Inverse Design Framework With Invertible Neural Networks for Passive Vibration Suppression in Phononic Structures

Manaswin Oddiraju
Department of Mechanical and Aerospace Engineering,
University at Buffalo,
Buffalo, NY 14260
e-mail: moddiraj@buffalo.edu

Amir Behjat
Department of Mechanical and Aerospace Engineering,
University at Buffalo,
Buffalo, NY 14260
e-mail: amirbehj@buffalo.edu

Mostafa Nouh
Associate Professor
Department of Mechanical and Aerospace Engineering,
University at Buffalo,
Buffalo, NY 14260
e-mail: mnouh@buffalo.edu

Souma Chowdhury[1]
Associate Professor
Department of Mechanical and Aerospace Engineering,
University at Buffalo,
Buffalo, NY 14260
e-mail: soumacho@buffalo.edu

Automated inverse design methods are critical to the development of metamaterial systems that exhibit special user-demanded properties. While machine learning approaches represent an emerging paradigm in the design of metamaterial structures, the ability to retrieve inverse designs on-demand remains lacking. Such an ability can be useful in accelerating optimization-based inverse design processes. This paper develops an inverse design framework that provides this capability through the novel usage of invertible neural networks (INNs). We exploit an INN architecture that can be trained to perform forward prediction over a set of high-fidelity samples and automatically learns the reverse mapping with guaranteed invertibility. We apply this INN for modeling the frequency response of periodic and aperiodic phononic structures, with the performance demonstrated on vibration suppression of drill pipes. Training and testing samples are generated by employing a transfer matrix method. The INN models provide competitive forward and inverse prediction performance compared to typical deep neural networks (DNNs). These INN models are used to retrieve approximate inverse designs for a queried non-resonant frequency range; the inverse designs are then used to initialize a constrained gradient-based optimization process to find a more accurate inverse design that also minimizes mass. The INN-initialized optimizations are found to be generally superior in terms of the queried property and mass compared to randomly initialized and inverse DNN-initialized optimizations. Particle swarm optimization with INN-derived initial points is then found to provide even better solutions, especially for the higher-dimensional aperiodic structures.
[DOI: 10.1115/1.4052300]

Keywords: design automation, machine learning, invertible neural network, inverse design, metamaterial

1 Introduction

Inverse design refers to a process of finding the inputs x which produce the desired value of the outputs y constrained to the following mapping $f: x \rightarrow y$, while also minimizing other cost functions in some cases [1]. Many on-demand designs and hierarchical design problems can be formulated as inverse design problems; a few notable examples close to the topic area of this paper include the design of metamaterials [2–5], nanophotonics [6], and new molecular structures [7–9]. Optimization techniques are by far one of the most popular approaches to solve inverse design problems [9,10], which are often modeled as a constrained optimization problem. Solving inverse problems using optimization for complex nonlinear systems is in general computationally expensive and in some cases intractable due to the ill-posed nature of inverse problems [9,11]. There is also a growing body of machine learning applications to support inverse design [12,13], although their potential use for on-demand design retrieval remains under-explored. In this paper, we address these challenges by introducing a novel invertible learning approach to retrieve approximate inverse solutions on-demand and use them to accelerate design optimization processes. This new inverse design framework is particularly tailored to automated design of metamaterials for vibration regulation, and a 1D phononic structure problem is used to analyze and demonstrate the framework's performance in this paper.

In the rest of this section, we present a brief survey of existing methods to systematically design phononic structures and related metamaterial systems, followed by a survey of existing work in using machine learning (ML) techniques for inverse design of such systems. Thereafter, we introduce the primary contributions of our work and converge on the key objectives of this paper.

1.1 Design of Phononic Structures and Metamaterials.

In recent years, the notion of using passive periodic, i.e., self-repeating, individual cells as building blocks for composite-like structures has significantly grown [14]. Notable among these are phononic structures in which the building blocks typically take the form of alternating materials, geometries, or a combination thereof, as well as metamaterials which comprise locally resonant inclusions within a host medium. It has been shown that such systems can be engineered to possess unique characteristics such as their ability to exhibit extended regions of forbidden wave propagation known as bandgaps [15–17], superior dissipative properties [18,19], negative effective stiffness [20], unidirectional vibrations [21], among others. With these capabilities, these systems have become increasingly popular, and despite manufacturing challenges [22–24], they are being translated to diverse applications ranging from acoustics [25,26] and photonics [27,28] to sensing and other end uses [29,30].

Suppression and mitigation of undesirable structural vibrations are everlasting challenges in oscillating mechanical platforms and structures exhibiting elastic deformations. Over the years, this has

[1]Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received April 23, 2021; final manuscript received August 27, 2021; published online September 23, 2021. Assoc. Editor: James T. Allison.

garnered significant attention with research efforts spanning various passive and active vibration control methods [31–34]. Phononic structures and metamaterials present an attractive solution to this problem because of their ability to exhibit bandgaps. The search for novel features and functionalities in such systems typically requires analyses that dive into the fundamental constitutive equations of new architectures or unit cell configurations [35]. However, the design of such structures for engineering applications or the on-demand discovery of novel structural configurations is usually more exhaustive and often computationally prohibitive if a classical iterative approach is taken. Notable work in speeding up the search of metamaterial design includes the use of simplified models to design periodic nano-antennas [36] and the use of group theory to design bianisotropic metamaterials [37], where candidates are grouped based on their similarity in properties and this group space is searched based on desired properties given by an user. However, simplified models may not be available in every problem or provide adequate fidelity of outcomes. Furthermore, an increase in the number of properties of interest or non-linearity of the design/property relationship will limit the application of the latter approach. These challenges are expected to further increase if aperiodic designs are to be explored [38–40], since the number of design variables could increase by an order of magnitude compared to their periodic counterparts.

1.2 Machine Learning Approaches to Inverse Design. The limitation of classical inverse design methods and the significant progress made in the field of machine learning in the recent years have positioned tools such as deep neural networks (DNNs) [5,41–44] and generative adversarial networks (GANs) [41,45–48] as valuable alternatives to classical methods. Training them offline can be (albeit, not guaranteed to be) computationally less expensive compared to running multiple optimizations, and once trained, they can be used to answer queries instantaneously.

For instance, Ma et al. [49] used a variational autoencoder-based generative model to design photonic metamaterials. However, this approach remains limited to relatively simple designs with less than ten parameters and requires a significant amount of training data. Similarly, approaches based on GANs while effective in inverse design (e.g., nanophotonic structures [45]) usually require enormous data sets to train. In another work, Chan et al. [50] proposed an approach to accelerate the metamaterial design process by creating unit cell databases, which can then be used to quickly design structures by using combinatorial optimization; this approach is however not amenable to on-demand design retrieval. Morris et al. [51] used Bayesian network classifiers to design negative-stiffness metamaterials. This method can find a set of feasible designs and also allows uncertainty analysis. Similarly, Bostanabad et al. [52] proposed an approach called globally approximate Gaussian process for the design of orthotropic metamaterials, which uses multiple independent Gaussian process (GP) models trained over big data sets. As noted above, the provision to retrieve a single design on-demand is not readily evident in most of these approaches, and majority of them work when the feature space is small (as in periodic metamaterials). The former provision is however available with more recent approaches using pairs of DNNs and conditional machine learning models, which are further discussed below.

DNNs are also being increasingly used in the space of metamaterial design, with a few examples reported by Ma et al. [53] and Finol et al. [54] in the photonic and phononic domains, respectively. A notable (and likely the only one of its kind) DNN-based approach that explicitly allows both forward and inverse predictions in metamaterial design is evident in the work of Malkiel et al. on nanophotonic structures [42–44]. This approach uses two separate DNNs to perform forward and inverse modeling. However, the sample requirement to train the inverse model is quite high, and no implicit capability of dealing with the non-uniqueness of the inverse mapping is provided. Moreover, the relation between the forward and inverse mapping is not rigorously preserved, i.e., with this approach, satisfaction of $f_{\mathrm{Inv}}(f_{\mathrm{Fwd}}(x)) = x$ is not guaranteed. This makes it difficult to reliably retrieve multiple designs in a close neighborhood of a target property value for analysis and optimization purposes.

Here, we are interested in investigating an inverse design framework that allows on-demand retrieval of designs (given user-specific desired properties), guarantees the $f_{\mathrm{Inv}}(f_{\mathrm{Fwd}}(x)) = x$, relation, and can work on fairly large input feature spaces as given by aperiodic structures.

Another emerging class of models being used for inverse design is conditional generative models. These models improve the efficiency of generative models by conditioning the generator and discriminators on data/class labels for its training. Hodge et al. [55] used conditional GANs or cGAN to design tensorial meta-surfaces for radio frequency applications. Here, the spectral response is added as a condition to limit the output of the generative model, which however requires further post processing by a simulator network. Yilmaz and German used a similar structure for inverse design of airfoils [56], conditioned with parameters such as the range of the angle of attack for stall instances, and drag and lift coefficients. Nobari et al. proposed another idea of range-constrained GAN where the conditions are in the form of the range of the variables [57], which is claimed to provide the advantage of generating a uniformly distributed design. The main limitation of this approach is the reliance on a label estimator, which must also be differentiable and exact. Another related approach uses performance conditioned diverse GAN (PcDGAN) [58], which enhances the diversity in output to improve the coverage over design space while maintaining continuous conditions. This method also requires a differentiable estimator, which in this case is based on regression and thus affected by the regression accuracy. Thus overall, conditioned GANs provide a promising approach to inverse design, while being limited by the availability and cost of evaluating conditions in real-world problems (e.g., lift on complex 3D airfoil sections) and their special training needs (e.g., estimators) in general.

1.3 Research Objectives of This Work. In this paper, which is a methodological and empirical extension of our 2020 ASME IDETC paper [59], we explore a new modeling concept—provably ***invertible*** *neural networks* or INN—to enable on-demand design and retrieval of both periodic and aperiodic phononic structures. Here, the design efforts are focused on achieving on-demand vibration characteristics, namely, desired frequency range(s) with no resonant peaks, i.e., non-resonant frequency ranges (NRFRs). *Once an INN is trained over samples, each containing a phononic structure design and its frequency response, the INN can not only be used to predict the frequency response of other unseen designs but also executed in reverse to predict a design that corresponds to a given frequency response.* To our knowledge, this is the first formal exploration of INN for inverse design of structural metamaterial systems. Thus, we also develop standard DNN models of the forward problem to compare and contrast the prediction performance of the INNs in this application context. In addition, we develop a forward optimization process for designing such phononic structures while minimizing the structural mass and illustrate how the novel inverse retrieval approach can significantly accelerate or improve the performance of the forward optimization, compared to conducting a standard constrained forward optimization (with random initialization) for inverse design.

