



## LETTER

## OPEN ACCESS

RECEIVED  
21 March 2023REVISED  
19 April 2023ACCEPTED FOR PUBLICATION  
16 May 2023PUBLISHED  
31 May 2023

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.



# Tackling multimodal device distributions in inverse photonic design using invertible neural networks

Michel Frising<sup>1,\*</sup> , Jorge Bravo-Abad<sup>2</sup> and Ferry Prins<sup>1</sup> <sup>1</sup> Condensed Matter Physics Center (IFIMAC) and Department of Condensed Matter Physics, Autonomous University of Madrid, 28049 Madrid, Spain<sup>2</sup> Condensed Matter Physics Center (IFIMAC) and Department of Theoretical Condensed Matter Physics, Autonomous University of Madrid, 28049 Madrid, Spain

\* Author to whom any correspondence should be addressed.

E-mail: [michel.frising@uam.es](mailto:michel.frising@uam.es)**Keywords:** nanotechnology, inverse design, multimodality, machine learning, invertible neural networks, nanophotonics, generative modelingSupplementary material for this article is available [online](#)

## Abstract

We show how conditional generative neural networks can be used to efficiently find nanophotonic devices with desired properties, also known as inverse photonic design. Machine learning has emerged as a promising approach to overcome limitations imposed by the dimensionality and topology of the parameter space. Importantly, traditional optimization routines assume an invertible mapping between the design parameters and response. However, different designs may have comparable or even identical performance confusing the optimization algorithm when performing inverse design. Our generative modeling approach provides the full distribution of possible solutions to the inverse design problem, including multiple solutions. We compare a commonly used conditional variational autoencoder (cVAE) and a conditional invertible neural network (cINN) on a proof-of-principle nanophotonic problem, consisting in tailoring the transmission spectrum through a metallic film milled by subwavelength indentations. We show how cINNs have superior flexibility compared to cVAEs when dealing with multimodal device distributions.

## 1. Introduction

Inverse design is the process of matching a device or process parameters to produce a desired performance. The possibility of enabling new materials and nanodevices by ‘reverse-engineering’ them from the desired properties and characteristics has drawn a great deal of both fundamental and applied interest. Inverse design is particularly popular in the field of nanophotonics, where the ongoing quest for miniaturization requires the exploration of extremely large parameter spaces spanned by freeform geometries and a wide variety of materials combinations. To efficiently explore these vast parameter spaces, the inverse-design problem is commonly approached using computational optimization techniques capable of identifying solutions beyond human intuition. Techniques such as gradient-based topology optimization [1–5], evolutionary design [6, 7] and more recently artificial neural networks [8–11] and global optimization nets [12, 13] have been used successfully to design devices that vastly outperform designs based on human intuition.

To date, however, most inverse design approaches rely on the assumption that a unique one-to-one mapping exists between the device and a given design target. In reality, it is often the case that multiple device designs exhibit comparable or even identical performance, yielding a multimodal device distribution. One clear example of this phenomenon are systems or devices with symmetries. For example, the structure shown in figure 1 (which we will use throughout this manuscript for illustrative purposes) consists of a central slit in a thin metal film flanked on each side by different gratings consisting of periodic indentations in the metal. The system has a natural plane of symmetry in the middle of the slit as it does not matter for the spectral

response if the gratings are swapped. While this may be obvious to the reader, this situation is less clear to an optimization routine which tries to minimize a target and assumes that there is, at least locally, a minimum. Consequently, the optimization routine may oscillate between the two solutions, preventing the algorithm from converging. Here we chose a photonic structure as a testbed since models of photonic systems show very good agreement with experiments [14, 15]. In fact, the agreement is so well that we can use the model as a surrogate without need for further experimental benchmarking.

