



Artificial neural network discovery of a switchable metasurface reflector

J. R. THOMPSON,^{1,2} J. A. BURROW,³ P. J. SHAH,^{1,4} J. SLAGLE,¹ E. S. HARPER,¹ A. VAN RYNBACH,¹ I. AGHA,³ AND M. S. MILLS^{1,*}

¹Air Force Research Laboratory, 2179 12th St., Wright-Patterson Air Force Base, OH 45433-7132, USA

²Azimuth Corporation, 2970 Presidential Dr. # 200, Fairborn, OH 45324, USA

³Electro-Optics Department, University of Dayton, 300 College Park Ave., Dayton, OH 45469, USA

⁴Apex Microdevices LLC, 4871 Misrach Ct., West Chester, OH 45069, USA

*matthew.mills.25@us.af.mil

Abstract: Optical materials engineered to dynamically and selectively manipulate electromagnetic waves are essential to the future of modern optical systems. In this paper, we simulate various metasurface configurations consisting of periodic 1D bars or 2D pillars made of the ternary phase change material $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST). Dynamic switching behavior in reflectance is exploited due to a drastic refractive index change between the crystalline and amorphous states of GST. Selectivity in the reflection and transmission spectra is manipulated by tailoring the geometrical parameters of the metasurface. Due to the immense number of possible metasurface configurations, we train deep neural networks capable of exploring all possible designs within the working parameter space. The data requirements, predictive accuracy, and robustness of these neural networks are benchmarked against a ground truth by varying quality and quantity of training data. After ensuring trustworthy neural network advisory, we identify and validate optimal GST metasurface configurations best suited as dynamic switchable mirrors depending on selected light and manufacturing constraints.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Bodies of scientific literature exist describing various methods that can be utilized for both dynamically and selectively controlling light through a material. Dynamic behavior refers to an induced change such that subsequent light-matter responses are altered. Selectivity, in our context, refers to a material responding to different initial conditions of light in a pre-engineered way. There are many well-known ways to induce dynamic behavior; to name a few: the photo-elastic, thermo-optic, and photo-refractive effect, the Pockels and Kerr effect, and the re-orientation of liquid crystals [1,2]. The focus of this paper exploits another intriguing candidate for creating dynamic behavior: chalcogenide-based phase change materials (PCMs)—a glassy material class with a host of amenable switching-related properties [3,4]. In contrast to standard glassy materials, PCMs uniquely experience high-speed phase transitions, long-lived thermal stability, sharp resistance changes, overall chemical stability, and repeatable reversible transitions [5,6]. The concept behind PCMs is straightforward. Like many glassy materials, the solid form of PCMs can exist in both a crystalline and an amorphous state. When the PCM reaches a threshold temperature, the material crystallizes. On the other hand, if the glass is quenched before this threshold, then it solidifies with an amorphous structure. Laser light or electrical pulses can be used to illuminate and heat the sample past the glass-transition temperature. If a short and intense stimulus is used, the heat dissipation will be fast enough to lead to the amorphous state; otherwise, the crystalline state is formed [7]. From an optics viewpoint, the most notable trait is a large low-loss refractive index change, usually on the order of unity, between the crystalline and amorphous states stemming from a drastic atomic rearrangement. The amalgamation of these properties

has elevated PCMs into a special class of dynamic materials first pursued and commercialized for optical data storage [8–12]. Although various PCM candidates have been studied [13–15], one of the most successful to date in terms the aforementioned properties, is ternary Ge-Sb-Te compounds, specifically the $\text{Ge}_2\text{Sb}_2\text{Te}_5$ mixture (GST) which boasts amorphization times of less than a nanosecond and is stable over millions of switching cycles [7,16,17]. Consequently, GST fabrication capabilities have matured over the years.

With respect to selective optical behavior, one often relies on inhomogeneities within a material system. Classic cases are optical gratings and phase masks which split light spatially and/or spectrally due to periodically engineered material distributions with size scales on the order of the wavelength. Dispersion itself can be viewed as a source of optical selectivity. Material dispersion arises from an inherent variation of refractive index with wavelength, and because this behavior is complex-valued, any material will innately be frequency selective in terms of its refractive, absorptive, reflective, and transmissive optical properties. Waveguide dispersion and modal dispersion seen in optical fibers results from the geometrical arrangement of refractive index and is often prudently modified for optimal waveguiding [18]. In short, customized selectivity of optical properties becomes possible by arranging materials into well thought-out geometries at the micro- and nano-scale. These approaches seek to exploit resonance features, photonic bandgaps, and waveguiding properties. Examples include hollow-core photonic crystal fibers [19,20], metamaterials [21–25], and dielectric multilayer thin films [26–30]. One of the earliest scenarios of prescribing optical selectivity for reflection and transmission spectra can be found in the field of multilayer optical coatings. Attributable to many degrees of freedom, the thickness and material composition of every layer can often be adjusted such that a desired reflection or transmission spectra is obtained [29,31]. Numerical methods have been developed which inversely design these thin film stacks to fit any desired reflection profile [32–34]. Unfortunately, multilayer coatings have limitations. In order to achieve a given spectral specification, sometimes hundreds of layers are required leading to a lengthy and costly fabrication process. When designs are needed for longer wavelength regimes (e.g. infrared), the individual layers must necessarily become thicker. This leads to longer deposition times as well as the introduction of secondary problems such as stress delamination or spallation [35,36]. Without proper measures, multilayer stacks can also suffer from a blue-shift phenomenon whereby light incident at oblique angles cause the prescribed reflection bands to shift toward shorter wavelengths [37].

More recently, metamaterials have come into focus as an alternative means of selectively tailoring the spectral response of light [21,38,39]. Because of the near infinite degrees of design freedom, metamaterial devices hold promise as a way to obtain multi-functional selectivity. Composed of nanoscale scatterers, these engineered materials have since demonstrated novel control over several properties of light such as polarization [40], absorption [41], reflection and transmission [42], and refraction [43,44]. Progress into all-dielectric metamaterials has gained traction motivated in part by the need to avoid losses characteristic of metallic-based metamaterials [24]. A simpler version of the metamaterial is the 2D subclass, the metasurface. Metasurfaces are composed of a single patterned layer grown on a substrate and have been investigated for a variety of applications [45–47]. In particular, all-dielectric metasurfaces have demonstrated exceptional performance in regards to high quality low-loss reflectors [42,48] and transmissive achromatic lenses [8].

The potential of identifying dynamic and selective reflectors by combining the dynamic phase change characteristics of GST with the selectivity offered by metasurfaces is apparent; however, a key problem stands in the way: the sheer amount of possible designs makes optimizing to a desired reflection spectra intractable. Even a relatively simple metasurface has a near infinite number of ways to arrange the surface pattern. This is further complicated by the conditions of the incident light, which can be described by polarization, beam shape, wavelength, and inclination/azimuthal angle. Each of these properties affect the response of the metasurface. Furthermore, when

utilizing a dynamic material like GST, each and every metasurface configuration must link the dispersive properties of both the crystalline and amorphous states. Even with modern computing resources and highly optimized computational methods, comprehensively searching this immense parameter space is infeasible. For this reason, any ability to design metamaterials with a prescribed functionality has required a keen understanding of the underlying physics; for example, past analyses have looked at Fano resonances [49], adjoint methods [50], or Mie scattering [22,24]. Yet these approaches are always bespoke to a particular meta structure and can prove difficult to implement for non-standard geometries. Thus, these usual options necessarily move away from a general methodology. However, with the recent resurgence of machine learning approaches, numerical methods in combination with artificial neural networks (ANNs) are now being leveraged to generally approach this problem set [51–56]. Various techniques exist to find and optimize optical devices with specific properties and/or task driven applications. Some such techniques include the direct binary search (DBS) method [57,58], objective-first (OB-1) algorithm [59–61], topology optimization (TO) [62], and genetic algorithms (GA) [63]. One particularly promising study leverages ANNs as highly efficient function approximators. By providing a training set of parameterized metasurfaces linked to simulated output reflection and transmission spectra, an ANN learns to accurately approximate the numerical method used to generate the data. This spectra predicting network (SPN) [51,52] enables a key benefit as, once trained, it is many orders of magnitude faster at simulating reflection and transmission spectra when compared to the original computational method. The speed-up allows one to comprehensively explore large parameter spaces and ultimately identify devices subject to a set of constraints.

In this paper, we investigate dynamic and selective GST metasurfaces consisting of either periodic bars (1D case) or cylinders (2D case). We judiciously sample geometries within fabrication tolerances parameterized by lattice spacing, unit cell feature height, and unit cell feature width/radius. For each arrangement, output reflection and transmission coefficients are computed with rigorous coupled wave analysis (RCWA) over a range of IR wavelengths, angles of incidence, polarizations, and both GST phase-states. These results are used to train spectra predicting artificial neural networks capable of learning the relationship between the input conditions and the output reflection and transmission values. A detailed study of the SPN training requirements (quality and quantity of necessary training data) is performed for the computationally inexpensive 1D metasurface case and then harnessed for the slower to compute 2D metasurface case. Merit functions are applied across the SPN predictions identifying metasurface configurations with optimal infrared reflection/transmission switching properties. Several optimal designs are identified for different incident light conditions and then re-computed with RCWA to verify their behavior. While other GST and non-GST switchable devices exist [64–71], to the best of our knowledge we were able to find for the first time an angle and polarization independent high-contrast switchable reflective/transmissive metasurface for SWIR. And although this work focuses on GST metasurfaces, the methodology introduced is generally applicable to other active material classes.

2. Problem setup, training data collection, and ANN architecture

Consider two metasurfaces consisting of either 1D periodic GST bars or 2D cylindrical pillars surrounded by air on an alumina substrate (Fig. 1). Regardless of this metasurface dimension, \mathcal{D} , both systems can be parameterized by a lattice spacing, Λ , a feature height, H , and a feature width, X (Fig. 1(a)). The parameter X represents the half width of the bar in the $\mathcal{D} = 1$ case (Fig. 1(b)) and the radius of the pillar in the $\mathcal{D} = 2$ case (Fig. 1(c)). The lattice spacing in the $\mathcal{D} = 2$ case is assumed to be square and thus, $\Lambda_x = \Lambda_y = \Lambda$. Light of wavelength, λ , is incident with a polarization, \mathcal{P} , and inclination angle, θ . The material dispersion and losses, $\eta_{\text{GST}}(\lambda)$, depend on whether the crystalline ($\eta_{\text{cGST}}(\lambda)$) or amorphous state ($\eta_{\text{aGST}}(\lambda)$) is present. See Appendix A for

a list of important symbols used in the manuscript and Appendix B for the material dispersion curves used in our simulations.

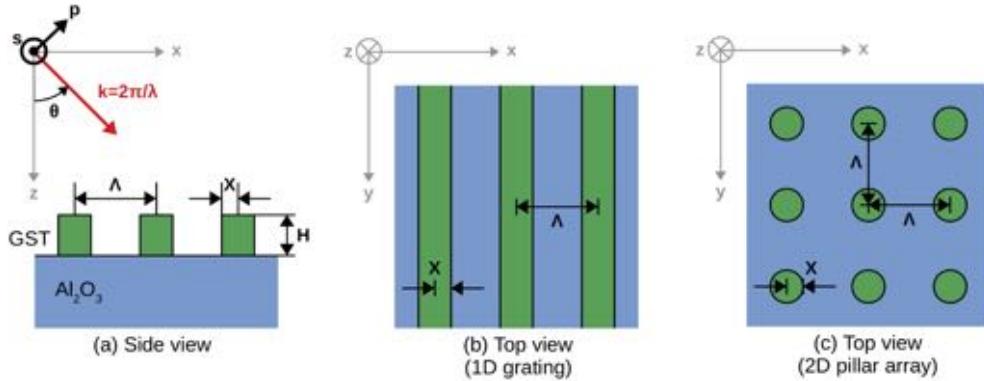


Fig. 1. Schematic of GST metasurfaces on alumina substrates. (a) Side view illustrating the parameterized surface and the initial light conditions. (b) Top view of the 1D GST metasurface. (c) Top view of the 2D GST metasurface.

For each and every combination of geometric parameters and light conditions in the set: $\{X, H, \Lambda, \mathcal{D}, \theta, \mathcal{P}, \lambda, \eta_{\text{GST}}(\lambda)\}$, a single reflectance and transmittance coefficient, R and T , is able to be computed via RCWA. Each pair of computed coefficients satisfy $R + T \leq 1$ with the equality occurring when losses are insignificant. Because brute-force sweeping the configuration space is infeasible, our goal is to train ANNs capable of interpolating this full design space from a small random sampling. Although training data could be collected by sampling from this 8 dimensional input space, such an approach leads to superfluous data collection and hamstrings the ANN training, analysis, and optimization. A more prudent approach is to reserve the input configuration space to only the geometric parameters $\Xi = \{X, H, \Lambda\}$ and compute for each combination of $\{\mathcal{D}, \theta, \mathcal{P}\}$, a reflection and transmission spectra for both GST states, $\Phi = \{R(\lambda, \eta_{\text{aGST}}(\lambda)), T(\lambda, \eta_{\text{aGST}}(\lambda)), R(\lambda, \eta_{\text{cGST}}(\lambda)), T(\lambda, \eta_{\text{cGST}}(\lambda))\}$. The sets, Ξ and Φ , are introduced to elucidate the input and output training data format that a single training datum, $\Psi = \{\Xi, \Phi\}$, assumes. In addition, the notation $\tilde{\Phi} = \{\tilde{R}(\lambda, \eta_{\text{aGST}}(\lambda)), \tilde{T}(\lambda, \eta_{\text{aGST}}(\lambda)), \tilde{R}(\lambda, \eta_{\text{cGST}}(\lambda)), \tilde{T}(\lambda, \eta_{\text{cGST}}(\lambda))\}$ will be used to distinguish between neural network spectra predictions and RCWA simulated spectra, Φ . Note that in the form Ψ , a single training datum contains the spectral switching information caused by the GST phase-state change for each uniquely chosen design Ξ . Furthermore, this grouping is amenable to our situation since values stored in Ψ will take on a range of values while the remaining initial conditions, $\{\mathcal{D}, \theta, \mathcal{P}\}$, will take on discrete choices. We are left with two main setup tasks; the first is to define the combinations of $\{\mathcal{D}, \theta, \mathcal{P}\}$ and to wisely set bounds and resolutions on the values within Ξ and Φ ; the second is to define a procedure which will allow for high throughput collection of training data in the form, Ψ .