Thus, the specific objectives of this paper include:

(1) Develop an INN-based modeling approach to represent the vibration characteristics of phononic structures, along with DNN models for comparison, both trained on data generated by a transfer matrix method (TMM).
(2) Develop a TMM-based optimization framework that can be efficiently initiated with INN-retrieved approximate solutions for the accelerated design of 1D phononic structures.

(3) Demonstrate the performance of the INN-based inverse design framework in generating periodic and aperiodic structures for 1D phononic systems with on-demand lateral vibration characteristics.

The rest of this paper is structured as follows. In Sec. 2, we describe the main components of our new inverse design framework. Section 3 presents the problem formulation as applied to the 1D phononic structure case study. Section 4 presents and discusses the model prediction and inverse design results, along with further comparative analysis of periodic versus aperiodic designs obtained thereof. Section 5 summarizes the concluding remarks.

2 Invertible Neural Network-Based Framework to Design the Phononic Structure

Figure 1 illustrates the overall design automation framework. This framework aims to combine INN-based inverse retrieval and simulation-based optimization to find the optimal configuration of a given system subject to specific (on-demand) characteristics. A phononic structure inspired by drill pipes is chosen as a case study to analyze this framework's performance and the benefits of the underlying new inversion concepts. The main components of the framework as shown in Fig. 1 are as follows:

(i) Generation of samples using a medium fidelity forward model; in this case, we use the TMM to model the vibration response of 1D phononic structures.
(ii) Training of the INN on these samples to map periodic and aperiodic structural designs (inputs) to the output quantity of interest.
(iii) Instantaneous generation of approximate design solutions given a desired value of the output, by performing inverse prediction through the INN.
(iv) Optimization process for inverse design, with the desired output values set as constraints, and initialized by the INN-derived approximate inverse solution.

While for the case studies in this paper the TMM is used to generate the samples, the automated inverse design framework can be applied to other metamaterial problems beyond the phononic structures studied here using a pertinent physics-based simulation to generate the samples. The fundamental concept of inverse retrieval of solutions via the INN and its integration with the forward optimization process is thus expected to work in a wide variety of on-demand design problems. In any such application, the INN must be trained on the generated samples to model the "design → properties" relationship, and it automatically also learns the reverse "properties → design" relation. For the application considered here, the property of interest is defined to be the frequency response of lateral vibrations. The remainder of this section describes the two main modeling components of our current automated inverse design framework, namely, the INN and TMM.

2.1 Invertible Neural Networks.

INNs are a class of neural networks that have a unique architecture, which enables their invertibility [60]. When an INN is trained on the well-understood forward problem, it can capture the inverse model implicitly [60]. A handful of notable INN architectures have been reported in recent years [61], with preliminary applications to image reconstruction [62,63], parameter estimation [60,64], and generative flow modeling [65]. In this paper, we chose the INN architecture proposed in Ref. [60] since it ensures an analytically computable inverse by preserving the nonsingular nature of the Jacobian that encapsulates the output to input derivatives for each layer of the network.

Figure 2 illustrates the general architecture of our INN implementation, including the representative invertible INN block showing information flow in both forward and inverse directions (Figs. 2(*a*) and 2(*b*), respectively), as well as the overall network structure comprising multiple such blocks (Fig. 2(*c*)). The transformations encapsulated by an INN block can be expressed as

$$\begin{aligned} v_1 &= u_1 \odot \exp\left(s_2(u_2)\right) + t_2(u_2) \\ v_2 &= u_2 \odot \exp\left(s_1(v_1)\right) + t_1(v_1) \end{aligned} \tag{1}$$

where $\odot$ represents element-wise multiplication, such that $[a_{i,j}] \odot [b_{i,j}] = [a_{i,j} \times b_{i,j}]$, with $[a_{i,j}]$ and $[b_{i,j}]$ being two matrices of equal dimensions. Here, u and v are the inputs and outputs of the INN and the subscripts 1 and 2 depict two splits of the input vector. This INN architecture requires the input vector to be split in order to keep the layer-wise Jacobian a triangular matrix; as in

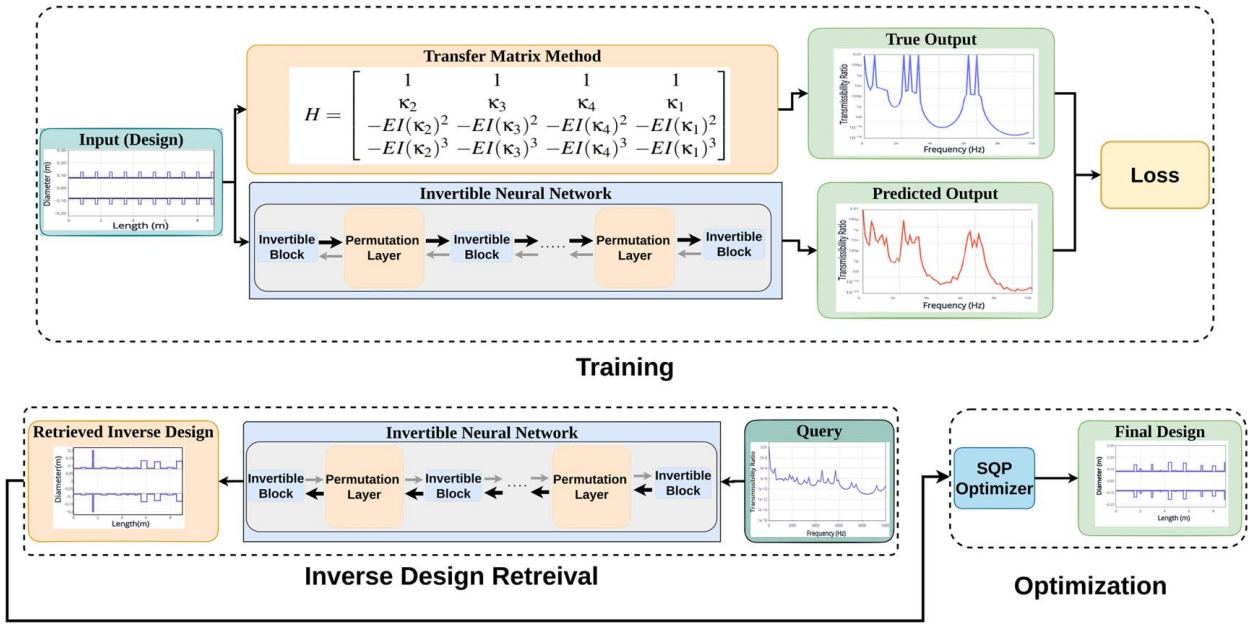

Fig. 1 Inverse design framework for phononic structures with on-demand frequency response. *Top:* Training of the INN model using samples given by TMM, *bottom:* inverse design process, involving approximate inverse solution retrieval via INN (executed backwards) given a frequency response query, and optimization process initialized with this solution.

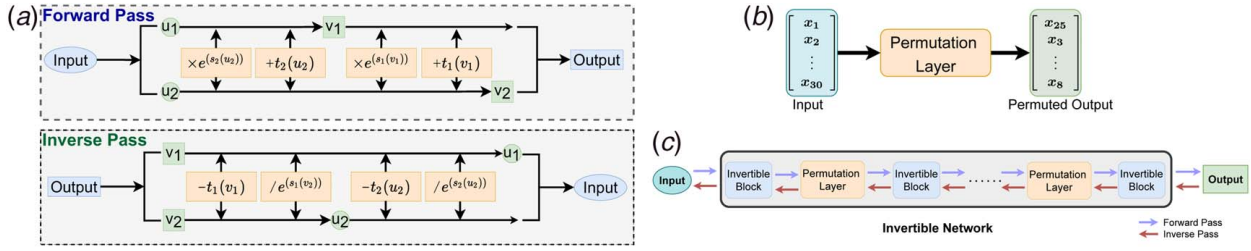

Fig. 2 INN architecture: (*a*) invertible blocks, (*b*) permutation block, and (c) entire INN

Ref. [60], for an even-sized input vector (of size $2m$, $\sim m \in Z^+$) the splits are equal, and for an odd sized vector (of size $2m+1$, $\sim m \in Z^+$), the splits get assigned m and $m+1$ input elements, respectively. The terms s_1, t_1 and s_2, t_2 can be any arbitrarily complicated functions or transformations of u_2 and v_1, respectively. Here, they are embodied by neural networks with fully connected layers and leaky ReLU activation functions.

The above expressions can be easily inverted if the output $v = [v_1,\ v_2]$ is known based on the expressions:

$$\begin{aligned} u_2 &= (v_2 - t_1(v_1)) \odot \exp(-s_1(v_1)) \\ u_1 &= (v_1 - t_2(u_2)) \odot \exp(-s_2(u_2)) \end{aligned} \tag{2}$$

The structure shown in Fig. 2(*a*) constitutes a single representative INN block. Now, as shown in Fig. 2(*c*), the overall INN model is essentially a deeper network composed of a series of these blocks with permutation layers in between to shuffle the input to the next block in a pseudo-random manner. This pseudo-random permutation causes the input vector splits, u_1 and u_2, to vary between layers, thereby increasing the inter-variable interaction [60]. The expressed architecture requires that the model's input and output vectors have the same size. If input and output vectors have different dimensions, as is the case in our problems, it can be handled as follows: (i) if the size of input > size of output, another set of variables called the *latent variables* are generated from a normal distribution and concatenated to augment the output vector; the size of the latent variable vector is equal to the difference of the input and output vector sizes; these latent variables can also be used to impose additional constraints to mitigate the ill-posed nature of the inverse prediction. (ii) If the size of input < size of output, the input vector is either padded with zeros or repeated multiple times to match the dimension of the output vector [60].

For the phononic structure problem in this paper, the number of outputs (80) far exceeds the number of the inputs (30 for an aperiodic system and 3 for a periodic system). Therefore, we repeat the design variable vector multiple times to pad the input vector and make its dimension equal to that of the output vector. The other model parameters used in the INNs constructed in this paper are listed in Table 3.