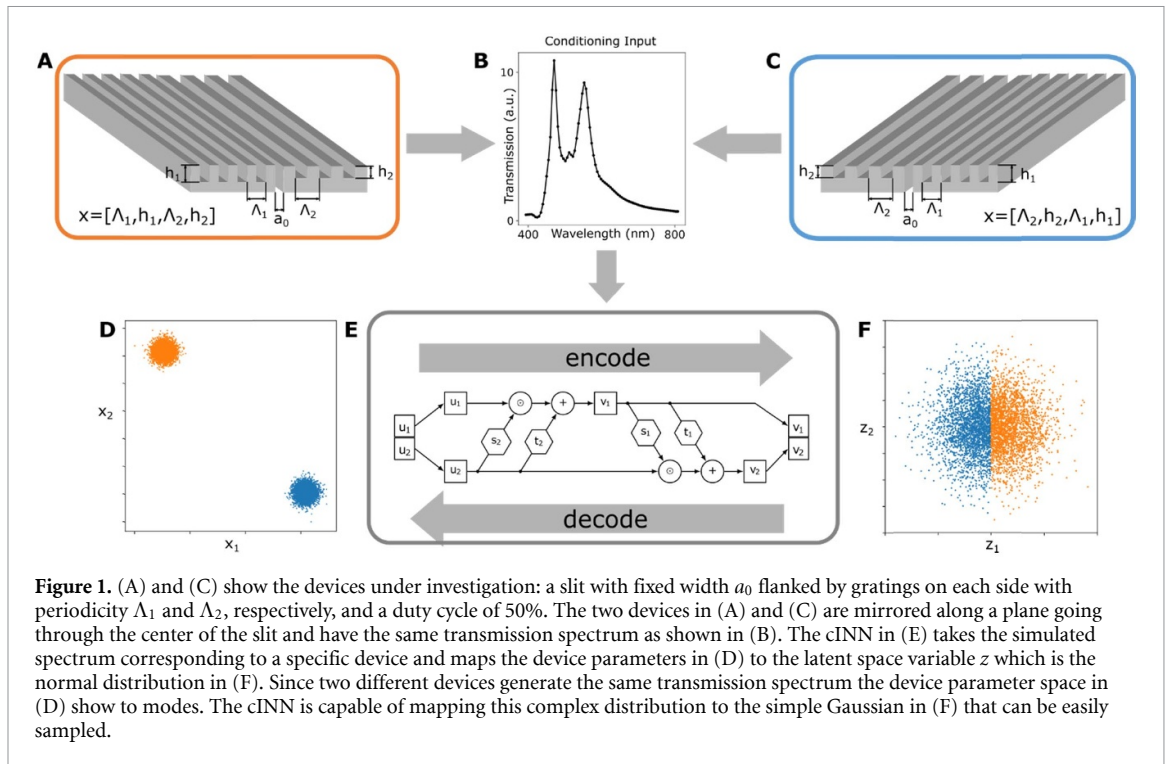
Several studies have attempted to address the issue of multimodality, though with limited success. Liu *et al* [9] observed the problem of multimodality when trying to teach a neural network to reverse-engineer stacks of dielectrics to exhibit desired transmission spectrum. While the forward simulation predicting the transmission spectrum from device parameters could be trained with ease, they observed that matching a stack to a given spectrum (i.e. the inverse pass) was impossible. They attributed this problem to multimodality, meaning that different dielectric stacks map to the same response, preventing the network from converging. They circumvented this issue by using a tandem network in which the forward and backward passes feed into each other to stabilize training. While this approach allows the network to converge despite the multimodality of the training set, the trained network only offers a single solution while other possible solutions are lost.

To truly move beyond one-to-one mapping, recent approaches have explored the use of generative models, including for example VAEs [16] and generative adversarial networks (GANs) [10, 17]. Rather than assuming a simple one-to-one mapping, these generative methods model a distribution of possible devices and their design parameters allowing for the identification of multiple solutions. Ma *et al* [18] employed such a generative model for inverse photonic design, training a VAE to generate proposals for a unit cell of a periodic metamaterial exhibiting a set of desired transmission and reflection properties. However, VAEs are restricted to simple parametrized distributions in their latent space, limiting their expressive power when the device distribution is multimodal. Normalizing flows have emerged as a powerful tool to construct more expressive distributions [19–23] beyond Gaussians to model the true data distribution, but have so far not been used in inverse photonic design. Conditional GANs (cGANs) [24] are also in principle capable of modeling multimodal distributions of parameters but are very prone to mode collapse [25] meaning that the multimodal data distribution is mapped to only one of the modes while still ignoring other possible solutions.

Here, we propose the use of conditional invertible neural networks (cINNs) [26, 27] to tackle the issue of multimodal device parameters in inverse photonic design. cINNs have been recently introduced as a highly versatile platform for inverse design. They belong to the family of flow-based techniques [26, 28], which learn mappings between different distributions. Since they are trained with maximum likelihood loss, mode collapse as observed in GANs, is virtually impossible [26]. Moreover, since the same network is used for the forward and backward pass, only half of number of parameters is necessary. cINNs have been shown to work in a wide range of applications, from improving the robustness in medical imaging and inverse kinematics [26] to changing the style of images [29]. Here, we apply cINNs to the field of photonics, demonstrating their capability to effectively deal with multimodal device parameters in inverse photonic design. As an example of our approach, we train the cINN to find the geometrical parameters of the aforementioned slit in a thin metal film flanked by periodic grooves [15, 30] to match a desired transmission spectrum. The slit flanked by gratings is particularly suited for this purpose, as the symmetry of the structure intrinsically introduces a multimodal distribution in the parameter space of the device geometry. In contrast to past approaches, we show that the cINN provides all possible solutions to the inverse design problem. We emphasize this by comparing our results to a VAE network trained on the same data.

The structures under study are shown schematically shown in figure 1. A central subwavelength aperture of width  $a_0$  in a silver film of thickness  $t$  is surrounded on both sides by gratings made from a finite periodic array of grooves. The periodicity  $\Lambda$  and height  $h$  of the grooves can be different on each side of the slit. The structure is completely described by the parameter vector  $x = [\Lambda_1, \Lambda_2, h_1, h_2]$ . The width of the grooves is fixed to always be half of the corresponding periodicity, i.e. a duty cycle of 50%, and the entire structure is assumed to be embedded in air with a refractive index  $n = 1$ .