2.1. Constraining the metasurface sampling space

Defining $\{\mathcal{D}, \theta, \mathcal{P}\}$ for our simulations is straightforward. \mathcal{D} can take on two values, $\mathcal{D} \in \{1, 2\}$, and the polarization can be either s- or p- polarized, $\mathcal{P} \in \{s, p\}$. Although the inclination angle could take on a range of values, we restrict ourselves to two discrete cases, $\theta = \{0^\circ, 45^\circ\}$, in order to identify normally incident mirrors or folding mirrors. Thus there are $2^3 = 8$ possible combinations of $\{\mathcal{D}, \theta, \mathcal{P}\}$. For each individual combination, the collection of training data is in a format given by Ψ . Since we are ultimately interested in identifying switchable IR reflectors, we set a wavelength range of $1 \text{ } \mu\text{m} \leq \lambda \leq 3 \text{ } \mu\text{m}$ with a step size of $\Delta\lambda = 4 \text{ nm}$. Thus, the

length of each Φ containing the reflection and transmission spectra of both GST phase-states is $501 \times 4 = 2004$ elements long.

Next we set the bounds and resolutions for values contained in Ξ . The simplest of these values are the feature heights, H , since they are independent of all other parameters. Selection ranges are capped between $0.05 \mu\text{m} \leq H \leq 1 \mu\text{m}$ with a resolution of $\Delta H = 10 \text{ nm}$ dictated by known fabrication tolerances. As an extra measure, we account for a 6% reduction in feature height during the transition between amorphous to crystalline phases. When defining the range of the feature width, X , one must only ensure that no overlap with the next unit cell occurs. This is best accomplished by locking the range such that $0 \leq X' \leq 0.5$, where $X' = X/\Lambda$ is the feature width scaled by the lattice spacing. This scaled feature width is set to a resolution of $\Delta X' = 0.005$ stemming from the fabrication tolerances of Λ . Like the feature width, the lattice spacing, Λ , can assume any value as long as it does not cause overlap in the unit cell; however, because our eventual goal is to identify highly switchable IR reflectors, we are only interested in Λ size scales where this is possible.

Previous studies of the electromagnetic modes within all-dielectric metasurfaces [23,72,73] offer the perspective that metasurfaces can be understood in terms of coupling and interference of propagating waveguide array modes (WGAMs) at the layer interfaces. This analysis leads to the conclusion that metasurface configurations boasting high reflection or transmission are the result of the interference of a small number of WGAMs – termed the high contrast metasurface (HCM) regime [72]. By computing these WGAMs, one can restrict the metasurface lattice spacing to only the regions capable of high reflection and transmission. Photonic crystal band diagrams arising from the periodic structuring of the surface geometry are equivalently the cutoff frequencies of these WGAMs as a function of X' , inclination angle, polarization, η_{GST} , and wavelength. Therefore, by extracting the photonic crystal band structure, one can determine the number of WGAMs able to exist within the metasurface. The lowest band solution can be correlated to a minimum lattice spacing, $\Lambda_{\text{low}}(X', \theta, \mathcal{P}, \eta_{\text{GST}}, \lambda)$, where all $\Lambda < \Lambda_{\text{low}}(X', \theta, \mathcal{P}, \eta_{\text{GST}}, \lambda)$ have a single WGAM propagating. This defines the so-called effective medium (EM) regime characteristic of metamaterial studies aiming to change effective constitutive properties of a material. Floquet-Bloch theorem can then be used to correlate a maximum lattice spacing, $\Lambda_{\text{high}}(\theta, \lambda) = \lambda/(1 + \sin \theta)$ [73], where all $\Lambda \geq \Lambda_{\text{high}}(\theta, \lambda)$ results in more than one diffraction order and is therefore not useful for single beam high reflection or transmission. By following this general procedure, the simulation region of interest can be reduced down to lattice constants between $\Lambda_{\text{low}}(X', \theta, \mathcal{P}, \eta_{\text{GST}}, \lambda) \leq \Lambda < \Lambda_{\text{high}}(\theta, \lambda)$ ultimately decreasing the amount of training data to be collected. Again, the resolution is set based on fabrication tolerances, $\Delta \Lambda = 10 \text{ nm}$. Table 1 organizes the results of our metasurface design sampling space for the GST metasurfaces.

Table 1. Metasurface sampling space for values in Ξ and the wavelength range for each computed spectra in Φ .

Input Parameter	Range	Resolution
Feature Height, H	$0.05 \mu\text{m} \leq H \leq 1 \mu\text{m}$	$\Delta H = 10 \text{ nm}$
Scaled Width, X'	$0.0 \leq X' \leq 0.5$	$\Delta X' = 0.005$
Lattice Spacing, Λ	$\Lambda_{\text{low}}(X', \theta, \mathcal{P}, \eta_{\text{GST}}, \lambda) \leq \Lambda < \Lambda_{\text{high}}(\theta, \lambda)$	$\Delta \Lambda = 10 \text{ nm}$
Wavelength, λ	$1 \mu\text{m} \leq \lambda \leq 3 \mu\text{m}$	$\Delta \lambda = 4 \text{ nm}$

2.2. Procedure for collecting training data

For each $\{\mathcal{D}, \theta, \mathcal{P}\}$, sampling a metasurface design is done by randomly assigning values to H , X' , and Λ within the ranges and resolutions defined in Table 1. Since the bounds of Λ are dependant on other parameters, a lookup table is generated containing possible values of Λ_{low} and Λ_{high} . To calculate the values of Λ_{low} the open source program MIT Photonic Bands (MPB) [74]

was leveraged (see Appendix C for an example calculation). To simplify the size of this lookup table, we set $\lambda = 2 \mu\text{m}$, which locks the value of $\eta_{\text{GST}}(\lambda = 2 \mu\text{m})$. Although this assumption can lead to instances of Ξ lying outside the HCM regime for $\lambda \neq 2 \mu\text{m}$, the small amount of superfluous data collected is worth the time saved in pre-processing the lookup table. Moreover, in this manuscript we desire to find an optimized switchable mirror for $\lambda = 2 \mu\text{m}$. The parameter X' is then randomized and used to set the appropriate bounds of Λ , which can then be randomly assigned itself. X is then reverted back from the scaled form by the simple relationship $X = X'\Lambda$. Finally, these values are grouped in the format $\Xi = \{X, H, \Lambda\}$. This process is repeated until a sufficient amount of input training data is obtained. The question of what constitutes a sufficient amount is analyzed in Sec. 3.

Equipped with the randomly sampled input values constrained to the HCM regime for a given $\{\mathcal{D}, \theta, \mathcal{P}\}$, we begin the computation of Φ for each Ξ . To do this, we implement our self-built Machine Accelerated Nanoscale Targeted Inhomogenous Structures (MANTIS) framework. The workflow is summarized as such:

1. An instance of Ξ with settings $\{\mathcal{D}, \theta, \mathcal{P}\}$ is loaded representing a unique metasurface design. A twin design is created and the crystalline/amorphous GST dispersion data is set respectively.
2. Stanford's Stratified Structure Solver (S^4) [75] is leveraged to perform RCWA on the two designs. The computational accuracy of RCWA robustly converges with increased spatial harmonics, n_H . In our context, $n_H = 15$ is accurate to within $\approx 0.9\%$ of any $n_H = 51$ simulated spectra. Furthermore, $n_H = 51$ is found to approximate the true spectra for the $\mathcal{D} = 1$ case as adding more harmonics does not change the resulting spectra. For $\mathcal{D} = 2$, the amount of spatial harmonics must be squared to capture the same accuracy (i.e. $n_H = 15^2$ and $n_H = 51^2$). We only used $n_H = 15^2$ for the $\mathcal{D} = 2$ case since we must rely on the SPN for results.
3. Post-processing extracts resulting reflection and transmission spectra on the instance, Ξ , for both GST states. These results are stored in the form of Φ .
4. The Signac data management framework [76] is used to bundle the input parameters with the related output spectra in the form of Ψ . This is compressed as a binary file with metadata for later recall.
5. Steps 1-4 are looped through each uniquely generated value of Ξ .
6. The bounding procedure described in this section and step 5 is repeated for 7 possible combinations of $\{\mathcal{D}, \theta, \mathcal{P}\}$. There is a redundancy in calculating both s- and p-polarized light incident at $\theta = 0^\circ$ for $\mathcal{D} = 2$ due to azimuthal symmetry, thus the total number of combinations (eight) is reduced by one.

To summarize, following the protocol of this section results in seven separate categories of training data. Each collection covers all combinations of metasurface dimension, polarization, and inclination angle, while also reducing any redundancies. Within each collection, training data exists in a format $\Psi = \{\Xi, \Phi\}$ suitable for ANN training. The values of each Ξ contain the geometric parameters of a unique metasurface design which was constrained to be within (or near) the HCM regime and within fabrication tolerances. For each unique Ξ , the values of Φ contain the RCWA computed reflection and transmission spectra for both GST states. Details on the necessary accuracy on the RCWA computed spectra is addressed in Sec. 3.

2.3. Spectra predicting neural network architecture and training

The acquired training data can then be piped into a spectra predicting neural network (SPN) architecture [51,52]. For ease of interpretability and since all the values in the training set Ψ are on similar scales near unity, the raw device parameter, Ξ , and spectral data, Φ , were used for training. With adequate training data, this type of ANN is able to mimic the original RCWA computations, but benefits from a $\times 10^6$ speed-up [52]. Our SPN is created using the Keras API with a TensorFlow back-end. Illustrated in Fig. 2, the architecture consists of seven dense layers. The input layer is sized to accept the geometric parameter vector Ξ . The network topology then expands and segments twice into four different routes. The first segmentation occurs between the first and second hidden layers, where the 1002 nodes in the first hidden layer are densely connected to two separate second hidden layers, each with 2004 nodes. The two second hidden layers are independent of each other and represent the aGST and cGST phase-states. Each of the hidden second layers then split again and the 2004 nodes in each second hidden layer are densely connected to two separate 4008 node third hidden layers, representing the splitting between reflection and transmission spectra. Thus, there are four independent routes that all share the same first hidden layer. Each of these four segments terminate with a 501 node output layer correlating to the wavelength and GST-state dependent reflection and transmission spectra values stored in $\hat{\Phi}$, where $\hat{\Phi}$ is sought to match Φ . The rectified linear unit (ReLU) activation function is used between each layer except for the last, which is sigmoidal in order to restrain output node values between [0,1]. The training batch size is set to 256 and the training-validation ratio is 90%/10%. The SPN implements the Adam optimization routine with mean square error (MSE) for the loss function. Initial Adam learning rate hyper-parameters are set to $\gamma_{LR} = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The SPN is continually trained until 50 epochs pass without improving

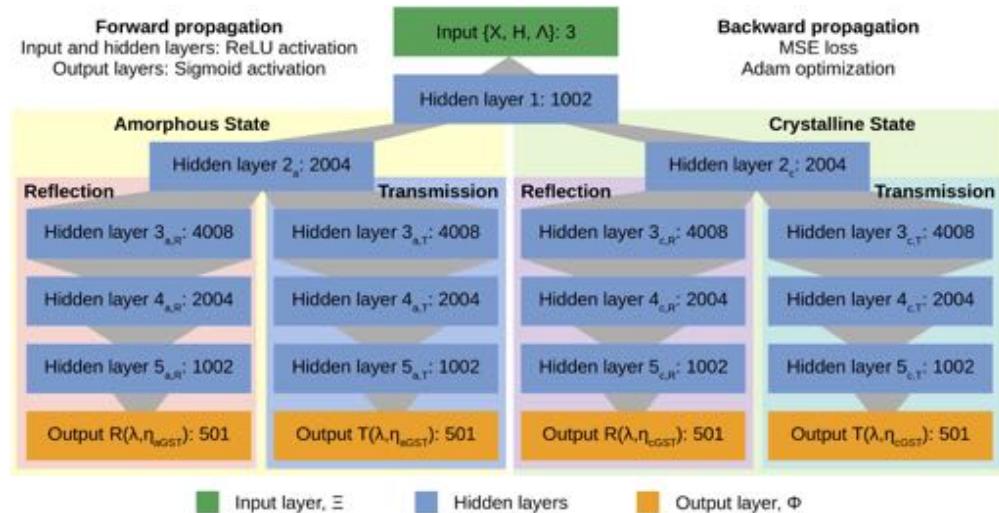


Fig. 2. The architecture of the implemented spectra predicting network. The SPN we use is a segmented dense ANN consisting of seven layers: an input, an output, and five hidden. The first hidden layer branches into two segmented paths for the purpose of separating the aGST and cGST phase-state information (subscripted a and c). Each of these in turn branches again at the third layer in order to separate reflection and transmission spectra (subscripted R and T). The input layer has three nodes representative of Ξ , whereas the collection of four outputs is either Φ or $\hat{\Phi}$ depending whether the SPN is being trained or used to make predictions. Each of the four outputs has 501 nodes containing the spectral value for all λ described in Table 1.

the MSE of the validation set by more than 10^{-6} , after which the weights of the SPN that yields the smallest MSE for the validation data are kept to ensure there is no overfitting.

Various training set sizes, α , can be used to converge a SPN. We explore the SPN predictive ability resulting from five different training set sizes, $\alpha \in \{0.1\%, 0.5\%, 1\%, 2\%, 3\%\}$ each containing either coarse ($n_H = 15$) or fine ($n_H = 51$) spectral data. The percentages represent the fraction of the randomly sampled designs to the total design space. Each trained SPN is then used to interpolate the full HCM regime. An L_2 normalization merit function is then applied to pick out optimal switching metasurface designs. All SPN identified optimal designs are then re-confirmed with high spatial harmonic RCWA simulations.

3. 1D GST Case: SPN benchmarking and identifying optimal switching mirrors

Recall that the purpose of the SPNs is to accurately mimic the computations of RCWA for each and every design within a parameterized space. Two frequently asked questions result: (1) how much training data does the SPN require and (2) how accurate must each training datum be in order to make trustworthy predictions. To our knowledge, these two questions are often ignored at first and later on justified by verifying an identified SPN predicted design with the original CEM method. This ansatz-like approach is certainly valid and often the necessary method when exploring an immense or computationally expensive parameter space. Fortunately, the $\mathcal{D} = 1$ GST metasurface case provides an opportunity to truly answer these two questions because brute-force computation of the entire parameter space with a dense resolution is possible. The high resolution collections of RCWA computations provide a "ground truth" to compare different SPNs against. In this section, we perform this comparative analysis for different SPNs trained on different set sizes. After confirming a sufficiently trained SPN, we harness it to identify an optimal switching mirror at $\lambda = 2 \mu\text{m}$.