In our application, owing to the manner in which the output space is encoded, as discussed later in Sec. 3.3, the mapping was observed to be mostly of one-to-one nature. This gave us the opportunity to impose a direct loss term for the inverse direction, which is found to improve training convergence and the accuracy of inverse predictions, compared to the more typical INN loss function that uses *maximum mean discrepancy* (MMD) in the inverse direction. This chosen loss function with the first term representing the forward prediction loss in terms of *mean squared error* (MSE) and the second term representing the direct MSE in the inverse direction can be expressed as given in Eqs. (3)–(5):

$$\text{Loss} = E_{\text{MSE}}(Y_{ij}, \hat{Y}_{ij}) + E_{\text{MSE}}(X_{ij}, \hat{X}_{ij}) \tag{3}$$

$$E_{\text{MSE}}(X_{ij}, \hat{X}_{ij}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\left(\hat{X}_{ij} - X_{ij}\right)^2 \tag{4}$$

$$E_{\text{MSE}}(Y_{ij}, \hat{Y}_{ij}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\left(\hat{Y}_{ij} - Y_{ij}\right)^2 \tag{5}$$

Here, Y_{ij} and $\hat{Y}_{ij}$, respectively, represent the true output and the network prediction of the jth dimension of the ith sample and X_{ij} and $\hat{X}_{ij}$, respectively, represent the true input and the inverse network prediction of the jth dimension of the ith sample. Here, N and M, respectively, depict the number of samples and the dimension of the output vector. Similarly, the second term in Eq. (3) can be obtained by replacing $\hat{Y}_{ij}$ and Y_{ij} with $\hat{X}_{ij}$ and X_{ij}, respectively.

The INN can also be used in cases where there is a many-to-one mapping, i.e., in cases where multiple inputs map to the same output. In that case, the loss function used for training the INN as specified in Ref. [60] is a weighted combination of two terms: (1) the forward loss and (2) a penalty applied to capture how the distribution of the inputs predicted when the INN is executed in the reverse direction deviate from the distribution of the actual inputs. This overall loss function is thus given by

$$\text{Loss} = E_{\text{MSE}}(Y_{ij}, \hat{Y}_{ij}) + \lambda E_{\text{MMD}}(\hat{X}_{ij}, X_{ij}) \tag{6}$$

where λ is the weight assigned to the second term.

To capture the deviation of the inversely predicted input distribution, a kernel-based loss called MMD is used, which is the second term in Eq. (6). A brief mathematical description of MMD is given in Appendix A, and further details can be found in Ref. [60]. This loss function given by Eq. (6) is particularly useful where the inverse is non-unique (i.e., many-to-one X–Y mapping clearly exists), and thus latent parameters are used in the INN to enable the inverse mapping to capture the input distribution and allow the retrieval of multiple designs by sampling over latent parameters.

2.2 Approximation Model Baselines: Deep Neural Nets. For the purpose of comparing the INN's performance in forward and inverse predictions, and later its further impact in inverse design retrieval, we develop two different types of DNN models. These are standard DNN models that can only predict a mapping in one direction and do not provide bijectivity or explicit invertibility. The first type of the DNN model, to be called as forward DNN or DNN-`F` here onward, maps from the design input X to the property output Y and trains over the loss function given by the first term of Eq. (3). DNN-`F` is used to benchmark the forward modeling accuracy of the INN to lay to rest the typical concerns regarding the loss of expressibility in the INN's forward predictions, thereby preserving the utility of INN's bijectivity, e.g., in retrieving multiple inverse designs (refer Sec. 3.4). The second type of the DNN model, to be called as inverse DNN or DNN-`I` here onward, maps from the property output Y to the design input X, and trains over the loss function given by the second term of Eq. (3). DNN-`I` is used as a comparative baseline to analyze the inverse design retrieval performance of the INN.

2.3 Transfer Matrix Method. The TMM is an analytical method commonly used to model and depict the dynamics of elastic structures made up of adjacent blocks [66]. At its core, the

TMM tracks the transmission of different forces (or moments as applicable) and displacements (or rotations as applicable) from one segment of a structural system to the neighboring one [67]. Therefore, to capture the performance of a tapered elastic structure or one of varying cross section, it must be discretized to different cylindrical sub-bodies. This discretization may lead to a loss of model fidelity due to an insufficient number of discrete segments. Another limitation of the TMM method are numerical errors, which are often associated with calculating the analytical response depending on matrix conditions. In some cases, solving the TMM requires a high degree of numerical precision in order to remain computationally stable, which results in increased computational time.

Here, TMM is used to compute the transmissibility ratio of the 1D phononic structures with respect to lateral vibrations. The term "transmissibility ratio" refers to the amount of vibrations transmitted from one end of a structure to the other end when subjected to an external incident excitation. The excitation can be modeled as a displacement or force based on the boundary conditions. While the transmissibility ratio is typically below unity within non-resonant regimes, it jumps significantly during resonance, which can be detrimental to the health of the structure, delicate components, and sensitive payloads.

For an Euler–Bernoulli beam with free-free boundary conditions, the transmissibility ratio (T_R) as given by the TMM method can be expressed as

$$\begin{aligned} T_R(\omega) &= (A \times E) - (B \times E) - C + D \\ A &= \bar{T}[1,1] \\ B &= \frac{\bar{T}[1,2] \times \bar{T}[4,1]}{\bar{T}[4,2]} \\ C &= \frac{\bar{T}[1,2] \times \bar{T}[4,4]}{\bar{T}[4,2]} \\ D &= \bar{T}[1,4] \\ E &= \frac{\bar{T}[3,2] \times \bar{T}[4,4] - \bar{T}[3,4] \times \bar{T}[4,2]}{\bar{T}[3,1] \times \bar{T}[4,2] - \bar{T}[3,2] \times \bar{T}[4,1]} \end{aligned} \tag{7}$$

Here, $\bar{T}$ is the transfer matrix and ω is the angular frequency (expressed in rad/s) of the pertinent degree-of-freedom; this is the lateral displacement in our case studies, as depicted in Fig. 3(*a*). Our detailed derivation of the above equations and the composition of $\bar{T}$ for the problem considered here are provided in a public repository.[2] Here onward, the TMM model will be referred to as the high-fidelity forward model.

3 Phononic Structure Problem

As our application problem, we chose a cylindrical base-structure similar to a drill pipe, as shown in Fig. 3(*a*). The long and slender geometry of drill pipes makes it susceptible to vibrations. To instill vibration suppression capabilities, drill pipes can be constructed as 1D phononic structures by adding inserts in the form of annular rings to their periphery in order to give rise to frequency bandgaps, i.e., extended regions of forbidden vibrations, as shown in Ref. [68]. The addition of inserts of appropriate geometry, when optimally placed, was experimentally shown to mitigate high amplitude vibrations, thereby minimizing the possibility of excessive damage to the drill pipe during operation [69]. Along with the increased design complexity of the drill pipe augmented with annular rings, this problem also presents significant modeling complexity as the vibration characteristics are challenging to represent due to resonances. These resonances appear as sudden spikes in the transmissibility ratio, adding numerical irregularity to the data and rendering the problem a good test-bed for the analysis shown here. In the remainder of this section, we provide the problem description of our case studies based on the drill pipe example, the formulation of the optimization problem for inverse design of this 1D phononic structure, and the inverse modeling setup used to train INNs.

[2]https://github.com/adamslab-ub/1D-Phononic-Structure-Design-with-INN

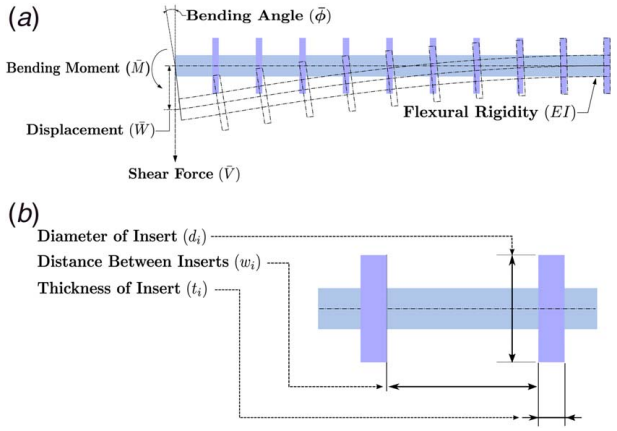

Fig. 3 1D phononic structure (drill pipe with inserts): (*a*) geometry and physical parameters and (*b*) design variables

3.1 Periodic and Aperiodic Structures. In addition to periodic designs as proposed in Ref. [68], we also explore aperiodic designs that lift the periodic constraints from the inserts. This enables greater flexibility, while at the same time leading to an n-fold increase in the dimensionality of the design space that must be searched (n being the number of inserts), both in the forward and inverse problems. The increased flexibility thereof could provide further reduction in mass over a periodic structure, while seeking to match the desired non-resonant frequency characteristics of the latter. The aperiodic structure could also hypothetically offer opportunities for reconfiguration during the system's lifetime in order to target different frequency ranges where resonance is to be mitigated. A uniform self-repeating pattern, which is typically the case in phononic and metamaterial structures, does not inherently provide this flexibility.

Figure 3(*a*) shows the drill pipe structure with the inserts and Fig. 3(*b*) (with the radial dimension enhanced) shows the design variables to be optimized. Throughout this paper, the number of inserts is kept fixed at 10. As seen from the figure, the aperiodic structure is parameterized in terms of d_i, w_i, and t_i, where d_i and t_i are the outer diameter and the thickness of the ith insert, respectively, and w_i is the distance between two adjacent inserts. For the corresponding periodic structure, we therefore only have three design variables, d, t, and w. The geometry of the base cylindrical structure and the material properties of the system are listed in Table 1.

3.2 Optimization Formulation. Our optimization formulation aims to minimize the mass of the phononic structure while maintaining a range of frequencies within which no resonance occurs. As such, the constraint violation is represented in terms of a counter $h(x, \omega_L, \omega_U)$ that counts the number of peaks in the frequency response between ω_L and ω_U, which are, respectively, the

Table 1 Geometrical and material properties

Properties of the phononic structure	Value
Length (l)	9 m
Density (ρ)	1800 kg/m^3
Elastic modulus (E)	193 GPa
Shear modulus (G)	77.2 GPa
Inner diameter of insert (d_{in})	0.16 m

Table 2 Design variable bounds

Variable	Bounds
d_i	[0.16 m, 0.4 m]
w_i	[0.075 m, 0.3750 m]
t_i	[0.15 m, 2.25 m]

start and end frequencies of the desired (on-demand) non-resonant range. In the rest of the paper, $[\omega_L, \omega_U]$ is referred to as the NRFR query. The optimization problem can now be expressed as

$$\min_x \quad m(x) = \rho \times \sum_{i=1}^{10} \frac{\pi}{4}(d_i^2 - d_\triangleright)w_i$$
$$\text{s.t.} \quad h(x, \omega_L, \omega_U) = \sum_{\omega_1}^{\omega_2} \delta(x, \omega_i) = 0 \text{ where} \tag{8}$$

$$\delta(x, \omega_i) = \begin{cases} 1 & \text{if } T_R(x, \omega_i) \geq T_R(x, \omega_{i\pm1}) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

In the above equations, ρ is the density of the material of the pipe and its inserts, and $d_\triangleright$ is the inner diameter of the base pipe. Here, δ is a binary operator that signifies the existence of a peak in the transmissibility ratio at a given frequency. Analogical to a bracketing approach in line search, the first condition in Eq. (9), i.e., $T_R(x, \omega_i) \geq T_R(x, \omega_{i\pm1})$, when true indicates the existence of a peak within the $[\omega_{i-1}, \omega_{i+1}]$ range, which is designated as a peak at ω_i for ease of modeling and illustrating the resonant peaks. Figure 3(*b*) shows the design variables per insert to be optimized, and Table 2 gives the upper and lower bounds of these variables. Since we consider 10 rings, there are only three variables for the periodic structure (which involves uniformly spaced identical rings), i.e., $[d \quad w \quad t]$; and a total of 30 design variables in the aperiodic structure, i.e., $[d_1 \quad w_1 \quad t_1 \quad d_2 \quad w_2 \quad t_2 \quad \ldots \quad d_{10} \quad w_{10} \quad t_{10}]$.