Extraordinary optical transmission occurs when surface plasmon polaritons (SPPs) excited by the incident light at the air-metal interface constructively interfere at the subwavelength aperture. By introducing variations in the pitch of the periodic corrugations, the wavelengths at which SPPs can be excited can be modulated, allowing for control over the transmission spectrum of the structure. For a slit which is flanked by the same gratings on each side (i.e.  $\Lambda_1 = \Lambda_2$  and  $h_1 = h_2$ ), a single transmission peak dominates the transmission spectrum. For an asymmetric configuration, i.e.  $\Lambda_1 \neq \Lambda_2$  and  $h_1 \neq h_2$ , more complex transmission spectra can occur, for example yielding two separate transmission peaks at different wavelengths, each corresponding to one of the two periodicities of the gratings. Intuitively, the mirror image of each asymmetric configuration has an identical transmission spectrum, illustrating the intrinsic



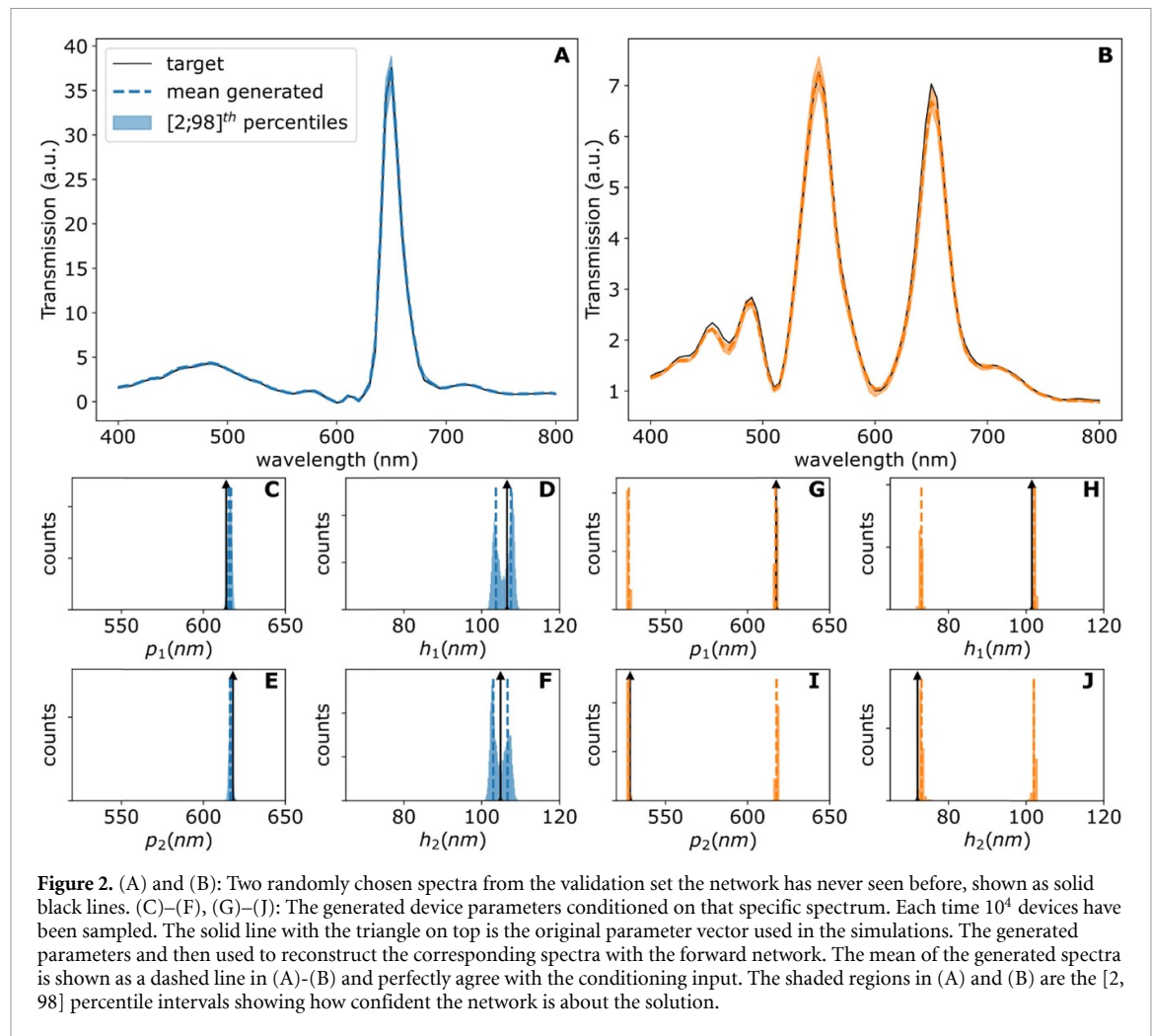
multimodality that is present as a result of symmetries in the geometry. Importantly, while this type of multimodality of mirrored designs is straightforward for human intuition, it causes severe problems in inverse photonic design where one target spectrum suddenly has multiple design solutions associated with it.

