

# Global Optimization of Dielectric Metasurfaces Using a Physics-Driven Neural Network

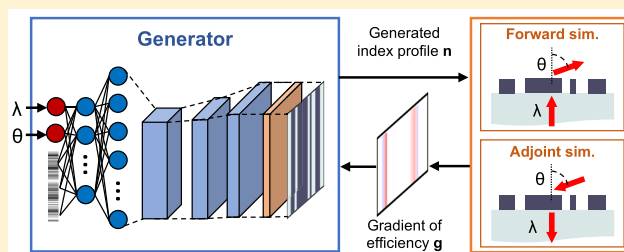
Jiaqi Jiang<sup>1</sup> and Jonathan A. Fan<sup>1\*</sup>

Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States

## Supporting Information

**ABSTRACT:** We present a global optimizer, based on a conditional generative neural network, which can output ensembles of highly efficient topology-optimized metasurfaces operating across a range of parameters. A key feature of the network is that it initially generates a distribution of devices that broadly samples the design space and then shifts and refines this distribution toward favorable design space regions over the course of optimization. Training is performed by calculating the forward and adjoint electromagnetic simulations of outputted devices and using the subsequent efficiency gradients for backpropagation. With metagratings operating across a range of wavelengths and angles as a model system, we show that devices produced from the trained generative network have efficiencies comparable to or better than the best devices produced by adjoint-based topology optimization, while requiring less computational cost. Our reframing of adjoint-based optimization to the training of a generative neural network applies generally to physical systems that can utilize gradients to improve performance.

**KEYWORDS:** Global optimization, generative neural networks, machine learning, adjoint variable method, dielectric metasurfaces, metagrating



Metasurfaces are subwavelength-structured artificial media that can shape and localize electromagnetic waves in unique ways.<sup>1–3</sup> These technologies are useful in imaging,<sup>4–6</sup> sensing,<sup>7</sup> and optical information processing applications,<sup>8</sup> among others, and can operate at wavelengths spanning the ultraviolet to radio frequencies.<sup>9–11</sup> A central research thrust in the field has been the identification of effective and computationally efficient ways to design high-performance metasurfaces, given a desired electromagnetic response.<sup>12</sup> In this aim, inverse design based on optimization has shown great promise. These methods range from heuristic swarm<sup>13</sup> and genetic algorithms<sup>14,15</sup> to adjoint-based topology optimization,<sup>16,17</sup> and they have led to metagratings,<sup>18,19</sup> metasurfaces,<sup>20–22</sup> and other nanophotonic devices<sup>23–25</sup> with exceptional performance. However, they are computationally costly, making it difficult and even intractable to scale these methods to large ensembles of devices or large area devices.

To address this computational roadblock, concepts in machine learning that augment the device design process have been investigated.<sup>26–28</sup> In the current manifestations of machine learning-enabled photonics design, a training set of device geometries and their associated optical properties is first produced. These data are then used to train a neural network, which “learns” the relationship between device geometry and optical response. A properly trained network can then produce new device designs beyond the training data set at low computational cost. To date, a range of machine learning concepts, including deep neural networks with fully connected

networks, convolutional networks, and generative adversarial networks (GANs), have been proposed.<sup>29–31</sup>

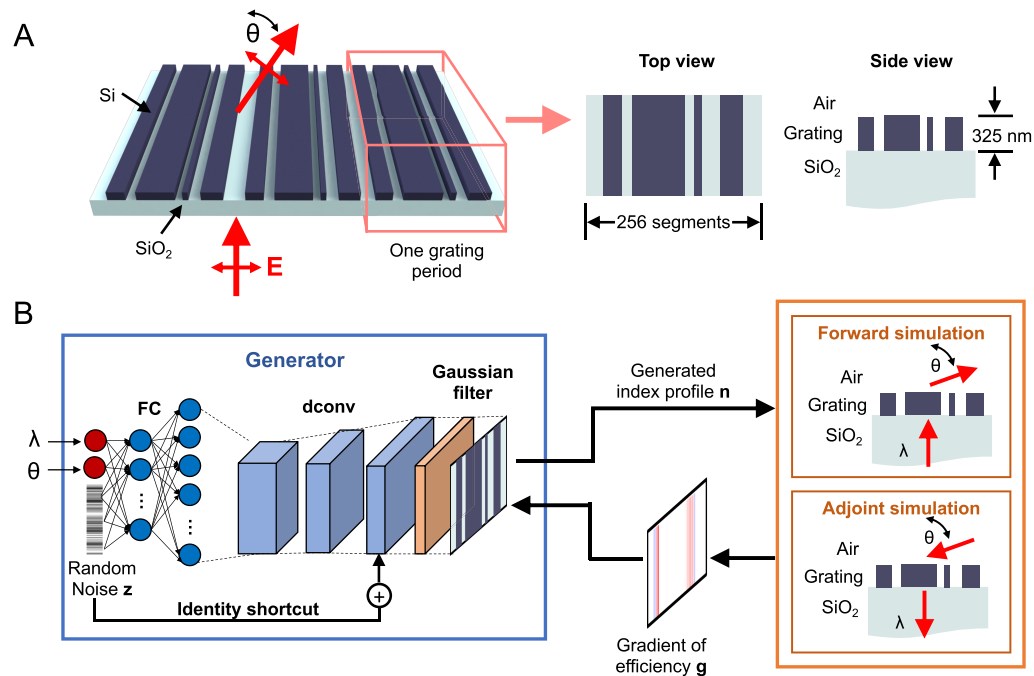
These initial demonstrations show that neural networks have the potential to learn the relationship between structural geometry and optical response, but they also highlight key challenges to the approach.<sup>26,29</sup> One challenge is that the computational cost of creating the training data set itself can be immense. Networks for structures described by even a few geometric parameters require tens to hundreds of thousands of devices for training. GAN-based design strategies have the potential to work with relatively less training data,<sup>31</sup> but these data are preoptimized and are computationally costly to produce. Another challenge is that many nanophotonic devices have complex geometries and reside in a high dimensional design space, making it difficult for even the best networks to learn the nuanced relationship between device geometry and response. New approaches that extend beyond standard machine learning approaches are required for neural networks to be practically useful in the electromagnetics design process.

