Physics-informed machine learning for inverse design of optical metamaterials

Sulagna Sarkar,[1,*] Anqi Ji,[2] Zachary Jermain,[3] Robert Lipton,[3]
Mark Brongersma,[4] Kaushik Dayal,[1,5,6,7] and Hae Young Noh[1,8,†]

[1]*Department of Civil and Environmental Engineering, Carnegie Mellon University*
[2]*Department of Electrical Engineering, Stanford University*
[3]*Department of Mathematics, Louisiana State University*
[4]*Geballe Laboratory for Advanced Materials, Stanford University*
[5]*Center for Nonlinear Analysis, Department of Mathematical Sciences, Carnegie Mellon University*
[6]*Department of Mechanical Engineering, Carnegie Mellon University*
[7]*Pittsburgh Quantum Institute, University of Pittsburgh*
[8]*Department of Civil and Environmental Engineering, Stanford University*
(Dated: October 8, 2023)

Keywords: inverse design, optical metamaterials, physics-informed learning, deep learning

Optical metasurfaces consist of densely arranged unit cells that manipulate light through various light confinement and scattering processes. Due to its unique advantages, such as high performance, small form factor and easy integration with semiconductor devices, metasurfaces have been gathering increasing attention in fields such as displays, imaging, sensing and optical computation. Despite advances in fabrication and characterization, a viable design prediction for suitable optical response remains challenging for complex optical metamaterial systems. The computation cost required to obtain the optimal design exponentially grows as the design complexity increases. Furthermore, the design prediction is challenging since the inverse problem is often ill-posed. In recent years, deep learning (DL) methods have shown great promise in the area of inverse design. Inspired by this and the capability of DL to produce fast inference, we introduce a physics-informed DL framework to expedite the computation for the inverse design of metasurfaces. Addition of the physics-based constraints improve generalizability of the DL model while reducing data burden. Our approach introduces a tandem DL architecture with physics-based learning to alleviate the non-uniqueness issue by selecting designs that are scientifically consistent, with low error in design prediction and accurate reconstruction of optical responses. To prove the concept, we focus on the inverse design of a representative plasmonic device that consists of metal gratings deposited on a dielectric film on top of a metal substrate. The optical response of the device is determined by the geometrical dimensions as well as the material properties. The training and testing data are obtained through Rigorous Coupled-Wave Analysis (RCWA), while the physics-based constraint is derived from solving the electromagnetic (EM) wave equations for a simplified homogenized model. We consider the prediction of design for the optical response of a single wavelength incident or a spectrum of wavelength in the visible light range. Our model converges with an accuracy up to 97% for inverse design prediction with the optical response for the visible light spectrum as input. The model is also able to predict design with accuracy up to 96% and optical response reconstruction accuracy of 99% for optical response of a single wavelength of light as input.

---

[*] sulagnas@andrew.cmu.edu

[†] noh@stanford.edu

1. Introduction

In this study, we aim to develop a physics-informed data-driven approach to model the inverse design of optical metamaterials. Inverse design of materials have a property-to-structure objective, wherein, an algorithm provides valid design of materials for achieving target properties [1]. In other words, inverse design calculates cause from effect [2]. Optical metamaterials have various light modulation applications like, cloaking [3, 4], optical communication, imaging [5], and holography [6] owing to electromagnetic characteristics unlike those of any conventional materials and ability to manipulate optical response through interaction with light [6, 7]. Given the wide range of application of optical metamaterials, there arises a need to solve the inverse problem, that is, prediction of design parameters given the desired optical response [8, 9]. However, inverse problems are often touted as "ill-posed" problems, which is numerically difficult to solve for as there is no closed form solution . Optical metamaterial systems often encompass a wide range of possible design candidates. These systems are usually comprised of ultra-thin array of periodic sub-wavelength structures, called meta-atoms that mimic atoms in ordinary materials [10, 11]. These meta-atoms have multi-dimensional design parameters that are challenging to be determined by inverse design problem.

The inverse design of metamaterials using semi-analytical methods such as Rigorous Coupled Wave Analysis (RCWA), or numerical methods such as Finite Difference Time Domain (FDTD) and Finite Element Method (FEM), is computationally expensive due to the immense size of the design space [8, 9]. These methods often involve iterative searches and parameter sweeps through a multidimensional space, making the design process time-consuming. Additionally, when these approaches are used, the design geometry is often limited to simple design parameters, which does not capture the full potential of the optical metamaterial design. Few studies have also used evolutionary methods like genetic algorithm to facilitate metamaterial design [12]. However, these methods are less suitable for inverse design tasks as they typically require high number of evolutions and have limited design exploration capabilities [13–16]. Furthermore, these algorithms necessitate the constant execution of physics-based simulations to evaluate the objective function, adding to their computational demands and making them less efficient for designing complex metamaterial systems [17, 18].

Recent advancements in computer architectures, algorithms, computer hardware, and the availability of large datasets have enabled the use of deep learning (DL) methods to solve various mechanics problems, including the inverse design of materials [19, 20]. DL algorithms utilize multiple layers of non-linear transformations to learn complex functions from data [21]. However, despite its success in solving complex problems, DL still faces several issues, such as high data burden, lack of robustness, and difficulty in interpretation due to learning high dimensional complex functions [22]. Moreover, DL solutions often face convergence issues when multiple valid solutions exist, as it is deterministic in nature. The electromagnetic (EM) wave equations governing the relationship between design and optical response are highly non-linear, and even a relatively small error in the design prediction can result in considerable deviation in the optical response. Additionally, the one-to-many mapping from response to design results in nearly identical optical responses being produced by different design structures, which further complicates the inverse problem solution using DL methods and renders it unstable and continuously dependent on initial conditions [23].

Thus, solving the inverse problem by DL methods require adoption of either (i) a probabilistic approach instead of deterministic model, or, (ii) a regularizer to constrain the network parameters to a specific domain.

DL powered inverse design for metamaterials have adopted generative models like Generative Adversarial Network (GAN) and Variational Autoencoder (VAE) [24]. GANs were employed by Liu et al [25] and Jiang et al [26] to predict structural images for a given transmission spectra. These networks can generate novel structural patterns, which can provide insight into new structures beyond human intuition built on experience and knowledge. VAE, a semi-supervised strategy, has been utilized by Ma et al [27] and Liu et al [28]. VAE models have a decoder that reconstructs the structure geometry from compressed latent variables. The encoder learns the parameters of the distribution of the latent variables. The decoder output generates multiple candidate designs for a spectrum, from which a single design is chosen based on fabrication

requirements.

The tandem model architecture developed by Liu et al. for optical metamaterial design aims to predict structure design that leads to proper reconstruction of optical response [29]. The model concatenates an inverse DL model architecture, which predicts structural design from response data, with a pre-trained forward DL model that predicts optical response from the predicted design. The model then minimizes the difference between the input response spectrum and the corresponding spectrum produced by the predicted design parameter. This method functions like a pseudo auto-encoder model, requiring fewer parameters to train and achieving stable convergence.

Another approach uses gradients from training the forward model for inverse design. Starting with an initial design, the loss is calculated as the difference between the chosen design's reconstructed spectrum and the actual spectra. Gradients of the loss function wrt design parameters from the already trained forward model is used to guide design updates through iterations to minimize the loss. They enable exploration of large number of designs simultaneously and thus can yield multiple designs for the inverse problem. However, this technique risks converging into network singularities, as they actively search for extremas in the parameter space [30–32].

A comparison of model performances of tandem neural network with conditional GANs and conditional VAEs for inverse design of metamaterials is performed by Ma et al.[33]. In this study it is seen that the tandem model slightly outperforms the generative models in terms of target reconstruction for low degree of freedom (DOF) structure. Since tandem networks have a relatively simple architecture, they are able to capture the response-design relationship with less data requirement [17, 34] and easier hyper-parameter tuning [35] than the generative models.

However, the tandem model introduced by Liu et al. [29] lacks a constraint on the design parameter in the loss function. Consequently, the inverse model is learned by minimizing the optical reconstruction cost function, which can result in design parameters that are far from the ground truth. Not imposing constraints on the structural parameters leads to limited exploration of design options and the generation of impractical structures [36–38]. To address this limitation, researchers have used hyper-parameter driven design loss, serving as a penalty term, along with the reconstruction loss term in the model's loss function. This addition enhances network robustness and ensures that the retrieved parameters closely align with the dataset [36, 38]. Moreover, the design loss inclusion facilitates the consideration of any fabrication prerequisites when applying inverse design in practical applications. This approach introduces a small regularization loss, guiding predictions towards at least one potential design parameter within the training dataset. Although this may slightly impact convergence due to a marginally higher loss for alternate candidate design parameters, the concurrent presence of the reconstruction loss aids in maintaining algorithmic stability.

Furthermore, the tandem model is a data-driven DL model whose internal working is often difficult to comprehend. DL models have multiple hidden layers with intricate activation functions, making it challenging to fully comprehend the learned relationships. Therefore, DL models are often referred to as "black box" models. Moreover, DL models require a large amount of data and have poor generalization, meaning they may not perform well on inputs outside the range of the training set [23, 24]. A naive solution, like, simplification of the DL model or using simple linear/non-linear ML algorithms with less parameters to learn would not work in practice. Even though the simplified model is easier to explain, the approximation power of DL is lost which decreases the accuracy of the model. Another approach to tackle the issues involve increasing the design space by data augmentation strategies. However, this too requires producing new labeled data which is resource-intensive.