3.3 Inverse and Forward (Approximation) Modeling. The INN model is constructed to map the design of the phononic structure to the frequency response of its lateral vibrations. More specifically, the vector of design variables (d_i, t_i, w_i) serve as inputs, and a vector of 80 transmissibility ratio values $(T_{R_i}, i = 1, 2, \ldots, 80)$ measured at equally spaced frequency intervals between 0.1 and 10,000 Hz are treated as the outputs. The forward mapping performed by the INN can be expressed as

$$f: \begin{bmatrix} \mathbf{D}^\mathrm{T} \\ \mathbf{W}^\mathrm{T} \\ \mathbf{T}^\mathrm{T} \end{bmatrix} \rightarrow \begin{bmatrix} T_{R_1} \\ \vdots \\ T_{R_{80}} \end{bmatrix} \tag{10}$$

where $\mathbf{D} = [d_1, d_2, \ldots, d_{10}]$, $\mathbf{W} = [w_1, w_2, \ldots, w_{10}]$, and $\mathbf{T} = [t_1, t_2, \ldots, t_{10}]$ are the diameter of each ring, their thickness, and the distance between two rings. The inverse modeling enabled by the INN is simply the reverse of the mapping shown in Eq. (10), i.e., given a sample output vector, it can retrieve a prediction of the corresponding input vector. Note that the design variables are repeated in order to meet the requirement of the INN to have equal sized input and output vectors as discussed earlier. Thus, for aperiodic structures, the input vector is given by $[\mathbf{D}, \mathbf{W}, \mathbf{T}, \mathbf{D}, \mathbf{W}, \mathbf{T}, \mathbf{D}, \mathbf{W}]^\mathrm{T}$. In the case of periodic structures, the three variable input vectors are repeated to construct the padded 80-sized input vector. For comparison purposes, standard DNNs with similar node activation type are also trained. The INN and DNN models for the periodic and aperiodic structures are trained using the same optimizer and hyper-parameter setup, which along with further details of networks are summarized in Table 3.

For the neural network models to be effective in forward prediction and inverse design thereof, it is critical for them to capture the resonant peaks in the frequency response. We use 80 bins to discretize the 0.1–10000 Hz frequency space, as an acceptable compromise between fidelity of representing the frequency response and keeping the overall cost of its TMM-based evaluation tractable. To encourage the INN to accurately capture the resonant peaks found using the bracketing type approach (given in Eq. (9)) over this discretized frequency space, we artificially boost the magnitude of transmissibility ratio at the resonant peaks. More specifically, while the T_R at the identified resonant peaks are observed to be roughly 2 orders of magnitude greater than the average base T_R in their neighborhood as computed by TMM, we artificially boost the magnitude of T_R at the peaks to a fixed high value of 10^{-2}. From our numerical experiments, we found this adjustment to improve the modeling accuracy of the INN and DNN, which are purposed with identifying the location of the peaks (their exact magnitude being of marginal relevance).

Table 3 Neural network settings

	Periodic		Aperiodic	
Parameter	INN	DNN	INN	DNN
Input size	80			
Output size	80			
Hidden layers	3	5	5	5
Invertible blocks	10	–	10	–
Type	Fully connected			
Activation func.	Leaky ReLU			
Learning rate	10^{-4}			
Optimizer	ADAM			
Hyper-param optim.	Bayesian optimization			
Training data size	5000			
Testing data size	1000			

The INNs used later for inverse design retrieval to initiate forward optimizations are trained with the loss function given by Eq. (3); this approach provided better inverse prediction accuracy. The use of latent variables and training with the loss function given by Eq. (6) was also explored. In that case, a weight setting of $\lambda = 10^{-2}$ was found to work best. Further discussion on this second training approach and associated outcomes are given in Appendix B. For training the baseline DNN-`F` model, only the first term of Eq. (3) is used as the loss function. For training the baseline DNN-`I` model, only the second term of Eq. (3) is used as the loss function.

3.4 Inverse Design Retrieval Using Invertible Neural Network. *Retrieving Single Design*: For the 1D phononic structure problem studied here, user query can be in practice expected to be given in terms of a NRFR, $[\omega_L, \omega_U]$, as expressed in Eq. (9). However, the INN is designed to output the entire frequency response, i.e., the vector of 80 transmissibility ratios $([T_{R_1}, T_{R_2}, \ldots, T_{R_{80}}])$; this output choice was made to partly alleviate the non-uniqueness of the inverse mapping without having to resort to latent parameters. Hence, the INN needs the entire T_R vector to be fed from the output end for the inverse prediction to work. For ease of results generation corresponding to each case study discussed later in Sec. 4.2, we therefore simply picked an output vector each from the test set that satisfies the $[\omega_L, \omega_U]$ query for each case. In practice, an entire frequency response, containing the query, could be provided by a domain expert, or in other applications might be readily available as the query itself. While the primary results shown here uses the above-stated approach to facilitate the inverse retrieval and the optimizations initialized thereof (for brevity of illustration), we have also developed a more generalizable procedure to translate the query to the entire output vector; this is summarized in the supplemental document,[3] along with example results following from that procedure.

[3]See Note 2.

Retrieving Multiple Designs: The explicit invertibility and bijectivity of the INN can be uniquely used to generate multiple diverse designs reliably in the close neighborhood of a user-specified target output. Since a measure of forward error of the INN is available based on training or cross-validation, this information can be used to determine a range or confidence interval of the output over which multiple inputs can be sampled. More specifically, we derive a set of sample queries from a normal distribution with the user-queried output as the mean and the training error as standard deviation.

These sampled queries are then fed into the output side of the trained INN, which is then executed in reverse to generate a set of sample input designs. We will use this capability later in the next section to initiate portion of the population of the particle swarm optimization (PSO) algorithm used to solve the inverse design problem given by Eqs. (8) and (9).

4 Results and Discussion

4.1 Modeling Performance of the Invertible Neural Network. We train two INNs each for the aperiodic and the periodic structure. In both cases, a training set of 5000 samples are generated using Latin hypercube sampling or LHS (with *maximin* criterion) with their function values evaluated using the TMM. The choice is based on numerical experiments with various training set sizes, details of which are provided in Appendix C. To compare the modeling accuracy of the two INNs, we also train two corresponding pairs of DNNs (DNN-F and DNN-I) on the same sample sets. The prediction performance of the models is tested on 1000 unseen samples each for the aperiodic and periodic structures, which are also generated using LHS. The training settings for the NN models are summarized in Table 3. Figure 4 illustrates the testing error for the INN in the forward and inverse directions and those of the corresponding DNN-F and DNN-I models for the periodic and aperiodic cases. The error is represented in terms of the MSE of the output vector per sample, shown as a distribution over the entire test data set. Here, the MSE over each (range-normalized) test sample is given by

$$\text{Forward:} \quad \text{MSE}_{\text{Fwd}}(X_i) = \frac{1}{80}\sum_{j=1}^{80}\left(\hat{Y}_{ij} - Y_{ij}\right)^2 \tag{11}$$

$$\text{Inverse:} \quad \text{MSE}_{\text{Inv}}(Y_i) = \frac{1}{80}\sum_{j=1}^{80}\left(\hat{X}_{ij} - X_{ij}\right)^2 \tag{12}$$

As seen from Fig. 4, for the periodic case, the forward prediction error of INN (median MSE_{Fwd} of 0.006) is slightly better than that of the DNN-F (median MSE_{Fwd} of 0.017), with the INN also providing better robustness (lower variance). For the aperiodic case, the forward prediction accuracy of INN and DNN-F are comparable. Unlike in the case of forward prediction, Fig. 4 shows that for inverse prediction, the model errors are noticeably greater in the aperiodic case compared to the periodic case. The better accuracy achieved in the periodic case could be attributed to the presence of distinct bandgaps in the corresponding periodic structures, and that the change in the output (T_R) is highly sensitive to changes in any input. The latter is because there are only three variables that regulate the shape of all the identical inserts in the periodic case, in contrast to 30 in the aperiodic case. For this inverse prediction, the INN performs better than the DNN-I for the periodic and aperiodic cases, in terms of median MSE_{Inv} values, as seen from Fig. 4.

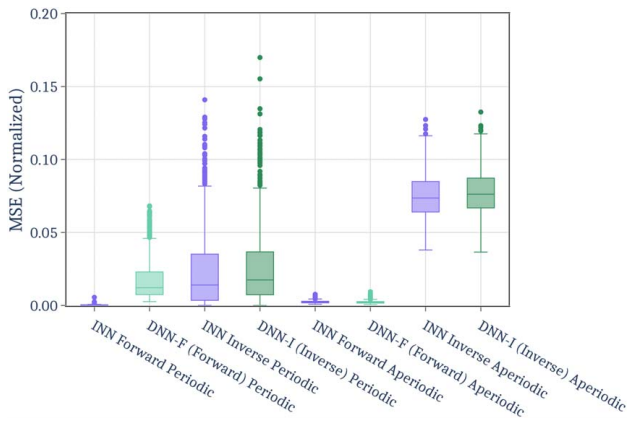

Fig. 4 Modeling error (MSE) of INNs and DNNs on test samples

Figure 5 illustrates the frequency response predicted by the INN and DNN-F models over a representative periodic and aperiodic test sample each, showing the predicted transmissibility ratio and the corresponding ground truth given by TMM. It can be seen that, in the periodic case (Fig. 5(*a*)), there is a large variance between the non-peak values, thereby favoring the model to learn the difference between a normal frequency range and a bandgap. In the aperiodic case (Fig. 5(*b*)), the difference between non-peak values is less distinct. Hence, both the INN and DNN models do better in the periodic case (compared to the aperiodic case), in terms of tracking the frequency response. In both cases, while the INN predicts a few spurious peaks in frequency ranges where none exist (e.g., between ~1.0 and 2.2 kHz in Fig. 5(*a*)), it does well in capturing almost all of the true peaks that do exist.