In this Letter, we introduce a new concept in electromagnetic device design by incorporating adjoint variable calculations directly into generative neural networks. Termed global topology optimization networks (GLOnets), our approach is capable of generating high-performance top-

**Received:** May 6, 2019

**Revised:** July 5, 2019

**Published:** July 11, 2019



**Figure 1.** Global optimization based on a generative neural network. (A) Schematic of a silicon metagrating that deflects normally incident TM-polarized light to the outgoing angle  $\theta$ . The metagrating consists of 325 nm thick Si ridges in air on a  $\text{SiO}_2$  substrate. In the generative neural network, the device is specified by a  $1 \times 256$  vector,  $\mathbf{n}$ , which represents the refractive index profile of one period of the grating. (B) Schematic of the conditional GLOnet for metagrating generation. The generator is built on fully connected layers (FC), deconvolution layers (dconv), and a Gaussian filter. An identity shortcut connection is also used and adds  $\mathbf{z}$  to the output of last deconvolution layer. The input is the device wavelength  $\lambda$ , deflection angle is  $\theta$ , a 256-dimensional noise vector of  $\mathbf{z}$ , and the output is the device vector  $\mathbf{n}$ . During each iteration of training, a batch of devices is generated and efficiency gradients  $\mathbf{g}$  are calculated for each device using forward and adjoint electromagnetic simulations. These gradients are backpropagated through the network to update the weights of the neurons.

ology-optimized devices spanning a range of operating parameters with modest computational cost. GLOnets work by initially evaluating a distribution of devices spanning the design space and then continuously optimizing this device distribution until it converges to a cluster of high-efficiency devices. Physics-based gradients are utilized for backpropagation to ensure that network training is directly tied with enhancing device efficiency. Unlike other manifestations of machine learning-enabled photonics design, our approach does not use or require a training set of known devices but instead learns the physical relationship between device geometry and response directly through electromagnetic simulations. We note that our network performs a global search for the globally optimal device within the design space, but it does not guarantee that the final generated devices are globally optimal. In general, it is not possible to guarantee globally optimal solutions in nonconvex optimization problems, including our problem.

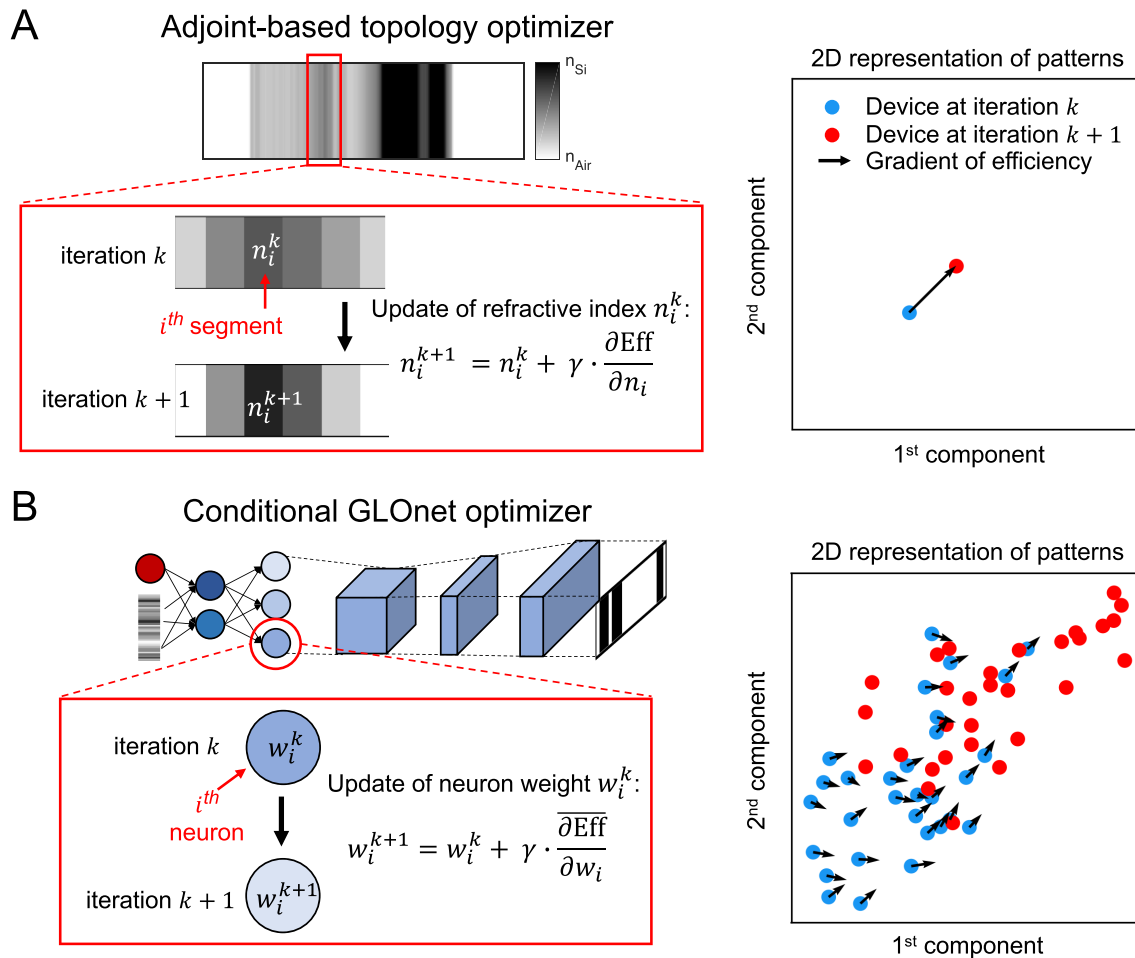
**Methods.** For this study, we will focus on conditional GLOnets that have wavelength and deflection angle as inputs, and we will simultaneously design an ensemble of silicon metagratings that operate across a range of wavelengths and deflection angles. This concept builds on our analysis of unconditional GLOnets,<sup>32</sup> which can optimize only a single device in a training session. Our metagratings consist of silicon nanoridges and deflect normally incident light to the +1 diffraction order (Figure 1A). The thickness of the gratings is fixed to be 325 nm and the incident light is TM-polarized. For each device, the metagrating period is subdivided into  $N = 256$  segments and each segment possesses a refractive index value between silicon and air. These refractive index values are the