Recently, scientists have adopted a method to improve generalization power of DL models while reducing the large data requirement and producing scientifically consistent predictions. Domain knowledge, like, the governing physics are integrated in a DL model by a variety of methods, which include, incorporation in the loss function, residual modeling, and initialization of model parameters or architecture of the DL model [2]. They are collectively referred to as physics-informed neural network [1, 23, 39–43]. Research conducted by Lu et al [44]. and Pestourie et al [45] employ neural networks in conjunction with a low-fidelity

physics solver (i.e., simplified physics model) to alleviate data requirements and enhance computational efficiency. The incorporation of physics principles ensures the preservation of the conservation laws and symmetry requirements. In our work, we aim to leverage this multi-fidelity model, which combines a low-fidelity physics-based model with a data-driven neural network for the purpose of inverse design of optical metamaterials. This model has the capability to predict unique design parameters that align with the physics principles integrated by the simplified physics model. This integration enhances the interpretability of the predictions, offering an alternative to "black box" DL models. Moreover, the combined output of the neural network and the simplified physics model can closely match the results obtained from high-fidelity or full-feed physics simulations [45].

To integrate physics into the DL model, we utilize two approaches: (a) physics-informed loss function, which includes a constraint based on the governing physics in the model's loss function, and (b) physics-informed design of architecture, where physics-based features are embedded into the neural network design via intermediate layers. Specifically, we solve Maxwell's electromagnetic wave (EM) equations for a simplified homogenized structure to obtain the physics knowledge. We penalize the final design predictions that are not physics-consistent in the loss function, or guide the design towards physics-consistency through the architecture. This physics knowledge acts as a regularizer during the training of the DL model which reduces the search space of the model parameters. Hence, we predict design parameters that are explainable without decrease in prediction accuracy with less labelled data. Furthermore, the DL models have more generalization power for out-of-sample scenarios. Thus, the addition of physics in the DL model makes them models more lucrative to domain scientists. Our study demonstrates that the physics-informed DL models outperform the purely data-based DL method in terms of design prediction and reconstruction accuracy.

The rest of the paper is organized as follows. Section 2 discusses about the representative structure of an optical metamaterial and the underlying mechanics that this structure follows. Section 3 describes the specific deep learning components used for inverse design as well as the integration of physics in the deep learning model. Section 4 evaluates and characterizes the model performances with concluding remarks discussed in Section 5.

2. Physics-based model

In this section, the geometry of a unit optical metamaterial cell, the governing physics of optical metamaterials, and the simplification of the structure by homogenization are discussed.

2.A. Description of optical metamaterial structure

Metasurfaces have various types of structure including spherical shell meta atoms [46, 47], stratified medium of metal, and dielectric [48, 49], and square or circular split ring resonator. The most commonly studied representative structure of optical metamaterials is the thin polymer or dielectric film over a thick substrate and a periodic grated metal bar stacked on top of the polymer thin film [50–52]. These structures effectively modulate the incident light, such that, the outgoing light waves have the desired amplitude and phase [50]. The gratings can have a variety of designs ranging from simple structures like cylindrical or rectangular to complicated structures like gyroid inspired by scales of butterflies, bow, H or cross [51, 53, 54].

In this study, rectangular gratings are analyzed that are periodic in the x_1-direction and homogeneous in the x_2-direction. These gratings are stacked on an insulating polymer film on a substrate, as depicted in Fig. 1a, with the stacking taking place along the x_3-direction. The incoming light propagates along the x_3-direction with normal incidence. The design parameters of this metamaterial system consist of the width of the grating, w, the period of the grating, p, the thickness of the grating, t_1, and the thickness of the polymer film, t_2. The substrate and grating material are made of metal, specifically gold (Au) due to its high absorption coefficient [55], and the dielectric film consists of an insulating substance, such as liquid crystal elastomer (LCE), with tunable properties. Such structures are easy to manufacture while being able to produce a wide range of optical responses [51]. This plasmonic device is representative of a structure that manipulates light through various light confinement and scattering processes [10, 11]. The parameters that

(a) Metal-polymer-metal optical metamaterial structure

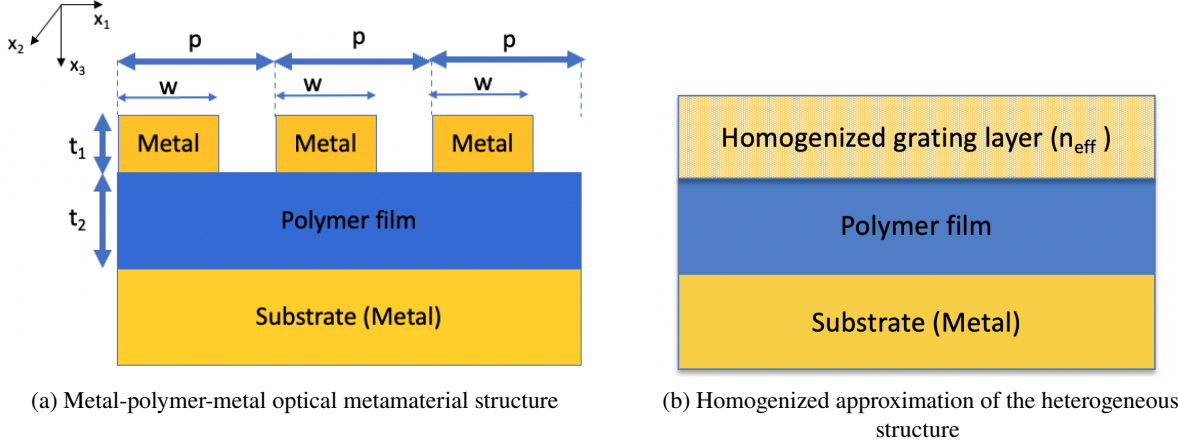(b) Homogenized approximation of the heterogeneous structure

Figure 1. Approximation of the heterogeneous layer in the metal-polymer-metal structure, (a), with a layer having "average" property, (b), that simplifies computation.

vary during optical metamaterial response modeling include the parameters of the incident light, such as the wavelength, angle of incidence, and polarization, as well as the design parameters, material properties (e.g., refractive index), and optical response.

2.B.   Governing physics

Optical metamaterials follow Maxwell's EM equations, Eq. 2.1 and 2.2, where $\boldsymbol{E}, \boldsymbol{H}, \omega, \mu_0, \epsilon_0$ and ϵ are electric field, magnetic field, angular frequency of the wave, vacuum permeability, vacuum permittivity, and relative permittivity respectively [56]. These equations are solved for a particular design to obtain response in the form of electric field and magnetic field with an incident EM wave of λ_0 wavelength.

$$\nabla \times \boldsymbol{E} = i\omega\mu_0\boldsymbol{H} \tag{2.1}$$
$$\nabla \times \boldsymbol{H} = -i\omega\epsilon_0\epsilon\boldsymbol{E} \tag{2.2}$$

The optical response is computed as the ratio of the resultant field intensity to the incident field intensity, which determines the reflection, transmission, or absorption of the incident wave. To solve the inverse design problem using a physics-informed machine learning approach, the machine learning algorithm is guided by the consistency of the response obtained by solving electromagnetic wave equations for the predicted design. However, solving these equations for a complex structure with high dielectric contrast, such as the one shown in Fig. 1a, is computationally expensive and requires numerical or semi-analytical methods.

To enable inexpensive computation and incorporate physics as a guide for the machine learning model, we use a simplified structure that homogenizes the top layer. This approach reduces the design parameter space and makes it easier and faster to calculate the response by solving electromagnetic equations. The homogenization results in an effective material property that approximates or "averages" the property of the original metal gratings in the top layer, in accordance with effective medium theory (discussed in Section 2.C). This simplification allows the response to be calculated using analytical solutions and seamlessly incorporated as guiding physics for the DL algorithm.

2.C.   Simplified physics model through homogenization

Effective medium theory simplifies the design of complicated structures, which provides computational efficiency and straightforward calculation of the forward equation. When the conditions for the effective medium theory holds, the resultant simplified structure has an equivalent response as the original complex structure. The effective medium theory in conjunction with analytical calculation of the optical response

has been used in many studies of photonics [57–61], elasticity [62], and acoustics [63]. We incorporate the physics by considering a stratified medium, wherein, the top layer, containing gratings, is homogenized in accordance to effective medium theory. We utilize this simplified model to introduce a scientific consistency penalty in the DL algorithm, which reduces the search space to design prediction that are consistent with the governing physics. A homogeneous layer is considered with effective refractive index n_{eff} that is intermediate between n_{au} and n_{air}.

The effective index is dependent on the polarization of the incident light. We consider normal incidence, with the incident light travelling in the x_3 direction with transverse electric (TE) mode of wave propagation where ($\boldsymbol{E} \perp \boldsymbol{K}$) where $\boldsymbol{K}$ is the incident light's wave vector and $\boldsymbol{E}$ is the electric field of the EM wave. The effective refractive index takes into consideration the fill factor (f) which is the volume-fraction of metal present in the top layer. In TE mode, considering $\lambda_0 \gg p$, and $\boldsymbol{E}$ as approximately continuous across the boundary, from discontinuity of $\boldsymbol{H}$ across the boundary [60, 61], we get, the effective refractive index in Eq. 2.3.

$$n_{eff}^{TE} = \left(\epsilon_{eff}^{TE}\right)^{\frac{1}{2}} = \left(n_{au}^2 f + n_{air}^2 (1 - f)\right)^{\frac{1}{2}} \tag{2.3}$$

When the dimensions of the structure are comparable to the wavelength of the incident light, diffraction effects become significant and impact the optical properties. However, in this regime, the above homogenization theorem cannot be used to calculate the effective optical response [60]. Rytov et al. [64] derived transcendental wave equations for the TE mode of propagation of plane waves in an infinite periodic layered medium, given by Eq.2.4. The solution of this equation involves higher-order refractive indices. For $\lambda_0 \gg p$, the tangent term in Eq.2.4 can be truncated, and the homogeneous effective index in Eq.2.3 can be recovered. However, for design parameters that are mostly sub-wavelength, i.e., $\lambda_0 > p$ and not $\lambda_0 \gg p$, the second-order solution given by Rytov et al. [64], which truncates the $\tan x$ series at the cubic term, is more appropriate, as shown in Eq.2.5.

$$\left(n_{air}^2 - n_{TE}^2\right)^{\frac{1}{2}} \tan\left(\pi \frac{p}{\lambda_0}(1 - f)(n_{air}^2 - n_{TE}^2)^{\frac{1}{2}}\right) = -(n_{au}^2 - n_{TE}^2)^{\frac{1}{2}} \tan\left(\pi \frac{p}{\lambda_0} f(n_{au}^2 - n_{TE}^2)^{\frac{1}{2}}\right) \tag{2.4}$$

$$n_{TE}^{(2)} = \left(n_{TE}^{(1)2} + \frac{1}{3}\left(\pi \frac{p}{\lambda_0} f(1 - f)\right)^2 (n_{au}^2 - n_{air}^2)^2\right)^{\frac{1}{2}} \tag{2.5}$$

After homogenizing the top grating layer of our representative structure in Fig.1a, the metamaterial structure gets converted to a stratified medium with effective properties in the top layer. This homogenization is depicted in Fig. 1b.

