



PAPER

OPEN ACCESS

RECEIVED
13 July 2023REVISED
16 February 2024ACCEPTED FOR PUBLICATION
14 March 2024PUBLISHED
3 April 2024

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Generative adversarial networks for data-scarce radiative heat transfer applications

J J García-Esteban , J C Cuevas and J Bravo-Abad*

Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, E-28049 Madrid, Spain

* Author to whom any correspondence should be addressed.

E-mail: jorge.bravo@uam.es**Keywords:** generative adversarial networks, near-field radiative heat transfer, spectral applications, data scarcity

Abstract

Generative adversarial networks (GANs) are one of the most robust and versatile techniques in the field of generative artificial intelligence. In this work, we report on an application of GANs in the domain of synthetic spectral data generation for data-scarce radiative heat transfer applications, an area where their use has not been previously reported. We demonstrate the proposed approach by applying it to an illustrative problem within the realm of near-field radiative heat transfer involving a multilayered hyperbolic metamaterial. We find that a successful generation of spectral data requires two modifications to conventional GANs: (i) the introduction of Wasserstein GANs (WGANs) to avoid mode collapse, and, (ii) the conditioning of WGANs to obtain accurate labels for the generated data. We show that a simple feed-forward neural network (FFNN), when augmented with data generated by a CWGAN, enhances significantly its performance under conditions of limited data availability. In addition, we show that CWGANs can act as a surrogate model with improved performance in the low-data regime with respect to simple FFNNs. Overall, this work contributes to highlight the potential of generative machine learning algorithms in scientific applications beyond image generation and optimization.

1. Introduction

Machine learning, a rapidly expanding area within computer science, focuses on advancing the foundations and technology that enables machines to learn from data [1–4]. Deep learning, on the other hand, is a subset of machine learning techniques that uses artificial neural networks (ANNs) to model and solve complex data-driven problems, such as image and speech recognition [5, 6], natural language processing [7], and autonomous driving [8], among many others. With the current explosion of data and the rapid development of improved hardware and algorithms, machine learning and deep learning are becoming crucial tools in many industries, including healthcare, finance, and manufacturing.

Motivated by this success, machine learning and deep learning techniques are attracting increasing attention from a variety of scientific disciplines beyond computer science, revolutionizing traditional approaches to the modeling and analysis of data-driven scientific problems. In physics, these techniques are employed to tackle complex problems [9], including the representation of quantum many-body wave functions [10], the discovery and identification of phase transitions in condensed-matter systems [11–13], the solution of statistical problems [14], the development of novel quantum information technologies [15], the modeling of gravitational waves [16], and the design of nanophotonic devices with novel or improved functionalities [17–20]. Machine learning algorithms have been effectively used for the accelerated discovery and design of new materials and molecules in the fields of materials science and chemistry [21, 22], being instrumental in molecular dynamics simulations [23], in predicting chemical reactions [24], and in modeling the quantum mechanical energies of molecules [25]. In the realm of biology, the applications of machine learning and deep learning are also vast, including breakthroughs in gene expression prediction tasks, prediction of micro-RNAs targets, and novel single-cell methods [26]. Overall, despite significant

challenges, such as the interpretability and transferability, machine learning and deep learning are playing an increasingly pivotal role in the advance of a broad variety of scientific research methodologies.

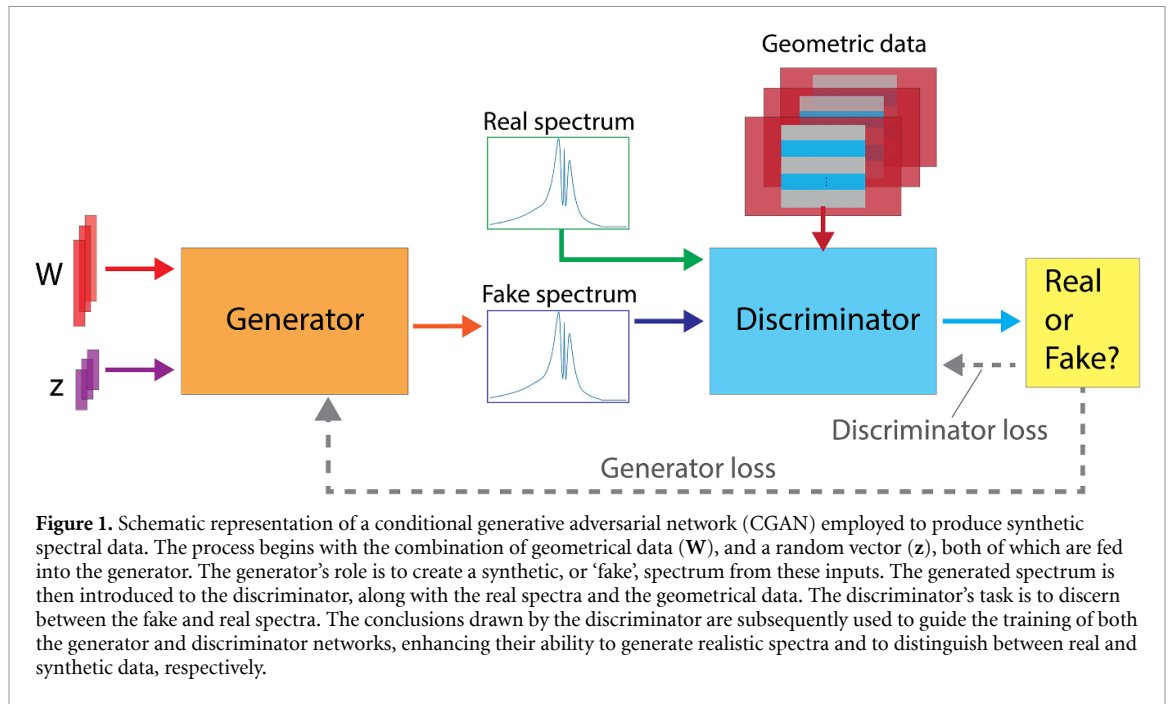
In this context, generative adversarial networks (GANs), a powerful subset of generative machine learning, have emerged as a versatile tool for creating new data instances that closely resemble a given training set [27]. This innovative paradigm involves a two-player adversarial setup where a generative ANN strives to produce data instances that are indistinguishable from the training set, while a discriminative network attempts to distinguish between the instances generated by the generative network and the real data [28]. The competing nature of these two networks drives the generative network to generate increasingly realistic data, pushing the boundaries of what is achievable with generative models. This technology has been instrumental across a broad range of scientific disciplines, including physics, chemistry, and biology. In physics, GANs have been used for simulating complex systems and predicting outcomes of experiments, with examples in high energy physics [29], condensed matter physics [30–32], nanophotonics [33, 34], and cosmology [35]. In the field of chemistry, GANs have been harnessed to generate novel chemical structures and predict their properties [36], thereby accelerating the process of drug discovery and materials design [37]. In biology, GANs have been employed in a variety of tasks, including protein engineering [38] and generate biological imaging data [39]. This myriad of applications demonstrate the significant potential of GANs in transforming scientific research by providing a powerful tool for hypothesis generation, experimental design, and data augmentation where empirical data is scarce or expensive to obtain. Despite their significant success in image generation, the use of GANs has been mostly limited to this area. It would be highly desirable to expand their application for the generation of scientific numerical data. In this regard, recent efforts have been devoted to adapt GANs for the generation of synthetic spectral data [40–43], but the reported methods are often not well suited for application in many scientific areas.

In this work, we introduce a novel application of GANs for synthetic spectral data generation for data-scarce radiative heat transfer applications. We particularly focus on an illustrative problem in the research area of near-field radiative heat transfer, involving a multilayered hyperbolic metamaterial. We explore the use of a Conditional Wasserstein GAN (CWGAN) for data augmentation and investigate its impact on the predictive capabilities of a feed-forward neural network (FFNN). We find that the successful production of spectral data requires two main changes to standard GANs. Firstly, the implementation of Wasserstein GANs (WGANs) is necessary to counteract mode collapse, and secondly, these WGANs need to be conditioned to yield accurate labels for the generated data. We demonstrate that a simple FFNN, when augmented with data produced by a CWGAN, notably improves its performance under conditions of data scarcity. Furthermore, we illustrate that CWGANs have the ability to serve as efficient surrogate model in low-data regimes. Overall, this research work contributes to highlighting generative AI algorithms' potential in applications extending beyond the conventional realm of image generation.