design variable in our problem and are specified as  $\mathbf{n}$  (a  $1 \times N$  vector). Index values in the vector are normalized to a range of  $-1$ , which represents air, and  $+1$ , which represents silicon. The optimization objective is to maximize the deflection efficiency of the metagrating given an operating wavelength ranging from 600 to 1300 nm and an outgoing angle ranging from 40 to 80°.

A schematic of our conditional GLOnet is presented in Figure 1B. The input is the operating wavelength  $\lambda$ , the desired outgoing angle  $\theta$ , and an  $N$ -dimensional noise vector  $\mathbf{z}$ , which is a uniformly distributed random variable. The output is the refractive index profile of the device,  $\mathbf{n}$ . The weights of the neurons are parametrized as  $\mathbf{w}$ . The generator, conditioned on  $(\lambda, \theta)$ , maps different  $\mathbf{z}$  onto different device instances:  $\mathbf{n} = G_{\mathbf{w}}(\mathbf{z}; \lambda, \theta)$ . The ensemble of all possible  $\mathbf{z}$  and corresponding  $\mathbf{n}$ , given  $(\lambda, \theta)$  as inputs, are denoted as  $\{\mathbf{z}\}$  and  $\{\mathbf{n} | \lambda, \theta\}$ , respectively.

An important feature of our use of neural networks is that we can readily incorporate layers of neurons at the output of the network that can perform mathematical operations on the outputted device. In our case, we set the last layer of the generator to be a Gaussian filter, which eliminates small, pixel-level features (Figure S2) that are impractical to fabricate. The only constraint with these mathematical operations is that they need to be differentiable, so that they support backpropagation during network training.

Proper network initialization is required to ensure that the network at the start of training maps the noise vectors  $\{\mathbf{z}\}$  to the full design space. We take two steps to initialize our network. First, we randomly assign the weights in the network with small values,<sup>33</sup> which sets the outputs of our last deconvolution layer to be close to 0. Second, we directly add



**Figure 2.** Comparison between adjoint-based topology optimization and conditional GLOnet optimization. (A) Adjoint-based topology optimization uses efficiency gradients from an individual device to improve its performance within the local design space. A visualization of the device in a 2D representation of the design space illustrates that from iteration  $k$  to  $k + 1$ , the device moves incrementally to a nearby local maxima, indicated by its local gradient. (B) Conditional GLOnets use a neural network to map random noise to a distribution of devices. Gradients of efficiency, averaged over a batch of devices, are backpropagated to update the weights of the neurons and deconvolution kernels, which improves the average efficiency of the generated device distribution. A visualization of the device distribution illustrates that from iteration  $k$  to  $k + 1$ , the efficiency gradients from individual devices (black arrows) are used to collectively bias the device distribution toward high-efficiency regions of the design space.

the noise vector  $\mathbf{z}$  to the output of last deconvolution layer using an “identity shortcut”.<sup>34</sup> To facilitate this second step, it is important that the dimensionality of  $\mathbf{z}$  matches with  $\mathbf{n}$ . In combining these two initialization steps, we get that the initial ensemble of all possible generated device instances  $\{\mathbf{n} | \lambda, \theta\}$  has approximately the same distribution as the ensemble of noise vectors  $\{\mathbf{z}\}$ , and it therefore spans the full device design space.

During network training, the objective is to iteratively optimize  $\mathbf{w}$  to maximize the efficiencies of  $\{\mathbf{n} | \lambda, \theta\}$  for all possible  $(\lambda, \theta)$  within the target range. In other words, we aim to maximize the probability of generating high-efficiency devices. An ideal, perfectly trained network would map  $\{\mathbf{z}\}$  to  $\{\mathbf{n} | \lambda, \theta\}$  containing only the globally optimized device. To improve  $\mathbf{w}$  with each iteration, a batch of  $M$  devices,  $\{\mathbf{n}^{(m)} | m = 1, 2, \dots, M\}$ , is initially generated by sampling  $\mathbf{z}$  from the noise vector distribution,  $\lambda$  from the target wavelength range, and  $\theta$  from the target outgoing angle range. Our loss function for the conditional generator is as follows:

$$L = -\frac{1}{M} \sum_{m=1}^M \exp\left(\frac{\text{Eff}^{(m)} - \text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})}{\sigma}\right) \mathbf{n}^{(m)} \cdot \mathbf{g}^{(m)} \quad (1)$$