Transfer Matrix Method (TMM) is used to solve EM equations in a multi-layer system subject to a uniform incident field [65, 66]. The field in the medium is divided into two components, the forward (transmitted) component and the backward (reflected) component. The amplitudes of the field across an interface (say from material A to B) are related by the Fresnel transmission (t_{AB}) and reflection coefficient (r_{AB}). The phase shift across the medium (say, B) is controlled by a factor composed of the wave number (k), refractive index of each medium (n_B), and the thickness of the layers (t_B). The system transfer matrix is defined by combining these factors for each interfaces and mediums, which determines the amplitude of the field in each layer. The reflection/ absorption/ transmission coefficient is computed from the elements of the transfer matrix.

Our physics-informed machine learning algorithm examines the consistency between the predicted design's reflection coefficient calculated from the aforementioned simplified method, and the reflection

coefficient corresponding to the true physics, which is the input to the DL model. This would guide the algorithm and predict design parameters which is consistent with the governing physics.

3. Physics-informed DL for inverse design

In this section, we discuss the physics-informed DL framework that we have developed to solve the inverse design problem. The DL framework consists of a tandem architecture that has an inverse model in conjunction with a pre-trained forward model [29]. Training this network to predict design parameters ensures that the predicted design reconstructs the optical response. Furthermore, we introduce the simplified physics which is used as a constraint to guide the DL model to produce scientifically-consistent results. This simplified physics is applied to an effective structure which is obtained by homogenizing the complicated grating layer of the predicted metamaterial design. The physics equations are then solved to obtain the optical response, which is compared to the initial input optical response to guide the DL model. The simplified physics is incorporated as a part of the loss function of the DL model or as one of the layers in the DL model architecture. The overall structure of the method schematic is depicted in Fig. 2. We use the physics-informed deep learning approach to facilitate inverse design for a 1-D periodic grating metamaterial structure, as illustrated in Fig.1a. Nevertheless, it is interesting to note that this methodology can also be extended to 2-D periodic grating, multi-layer optical metamaterial structures.
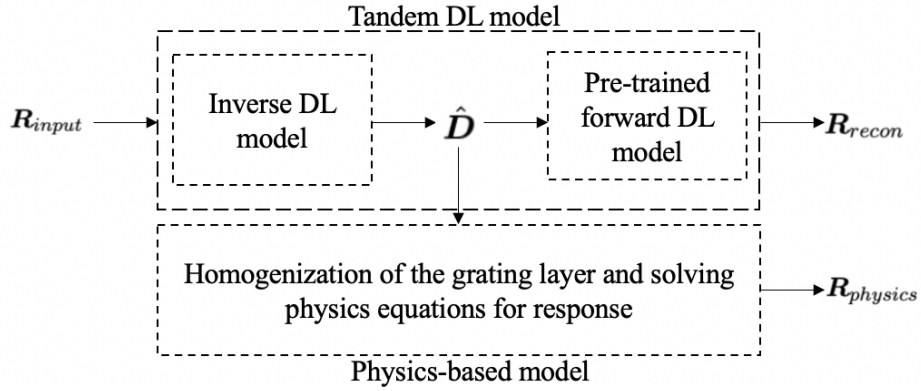

Figure 2. Overview of the components of the physics-informed DL model. R_{input} is the input optical response and the design, $\hat{D}$, predicted from the inverse model. The predicted design, $\hat{D}$, is then fed into a pre-trained forward model to ensure proper reconstruction of the optical response, R_{recon}. The structure of $\hat{D}$ is then simplified according to the homogenization principle and the physics equations are solved to obtain $R_{physics}$ that ensures physics-informed learning.

3.A. Components of the physics-informed DL model's architecture

In this subsection, we introduce the architecture of the physics-informed DL model. We formulate the inverse problem, introduce the reconstruction constraints and the physics information to enable proper physics-consistent prediction and optical reconstruction of the predicted design.

For our study, we consider the inverse of the EM equations as the target function. Given input and output pair of response and design, $\{R_i, D_i\}_i^n$, we train a DL model to learn the relationship between R_i and D_i. For the forward EM equation, $R_i = f(D_i)$, response, R_i is calculated from design, D_i. The DL model is trained to approximate the inverse function, f^{-1} using training data that follows, $D_i = f^{-1}(R_i)$, without an iterative search procedure.

As discussed in Sec.1, due to the ill-posed nature of inverse problems and high non-linearity of the EM equations, a plain inverse model cannot ensure prediction of design parameters with accurate reconstruction of optical response. Hence, we develop a tandem architecture wherein a DL model approximating the

forward EM equation is appended with the inverse DL model, such that, the output of the inverse model, $\hat{\boldsymbol{D}}$, is fed into the forward DL model.

The forward DL model approximates the function that calculates $\boldsymbol{R}$ from $\boldsymbol{D}$. The loss of the forward model is defined as the difference between predicted response from the forward DL model, $\boldsymbol{R}_{recon}$, and response produced by true design, $\boldsymbol{R}$. The forward model is trained such that the difference between $\boldsymbol{R}$ and $\boldsymbol{R}_{recon}$ is minimised.

The tandem model architecture consist of the inverse DL model in conjunction with the pre-trained forward DL model. After training the forward model, the weights of the model are frozen, i.e., they are considered as non-trainable parameters during the training of the tandem model. The inverse model predicts the design parameters and this output is fed into the pre-trained forward DL model, as depicted in Fig. 2, to ensure proper reconstruction of optical response. The entire structure is trained to minimize the design prediction error (deviation of predicted design from the inverse model, $\hat{\boldsymbol{D}}$, from the ground truth design, $\boldsymbol{D}$) and the reconstruction error (deviation of the response reconstructed by the predicted design, $\boldsymbol{R}_{recon}$, from the input response, $\boldsymbol{R}_{input}$).

We introduce the physics-based constraint to guide the tandem DL architecture for better design prediction. The physics-based constraint is calculated by solving for the optical response from the physics equations introduced in Sec.2 from the predicted design parameters. The optical response calculated is then compared with the true optical response. This model is named physics-informed loss (PIL). However, the simplified physics model is an approximation of the true model and the effective homogenization principles are only followed by design parameters that have $\lambda_0/p > 2$. Thus, we cannot incorporate the physics-based constraints for every design parameter in our domain.

To tackle this issue and to ensure that the physics knowledge for all observations is utilized, we assign the design parameters that follow the simplified physics model to an intermediate layer in the NN model. The intermediate layer outputs the design parameters that produce response as per the simplified physics-based model. The intermediate layer predicts the thickness parameters, (t_1, t_2), and fill factor (f), from which the effective refractive index is calculated according to Eq. 2.3. The optical response from the simplified structure is calculated from the physics equations and compared to the true response for the physics-based constraint term. It's crucial to emphasize that the design parameters are not expressly bound; rather, they are configured to yield the authentic optical response following the 1-D TMM equations. As previously highlighted, the true response fluctuates with p, but varying p can yield an identical f, and the identical f through the simplified physics response would yield the same response for different p which would deviate greatly from the true response. Consequently, the predicted f' by the DL model is tailored to generate the true response through simplified physics calculations. This adjusted f' diverges from the genuine f since it generates distinct responses for varying p values. After the intermediate layer, the NN consists of a few layers that learn the true design parameters, i.e., they produce response as per the true physics model. This portion of the NN learns the change in design parameter due to introduction of the grating in the top layer. The final output predicts the width (w) and period (p) of gold bars, thickness of gold bar $(t1)$ and film $(t2)$. Hence, in this method the physics is incorporated in the DL architecture and can guide the DL by using simplified physics yet not constraint the final design as per requirements of the simplified model. This modification enables the use of physics knowledge for all the design parameters. The DL architecture of the PIA model is depicted in Fig. 3. This model is called physics-informed architecture (PIA). The initial portion of PIA learns the inverse function consistent with the physics of the homogenized model and guides the later part of the model to learn the effect of introducing rectangular metal blocks with air gaps instead of homogeneous metal block.