4.2 Optimization Initialized With Inverse Design Retrieval. *Inverse Design Case Studies*: Now the INN can be used to on-demand retrieve the phononic structure design corresponding to a (user-queried) frequency response. This is achieved by feeding the desired output (user query) to the INN, and running it backwards, as shown in Fig. 2*c*. Here, we explore the effectiveness of using such a retrieved design to initialize a forward optimization process for finding a more accurate inverse design that satisfies the user-requested properties as well as minimizes mass, i.e., solves Eqs. (8)–(9). The high-fidelity TMM model is used for all forward optimizations performed in this paper. We consider two cases, each connected with a distinct (desired) *NRFR* of a span of 1.5 kHz, which are listed in Table 4. For each case, the user query is first translated to an output frequency response, as discussed in Sec. 3.4. The trained INNs for the aperiodic and periodic structures are then executed in reverse to predict the aperiodic and periodic designs corresponding to that demanded frequency response.

Forward Optimization Approach: Two different optimization approaches are used here. Primarily we use a standard gradient-based solver, the sequential quadratic programming (SQP) [70], and further analyses is provided only for the results obtained with SQP. This SQP is implemented using MATLABS built-in optimization toolbox. Our choice of SQP is deliberate. We intended to eliminate any additional stochastic effects attributed to the optimizer (which would otherwise make further analyses challenging). However, as we will see later, the problem tends to be multi-modal. Hence, to provide insightful global comparisons and allow the use of good training samples (those with the smallest violation of the user's *NRFR* query) to create a more effective baseline, we also apply a well-known implementation of the particle swarm optimization (PSO) algorithm suited for constrained high-dimensional problems [71]. To account for its stochastic effects, the PSO algorithm is run 10 times each with the INN-based initialization and corresponding baseline (discussed later). The settings used for the SQP optimization and PSO solvers are given in Appendix D. Note that, this inverse design framework can readily incorporate other approaches for forward optimization, including genetic algorithms, Bayesian optimization, or surrogate-based optimization [72] with the INN used as the forward surrogate model.

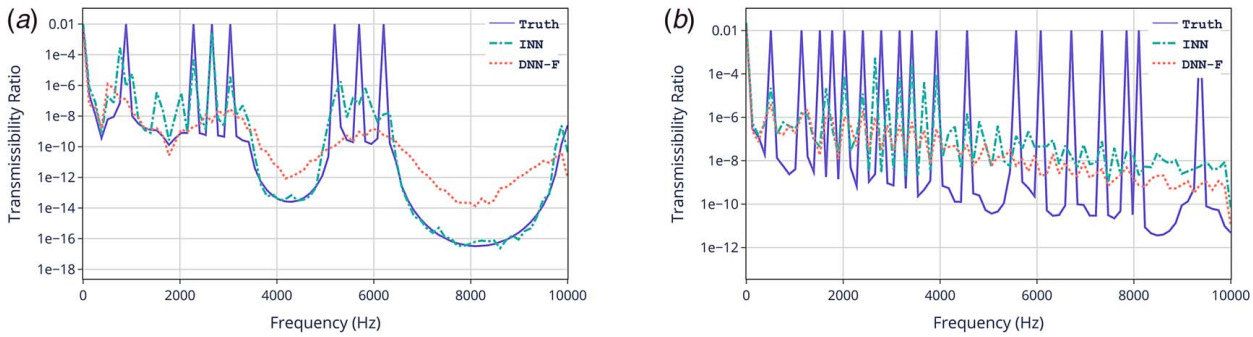

Fig. 5 Transmissibility ratios showing the frequency response of the phononic structure's vibrations as a representative example of INN forward prediction on (*a*) periodic design and (*b*) aperiodic design

Table 4 SQP optimization results for the inverse design of the 1D phononic structure given *NRFR* query—single run each with INN and DNN-*I* initializations, and best/median of 10 runs with random initialization

		Periodic: final solution		Aperiodic: final solution	
NRFR query ω_L to ω_U (Hz)	Optimization initialization[a]	Constraint violation	Mass (kg)	Constraint violation	Mass (kg)
4000–5500 Hz	Random, best[a]	**1** (infeasible)	**735.9**	2 (infeasible)	376.6
	Random, median[b]	3 (infeasible)	514.9	5 (infeasible)	181.3
	INN	**0**	**671.2**	2 (infeasible)	370.9
	DNN	0	921.8	2 (infeasible)	475.4
6500–8000 Hz	Random, best[a]	**0**	**566.2**	3 (infeasible)	433.2
	Random, median[b]	4 (infeasible)	232.3	4 (infeasible)	367.1
	INN	0	583.2	**3** (infeasible)	**246.9**
	DNN	0	691.14	3 (infeasible)	353.5

[a]Random, best: the best results, sorted in terms of feasibility and then mass, out of the 10 optimization runs that use random initialization.
[b]Random, median: the median results, sorted in terms of feasibility and then mass (among solutions with median feasibility), out of the 10 runs.

Baselines for Comparison: To conduct comparative analyses of the performance of the INN-initialized optimization approach to inverse design, we have created a suite of three baselines. The first two are used with SQP and are defined as: (i) *random initialization*: we run 10 SQP optimizations with 10 different random designs, each drawn from a random uniform distribution over the design space. (ii) *Inverse DNN (DNN-I) initialization*: we run a single SQP optimization with the initial point retrieved by executing the trained inverse DNN-I model. These baselines are compared with the outcomes of the SQP optimization initialized with a single retrieved design given by the INN for each user query case. The third baseline called *prior sample set initialization*, which is implemented with PSO, uses the following initial population: 10 designs are selected from the training sample set (used for training the INN and DNNs) based on zero or the smallest constraint violation with respect to the user query, which is combined with 50 random designs given by Sobol-sequence sampling used as default in the chosen PSO algorithm [71] for every run of the PSO algorithm. This baseline is compared with an INN-initialized PSO, where 10/60 of the initial points are generated through the INN as described in Sec. 3.4, and the remaining 50/60 initial points are generated randomly as in the above-stated baseline, both for every run of the PSO algorithm with this initialization. The optimum designs resulting from the baselines and the corresponding INN-initialized optimizations are compared mainly in terms of the constraint violation (i.e., how well they satisfy the user query) and secondarily in terms of mass (which is premised to matter only when comparing feasible designs). All optimization runs are terminated at 300 (TMM-based) function evaluations.

SQP Optimization Results: Table 4 lists the SQP optimization results for the two *NRFR* query cases. Figure 6 shows the constraint violation of the initial designs generated using the INN and the other baselines. From Fig. 6, it can be seen that both the INN and the DNN-I predict better initial points in the periodic space as compared to the aperiodic space. Their performances in retrieving inverse designs are comparable and are in general better than random initialization, as seen from Fig. 6. From Table 4, it is apparent that in terms of feasibility of the final solutions, i.e., whether the *NRFR* query is met or not, the proposed INN-initialized optimization approach performs much better than the median performing run of randomly initialized optimizations. For periodic structures, the median random initialization fails to provide feasible solutions in all cases, while the INN and DNN-I initialized optimizations lead to feasible solutions in all cases. Moreover, at similar feasibility

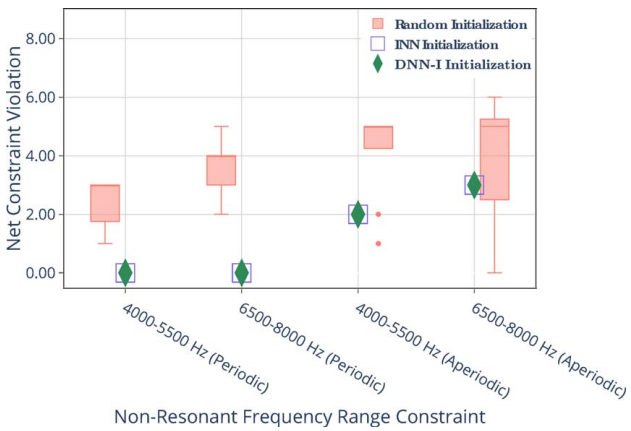

Fig. 6 Net constraint violation (i.e., number of resonant peaks in the user-specified *NRFR*) of the solution used to initialize SQP optimization

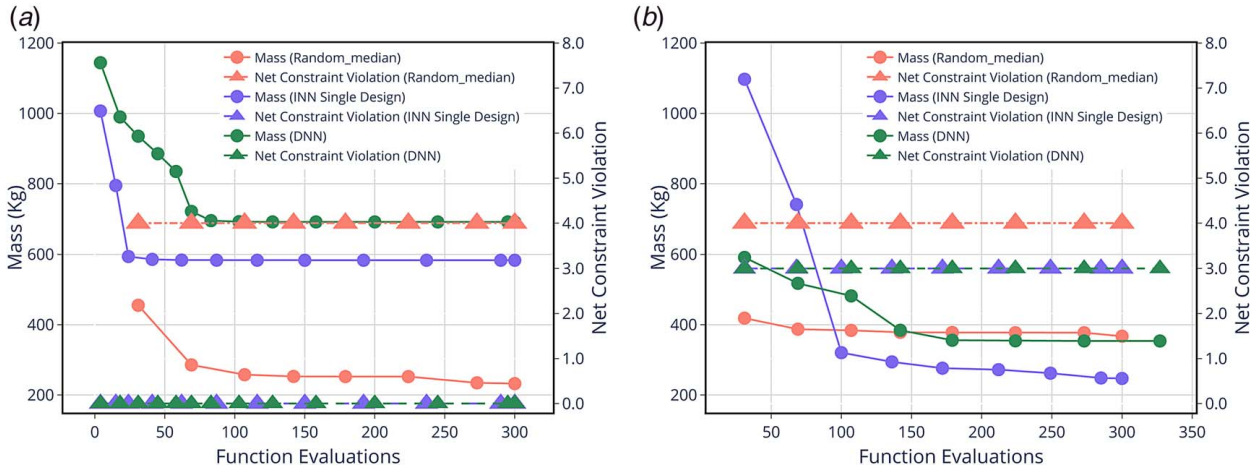

Fig. 7 Convergence history of SQP optimizations with different initialization for the *NRFR* query of 6.5–8.0 kHz: (*a*) periodic structure and (*b*) aperiodic structure

Table 5 PSO results (best/median of 10 runs) for the inverse design of the phononic structure given *NRFR* query

		Periodic: final solution				Aperiodic: final solution			
		Constraint violation		Mass		Constraint violation		Mass	
NRFR query ω_L to ω_U (Hz)	Optimization initialization	Median	Best	Median	Best	Median	Best	Median	Best
4000–5500 Hz	Prior sample set	1.0000	0.0000	625.3415	679.0066	0.0000	0.0000	356.2176	349.9246
	INN	1.0000	**0.0000**	530.4582	**674.0474**	0.0000	**0.0000**	365.1117	**346.8808**
6500–8000 Hz	Prior sample set	0.0000	0.0000	640.6702	634.1198	0.0000	0.0000	559.2052	557.6745
	INN	0.0000	**0.0000**	618.62902	**611.4982**	0.0000	**0.0000**	560.3241	**534.2124**

performance, the INN initialized optimizations lead to smaller mass values compared to those initialized by DNN-I, across all cases. In the aperiodic case, the best of random initialization runs perform better than the INN and DNN-I initialization in the 4.0–5.5 kHz range; the relatively poorer performance of both INN and DNN-I in aperiodic cases can be attributed to the poorer accuracy in retrieving inverse designs, as discussed earlier in Sec. 4.1. Figure 7 shows the convergence history of the SQP optimizations for the 6.5–8.0 kHz constraint case to elucidate how the INN and DNN-I initializations help in the subsequent optimization performance, compared to the median run of random initialization.