For the  $m$ th device, the gradients of efficiency with respect to  $\mathbf{n}^{(m)}$ , denoted by  $\mathbf{g}^{(m)}$ , specifies how the device refractive indices can be modified to improve the efficiencies. These efficiency gradients  $\mathbf{g}^{(m)}$  are calculated using the adjoint variables method<sup>19,24</sup> and are calculated from electric and magnetic field values taken from forward and adjoint electromagnetic simulations. The gradients we use in this work are taken from those we previously developed for the topology optimization of metagratings.<sup>16,17,19,35</sup>

The term  $\text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})$  is the theoretical maximum efficiency for each wavelength and angle pair. In practice,  $\text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})$  is unknown, as it represents the efficiencies of the globally optimal devices for which we are trying to solve. Over the course of network training, we estimate  $\text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})$  to be the highest efficiency calculated from the batches of generated devices.  $\text{Eff}^{(m)}$  is the efficiency of the  $m$ th device and is directly calculated with the forward electromagnetic

simulation. The expression  $\exp\left(\frac{\text{Eff}^{(m)} - \text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})}{\sigma}\right)$  represents a bias term that preferentially weighs higher efficiency devices during network training and reduces the impact of low efficiency devices that are potentially trapped in undesirable local optima. The magnitude of this efficiency biasing term can be tuned with the hyperparameter  $\sigma$ . A detailed derivation of the loss function from first-principles can be found elsewhere<sup>32</sup> and in the [Supporting Information](#).

The gradient of the loss function with respect to the indices for the  $m$ th device is  $\frac{\partial L}{\partial \mathbf{n}^{(m)}} = -\frac{1}{M} \exp\left(\frac{\text{Eff}^{(m)} - \text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})}{\sigma}\right) \mathbf{g}^{(m)}$ . In this form, minimizing the loss function  $L$  is equivalent to maximizing the device efficiencies in each batch. To train the network and update  $\mathbf{w}$ , we use backpropagation to calculate  $\frac{\partial L}{\partial \mathbf{w}} = \sum_{m=1}^M \frac{\partial L}{\partial \mathbf{n}^{(m)}} \cdot \frac{\partial \mathbf{n}^{(m)}}{\partial \mathbf{w}}$  each iteration.

To ensure that the generated devices are binary (i.e., all silicon or air), we add  $-\ln^{(m)} |2 - \ln^{(m)}|$  as a regularization term to the loss function. This term reaches a minimum when  $|\mathbf{n}^{(m)}| = 1$  and the device segments are either silicon or air. This binarization condition serves as a design constraint that limits metagrating efficiency, as the efficiency enhancement term (eq 2) favors grayscale patterns. To balance binarization with efficiency enhancement in the loss function, we include the tunable hyperparameter  $\beta$ . Our final expression for the loss function is:

$$L = -\frac{1}{M} \sum_{m=1}^M \left[ \exp\left(\frac{\text{Eff}^{(m)} - \text{Eff}_{\max}(\lambda^{(m)}, \theta^{(m)})}{\sigma}\right) \mathbf{n}^{(m)} \cdot \mathbf{g}^{(m)} + \beta \ln^{(m)} |2 - \ln^{(m)}| \right] \quad (2)$$

We can view conditional GLOnets, in which the nonlinear mapping between  $(\mathbf{z}, \lambda, \theta)$  and device layout is iteratively improved using physics-driven gradients, as a reframing of the adjoint-based optimization process. We want to be clear, however, that conditional GLOnets are qualitatively different from adjoint-based topology optimization. To conceptualize these differences, we discuss each optimization strategy in more detail. Adjoint-based topology optimization applies to a single device and is a local optimizer. The algorithm takes an initial dielectric distribution and enhances its efficiency by adjusting its refractive indices at each segment using gradient ascent (Figure 2A). This method is performed iteratively until the device reaches a local maximum in the design space. The performance of the final device strongly depends on the choice of initial dielectric distribution.<sup>36</sup> More of the design space can be explored with this approach by performing topology optimization on many devices, each with different initial dielectric distributions. Devices that happen to have initial dielectric distributions in favorable design space regions will become high performing.

Local optimizers are an effective tool to designing a wide range of photonic devices. However, their usage is accompanied by a number of caveats. First, they require significant computational resources. Hundreds of electromagnetic simulations are required to topology optimize a single device, and for multifunctional devices this number of simulations can scale to very large numbers. Second, the sampling of the design space is limited to the total number of devices being optimized. For complex devices described by a very high-dimensional