This work is organized as follows. In section 2, we review the fundamentals of the primary generative adversarial frameworks underlying this work. Section 3 discusses the basic principles of the particular physical problem we utilize to exemplify our approach. In section 4, we present and discuss the results obtained from implementing the generative adversarial method detailed in section 2 to the specific example problem outlined in section 3. Finally, in section 5 we sum up the conclusions of this work.

2. Generative adversarial route to synthetic spectral data generation

In this section, we review the basics of the three main generative adversarial frameworks that form the basis of this work, namely, GANs, WGANs and CWGANs. Here, we assume that the origin of the real spectra used in these approaches is completely general, coming from any physical, chemical or biological process (in the following sections we discuss the application to a particular problem within the realm of near-field radiative heat transfer). Figure 1 shows a schematic representation of the underlying architecture of a general GAN algorithm. It comprises two main parts: a generator network and a discriminator network (represented as orange and blue rectangles, respectively, in figure 1). The generator and the discriminator are two interconnected networks in the GAN system. The generator generally begins with random noise (\mathbf{z} in figure 1) and uses it to create new data samples (new spectra in our application). On the other hand, the discriminator takes these spectra and calculates the likelihood that each one originates from the actual training data set. The two networks have competing objectives. The generator's goal is to create spectra that perfectly mirrors the distribution of the training data. In doing so, it aims at generating spectral data so convincingly authentic that the discriminator cannot tell it apart from the real training data. Meanwhile, the discriminator aims at distinguishing between the actual training data and the data fabricated by the generator. Both networks are trained together in a competition until a Nash-type equilibrium is reached and the training process ends [27]. At that stage, the generator is producing spectra that the discriminator can no longer reliably classify as 'real' or 'fake', signaling the end of the training process.



Mathematically, the GAN configuration can be formulated as a minimization problem of the generator and discriminator loss functions, suitably written in terms of differentiable functions representing the discriminator and the generator network models, $D(\mathbf{x}; \theta_d)$ and $G(\mathbf{z}; \theta_g)$, respectively (θ_d and θ_g are the parameters of the corresponding ANNs—in what follows, for the sake of clarity, we do not include this dependence in the equations). We start by focusing on the loss function of the discriminator, $L_D(G, D)$, which can be written as [27]

$$L_D(G, D) = -\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log(D(\mathbf{x}))] - \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] , \quad (1)$$

where \mathbb{E} stands for expectation over either the training samples, \mathbf{x} , or the input noise variables, \mathbf{z} (characterized respectively by a probability distribution $p_{\text{data}}(\mathbf{x})$ and a prior distribution $p_z(\mathbf{z})$). From equation (1), and considering that D outputs a single scalar representing a probability, we can see that the training of the discriminator is trying to minimize the likelihood of mistaking a real sample for a fake one or a fake sample for a real one (first and second terms, respectively, of the right-hand side of equation (1)).

However, there are significant performance issues associated to this original choice of the loss function $L_D(G, D)$, including the difficulty to reach the above-mentioned Nash equilibrium state (each network is updated independently, and given the competitive nature of the generator and the discriminator, in general, there is no clear point to stop the training) or the so-called *mode collapse* (arising when a network fails to generalize accurately to all regions of the training data distribution). To address these issues, a different strategy to train GANs was introduced, the so-called WGANs [44]. One of the main modifications of WGANs with respect to the original GAN architecture is the presence of a critic model, $C(\mathbf{x})$, instead of the discriminator model. Importantly, $C(\mathbf{x})$ outputs a score instead of a probability, which, in turn, allows us to define a new loss function for the critic, $L_C(G, C)$, given by

$$L_C(G, C) = -\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [C(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [C(G(\mathbf{z}))]. \quad (2)$$

This loss function summarizes well some of the main advantages of the WGANs. WGANs help address the training challenges of GANs by using the Wasserstein distance (or Earth Mover's distance) as the loss function instead of the Jensen–Shannon divergence used in original GANs. In addition, it provides a meaningful loss metric, i.e. the value of the critic in WGANs provides a meaningful measure of the distance between the real and generated data distributions. This is unlike the original GANs where the discriminator's output does not correlate well with the quality of the generated samples.

A pivotal aspect of WGANs is the need for the Critic to operate within the set of the so-called *1-Lipschitz* functions, a critical component of the model. Lipschitz functions are mathematical functions possessing a property where there exists a real-valued constant such that, for every pair of points, the absolute difference in function values can be bounded by this constant times the absolute difference of input values. When this constant is 1, the functions are known as 1-Lipschitz. The 1-Lipschitz constraint is crucial as it bounds how

much a function's output can change with small variations in input, ensuring the function does not change too abruptly. In WGANs, this is vital to ensure that the Critic provides meaningful and stable gradients for the generator to learn from, facilitating a more reliable learning process. Originally, weight clipping was proposed to enforce this 1-Lipschitz condition. However, this sometimes resulted in convergence failure. To counter these issues, a more robust technique known as the *gradient penalty* was introduced. This method involves adding a loss term to maintain the L2 norm (a measure of the vector length of parameters or weights) of the Critic close to a value of 1 [45]. This approach assists in keeping the Critic's function within the 1-Lipschitz constraint, enhancing the stability and performance of WGANs. Incorporating these improvements, the full loss for the critic in a WGAN now reads

$$L_C(G, C) = -\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [C(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [C(G(\mathbf{z}))] + \lambda \mathbb{E}_{\hat{\mathbf{x}} \sim p_{\hat{\mathbf{x}}}} \left[\left(\|\nabla_{\hat{\mathbf{x}}} C(\hat{\mathbf{x}})\|_2 - 1 \right)^2 \right], \quad (3)$$

where $\|\cdot\|_2$ is the L2 norm and λ is a weight parameter for the gradient penalty (throughout this work we assume $\lambda = 10$, and that $\hat{\mathbf{x}} = \mathbf{x} + \alpha(G(\mathbf{z}) - \mathbf{x})$ is an interpolated point between a real sample and generated sample on which to calculate the gradient, with $\alpha \in [0, 1]$). As a final remark, proper training of the WGAN requires the critic to be trained ahead of the generator, so that for each training step of the generator the critic is updated n_{train} times (following [44], we chose $n_{\text{train}} = 5$ in all our models).

As for the generator loss function, $L_G(G, C)$, an important aspect to realize is that for synthetic spectral data generation our focus is on a regression problem. This implies that we need a greater control over the output than that obtained just by acquiring a random but realistic sample. It is therefore crucial to ensure that the generated data accurately corresponds to the correct system parameters that yield that response. To achieve this, we need to *condition* the WGAN, leading to the creation of a CWGAN [46]. In this work, we will implement that by adding an extra loss term quantifying the mean absolute error (MAE) between the conditioned generated example and the training example corresponding to the *ground truth* of the condition. Accounting for this conditioning, $L_G(G, C)$ can be expressed as

$$L_G(G, C) = \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z}), \mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [|\mathbf{x} - G(\mathbf{z}|\mathbf{W})|] - \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [C(G(\mathbf{z}))], \quad (4)$$