3.B.   Training of the physics-informed DL model

Having discussed the architecture of the physics-informed model, in this section we introduce the training of the DL model by describing the following: (a) objective function that consist of data-based and physics-based loss (b) training the DL model by a non-convex optimization technique and backpropagation.
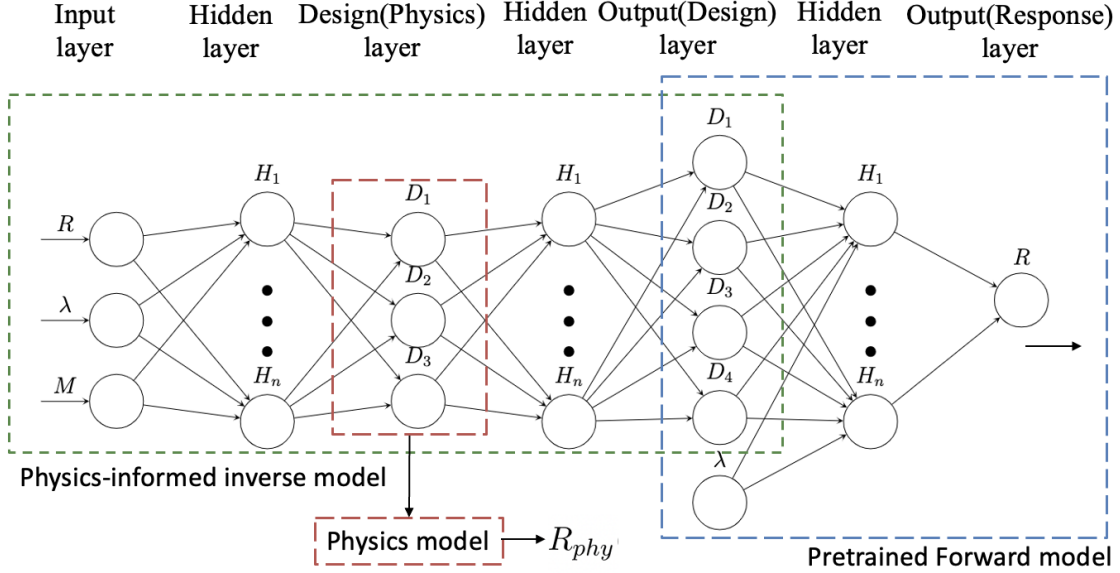
Figure 3. **Physics-informed DL architecture (PIA).**The "Design (Physics) layer" includes intermediate variables that are consistent with the underlying physics. These variables are calculated based on the design parameters that result in a certain response, as predicted by the EM equations for the simplified homogenized structure. The "Output (Design) layer" then uses these intermediate variables to predict the original complex design parameters.

3.B.1. Loss function of DL model

To train our physics-informed DL model, we designed an objective function, Eq. 3.1, that is to be minimized to learn the optimal model parameters. The objective function of the inverse model minimizes the loss associated with the design parameter, which is the first term in Eq. 3.1. The tandem model architecture adds a reconstruction loss which is the second term in Eq. 3.1. This term calculates the deviation of the optical response produced by the predicted design, $\hat{D}$, from the desired response, i.e., difference between the reconstructed response which is the output of the pre-trained forward model, $\boldsymbol{R}_{recon}$, and the input optical response, $\boldsymbol{R}_{input}$.

The physics is incorporated in the third term in Eq. 3.1 as a penalty for final output layer or intermediate layer for PIL and PIA respectively. The penalty is the difference between input optical response,$\boldsymbol{R}_{input}$, and the response calculated by solving EM equations for a simplified design structure, $\boldsymbol{R}_{physics}$. The simplified design corresponds to homogenizing the top grating layers of the representative structure. For PIA, the final layer design output is only used for the data-based loss function term and the intermediate layer design output is used for the physics-based loss function term. Whereas, for PIL, the final layer design output is used for both the data-based and physics-based loss function terms.

$$Loss = \mathrm{w}_{data_1}\mathrm{MAE}(\boldsymbol{D}, \hat{\boldsymbol{D}}) + \mathrm{w}_{data_2}\mathrm{MAE}(\boldsymbol{R}_{input}, \boldsymbol{R}_{recon}) + \mathrm{w}_{phy}(\boldsymbol{D})\mathrm{MAE}(\boldsymbol{R}_{input}, \boldsymbol{R}_{physics}) \quad (3.1)$$

The weightage of each loss term is a hyper-parameter. w_{data_1} and w_{data_2} is chosen based on grid search hyper-parameter tuning that gives best model performance on the validation set. w_{phy} is proportional to λ_0/p which ensures that more weightage is given to the physics-based loss function term when $\lambda_0 \gg p$ and the effective medium theory of the homogenization holds true.

3.B.2. Training the DL model

DL models typically consist of multiple layers of neurons with each layer consisting of hundreds of neuron. The number of layers and neurons depends on the type and the complexity of problem that is to be solved.

Since we are dealing with a regression problem with continuous input, we chose <10 layers of neurons and each layer has a few hundred neurons [21, 67].

Each neuron has a non-linear transformation of the input variables, or, an activation function, to approximate the relationship between input and output variables. The most common activation function used for this problem is Rectified Linear Unit (RelU) [21, 68], Sigmoid and Hyperbolic Tangent (Tanh) [13, 26]. However, RelU activation function, $max(0, z_k)$, often encounters the "dying RelU" problem where the neuron associated only outputs zero [69]. This can be solved by using variants of RelU, eg., smooth and continuous function- Sigmoid-Weighted Linear Units (SilU) [70] which gets rid of the point of inflection and has non-zero slope segments. SilU activation function has the functional form of sigmoid function multiplied by its input $z_k\sigma(z_k)$ [70]. We use SilU as the activation function for most layers as it has an advantage of non-saturation of gradient while being smooth and continuous at all points. We also use Tanh as the activation function for some layers as they preserve negative inputs and have strong gradients which lead to big learning steps and faster convergence.

The model parameters of the physics-informed DL model is determined by minimization of the objective function defined in Sec. 3.B.1. The objective function is optimized by backpropagation with "AdamW" optimizer which uses the adaptive momentum technique along with weight decay. This optimizer is chosen because momentum helps in faster convergence and weight decay provides additional regularization to prevent overfitting. The physics principles are leveraged as additional regularization terms that steers the learning of the model parameters such that the model predictions are consistent with the governing physics.

4. Evaluation

In this section, we discuss the evaluation of our developed algorithm and the description of the dataset used.

4.A. Data description

4.A.1. Variables of the physics-informed DL model

In order to model the complicated relationship between the design parameters and the EM response, there are a number of variables that must be considered. They can be grouped into the following categories:
- Parameters of incident light: wavelength of the incident light
- Design parameters: width (w), period (p), thickness (t_1) of metal grating, thickness (t_2) of polymer
- Optical response: reflection and absorption of the incident light

Since we used an absorbing material like gold as the substrate and the metal grating, the metamaterial structure only absorbs and reflects the incident light. Since the transmission of light is not considered, the complex reflection coefficient and absorption of light are the input optical response to the DL model. Absorption of light is the intensity attenuation as light passes through the material. The reflection coefficient has real and imaginary components. The complex reflection presents the existence of phase shift between incident and reflected EM waves. Due to the conservation of energy, the amplitude of reflection coefficient and absorption coefficient follow Eq. 4.1.

$$A + R = 1 \tag{4.1}$$

Wavelength of light is considered in the visible range. For this study, we have developed our model for different input cases - *fixed single wavelength*, *variable single wavelength*, and *multiple wavelengths*. For *fixed single wavelength* model, the optical response- reflection and absorption, for a fixed wavelength value is considered for inverse design prediction. For *variable single wavelength* model, the optical response- reflection and absorption, particular wavelength, and material information is provided as input to the inverse DL model. For *multiple wavelengths* model, the reflection and absorption for 40 equidistant wavelength points in the wavelength range 500-700nm are input to the inverse DL model. The standardised optical response for a particular set of design parameters is given in Fig. 4. The design parameters span the

dimensions within the range of 50 - 500 nm for period of the grating, 30-200 nm for width of grating, 10-60 nm for thickness of grating, and 10- 100 nm for thickness of polymer film.
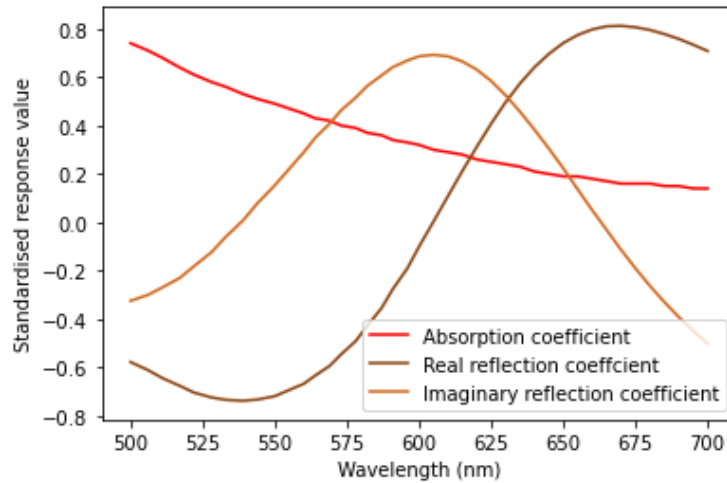

Figure 4. Optical response spectra for wavelength range 500-700 nm for design parameters: p= 280 nm, t_1, t_2= 10nm.

4.A.2. Dataset preparation

The dataset is prepared by a semi-analytical approach that solve the EM equations called Rigorous Coupled-Wave Analysis (RCWA). RCWA offers a computationally efficient and numerically stable method to provide exact solutions to Maxwell's equations for multi-layer periodic structures [37, 71, 72]. In this algorithm, infinite periodic structures are calculated with Fourier harmonic basis. For the purpose of this study 10 Fourrier harmonics orders are considered. 2D RCWA is performed that consider stacking along x_3-direction and periodic grating along x_1-direction. The inputs to the RCWA model is design dimensions, material parameters and parameters of incident light. The model solves the forward model and provides the reflection and absorption coefficient.