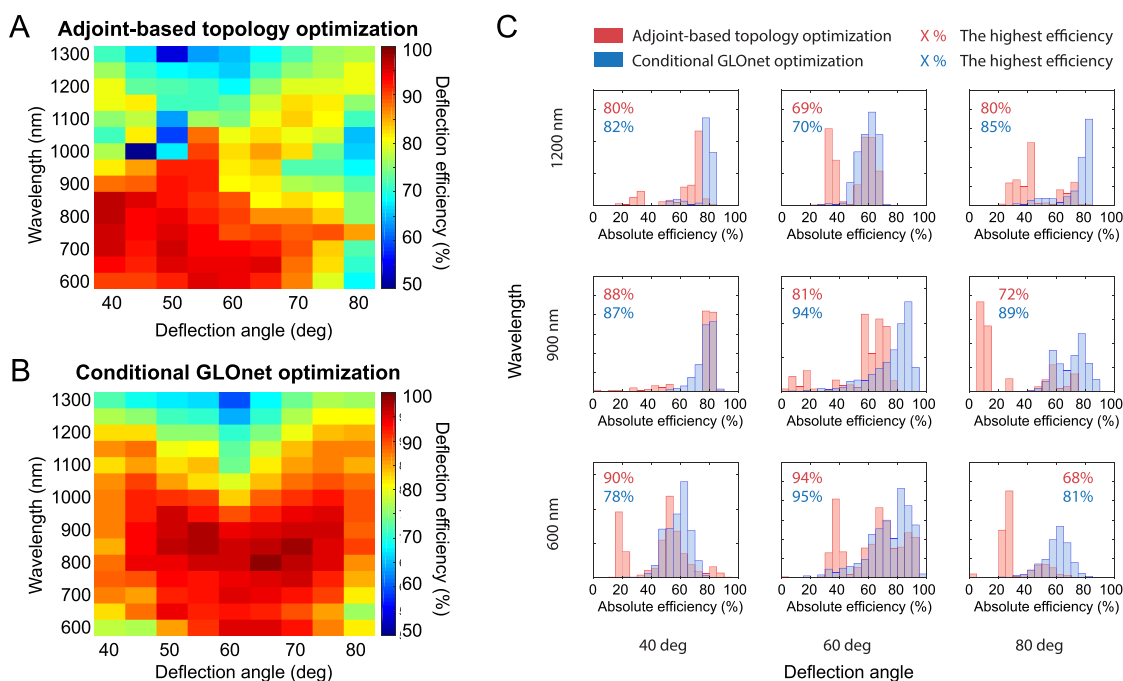
design space, the required sampling may be extremely large. Third, devices are optimized independently of one another, and gradient information from one device does not impact other devices. Fourth, multiple topology optimizations are required to produce different devices with different operating parameters.

Conditional GLOnets are qualitatively different in that they optimize an entire distribution of device instances, as mediated by the noise vector  $\mathbf{z}$ . The starting point of each conditional GLOnet iteration is similar to adjoint optimization and involves the calculation of efficiency gradients for individual devices using the adjoint method. However, the difference arises when these gradients are backpropagated into the network. Consider a single device defined by inputs  $(\mathbf{z}, \lambda, \theta)$  that produces a gradient  $\mathbf{g}$  for network backpropagation: all the weights in the network get updated, thereby modifying the complete mapping of  $\{\mathbf{z}\}$  to  $\{\mathbf{n} | \lambda, \theta\}$  (Figure 2B). This points to the presence of crosstalk between all device instances during the network learning process. Crosstalk can be useful when devices in promising parts of the design space bias the overall distribution of device instances to these regions. Regulation of the amount of crosstalk between devices, which is important to stabilizing the optimization method, is achieved from the nonlinearity intrinsic to the neural network itself.

Another advantage of our conditional GLOnet is that it is effective at globally surveying the design space, enhancing the probability that optimal regions of the design space are sampled and exploited. Such global surveying is made possible in part because the initial network maps  $\{\mathbf{z}\}$  onto the full device design space, and in part because different  $(\mathbf{z}, \lambda, \theta)$  are sampled each iteration, leading to the cumulative sampling of different regions of the design space during training. Conditional GLOnets also enable the simultaneous optimization of devices designed for operating parameters that span a broad range of values, over a single network training session. For our metagratings, these parameters are the outgoing angle and wavelength, but they can generally involve any combination of design parameters in the problem including device thickness, refractive index, or light polarization, among others. This codesign leads to a substantial reduction in computation time per device, which is because these devices operate with related physics and strongly benefit from crosstalk from the network training process.

**Results and Discussion.** To benchmark devices designed from our conditional GLOnet, we first perform adjoint-based topology optimization on metagratings operating across our desired range of wavelengths and angles. Details pertaining to this calculation can be found elsewhere.<sup>27,28</sup> These devices operate across a wavelength range between 600 and 1300 nm in increments of 50 nm and across a deflection angle range between 40 and 80° in increments of 5°. For each wavelength and angle pair, we optimize 500 devices, each with random grayscale patterns serving as initial dielectric distributions. A total of 200 iterations is performed for each optimization, and the deflection efficiencies of the optimized devices are calculated using a rigorous coupled-wave analysis (RCWA) solver.<sup>37</sup> The efficiencies of the best device for each wavelength and angle pair are plotted in Figure 3A.

With our fully trained conditional GLOnet, we generate 500 devices for each wavelength and angle pair by fixing  $(\lambda, \theta)$  at the network input and sampling  $\mathbf{z}$  500 times. Details pertaining to the architecture and training parameters are in the [Supporting Information](#). The efficiencies of the best devices



**Figure 3.** Performance comparison of adjoint-based topology optimization and conditional GLOnet optimization. (A) Plot of metagrating efficiency for devices operating with different wavelength and angle values, designed using adjoint-based topology optimization. For each wavelength and angle combination, 500 individual optimizations are performed and the highest efficiency device is used for the plot. (B) Plot of metagrating efficiency for devices designed using the conditional GLOnet. For each wavelength and angle combination, 500 devices are generated and the highest efficiency device is used for the plot. (C) Efficiency histograms of devices designed using adjoint-based topology optimization (red) and conditional GLOnet optimization (blue). The highest device efficiencies in each histogram are also displayed. For most wavelength and angle values, the efficiency distributions from the conditional GLOnet are narrower and have higher maximum values compared to those from adjoint-based topology optimization.

