Generating extreme quantum scattering in graphene with machine learning

Chen-Di Han [1] and Ying-Cheng Lai [1,2,*]

[1] *School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA*

[2] *Department of Physics, Arizona State University, Tempe, Arizona 85287, USA*

(Received 1 September 2022; revised 5 December 2022; accepted 7 December 2022; published 14 December 2022)

Graphene quantum dots provide a platform for manipulating electron behaviors in two-dimensional (2D) Dirac materials. Most previous works were of the "forward" type in that the objective was to solve various confinement, transport, and scattering problems with given structures that can be generated by, e.g., applying an external electrical field. There are applications such as cloaking or superscattering where the challenging problem of inverse design needs to be solved: Finding a quantum-dot structure according to certain desired functional characteristics. A brute-force search of the system configuration based directly on the solutions of the Dirac equation is computationally infeasible. We articulate a machine-learning approach to addressing the inverse-design problem where artificial neural networks subject to physical constraints are exploited to replace the rigorous Dirac equation solver. In particular, we focus on the problem of designing a quantum-dot structure to generate both cloaking and superscattering in terms of the scattering efficiency as a function of the energy. We construct a physical loss function that enables accurate prediction of the scattering characteristics. We demonstrate that, in the regime of Klein tunneling, the scattering efficiency can be designed to vary over two orders of magnitude, allowing any scattering curve to be generated from a proper combination of the gate potentials. Our physics-based machine-learning approach can be a powerful design tool for 2D Dirac material-based electronics.

DOI: 10.1103/PhysRevB.106.214307

I. INTRODUCTION

Two-dimensional (2D) Dirac material systems such as graphene, topological insulators, molybdenum disulfide, and topological Dirac semimetals [1–5] constitute a research frontier in condensed matter physics and materials science. A common feature of these materials is that their energy bands contain a Dirac cone structure that gives rise to a linear energy-momentum relation (dispersion relation) for low-energy excitations, which is characteristic of relativistic quantum motions governed by the Dirac equation. The Dirac cone structure can be exploited for storing and transferring information with applications in Dirac material-based electronics and spintronics [6–9].

Given a 2D Dirac material, the system structure and the applied gate potentials, a combination of these can generate rich quantum behaviors. For example, in the field of relativistic quantum chaos [10–13], electrons can be confined in a graphene sheet of certain geometric shape, generating billiard systems that provide a platform to study the relativistic quantum manifestations of different kinds of classical dynamics. In scattering, a combination of gate potentials can be applied to a graphene sheet to create a quantum-dot structure, which can be experimentally realized through STM tips or doping [14–19]. Graphene quantum dots are also a paradigm to study various transport phenomena [20–22]. Given a scattering structure, the tight-binding Hamiltonian or the Dirac equation can be solved to yield the scattering functions so that the performance of the device can be assessed. A common feature among the previous works is their "forward" nature: Studying various quantum behaviors or performance of the underlying system with a *given* structure.

Inverse design addresses the opposite problem: How to design a Dirac material system to generate certain desired functional characteristics. For example, suppose we wish to design a graphene device to generate the desired scattering functions by using a 2D multilayer structure of concentric circular graphene ribbons, where the layers can have different widths and each layer is made physically distinct from the others through a vertical electric field (gate potential). A brute-force approach to searching the optimal multilayer structure to generate the desired scattering functions would be to test a large number of combinations of the geometric parameters of the various layers as well as the values of the gate potentials. With luck, it may be possible to find a specific structure and a set of gate potentials such that the resulting scattering curves approximately match the desired functions. However, such a brute-force approach is generally practically infeasible because of the extremely large parameter space that needs to be searched. There is in fact no guarantee that this approach would be successful, as the existence of a device with the desired response is unknown *a priori*. Aside from quantum scattering [23], inverse-design problems of this kind occur in other fields such as quantum information [24], biology [25], molecular design [26], and photonics [27].