By gradually increasing the spatial harmonics in the RCWA simulations, the computed reflection and transmission values converge. For the geometries in question, we determined that $n_H = 51$ was sufficient to numerically approximate the true curve and what we consider for the rest of the manuscript as the "ground truth" for the 1D GST case. Likewise, we found that $n_H = 15$ leads to an average spectral error magnitude of 0.9% from this ground truth. In all test cases, $n_H = 15$ captured all key spectral features, and thus was the minimum acceptable RCWA resolution we imposed. Following the protocols put forth in the previous sections, training data was gathered for the $\mathcal{D} = 1$ case using both $n_H = 51$ and $n_H = 15$. This resulted in eight data collections taking the form of Ψ encompassing both RCWA resolutions for each of the four combinations of polarization and inclination angle. Applying only the upper bound, Λ_{high} , the total number of possible GST bar designs that were non-diffractive (i.e. existing in either the effective medium or high contrast regime) was $\approx 2.02 \times 10^6$ and $\approx 1.19 \times 10^6$ for light incident at $\theta = 0^\circ$ and $\theta = 45^\circ$, respectively. By using the WGAMs to bound Λ_{low} , the total number of necessary simulations was reduced to $\approx 56.4\%$ and $\approx 54.6\%$ for $\theta = 0^\circ$ and $\theta = 45^\circ$. This effectively cut the total computation time in half. Using this bounding, brute-force computing the entire parameter space was on the order of hours for each $n_H = 15$ collection and days for each $n_H = 51$ collection (about a $40\times$ difference) for a system parallelized across 128 real cores. The amount of training data and the computation time for each collection is summarized in Table 2, where the values in bold are the number of designs simulated and the appropriate computation times. Note for the $\mathcal{D} = 1$ case, polarization does not affect the amount of possible designs (discussed in Appendix C.), but does affect the spectra computations.

3.1. SPN Benchmarking: Training set size and resolution

A different SPN is trained for every combination of α and n_H . Each SPN is used to interpolate reflection and transmission spectra for every geometric parameter, Ξ , and GST phase-state combination, which can then be compared against a ground truth (i.e. RCWA simulations that

Table 2. Training data collection: 1D grating of GST bars.

θ ($^{\circ}$)	P {s, p}	Amount of obtained data in format Ψ			Compute time	
		ND ^a (million)	HCM ^b (million)	HCM/ND (%)	$n_H = 15$ (days)	$n_H = 51$ (days)
0	<i>s</i>	2.02	1.14	56.4	0.1	4.0
0	<i>p</i>	2.02	1.14	56.4	0.1	4.0
45	<i>s</i>	1.19	0.65	54.6	0.6	2.3
45	<i>p</i>	1.19	0.65	54.6	0.6	2.3

^aND: nondiffractive regime consists of EM and HCM; EM: effective medium regime;

^bHCM: high contrast metasurface regime.

used $n_H = 51$). This procedure and analysis is repeated for all θ and \mathcal{P} . This immense amount of information is needed for complete benchmarking, but displaying each and every comparison would overwhelm this manuscript. Thus, in this section we will simply show the results for $\{\theta = 0^{\circ}, \mathcal{P} = s\}$. For clarity, it is useful to invoke the notation $\Phi(n_H)$ and $\widehat{\Phi}(n_H, \alpha)$ in order to distinguish between RCWA spectral data computed from n_H harmonics and spectral data predicted from a SPN trained on a fraction, α , of the HCM parameter space using $\Phi(n_H)$ sampled data. For example, $\widehat{\Phi}(15, 1\%)$ would refer to SPN generated spectra having been trained on a random assortment of $\approx 11,400$ designs (i.e. 1%) from the 1D HCM parameter space using the RCWA with $n_H = 15$ resolved GST spectra.

Figure 3 shows the reflectance of amorphous GST spectral data, $R(2 \mu\text{m}, \eta_{\text{aGST}})$, at a height and wavelength cross-section, $H = 0.5 \mu\text{m}$ and $\lambda = 2 \mu\text{m}$. We remind the reader that our choice to display the data along $H = 0.5 \mu\text{m}$ and $\lambda = 2 \mu\text{m}$ is to provide an example visualization; however, the full benchmarking accounts for all heights and wavelengths bounded by Table 1. The first row (Figs. 3(a.0-c.0)) shows the partially converged RCWA $\Phi(15)$ aGST reflectance approximation, the ground truth $\Phi(51)$, and the absolute difference between these two spectra. We found that $\Phi(15)$ accurately captured the main spectral features of the ground truth with error $<2\%$, but suffered from fictitious high frequency oscillations around regions of large reflectance gradients that lead to errors $>10\%$ (Fig. 3(b.0)). Figures 3(a.1-a.5) graphically show the sparse random sampling of coarse RCWA data $\Phi(15)$ for the different set sizes, $\alpha \in \{0.1\%, 0.5\%, 1\%, 2\%, 3\%\}$. This sampled data was used to train five different SPNs, $\widehat{\Phi}(15, \alpha)$. Figures 3(b.1-b.5) illustrate the resulting predictive capability of each of these. Figures 3(c.1-c.5) then compare the absolute error of these predictions with the ground truth. Remarkably, even the sparsest sampling of $n_H = 15$ data, $\widehat{\Phi}(15, 0.1\%)$, resulted in features consistent with the true $\Phi(51)$ spectra. As a general trend, increasing the amount of coarse training data improved the SPN predicted spectra. Furthermore, as the coarse training data increased, the absolute difference of the SPN predictions with the ground truth approached the absolute difference between $\Phi(51)$, and $\Phi(15)$. This gives evidence that a SPN is indeed able to mimic the original computational method, but is ultimately limited in predictive power by the resolution of the RCWA training data.

We can gain valuable insight investigating the moments of the absolute differences between SPN and RCWA simulations (i.e. mean, μ ; variance, σ^2 ; skewness, γ ; and kurtosis, κ). The first moment is the mean absolute error (MAE), μ , which is an L_1 norm and is plotted as a function of training set size in Fig. 4(a). The MAE averages the spectral error magnitudes over all wavelengths, λ ; phase-states, η_{GST} ; and device geometries, n . Written more explicitly,

$$|\widehat{\Phi}(n_H, \alpha) - \Phi(n_H)| \propto \sum_n \sum_{\lambda} \sum_{\eta_{\text{GST}}} |\widehat{R}_n(n_H, \alpha; \lambda, \eta_{\text{GST}}) - R_n(n_H; \lambda, \eta_{\text{GST}})| + |\widehat{T}_n(n_H, \alpha; \lambda, \eta_{\text{GST}}) - T_n(n_H; \lambda, \eta_{\text{GST}})|. \quad (1)$$

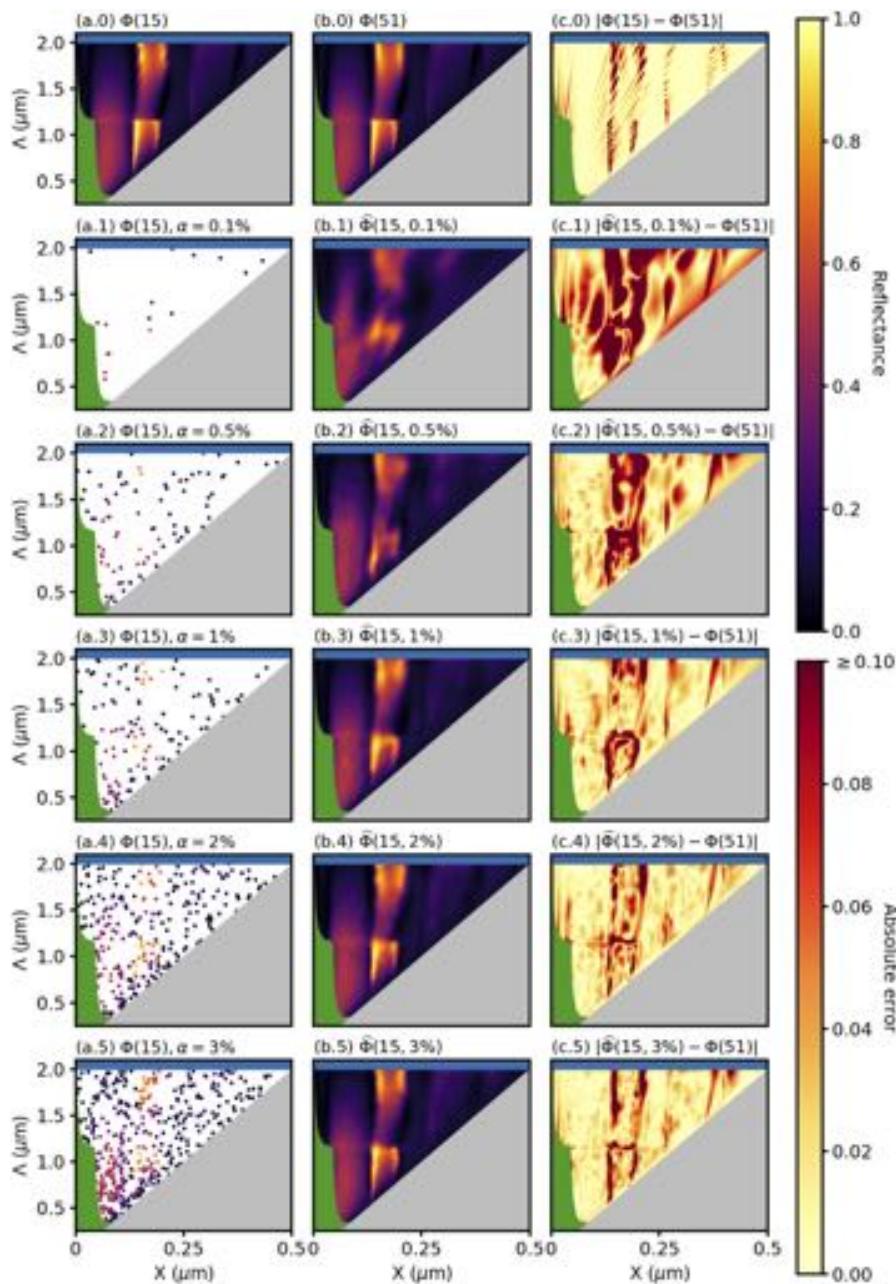


Fig. 3. Simulations of the reflectance (columns (a) and (b)) and absolute error (column (c)) for an $H = 0.5 \mu\text{m}$ cross-section of the HCM parameter space with $\lambda = 2 \mu\text{m}$ s-polarized light incident normal on the 1D GST grating in the amorphous phase-state. The HCM parameter space is bounded by the EM regime (green), diffraction regime (blue), and the unphysical parameter space (gray). In column (a), (a.0) is the RCWA simulated reflectance using $n_H = 15$ and all plots below (i.e. (a.1-a.5)) show the various samplings α used for SPN training. In column (b), (b.0) is the RCWA simulated reflectance using $n_H = 51$ and all plots below (i.e. (b.1-b.5)) are the SPN predicted reflectance that were trained using the data from (a.1-a.5). In column (c), (c.0) is the absolute error between the two RCWA simulations and all plots below (i.e. (c.1-c.5)) are the SPN absolute errors.

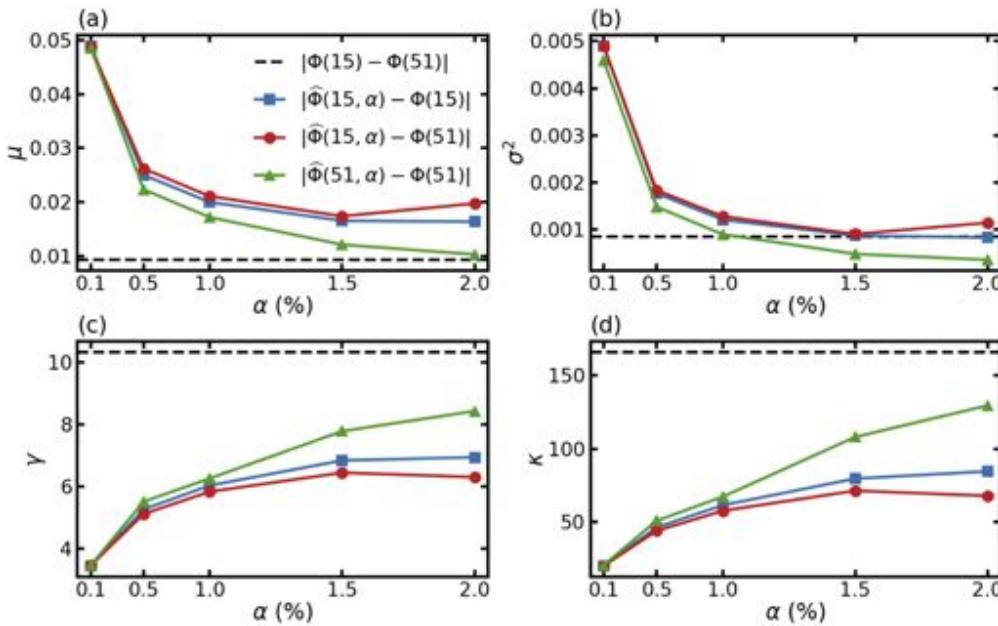


Fig. 4. The first four ($m \in \{1, 2, 3, 4\}$) moments, $E[|\widehat{\Phi} - \Phi|^m]$, of the absolute error for the state $\{1D, 0^\circ, s\}$. The subplots show (a) the mean, μ ; (b) the variance, σ^2 ; (c) the skewness, γ ; and (d) the kurtosis, κ , of the absolute error between SPN and RCWA spectral simulations of various n_H (red, blue, and green lines) and α , as well as the absolute error of RCWA $n_H = 15$ with respect to the ground truth RCWA $n_H = 51$ (black dashed line).

Dividing the RHS of Eq. (1) by the product of the total number of device designs, wavelengths, and phase-states will yield the exact MAE. The total number of devices can be found in the HCM column of Table 2, while the total number of wavelengths is 501, as determined from Table 1, and the total number of phase-states is two (e.g. aGST and cGST). For $\{\theta = 0^\circ, s\}$, this value is $1.14 \times 10^6 \times 501 \times 2 = 1.14 \times 10^9$. As expected the MAE of $\widehat{\Phi}(51, \alpha)$ vs. $\Phi(51)$ continues to decrease for larger training set sizes. However, this is not true for SPNs trained on various fractions of $\Phi(15)$ data, $\widehat{\Phi}(15, \alpha)$. Somewhat unexpectedly, the $\widehat{\Phi}(15, \alpha)$ vs. $\Phi(51)$ MAE decreases up to the training set size $\alpha = 2\%$. Above this, the predictions become further from the ground truth. Yet, we see that $\widehat{\Phi}(15, \alpha)$ continues to decrease its MAE when compared to $\Phi(15)$ simulations. The conclusion drawn is that as the training set size increases past 2%, $\widehat{\Phi}(15, \alpha > 2\%)$ begins to over-predict the partially converged $\Phi(15)$ data. By this, we mean that it "learns" the numerical oscillation errors of $\Phi(15)$ and therefore begins to more poorly predict the ground truth. This trend is further confirmed from the variance of the absolute error, σ^2 , shown in Fig. 4(b). This suggests that not only does the SPN data MAE decrease up to $\alpha \leq 2\%$ before it begins to overfit the approximate $\Phi(15)$ spectra, but so too does the spread of the error. In addition, the variance of the $\widehat{\Phi}(15, 2\%)$ error is the same as $\Phi(15)$.