The transmission through the slit flanked by gratings is simulated using the coupled mode theory (CMT) framework [15]. The CMT framework has been used extensively to simulate transmission and extraordinary optical transmission (EOT) phenomena in a variety of systems [15, 30] and is known to provide results with excellent agreement to the experiments at minimal computational cost. As a result, the generation of a full training set of 60 000 different structures can be performed efficiently. The training set is divided between 45 000 structures for training, and 15 000 structures for validation.

We modeled our cINN after Ardizzone *et al* [27], using their FrEIA framework, the technical details are described in the Methods section and the Supplementary Information. Briefly, the cINN takes the device parameter vector and maps it to a latent space variable  $z$ , which can be sampled conveniently to generate new devices by running the cINN in reverse. The cINN additionally takes a conditional vector extracted from the spectrum corresponding to the device as an input. The conditioning network consists of a ResNet-34 [31]. To benchmark the performance of our cINN, we compare it to a commonly used conditional Invertible Neural (cVAE) [32] and two Bayesian methods.

## 2. Results and discussion

The power and flexibility of the cINN for this class of problems is showcased in figure 2. Two target spectra from the validation set are chosen (black solid lines in (A) and (B)). Each time,  $10^4$  device parameters are generated by sampling from the trained cINN with the spectra in (A) and (B) as conditioning inputs. The resulting distributions of parameters of the devices exhibiting these transmission spectra are shown as histograms in (C)–(F) and (G)–(J). The solid black line with the triangle on top marks the original parameter vector that has been used in the simulation. The generated parameters are then passed to a forward network which has been trained previously. The forward network is a fully connected dense network which takes device parameters as input and outputs the transmission spectrum of that device and its sole purpose is to speed up the generation process. In the Supporting Information evidence is presented that this forward network indeed models the generative process with high fidelity. The regenerated spectra show excellent agreement with the target spectrum. The mean of the  $10^4$  regenerated spectra fall exactly on top of the target and the shaded regions present the 2nd to 98th percentile region, meaning that 96% of the regenerated spectra lie within that region. The cINN correctly learns that the spectrum in A has been generated from gratings with the same periodicity on both sides of the slit (panels (C) and (E) in figure 2), while the spectrum in B requires two different periodicities (panels (G) and (I) in figure 2). Also note that, as