Recently, machine learning has been introduced to inverse optical design where the goal is to find the best

*Ying-Cheng.Lai@asu.edu

structure of a multilayer dielectric sphere to generate the desired electromagnetic response by approximating the Maxwell equations with a trained artificial neural network [28]. The approximation can simplify the original optimization problem without loss of details. Subsequently, the approach has been extended to designing optical metasurfaces [29], metagrating [30], and multifunctional devices [31,32]. Machine learning has thus offered a general solution to the inverse-design problem.

This work concerns designing systems that can generate *extreme* scattering characteristics. In particular, the performance of any scattering system, optical or electronic, can be conveniently characterized by the scattering cross section as a function of a basic quantity such as the wavelength (in optics) or the Fermi energy (in electronic systems). The two opposite extremes correspond to a small and a large scattering cross section, respectively, with the former representing cloaking and the latter signifying superscattering. In optics, one approach to cloaking is through scattering cancellation [33] based on the the idea that, in a multilayer structure, polarization from different layers can cancel each other to generate an exceedingly small scattering cross section. Multilayer structures can also be exploited to produce superscattering [34]. There were experimental results on cloaking or superscattering in optics [34–36]. For those problems, a basic physical constraint is that it is generally not possible to generate cloaking or superscattering for all kinds of incident waves [37,38], so these exotic phenomena can occur only for a certain type of incident waves. In optics, both cloaking and superscattering can occur in a multilayer structure of dielectric materials, so it is possible to generate cloaking and superscattering in the same device. For example, it has been recently demonstrated that a single device based on Ag-semiconductor multilayer sphere can produce large and small scattering cross sections [28,39]. Without changing the dimension of the device, a superscattering-cloaking transition was reported in a two-layer system where a significant change in the scattering cross section can occur with respect to the incident angle [40]. One optical material capable of such a transition is GeTe [41,42], where, under different environmental conditions such as the temperature [43], the material exhibits different physical properties. Related optical scattering problems were investigated with graphene-coated nanosphere [44,45] and its design [46]. The analogy between matter and electromagnetic waves stimulated the idea of generating electron cloaking through a multilayer structure [47–49], which subsequently was extended to graphene for cloaking relativistic electrons [50–52].

Our study focuses on the following question: Is it possible to generate the two extreme scattering behaviors, cloaking and superscattering, in the same 2D Dirac material system? Because of the relative easiness to experimentally tune the gate potential through varying an applied electrical field [53], graphene stands out as a viable candidate for generating superscattering and cloaking as well as a transition between them in the same fixed structure. To be concrete, we consider a multilayer scattering structure that consists of N concentric circles, where a different gate potential is applied to each distinct circular layer. To reduce the dimension of the parameter space so as to make the inverse-design problem feasible, we fix the geometric structure of the scattering system and allow the set of gate potentials to be tunable. We develop a class of neural networks subject to physical constraints with the goal to find a set of gate potentials to generate electronic cloaking and another set to generate superscattering in the *same device*. A key component of the inverse design is our articulation of a physical loss function to significantly reduce the training and testing errors and to eliminate nonphysical solutions. We demonstrate that, even in the regime of Klein tunneling, the scattering efficiency (to be defined in Sec. II A) can vary over two orders of magnitude. Compared with the corresponding optical system where the material for each layer is fixed and the width for each shell can be changed, in our graphene scattering system only the gate potential is changed with the device structure intact, rendering it experimentally feasible. Our physics-based machine-learning approach can be a powerful design tool for graphene-based electronic devices and is generalizable to solving inverse problems in other areas of science.