Surprisingly, this suggests when training SPNs using coarser computational data, it is better practice to under-sample rather than over-sample the space. The most accurate SPN requiring the least computational time was $\widehat{\Phi}(15, 2\%)$ having an MAE of $(1.74 \pm 2.99)\%$. While $\widehat{\Phi}(51, \alpha)$ clearly shows greater accuracy for all α , in order to achieve a similar MAE as $\widehat{\Phi}(15, 2\%)$ would require $\alpha \approx 1\%$. As a result, $\widehat{\Phi}(15, 2\%)$ and $\widehat{\Phi}(51, 1\%)$ give roughly the same accuracy to the ground truth; however, obtaining training data for $\widehat{\Phi}(15, 2\%)$ is 20 \times faster as it only requires $n_H = 15$. Interestingly, while the first and second moments show minimization at $\alpha = 2\%$, the third and fourth moments, skewness γ and kurtosis κ , continue to increase as α increases (Figs.

4(c,d)). Since γ and κ are normalized with respect to σ , only errors greater than σ will contribute to the skewness and kurtosis. As the SPN data gets better at predicting, such that the MAE and its deviation decreases, the rate of decrease in larger errors is slower than that of the rate of variance decrease. Consequently, the "tailedness" of the error distribution grows and either more outliers exist outside of the deviation or those that lie outside of the deviation lie further from the deviation as the error variance decreases. However, the $\Phi(15)$ absolute errors have an even heavier tail than the SPN data due to the fact that most of $\Phi(15)$ is converged, leading to small μ and σ , except for regions of large spectral gradients (seen in Fig. 3(c.0)), which lead to large errors that act as outliers. Further discussion of the absolute error distributions are discussed in Appendix D.

3.2. SPN Benchmarking: Predictions from a merit function

In the previous section, SPNs were trained to mimic high resolution RCWA simulation with average accuracy of $\approx 98\%$. Interestingly, achieving high predictive power only required a sparse sampling of coarsely resolved RCWA training data. However, the imprecision was most severe around areas of high reflectance and transmittance gradients. This can be potentially problematic when trying to identify designs with specific desired qualities. It is feasible that an optimal SPN that exhibits a low average spectral error may still diverge greatly at $\lambda = 2 \text{ } \mu\text{m}$ and consequently miss the optimal switchable mirror design. Ultimately, we want to decipher whether such SPNs are still viable for finding desired devices, such as our switchable mirror at $\lambda = 2 \text{ } \mu\text{m}$, and if enough predictive robustness has been encoded so that they can still identify a neighborhood of designs subject to high reflection and transmission merit functions. To do this, we define an L_2 normalized merit function so that one GST phase-state is compared to an optimal reflectance of 1 and the opposite phase-state is compared an optimal transmittance of 1. This device performance merit function is defined as

$$\|\Phi(\lambda)\|_2 = \min \left\{ \sqrt{\frac{1}{2} [(1 - R(\lambda, \eta_{aGST}))^2 + (1 - T(\lambda, \eta_{cGST}))^2]}, \sqrt{\frac{1}{2} [(1 - R(\lambda, \eta_{cGST}))^2 + (1 - T(\lambda, \eta_{aGST}))^2]} \right\}. \quad (2)$$

Since we do not know beforehand which GST phase-state is best suited as a mirror or window, Eq. (2) calculates both possibilities and keeps the minimum. In addition, the L_2 merit function conveys GST-state dependent reflection and transmission switching ability across the entire wavelength range. Since our interest is in a switching mirror at $\lambda = 2 \text{ } \mu\text{m}$, we only compute the value at $\|\Phi(\lambda = 2 \text{ } \mu\text{m})\|_2$. Effectively, this results in a single numerical value bound between $[0,1]$ for each Ξ , where 0 represents a perfect switching mirror at $\lambda = 2 \text{ } \mu\text{m}$. We calculate this value for every ground truth design, $\Phi(51)$, and the corresponding SPNs, $\widehat{\Phi}(15, \alpha)$. The designs can then be sorted from minimum to maximum $\|\Phi\|_2$, and clustered in the geometric design space, Ξ . After sorting the ground truth designs, we grab the top N devices. We then repeat this for the SPN predictions, where M SPN predicted optimal designs are corroborated with the ground truth N designs. We expect $M \leq N$ with boundary cases $M = N$ representing perfect SPN switchable mirror predictive power and $M = 0$ conveying that none of the SPN's top N designs are corroborated by the ground truth.

The top N switchable mirror designs were investigated within the range $1 \leq N \leq 5 \times 10^4$. Figure 5(a) plots an example the ground truth best $N = 1000$ switchable mirror designs (cluster in blue) using $\Phi(51)$ data compared to the top 1000 designs predicted using various SPN data (cluster in red). M is the number of designs that exist in both the $\Phi(51)$ and $\widehat{\Phi}(15, \alpha)$ top 1000 designs (i.e. cluster overlap). The ratio M/N is plotted in Fig. 5(b). The solid black line with squares shows how much overlap there is between $\Phi(15)$ and $\Phi(51)$ top N designs. We see $\Phi(15)$ clearly performs better (i.e. has better cluster overlap for each value of N) than all $\widehat{\Phi}(15, \alpha)$, which is to be expected since the $\Phi(15)$ data was used to train each of the SPNs.

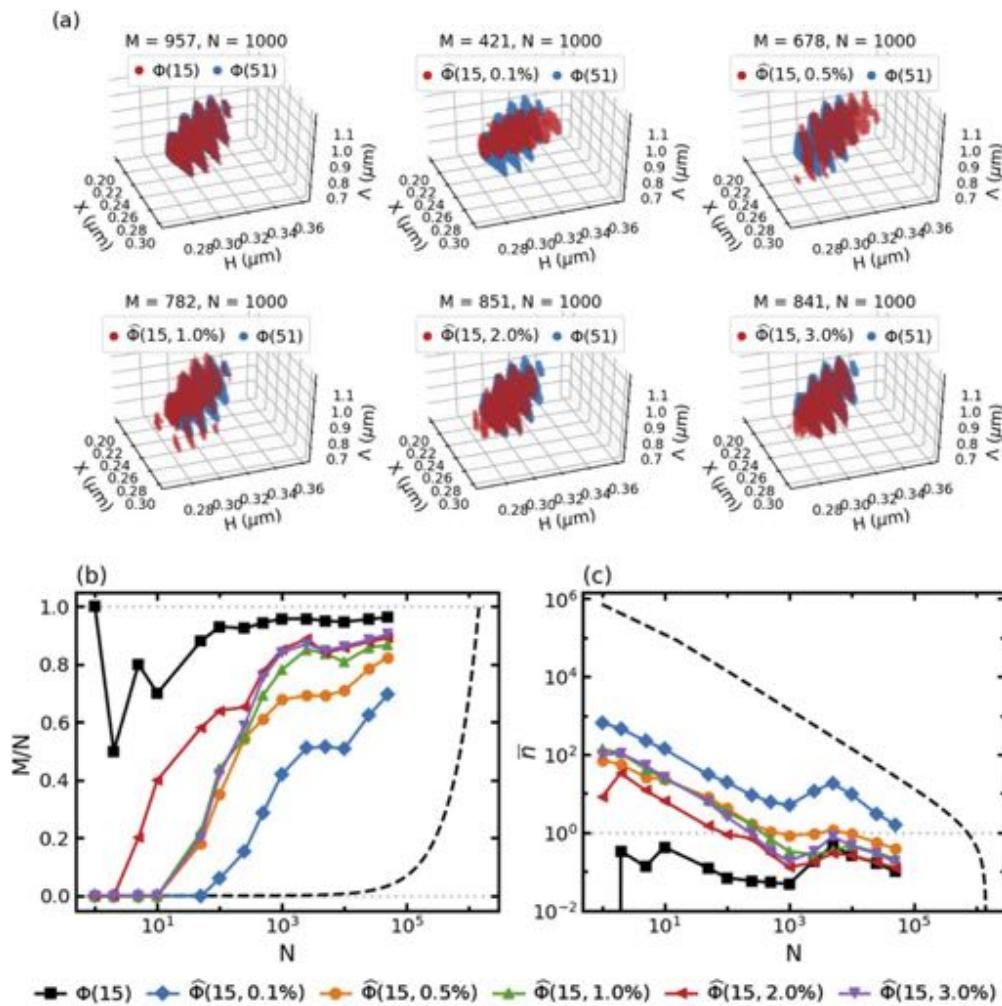


Fig. 5. (a) An example of the clustering of 1D GST bars geometric parameters for the top $N = 1000$ best switchable mirror designs based on RCWA and SPN spectral data. SPN and RCWA top N designs (b) complete cluster overlap ratio M/N and (c) "closeness" to overlap WMID $\bar{n}(\alpha, N)$ are shown. The dashed black lines are the expected values if N designs were randomly selected from the HCM parameter space.

For $2 < N < 100$ we see that in Fig. 5(b) $\widehat{\Phi}(15, 2\%)$ has larger M/N values when compared to the other SPNs and begins to correctly identify top designs for smaller cluster sizes. Training set sizes of 0.5%, 1%, and 3% all share similar M/N plots, requiring $N > 10$ before we see any overlap with the ground truth ($M/N > 0$), while $\alpha = 0.1\%$ clearly performs the worst and requires larger cluster sizes of $N > 100$ before we begin to see any overlap with the ground truth's (i.e. $\Phi(51)$) top N designs. As N increases we would expect all SPNs to have a larger overlap with the ground truth and for M/N to converge to one, which explains why for $N > 100$ the training set sizes 0.5%, 1%, 2%, and 3% all perform roughly the same with a $M/N > 50\%$ overlap with $\Phi(51)$. As N increases even further, we would eventually expect $\widehat{\Phi}(15, 0.1\%)$ to have a similar M/N as the other SPNs. We would expect this trend even if we were to just randomly select N device geometries and compare them to the ground truth's top N designs. The probability of randomly selecting any device design out the HCM parameter space is $1/S$, (S being the total

number of device designs taken from Table 2) so that the probability of all randomly selected N designs are the $\Phi(51)$ top N designs is N/S . Thus, for randomly selecting N designs we would expect $M/N = N/S$, which is plotted in Fig. 5(b) as a dashed black line. While this shows that even randomly selecting N designs will show greater M/N cluster overlap as N increases, since the number of designs is $S \approx 10^6$, N would need to be larger than 10,000 in order to even show a 1% cluster overlap with the ground truth's top N designs. This shows that all SPNs begin to show M/N cluster overlap at cluster sizes many orders of magnitude smaller than if one were to randomly select designs from the HCM parameter space.

The cluster overlap M/N corroborates what we saw with the MAE in Sec. 3.1 and gives us confidence that not only does $\widehat{\Phi}(15, 2\%)$ on average best predict the overall spectra, but this spectra can be trusted specifically at $\lambda = 2 \mu\text{m}$ in order to select several optimal switchable reflectors at the desired wavelength. Notice, however, that $M/N = 0$ for the first few top N designs. This suggests that using the SPN to find a few of the top best performing switchable designs (i.e. small N) may fall short and not in fact find any of the ground truth best designs determined from the $\Phi(51)$ data. Nevertheless, relying solely on M/N is limited because it only compares complete cluster overlap and doesn't take into account how close the SPN predicted designs are to acceptable designs. As a result, we will show that while the SPN may not find the absolute best performing devices, the top designs determined from the SPN data are "close enough", performing similar to the ground truth designs and the differences in metasurface geometries may even in fact fall well within any sort of fabrication tolerances.

To measure the "closeness" between the SPN predicted and ground truth top designs, we calculate the average ranking number, \bar{n} , of the top N designs and find the weighted difference between that of the ground truth, $\Phi(51)$, and the various SPNs shown in Fig. 5(c). All designs in the HCM parameter space are sorted and indexed using L_2 normalization (Eq. (3)) of the $\Phi(51)$ spectral data. Thus, the ground truth top N designs are perfectly sorted such that their average index is given by $\bar{n}_{\Phi(51)}(N) = \sum_{n=1}^N n/N = (N + 1)/2$. Next the top N designs predicted using the SPN spectral data is compared to the ground truth and indexed appropriately. The average of the SPN's indices $\bar{n}_{\widehat{\Phi}(15,\alpha)}(N)$ depends on the various SPNs used, where $\alpha \in \{0.1\%, 0.5\%, 1\%, 2\%, 3\%\}$. The weighted mean index difference (WMID) of the top N sorted design indices between RCWA and SPN data is given by

$$\bar{n}(\alpha, N) = \frac{\bar{n}_{\widehat{\Phi}(15,\alpha)}(N) - \bar{n}_{\Phi(51)}(N)}{\bar{n}_{\Phi(51)}} = \frac{\bar{n}_{\widehat{\Phi}(15,\alpha)}(N)}{\bar{n}_{\Phi(51)}} - 1 = \frac{2\bar{n}_{\widehat{\Phi}(15,\alpha)}(N)}{N + 1} - 1. \quad (3)$$

For example, the true top $N = 5$ designs determined from $\Phi(51)$ are indexed as $\{1, 2, 3, 4, 5\}$ and their average is $\bar{n}_{\Phi(51)}(5) = (5 + 1)/2 = 3$. However, for $\widehat{\Phi}(15, 2\%)$ the top 5 predicted designs are indexed as $\{8, 93, 18, 66, 4\}$. This means that the SPN's predicted 1st best design was in fact the 8th best design, the SPN's predicted 2nd best design was in fact the 93rd best design, and so forth. The average SPN sorted index is $\bar{n}_{\widehat{\Phi}(15,2\%)}(5) = 37.8$ and the WMID given by Eq. (2) is $\bar{n}(2\%, 5) = 2(37.8)/(5 + 1) - 1 = 11.6$. This implies that the difference between the true top $N = 5$ best designs based on $\Phi(51)$ data and that predicted using $\widehat{\Phi}(15, 2\%)$ was off by 11.6 \times . This method is plotted in Fig. 5(c) and gives another way to compare clusters beyond strict overlap (M/N). Smaller $\bar{n}(\alpha, N)$ means the SPN's top designs are more tightly clustered around the true top designs, even if there is not complete overlap. As can be seen in Fig. 5(b), while none of the SPN's predicted best design ($N = 1$) are the true best design, we see that $\widehat{\Phi}(15, 2\%)$ picked a top design that was closer to the true best design compared to any other SPN based on the WMID of their sorted index values (Fig. 5(c)). In particular, the single best design determined using $\widehat{\Phi}(15, 2\%)$ has a WMID an order of magnitude closer to the ground truth compared to $\widehat{\Phi}(15, 0.5\%)$, $\widehat{\Phi}(15, 1\%)$, and $\widehat{\Phi}(15, 3\%)$, and two orders of magnitude closer to the ground truth compared to $\widehat{\Phi}(15, 0.1\%)$. In addition, $\widehat{\Phi}(15, 2\%)$ continues to have at least

an order of magnitude smaller WMID compared to other the SPNs up to $N = 500$. For $N \geq 500$ all SPNs with $\alpha < 3\%$ show similar clustering with small WMIDs, $\bar{n}(\alpha < 3\%, N \geq 500) \leq 1\%$.

