413.
(56) Tong, J.; Chen, S. Computation improvement for the rigorous coupled-wave analysis with GPU. *2012 Fourth International Conference on Computational and Information Sciences* **2012**, 123−126.
(57) Xiong, Q.; Chen, S.; Hu, Y. Efficient computation scheme of grating structure using RCWA and FDFD. *2011 International Conference on Computational Problem-Solving (ICCP)* **2011**, 5−8.
(58) Vouvakis, M. N.; Cendes, Z.; Lee, J.-F. A FEM domain decomposition method for photonic and electromagnetic band gap structures. *IEEE Transactions on Antennas and Propagation* **2006**, *54*, 721−733.
(59) Zhao, K.; Rawat, V.; Lee, J.-F. A domain decomposition method for electromagnetic radiation and scattering analysis of multi-target problems. *IEEE Transactions on Antennas and Propagation* **2008**, *56*, 2211−2221.
(60) Hughes, T. W.; Williamson, I. A.; Minkov, M.; Fan, S. Wave physics as an analog recurrent neural network. *Sci. Adv.* **2019**, *5*, eaay6946.
(61) Peurifoy, J.; Shen, Y.; Jing, L.; Yang, Y.; Cano-Renteria, F.; DeLacy, B. G.; Joannopoulos, J. D.; Tegmark, M.; Soljačić, M.