II. METHODS: MULTILAYER GRAPHENE QUANTUM DOT, MACHINE LEARNING, AND INVERSE DESIGN

We first clarify the physical meanings of the terms "cloaking" and "superscattering" adopted in this work. Our goal of inverse design is to realize a structure that can generate these two opposite extremes of quantum scattering. Ideal cloaking is characterized by zero scattering cross section or efficiency. Since near-zero scattering efficiency can trivially occur for a scatterer of arbitrarily weak strength $\bar{V}\bar{R} \to 0$, where $\bar{V}$ and $\bar{R}$ are the average gate potential and the mean size of the quantum dot, respectively, cloaking is meaningful only for scatterers with a reasonably large scattering strength, e.g., the quantum dot of size R_B in Fig. 1(b). For this case to be qualified as cloaking, the resulting scattering cross section should be as small as that from a scatterer of much smaller size such as the dot of radius R_C in Fig. 1(b). In the opposite limit, strong scattering can naturally occur if the scattering strength is high. By superscattering we mean that, even when the scatterer is relatively weak, the resulting scattering cross section can be as large as that from a scatterer of much larger size, such as the quantum dot of size R_A in Fig. 1(b).

A. Graphene quantum dot

We consider a 2D circular graphene quantum dot of radius R_N subject to a gate potential profile $V(\mathbf{r})$. The Hamiltonian of the scattering system is

$$H = v_g \boldsymbol{\sigma} \cdot \mathbf{p} + V(\mathbf{r}), \tag{1}$$

where v_g is the Fermi velocity and $\boldsymbol{\sigma} \equiv [\sigma_x, \sigma_y]$ is the vector of Pauli matrices. For simplicity, we assume that the gate potential has no angular dependence and write $V(\mathbf{r}) = V(r)$, for $0 < r \leqslant R_N$. We further assume that $V(r)$ is a piecewise constant function with N distinct values, which experimentally can be realized by STM tips or doping in the substrate

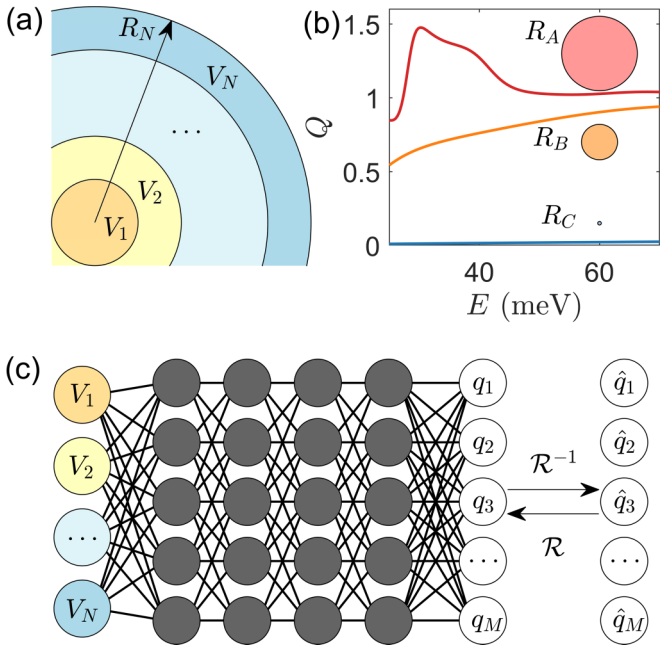

FIG. 1. Illustration of a multilayer graphene quantum dot, scattering efficiency, and the physics-constrained machine-learning scheme. (a) A multilayer circular graphene quantum dot generated by the gate potential profile $V(r)$ in Eq. (2). (b) For the special case of a single-layer quantum dot, scattering efficiency versus the Fermi energy E for three different sizes of the quantum dot: $R_A = 63$ nm, $R_B = 30$ nm, and $R_c = 3$ nm. Because of the simple structure, the overall scattering efficiency increases as the dot becomes larger, as expected, indicating naturally that a single-layer structure is not capable of generating extreme scattering behaviors. (c) The neural-network architecture underlying our proposed scheme of physics-constrained machine learning. The basic component of the neural network is a multilayer perceptron, where the input is the set of constant gate potential values $(V_1, V_2, \ldots, V_N)$ and the output is the properly discretized scattering efficiency function as exemplified in (b). The function is defined in a fixed energy range and is uniformly sampled with M discrete points, leading to an M-dimensional output vector: $\boldsymbol{Q} \equiv (q_1, q_2, \ldots, q_M)^T$. The rightmost layer with the output vector $\hat{\boldsymbol{Q}} \equiv (\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)^T$ is introduced to ensure that the whole neural-network architecture "respects" the basic physics of quantum scattering (see text for details).