If one were to randomly select N designs from the HCM parameter space, the expected average index would be $(S + 1)/2$, where again S is the total number of designs in the HCM space. Thus, the WMID of N randomly selected designs is $\bar{n}_{\text{rand}}(N) = ((S + 1)/2 - \bar{n}_{\Phi(51)}(N))/\bar{n}_{\Phi(51)}(N) = (S - N)/(N + 1)$, shown as a black dashed line in Fig. 5(c). Since $S \approx 10^6$, then $\bar{n}_{\text{rand}}(1) \approx 5 \times 10^5$ and all SPN's WMIDs are several orders of magnitude closer to the $\Phi(51)$ best design. As N increases, all SPN's $\bar{n}(\alpha, N)$ decrease and converge toward zero, which is to be expected even of randomly sampling. Though, we can safely say that all SPN's perform better than random sampling up to $N = 10^6$, at which point $\bar{n}_{\text{rand}}(N \geq 10^6) \leq 1$ and comparable to the SPNs. Figure 5(c) also corroborates the findings of Sec. 3.1, where a training set size of $\alpha = 2\%$ produces an SPN that not only predicts spectra with the minimum MAE, but also gives the best switchable mirror design at $\lambda = 2 \mu\text{m}$ when selected using the $\|\Phi(2 \mu\text{m})\|_2$ merit function (Eq. (2)). While the SPN's best predicted design may not be the true best design based on $\Phi(51)$ data, we can accept with confidence that the predicted best design will be comparable in performance to that of the ground truth.

3.3. Retrieving optimal 1D grating switchable mirror designs

The spectra of optimal switchable mirror designs for s- and p-polarized light of $\lambda = 2 \mu\text{m}$ incident at angles $\theta \in \{0^\circ, 45^\circ\}$ on a 1D grating of GST bars are shown in Fig. 6 in four quadrants—I: $\{0^\circ, \text{s}\}$, II: $\{0^\circ, \text{p}\}$, III: $\{45^\circ, \text{s}\}$, IV: $\{45^\circ, \text{p}\}$. Each $\{\theta, \mathcal{P}\}$ quadrant shows four plots, of which the left column (purple and green lines) is the spectra of the best switchable mirror design determined from $\Phi(51)$ and the right column (red and green lines) is the spectra of the best switchable mirror design determined from $\widehat{\Phi}(15, 2\%)$. The device geometry $\Xi = \{X, H, \Lambda\}$ of the associated aGST and cGST spectra is labeled above each paired plot. Also, $\Phi(51)$ spectral simulations of the top design predicted from $\widehat{\Phi}(15, 2\%)$ are plotted (black dashed and dash-dotted lines) with the SPN predictions for comparison. As can be seen in Fig. 6, the SPN does a remarkable job at predicting the spectra with an enough accuracy so that by using the L_2 merit function to select the top switchable mirror design we were able to find designs that were reasonably close to the ground truth top designs, with the exception to $\{45^\circ, \text{s}\}$ in quadrant III. For quadrants I, II, and IV we see that the feature sizes of the top designs determined from $\widehat{\Phi}(15, 2\%)$ data are within several tens of nanometers of the true top designs. Moreover, for all device geometries shown and at $\lambda = 2 \mu\text{m}$, the amorphous GST phase-state has a maximum normalized power flux (regardless of reflectance or transmittance) of $>95\%$, whereas the crystalline GST phase-state has a maximum normalized power flux of $>80\%$. The cGST phase-state suffers performance due to having more absorption than the aGST phase-state (refer to Appendix B.). Depending on the incident light inclination angle and polarization, either of the aGST or cGST states can act as a mirror or window.

The outlier in the SPN and RCWA predicted best switchable mirror designs is for the $\{45^\circ, \text{s}\}$ case. We see in quadrant III of Fig. 6 that the SPN's predicted top design has a lattice spacing off by nearly 100 nm and a height off by 200 nm from the true best design. By observing the true best design spectra (green and purple lines), we see that there is a resonance at $\lambda = 2 \mu\text{m}$ for the aGST state. Due to the high-Q quality factor of the spectra at this wavelength and the associated large spectral gradient, the SPN misses this feature. This is seen in the adjacent aGST plot, where the SPN predicted spectra (red and blue lines) do not show the sharp resonance that truly exists near $\lambda \approx 1.75 \mu\text{m}$ (black dashed and dash-dotted lines). For this reason, when relying on the SPN spectral data to select the optimal switchable mirror design, we see a different device geometry that shifts the spectra so that the broader transmission peak (in blue) lies at $\lambda = 2 \mu\text{m}$ and the missed high resonance peak is shifted to a different λ . Nevertheless, while the SPN's

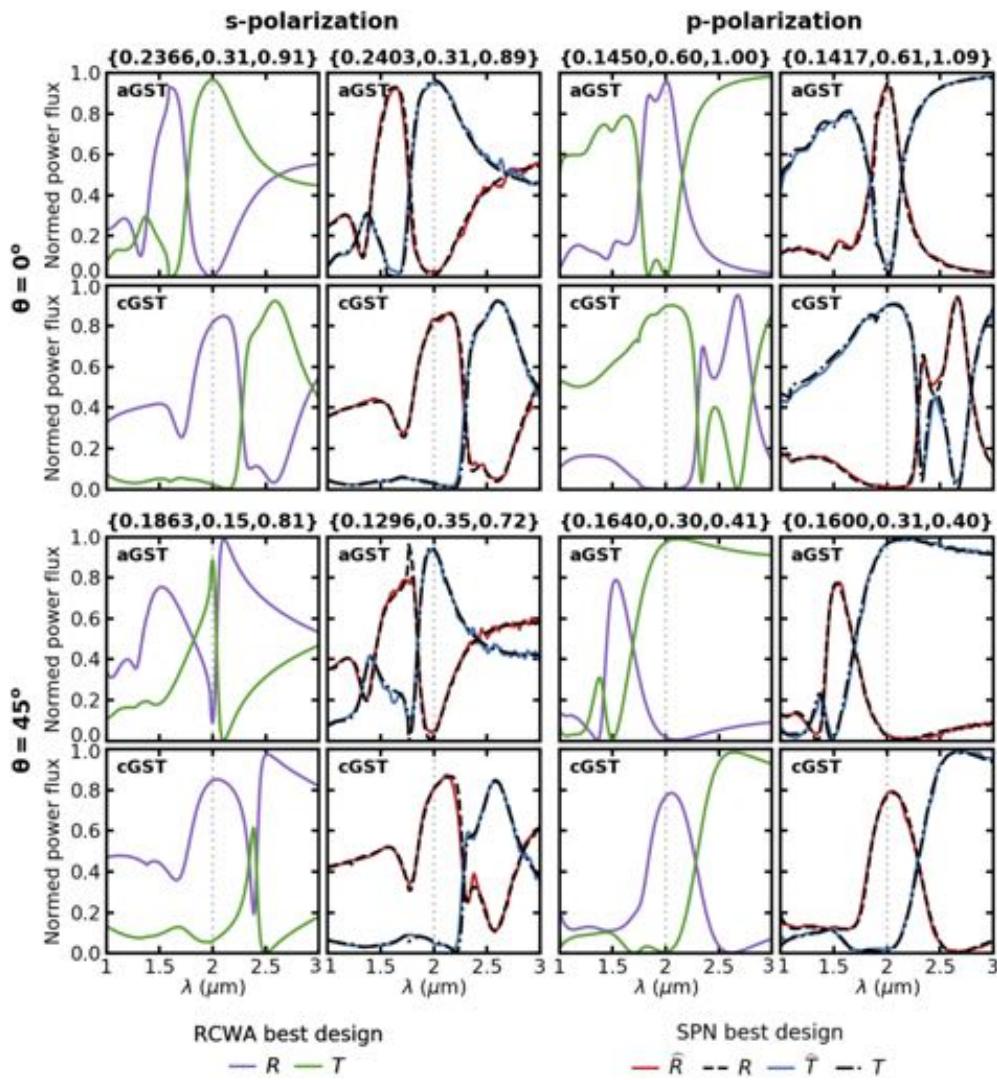


Fig. 6. Spectra comparisons of the true best switchable mirror designs using $\Phi(51)$ data (purple and green lines) to that predicted using $\widehat{\Phi}(15, 2\%)$ data (red and blue lines) for the four permutations of incident light angle and polarization. The $\Phi(51)$ spectra for the SPN predicted best design is overlaid (black dashed and dash-dotted lines). The metasurface geometry $\Xi = \{X, H, \Lambda\}$ is given above each aGST/cGST paired spectra in units of μm .

predicted top design differs from the true top design, the design predicted using $\widehat{\Phi}(15, 2\%)$ still produces comparable switching capabilities as the true top design.

4. 2D GST case: Retrieving the optimal polarization and angle independent pillar array switchable mirror design

We randomly sampled the HCM regime parameter space of the 2D pillar array for normal incidence, as well as for 45° , for s- and p-polarizations, where we neglected simulating the $\{0^\circ, p\}$ case because it would have been identical to the $\{0^\circ, s\}$ case. Table 3 shows the number

of simulations necessary to fill the HCM regime for the spectral range $1 \mu\text{m} \leq \lambda \leq 3 \mu\text{m}$ and pillar height range $0.01 \mu\text{m} \leq H \leq 1 \mu\text{m}$. Like the 1D case, the 2D case also has a reduction in the non-diffractive regime of $\approx 50\%$ when using modal analysis and restricting the parameter space to the HCM regime. While the size of the the HCM parameter space of the 2D GST pillar array arranged in a square lattice is comparable to that of the 1D GST grating, due to the increases dimensionality of the 2D case, what would have taken n_H harmonics to lead to spectral convergence for the 1D case would require at least n_H^2 harmonics to do the same for the 2D case. Therefore, we need many more harmonics ($15^2 = 225$ as determined from Sec. 3) to approach the same level of convergence seen with the 1D grating RCWA simulations. Moreover, since the time of convergence is proportional to the number of harmonics cubed ($t \propto n_H^3$) as also demonstrated in Sec. 3, then $t_{n_H^2}/t_{n_H} = (n_H^2)^3/n_H^3 = n_H^3$, where $n_H = 15$ and $n_H^2 = 225$ for the 1D and 2D cases, respectively. This means that the 2D case takes $\approx 3375\times$ longer to simulate. Hence, what had taken $\Phi(15)$ (1D case) hours to simulate the HCM parameter space would have taken $\Phi(15^2)$ (2D case) about a year to simulate. Fortunately, we needed only to sample a small fraction of the 2D pillar array HCM parameter space to train our SPN. We chose to randomly sample $\alpha \approx 2\%$ of the 2D pillar array HCM parameter space since our 1D simulations had suggested this to be optimal for minimizing SPN predicting inaccuracies (i.e. minimal MAE). Thus, after days of RCWA simulations, we were able to train our SPN and fill the rest of the 2D pillar array HCM parameter space using the SPN in minutes. In Table 3 the number of simulations that we actually performed, and time to complete these simulations, are shown in bold.

Table 3. Training data collection: 2D GST pillar array metasurface.

θ ($^\circ$)	P {s, p}	Amount of obtained data in format Ψ					Computetime	
		ND ^a (million)	HCM ^b (million)	HCM/ND (%)	RS ^c (thousand)	RS/HCM (%)	HCM (days)	RS (days)
0	s	2.03	1.19	58.6	25.21	2.2	363.9	7.7
45	s	1.19	0.49	41.2	13.67	2.8	149.8	4.2
45	p	1.19	0.55	46.2	14.41	2.6	168.2	4.4

^aND: nondiffractive regime consists of EM and HCM; EM: effective medium regime;

^bHCM: high contrast metasurface regime;

^cRS: random sampling of HCM.

The three SPNs—one for each of the $\{0^\circ, \text{s}\}$, $\{45^\circ, \text{s}\}$, and $\{45^\circ, \text{p}\}$ cases—were trained on 90% of the randomly sampled (RS) design data (Table 3) and 10% was put aside to validate that the SPNs were not overfitting the training data. Using the validation data we were able to compare the SPN predictions to the ground truth. Unlike the 1D case, we are unable to simulate the entire HCM parameter space with RCWA and must assume that the validation data is an accurate representation of the entire HCM space. For the $\{0^\circ, \text{s}\}$, $\{45^\circ, \text{s}\}$, and $\{45^\circ, \text{p}\}$ cases we found the MAEs to be $(1.69 \pm 3.38)\%$, $(1.69 \pm 3.74)\%$, and $(1.57 \pm 3.67)\%$, respectively. This is comparable to the accuracies seen by the 1D GST grating SPNs in Sec. 3.1, Fig. 4. Using the L_2 error function defined in Eq. (3), we sorted the designs from best to worst based on their switchable mirror capabilities. Similarity in predicted top switchable mirror design selected for each of the three cases (refer to Appendix E.) suggested there existed a single design that was polarization and angle independent. Consequently, a new L_2 function was defined as an average of the individual error functions, $\overline{||\Phi||_2} = (||\Phi_{\{0^\circ, \text{s}\}}||_2 + ||\Phi_{\{45^\circ, \text{s}\}}||_2 + ||\Phi_{\{45^\circ, \text{p}\}}||_2) / 3$. This allowed us to comb through all the data from the three SPN predicted HCM spectra and select a design that optimized switching capabilities for all three cases. From this we determined a single optimal design $X = 0.195 \mu\text{m}$, $H = 0.33 \mu\text{m}$, $\Lambda = 0.45 \mu\text{m}$ that exhibited switching capabilities for the three angle and polarization combinations, whose spectra is shown in Fig.