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871

Nanophotonic particle simulation and inverse design using artificial neural networks. *Sci. Adv.* **2018**, *4*, eaar4206.
(62) Wiecha, P. R.; Muskens, O. L. Deep learning meets nanophotonics: a generalized accurate predictor for near fields and far fields of arbitrary 3D nanostructures. *Nano Lett.* **2020**, *20*, 329−338.
(63) Lino, M.; Cantwell, C.; Bharath, A. A.; Fotiadis, S. Simulating Continuum Mechanics with Multi-Scale Graph Neural Networks. *arXiv* **2021**, 2106.04900.
(64) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L. u.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inf. Process. Syst.* **2017**, 5998−6008.
(65) Malkiel, I.; Mrejen, M.; Nagler, A.; Arieli, U.; Wolf, L.; Suchowski, H. Plasmonic nanostructure design and characterization via deep learning. *Light Sci. Appl.* **2018**, *7*, 60.
(66) Ma, W.; Cheng, F.; Liu, Y. Deep-learning-enabled on-demand design of chiral metamaterials. *ACS Nano* **2018**, *12*, 6326−6334.
(67) Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. *Proceedings of the 36th International Conference on Machine Learning*; 2019; pp 7354−7363.
(68) Zhang, C.; Jin, J.; Na, W.; Zhang, Q.-J.; Yu, M. Multivalued neural network inverse modeling and applications to microwave filters. *IEEE Transactions on Microwave Theory and Techniques* **2018**, *66*, 3781−3797.
(69) Jiang, J.; Chen, M.; Fan, J. A. Deep neural networks for the evaluation and design of photonic devices. *Nat. Rev. Mater.* **2021**, *6*, 679−700.
(70) Tumbleston, J. R.; Shirvanyants, D.; Ermoshkin, N.; Janusziewicz, R.; Johnson, A. R.; Kelly, D.; Chen, K.; Pinschmidt, R.; Rolland, J. P.; Ermoshkin, A.; et al. Continuous liquid interface production of 3D objects. *Science* **2015**, *347*, 1349−1352.
(71) Cumpston, B. H.; Ananthavel, S. P.; Barlow, S.; Dyer, D. L.; Ehrlich, J. E.; Erskine, L. L.; Heikal, A. A.; Kuebler, S. M.; Lee, I.-Y. S.; McCord-Maughon, D.; et al. Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. *Nature* **1999**, *398*, 51−54.
(72) Wu, X.; Brown, L.; Kapolnek, D.; Keller, S.; Keller, B.; DenBaars, S.; Speck, J. Defect structure of metal-organic chemical vapor deposition-grown epitaxial (0001) GaN/Al2O3. *Journal of applied physics* **1996**, *80*, 3228−3237.
(73) Capodieci, L. From Optical Proximity Correction to Lithography-Driven Physical Design (1996−2006): 10 years of Resolution Enhancement Technology and the roadmap enablers for the next decade. *Opt. Microlithogr.* **2006**, *XIX*, 615401.
(74) Mogab, C. The loading effect in plasma etching. *J. Electrochem. Soc.* **1977**, *124*, 1262.
(75) Dogan, G.; Demir, S. O.; Gutzler, R.; Gruhn, H.; Dayan, C. B.; Sanli, U. T.; Silber, C.; Culha, U.; Sitti, M.; Schutz, G.; et al. Bayesian Machine Learning for Efficient Minimization of Defects in ALD Passivation Layers. *ACS Appl. Mater. Interfaces* **2021**, *13*, 54503−54515.
(76) Lu, Y.; Zhao, Y.; Li, M.; Yuan, W.; Peng, X.; Hu, H.; Yao, S.; Liu, Z.; Tian, Y.; Gao, Y.; et al. Accurate etch modeling with massive metrology and deep-learning technology. *Opt. Microlithogr.* **2020**, *XXXIII*, 113270B.
(77) Bhowmik, S.; Govind Rajan, A. Chemical Vapor Deposition of 2D Materials: A Review of Modeling, Simulation, and Machine Learning Studies. *iScience* **2022**, *25*, 103832.
(78) Luo, R. Optical proximity correction using a multilayer perceptron neural network. *Journal of Optics* **2013**, *15*, 075708.
(79) Watanabe, Y.; Kimura, T.; Matsunawa, T.; Nojima, S. Accurate lithography simulation model based on convolutional neural networks. *Opt. Microlithogr.* **2017**, *XXX*, 101470K.
(80) Yang, H.; Li, S.; Deng, Z.; Ma, Y.; Yu, B.; Young, E. F. GAN-OPC: Mask optimization with lithography-guided generative adversarial nets. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **2020**, *39*, 2822−2834.
(81) Wang, C.; Tan, X.; Tor, S.; Lim, C. Machine learning in additive manufacturing: State-of-the-art and perspectives. *Additive Manufacturing* **2020**, *36*, 101538.
(82) Chowdhury, S.; Mhapsekar, K.; Anand, S. Part build orientation optimization and neural network-based geometry compensation for additive manufacturing process. *J. Manuf. Sci. Eng.* **2018**, *140*, 031009.
(83) Baturynska, I.; Semeniuta, O.; Wang, K. Application of machine learning methods to improve dimensional accuracy in additive manufacturing. *International Workshop of Advanced Manufacturing and Automation*; Springer, 2018; pp 245−252.
(84) Shao, H.-C.; Ping, H.-L.; Chen, K.-s.; Su, W.-T.; Lin, C.-W.; Fang, S.-Y.; Tsai, P.-Y.; Liu, Y.-H. Keeping Deep Lithography Simulators Updated: Global-Local Shape-Based Novelty Detection and Active Learning. *arXiv* **2022**, 2201.09717.
(85) Lin, Y.; Li, M.; Watanabe, Y.; Kimura, T.; Matsunawa, T.; Nojima, S.; Pan, D. Z. Data efficient lithography modeling with transfer learning and active data selection. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **2019**, *38*, 1900−1913.
(86) Nimier-David, M.; Speierer, S.; Ruiz, B.; Jakob, W. Radiative backpropagation: adjoint method for lightning-fast differentiable rendering. *ACM Trans. Graphics (TOG)* **2020**, *39*, 146.
(87) Shi, R.; Kross, J. Differential ray tracing for optical design. *Proc. SPIE* **1999**, *3737*, 149−160.
(88) Cuillerier, A. C.; Borne, J.; Dallaire, X.; Thibault, S. Toward hybrid refractive and metalens design. *Int. Opt. Des. Conf.* **2021**, 120781W.
(89) Nikolov, D. K.; Bauer, A.; Cheng, F.; Kato, H.; Vamivakas, A. N.; Rolland, J. P. Metaform optics: bridging nanophotonics and freeform optics. *Sci. Adv.* **2021**, *7*, eabe5112.
(90) Kamali, S. M.; Arbabi, A.; Arbabi, E.; Horie, Y.; Faraon, A. Decoupling optical function and geometrical form using conformal flexible dielectric metasurfaces. *Nat. Commun.* **2016**, *7*, 11618.
(91) Chen, W. T.; Zhu, A. Y.; Sisler, J.; Huang, Y.-W.; Yousef, K. M.; Lee, E.; Qiu, C.-W.; Capasso, F. Broadband achromatic metasurface-refractive optics. *Nano Lett.* **2018**, *18*, 7801−7808.
(92) Metzler, C. A.; Ikoma, H.; Peng, Y.; Wetzstein, G. Deep optics for single-shot high-dynamic-range imaging. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020; pp 1375−1385.
(93) Wetzstein, G.; Ozcan, A.; Gigan, S.; Fan, S.; Englund, D.; Soljačić, M.; Denz, C.; Miller, D. A.; Psaltis, D. Inference in artificial intelligence with deep optics and photonics. *Nature* **2020**, *588*, 39−47.
(94) Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier neural operator for parametric partial differential equations. *arXiv* **2020**, 2010.08895.

https://doi.org/10.1021/acsphotonics.2c00612
ACS Photonics 2022, 9, 2860−2871