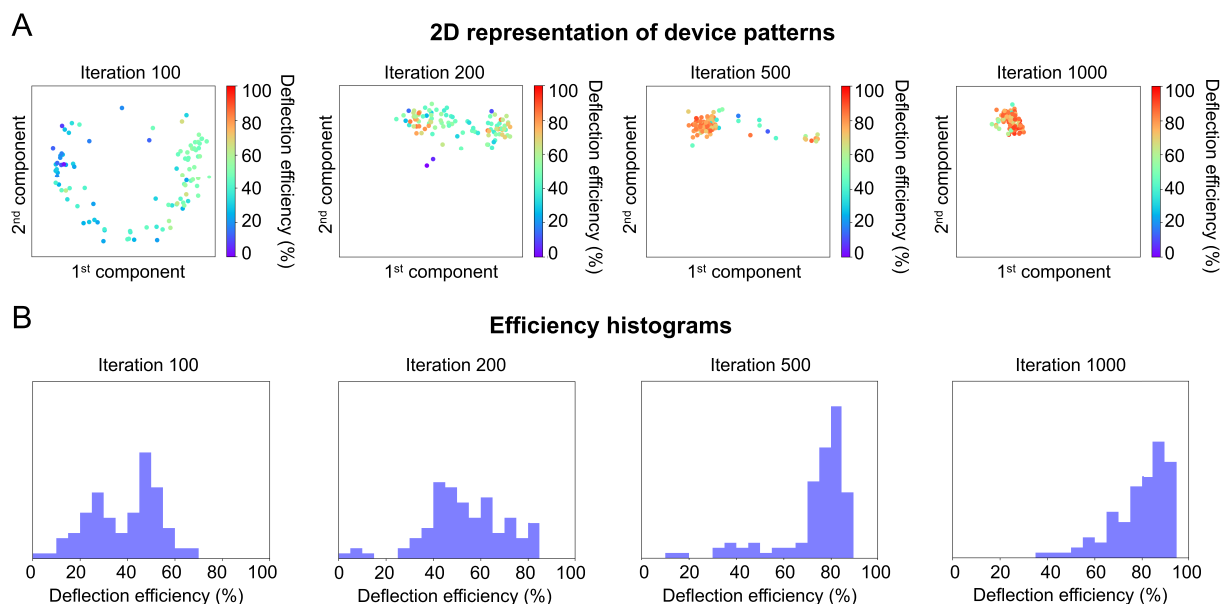
for the same wavelengths and deflection angles displayed in Figure 3A are plotted in Figure 3B. These efficiency values indicate that the best devices from the conditional GLOnet compare well with or are better than the best devices from adjoint-based optimization. Statistically, 75% of devices from the conditional GLOnet have efficiencies higher than those from adjoint-based optimization, and 92% of devices from the conditional GLOnet have efficiencies higher than or within 5% those from adjoint-based optimization. Although our conditional GLOnet performs well for most wavelength and angle values, it does not optimally perform in certain regimes, such as that at short wavelengths and small deflection angles. We hypothesize that these nonidealities can be improved with further refinement of the network architecture and training process, and this will be the topic of future study.

The efficiency histograms from adjoint-based topology optimization and the conditional GLOnet for select wavelength and angle pairs are displayed in Figure 3C. A more complete set of histograms is in Figure S2. The histograms show that adjoint-based topology optimization generates devices with highly variable efficiencies. This indicates that the initial dielectric distributions of these devices broadly span the design space, and with each device being locally optimized, the result is a set of devices with a wide range of layouts and efficiencies. The conditional GLOnet-generated devices, on the other hand, tend to have more devices clustered at the high-efficiency end of the distribution. An examination of the layouts of these devices indicate that many have very similar geometries (Figure S3). This trend is consistent with the

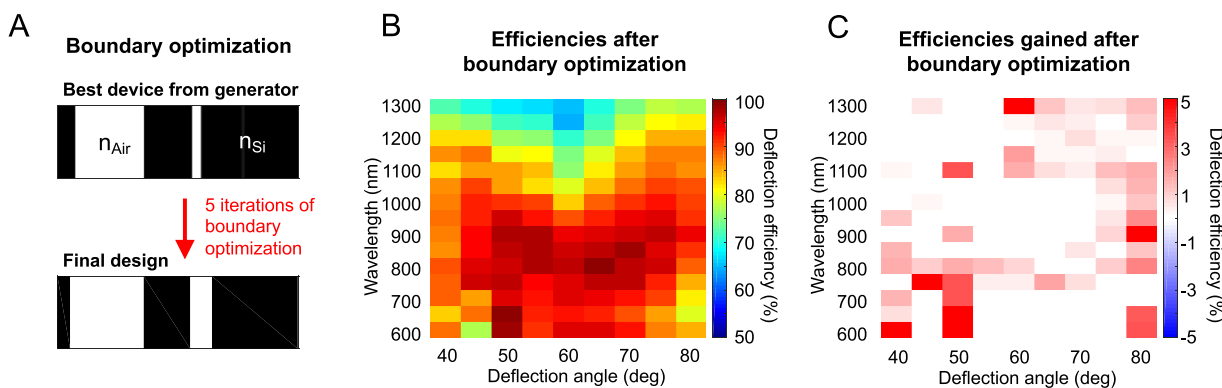
objective of the conditional GLOnet, which is to optimize the efficiency of  $\{\mathbf{n}|\lambda, \theta\}$ .

To help visualize the device optimization process with our conditional GLOnet, we show how the distribution of devices in the design space, together with its corresponding efficiency histogram, evolves over the course of network training. The devices in this example all operate at  $\lambda = 900$  nm and  $\theta = 60^\circ$ , and 100 devices are randomly generated during each iteration of training for visualization. The high-dimensional design space is visualized by performing a principle component analysis on the 500 binary metagratings with  $\lambda = 900$  nm and  $\theta = 60^\circ$  produced by adjoint-based optimization (used for Figure 3A) and then reducing the dimensionality of the space to two dimensions. The results are displayed in Figure 4 and Movie S1. Initially, the distribution of generated devices is spread broadly across the design space and the efficiency histogram spans a wide range of values with most devices exhibiting low to modest efficiencies. As network training progresses, the distribution of generated devices more tightly clusters and the efficiency histogram narrows at high-efficiency values. By the 1000 iteration mark, the generated devices have very high efficiencies and the histogram is strongly skewed toward high-efficiency values. A similar analysis for devices operating with other wavelength and angle combinations is presented in Figure S4.