7(a). As can be seen, the SPN does a remarkable job at predicting the reflection and transmission spectra (red and blue lines, respectively) when compared to the RCWA simulations (dashed and dash-dotted). Figure 7(b) shows the polarization, \mathcal{P} , inclination angle, θ , and azimuthal angle, ϕ , dependence of the aGST state transmittance and the cGST state reflectance. As can be seen, there is no ϕ dependence and very little θ dependence. For all ϕ and for $\theta \leq 45^\circ$ this metasurface design shows s-polarized light has $88\% \lesssim T \lesssim 95\%$ for aGST and $78\% \lesssim R \lesssim 83\%$ for cGST, and p-polarized light has $95\% \lesssim T \lesssim 98\%$ for aGST and $69\% \lesssim R \lesssim 78\%$ for cGST. Using the SPN to exhaustively fill the HCM parameter space allowed us to discover a GST pillar array

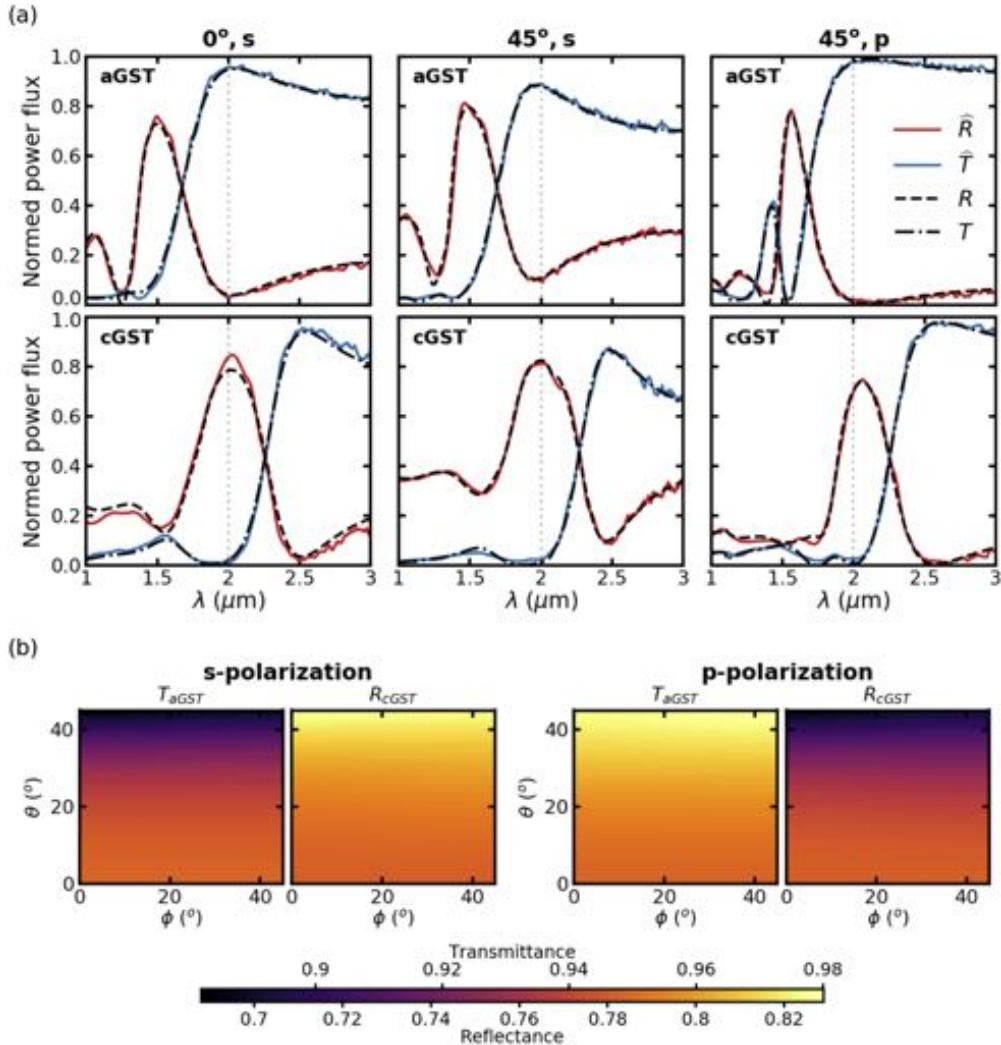


Fig. 7. Spectral plots of the 2D pillar array metasurface ($X = 0.195 \mu\text{m}$, $H = 0.33 \mu\text{m}$, $\Lambda = 0.45 \mu\text{m}$) optimized as an angle and polarization independent switchable mirror using $\widehat{\Phi}(15^2, \approx 2\%)$ data. (a) SPN predicted reflectance (red line) and transmittance (blue line), along with the approximated ground truth $\Phi(15^2)$ reflectance (black dashed line) and transmittance (dash-dotted line). (b) 2D pillar array of RCWA $\Phi(15^2)$ optimal switchable mirror design spectral plots for $\lambda = 2 \mu\text{m}$ s- and p-polarized light. Shown are the amorphous state transmittance and crystalline state reflectance.

metasurface design that is capable of acting as a switchable mirror for any light polarization incident at angles $\theta \leq 45^\circ$.

5. Conclusions

We have investigated a general approach for using an SPN to discover novel dynamic and selective optical components, and demonstrated its utility by optimizing a GST metasurface for use as a switchable mirror. This approach affords us the opportunity to search a parameterized feature space many orders of magnitude faster than with conventional approaches. For our dynamic mirror baseline metasurface features, we made use of both a linear bar and a cylindrical pillar geometry and trained an ANN to simulate their optical behavior as a switchable mirror. Parameter spaces were constrained to the high-contrast metasurface regime using modal analysis, effectively reducing our parameter space in half and limiting our search area only to regions that could produce high reflectivity and transmissivity. Through a comprehensive study of the 1D GST grating we found that a coarser simulation that used less harmonics ($n_H = 15$), and a sampling size of only about a 2% of the HCM parameter space, was required to train the SPN to accurately interpolate the full range of the design space with an MAE of $<2\%$. Knowledge from the 1D GST study was then used for the 2D case, where less harmonics and small HCM sampling, along with the SPN predicting capabilities, greatly improved simulation times and allowed us to simulate the entire 2D GST pillar array metasurface parameter space when it would have been impossible to do otherwise. The ability to completely fill a large multidimensional parameter space allowed us to then select the best switchable mirror design possible. Four separate 1D GST grating designs were found for the four permutations of light inclination angles, $\{0^\circ, 45^\circ\}$, and polarizations, $\{s, p\}$, which showed roughly between 80% – 95% reflectance and transmittance between the two phase-states for $\lambda = 2 \mu\text{m}$. In addition, a single 2D GST pillar array was found that showed roughly between 75% – 95% reflectance and transmittance between the two phase-states for all azimuthal angles, ϕ , and for inclination angles $\theta \leq 45^\circ$.

Though, these trained SPNs have the added benefit that they require less memory to store than saving the entire RCWA parameter space spectral data. Our SPNs had nearly 80 million trained weights, whereas the entire HCM spectra would require roughly 2 billion data points—each individual spectra has about 2000 elements and there are about one million metasurface designs in the HCM space per simulation group—making the SPNs about 25 \times smaller. In addition, it takes the SPN only minutes to re-simulate the entire HCM spectra. This given flexibility would be lost had we taken a genetic algorithm approach to find an optimal switchable mirror device. Since the entire HCM spectra can be quickly retrieved via the SPN, one can define any number of merit functions to find a device with a specific purpose and performance without the need to do time consuming physics simulations.

Appendix A List of important symbols

λ	Excitation planewave wavelength in range [1 μm , 2 μm]; refer to Table 1
θ	Incident light inclination angle in set $\{0^\circ, 45^\circ\}$
ϕ	Incident light azimuthal angle
\mathcal{P}	Collection of polarization states, {s, p}
\mathcal{D}	Dimensions of GST metasurface: 1D for grating, 2D for pillar array
X	Feature half-width (radius) for 1D (2D) metasurface; refer to Table 1
H	Feature height of the GST metasurface in range [0.05 μm , 1 μm]; refer to Table 1
Λ	Lattice spacing of the GST metasurface; refer to Table 1
X'	Scaled half-width (radius), X/Λ in range [0, 0.5]; refer to Table 1
$\eta_{\text{GST}}(\lambda)$	The set of the wavelength dependent complex refractive index for amorphous and crystalline GST phase-states, $\{\eta_{\text{aGST}}(\lambda), \eta_{\text{cGST}}(\lambda)\}$
n_{H}	Number of spatial harmonics used in RCWA simulation
α	Fraction of the HCM parameter space used for SPN training
Ξ	Collection of device geometric parameters, $\{X, H, \Lambda\}$, used for SPN inputs
$\Phi(n_{\text{H}})$	Collection of RCWA spectra simulated with n_{H} harmonics and used for training the SPN outputs, $\{R(\lambda, \eta_{\text{aGST}}), T(\lambda, \eta_{\text{aGST}}), R(\lambda, \eta_{\text{cGST}}), T(\lambda, \eta_{\text{cGST}})\}$
$\widehat{\Phi}(n_{\text{H}}, \alpha)$	Collection of SPN spectra that was trained with α fraction of $\Phi(n_{\text{H}})$ data, $\{\widehat{R}(\lambda, \eta_{\text{aGST}}), \widehat{T}(\lambda, \eta_{\text{aGST}}), \widehat{R}(\lambda, \eta_{\text{cGST}}), \widehat{T}(\lambda, \eta_{\text{cGST}})\}$
Ψ	Collection of SPN input and output training data, $\{\Xi, \Phi\}$
μ	Mean absolute error (MAE) between SPN and RCWA spectra defined by Eq. (1), aka L_1 norm
σ	Standard deviation of absolute errors
γ	Skewness of absolute errors
κ	Kurtosis of absolute errors
$ \Phi(15) - \Phi(51) $	The MAE, μ , between partially converged RCWA spectra and the ground truth
$ \widehat{\Phi}(n_{\text{H}}, \alpha) - \Phi(51) $	The MAE, μ , between SPN spectra and the ground truth
$ \Phi(\lambda) _2$	L_2 norm merit function used to rank optimal switchable mirror designs at wavelength λ , defined by Eq. (2)
$\overline{ \Phi _2}$	L_2 norm merit function used to rank optimal angle and polarization independent 2D switchable mirror designs at $\lambda = 2 \mu\text{m}$, $(\Phi_{\{0^\circ, \text{s}\}} _2 + \Phi_{\{45^\circ, \text{s}\}} _2 + \Phi_{\{45^\circ, \text{p}\}} _2)/3$
N	Number of top designs selected after sorting using L_2 norm merit function, $ \Phi(\lambda) _2$
M	Number of SPN's top N designs that are in common with the ground truth
S	Total number device designs in HCM parameter space; refer to Table 2
$\bar{n}_{\Phi(51)}(N)$	Average sorted index of top N designs, $(N + 1)/2$
$\bar{n}_{\widehat{\Phi}(15, \alpha)}(N)$	Average sorted index of SPN predicted top N designs
$\bar{n}(\alpha, N)$	Weighted mean index difference (WMID) between SPN and ground truth, Eq. (3)

Appendix B GST material dispersion

GST can refer to several different ternary mixtures of germanium, antimony, and tellurium (Fig. 8(a)). The material dispersion implemented in our simulations is based off a common $\text{Ge}_2\text{Sb}_2\text{Te}_5$ formulation. The wavelength dependent real and imaginary refractive index values of both phase-states, $\eta_{\text{aGST}}(\lambda)$ and $\eta_{\text{cGST}}(\lambda)$, are displayed in Figs. 8(b,c), along with the refractive index values of the alumina substrate, Al_2O_3 . The Al_2O_3 refractive index values were taken from Ref. [77].

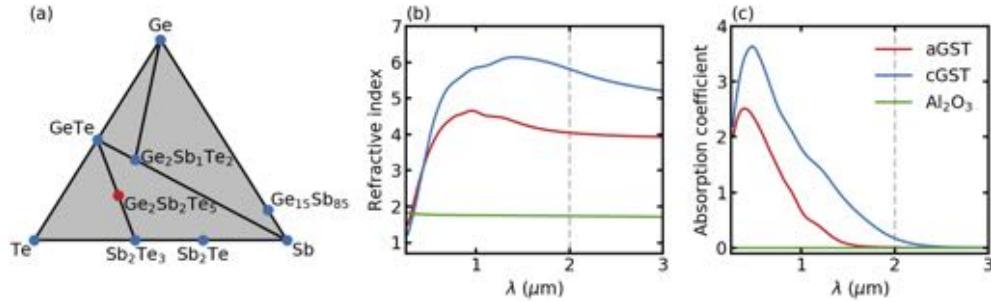


Fig. 8. Material Dispersion curves used in our simulations. (a) The Ge-Sb-Te ternary diagram, where the specific GST mixture we used in this manuscript, $\text{Ge}_2\text{Sb}_2\text{Te}_5$, is displayed in red. The (b) real and (c) imaginary refractive index values for both the amorphous and crystalline GST phase-states, as well as Al_2O_3 , are shown as red, blue, and green lines, respectively.

Appendix C Parameter space restriction using PhC bands

As briefly discussed in Sec. 2.1, the HCM regime exhibits high reflectivity and transmissivity due to the interference of WGAMs at the metasurface interface. A more thorough explanation can be found in Refs. [23,72,73]. Figure 9 shows an example of a dispersion relationship between normalized wavevectors $k_x\Lambda/(2\pi)$, $k_z\Lambda/(2\pi)$, and Λ/λ for the first six modes using the MPB plane-wave expansion method. In this example we simulate a 1D waveguide array (infinite in y and z , and periodic in x) of alternating air and GST layers that has a scaled width $X/\Lambda = 0.5$ and uses the GST refractive index $\eta_{\text{aGST}}(\lambda = 2 \mu\text{m})$. MPB scales all dimensions by the lattice spacing Λ . However, GST is dispersive and for our modal analysis we used $\eta_{\text{aGST}}(\lambda = 2 \mu\text{m})$. Thus, when we look at cutoff values of Λ/λ in MPB, we are actually describing cutoff values for Λ since λ is fixed. This is how we retrieved the Λ_{low} discussed in Sec. 2.1.

Figure 9(b) shows the WGAM dispersions. For light incident at $\theta = 0^\circ$ on the 1D GST grating, WGAMs would be excited propagating in the k_z direction, where $k_x = 0$. This is shown for s- and p-polarizations as red and blue lines, respectively. However, for light incident at $\theta = 45^\circ$ on the 1D GST grating, WGAMs would be excited propagating in the k_z direction, but there would also be a k_x component determined by $k_x = (2\pi/\lambda) \sin \theta$. This changes the dispersion slightly as shown by the black dashed and dotted lines for s- and p-polarizations, respectively. While the angle of incidence causes a change in the modal Λ/λ cutoff values, shown in Fig. 9(b) as circles for 0° and squares for 45° , the polarization does not. This is better understood by noticing the relationship between WGAMs and photonic crystal (PhC) modes.