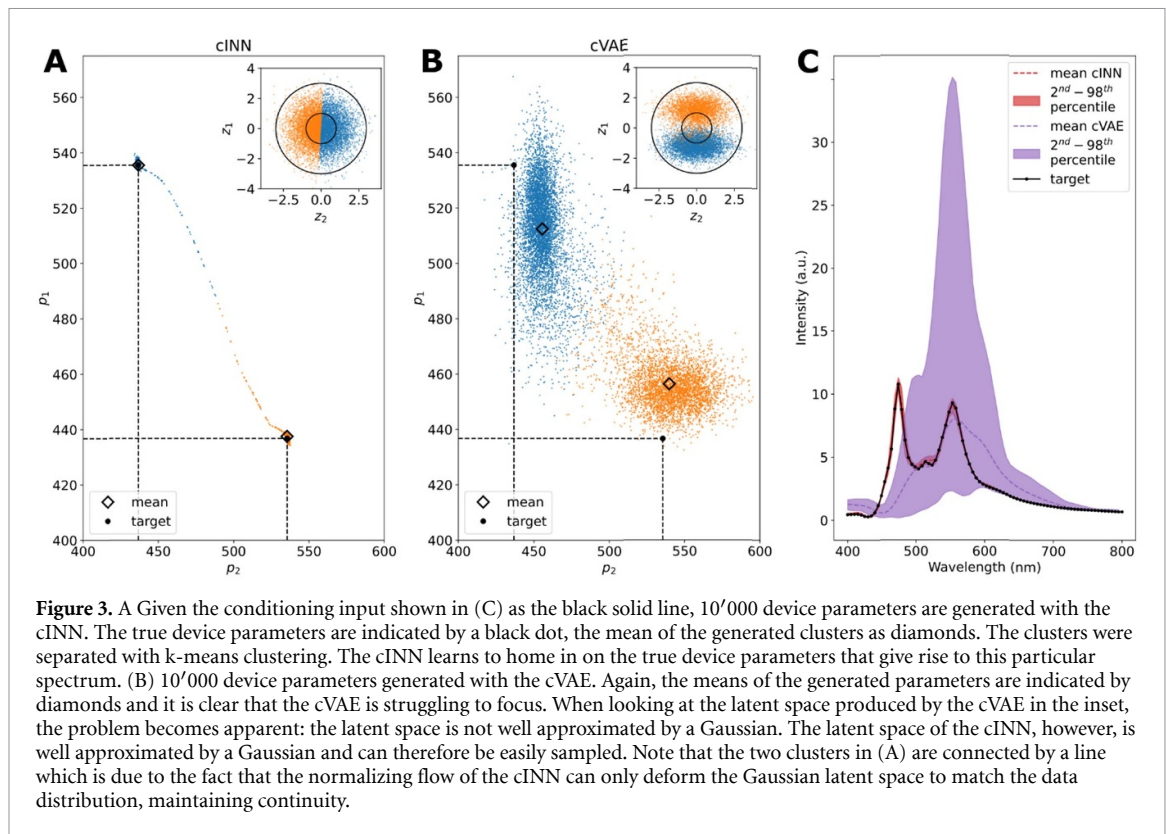


previously mentioned, the dataset was constructed randomly and the network was still able to fully capture the underlying symmetry properties of the problem. It is also interesting to note that in panels (D) and (F) the network generated two connected peaks for the groove depth. Since the regenerated spectra show excellent agreement it seems that the network is not so sensitive to groove depth for that specific conditioning input. In panels (H) and (G) of figure 2, the proposed groove depths are clearly separated.

The added flexibility of the normalizing flow is apparent when comparing the latent space generated by the cINN with the latent space of a cVAE [18, 32] in figure 3. In our example structure, the latent space has dimensionality 4 and we chose one spectrum as conditioning input and sampled 10 000 samples as before. Figure 3 shows scatter plots of generated periodicities  $p_1$  and  $p_2$  for the cINN (figure 3(A)) and the cVAE (figure 3(B)). In the case of the cINN, the vast majority of solutions concentrates around the target values for  $p_1$  and  $p_2$ , demonstrating the excellent performance of this network. Please note that the two clusters are connected by a small number of points due to the fact that the specific normalizing flow that we used maintains continuity and cannot split a unimodal base distribution. The result, however, is a good enough approximation, since these points correspond to 400 samples out of  $10^4$  or 4%. The cVAE, however, has a much larger spread in the latent space, showing both less precise and less accurate performance as compared to the cINN.

To understand the large difference in the performance of the cINN and the cVAE, we can have a closer look at their respective latent spaces shown as the insets of figures 3(A) and (B). The complete latent space is reproduced in the Supporting Information, here only  $z_1$  and  $z_2$  are shown since they capture the relevant information. While the latent space of the cINN for the two solutions is a Gaussian, the cVAE is far from normally distributed. The resulting limitation for sampling becomes obvious: With the cINN we can directly sample from a Gaussian with zero mean and a variance of one in the latent space and obtain accurate results from the inverse pass of the cINN. If samples are drawn with the same method from the latent space of the cVAE, a large number of out-of-distribution samples are generated, resulting in incorrect predictions. In both cases the generated parameter vectors were used to re-simulate the transmission spectra of the proposed





devices as shown in figure 3(C). The cINN shows excellent agreement as before, while the results from the cVAE are neither accurate nor precise. Please note there are techniques to alleviate this problem by sampling from the aggregated posterior as suggested by Tomczak and Welling [33], or by generating another VAE for sampling as suggested by Dai and Wipf [34], but that would add an additional layer of complexity. As mentioned before, normalizing flows as described in Kingma *et al* [19], would theoretically allow to model more flexible posterior distributions but add more complexity as well. In contrast, the cINN offers a powerful yet simple framework to model complex data distributions.

At this point it is important to highlight how the model assumptions and loss functions are hindering training on multimodal distributions or even leading it to failure. First, consider the commonly used mean squared error (MSE) which is used in a regression setting. When using MSE in this context the implicit assumption is made that there is a one-to-one mapping between devices and responses. However, if two devices map to the same response, the gradients propagated through the network will point into different directions and the network will not be able to choose between the two and the result will be a solution that is neither or. As discussed earlier, a cVAE struggles with multimodal datasets because the Gaussian for the posterior, while being easy to sample, is very restrictive and the model ends up lacking expressive power [35] or the sampling becomes more tedious. And while a cVAE can be augmented with normalizing flows to model multimodal distributions, this addition makes them harder to train and will use more parameters. Using a mixture of Gaussians [36] for the latent space or a VampPrior [33] allows to build more expressive posteriors but also sacrifices simplicity and ease to train. The cINN, by construction, leads itself to the modeling of complicated data distributions, no assumptions need to be made about the posterior and finally the same network is used for the forward and inverse pass, reducing the amount of parameters compared to the cVAE.

In addition to the comparison with the cVAE we have also compared our cINN approach with traditional Bayesian methods. The first Bayesian method is approximate Bayesian computation with a sequential Monte Carlo sampler [37], which is likelihood-free and therefore conceptionally closer to our neural network (NN) approaches. The second method is based on nested sampling which is the state-of-the-art in model comparison and parameter estimation but requires an approximation of the likelihood [38–40]. Even though both methods recover valid parameter distributions including multimodality, the computational cost is exceedingly high. Both Bayesian methods are iterative by construction and require large numbers of function evaluations on the order of the size of the training set for the NNs to recover the parameters of just one target spectrum. A detailed comparison between the different methods is provided in the supporting information.