In the *variable single wavelength* model, we recorded a total of 60,000 observations for TE polarization. This dataset included the optical response measured at 40 evenly spaced wavelength points ranging from 500nm to 700nm. Each wavelength point was evaluated for 1500 unique design parameters. We used these 1500 distinct design parameters along with their corresponding optical response spectra measured at 40 different wavelength points as the dataset for the *multiple wavelength* model. For the *fixed single wavelength* model, we collected 40,000 observations for TE polarization at a fixed incident wavelength of 450nm. The dataset was divided into a 75%-25% train-test split, and within the training data, a further 75%-25% train-validation split was applied. This resulted in a total of 22,500 observations used for training, 10,000 for testing, and 7,500 for validation in the *fixed single wavelength* model. In contrast, for the *variable single wavelength* model, we utilized a larger dataset, which consisted of 33,750 observations for training, 15,000 observations for testing, and 11,250 observations for validation.

4.B. Model description

We implemented our physics-informed DL model using Pytorch. As discussed in Sec.4.A.1, the dimensions of the design parameters have different ranges. Hence, to avoid biases we normalise the data to min-max scale that ensures data is between 0 and 1. To tackle non-linearity we used a mix of non-linear activation function like SilU [70] and Tanh. The output layer of our model has sigmoid activation. This is to ensure that the output ranges between 0 and 1 in accordance with the range of the variables considered in the model. All objective functions are computed using Mean Absolute Error (MAE) that calculates the L_1 deviation. We
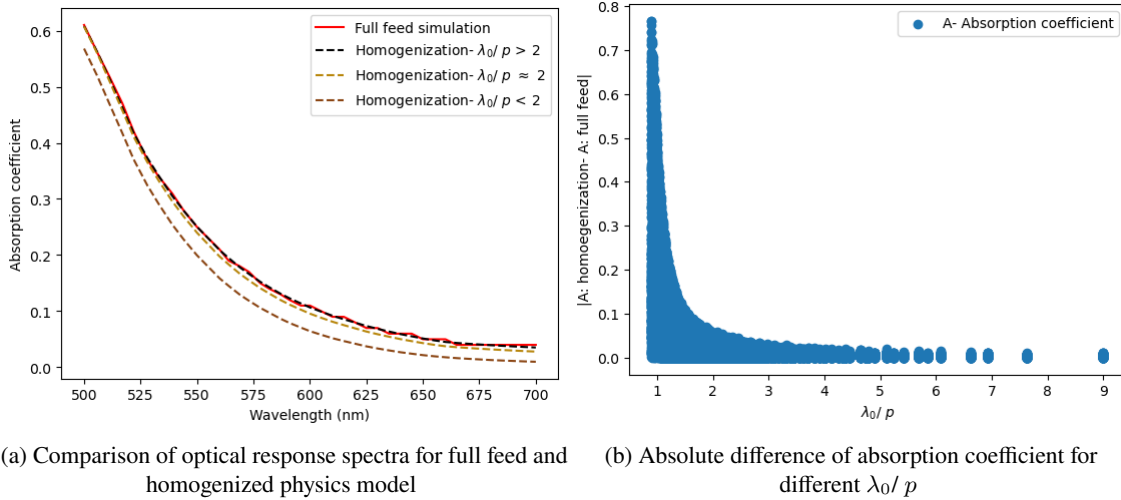
(a) Comparison of optical response spectra for full feed and homogenized physics model

(b) Absolute difference of absorption coefficient for different λ_0/p

Figure 5. (a) Comparison of the absorption spectrum from full feed simulation and homogenized model for different λ_0/p. The difference between homogenized model and full feed simulation increases with decreasing λ_0/p. (b) The absolute difference of absorption coefficient from full feed simulation by RCWA and our physics-based model that homogenizes the metal grating in the top layer and calculates the absorption coefficient by TMM.

use MAE loss to avoid heavy penalty of outliers. The simplified physics model is introduced as a regularizer to train the DL model. The efficacy of the simplified physics for the metamaterial structure in Fig. 1a is evaluated in Fig. 5. We observed that our dataset validates the expected behaviour of the homogenized model on comparison with the full feed simulation. The spectra graphs of Fig. 5 denote that as λ_0/p decreases, the absorption of the incident wavelength spectra for the homogenized model increasingly diverges from the full feed simulation. Fig. 5b shows that the absolute difference in the computed absorption coefficient from homogenized model and from full feed model (y-axis) decreases with the increase of λ_0/p (x-axis). Therefore, our dataset corroborates the limitation of using effective medium theory based physics model, which restricts the homogenization of a periodic optically responsive structure to large values of λ_0/p. Thus, it is necessary to introduce a weight (w_{phy}) for the physics-based model, which is proportional to λ_0/p and is set to 0 when $\lambda_0/p < 2$ in the PIL model.

4.C. Results and discussion

We evaluate our physics-informed DL model for both *fixed single wavelength*, *variable single wavelength*, and *multiple wavelengths* of incident light. For the *fixed single wavelength* model, the entire model is trained on a fixed wavelength of incident light. For *variable single wavelength*, the model predicts design parameters given optical response and wavelength information for a single wavelength of incident light. For *multiple wavelength*, the model predicts design parameters given the optical response for a spectrum of wavelength of incident light in the visible range.

4.C.1. Overall performance of the tandem physics-informed DL model

We evaluate our physics-informed DL models to solve the inverse problem for the dataset described in Sec. 4.A.2. For all three aforementioned input cases, the model is able to perform inverse design with proper reconstruction of optical response. The inverse design model for *fixed single wavelength* trained with $\lambda = 450nm$, converges with $\approx 7\%$ mean average error (MAE). For *variable single wavelength* and *multiple wavelength*, our model converges with $\approx 4\%$ MAE and 2.6% MAE respectively for inverse design. Furthermore, our model is able to achieve $< 1\%$ reconstruction error for single wavelength model and $\approx 3\%$ reconstruction error for multiple wavelength model. The comparison of performance for inverse design and

reconstruction by our physics-informed architecture model (PIA) with tandem neural network(tandem-NN) and physics-informed loss (PIL) model is depicted in Fig. 6.

We validate and quantify the expected behavior from various studies [29, 73] of the simple feed-forward neural network (NN) and tandem-NN model using our dataset. We observe that for *single wavelength response* input cases, Fig.6a,6b, the reconstruction error significantly decreases on using tandem-NN model. As the design error is also a component of the objective function (Eq. 3.1), the tandem-NN model improves the reconstruction while maintaining the accuracy of the inverse design.

In spite of the improvement in performance in tandem-PIL, physics knowledge is not utilised during training of all design parameters to prevent erroneous prediction as the simplified physics deviates substantially from the true governing physics. Thus, to make better utilization of the simplified physics model, we develop the PIA model. In this model, we embed the simplified physics-obeying design parameters into an intermediate layer of neurons in the DL model thus enabling us to model the whole design space. PIA does not constraint the final design output to follow the simplified physics, rather the final design output only contributes to the data-based loss term in the objective function in Eq. 3.1. Instead, PIA constraints output of an intermediate layer to follow the simplified physics as discussed in Sec.3.A. As shown in the reconstruction error figures, Fig.6a, 6b, and 6c, tandem-PIA provides better reconstruction than tandem-PIL as the physics-based constrained variables in the intermediate layer guides the DL model towards capturing the governing physics for the whole design space without causing any significant change in the design accuracy for the inverse model as noted in the design error figures, Fig.6d, 6e, and 6f.

4.C.2.   Performance on *fixed single wavelength* input

For *fixed single wavelength*, we see that PIA performs best in terms of reconstruction of optical response and design prediction error. From Fig. 6a, we see that tandem-NN outperforms vanilla neural network (NN) by achieving a 3/5th reduction in response reconstruction error. This improvement is expected as a simple feed forward neural network does not perform well for inverse design problems. The reconstruction error is improved by 1/4th from the tandem-NN model by using the PIA model. The inverse design error is seen to reduce by addition of physics-based constraint with comparable error values for PIL and PIA as seen in Fig. 6d.

Fig. 6a and 6d depicts the reconstruction error and design prediction error for the DL models for light with wavelength of 450nm incident on the structure. Inverse design is more challenging for shorter wavelength due to emergence of multiple diffraction orders that increases complexity of the underlying physics. Consequently, we present the result of our model being evaluated on 450nm incident light for all *fixed single wavelength* evaluations. However, it is to be noted that the same model architecture is used for inverse design for higher wavelengths within the visible light region as well.

4.C.3.   Performance on *variable single wavelength* input

PIA performs design prediction with lowest reconstruction error for *variable single wavelength* input. Fig. 6b and 6e depicts the optical response reconstruction error and design prediction error for the DL models for a single wavelength of light incident with wavelengths between 500-700nm. Fig. 6b shows that tandem-NN outperforms vanilla neural network (NN) by achieving a 1/4th reduction in response reconstruction error, as anticipated given the challenges faced by basic feed-forward neural networks for inverse design. Reconstruction error improves by 1/10th from the tandem-NN model by using the PIA model. Inverse design error reduces by a 1/10th in value due to addition of physics-based constraint with comparable error values for PIL and PIA as seen in Fig. 6d.

4.C.4.   Performance on *multiple wavelength* input

The predictive power of the physics-based inverse design and response reconstruction is analysed for *multiple wavelength* input, i.e., wavelength spectrum response as input. From Fig. 6f, we see that physics-informed DL improves on design prediction accuracy over data-based DL, reducing the prediction error

by 1/5th. However, we also note that the introduction of tandem-NN does not cause any improvement in reconstruction of optical response unlike the other single wavelength input cases. The result is so because, it is not likely for different designs to produce identical response for all optical parameters for each wavelength value in the spectrum. Since there are multiple optical response parameters (complex reflection and absorption coefficient) for each of the 40 wavelength points, the dataset we considered do not pose a non-unique inverse design problem for the response of a wavelength spectrum scenario. Hence, the introduction of the tandem-NN model does not increase reconstruction accuracy for the *multiple wavelength* case as seen in Fig. 6c.