The 1D GST WGA can also be treated as a 1D PhC (Bragg stack) by observing the dispersion of k_x through the material and keeping $k_z = 0$. This results in the periodic PhC band diagram shown in Fig. 9(c). Since $k_z = 0$, the PhC bands map out the modal Λ/λ cutoff values of the WGAMs for different incident angles, θ . By extending the PhC band beyond the irreducible Brillouin zone (IBZ) and using the relationship $k_x = (2\pi/\lambda) \sin \theta$, one can determine Λ/λ cutoff

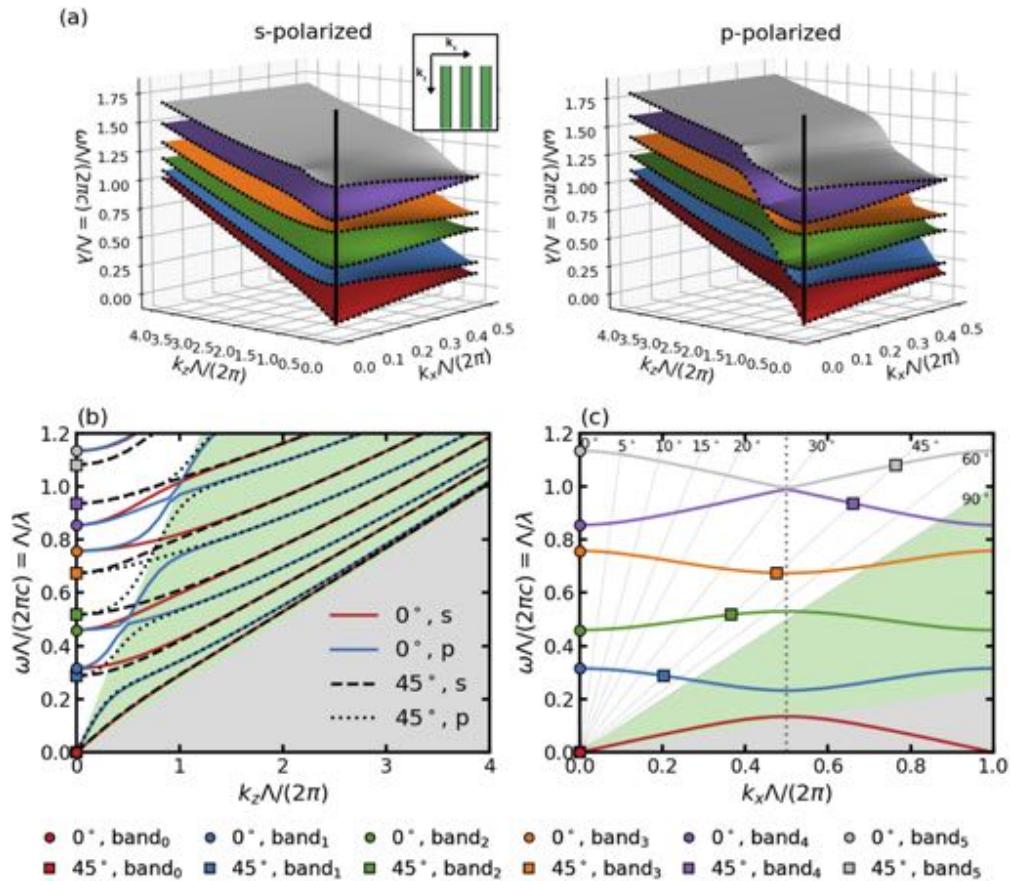


Fig. 9. Band diagrams for a 1D aGST waveguide array (aka Bragg stack, 1D photonic crystal) that is infinite in y and z , and periodic in x , with $\lambda = 2 \mu\text{m}$ and $X/\Lambda = 0.5$. (a) The dispersion relationship between $k_z\Lambda/(2\pi)$, $k_x\Lambda/(2\pi)$, and Λ/λ for s- and p-polarized light, where the different colored sheets represent different modal bands. The insert shows a schematic of the WGA. (b) The WGAMs for $k_x\Lambda/(2\pi) = 0$ (colored lines) and $k_x\Lambda/(2\pi) = (\Lambda/\lambda) \sin(45^\circ)$ (black dashed and dotted lines). The WGAMs' Λ/λ cutoff values are marked as colored circles and squares, as also shown in the associated diagram (c). The gray region marks the area below the aGST light-line and the green region marks the area below the air light-line. (c) The photonic crystal band diagram for $k_z = 0$. Above the air light-line (white region) shows the k_x angle dependence, $k_x\Lambda/(2\pi) = (\Lambda/\lambda) \sin \theta$, where the $\theta = 0^\circ$ circles and $\theta = 45^\circ$ squares mark the WGAMs' Λ/λ cutoff values seen in (b).

for each WGAM from $k_x \Lambda / (2\pi) = (\Lambda / \lambda) \sin \theta$. The values of Λ / λ for the various modes with light incident at $\theta = 0^\circ$ and 45° are marked as circles and squares, respectively, in Fig. 9(c). These marked values are the cutoff Λ / λ values seen in Fig. 9(b). Thus, when trying to determine the smallest value of Λ_{low} such that there are at least two WGAMs that propagate within the 1D GST grating metasurface, one simply needs to solve for the first excited PhC band (blue line in Fig. 9(c)) and take the value of Λ / λ at $k_x \Lambda / (2\pi) = (\Lambda / \lambda) \sin \theta$, where in our example $\lambda = 2 \mu\text{m}$ and $\theta \in \{0^\circ, 45^\circ\}$.

In addition, since the cutoff value for Λ_{low} is determined in the PhC scheme when $k_z = 0$, there is only propagation in k_x , and hence, transverse electric and magnetic fields will lie solely in the yz -plane. However, the 1D GST PhC is invariant in the yz -plane and thus behaves identically for s- and p-polarizations. Again, this is seen in Fig. 9(b), where s- and p-polarizations of the same incident angle, θ , have different WGA dispersions, but share the same cutoff Λ / λ values. Thus, s- and p-polarizations with the same θ will have the same HCM parameter space size (shown in Table 2). Though, in the case of the 2D cylindrical GST WGA (i.e. cylinders that are periodic in x and y , and infinite in z), symmetry between the y and z dimension is broken, and the PhC band diagram becomes nondegenerate. There is a split between s- and p-polarization Λ / λ cutoff values, resulting in different HCM parameter space sizes that can be observed in Table 3.

Appendix D SPN error dispersion

As in Sec. 3.1, we will only discuss the {1D, 0° , s} case. The absolute errors of the $\Phi(15)$ and $\widehat{\Phi}(15, \alpha)$ with respect to the ground truth, $\Phi(51)$, were binned between [0,1] with bin sizes of 0.001, and the counts normalized so that the integration over all bins equaled one. This probability density function is shown in Fig. 10(a). As expected, the insert shows that $\Phi(15)$ has a much higher probability for small errors (<0.3%) when compared to any of the SPNs. However, $\Phi(15)$ also shows a greater probability for extremely large errors (>75%) when also compared to any of the SPNs—albeit the probability of $\Phi(15)$ having larger errors is roughly three orders of magnitude less than the probability of having smaller errors. By referring to the cumulative distribution function in Fig. 10(b), we see that nearly 50% of the $\Phi(15)$ absolute error

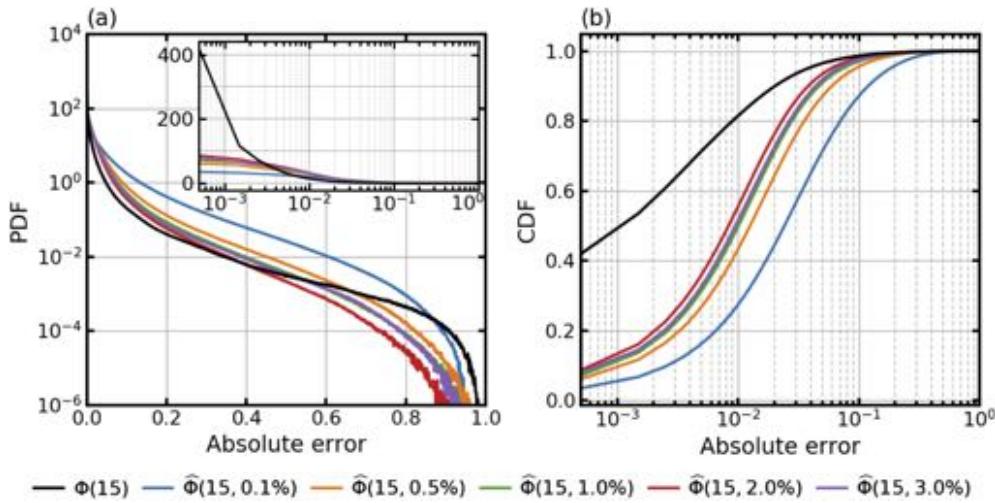


Fig. 10. (a) The probability density function of the absolute error of the approximate $\Phi(15)$ (black line) and $\widehat{\Phi}(15, \alpha)$ (colored lines) with the ground truth $\Phi(51)$. The insert emphasizes small absolute error, whereas the main figure emphasizes the PDF tail. (b) The cumulative distribution function of the various absolute errors.

was $<0.1\%$ and about 80% had error $<1\%$. Yet $\Phi(15)$ has a MAE of $\approx 0.9\%$ and a variance comparable to that of $\widehat{\Phi}(15, 2\%)$ (Fig. 4(a,b)). This shows that the vast majority of the $\Phi(15)$ has converged to the ground truth $\Phi(51)$, but that there also exists outliers that exhibit very large error magnitudes, which can be seen in regions that have large spectral gradients (Fig. 3(a.0-c.0)). These outliers contribute to the larger variance seen by $\Phi(15)$. The SPNs show a greater spread in error magnitudes, insofar that the probability of absolute errors between 0.3% and 75% is greater than that of $\Phi(15)$. Moreover, by increasing the training set size, α , of the SPNs up to 2%, the PDF tail appears to decrease. This suggests that the SPN is increasing its accuracy with respect to the ground truth and not fitting the portions of the $\Phi(15)$ spectra that have large errors. Nevertheless, the kurtosis of $\widehat{\Phi}(15, \alpha)$ does consistently increase as α increases due to the rapid decrease in the variance (Fig. 4(d)). From Fig. 10(b) we see that roughly 95% of the $\widehat{\Phi}(15, 2\%)$ spectra have an absolute error less than 5.7%. In other words, 95% of the spectral data lies within one standard deviation from the MAE, $(2.1 \pm 3.6)\%$, taken from Figs. 4(a,b).

Appendix E Optimal 2D GST pillar array switchable mirrors

As discussed in Sec. 4, three separate SPNs were trained for the 2D pillar array metasurface, one for each of the states $\{\theta = 0^\circ, \mathcal{P} = s\}$, $\{45^\circ, s\}$, $\{45^\circ, p\}$, where $\{0^\circ, p\}$ was neglected due to symmetry. These SPNs were then used to fill the entire HCM parameter space and with the L_2 norm function the optimal design for each state was determined. Figure 11 shows the spectra for the top design for each of the three states. The optimal 2D metasurface geometries for the $\{0^\circ, s\}$, $\{45^\circ, s\}$, $\{45^\circ, p\}$ states were $\Xi = \{0.2 \text{ } \mu\text{m}, 0.33 \text{ } \mu\text{m}, 0.48 \text{ } \mu\text{m}\}$, $\{0.21 \text{ } \mu\text{m}, 0.33 \text{ } \mu\text{m}, 0.5 \text{ } \mu\text{m}\}$, and $\{0.18 \text{ } \mu\text{m}, 0.34 \text{ } \mu\text{m}, 0.43 \text{ } \mu\text{m}\}$, respectively, where $\Xi = \{X, H, \Lambda\}$. As can be seen, each of the GST metasurface parameters of these three designs are within a few tens of nanometers of each other. This motivated us to define a new L_2 error function and search for a single 2D pillar array that acted as a switchable mirror for all angles and polarizations.

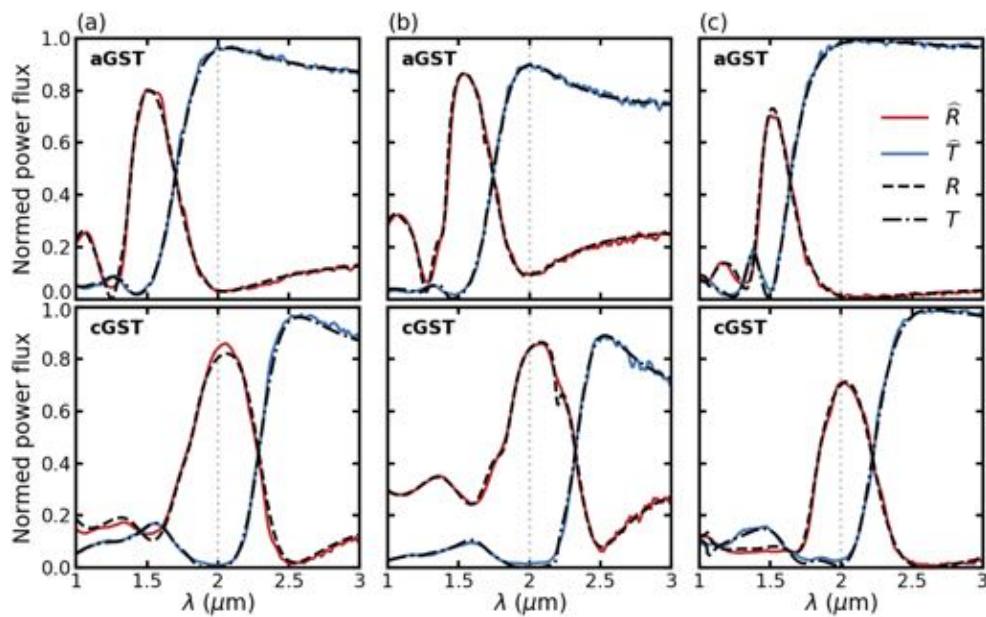


Fig. 11. Optimal switchable mirror 2D pillar array designs and spectra for $\lambda = 2 \mu\text{m}$ determined from $\hat{\Phi}(15^2, \approx 2\%)$ data. For light incident parameters (a) $\{0^\circ, \text{s}\}$ the optimal design is $\Xi = \{0.2 \mu\text{m}, 0.33 \mu\text{m}, 0.48 \mu\text{m}\}$; (b) $\{45^\circ, \text{s}\}$ the optimal design is $\Xi = \{0.21 \mu\text{m}, 0.33 \mu\text{m}, 0.5 \mu\text{m}\}$; and (c) $\{45^\circ, \text{p}\}$ the optimal design is $\Xi = \{0.18 \mu\text{m}, 0.34 \mu\text{m}, 0.43 \mu\text{m}\}$.

Funding

Air Force Research Laboratory (FA8650-16-D-5404-0013).

Acknowledgments

We would like to thank J. Vernon for discussions surrounding the manuscript. We would also like to thank the reviewers for their inquiries and contributions toward the clarity of the manuscript.

Disclosures

The authors declare no conflicts of interest.