### 3. Conclusion

In summary, we have shown how a multimodal device distribution can lead to pitfalls with commonly used generative models to learn these device distributions. Furthermore, we demonstrated the flexibility of cINNs how they can, with no additional knowledge about the device distribution, learn a mapping that can be used to sample new structures, providing the full posterior of the device distribution, meaning all the possible solutions to the inverse design problem. In general, adopting a probabilistic approach provides a more complete picture of all the possible solutions to an inverse design problem and how confident the algorithm is. Here we note that very recently the use of cINNs has been mentioned in the context of benchmarking different deep learning approaches to inverse models for designing artificial electromagnetic materials [41], but without considering multimodal device distributions. On a similar note, GANs have also been employed for inverse photonic design [10, 17] but have not been thoroughly explored to generate distributions of devices in the context of generative modeling.

The cINN provides solutions with high accuracy and precision on all solutions in the design space, whereas the cVAE only captures the fact that the solution space is multimodal, but with low precision and accuracy. The reason for that is the limited flexibility unimodal Gaussian latent variables provide. While flow-based models and more expressive priors could alleviate that problem, the cINN offers a simple and straightforward framework and simple to train solution. Finally, it is important to emphasize that while we have focused on the slit flanked by periodic corrugations as a proof-of-principle problem the advantage of the cINN in solving multimodal device distributions is generic and may help solve a large variety of inverse design problems to explore the design space of nanophotonic devices.

### 4. Methods

We implemented a cINN following the approach of Ardizzone *et al* [27], using their FrEIA framework. As shown schematically in figure 1(E), the basic building block of the cINN is the (conditional) affine coupling block first proposed by Dinh *et al* [28]. The network models a change of variables that maps a latent variable  $z$  to a sample  $x$  with the conditional input  $c$ :

$$p_x(f(z, c)) = p_z(z) \left| \det \frac{\partial f(z, c)}{\partial z} \right|^{-1}. \quad (1)$$

The conditioning input, in this case the spectrum  $y$  that corresponds to a certain device parameter vector  $x$ , is fed through an additional conditioning network consisting of a ResNet-34 [31] to extract meaningful features from the spectrum. More details for the hyperparameters and data preparation of the network can be found in the supporting information.

The cVAE used for comparison is implemented similarly to Ma *et al* [18] and Sohn *et al* [32] and consists of two simple dense networks with five hidden layers and 256 neurons each. The conditioning input is concatenated to the input of the encoder and decoder, respectively. One network maps the device parameter along with the simulated spectrum to the latent variable  $z$  (encoder), while the other network tries to reconstruct the parameter vector (decoder) from the latent variable  $z$  and the simulated spectrum.

The main difference between the two networks resides in the way they are trained. The cINN is trained with a modified maximum likelihood loss function including the Jacobian from the change of variables as suggested by Kruse *et al* [42]:

$$\mathcal{L}(z) = \frac{z^2}{2} - \log |\det J_{x \rightarrow z}| \quad (2)$$

while the cVAE is trained with the evidence lower bound (ELBO) [32]:

$$\mathcal{L}(z, y) = -D_{\text{KL}}(q_\theta(z|x_i, y_i) || p(z|y)) + E_{z \sim q_\theta(z|x_i, y_i)} [\log p_\phi(x_i|z, y_i)] \quad (3)$$

where  $D_{\text{KL}}$  is the Kullback–Leibler divergence, which measures the difference between distributions  $q_\theta(z|x_i, y_i)$  and  $p(z|y)$ , the approximated posterior and the prior, respectively. The dependence on  $y$  in the prior  $p(z|y)$  has been dropped to simplify training as suggested by Sohn *et al* [32].

$E_{z \sim q_\theta(z|x_i, y_i)} [\log p_\phi(x_i|z, y_i)]$  is the reconstruction loss. Under the constraint that the latent space should be Gaussian, meaning that the term  $D_{\text{KL}}(q_\theta(z|x_i, y_i) || p(z|y))$  acts as a regularization term [43, 44]. If  $q_\theta(z|x_i, y_i)$

is a Gaussian with mean  $\mu_q$  and variance  $\sigma_q$  and  $p(z|y)$  a Gaussian with mean  $\mu_p$  and variance  $\sigma_p$ , the Kullback–Liebler divergence simplifies to its well known form [16, 44]:

$$-D_{\text{KL}}(q_{\theta}(z|x_i, y_i) || p(z|y)) = \frac{1}{2} [1 + \log \sigma_q^2 - \sigma_q^2 - \mu_q^2] \quad (4)$$

where  $\mu_q$  and variance  $\sigma_q$  are estimated by the encoder network [16].

The loss functions in equations (2) and (3) are very different, reflecting the different objectives to minimize. While the goal of equation (2) is to maximize the likelihood after the change of variables, equation (3) tries to minimize reconstruction loss while imposing the restriction of a normal distribution on the latent space, like a regularization term [43]. Since equation (2) is a form of maximum likelihood loss function, mode collapse is virtually impossible [29]. The objective of the ELBO is to learn a representation of the data that is as close as possible to the true distribution which is especially useful when the data is very high dimensional as normalizing flows typically do not scale well to high dimensional data such as images.