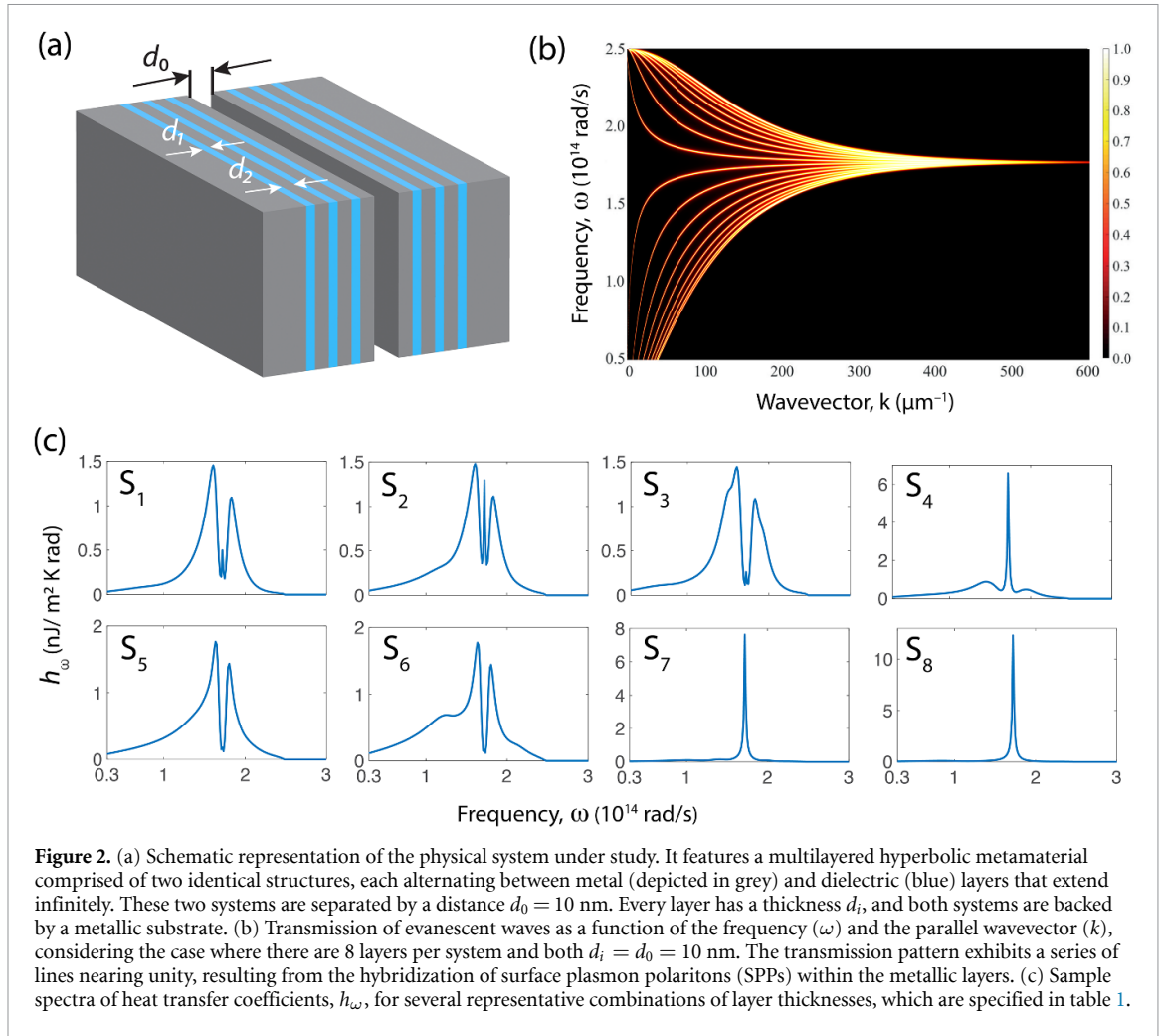
where \mathbf{x} is the training example corresponding to the system parameters \mathbf{W} , and $G(\mathbf{z}|\mathbf{W})$ is a generated example conditioned on the same parameters \mathbf{W} . The first term in the r.h.s. of equation (4) corresponds to the above discussed conditioning procedure, while the second term is associated to the coupling of the generator and the critic.

3. Illustrative problem: near-field radiative heat transfer spectra in multilayer hyperbolic metamaterials

In this section we provide an overview of the fundamentals of the specific problem we use to illustrate the proposed approach. We have chosen a physical problem in the context of near-field radiative heat transfer involving multilayer hyperbolic metamaterials [47]. Despite the specific character of this class of systems, the chosen problem can be considered both representative of the types of problems that our approach can address effectively and complex enough to showcase the versatility of our method.

In the realm of thermal radiation, a key breakthrough has been achieved with the experimental verification that the limit established by Stefan–Boltzmann's law for radiative heat transfer between two bodies can be significantly exceeded by reducing their distance to less than the thermal wavelength λ_{Th} ($\sim 10 \mu\text{m}$ at room temperature) [48]. At such near-field distances, the heat transfer process is enhanced through evanescent waves (photon tunneling), an effect not encompassed by Stefan–Boltzmann's conventional framework. This mechanism governs the near-field radiative heat transfer (NFRHT) when the intervening separations are exceedingly narrow [49–52]. To further augment NFRHT, numerous strategies have been introduced, with a particular emphasis on exploiting surface modes found within multilayered structures. The creation of *hyperbolic metamaterials*, achieved by alternating layers of dielectric and metallic materials, has garnered considerable attention [53–62]. It has been demonstrated that hybridizing surface modes across metal-dielectric interfaces significantly enhances NFRHT, surpassing that observed with two parallel infinite plates [60].

The model system considered in this work is sketched in figure 2(a). It consists of two identical multilayered structures separated by a gap d_0 . Each structure comprises N layers, alternating between a metal with permittivity ϵ_m and a lossless dielectric with permittivity ϵ_d ($\epsilon_d = 1$ is assumed throughout this work). The metal layers are characterized by a Drude model permittivity: $\epsilon_m(\omega) = \epsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\gamma)}$, where ϵ_∞ represents the infinite frequency permittivity, ω_p is the plasma frequency, and γ is the damping rate. We set the parameters $\epsilon_\infty = 1$, $\omega_p = 2.5 \times 10^{14} \text{ rad s}^{-2}$, and $\gamma = 1 \times 10^{12} \text{ rad s}^{-2}$, which yield a surface plasmon



frequency that closely matches the surface phonon polariton frequency observed at the interface between SiC and vacuum. The dimension of each layer, indexed as i , is represented by d_i , which can vary within a predefined range (details to be provided subsequently).

To study the radiative heat transfer across the gap of the system, we adopt the analysis of [47, 60] and use the well-established theory of fluctuational electrodynamics [63, 64], with a particular emphasis on near-field effects. In such a regime, radiative heat transfer takes place primarily through TM- or p -polarized evanescent waves. The heat transfer coefficient (HTC) between the structures, representing the linear radiative thermal conductance per unit area, can be expressed as [65]

$$h = \frac{\partial}{\partial T} \int_0^\infty \frac{d\omega}{2\pi} \Theta(\omega, T) \int_{\omega/c}^\infty \frac{dk}{2\pi} k \tau_p(\omega, k), \quad (5)$$

where ω is the frequency and k is the magnitude of the wave vector parallel to the surface planes. Here, $\Theta(\omega, T) = \frac{\hbar\omega}{e^{\hbar\omega/k_B T} - 1}$ is the average thermal energy of a mode at frequency ω (T is the temperature). The transmission of p -polarized evanescent modes is given by [60]

$$\tau_p(\omega, k) = \frac{4 [\text{Im} \{r_p(\omega, k)\}]^2 e^{-2q_0 d_0}}{|1 - r_p(\omega, k)^2 e^{-2q_0 d_0}|^2} \quad (6)$$

where the function $r_p(\omega, k)$ represents the Fresnel reflection coefficient for p -polarized evanescent waves, transitioning from vacuum to a layer, and $q_0 = \sqrt{k^2 - \omega^2/c^2}$ is the normal component of the wave vector in vacuum. The coefficients were numerically computed using a scattering matrix method [66]. Although we considered the impact of s -polarized modes, they are negligible for the gap sizes investigated.