[14–19]. The potential profile is

$$V(r) = \begin{cases} V_1, & r < R_1 \\ V_2, & R_1 < r < R_2 \\ \cdots & \cdots \\ V_N, & R_{N-1} < r < R_N \\ 0, & r > R_N \end{cases} \tag{2}$$

leading to a multilayer scattering structure with N distinct concentric circular layers, where the gate potential in each layer is a constant, as schematically illustrated in Fig. 1(a). An advantage of the multilayer structure is that the structural parameters $(\mathbf{R}, \mathbf{V}) \equiv ([R_1, R_2, \ldots, R_N]^T, [V_1, V_2, \ldots, V_N]^T)$ can be relatively readily adjusted in experiments. This should be contrasted to a multilayer structure in optics [28–32], where only the set of radii can be changed once the dielectric materials are fixed.

Our goal is to design the potential profile $V(r)$ to generate extreme scattering behaviors for a fixed set of radii $(R_1, R_2, \ldots, R_N)$. To gain insights, we first consider the relatively simple case of a single-layer scatterer: $N = 1$. Matching the boundary condition at $r = R_1$ and setting the Fermi velocity to be $v_g = 10^6$ m/s, we obtain the solutions of the 2D Dirac equation [54] in terms of the scattering spinor wave function (Appendix A). The scattering behavior can be conveniently characterized by the scattering efficiency defined as

$$Q = \frac{\text{scattering cross section}}{\text{geometric size}} = \frac{2}{kR_1} \sum_{l=-\infty}^{\infty} |A_l|^2, \tag{3}$$

where A_l is the coefficient for scattering wave in the polar coordinates associated with angular momentum l. Intuitively, a large scattering efficiency indicates that the object is more "visible." Three examples of the scattering efficiency are shown in Fig. 1(b) for $V_1 = 87.5$ meV and for dot size $R_A = 63$ nm, $R_B = 30$ nm, and $R_C = 3$ nm, corresponding to strong, intermediate, and weak (all in the relative sense) scattering, respectively. Note that the Q value for R_B is approximately two orders of magnitude larger than that for R_C. Note also that the incident energy is approximately one-quarter to three-quarter of the potential height, so the particle is in the Klein tunneling regime [55]. A recent work [56] indicated that, for a graphene quantum dot, in the limit $(kR_1) \to 0$, for a fixed Fermi energy the scattering efficiency scales with the size of the scatterer as $Q \propto (VR_1)^3$. However, the scaling breaks down for $R_1 \approx 10$ nm at which the scattering efficiency tends to saturate and increases only slowly with further increase in R. Overall, Fig. 1(b) indicates that, as the size of the scatterer increases, scattering is enhanced, as can be intuitively anticipated. The simple single-layer structure is thus not capable of generating extreme scattering behaviors.

For a graphene multilayer circular quantum dot with given parameters $(\mathbf{R}, \mathbf{V})$, the scattering problem can be solved by the method of transfer matrix [57] through matching the spinor wave functions at all the layer boundaries (Appendix A). After the spinor wave functions in all layers have been obtained, we can calculate the scattering efficiency according to

$$Q = \frac{2}{k_{N+1}R_N} \sum_{l=-\infty}^{\infty} \left|A_l^{N+1}\right|^2, \tag{4}$$

where $k_N \equiv |E|/v_g$. Given a scattering structure, it is thus straightforward to calculate the scattering efficiency (the solutions of the forward problem).