PSO Results: Table 5 summarizes the optimization results obtained in the best and median performing runs out of the 10 runs of the PSO algorithm, respectively, with the baseline and the INN initialization. Comparing the results of the best and median performing runs shows that the PSO provides robust results (i.e., low variance) in this application. In this case, while INN provided suitable initializations (with constraint violation of 0 or 1) in some runs, in some of the other runs one of the random initial points in the population was observed to have the lowest constraint violation. As a result, the performance of INN initialization and the baseline (with top training samples), both of which use 50/60 random initial points, are observed to be similar in terms of finding feasible solutions. However, as seen from Table 5, INN initializations result in optimal solutions with lower mass values compared to the baseline. It is also important to note that the PSO approach performs substantially better than the SQP approach (irrespective of the initialization) in finding feasible solutions for aperiodic structures.

4.3 Periodic Versus Aperiodic Designs. As seen from the results listed in Table 4 and convergence histories in Fig. 7, it is easier to find feasible *periodic* structures (i.e., with no resonance peaks) compared to feasible *aperiodic* structures, given the same *NRFR* query. This observation is also evident from the frequency response of the optimum designs obtained with INN-initialized optimization—with the optimized periodic design showing a large bandgap extending well beyond the queried *NRFR* range (Fig. 8), while the aperiodic design just about satisfies the constraint (Fig. 8). This is understandable given that bandgap formation in elastic structures at the finite level is rooted in multiplicity effects that require a large number of self-repeated cells, as described in Ref. [73]. While aperiodic designs have no such unique vibration transmission features, they are helpful in cases where the required non-resonant range is not very wide.

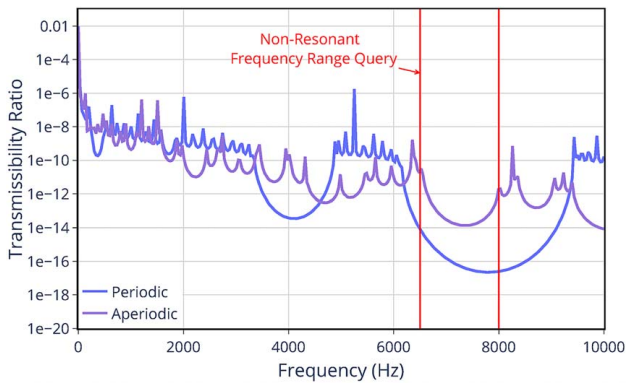

Fig. 8 INN-initialized SQP optimization results for the *NRFR* query of 6.5–8.0 kHz: frequency response (T_R versus freq.) of the periodic and aperiodic designs

On the other hand, by lieu of the flexibility inherent to the aperiodic structure, we are able to find aperiodic designs with significantly smaller mass than their periodic counterparts as shown in Table 5. While, in principle, the aperiodic structure space is a superset of the periodic structure space, better solutions are not always found with the generalized aperiodic problem definition. This is mainly because of the significantly increased computational cost of both INN modeling and forward optimization of aperiodic structures, attributed to their higher-dimensional design space (30 versus 3 in our case studies). Note that, in this paper, both the periodic and aperiodic cases were allowed similarly sized data set for INN training, and the same number of maximum function evaluations during optimization, thus limiting the achievable performance with the aperiodic case.

5 Concluding Remarks

In this paper, we studied the application of INNs as a nascent machine learning tool to enable on-demand inverse design of phononic structures. The specific type of INN architecture exploited here allows guaranteed invertibility, i.e., $f_{\text{Inv}}(f_{\text{Fwd}}(x)) = x$, with the backward (inverse) mapping learned for free when the INN is trained to perform forward prediction. For demonstration purposes, a 1D phononic structure embodying drill pipes with annular rings was considered, where the inverse design goal is to mitigate resonance peaks in lateral vibrations, over user-queried (demanded) frequency ranges. To further highlight the effectiveness of INN in modeling large-sized design-feature spaces, we considered an aperiodic structure involving 30 design variables, in addition to the typical periodic structure with a self-repeating ring pattern (involving three design variables). We employed the TMM, which is an analytical approach to model the dynamic behavior of phononic structures and locally resonant metamaterials, to compute the frequency response of the set of designs used for training and testing the INN models.

The INN model provides promising accuracy in forward and inverse prediction, especially in comparison to corresponding DNN models trained with similar settings. The inverse prediction errors, with a median MSE of 0.07 over unseen normalized test samples, while slightly higher than the forward prediction errors, are acceptable—as later demonstrated through the INNs' effective usage in inverse design retrieval. In this first-of-its-kind application of INN, we used the entire frequency response (represented by a vector of transmissibility ratios) as the output to facilitate reduced non-uniqueness of the inverse mapping. To directly model more compact and pragmatic property specifications, e.g., specified bandgaps, and allow additional conditions (e.g., mass), future work will explore the concept of latent variables used in training the INN, such that the quality of the inverse mapping is still preserved. How the use of latent variables might affect the cost and stability of the INN training process however remains an open question. For example, in problems where the output-to-input mapping is non-unique, we recommend the use of latent variables in the network and MMD in the loss function, as opposed to directly using the MSE in the reverse direction. Now while the latter could lead to mode collapse and oscillations during training, the former often compromises the accuracy of the inverse predictions. Thus, better trade-offs combining MMD and MSE in the reverse direction must be explored in the future, especially for problems presenting many-to-one *X*-*Y* data.

Our inverse design framework also integrates the INN model and the inverse DNN model with forward optimization in a manner where the INN (or the inverse DNN) can be used to instantaneously retrieve an approximate inverse design solution given a frequency response containing a queried non-resonant range. This initial point is then used to start a SQP-based optimization process that implements the user query as a constraint and minimizes the structural mass. Over two different NRFR specifications, we show that INN provides effective initialization for the optimization process especially for periodic structures, while the higher inverse prediction errors in the case of aperiodic structures continue to pose barriers to the search of feasible solutions. Moreover, INN performance is found to be in general superior in terms of feasibility (i.e., the absence of resonance peaks in the specified range) and mass compared to optimized designs respectively generated with random initialization and the inverse DNN model. A second optimization implementation based on PSO was found to provide even better results (e.g., feasible solutions in all aperiodic structure cases), when initialized with multiple designs generated via INN, applied in reverse on the specified *NRFR* query with added noise. In these inverse design retrieval contexts, note that every user query is not guaranteed to have a feasible solution within the given problem setup, a common issue in inverse retrieval. With such an infeasible query, our INN is expected to predict a design that gives relatively small violations of the queried property. Therefore, exploring new techniques to identify feasible/infeasible queries is an important direction of further research for the inverse design community.

Here, we have used a fairly straightforward optimization approach, in order not to dilute the main contribution of this paper, which lies in demonstrating the construction and use of INN for inverse design retrieval. Nevertheless, in future, it would be critical to explore more creative integration of INN and multi-fidelity optimization, e.g., where the INN can also be both used for forward evaluations (like a metamodel) during optimization, and refined in situ for intermediate inverse retrieval as and when needed—this would be to further accelerate the overall inverse design process. In addition, while this paper presented initial evidence of the effectiveness of INNs in the inverse design of large-featured aperiodic structures, application to more complex problems such as 2D and 3D aperiodic metamaterial systems (and comparison with other emerging inverse modeling techniques such as GANs) are needed to further elucidate the usability and limitations of the INN-based approach.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The data and information that support the findings of this article are freely available.[4]

Appendix A: Maximum Mean Discrepancy in Invertible Neural Network

The loss function over which INN is trained using the MMD term to facilitate the accuracy of the reverse mapping. This MMD can be expressed as

$$MMD(p(z)\|q(z)) = \mathbb{E}_{p(z),p(z')}\left[k\left(z, z'\right)\right] + \mathbb{E}_{q(z),q(z')}\left[k\left(z, z'\right)\right] - 2\mathbb{E}_{p(z),q(z')}\left[k\left(z, z'\right)\right] \tag{A1}$$

where $p(z)$ and $q(z)$ are the two distributions to be compared, $\mathbb{E}$ is the expectation operator, and $k(z, z')$ can be any universal kernel. In this paper, we use the inverse multiquadratic kernel, $k(x, x') = 1/(1 + \|(x - x')/h\|_2^2)$.

Appendix B: Invertible Neural Network With Latent Parameters and Maximum Mean Discrepancy Loss Term

Effect of Varying λ on INN Performance: Here, we use the loss function given by Eq. (6). To study the impact MMD has on the

[4]See Note 2.