An examination of total computation time indicates that our conditional GLOnet is computationally efficient at simultaneously optimizing a broad range of devices operating at different wavelengths and angles. A detailed analysis, presented



**Figure 4.** Evolution of device patterns and efficiency histograms as a function of conditional GLOnet training. (A) Visualization of 100 device patterns generated by the conditional GLOnet at different iteration numbers, depicted in a 2D representation of the design space. All devices are designed to operate at a wavelength of 900 nm and an angle of 60°. The distribution of generated devices is initially spread out in the design space at the early stages of training and converges to a high efficiency cluster by the 1000 iteration mark. (B) Efficiency histogram of generated devices at different iteration numbers. The efficiency histogram is initially broad and converges to a distribution of devices biased toward high efficiencies by the 1000 iteration mark.



**Figure 5.** Device refinement using boundary optimization. (A) Boundary optimization uses efficiency gradients to refine the boundaries of binary structures. (B) Efficiency plot of devices generated by the conditional GLOnet and then refined with 10 iterations of boundary optimization. For this plot, the best device from Figure 3B for each wavelength and angle combination is used for boundary optimization. (C) Plot of gains in efficiency after boundary optimization, calculated from the data in Figure 3B and Figure 5B. Most devices experience modest boosts in efficiency and 4% of devices exhibit over a 5% efficiency improvement.

in the Supporting Information, indicates that the conditional GLOnet uses 10 $\times$  less computational cost compared to our benchmark adjoint-based topology optimization calculations. We note that as the number of accessible computing nodes scales up, the efficacy and power of conditional GLOnets can be enhanced by implementing more simulations in parallel and scaling the batch sizes up. Such scaling is particularly amenable to existing cloud and server computing infrastructure, which generally enable access to large numbers of computing nodes.

Finally, we attempt to refine the generated devices from the conditional GLOnet using adjoint-based boundary optimization (Figure 5A). In this algorithm,  $\mathbf{g}$  is calculated by conducting a forward and adjoint simulation, which is consistent with topology optimization. However, we only consider the gradients at the silicon–air boundaries of the device and fix the device refractive indices to be binary

throughout the optimization. For this analysis, we perform 10 iterations of boundary optimization on the highest efficiency generated device for each wavelength and angle pair (Figure 3B). The final device efficiencies after boundary optimization are shown in Figure 5B and the differential changes in efficiency are shown in Figure 5C. Most of the efficiency changes are relatively modest and only 4% of devices have efficiency gains larger than 5%, indicating that devices from the conditional GLOnet are already at or near local optima.

**Conclusions.** In summary, we have shown that conditional GLOnets are an effective and computationally efficient global topology optimizer for metagratings. A global search through the design space is possible because the generative neural network optimizes the efficiencies of device distributions that initially span the full design space. The best devices generated by the conditional GLOnet compare well with the best devices

generated by adjoint-based topology optimization. By conditioning GLOnets with a continuum of operating parameters, ensembles of devices can be simultaneously optimized, further reducing overall computational cost. Future work will focus on extending conditional GLOnets to other metasurface systems, including aperiodic broadband devices. The loss function for those design problems can be defined in the same way here, but with  $\mathbf{g}$  tailored to the specific optimization target. For broadband devices, for example,  $\mathbf{g}$  should consist of the weighted summation of efficiency gradients at different wavelengths. Given the generality of our approach, we envision that conditional GLOnets can apply to the design of other classes of photonic devices and more broadly to other physical systems in which device performance can be improved using gradients.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.nanolett.9b01857](https://doi.org/10.1021/acs.nanolett.9b01857).

Network architecture and training process of GLOnet; formulation of the loss function for the conditional GLOnet; computational cost of the conditional GLOnet versus adjoint-based topology optimization; Figures S1 to S4 (PDF)

Evolution of device patterns and efficiency histograms as a function of GLOnet training for wavelength 900 nm and deflection angle  $60^\circ$  (MP4)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [jonfan@stanford.edu](mailto:jonfan@stanford.edu).

### ORCID

Jiaqi Jiang: [0000-0001-7502-0872](https://orcid.org/0000-0001-7502-0872)

Jonathan A. Fan: [0000-0001-9816-9979](https://orcid.org/0000-0001-9816-9979)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The simulations were performed in the Sherlock computing cluster at Stanford University. This work was supported by the U.S. Air Force under Award Number FA9550-18-1-0070, the Office of Naval Research under Award Number N00014-16-1-2630, and the David and Lucile Packard Foundation.