To generate extreme scattering behaviors, i.e., cloaking or superscattering, it is necessary to define the input properly by specifying a range for the gate potentials because, if the potentials are too weak, the scattering efficiency will be trivially near zero and this does not lead to cloaking. Likewise, unreasonably large potential values can lead to strong scattering, but this is not superscattering. Making use of the scaling relation for the scattering efficiency for weak scatterers [56],

$$Q \propto \left[\sum |V_i|(R_{i+1} - R_i)\right]^3,$$

we have $Q \propto (\langle|\mathbf{V}|\rangle R_B)^3$ because $R_i = iR_N/N$. It is thus reasonable to choose the gate potentials such that $\langle|\mathbf{V}|\rangle \approx V_0$,

where $\langle|\mathbf{V}|\rangle$ is the average gate potential applied to the multilayer structure and V_0 is the potential applied to a single-layer structure of the same size.

For a single-layer scatterer, the problem of finding a scattering configuration to generate the desired scattering efficiency Q_{desired} can be formulated as the following optimization problem:

$$\min_{\mathbf{V}} \left\| \frac{Q(\mathbf{V}) - Q_{\text{desired}}}{Q_{\text{desired}}} \right\|^2, \tag{5}$$

where $\|\cdot\|^2$ denotes the mean-square error (MSE). Because of the simple geometry, it is only necessary to optimize a small number of parameters, making it relatively straightforward to obtain the solutions of the optimization problem. In particular, it is only necessary to discretize each parameter dimension and perform a grid search. Difficulties arise when the scatterer has many layers, making the function $Q(\mathbf{V})$ highly nonlinear with many parameters that need to be optimized. For example, for a multilayer sphere of 10 layers, even if only four parameters need to be determined for each layer, the number of parameter combinations will be larger than $4^{10} \sim 10^6$. This issue of computational complexity motivated us to exploit machine learning to solve the optimization problem.

B. Physics-constrained machine learning

Machine learning provides an effective platform to approximate mathematical functions arising from physics [58]. For example, in photonic design, neural networks can be used to approximately represent the scattering cross section from a variety of devices [28,29]. To exploit machine learning to generate extreme scattering behaviors in a graphene quantum dot, an essential step is to train a multilayer perceptron such that it mimics the scattering process. To achieve this, we introduce a physics-constrained neural network architecture, as illustrated in Fig. 1(c). The concrete design of the architecture is as follows. The input vector has the dimension N, a set of constant gate potential values $\boldsymbol{V} \equiv (V_1, V_2, \ldots, V_N)^T$. The output is the scattering efficiency function as exemplified in Fig. 1(b), which is defined in a given energy range and is uniformly sampled with M discrete points, leading to an M-dimensional output vector: $\boldsymbol{Q} \equiv (q_1, q_2, \ldots, q_M)^T$. In this work, we use $M = 200$ (a rather arbitrary choice). Aside from the input and output layers, the neural network has four hidden layers, each containing 200 neurons. There are thus five layer-to-layer transforms from the input to the output. The activation for the first four transforms is set to be the standard rectified linear unit (ReLU) and there is no activation from the fourth hidden layer to the output layer (a simple linear transform). Note that, before any activation, the neural network uses a linear transformation between two adjacent layers, which has approximately 4×10^4 parameters. Thus, the number of training parameters in the architecture from the input vector $\boldsymbol{V}$ to the output vector $\boldsymbol{Q}$ is approximately 10^5.

To train the neural network, a certain amount of ground truth data from the graphene scattering system is required. To generate the data, it is necessary to determine the range of the input, which in our case is the range of the externally applied gate potential. To be concrete, we assume that the mean value of the gate potential in the whole quantum-dot scattering region is $\langle V(r)\rangle \approx V_0$, where $V_0 = 87.5$ meV, so we choose $|V_i| \in (75$ and 100 meV). The training data are generated by randomly select V_i in this range, where V_i can be either positive or negative, i.e., -100 meV $< V_i < -75$ meV or 75 meV $< V_i <$ 100 meV. For a larger number N of layers, the multilayer quantum-dot structure is geometrically more sophisticated, thereby requiring more training data. Empirically, we choose the number of training data points to be $10^3 N$. Say the scattering structure has 10 circular layers: $N = 10$. Compared with a brute-force grid search in 10 dimensions, our choice of 10^4 data points corresponds to search in a parameter space of a significantly reduced dimension: About three. Our machine-learning approach thus requires far less data amount than a brute-force search would.