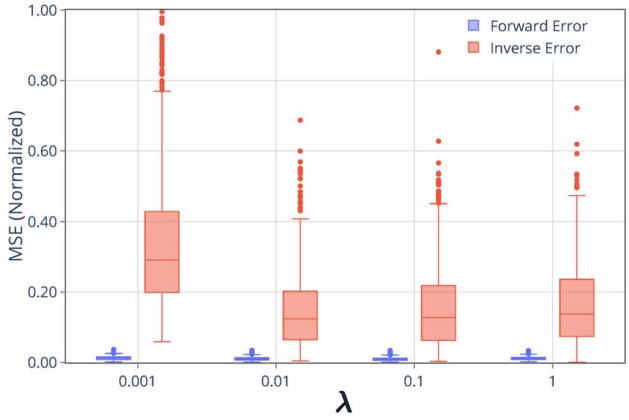

Fig. 9 Forward modeling error of INN with latent variables on periodic structure data, with different values of λ (i.e., the weight of MMD in the overall loss function used to train the INN)

forward and inverse prediction errors of the INN, we trained the INN using different values of λ on the periodic structure samples, and the corresponding testing errors (MSE) on unseen 300 samples are presented in Fig. 9. In this case, we also use 10 latent variables, and as the number of outputs is greater than the number of inputs, we replicated the information in the input layer to match the sum of output dimensions and 10 latent variables. From the testing results, it can be seen that the error decreases when λ is increased from 0.001 to 0.01, but then increases when the value of λ is increased beyond 0.01. This variation might be because the higher weightage of the inverse MMD loss overshadows the direct forward loss function, and restricts the model to suboptimal configurations. On the other hand, having a low weight for the MMD loss leads to the model being free to learn routines which map out-of-domain inputs to outputs in the current domain, leading to a loss in the inverse performance.

Appendix C: Identifying Training Data Set Size

Figure 10 shows the change in the forward and inverse prediction accuracy of the INN, in terms of MSE, when the size of the training data on periodic samples is varied. In all of these cases, the training process is terminated at 30 iterations for time tractability and ease of comparison. The accuracy of the trained INN is measured on a test data set of 1000 samples. From Fig. 10, it can be observed that a training data set of 5000 samples provides a reasonable balance between accuracy and computational cost (the latter, as expected, scales linearly with the number of samples).

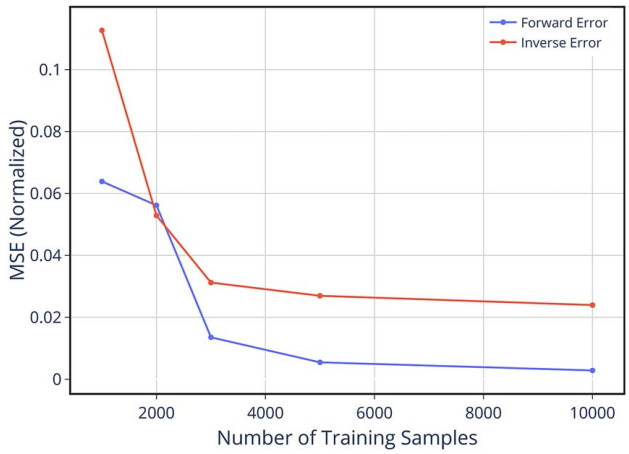

Fig. 10 Variation of INN testing accuracy with the size of the training data set

Table 6 Optimization settings

Parameter	Value
PSO[a]	
Population size	60
Maximum iterations	5
Maximum function evaluations	300
INN init. population	50
Random population	10
SQP[b]	
Maximum function evaluations	300
Optimality tolerance	10^{-6}

[a]Other parameters for PSO are set at default values reported in Ref. [71].
[b]Other parameters for SQP are set at default values used by the MATLAB fmincon solver.

Appendix D: Optimization Settings

Some of the high-level settings used in implementing the SQP (under MATLAB's in-built optimization toolbox) and PSO [71] in this paper are given in Table 6. Remaining settings are kept at their default values.

References

[1] Tanaka, M., and Dulikravich, G. S., 1998, *Inverse Problems in Engineering Mechanics*, Elsevier.
[2] Czech, C., Guarneri, P., Thyagaraja, N., and Fadel, G., 2015, "Systematic Design Optimization of the Metamaterial Shear Beam of a Nonpneumatic Wheel for Low Rolling Resistance," ASME J. Mech. Des., **137**(4), p. 041404.
[3] Ma, W., Cheng, F., and Liu, Y., 2018, "Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials," ACS Nano., **12**(6), pp. 6326–6334.
[4] Delissen, A., Radaelli, G., Shaw, L., Hopkins, J., and Herder, J., 2018, "Design of an Isotropic Metamaterial With Constant Stiffness and Zero Poisson's Ratio Over Large Deformations," ASME J. Mech. Des., **140**(11), p. 111405.
[5] Chen, Y., Zhu, J., Xie, Y., Feng, N., and Liu, Q. H., 2019, "Smart Inverse Design of Graphene-Based Photonic Metamaterials by an Adaptive Artificial Neural Network," Nanoscale, **11**(19), pp. 9749–9755.
[6] Liu, D., Tan, Y., Khoram, E., and Yu, Z., 2018, "Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures," ACS Photonics, **5**(4), pp. 1365–1369.
[7] Matthews, J., Klatt, T., Morris, C., Seepersad, C. C., Haberman, M., and Shahan, D., 2016, "Hierarchical Design of Negative Stiffness Metamaterials Using a Bayesian Network Classifier," ASME J. Mech. Des., **138**(4), p. 041404.
[8] Sanchez-Lengeling, B., Outeiral, C., Guimaraes, G. L., and Aspuru-Guzik, A., 2017, "Optimizing Distributions Over Molecular Space. An Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry (Organic)." ChemRxiv, 2017.
[9] Zunger, A., 2018, "Inverse Design in Search of Materials With Target Functionalities," Nat. Rev. Chem., **2**(4), pp. 1–16.
[10] Molesky, S., Lin, Z., Piggott, A. Y., Jin, W., Vucković, J., and Rodriguez, A. W., 2018, "Inverse Design in Nanophotonics," Nat. Photonics, **12**(11), pp. 659–670.
[11] Liang, D., Cheng, J., Ke, Z., and Ying, L., 2020, "Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks," IEEE Signal Process. Mag., **37**(1), pp. 141–151.
[12] Sanchez-Lengeling, B., and Aspuru-Guzik, A., 2018, "Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering," Science, **361**(6400), pp. 360–365.
[13] Long, Y., Ren, J., Li, Y., and Chen, H., 2019, "Inverse Design of Photonic Topological State Via Machine Learning," Appl. Phys. Lett., **114**(18), p. 181105.
[14] Mead, D., 1996, "Wave Propagation in Continuous Periodic Structures: Research Contributions From Southampton, 1964–1995," J. Sound. Vib., **190**(3), pp. 495–524.
[15] Hussein, M. I., Hamza, K., Hulbert, G. M., Scott, R. A., and Saitou, K., 2006, "Multiobjective Evolutionary Optimization of Periodic Layered Materials for Desired Wave Dispersion Characteristics," Struct. Multidiscipl. Optim., **31**(1), pp. 60–75.
[16] Al Ba'ba'a, H., Nouh, M., and Singh, T., 2017, "Formation of Local Resonance Band Gaps in Finite Acoustic Metamaterials: A Closed-Form Transfer Function Model," J. Sound. Vib., **410**(0022-460X), pp. 429–446.
[17] Callanan, J., and Nouh, M., 2019, "Emergence of Pseudo-Phononic Gaps in Periodically Architected Pendulums," Front. Mater., **6**, p. 119.