By comparing Fig. 6f and 6e, it is noted that the inverse design loss for *multiple wavelength* input is lower than the inverse model with single wavelength information as input. This is due to the higher level of information fed as input to the *multiple wavelength* model, as optical response information for 40 wavelength points is the input as opposed to optical response of one wavelength point in the single wavelength cases.

We also validated the performance of the PIA model in reconstructing the optical response. As seen in Fig. 7e, the optical response generated by the design predicted by our physics-informed DL model approximates the optical response calculated from full feed RCWA simulation for the true design very closely. The misfit between the reconstructed response spectrum and the input is the slight increase in variance of the reconstructed spectrum. The increased variance can be characterized due to the use of two DL models (PIA for inverse design and forward DL for prediction of response), as it is characteristic of DL model to overfit/ have high variance results. However, this difference is small ($\approx 3\%$) and the trend of the response is captured by the PIA model.

4.C.5. Robustness of the physics-informed architecture model

We analyze the ability of the PIA model to perform design prediction for low data and out-of-training sample cases.

In Fig. 7a, we observe that our PIA model is able to predict design parameters with very less training examples. The reported evaluations in Fig. 6 are based on DL models trained with > 20,000 data points. It is well known that DL models perform better function approximation and generalization with increased amount of training data. However, since collecting data involves expensive forward computation, we add external knowledge to limit training data requirement without loss of model performance. The introduction of the physics provides a source of knowledge of the behaviour of the function we aim to approximate. The addition of physics allows for better generalizability with low data dependency. Fig. 7a shows that for the model trained on *fixed single wavelength* of 450nm, the physics-informed DL model (PIA) performs better consistently with lower prediction error than data based model for less training examples. The difference of error grows with decreasing number of training samples. Therefore, addition of physics for inverse design DL model decreases the data burden and improves test accuracy.

In addition, we evaluated the model's ability to predict test examples beyond the range of the training data using the *fixed single wavelength* model trained on 450nm incident light. The generalizability of the model was tested on design parameters, grating period (p), polymer thickness (t_2), and input parameter, incident light absorption (A), which were outside the range of values used to train the DL model. Fig. 7b, 7c, and 7d display the model's performance for lower p and t_2 values and higher A values than those in the training set. The results show that, in all cases, PIA has a lower design prediction error than the data-based model.

4.D. Model characterization

We analysed the importance of the variables, as shown in Fig.8a, in the forward problem. This could provide a sense of weighing the error for the prediction of each design parameter in the inverse model based on the variable importance. The importance is computed from the response prediction error where each design variable does not contribute in the network to calculate the response. We analyse the importance of design, light and material parameters. For forward model, the average error is $\approx 0.4\%$. We observe that the presence
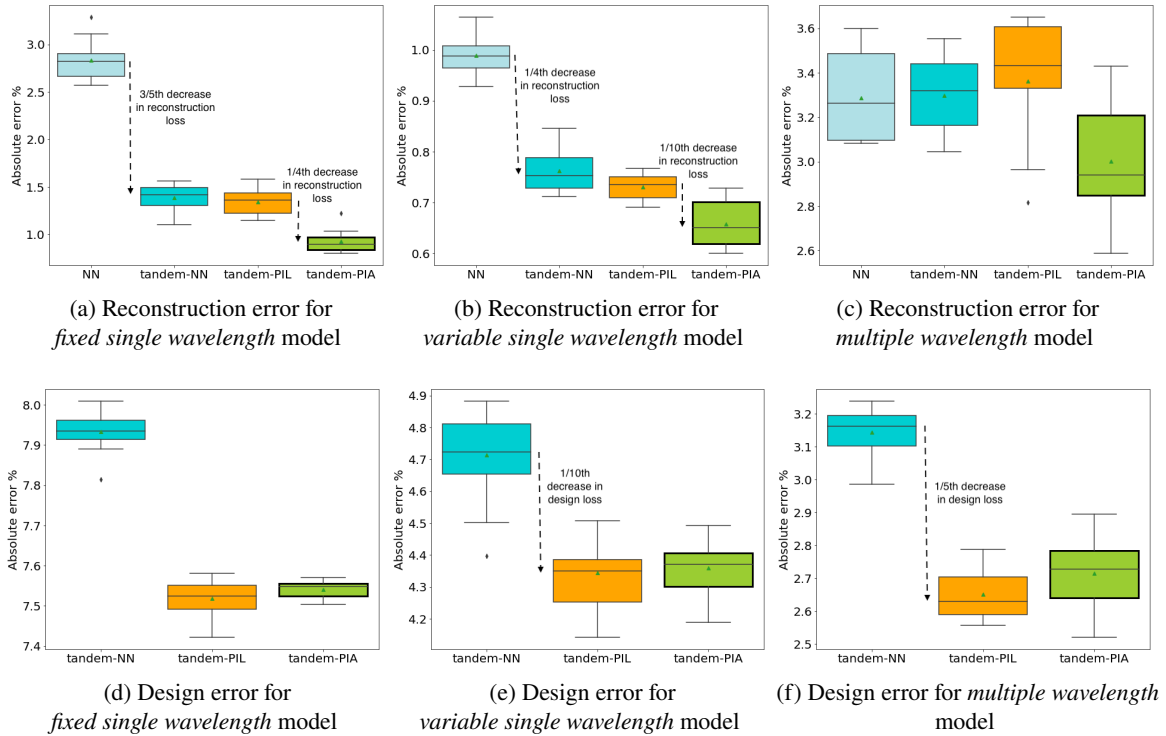
Figure 6. Design and reconstruction accuracy- (a),(d) Model trained for incident light of *fixed single wavelength*. (b),(e) Model trained for incident light of *variable single wavelength*. (c),(f) Model trained for incident light of *multiple wavelengths*. (a),(b),(c) Reconstruction accuracy improves on using tandem NN model instead of inverse NN model. Addition of physics-based constraint in loss (PIL) does not change the reconstruction error but improves further on using physics-informed architecture of DL (PIA) instead of PIL. (d),(e),(f) Design accuracy improves on addition of physics-informed loss. Design accuracy remains approximately same for PIA and PIL with increase of reconstruction accuracy.

of either refractive index information or the wavelength of the incident light is sufficient for the forward model to perform as good as the full model. However, if both the incident light as well as material parameter information is not provided, the model cannot predict the response properly, recording an average loss of $\approx 5.36\%$. Similarly, it is seen that the design parameters are paramount to calculating response, with the most important parameter being thickness of the film (MAE $\approx 13\%$) followed by period of grating (MAE $\approx 6.5\%$), thickness of metal grating MAE ($\approx 4.2\%$), and width of grating (MAE $\approx 2.3\%$)

In this study, we reported inverse design prediction results by the DL model for all the variable in the design space. That is to say, we considered that all the design parameters are unknown and the entire parameter set is to be determined that properly reconstructs response. However, in practice, during the design or fabrication of the photonic structures there are often constraints about multiple design parameters. In this case, not all design parameters need to be predicted by the inverse model. If the design information about a subset of design parameters are fed in, it reduces the uncertainty of prediction of the model and the target space, thus reducing the prediction error. We corroborated that in Fig. 8b, for *variable single wavelength* input, which shows that as we fix each design parameter for prediction model, the design error for the other variables reduces. If two parameters are to be predicted- thickness t_1, t_2 or material grating design, w, p, the design error is lower than the full design prediction and single design parameter prediction is even lower. Also, it is noted that the prediction of design parameters - period of grating, p and thickness of metal bars, t_1 are the most difficult.
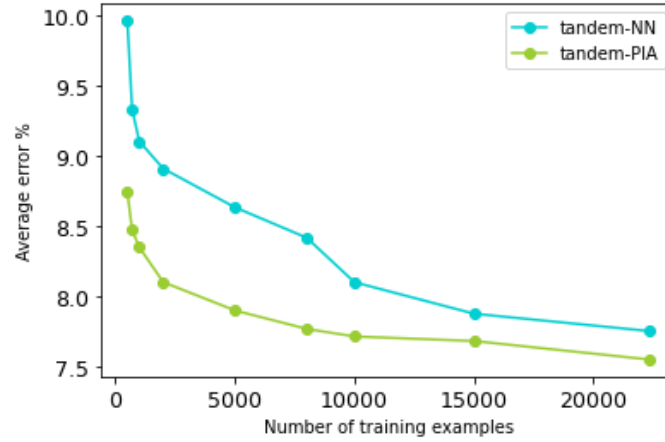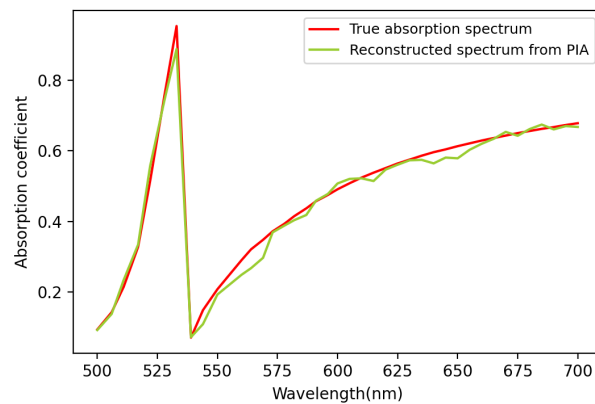
(a) Design error vs number of training examples for *fixed single wavelength* model.
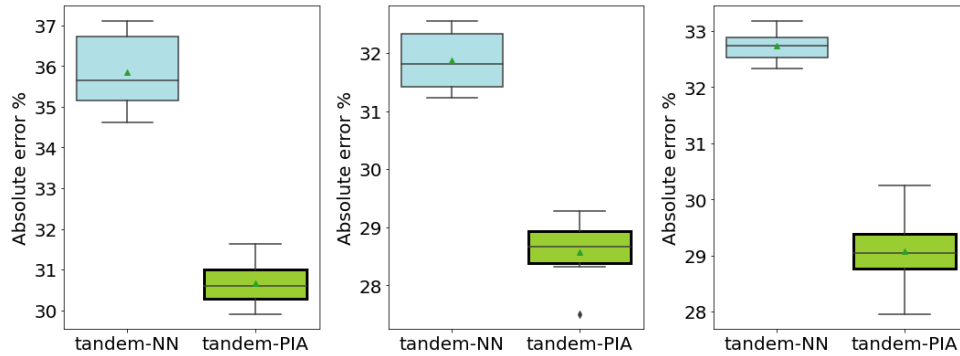
(b) Design error for period of metal grating, p, outside the training range