The significance of studying near-field radiative heat transfer (NFRHT) in these layered structures lies in the ability to tailor heat exchange by adjusting layer thicknesses. Specifically, in Drude metal parallel plates, NFRHT is predominantly influenced by two cavity surface modes [60], emerging from the hybridization of

Table 1. Thicknesses combinations of the representative samples of the spectral heat transfer coefficient, h_ω , shown in figure 2(c).

| Label | Layer thicknesses (nm) | | | | | | | |
|-------|------------------------|-------|-------|-------|-------|-------|-------|-------|
| | d_1 | d_2 | d_3 | d_4 | d_5 | d_6 | d_7 | d_8 |
| S_1 | 5.0 | 20.0 | 20.0 | 5.0 | 20.0 | 5.0 | 5.0 | 20.0 |
| S_2 | 5.0 | 12.5 | 12.5 | 5.0 | 12.5 | 5.0 | 5.0 | 12.5 |
| S_3 | 5.0 | 12.5 | 5.0 | 5.0 | 5.0 | 20.0 | 12.5 | 5.0 |
| S_4 | 5.0 | 5.0 | 12.5 | 5.0 | 5.0 | 5.0 | 20.0 | 12.5 |
| S_5 | 5.0 | 5.0 | 5.0 | 20.0 | 12.5 | 12.5 | 5.0 | 20.0 |
| S_6 | 5.0 | 5.0 | 5.0 | 12.5 | 5.0 | 20.0 | 5.0 | 12.5 |
| S_7 | 20.0 | 5.0 | 20.0 | 12.5 | 20.0 | 5.0 | 12.5 | 20.0 |
| S_8 | 12.5 | 12.5 | 5.0 | 12.5 | 5.0 | 5.0 | 12.5 | 5.0 |

the surface plasmon polaritons (SPPs) of the two metal-vacuum interfaces and yielding two near-unity contours in $\tau_p(\omega, k)$. Introducing internal layers of specific thicknesses allows for NFRHT enhancement through multiple surface states, as shown in figure 2(b) for an $N = 8$ layer structure (4 metallic and 4 dielectric) with a 10 nm vacuum gap at $T = 300$ K. This enhancement over a traditional bulk system is substantial across various gap distances, as detailed in [60].

Of special interest for this work is the spectral HTC, h_ω , defined as the HTC per unit of frequency: $h = \int_0^\infty h_\omega d\omega$. To create the initial dataset of real spectra (which later on will be augmented by our generative adversarial approach), we apply the above-described theoretical framework to compute a total of 6561 h_ω spectra. The thicknesses d_i of each layer were varied between 5 and 20 nm, and every spectrum contains 200 frequency points in the range $\omega \in [0.3, 3] \times 10^{14}$ rad s⁻¹. As discussed in the next section, that dataset of spectra will be split in different proportions to become training and test sets. Figure 2(c) shows several representative samples of h_ω spectra, corresponding to the following thicknesses combinations listed in table 1.

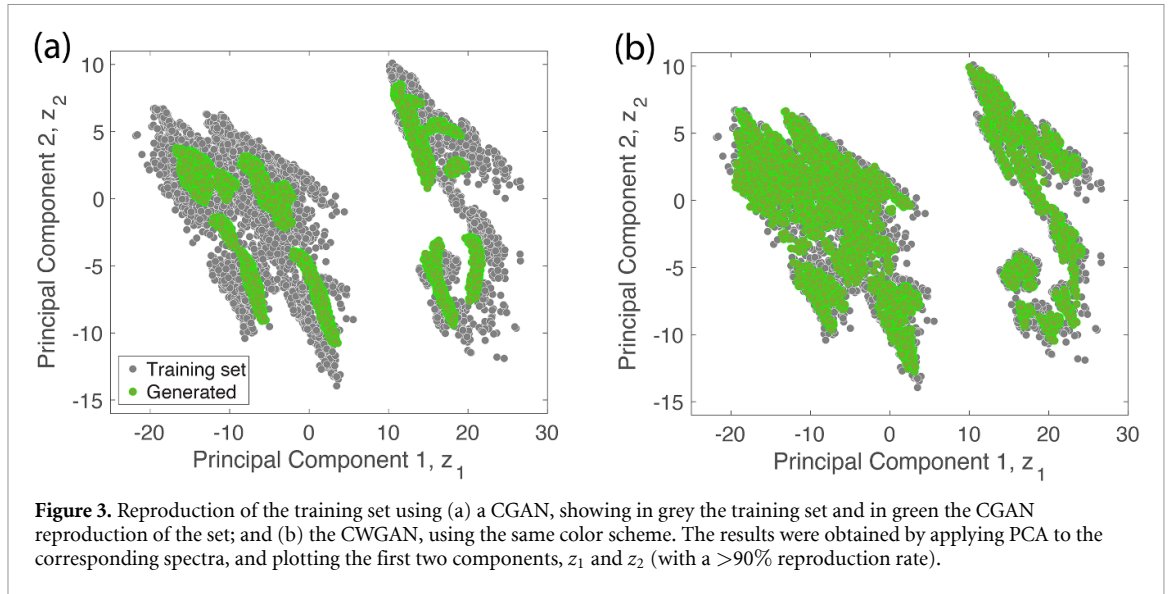
The spectra displayed in figure 2(c) show the broad variety of spectral features that can be obtained from the studied system (from double broad peaks with very narrow resonances in between, to single narrow peaks, or to two resonant peaks separated by a gap). This set of spectra essentially serves as a comprehensive blueprint for the whole adversarial approach, guiding it on the characteristics and features that should be exhibited in the synthetic data. Hence, this allows the proposed approach to capture a wider range of underlying patterns and relationships, which in turn allows it to generate a more realistic and diverse array of synthetic spectral data.

In this work we leverage the robust theoretical models of radiative heat transfer, which have been validated by experimental data across various studies (see, for instance [49–52]). The fidelity of these models provides a solid foundation for the datasets used to train our GANs, underscoring the reliability of our generated data. While we recognize the merit of experimental validation, our focus is on the application of GANs to address specific challenges within data-scarce scenarios in this field.

4. Results and discussion

We proceed in this section to report on the results obtained when applying the generative adversarial approach summarized in section 2 to the specific illustrative problem described in section 3. We begin by providing a strong quantitative justification of the necessity of using a trained CWGAN for this problem, instead of a plain CGAN (Conditional GAN—the details of the specific architecture used in each case are provided below). Figures 3(a) and (b) show, respectively, the ability to reproduce the training set of a trained CGAN and a trained CWGAN projecting the data on two dimensions via a principal component analysis (PCA) [67], which retains most of the training data structure due to its >90% reproduction rate (RR).

The PCA calculations shown in figure 3 were done as follows. First, we performed a singular value decomposition (SVD) of the covariance matrix Σ : $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{SVD}(\Sigma)$ with $\Sigma = \frac{1}{m} \mathbf{X}^T \mathbf{X}$ [68]. Here, \mathbf{U} and \mathbf{V} are unitary matrices, \mathbf{S} is the singular value matrix, m is the total number of data examples and \mathbf{X} is the data matrix, containing in each column one data example \mathbf{x} . To obtain the reduced 2-dimensional (2D) representation of the data, we calculated a reduced matrix $\mathbf{U}_{\text{reduced}}$ retaining the first 2 columns of the \mathbf{U} matrix obtained from the SVD, and use it as the projection matrix, $\hat{\mathbf{x}} = \mathbf{U}_{\text{reduced}}^T \mathbf{X}$, where $\hat{\mathbf{x}}$ is the 2-dimensional representation of the data. The RR of the whole 2D-PCA analysis is then defined as the ratio of the first 2 singular values over all the N singular values obtained:

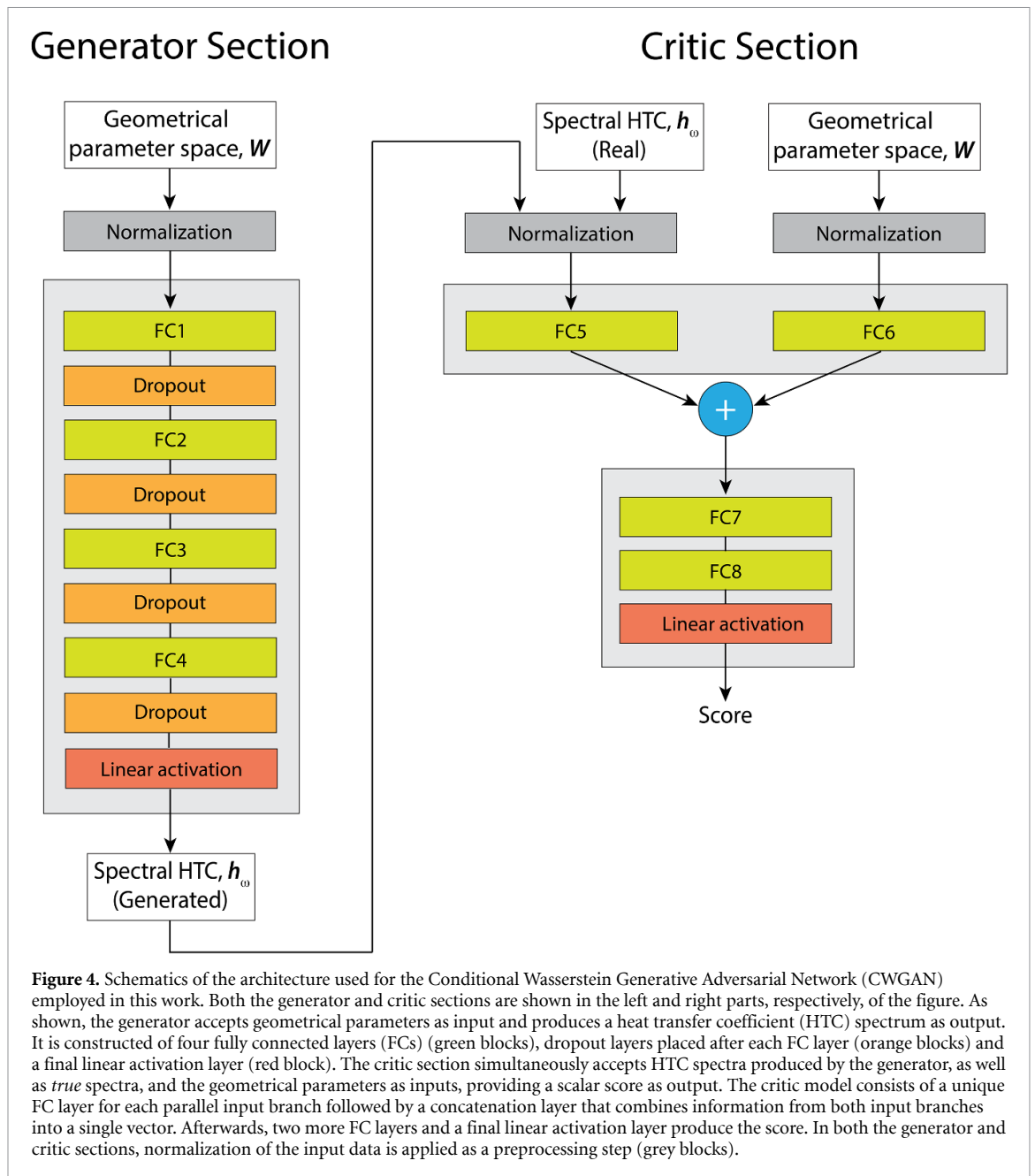


$$RR = \frac{\sum_{i=1}^2 S_{ii}}{\sum_{i=1}^N S_{ii}}. \quad (7)$$

The Conditional GAN and the Conditional WGAN share the same architectural design, the specifics of which are detailed further on. Both networks underwent identical training conditions, with the complete spectral dataset partitioned into 80% for training and 20% for validation. The key distinguishing factor lies in their respective loss functions ($L_D(G, D)$ for CGAN and $L_C(G, C)$ for CWGAN). The CGAN employs equation (1) along with a generator loss which is the sum of the first term in the r.h.s. of equation (4) (the conditional term) and the second term in the r.h.s. of equation (1) with opposite sign. Meanwhile, the CWGAN utilizes the loss defined in equation (4) for the generator and equation (3) for the critic. Notably, as depicted in figure 3(a), the CGAN manifests evident signs of mode collapse, rendering it unable to reproduce most examples beyond the principal cluster structures in the PCA. In contrast, the CWGAN shows the capability of replicating most of the complexities within the training set. These results can be considered as a novel illustration, in the context of synthetic generation of spectral data, of the key role played by Wasserstein's loss function to create more robust generative adversarial approaches.

Figure 4 summarizes the specific CWGAN architecture we have found as most efficient for the studied problem. The generator (left side of figure 4) is composed of 4 hidden fully connected layers of an increasing number of neurons in each layer (we consider 50, 100, 150, and 200 neurons for the four layers—represented as green blocks, labeled as FC1–FC4, in figure 4). The generator model takes the condition as input (the 8 values for the thicknesses of the layers) and returns a generated h_w spectrum (sampled by 200 frequencies). Consistently with the results in [69, 70], we found that we did not need to feed a random data distribution \mathbf{z} into the generator for operation: 20% dropout layers between all layers of the generator (i.e. connections between two consecutive neurons are dropped with a 20% chance) provide enough variability for this task (the dropout operations are represented as orange blocks in figure 4). On the other hand, the critic (right side of figure 4) takes both a sample spectrum and the condition (either generated or from the training distribution) in parallel processing lines of a single hidden layer (with 150 and 50 neurons, respectively—FC5 and FC6 in figure 4). Then, it concatenates and processes the information through two additional hidden layers (FC7 and FC8 in figure 4, with 100 and 50 neurons, respectively) to output a single number, the score. All hidden layers feature a scaled exponential linear unit (selu) activation function. In all models discussed in this work, as a pre-processing step, we will also calculate the logarithm of the spectra, subtract the mean of both the input parameters and the spectra, and divide both the system parameters and the spectra by the standard deviation (this normalization operation is represented by grey blocks in figure 4). Finally in both the generator and the critic a final linear activation layer is added to ensure the output has the correct size (red blocks in figure 4).

Note that, in the development of our CWGAN tailored for multilayer hyperbolic metamaterials, particular attention has been paid to the physical viability of the generated spectra. The training dataset comprises theoretically sound validated spectra, ensuring a foundation rooted in physical reality. In addition, the CWGAN architecture is designed to integrate physical parameters, which govern the generation process



to conform with the constraints of hyperbolic metamaterials, which further guarantees the generation of spectra that are physically viable.

In this context, we acknowledge various applications of GANs in spectral data generation across different domains. Thus, for instance, de Oliveira *et al* [43] indicated the limited utility of GANs compared to supervised generative networks in generating artificial gamma ray spectra, a finding that contrasts with our demonstration of GANs' efficiency in synthesizing spectral data for radiative heat transfer. Audebert *et al* [40] showcased the generation of annotated hyperspectral samples using GANs, yet their broad categorization approach differs from our precise conditioning using continuous variables. Teng *et al* [41] expanded a laser-induced breakdown spectroscopy spectral database with GANs without specific conditioning, unlike our method which integrates accurate geometrical parameters. Lastly, Zhu *et al* [42] employed boundary equilibrium generative adversarial networks for biological sample spectra synthesis, utilizing an autoencoder discriminator and lacking conditioning, a method not suited for our application. These studies collectively highlight the importance of domain-specific GAN adaptations, affirming that a method effective in one context may not translate directly to another due to fundamental differences in application scope, requirements and objectives.

We focus now on analyzing the evolution with the training steps of the loss functions of the discriminator and the generator ($L_C(G, C)$ and $L_G(G, C)$, respectively), as obtained by training the architecture displayed in