The training process is rather standard and is briefly described, as follows. The training data sets are generated by directly simulating quantum scattering based on the solutions of the Dirac equation in the setting of a multilayer quantum dot. Training is conducted by employing the standard stochastic gradient descent algorithm with batch size 128. The method of adaptive momentum (Adam) is used to minimize the loss function to find the parameters. The whole process from network construction and training to predicting the scattering-efficiency function is accomplished by using the open source package TENSORFLOW and KERAS [59,60].

A key to the success of a machine-learning architecture is the training loss. A commonly used loss function is the mean-square error (MSE)

$$\mathcal{L}_{\text{MSE}} = \|Q_{\text{pred}} - Q_{\text{true}}\|^2, \tag{6}$$

which typically works well when the data values are within the same order of magnitude. For the multilayer perceptron in Fig. 1(c), the weights and biases can be solved by the optimization algorithm based on the loss function. The structure from the input vector $\boldsymbol{V}$ to the output vector $\boldsymbol{Q}$ is mathematically designed without taking into account physical constraints of the underlying quantum scattering system. As a result, even with extensive training, nonphysical results can arise, which are in fact not uncommon. For example, to realize cloaking requires that the scattering efficiency has near zero values, but some components of the output vector $\boldsymbol{Q}$ can be negative, which is not physical, as illustrated in Fig. 2(a) and the inset for a graphene quantum dot that has $N = 9$ layers. This problem cannot be fixed through training. It is thus necessary to take physical constraints into account in the design of the neural-network architecture.

Incorporating physical principles and constraints into designing neural networks has become a recent subarea of research in machine learning. For example, the Hamiltonian structure has been built into the neural-network architectures to predict the dynamical behaviors of classical mechanical systems [61–69]. Physics-based neural networks have also been extended to other fields such as fluid dynamics [70,71] and metasurface design in optical [72] or quantum systems [73,74]. For our quantum scattering system, a basic requirement is that the scattering cross sections or efficiency not be negative. We address the issue of physical constraints by incorporating an additional output layer: The rightmost layer in Fig. 1(c) with the output vector $\hat{\boldsymbol{Q}} \equiv (\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)^T$. More specifically, the additional output layer takes the vector

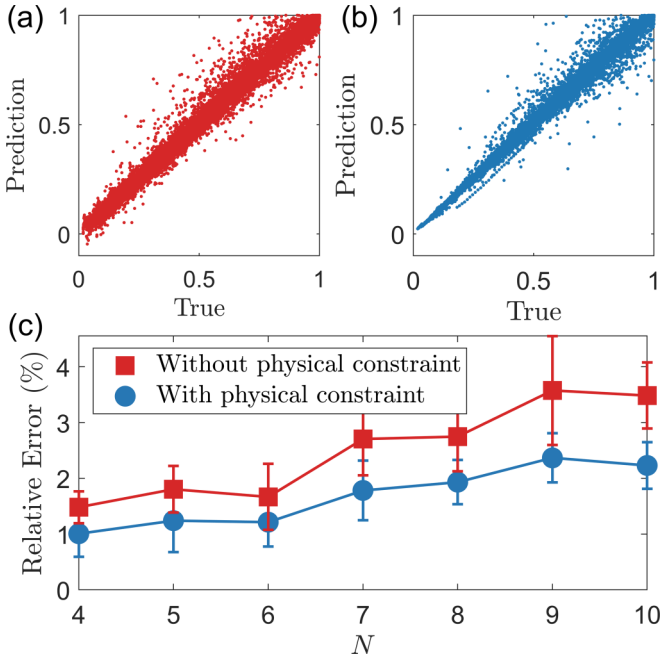

FIG. 2. The advantages of invoking a physics-constrained loss function. The performance of a given loss function can be conveniently visualized by the plot of the neural-network predicted values of the scattering efficiency function versus the true value. (a) The plot based on the mathematical loss function (6) (without physical considerations). There is a relatively wide spread of the predicted values about the true values and some predicted values of the scattering efficiency are negative (nonphysical). (b) With the physics-constrained loss function (7), the predicted values are all positive and are closer to the true values. (c) Relative prediction error versus the number of layers in the scattering structure. Compared with the case of loss function (6), the relative error associated with the physics-constrained loss function (7) is markedly reduced (by about 50%).