References

1. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, 2002).
2. P. J. Collings and M. Hird, *Introduction to Liquid Crystals: Chemistry and Physics* (CRC Press, 2017).
3. S. R. Ovshinsky, “Reversible electrical switching phenomena in disordered structures,” *Phys. Rev. Lett.* **21**(20), 1450–1453 (1968).
4. S. Chen, P. B. Gibbons, and S. Nath, “Rethinking database algorithms for phase change memory,” in *5th Biennial Conference on Innovative Data Systems Research (CIDR)*, (CIDR, 2011).
5. M. Wuttig and N. Yamada, “Phase-change materials for rewriteable data storage,” *Nat. Mater.* **6**(11), 824–832 (2007).
6. M. Wuttig, H. Bhaskaran, and T. Taubner, “Phase-change materials for non-volatile photonic applications,” *Nat. Photonics* **11**(8), 465–476 (2017).
7. A. V. Kolobov, P. Fons, and J. Tominaga, “Phase-change optical recording: Past, present, future,” *Thin Solid Films* **515**(19), 7534–7537 (2007).
8. M. Chen, K. A. Rubin, and R. W. Barton, “Compound materials for reversible, phase-change optical data storage,” *Appl. Phys. Lett.* **49**(9), 502–504 (1986).
9. I. Satoh and N. Yamada, “DVD-RAM for all audio/video, PC, and network applications,” in *Fifth International Symposium on Optical Storage (ISOS 2000)*, vol. 4085 (SPIE, 2001), pp. 283–290.

10. S. Raoux, F. Xiong, M. Wuttig, and E. Pop, "Phase change materials and phase change memory," *MRS Bull.* **39**(8), 703–710 (2014).
11. T. Ohta, "Phase-change optical memory promotes the DVD optical disk," *J. Optoelectron. Adv. Mater.* **3**, 609–626 (2001).
12. M. H. Lankhorst, B. W. Ketelaars, and R. A. Wolters, "Low-cost and nanoscale non-volatile memory concept for future silicon chips," *Nat. Mater.* **4**(4), 347–352 (2005).
13. N. Yamada, M. Takao, and M. Takenaga, "Te-Ge-Sn-Au phase change recording film For optical disk," in *Optical Mass Data Storage II*, vol. 0695 (SPIE, 1987).
14. E. Ohno, N. Yamada, T. Kurumizawa, K. Kimura, and M. Takao, "Tegesnau alloys for phase change type optical disk memories," *Jpn. J. Appl. Phys.* **28**(Part 1, No. 7), 1235–1240 (1989).
15. J. Feinleib, J. Deneufville, S. C. Moss, and S. R. Ovshinsky, "Rapid reversible light-induced crystallization of amorphous semiconductors," *Appl. Phys. Lett.* **18**(6), 254–257 (1971).
16. P. Guo, A. M. Sarangan, and I. Agha, "A review of germanium-antimony-telluride phase change materials for non-volatile memories and optical modulators," *Appl. Sci.* **9**(3), 530 (2019).
17. K. Shimakawa, A. Kolobov, and S. Elliott, "Photoinduced effects and metastability in amorphous semiconductors and insulators," *Adv. Phys.* **44**(6), 475–588 (1995).
18. G. Keiser, *Optical Fiber Communications* (McGraw-Hill Education, 2010), 4th ed.
19. S. Février, "Photonic crystal fibers," *Advanced Fiber Optics* **299**, 87–125 (2011).
20. F. Poletti, M. N. Petrovich, D. J. Richardson, and M. J. Li, "Hollow-core photonic bandgap fibers: Technology and applications," *Nanophotonics* **2**(5-6), 315–340 (2013).
21. N. I. Zheludev and Y. S. Kivshar, "From metamaterials to metadevices," *Nat. Mater.* **11**(11), 917–924 (2012).
22. C. M. Soukoulis and M. Wegener, "Past achievements and future challenges in the development of three-dimensional photonic metamaterials," *Nat. Photonics* **5**(9), 523–530 (2011).
23. P. Qiao, W. Yang, and C. J. Chang-Hasnain, "Recent advances in high-contrast metastructures, metasurfaces, and photonic crystals," *Adv. Opt. Photonics* **10**(1), 180 (2018).
24. S. Jahani and Z. Jacob, "All-dielectric metamaterials," *Nat. Nanotechnol.* **11**(1), 23–36 (2016).
25. M. Notomi, E. Kuramochi, and H. Taniyama, "Ultrahigh-Q nanocavity with 1D photonic gap," *Opt. Express* **16**(15), 11095 (2008).
26. K. Navrátil, J. Šík, J. Humlíček, and S. Nešpurek, "Optical properties of thin films of poly(methyl-phenylsilylene)," *Opt. Mater.* **12**(1), 105–113 (1999).
27. D. Hohlfeld and H. Zappe, "An all-dielectric tunable optical filter based on the thermo-optic effect," *J. Opt. A: Pure Appl. Opt.* **6**(6), 504–511 (2004).
28. A. Macleod, "Early days of optical coatings," *J. Opt. A: Pure Appl. Opt.* **1**(S), 779–783 (1999).
29. H. A. Macleod, *Thin-Film Optical Filters* (CRC Press, 2010), 4th ed.
30. W. H. Southwell, "Omnidirectional mirror design with quarter-wave dielectric stacks," *Appl. Opt.* **38**(25), 5464 (1999).
31. Y. Fink, J. N. Winn, S. Fan, C. Chen, J. Michel, J. D. Joannopoulos, and E. L. Thomas, "A dielectric omnidirectional reflector," *Science* **282**(5394), 1679–1682 (1998).
32. P. Baumeister, "Design of multilayer filters by successive approximations," *J. Opt. Soc. Am.* **48**(12), 955 (1958).
33. A. V. Tikhonravov, M. K. Trubetskoy, and G. W. DeBell, "Application of the needle optimization technique to the design of optical coatings," *Appl. Opt.* **35**(28), 5493 (1996).
34. J. A. Dobrowolski and R. A. Kemp, "Refinement of optical multilayer systems with different optimization procedures," *Appl. Opt.* **29**(19), 2876 (1990).
35. H. Mei, C. M. Landis, and R. Huang, "Concomitant wrinkling and buckle-delamination of elastic thin films on compliant substrates," *Mech. Mater.* **43**(11), 627–642 (2011).
36. J. Wang, R. L. Weaver, and N. R. Sottos, "A parametric study of laser induced thin film spallation," *Exp. Mech.* **42**(1), 74–83 (2002).
37. L. Frey, L. Masarotto, M. Armand, M.-L. Charles, and O. Lartigue, "Multispectral interference filter arrays with compensation of angular dependence or extended spectral range," *Opt. Express* **23**(9), 11799 (2015).
38. H. Chen, C. T. Chan, and P. Sheng, "Transformation optics and metamaterials," *Nat. Mater.* **9**(5), 387–396 (2010).
39. J. B. Pendry, D. Schurig, and D. R. Smith, "Controlling electromagnetic fields," *Science* **312**(5781), 1780–1782 (2006).
40. K. Aydin, V. E. Ferry, R. M. Briggs, and H. A. Atwater, "Broadband polarization-independent resonant light absorption using ultrathin plasmonic super absorbers," *Nat. Commun.* **2**(1), 517 (2011).
41. N. I. Landy, S. Sajuyigbe, J. J. Mock, D. R. Smith, and W. J. Padilla, "Perfect metamaterial absorber," *Phys. Rev. Lett.* **100**(20), 207402 (2008).
42. P. Moitra, B. A. Slovick, Z. Gang Yu, S. Krishnamurthy, and J. Valentine, "Experimental demonstration of a broadband all-dielectric metamaterial perfect reflector," *Appl. Phys. Lett.* **104**(17), 171102 (2014).
43. D. R. Smith, J. B. Pendry, and M. C. Wiltshire, "Metamaterials and negative refractive index," *Science* **305**(5685), 788–792 (2004).
44. V. M. Shalaev, "Optical negative-index metamaterials," *Nat. Photonics* **1**(1), 41–48 (2007).
45. D. Neshev and I. Aharonovich, "Optical metasurfaces: New generation building blocks for multi-functional optics," *Light: Sci. Appl.* **7**(1), 58 (2018).

46. Z. Bomzon, V. Kleiner, and E. Hasman, "Pancharatnam–Berry phase in space-variant polarization-state manipulations with subwavelength gratings," *Opt. Lett.* **26**(18), 1424 (2001).

47. N. Yu, P. Genevet, M. a. Kats, F. Aieta, J.-P. Tetienne, F. Capasso, and Z. Gaburro, "Light propagation with phase reflection and refraction," *Science* **334**(6054), 333–337 (2011).

48. B. Slovick, Z. G. Yu, M. Berding, and S. Krishnamurthy, "Perfect dielectric-metamaterial reflector," *Phys. Rev. B* **88**(16), 165116 (2013).

49. M. V. Rybin, K. B. Samusev, A. N. Poddubny, A. Hosseinzadeh, E. Semouchkina, G. Semouchkin, Y. S. Kivshar, and M. F. Limonov, "Fano resonances in all-dielectric metamaterials," in *2013 7th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics (Metamaterials 2013)*, (IEEE, 2013), pp. 226–228.

50. C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Opt. Express* **21**(18), 21693 (2013).

51. E. S. Harper, M. N. Weber, and M. S. Mills, "Machine accelerated nano-targeted inhomogeneous structures," in *2019 IEEE Research and Applications of Photonics in Defense Conference (RAPID)*, (IEEE, 2019), pp. 1–5.

52. E. S. Harper, E. J. Coyle, J. P. Vernon, and M. S. Mills, "Inverse design of broadband highly reflective metasurfaces using neural networks," *Phys. Rev. B* **101**(19), 195104 (2020).

53. S. Inampudi and H. Mosallaei, "Neural network based design of metagratings," *Appl. Phys. Lett.* **112**(24), 241102 (2018).

54. M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin, and K. Parsons, "Deep neural network inverse design of integrated photonic power splitters," *Sci. Rep.* **9**(1), 1368 (2019).

55. J. Li, H. Zhang, and J. Z. Chen, "Structural Prediction and Inverse Design by a Strongly Correlated Neural Network," *Phys. Rev. Lett.* **123**(10), 108002 (2019).

56. D. Liu, Y. Tan, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photonics* **5**(4), 1365–1369 (2018).

57. B. Shen, P. Wang, R. Polson, and R. Menon, "An integrated-nanophotonics polarization beamsplitter with 2.4 x 2.4 μm^2 footprint," *Nat. Photonics* **9**(6), 378–382 (2015).

58. B. Shen, R. Polson, and R. Menon, "Increasing the density of passive photonic-integrated circuits via nanophotonic cloaking," *Nat. Commun.* **7**(1), 13126 (2016).

59. A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vuckovic, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," *Nat. Photonics* **9**(6), 374–377 (2015).

60. J. Huang, J. Yang, D. Chen, X. He, Y. Han, J. Zhang, and Z. Zhang, "Ultra-compact broadband polarization beam splitter with strong expansibility," *Photonics Res.* **6**(6), 574–578 (2018).

61. J. Huang, J. Yang, D. Chen, W. Bai, J. Han, Z. Zhang, J. Zhang, X. He, Y. Han, and L. Liang, "Implementation of on-chip multi-channel focusing wavelength demultiplexer with regularized digital metamaterials," *Nanophotonics* **9**(1), 159–166 (2019).

62. L. F. Frellsen, Y. Ding, O. Sigmund, and L. H. Frandsen, "Topology optimized mode multiplexing in silicon-on-insulator photonic wire waveguides," *Opt. Express* **24**(15), 16866–16873 (2016).

63. Z. Yu, H. Cui, and X. Sun, "Genetically optimized on-chip wideband ultracompact reflectors and Fabry–Perot cavities," *Photonics Res.* **5**(6), B15–B19 (2017).

64. A. V. Pogrebnyakov, J. A. Bossard, J. P. Turpin, J. D. Musgraves, H. J. Shin, C. Rivero-Baleine, N. Podraza, K. A. Richardson, D. H. Werner, and T. S. Mayer, "Reconfigurable near-IR metasurface based on $\text{Ge}_2\text{Sb}_2\text{Te}_5$ phase-change material," *Opt. Mater. Express* **8**(8), 2264–2275 (2018).

65. A. Sarangan, J. Duran, V. Vasilyev, N. Limberopoulos, I. Vitebskiy, and I. Anisimov, "Broadband reflective optical limiter using GST phase change material," *IEEE Photonics J.* **10**(2), 1–9 (2018).

66. C.-Y. Hwang, S.-Y. Lee, Y.-H. Kim, T.-Y. Kim, G. H. Kim, J.-H. Yang, J.-E. Pi, J. H. Choi, K. Choi, H.-O. Kim, and C.-S. Hwang, "Switchable subwavelength plasmonic structures with phase-change materials for reflection-type active metasurfaces in the visible region," *Appl. Phys. Express* **10**(12), 122201 (2017).

67. C. R. de Galarreta, I. Sinev, A. M. Alexeev, P. Trofimov, K. Ladutenko, S. G.-C. Carrillo, E. Gemo, A. Baldycheva, J. Bertolotti, and C. D. Wright, "Reconfigurable multilevel control of hybrid all-dielectric phase-change metasurfaces," *Optica* **7**(5), 476–484 (2020).

68. K. J. Miller, R. F. Haglund, and S. M. Weiss, "Optical phase change materials in integrated silicon photonic devices: review," *Opt. Mater. Express* **8**(8), 2415–2429 (2018).

69. F. Ding, Y. Yang, and S. I. Bozhevolnyi, "Dynamic metasurfaces using phase-change chalcogenides," *Adv. Opt. Mater.* **7**(14), 1801709 (2019).

70. Z. L. Sámon, K. F. MacDonald, F. D. Angelis, B. Gholipour, K. Knight, C. C. Huang, E. D. Fabrizio, D. W. Hewak, and N. I. Zheludev, "Metamaterial electro-optic switch of nanoscale thickness," *Appl. Phys. Lett.* **96**(14), 143105 (2010).

71. B. Gholipour, J. Zhang, K. F. MacDonald, D. W. Hewak, and N. I. Zheludev, "An all-optical, non-volatile, bidirectional, phase-change meta-switch," *Adv. Mater.* **25**(22), 3050–3054 (2013).

72. C. J. Chang-Hasnain and W. Yang, "High-contrast gratings for integrated optoelectronics," *Adv. Opt. Photonics* **4**(3), 379 (2012).

73. C. J. Chang-Hasnain and W. Yang, "Integrated optics using high contrast gratings," in *Photonics: Scientific Foundations, Technology and Applications*, vol. III D. L. Andrews, ed. (Wiley, 2015), chap. 2, pp. 57–98.

74. S. G. Johnson and J. D. Joannopoulos, "Block-iterative frequency-domain methods for maxwell's equations in a planewave basis," *Opt. Express* **8**(3), 173–190 (2001).
75. V. Liu and S. Fan, "S⁴: A free electromagnetic solver for layered periodic structures," *Comput. Phys. Commun.* **183**(10), 2233–2244 (2012).
76. C. S. Adorf, P. M. Dodd, V. Ramasubramani, and S. C. Glotzer, "Simple data and workflow management with the signac framework," *Comput. Mater. Sci.* **146**, 220–229 (2018).
77. I. H. Malitson and M. J. Dodge, "Refractive index and birefringence of synthetic sapphire," *J. Opt. Soc. Am.* **62**, 1405 (1972).