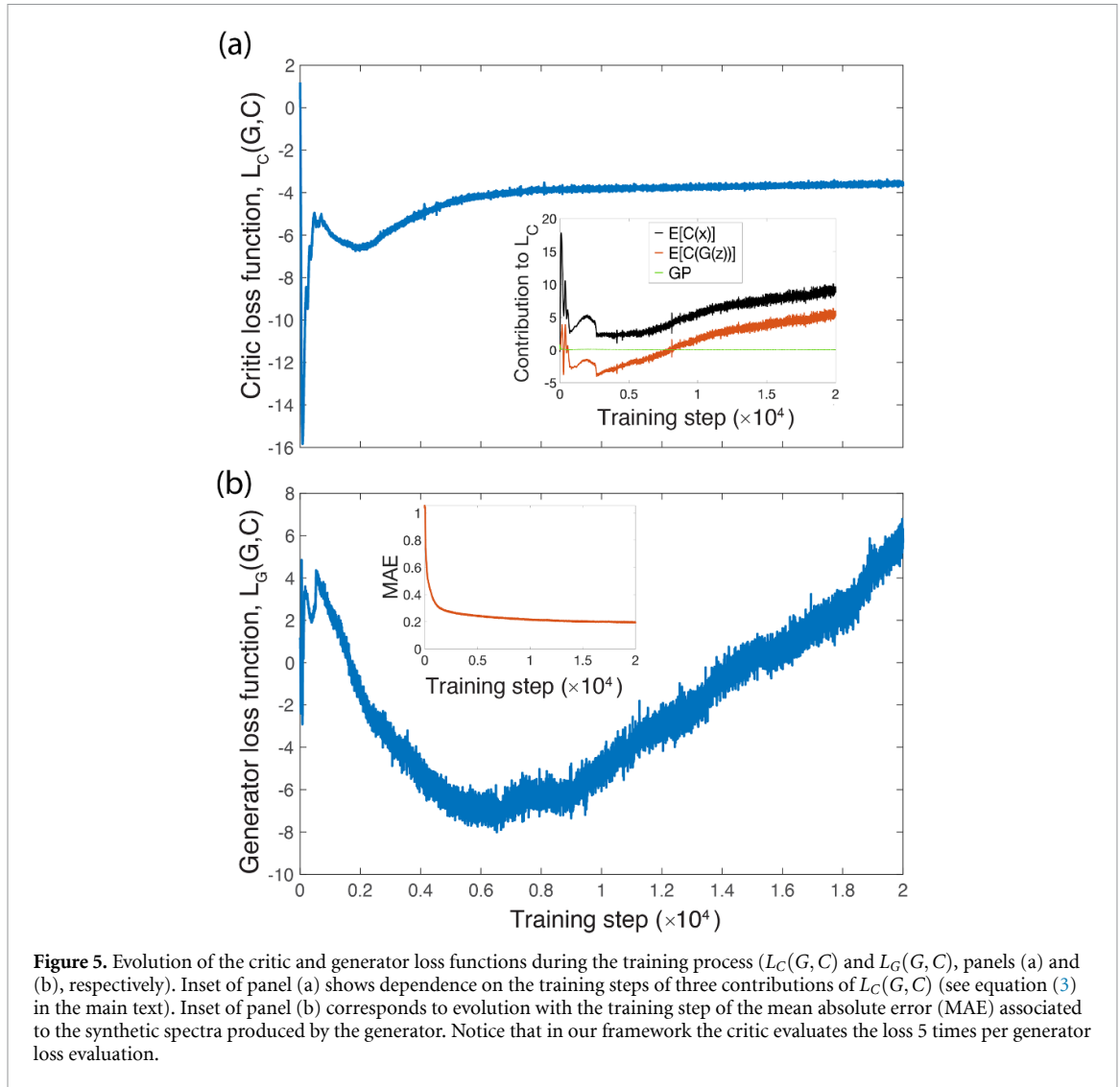


figure 4 with the 6561 h_ω spectra described in section 3. Monitoring the loss functions underlying our model during the training process is critical to understanding, diagnosing, and improving the whole generative adversarial framework studied in this work. The obtained numerical results are summarized in figure 5 (panel (a) corresponds to $L_C(G, C)$, while panel (b) displays the results for $L_G(G, C)$). As seen in figure 5(a), $L_C(G, C)$ initially displays a transient behavior based on a fast drop in value followed by a sudden increase and an oscillation (around 10^3 training steps), until it reaches at stationary value at approximately 10^4 training steps. To get additional insight into the numerical origin of this evolution, the inset of figure 5(a) shows the dependence with the training steps of the three different terms forming $L_C(G, C)$, namely, $\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})}[C(\mathbf{x})]$, $\mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[C(G(\mathbf{z}))]$, and the gradient penalty, $\lambda \mathbb{E}_{\hat{\mathbf{x}} \sim p_{\hat{\mathbf{x}}}}[(\|\nabla_{\hat{\mathbf{x}}} C(\hat{\mathbf{x}})\|_2 - 1)^2]$ (see equation (3)). As observed, $L_C(G, C)$ is dominated by the expected values of the score value produced by the critic model for both the true training samples, (\mathbf{x}), and the samples fabricated by the generator, $G(\mathbf{z})$ (black and red lines, respectively, in inset of of figure 5(a) —the green line corresponds to the gradient penalty term). A remarkable aspect of the overall evolution of both expected value contributions to $L_C(G, C)$ is the fact that, despite their complicated transient behavior, they lock their difference in values after around 10^4 training steps. This in turn allows $L_C(G, C)$ to reach a stationary value (notice the difference in sign between both terms in equation (3)), which marks the completion of the learning process by the critic.

Figure 5(b) shows the results for the evolution of the generator loss function, $L_G(G, C)$, during the training process. The inset of figure 5(b) displays the corresponding evolution of the MAE contribution to $L_G(G, C)$ (i.e. the evolution of the first term in the r.h.s. of equation (4)). As deduced by comparing the values reached by the MAE with those of $L_G(G, C)$, the generator loss function is dominated by the expected value term, $\mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[C(G(\mathbf{z}))]$ (second term in the r.h.s. of equation (4)). As described above, for large values of the training step ($\gtrsim 10^4$), this latter term necessarily follows its counterpart term in the critic loss function (so, as also pointed out above, their difference is maintained and a stationary value of $L_C(G, C)$ is reached).

Therefore, since we do not expect $L_G(G, C)$ to reach a stationary value, MAE becomes the key magnitude to monitor the quality of the learning process of the generator (note that that this is also in accordance with the overall goal of the proposed application: creating artificial spectral samples indistinguishable from the real ones). Indeed, as shown in inset of figure 5(b), the MAE computed for the studied problem converges to values <0.2 for the larger training step values considered in our calculations.

Next, we proceed to quantify our model's proficiency in reproducing the spectral heat transfer coefficient (HTC) of the studied physical system, based on the model architecture illustrated in figure 4. For this analysis, our focus is on two distinctive metrics, concentrating on two different aspects of the spectra. The first, the per-point relative mean error (L_{point}), gauges the model's capability to accurately represent each point in the spectrum through a relative error analysis. The second, the integral relative mean error (L_{integ}), is predominantly influenced by the model's accuracy in reproducing the primary spectral characteristics, notably the resonances. Their respective definitions are as follows:

$$L_{\text{point}} = \frac{1}{Nm} \sum_{i=1}^N \sum_{j=1}^m \frac{y_j^i - \hat{y}_j^i}{y_j^i}, \quad (8)$$

$$L_{\text{integ}} = \frac{1}{N} \sum_{i=1}^N \frac{\int (y^i - \hat{y}^i) d\omega}{\int y^i d\omega}, \quad (9)$$

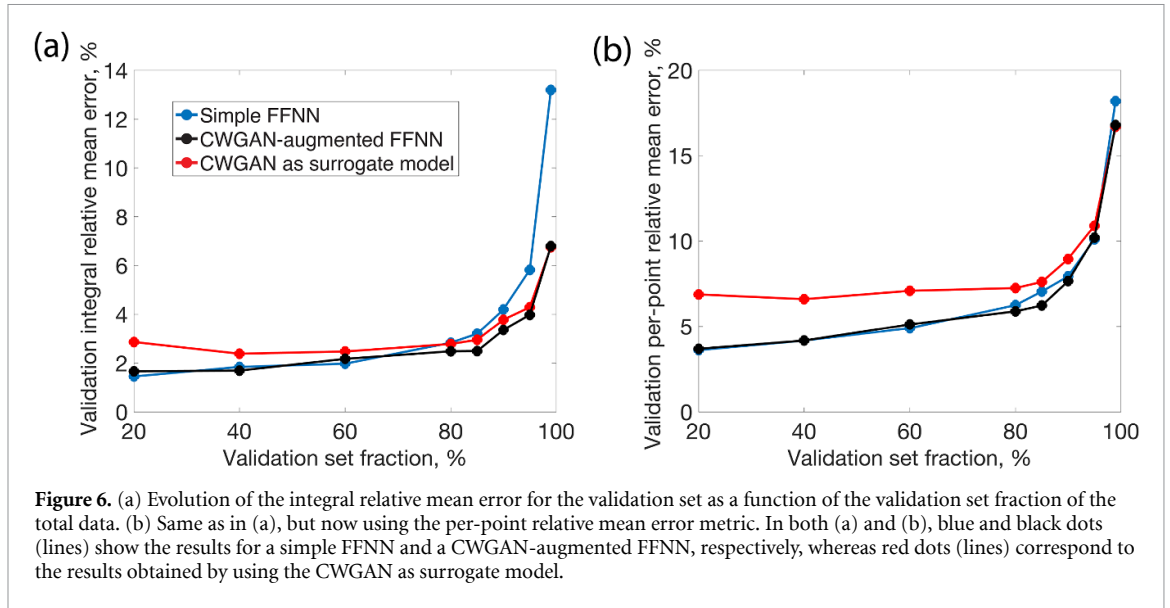
where N is the number of spectra, m the number of points in each spectrum, y^i is a data example and \hat{y}^i is the corresponding generated example by the model. Note that we undo all pre-processing operations to perform these calculations, and that all integrals are performed over frequency points.