[18] Chen, Y., Barnhart, M., Chen, J., Hu, G., Sun, C., and Huang, G., 2016, "Dissipative Elastic Metamaterials for Broadband Wave Mitigation at Subwavelength Scale," Composite Structures, **136**, pp. 358–371.
[19] Aladwani, A., and Nouh, M., 2021, "Strategic Damping Placement in Viscoelastic Bandgap Structures: Dissecting the Metadamping Phenomenon in Multiresonator Metamaterials," ASME J. Appl. Mech., **88**(2), p. 021003.
[20] Chronopoulos, D., Antoniadis, I., and Ampatzidis, T., 2017, "Enhanced Acoustic Insulation Properties of Composite Metamaterials Having Embedded Negative Stiffness Inclusions," Extreme Mechanics Letters, **12**, pp. 48–54.
[21] Attarzadeh, M., Callanan, J., and Nouh, M., 2020, "Experimental Observation of Nonreciprocal Waves in a Resonant Metamaterial Beam," Phys. Rev. Appl., **13**(2), p. 021001.
[22] Maruyama, T., Furuno, T., Oda, Y., Shen, J., and Ohya, T., 2010, "Analysis and Design of Metamaterial Reflectarray Using Combination of Multilayer Mushroom-Structure," 2010 IEEE Antennas and Propagation Society International Symposium, Toronto, ON, Canada, July 11–17, IEEE, pp. 1–4.
[23] Bückmann, T., Stenger, N., Kadic, M., Kaschke, J., Frölich, A., Kennerknecht, T., Eberl, C., Thiel, M., and Wegener, M., 2012, "Tailored 3d Mechanical Metamaterials Made by Dip-in Direct-Laser-Writing Optical Lithography," Adv. Mater., **24**(20), pp. 2710–2714.
[24] Yoon, G., Kim, I., and Rho, J., 2016, "Challenges in Fabrication Towards Realization of Practical Metamaterials," Microelectronic Engineering, **163**, pp. 7–20.
[25] Sui, N., Yan, X., Huang, T.-Y., Xu, J., Yuan, F.-G., and Jing, Y., 2015, "A Lightweight Yet Sound-Proof Honeycomb Acoustic Metamaterial," Appl. Phys. Lett., **106**(17), p. 171905.
[26] Mamaghani, A. E., Khadem, S., and Bab, S., 2016, "Vibration Control of a Pipe Conveying Fluid Under External Periodic Excitation Using a Nonlinear Energy Sink," Nonlinear Dyn., **86**(3), pp. 1761–1795.
[27] Liu, N., Guo, H., Fu, L., Kaiser, S., Schweizer, H., and Giessen, H., 2008, "Three-Dimensional Photonic Metamaterials at Optical Frequencies," Nat. Mater., **7**(1), pp. 31–37.
[28] Gansel, J. K., Thiel, M., Rill, M. S., Decker, M., Bade, K., Saile, V., von Freymann, G., Linden, S., and Wegener, M., 2009, "Gold Helix Photonic Metamaterial as Broadband Circular Polarizer," Science, **325**(5947), pp. 1513–1515.
[29] Chen, T., Li, S., and Sun, H., 2012, "Metamaterials Application in Sensing," Sensors, **12**(3), pp. 2742–2765.
[30] Al Ba'ba'a, H., and Nouh, M., 2019, "Control of Spatial Wave Profiles in Finite Lattices of Repelling Magnets," ASME J. Dyn. Syst. Meas. Control, **141**(11), p. 111015.
[31] Amezquita-Sanchez, J. P., Dominguez-Gonzalez, A., Sedaghati, R., de Jesus Romero-Troncoso, R., and Osornio-Rios, R. A., 2014, "Vibration Control on Smart Civil Structures: A Review," Mech. Adv. Mater. Struct., **21**(1), pp. 23–38.
[32] Alkhatib, R., and Golnaraghi, M., 2003, "Active Structural Vibration Control: A Review," Shock Vib. Dig., **35**(5), p. 367.
[33] Elias, S., and Matsagar, V., 2017, "Research Developments in Vibration Control of Structures Using Passive Tuned Mass Dampers," Annual Reviews in Control, **44**, pp. 129–156.
[34] Balaji, P., and SelvaKumar, K. K., 2020, "Applications of Nonlinearity in Passive Vibration Control: A Review," J. Vib. Eng. Technol., **9**(2), pp. 183–213.
[35] Delpero, T., Schoenwald, S., Zemp, A., and Bergamini, A., 2016, "Structural Engineering of Three-Dimensional Phononic Crystals," J. Sound Vib., **363**, pp. 156–165.
[36] Borneman, J., Chen, K.-P., Kildishev, A., and Shalaev, V., 2009, "Simplified Model for Periodic Nanoantennae: Linear Model and Inverse Design," Opt. Exp., **17**(14), pp. 11607–11617.
[37] Reinke, C. M., Teofilo, M., Su, M. F., Sinclair, M. B., and El-Kady, I., 2011, "Group-Theory Approach to Tailored Electromagnetic Properties of Metamaterials: An Inverse-Problem Solution," Phys. Rev. E, **83**(6), p. 066603.
[38] Yang, Q., Chen, X., Li, Y., Zhang, X., Xu, Y., Tian, Z., Ouyang, C., Gu, J., Han, J., and Zhang, W., 2017, "Aperiodic-Metamaterial-Based Absorber," APL Mater., **5**(9), p. 096107.
[39] Sakurai, A., Yada, K., Simomura, T., Ju, S., Kashiwagi, M., Okada, H., Nagao, T., Tsuda, K., and Shiomi, J., 2019, "Ultranarrow-Band Wavelength-Selective Thermal Emission With Aperiodic Multilayered Metamaterials Designed by Bayesian Optimization," ACS Central Sci., **5**(2), pp. 319–326.
[40] D'Alessandro, L., Krushynska, A. O., Ardito, R., Pugno, N. M., and Corigliano, A., 2020, "A Design Strategy to Match the Band Gap of Periodic and Aperiodic Metamaterials," Sci. Rep., **10**(1), pp. 1–13.
[41] So, S., Badloe, T., Noh, J., Bravo-Abad, J., and Rho, J., 2020, "Deep Learning Enabled Inverse Design in Nanophotonics," Nanophotonics, **9**(5), pp. 1041–1057.
[42] Malkiel, I., Nagler, A., Mrejen, M., Arieli, U., Wolf, L., and Suchowski, H., 2017, "Deep Learning for Design and Retrieval of Nano-Photonic Structures," preprint arXiv:1702.07949.
[43] Malkiel, I., Mrejen, M., Nagler, A., Arieli, U., Wolf, L., and Suchowski, H., 2018, "Deep Learning for the Design of Nano-Photonic Structures," 2018 IEEE International Conference on Computational Photography (ICCP), Pittsburgh, PA, May 4–6, IEEE, pp. 1–14.
[44] Malkiel, I., Mrejen, M., Nagler, A., Arieli, U., Wolf, L., and Suchowski, H., 2018, "Plasmonic Nanostructure Design and Characterization Via Deep Learning," Light: Sci. Appl., **7**(1), pp. 1–8.
[45] Liu, Z., Zhu, D., Rodrigues, S. P., Lee, K.-T., and Cai, W., 2018, "Generative Model for the Inverse Design of Metasurfaces," Nano Lett., **18**(10), pp. 6570–6576.
[46] Mao, Y., He, Q., and Zhao, X., 2020, "Designing Complex Architectured Materials With Generative Adversarial Networks," Sci. Adv., **6**(17), p. eaaz4169.
[47] Hodge, J. A., Mishra, K. V., and Zaghloul, A. I., 2019, "RF Metasurface Array Design Using Deep Convolutional Generative Adversarial Networks," 2019 IEEE International Symposium on Phased Array System & Technology (PAST), Waltham, MA, Oct. 15–18, IEEE, pp. 1–6.
[48] Jiang, J., Sell, D., Hoyer, S., Hickey, J., Yang, J., and Fan, J. A., 2019, "Free-Form Diffractive Metagrating Design Based on Generative Adversarial Networks," ACS Nano, **13**(8), pp. 8872–8878.
[49] Ma, W., Cheng, F., Xu, Y., Wen, Q., and Liu, Y., 2019, "Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model With Semi-Supervised Learning Strategy," Adv. Mater., **31**(35), p. 1901111.
[50] Chan, Y.-C., Ahmed, F., Wang, L., and Chen, W., 2021, "Metaset: Exploring Shape and Property Spaces for Data-Driven Metamaterials Design," ASME J. Mech. Des., **143**(3), p. 031707.
[51] Morris, C., Bekker, L., Haberman, M. R., and Seepersad, C. C., 2018, "Design Exploration of Reliably Manufacturable Materials and Structures With Applications to Negative Stiffness Metamaterials and Microstereolithography," ASME J. Mech. Des., **140**(11), p. 041251.
[52] Bostanabad, R., Chan, Y.-C., Wang, L., Zhu, P., and Chen, W., 2019, "Globally Approximate Gaussian Processes for Big Data With Application to Data-Driven Metamaterials Design," ASME J. Mech. Des., **141**(11), p. 111402.
[53] Ma, W., Liu, Z., Kudyshev, Z. A., Boltasseva, A., Cai, W., and Liu, Y., 2020, "Deep Learning for the Design of Photonic Structures," Nat. Photonics, **15**(2), pp. 77–90.
[54] Finol, D., Lu, Y., Mahadevan, V., and Srivastava, A., 2019, "Deep Convolutional Neural Networks for Eigenvalue Problems in Mechanics," Int. J. Numer. Methods Eng., **118**(5), pp. 258–275.
[55] Hodge, J. A., Mishra, K. V., and Zaghloul, A. I., 2019, "Joint Multi-Layer Gan-Based Design of Tensorial RF Metasurfaces," 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP), Pittsburgh, PA, Oct. 13–16, IEEE, pp. 1–6.
[56] Yilmaz, E., and German, B., 2020, "Conditional Generative Adversarial Network Framework for Airfoil Inverse Design," AIAA Aviation 2020 Forum, Virtual, June 15–19, p. 3185.
[57] Nobari, A. H., Chen, W., and Ahmed, F., 2021, "Range-Gan: Range-Constrained Generative Adversarial Network for Conditioned Design Synthesis." preprint arXiv:2103.06230.
[58] Nobari, A. H., Chen, W., and Ahmed, F., 2021, "Pcdgan: A Continuous Conditional Diverse Generative Adversarial Network for Inverse Design." preprint arXiv:2106.03620.
[59] Behjat, A., Oddiraju, M., Attarzadeh, M. A., Nouh, M., and Chowdhury, S., 2020, "Metamodel Based Forward and Inverse Design for Passive Vibration Suppression," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 84010, Virtual, Aug. 17–19, American Society of Mechanical Engineers, p. V11BT11A024.
[60] Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., Maier-Hein, L., Rother, C., and Köthe, U., 2018, "Analyzing Inverse Problems With Invertible Neural Networks." preprint arXiv:1808.04730.
[61] McCann, M. T., Jin, K. H., and Unser, M., 2017, "Convolutional Neural Networks for Inverse Problems in Imaging: A Review," IEEE Signal Process. Mag., **34**(6), pp. 85–95.
[62] Denker, A., Schmidt, M., Leuschner, J., Maass, P., and Behrmann, J., 2020, "Conditional Normalizing Flows for Low-Dose Computed Tomography Image Reconstruction." preprint arXiv:2006.06270.
[63] Xiao, M., Zheng, S., Liu, C., Wang, Y., He, D., Ke, G., Bian, J., Lin, Z., and Liu, T.-Y., 2020, "Invertible Image Rescaling," European Conference on Computer Vision, Virtual, Aug. 23–28, Springer, pp. 126–144.
[64] Adler, T. J., Ardizzone, L., Vemuri, A., Ayala, L., Gröhl, J., Kirchner, T., Wirkert, S., Kruse, J., Rother, C., Köthe, U., and Maier-Hein , L., 2019, "Uncertainty-Aware Performance Assessment of Optical Imaging Modalities With Invertible Neural Networks," Int. J. Comput. Assist. Radiol. Surg., **14**(6), pp. 997–1007.
[65] Song, Y., Meng, C., and Ermon, S., 2019, "Mintnet: Building Invertible Neural Networks With Masked Convolutions." preprint arXiv:1907.07945.
[66] Al Ba'ba'a, H., and Nouh, M., 2017, "An Investigation of Vibrational Power Flow in One-Dimensional Dissipative Phononic Structures," ASME J. Vib. Acoust., **139**(2), p. 021003.
[67] Han, L., Zhang, Y., Ni, Z.-Q., Zhang, Z.-M., and Jiang, L.-H., 2012, "A Modified Transfer Matrix Method for the Study of the Bending Vibration Band Structure in Phononic Crystal Euler Beams," Phys. B, **407**(23), pp. 4579–4583.
[68] Alsaffar, Y., Sassi, S., and Baz, A., 2018, "Band Gap Characteristics of Nonrotating Passive Periodic Drill String," J. Vib. Acoust., **140**(2), p. 021004.
[69] Sassi, S., Renno, J., Zhou, H., and Baz, A., 2021, "Experimental Investigation of the Vibration Control of Nonrotating Periodic Drill Strings," ASME J. Vib. Acoust., **143**(6), p. 061004.
[70] Kelley, C. T., 1999, *Iterative Methods for Optimization*, SIAM.
[71] Chowdhury, S., Tong, W., Messac, A., and Zhang, J., 2013, "A Mixed-Discrete Particle Swarm Optimization Algorithm With Explicit Diversity-Preservation," Struct. Multidiscipl. Optim., **47**(3), pp. 367–388.
[72] Ghassemi, P., Mehmani, A., and Chowdhury, S., 2020, "Adaptive In Situ Model Refinement for Surrogate-Augmented Population-Based Optimization," Struct. Multidiscipl. Optim., **62**, pp. 2011–2034.
[73] Al Ba'ba'a, H., Nouh, M., and Singh, T., 2017, "Pole Distribution in Finite Phononic Crystals: Understanding Bragg-Effects Through Closed-Form System Dynamics," J. Acoust. Soc. Am., **142**(3), pp. 1399–1412.