Even though we have access to relatively fast simulations to obtain the device response, we found it beneficial to train a simple fully connected model consisting of three hidden layers with 512 neurons each and ReLU activation for to predict the device response for a given parameter vector. This is not necessary but useful to quickly visualize the behavior of the devices generated by the cINN and the cVAE.

### Data availability statement

Codes to load the data and build, train and evaluate the models, as well as pretrained weights are provided here: [https://github.com/phond-uam/tackling\\_multimodality](https://github.com/phond-uam/tackling_multimodality).

The data that support the findings of this study are openly available at the following URL/DOI: <https://drive.google.com/file/d/1Z1yH2B0YC13l6w2tYq9HcAvy9qyLDLI/view?usp=sharing>.

### Acknowledgments

We graciously acknowledge the help of Lynton Ardizzone with feedback and suggestions on implementing our own cINN and helpful discussions to understand the INN framework. Furthermore, we want to thank Pablo Sanchez-Martin for helpful discussions about VAEs in general.

### Funding sources

The project that gave rise to these results received the support of a fellowship from the ‘la Caixa’ Foundation (ID 100010434). The fellowship code for M F is LCF/BQ/DI18/11660037. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant agreement No. 713673. The work was further supported by the Spanish Ministry for Science, Innovation, and Universities through the Europa Excelencia program (EUR2019-103826). We acknowledge the support from the ‘(MAD2D-CM)-UAM’ project funded by Comunidad de Madrid, by the Recovery, Transformation and Resilience Plan, and by NextGenerationEU from the European Union.

### Author contributions

The manuscript was written by M F with suggestions from F P and J B A. M F setup and trained the Neural Networks shown. J B A wrote the original codes to simulate the investigated devices and generated the training set based on device parameters selected by M F. F P and J B A provided helpful suggestions and discussions throughout the whole process.

### Conflict of interest

The authors declare that they have no competing interests.