We start our analysis by establishing a baseline outcome for comparison using a simple FFNN, comprised of five hidden layers, each hosting 200 neurons, characterized by a selu activation function, and a final linear activation layer. An analogous network has been previously shown to effectively model this identical physical system given ample data [47]—consequently, we anticipate that this particular network will serve as a suitable benchmark. Another important aspect is the apparent similarities between the generator and the FFNN. We aim at demonstrating that this specific simple design, given a sufficiently large dataset, can accurately replicate the spectral characteristics inherent to the type of systems under study, without resorting to any data augmentation technique. To accomplish this, we split the original dataset, which consists of 6561 h_ω spectra, allocating 80% for training set and the remaining 20% for the validation set. Our findings reveal that after 50 000 training iterations (epochs), and using the Adam optimization algorithm with a steady learning rate of 3×10^{-4} , this simple neural network is proficient in replicating the system's spectra, reaching values of L_{point} and L_{integ} for the validation set of 3.61% and 1.45%, respectively (note that these results are comparable to those found in [47]).

Then we proceed to assess the CWGAN capability as numerical engine for spectral data augmentation. To do that, we follow a two-step approach. First, once the CWGAN has been trained, we use it to generate a number of additional spectra to include on the training set. Second, we retrain the above-described FFNN with this new data set to create an *augmented* FFNN. In this work, we generate a total of 10 000 new spectra for this data augmentation process, which are added to the training set. This particular amount of additional spectra was chosen after observing that our results are converged when adding to the original dataset a number of extra spectra in the range 5000–10 000. Once the generator has been trained, this augmentation of the original dataset is computationally inexpensive. Note also that once the generator has been trained, it can be isolated from the overall model and employed for inverse design tasks, adhering to the methodology delineated in our prior research [47].

Following this approach, first, we found that when using the training/validation split considered until now (80%–20%), the FFNN and the augmented FFNN have similar performance, both in terms of per-point relative mean accuracy and integral relative mean error (respective error values of approximately 3.6% and 1.5% are obtained for both models). This suggests that, as anticipated, using a sufficiently large dataset minimizes the impact of augmenting the original dataset, making any data augmentation barely noticeable. However, that conclusion changes when a different data scenario is considered. To modify the amount of data available to the models, we increasingly reduced the size of the training set by transferring part of it to the validation set, and retrained both the simple and augmented FFNNs from scratch (different values of the split training/validation were considered in the range 80%–20%–1%–99%—note that the data augmentation is done separately for each split ratio).

Figure 6 summarizes the main results of this analysis using the two error metrics considered in this work (panels (a) and (b) correspond to L_{integ} and L_{point} , respectively—blue dots (lines) display the results for the simple FFNN, whereas black dots (lines) correspond to the CWGAN-augmented FFNN). As seen, both in



terms of L_{integ} and L_{point} , the simple FFNN and the augmented FFNN have similar performance when the validation set fraction is smaller than approximately 70% (we can therefore ascribe that interval to a sufficiently large training set scenario). However, in the case of L_{integ} (figure 6(a)), further increasing of the validation set fraction (i.e. further reduction of the training set size) leads to the augmented FFNN to perform increasingly better with training data reduction than the simple FFNN. As observed, values of $L_{\text{integ}} = 13.2\%$ are achieved using a simple FFNN in the limit case of a validation set fraction of 99%, whereas an augmented FFNN reduces that error close to a than a half ($L_{\text{integ}} = 6.8\%$). The performance improvement is not so dramatic in the case of the per-point relative mean error (figure 6(b)), but we still observe slight improvements in the low-data scenario (for the extreme case of a validation set fraction of 99%, $L_{\text{point}} = 18.2\%$ is obtained for the simple FFNN, while the augmented FFNN leads to $L_{\text{point}} = 16.8\%$). Overall, these numerical results suggest that models for radiative heat transfer trained on synthetic data created by the generator can be more accurate, under conditions of data scarcity, than models trained exclusively on real data. This superior performance of synthetically trained models has been also reported in the context of video representations based on synthetic data (see for instance [71]).

Finally, to conclude the present analysis, we compare the performance as surrogate model to generate h_ω spectra of the trained CWGAN with that of the simple and augmented FFNNs. This application arises from the conditioning modification we introduced in this work to conventional GANs (see equation (4) and the corresponding discussion in section 2). This conditioning allows us to accurately keep track of the geometrical parameters associated to each spectra created by the generator of our CWGAN model, so once the generator is trained, it can be decoupled from the whole architecture and used as a standalone surrogate model. The obtained numerical results are also displayed in figure 6 (red points and red lines corresponds to the results of a CWGAN used as surrogate model). As observed, for the integral relative mean error (figure 6(a)) the CWGAN displays worse performance than both the simple and the augmented FFNNs when the validation set fraction is below 70% approximately (i.e. the data regime that we previously labeled as of sufficiently large training set scenario for both the simple and the augmented FFNNs). However, in the low-data regime (values of the validation set fraction larger than 70%) we observe that the CWGAN surrogate gradually improves the results of a simple FFNN, until fully reproducing the improvement reached by an augmented FFNN in the extreme case of 99% validation set fraction. Regarding the per-point relative mean error (figure 6(b)), the CWGAN surrogate displays a significantly worse performance in comparison to both the simple and the augmented FFNNs for the validation set fractions below 70%, but from that point on, it starts converging to the results of an augmented FFNN in the low-data regime. Ultimately, this comparison between the surrogate role of all considered models reinforces our previous conclusion that the CWGAN shows higher resilience to the reduction of training data than the simple FFNN, becoming also in this context an efficient architecture in low-data scenarios.

5. Conclusions

In this work we have studied the application of GANs for synthetic spectral data generation in the domain of radiative heat transfer, an area where their use has not been previously reported. Our main focus has been an

illustrative problem in the domain of near-field radiative heat transfer involving a multilayered hyperbolic metamaterial. We have analyzed the use of a CWGAN for data augmentation and studied its influence on the predictive capacities of a FFNN. Our results reveal that generating spectral data effectively entails two main modifications to traditional GANs: firstly, incorporating WGANs to prevent mode collapse, and secondly, conditioning these WGANs to secure accurate labels for the generated data. It is demonstrated that a basic FFNN, augmented with data yielded by a CWGAN, substantially improves its efficiency in scenarios where data is scarce. Moreover, we present that CWGANs can function as a superior surrogate model when data is limited. Overall, this work aims at contributing to the research area of generative AI algorithms' applicability beyond the conventional field of image generation. We believe that our findings contribute to the understanding and applicability of generative AI algorithms in data-constrained contexts and could stimulate further research work in the application of generative AI in a variety of scientific scenarios.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/jbravoabad/spectralgen>.

Acknowledgments

J J G E was supported by the Spanish Ministry of Science and Innovation through a FPU Grant (FPU19/05281). J C C acknowledges funding from the Spanish Ministry of Science and Innovation (PID2020-114880GB-I00). J B A acknowledges funding from the Spanish Ministry of Science and Innovation (PID2022-139995NB-I00).