(c) Design error for thickness of polymer, t_2, outside the training range

(d) Design error for input parameter, absorption of light, outside the training range

(e) Reconstruction ability of the PIA model compared to the input response spectrum

Figure 7. Robustness of physics-informed learning vs data based learning - (a) With decrease in number of training examples PIA model performs increasingly better than tandem-NN (b), (c), (d) Generalization power of PIA is better than tandem-NN when the design parameters (period of grating, thickness of polymer) or input parameter (absorption of incident light) is out of range of the training samples (e) The PIA model has the ability to reconstruct the absorption response for a wavelength spectrum

(a) Variable importance for forward modeling

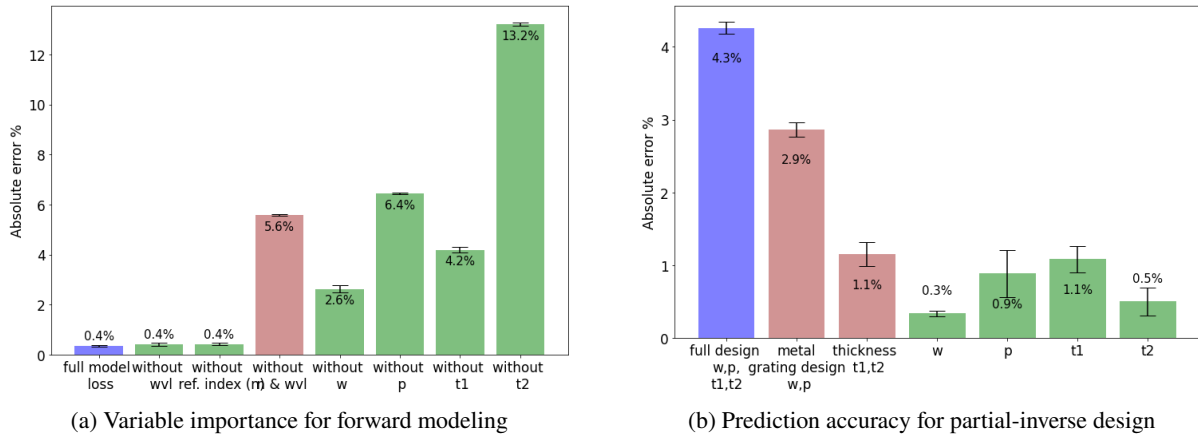(b) Prediction accuracy for partial-inverse design

Figure 8. (a) Analysing importance of variables in the forward modeling of optical response from input parameters. The figure shows the error in prediction of accurate optical response when any of the design, material or incident light parameter is absent. (b) The prediction power of the inverse DL model increases when partial information of design parameters is given.

5. Conclusion

In summary, to enable efficient and scientifically consistent design of metasurfaces in a supervised learning setting, we introduce a physics-informed machine learning model. This model leverages the power of neural networks to uncover the interdependence between device topology and optical response. To address the issues arising when using a deep neural network to solve inverse problems, such as non-unique predictions and high data burden, we propose a tandem architecture that predicts explainable and scientifically consistent design parameters while accurately reconstructing the optical response. By combining the forward and inverse models in the tandem architecture, we overcome the issue of non-unique prediction. Additionally, the inclusion of a physics-based penalty reduces data burden and increases generalizability while ensuring scientifically consistent prediction. To compute this physics-based term, we simplify the metamaterial structure into a stratified medium and solve for the optical response analytically. However, as our model deals with data where the dimensions are approximately equal to the wavelength, we restrict the physics-based loss function to sub-wavelength observations. To make use of physics knowledge for all observations, we developed a physics-informed DL architecture (PIA) that includes physics-consistent design parameters as intermediate neurons. This approach drives the model towards scientifically consistent prediction without constraining the final design output to be physics-consistent. Our proposed model achieves high accuracy in design prediction and optical response reconstruction, including up to 96% accuracy (4% MAE) for design prediction, 99.5% accuracy for the reconstruction of optical response for variable single wavelength input, and 97% accuracy in design prediction for multiple wavelength input.

Acknowledgments

[1] Z. Chen, A. Ogren, C. Daraio, L. C. Brinson, and C. Rudin, How to see hidden patterns in metamaterials with interpretable machine learning (2021), arXiv:2111.05949 [cs.LG].

[2] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, Integrating scientific knowledge with machine learning for engineering and environmental systems (2021), arXiv:2003.04919 [physics.comp-ph].

[3] P. Wang, F. Casadei, S. Shan, J. Weaver, and K. Bertoldi, Harnessing buckling to design tunable locally resonant acoustic metamaterials, Physical review letters **113**, 014301 (2014).

[4] L. Wang, J. Boddapati, K. Liu, P. Zhu, C. Daraio, and W. Chen, Mechanical cloak via data-driven aperiodic metamaterial design (2021), arXiv:2107.13147 [physics.app-ph].

[5] I. Arretche and K. H. Matlack, Effective phononic crystals for non-cartesian elastic wave propagation, Physical Review B **102**, 10.1103/physrevb.102.134308 (2020).

[6] C. Hu, Y. Xu, S. Zhu, and H. Chen, Phase-gradient metagratings via mode conversion, Phys. Rev. Applied **17**, 024023 (2022).

[7] D. R. Smith, J. B. Pendry, and M. C. K. Wiltshire, Metamaterials and negative refractive index, Science **305**, 788 (2004), https://www.science.org/doi/pdf/10.1126/science.1096796.

[8] J. A. Rodríguez, A. I. Abdalla, B. Wang, B. Lou, S. Fan, and M. A. Cappelli, Inverse design of plasma metamaterial devices for optical computing, Physical Review Applied **16**, 10.1103/physrevapplied.16.014023 (2021).

[9] 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**, 195104 (2020).

[10] C. Rockstuhl, C. Menzel, S. Mühlig, J. Petschulat, C. Helgert, C. Etrich, A. Chipouline, T. Pertsch, and F. Lederer, Scattering properties of meta-atoms, Phys. Rev. B **83** (2011).

[11] C. Rockstuhl, C. Menzel, T. Paul, and F. Lederer, Optical activity in chiral media composed of three-dimensional metallic meta-atoms, Physical Review B **79**, 35321 (2009).

[12] C. Liu, S. Maier, and G. Li, Genetic-algorithm-aided meta-atom multiplication for improved absorption and coloration in nanophotonics, ACS Photonics **XXXX** (2020).

[13] Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures, npj Computational Materials **6**, 12 (2020).

[14] A. Mall, A. Patil, D. Tamboli, A. Sethi, and A. Kumar, Fast design of plasmonic metasurfaces enabled by deep learning, Journal of Physics D: Applied Physics **53**, 49LT01 (2020).

[15] Y. Wang, Q. Zeng, J. Wang, Y. Li, and D. Fang, Inverse design of shell-based mechanical metamaterial with customized loading curves based on machine learning and genetic algorithm, Computer Methods in Applied Mechanics and Engineering **401**, 115571 (2022).

[16] H. Junjie, H. Lin, W. Xu, Y. Tian, Y. Wang, X. Shi, F. Deng, and L. Chen, Customized inverse design of metamaterial absorber based on target-driven deep learning method, IEEE Access **8**, 211849 (2020).

[17] Z. Liu, D. Zhu, L. Raju, and W. Cai, Tackling photonic inverse design with machine learning, Advanced Science **8**, 2002923 (2021), https://onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202002923.

[18] C.-T. Chen and G. Gu, Generative deep neural networks for inverse materials design using backpropagation and active learning, Advanced Science , 1902607 (2020).

[19] A. Shrivastava, J. Liu, K. Dayal, and H. Y. Noh, Predicting peak stresses in microstructured materials using convolutional encoder–decoder learning, Mathematics and Mechanics of Solids https://doi.org/10.1177/10812865211055504.

[20] X. Li, Y. Zhang, H. Zhao, C. Burkhart, L. C. Brinson, and W. Chen, A transfer learning approach for microstructure reconstruction and structure-property predictions, Scientific Reports **8**, 13461 (2018).

[21] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, Plasmonic nanostructure design and characterization via deep learning, Light: Science & Applications **7** (2018).

[22] M. Raissi, P. Perdikaris, and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics **378**, 686 (2019).

[23] O. Khatib, S. Ren, J. Malof, and W. Padilla, Deep learning the electromagnetic properties of metamaterials—a comprehensive review, Advanced Functional Materials **31** (2021).

[24] W. Ma, Z. Liu, Z. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, Deep learning for the design of photonic structures, Nature Photonics **15**, 1 (2020).

[25] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, and W. Cai, Generative model for the inverse design of metasurfaces, Nano Letters **18**, 6570–6576 (2018).

[26] J. Jiang, D. Sell, S. Hoyer, J. Hickey, J. Yang, and J. A. Fan, Free-form diffractive metagrating design based on generative adversarial networks, ACS Nano **13**, 8872–8878 (2019).