### ORCID iDs

Michel Frising  <https://orcid.org/0000-0003-3725-4824>

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

Ferry Prins  <https://orcid.org/0000-0001-7605-1566>

## References

- [1] Wambold R A, Yu Z, Xiao Y, Bachman B, Jaffe G, Kolkowitz S, Choy J T, Eriksson M A, Hamers R J and Kats M A 2021 Adjoint-optimized nanoscale light extractor for nitrogen-vacancy centers in diamond *Nanophotonics* **10** 393–401
- [2] Lin D, Melli M, Poliakov E, Hilaire P S, Dhuey S, Peroz C, Cabrini S, Brongersma M and Klug M 2017 Optical metasurfaces for high angle steering at visible wavelengths *Sci. Rep.* **7** 2286
- [3] Lu J and Vučković J 2013 Nanophotonic computational design *Opt. Express* **21** 13351–67
- [4] Jensen J S and Sigmund O 2011 Topology optimization for nano-photonics *Laser Photon. Rev.* **5** 308–21
- [5] Sell D, Yang J, Doshay S, Yang R and Fan J A 2017 Large-angle multifunctional metagratings based on freeform multimode geometries *Nano Lett.* **17** 3752–7
- [6] Huntington M D, Lauhon L J and Odom T W 2014 Subwavelength lattice optics by evolutionary design *Nano Lett.* **14** 7195–200
- [7] Kennedy J and Eberhart R 1995 Particle swarm optimization *Proc. ICNN'95—Int. Conf. on Neural Networks* vol 4 (Perth, WA: IEEE) pp 1942–8.
- [8] So S, Badloe T, Noh J, Bravo-Abad J and Rho J 2020 Deep learning enabled inverse design in nanophotonics *Nanophotonics* **9** 1041–57
- [9] Liu D, Tan Y, Khoram E and Yu Z 2018 Training deep neural networks for the inverse design of nanophotonic structures *ACS Photonics* **5** 1365–9
- [10] 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** 6570–6
- [11] Peurifoy J, Shen Y, Jing L, Yang Y, Cano-Renteria F, DeLacy B G, Joannopoulos J D, Tegmark M and Solja M 2018 Nanophotonic particle simulation and inverse design using artificial neural networks *Sci. Adv.* **4** 8
- [12] Jiang J and Fan J A 2019 Simulator-based training of generative neural networks for the inverse design of metasurfaces *Nanophotonics* **9** 1059–69
- [13] Jiang J and Fan J A 2019 Global optimization of dielectric metasurfaces using a physics-driven neural network *Nano Lett.* **19** 5366–72
- [14] Martín-Moreno L, García-Vidal F J, Lezec H J, Pellerin K M, Thio T, Pendry J B and Ebbesen T W 2001 Theory of extraordinary optical transmission through subwavelength hole arrays *Phys. Rev. Lett.* **86** 1114–7
- [15] García-Vidal F J, Martín-Moreno L, Ebbesen T W and Kuipers L 2010 Light passing through subwavelength apertures *Rev. Mod. Phys.* **82** 729–87
- [16] Kingma D P and Welling M 2014 Auto-encoding variational Bayes (arXiv:1312.6114) (Accessed 18 January 2022)
- [17] So S and Rho J 2019 Designing nanophotonic structures using conditional deep convolutional generative adversarial networks *Nanophotonics* **8** 1255–61
- [18] 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** 1901111
- [19] Kingma D P, Salimans T, Jozefowicz R, Chen X, Sutskever I and Welling M 2016 Improved variational inference with inverse autoregressive flow *29th Conf. on Neural Information Processing Systems NIPS 2016 (Barcelona, Spain)* pp 4743–51
- [20] Kingma D P and Dhariwal P 2018 Glow: generative flow with invertible 1x1 convolutions (arXiv:1807.03039) (Accessed 29 November 2020)
- [21] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S and Lakshminarayanan B 2019 Normalizing flows for probabilistic modeling and inference (arXiv:1912.02762) (Accessed 29 November 2020) (<https://doi.org/10.1063/5.0018903>)
- [22] Rezende D J and Mohamed S 2016 Variational inference with normalizing flows (arXiv:1912.02762) (Accessed 18 January 2022)
- [23] De Cao N, Titov I and Aziz W 2019 Block neural autoregressive flow (arXiv:1904.04676)
- [24] Mirza M and Osindero S 2014 Conditional generative adversarial nets (arXiv:1411.1784)
- [25] Goodfellow I 2017 NIPS 2016 Tutorial: generative adversarial networks (arXiv:1701.00160) (Accessed 01 December 2022)
- [26] Ardizzone L, Kruse J, Wirkert S, Rahner D, Pellegrini E W, Klessen R S, Maier-Hein L, Rother C and Köthe U 2019 Analyzing inverse problems with invertible neural networks (arXiv:1808.04730) (<https://doi.org/10.1007/s11548-019-01939-9>)
- [27] Ardizzone L, Kruse J, Lüth C, Bracher N, Rother C and Köthe U 2021 Conditional invertible neural networks for diverse image-to-image translation *Pattern Recognition (Pattern Recognition)* ed Z Akata, A Geiger and T Sattler (Cham: Springer) vol 12544 pp 373–87
- [28] Dinh L, Sohl-Dickstein J and Bengio S 2017 Density estimation using real NVP (arXiv:1605.08803) (Accessed 1 December 2020)
- [29] Ardizzone L, Lüth C, Kruse J, Rother C and Köthe U 2019 Guided image generation with conditional invertible neural networks (arXiv:1907.02392) (Accessed 26 January 2022) (<https://doi.org/10.1109/TNNLS.2020.3042395>)
- [30] Bravo-Abad J, Martín-Moreno L and García-Vidal F J 2006 Resonant transmission of light through subwavelength holes in thick metal films *IEEE J. Sel. Top. Quantum Electron.* **12** 1221–7
- [31] He K, Zhang X, Ren S and Sun J 2015 Deep residual learning for image recognition (arXiv:1512.03385) (Accessed 21 August 2020)
- [32] Sohn K, Lee H and Yan X 2015 Learning structured output representation using deep conditional generative models *28th Conf. on Neural Information Processing Systems NIPS 2015 (Montréal, Canada)* pp 3483–91
- [33] Tomczak J M and Welling M 2018 VAE with a VampPrior (arXiv:1705.07120) (Accessed 1 April 2022)
- [34] Dai B and Wipf D 2019 Diagnosing and enhancing VAE models (arXiv:1903.05789) p 12
- [35] Turner R E and Sahani M 2011 Two problems with variational expectation maximisation for time series models *Bayesian Time Series Models* ed D Barber, A T Cemgil and S Chiappa (Cambridge: Cambridge University Press) pp 104–24
- [36] Dilokthanakul N, Mediano P A M, Garnelo M, Lee M C H, Salimbeni H, Arulkumaran K and Shanahan M 2017 Deep unsupervised clustering with Gaussian mixture variational autoencoders (arXiv:1611.02648) p 13 (Accessed 9 April 2022)
- [37] Schälte Y, Klinger E, Alamoudi E and Hasenauer J 2022 PyABC: efficient and robust easy-to-use approximate Bayesian computation *J. Open Source Softw.* **7** 4304
- [38] Buchner J 2016 A statistical test for nested sampling algorithms *Stat. Comput.* **26** 383–92
- [39] Buchner J 2019 Collaborative nested sampling: big data vs. complex physical models *Publ. Astron. Soc. Pac.* **131** 108005
- [40] Buchner J 2021 UltraNest—a robust, general purpose Bayesian inference engine *J. Open Source Softw.* **6** 3001
- [41] Ren S, Mahendra A, Khatib O, Deng Y, Padilla W J and Malof J M 2022 Inverse deep learning methods and benchmarks for artificial electromagnetic material design *Nanoscale* **14** 3958–69
- [42] Kruse J, Ardizzone L, Rother C and Köthe U 2021 Benchmarking invertible architectures on inverse problems (arXiv:2101.10763) (Accessed 13 May 2021)
- [43] Doersch C 2021 Tutorial on variational autoencoders (arXiv:1606.05908) (Accessed 18 January 2022)
- [44] Odaibo S 2019 Tutorial: deriving the standard variational autoencoder (VAE) loss function (arXiv:1907.08956) (Accessed 3 March 2022)