ORCID iDs

J J García-Esteban  <https://orcid.org/0000-0002-4306-3453>

J C Cuevas  <https://orcid.org/0000-0001-7421-0682>

J Bravo-Abad  <https://orcid.org/0000-0001-6876-8022>

References

- [1] Mitchell T 1997 *Machine Learning* (McGraw-Hill)
- [2] LeCun Y, Bengio Y and Hinton G 2015 *Nature* **521** 436–44
- [3] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)
- [4] Aggarwal C 2018 *Neural Networks and Deep Learning* (Springer)
- [5] Krizhevsky A, Sutskever I and Hinton G 2012 *Advances in Neural Information Processing Systems* vol 25 pp 1097–105
- [6] Hinton G *et al* 2012 *IEEE Signal Process. Mag.* **29** 82–97
- [7] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 arXiv:1406.1078
- [8] Shalev-Shwartz S, Shammah S and Shashua A 2016 arXiv:1610.03295
- [9] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
- [10] Carleo G and Troyer M 2017 *Science* **355** 602–6
- [11] Wang L 2016 *Phys. Rev. B* **94** 195105
- [12] Carrasquilla J and Melko R G 2017 *Nat. Phys.* **13** 431–4
- [13] van Nieuwenburg E P L, Liu Y H and Huber S D 2017 *Nat. Phys.* **13** 435–9
- [14] Wu D, Wang L and Zhang P 2019 *Phys. Rev. Lett.* **122** 080602
- [15] Dunjko V and Briegel H J 2018 *Rep. Prog. Phys.* **81** 074001
- [16] Green S R, Simpson C and Gair J 2020 *Phys. Rev. D* **102** 104057
- [17] Peurifoy J, Shen Y, Jing L, Yang Y, Cano-Renteria F, DeLacy B G, Joannopoulos J D, Tegmark M and Soljačić M 2018 *Sci. Adv.* **4** eaar4206
- [18] So S, Badloe T, Noh J, Bravo-Abad J and Rho J 2020 *Nanophotonics* **9** 1041–57
- [19] Jiang J, Chen M and Fan J A 2021 *Nat. Rev. Mater.* **6** 679–700
- [20] Frising M, Bravo-Abad J and Prins F 2023 *Mach. Learn.: Sci. Technol.* **4** 02LT02
- [21] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 *Nature* **559** 547–55
- [22] Gómez-Bombarelli R, Wei J N, Duvenaud D, Hernández-Lobato J, Sánchez-Lengeling B, Sheberla D, Aguilera-Iparraguirre J, Hirzel T D, Adams R P and Aspuru-Guzik A 2018 *ACS Cent. Sci.* **4** 268–76
- [23] Zhang L, Han J, Wang H, Car R and Weinan E 2018 *Phys. Rev. Lett.* **120** 143001
- [24] Coley C W, Barzilay R, Jaakkola T S, Green W H and Jensen K F 2017 *ACS Cent. Sci.* **3** 434–43
- [25] Rupp M, Tkatchenko A, Müller K R and von Lilienfeld O A 2012 *Phys. Rev. Lett.* **108** 058301
- [26] Ching T *et al* 2018 *J. R. Soc. Interface* **15** 20170387
- [27] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 *Advances in Neural Information Processing Systems* vol 27 (*NIPS 2014*) pp 2672–80
- [28] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X 2016 *NIPS'16: Proc. 30th Int. Conf. on Neural Information Processing Systems* vol 29 pp 2234–42
- [29] Paganini M, de Oliveira L and Nachman B 2018 *Phys. Rev. D* **97** 014021
- [30] Carrasquilla J, Torlai G, Melko R G and Aolita L 2019 *Nat. Mach. Intell.* **1** 155–61

- [31] Koch R and Lado J 2022 *Phys. Rev. Res.* **4** 033223
- [32] Carracedo-Cosme J and Pérez R 2022 arXiv:2205.00447
- [33] So S and Rho J 2019 *Nanophotonics* **8** 1255–61
- [34] Christensen T, Loh C, Picek S, Jakobović D, Jing L, Fisher S, Ceperic V, Joannopoulos J D and Soljačić M 2020 *Nanophotonics* **9** 4183–92
- [35] Rodríguez A C, Kacprzak T, Lucchi A, Amara A, Sgier R, Fluri J, Hofmann T and Réfrégier A 2018 *Comput. Astrophys. Cosmol.* **5** 4
- [36] Anstine D M and Isayev O 2023 *J. Am. Chem. Soc.* **145** 8736–50
- [37] Sanchez-Lengeling B and Aspuru-Guzik A 2018 *Science* **361** 360–5
- [38] Repecka D et al 2021 *Nat. Mach. Intell.* **3** 324–33
- [39] Goldsborough P, Pawlowski N, Caicedo J C, Singh S and Carpenter A E 2017 *bioRxiv Preprint* <https://doi.org/10.1101/227645> (posted online 2 December 2017, accessed 15 February 2024)
- [40] Audebert N, Le Saux B and Lefevre S 2018 *Proc. IEEE Int. Geoscience and Remote Sensing Symp.* pp 4359–62
- [41] Teng G E, Wang Q Q, Kong J L, Dong L Q, Cui X T, Liu W W, Wei K and Xiangli W T 2019 *Opt. Express* **27** 6958–69
- [42] Zhu D, Xu L, Chen X, Yuan L M, Huang G, Li L, Chen X and Shi W 2020 *Opt. Express* **28** 17196–208
- [43] de Oliveira F M F, Daniel G and Limousin O 2023 *Nucl. Instrum. Methods Phys. Res. A* **1047** 167795
- [44] Arjovsky M, Chintala S and Bottou L 2017 arXiv:1701.07875
- [45] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A 2017 arXiv:1704.00028
- [46] Mirza M and Osindero S 2014 arXiv:1411.1784
- [47] García-Esteban J J, Bravo-Abad J and Cuevas J C 2021 *Phys. Rev. Appl.* **16** 064006
- [48] Polder D and Van Hove M 1971 *Phys. Rev. B* **4** 3303–14
- [49] Song B, Fiorino A, Meyhofer E and Reddy P 2015 *AIP Adv.* **5** 053503
- [50] Cuevas J C and García-Vidal F J 2018 *ACS Photonics* **5** 3896–915
- [51] Biehs S A, Messina R, Venkataram P S, Rodriguez A W, Cuevas J C and Ben-Abdallah P 2021 *Rev. Mod. Phys.* **93** 025009
- [52] Mittapally R, Lim J W, Meyhofer E, Reddy P and Song B 2023 *ACS Photonics* **10** 2474–80
- [53] Guo Y, Cortes C L, Molesky S and Jacob Z 2012 *Appl. Phys. Lett.* **101** 131106
- [54] Biehs S A, Tschikin M and Ben-Abdallah P 2012 *Phys. Rev. Lett.* **109** 104301
- [55] Guo Y and Jacob Z 2013 *Opt. Express* **21** 15014–9
- [56] Biehs S A, Tschikin M, Messina R and Ben-Abdallah P 2013 *Appl. Phys. Lett.* **102** 131106
- [57] Bright T J, Liu X L and Zhang Z M 2014 *Opt. Express* **22** A1112–27
- [58] Miller O D, Johnson S G and Rodriguez A W 2014 *Phys. Rev. Lett.* **112** 157402
- [59] Biehs S A and Ben-Abdallah P 2017 *Z. Nat. A* **72** 115–27
- [60] Iizuka H and Fan S 2018 *Phys. Rev. Lett.* **120** 063901
- [61] Song J, Cheng Q, Lu L, Li B, Zhou K, Zhang B, Luo Z and Zhou X 2020 *Phys. Rev. Appl.* **13** 024054
- [62] Moncada-Villa E and Cuevas J C 2021 *Phys. Rev. B* **103** 075432
- [63] Rytov S 1953 *Theory of Electric Fluctuations and Thermal Radiation* (Air Force Cambridge Research Center)
- [64] Rytov S, Kravtsov Y and Tatarskii V 1989 *Principles of Statistical Radiophysics* vol 3 (Springer)
- [65] Basu S, Zhang Z M and Fu C J 2009 *Int. J. Energy Res.* **33** 1203–32
- [66] Caballero B, García-Martín A and Cuevas J C 2012 *Phys. Rev. B* **85** 245103
- [67] Jolliffe I 2002 *Principal Component Analysis* (Springer)
- [68] James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning* (Springer)
- [69] Isola P, Zhu J Y, Zhou T and Efros A A 2016 arXiv:1611.07004
- [70] Ahmed S, Sánchez Muñoz C, Nori F and Kockum A F 2021 *Phys. Rev. Lett.* **127** 140502
- [71] Kim Y w, Mishra S, Jin S, Panda R, Kuehne H, Karlinsky L, Saligrama V, Saenko K, Oliva A and Feris R 2022 *Advances in Neural Information Processing Systems* vol 35 pp 35710–23