## ■ REFERENCES

- (1) Yu, N.; Capasso, F. *Nat. Mater.* **2014**, *13*, 139.
- (2) Kildishev, A. V.; Boltasseva, A.; Shalae, V. M. *Science* **2013**, *339*, 1232009.
- (3) Kuznetsov, A. I.; Miroshnichenko, A. E.; Brongersma, M. L.; Kivshar, Y. S.; Luk'yanchuk, B. *Science* **2016**, *354*, aag2472.
- (4) Zheng, G.; Mühlenbernd, H.; Kenney, M.; Li, G.; Zentgraf, T.; Zhang, S. *Nat. Nanotechnol.* **2015**, *10*, 308.
- (5) Colburn, S.; Zhan, A.; Majumdar, A. *Sci. Adv.* **2018**, *4*, No. eaar2114.
- (6) Arbabi, E.; Kamali, S. M.; Arbabi, A.; Faraon, A. *ACS Photonics* **2018**, *5*, 3132–3140.
- (7) Tittl, A.; Leitis, A.; Liu, M.; Yesilkoy, F.; Choi, D.-Y.; Neshev, D. N.; Kivshar, Y. S.; Altug, H. *Science* **2018**, *360*, 1105–1109.
- (8) Estakhri, N. M.; Edwards, B.; Engheta, N. *Science* **2019**, *363*, 1333–1338.
- (9) Kim, J. Y.; Kim, H.; Kim, B. H.; Chang, T.; Lim, J.; Jin, H. M.; Mun, J. H.; Choi, Y. J.; Chung, K.; Shin, J.; Fan, S.; Kim, S. O. *Nat. Commun.* **2016**, *7*, 12911.
- (10) Singh, R.; Cao, W.; Al-Naib, I.; Cong, L.; Withayachumnankul, W.; Zhang, W. *Appl. Phys. Lett.* **2014**, *105*, 171101.
- (11) Sell, D.; Yang, J.; Doshay, S.; Zhang, K.; Fan, J. A. *ACS Photonics* **2016**, *3*, 1919–1925.
- (12) Campbell, S. D.; Sell, D.; Jenkins, R. P.; Whiting, E. B.; Fan, J. A.; Werner, D. H. *Opt. Mater. Express* **2019**, *9*, 1842–1863.
- (13) Rogers, E. T.; Lindberg, J.; Roy, T.; Savo, S.; Chad, J. E.; Dennis, M. R.; Zheludev, N. I. *Nat. Mater.* **2012**, *11*, 432.
- (14) Yin, J. Y.; Wan, X.; Zhang, Q.; Cui, T. *Sci. Rep.* **2015**, *5*, 12476.
- (15) Jafar-Zanjani, S.; Inampudi, S.; Mosallaei, H. *Sci. Rep.* **2018**, *8*, 11040.
- (16) Jensen, J. S.; Sigmund, O. *Laser Photonics Rev.* **2011**, *5*, 308–321.
- (17) Molesky, S.; Lin, Z.; Piggott, A. Y.; Jin, W.; Vucković, J.; Rodriguez, A. W. *Nat. Photonics* **2018**, *12*, 659.
- (18) Sell, D.; Yang, J.; Wang, E. W.; Phan, T.; Doshay, S.; Fan, J. A. *ACS Photonics* **2018**, *5*, 2402–2407.
- (19) Yang, J.; Sell, D.; Fan, J. A. *Ann. Phys.* **2018**, *530*, 1700302.
- (20) Lin, Z.; Groever, B.; Capasso, F.; Rodriguez, A. W.; Lončar, M. *Phys. Rev. Appl.* **2018**, *9*, 044030.
- (21) Sell, D.; Yang, J.; Doshay, S.; Fan, J. A. *Adv. Opt. Mater.* **2017**, *5*, 1700645.
- (22) Xiao, T. P.; Cifci, O. S.; Bhargava, S.; Chen, H.; Gissibl, T.; Zhou, W.; Giessen, H.; Toussaint, K. C., Jr; Yablonovitch, E.; Braun, P. V. *ACS Photonics* **2016**, *3*, 886–894.
- (23) Piggott, A. Y.; Lu, J.; Lagoudakis, K. G.; Petykiewicz, J.; Babinec, T. M.; Vucković, J. *Nat. Photonics* **2015**, *9*, 374.
- (24) Hughes, T. W.; Minkov, M.; Williamson, I. A.; Fan, S. *ACS Photonics* **2018**, *5*, 4781–4787.
- (25) Lalau-Keraly, C. M.; Bhargava, S.; Miller, O. D.; Yablonovitch, E. *Opt. Express* **2013**, *21*, 21693–21701.
- (26) Peurifoy, J.; Shen, Y.; Jing, L.; Yang, Y.; Cano-Renteria, F.; DeLacy, B. G.; Joannopoulos, J. D.; Tegmark, M.; Soljačić, M. *Sci. Adv.* **2018**, *4*, No. eaar4206.
- (27) Inampudi, S.; Mosallaei, H. *Appl. Phys. Lett.* **2018**, *112*, 241102.
- (28) Ma, W.; Cheng, F.; Liu, Y. *ACS Nano* **2018**, *12*, 6326–6334.
- (29) Liu, D.; Tan, Y.; Khoram, E.; Yu, Z. *ACS Photonics* **2018**, *5*, 1365–1369.
- (30) Liu, Z.; Zhu, D.; Rodrigues, S. P.; Lee, K.-T.; Cai, W. *Nano Lett.* **2018**, *18*, 6570–6576.
- (31) Jiang, J.; Sell, D.; Hoyer, S.; Hickey, J.; Yang, J.; Fan, J. A. *arXiv preprint arXiv:1811.12436*, **2018** (accessed on Nov 29, 2018).
- (32) Jiang, J.; Fan, J. A. *arXiv preprint arXiv:1906.07843*, **2019** (accessed on June 18, 2019).
- (33) Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics* **2010**, 249–256.
- (34) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* **2016**, 770–778.
- (35) Phan, T.; Sell, D.; Wang, E. W.; Doshay, S.; Edee, K.; Yang, J.; Fan, J. A. *Light: Sci. Appl.* **2019**, *8*, 48.
- (36) Yang, J.; Fan, J. A. *Opt. Lett.* **2017**, *42*, 3161–3164.
- (37) Hugonin, J.; Lalanne, P. *Reticolo Software for Grating Analysis*; Institut d'Optique: Palaiseau, France, 2005.