[27] W. Ma, F. Cheng, Y. Xu, Q. Wen, and Y. Liu, Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy, Advanced Materials **31**, 1901111 (2019), https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201901111.

[28] Z. Liu, L. Raju, D. Zhu, and W. Cai, A hybrid strategy for the discovery and design of photonic structures, IEEE Journal on Emerging and Selected Topics in Circuits and Systems **10**, 126 (2020).

[29] D. Liu, Y. Tan, E. Khoram, and Z. Yu, Training deep neural networks for the inverse design of nanophotonic structures, ACS Photonics **5**, 1365 (2018), https://doi.org/10.1021/acsphotonics.7b01377.

[30] A. Khaireh-Walieh, D. Langevin, P. Bennet, O. Teytaud, A. Moreau, and P. R. Wiecha, A newcomer's guide to deep learning for inverse design in nano-photonics (2023), arXiv:2307.08618 [physics.optics].

[31] I. J. Goodfellow, J. Shlens, and C. Szegedy, Explaining and harnessing adversarial examples (2015), arXiv:1412.6572 [stat.ML].

[32] S. Huang, N. Papernot, I. Goodfellow, Y. Duan, and P. Abbeel, Adversarial attacks on neural network policies (2017), arXiv:1702.02284 [cs.LG].

[33] T. Ma, M. Tobah, H. Wang, and L. Guo, Benchmarking deep learning-based models on nanophotonic inverse design problems, Opto-Electronic Science **1** (2022).

[34] W. Li, P. Chen, B. Xiong, G. Liu, S. Dou, Y. Zhan, Z. Zhu, T. Chu, Y. Li, and W. Ma, Deep learning modeling strategy for material science: from natural materials to metamaterials, Journal of Physics: Materials **5**, 014003 (2022).

[35] A. Estrada-Real, A. Khaireh-Walieh, B. Urbaszek, and P. R. Wiecha, Inverse design with flexible design targets via deep learning: Tailoring of electric and magnetic multipole scattering from nano-spheres, Photonics and Nanostructures - Fundamentals and Applications **52**, 101066 (2022).

[36] x. xu, Y. Li, and W. Huang, An improved tandem neural network for the inverse design of nanophotonics devices, Optics Communications **481** (2020).

[37] X. Han, Z. Fan, Z. Liu, C. Li, and L. Guo, Inverse design of metasurface optical filters using deep neural network with high degrees of freedom, InfoMat **3** (2020).

[38] F. Gao, Z. Zhang, Y. Xu, L. Zhang, R. Yan, and X. Chen, Deep-learning-assisted designing chiral terahertz metamaterials with asymmetric transmission properties, J. Opt. Soc. Am. B **39**, 1511 (2022).

[39] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, Nature Reviews Physics **3**, 422 (2021).

[40] S. Pawar, O. San, B. Aksoylu, A. Rasheed, and T. Kvamsdal, Physics guided machine learning using simplified theories, Physics of Fluids **33**, 011701 (2021).

[41] Y. Tian, D. Livescu, and M. Chertkov, Physics-informed machine learning of the lagrangian dynamics of velocity gradient tensor (2021), arXiv:2108.07244 [physics.flu-dyn].

[42] Y. Chen, L. Lu, G. E. Karniadakis, and L. D. Negro, Physics-informed neural networks for inverse problems in nano-optics and metamaterials, Opt. Express **28**, 11618 (2020).

[43] Z. Fang and J. Z. Zhan, Deep physical informed neural networks for metamaterial design, IEEE Access **8**, 24506 (2020).

[44] L. Lu, R. Pestourie, S. G. Johnson, and G. Romano, Multifidelity deep neural operators for efficient learning of partial differential equations with application to fast inverse design of nanoscale heat transport (2022), arXiv:2204.06684 [physics.comp-ph].

[45] R. Pestourie, Y. Mroueh, C. Rackauckas, P. Das, and S. G. Johnson, Physics-enhanced deep surrogates for pdes (2022), arXiv:2111.05841 [cs.LG].

[46] R. Paniagua-Domínguez, F. López-Tejeira, R. Marqués, and J. A. Sánchez-Gil, Metallo-dielectric core–shell nanospheres as building blocks for optical three-dimensional isotropic negative-index metamaterials, New Journal of Physics **13**, 123017 (2011).

[47] J. Mun, S. So, and J. Rho, Spectrally sharp plasmon resonances in the near infrared: Subwavelength core-shell nanoparticles, Phys. Rev. Applied **12**, 044072 (2019).

[48] I. D. Rukhlenko, M. Premaratne, and G. P. Agrawal, Theory of negative refraction in periodic stratified metamaterials, Opt. Express **18**, 27916 (2010).

[49] O. V. Shramkova and G. P. Tsironis, Propagation of electromagnetic waves in pt-symmetric hyperbolic structures, Phys. Rev. B **94**, 035141 (2016).

[50] J. Park, J. Kang, S. J. Kim, X. Liu, and M. Brongersma, Dynamic reflection phase and polarization control in metasurfaces, Nano Letters **17** (2016).

[51] F. Ding, Y. Yang, R. A. Deshpande, and S. I. Bozhevolnyi, A review of gap-surface plasmon metasurfaces: fundamentals and applications, Nanophotonics **7**, 1129 (2018).

[52] J. S. Ho, B. Qiu, Y. Tanabe, A. J. Yeh, S. Fan, and A. S. Y. Poon, Planar immersion lens with metasurfaces, Physical Review B **91**, 10.1103/physrevb.91.125145 (2015).

[53] A. Kwadrin and A. F. Koenderink, Diffractive stacks of metamaterial lattices with a complex unit cell: Self-consistent long-range bianisotropic interactions in experiment and theory, Phys. Rev. B **89**, 045120 (2014).

[54] C. Pouya, J. Overvelde, M. Kolle, J. Aizenberg, K. Bertoldi, J. Weaver, and P. Vukusic, Characterization of a mechanically tunable gyroid photonic crystal inspired by the butterfly parides sesostris, Advanced Optical Materials **4**, n/a (2015).

[55] S. Ogawa and M. Kimata, Metal-insulator-metal-based plasmonic metamaterial absorbers at visible and infrared wavelengths: A review, Materials **11**, 10.3390/ma11030458 (2018).

[56] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light (Second Edition)*, 2nd ed. (Princeton University Press, 2008).

[57] R. Brunner, O. Sandfuchs, C. Pacholski, C. Morhard, and J. Spatz, Lessons from nature: Biomimetic subwavelength structures for high-performance optics, Laser & Photonics Review **6**, 641–659 (2012).

[58] H. Kikuta, Y. Ohira, H. Kubo, and K. Iwata, Effective medium theory of two-dimensional subwavelength gratings in the non-quasi-static limit, J. Opt. Soc. Am. A **15**, 1577 (1998).

[59] G. Bouchitte and R. Petit, Homogenization techniques as applied in the electromagnetic theory of gratings, Electromagnetics **5**, 17 (1985), https://doi.org/10.1080/02726348508908135.

[60] P. Lalanne and M. Hutley, Artificial media optical properties-subwavelength scale (2003).

[61] D. L. Brundrett, E. N. Glytsis, and T. K. Gaylord, Homogeneous layer models for high-spatial-frequency dielectric surface-relief gratings: conical diffraction and antireflection designs, Appl. Opt. **33**, 2695 (1994).

[62] Y. Bar-Sinai, G. Librandi, K. Bertoldi, and M. Moshe, Geometric charges and nonlinear elasticity of two-dimensional elastic metamaterials, Proceedings of the National Academy of Sciences **117**, 10195 (2020), https://www.pnas.org/content/117/19/10195.full.pdf.

[63] G. U. Patil and K. H. Matlack, Effective property evaluation and analysis of three-dimensional periodic lattices and composites through bloch-wave homogenization, The Journal of the Acoustical Society of America **145**, 1259 (2019), 31067925[pmid].

[64] S. M. Rytov, Electromagnetic properties of a finely stratified medium (2014).

[65] M. C. Troparevsky, A. S. Sabau, A. R. Lupini, and Z. Zhang, Transfer-matrix formalism for the calculation of optical response in multilayer systems: from coherent to incoherent interference, Opt. Express **18**, 24715 (2010).

[66] J. Hao and L. Zhou, Electromagnetic wave scatterings by anisotropic metamaterials: Generalized 4×4 transfer-matrix method, Phys. Rev. B **77**, 094201 (2008).

[67] C. Nadell, B. Huang, J. Malof, and W. Padilla, Deep learning for accelerated all-dielectric metasurface design, Optics Express **27**, 27523 (2019).

[68] W. Ma, F. Cheng, and Y. Liu, Deep-learning enabled on-demand design of chiral metamaterials, ACS Nano **12** (2018).

[69] L. Lu, Dying ReLU and initialization: Theory and numerical examples, Communications in Computational Physics **28**, 1671 (2020).

[70] S. Elfwing, E. Uchibe, and K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning (2017), arXiv:1702.03118 [cs.LG].

[71] M. M. Bay, S. Vignolini, and K. Vynck, PyLlama: A stable and versatile python toolkit for the electromagnetic modelling of multilayered anisotropic media, Computer Physics Communications **273**, 108256 (2022).

[72] W. Ji, J. Chang2, H.-X. Xu, J. R. Gao, S. Gröblacher, P. Urbach, and A. J. L. Adam, Recent advances in metasurface design and quantum optics applications with machine learning, physics-informed neural networks, and topology optimization methods (2023), arXiv:2307.09126 [physics.optics].

[73] P. Pillai, P. Pal, R. Chacko, D. Jain, and B. Rai, Leveraging long short-term memory (lstm)-based neural networks for modeling structure–property relationships of metamaterials from electromagnetic responses, Scientific Reports **11**, 18629 (2021).