$\boldsymbol{Q}$ as input and performs an exponential operation on each component of $\boldsymbol{Q}$ to generate the vector $\hat{\boldsymbol{Q}}$. Mathematically, the operation can be represented as $\mathcal{R}^{-1}(\cdot) \equiv \exp(\cdot)$, which maps Q to a strictly positive function $\hat{Q}$. The consideration has led us to introduce the following two-component loss function:

$$\mathcal{L}_{\text{physical Loss}} = \|Q_{\text{pred}} - \mathcal{R}(Q_{\text{true}})\|^2 + \|\hat{Q}_{\text{pred}} - Q_{\text{true}}\|^2, \quad (7)$$

where Q_{pred} is the scattering efficiency calculated from the vector $\boldsymbol{Q}$ and $\hat{Q}_{\text{pred}} = \mathcal{R}^{-1}(Q_{\text{pred}})$. Note that the squared difference between the true Q function and the transformed function $\hat{Q}$ is the conventional loss function. The idea to enforce the physical rule of the non-negativeness of the scattering cross section is to supply an additional term: The squared difference between the transformed true output function $\mathcal{R}(Q_{\text{true}})$ and the original output function Q. From the point of view of optimization, this loss function works as follows. If Q_{true} is close to zero, the operation $\mathcal{R}$ will return a relatively large value. When Q_{true} is large, the second term in the loss function will dominate. For small and large values of Q_{true}, the predicted values of the scattering efficiency will be positive.

The performance of the physics-constrained loss function is illustrated in Fig. 2(b), where various values of the predicted scattering efficiency function are plotted against the corresponding true values. Compared with Fig. 2(a) based on the mathematical loss function in Eq. (6), we see that the predicted values are all positive (as they should be) and are closer to the true values. The prediction performance can be characterized by the following relative error:

$$\text{Relative error} = \left| \frac{Q_{\text{true}} - Q_{\text{pred}}}{Q_{\text{true}}} \right|. \quad (8)$$

Figure 2(c) shows the error versus the number of layers in the graphene quantum-dot structure, where each data point is generated from an ensemble average of 100 independent neural-network realizations and, for each realization, the testing data set contains 1000 points. Two sets of results are shown: One according to the physics-constrained loss function (7) (blue points) and another based on the loss function (6) (red). We see that, as the number of layers in the quantum dot increases, the architecture becomes more complex, leading to some increase in the relative error. However, the relative error can be markedly reduced by employing the physics-based loss function (7).

C. Principle of inverse design

The physics-constrained neural network as shown in Fig. 1(c), once trained, takes an input vector $\boldsymbol{V}$ of the gate potentials to generate an output vector $\hat{\boldsymbol{Q}}$ representing the scattering efficiency that is the solution of the 2D Dirac equation. A well-trained neural network thus effectively functions as a Dirac equation solver. The main advantage of this substitution lies in the computational efficiency, as what it takes for the well-trained neural network to generate the solutions of the Dirac equation is simply matrix multiplications through a gradient descent. More specifically, let $\boldsymbol{Q}_{\text{desired}}$ be the desired function of the scattering efficiency in the same energy range as the training data sets, represented as a vector. The objective is to solve the following optimization problem:

$$\min_{\mathbf{V}} \left\| \frac{\hat{\boldsymbol{Q}}(\mathbf{V}) - \boldsymbol{Q}_{\text{desired}}}{\boldsymbol{Q}_{\text{desired}}} \right\|^2, \quad (9)$$

where $\hat{\boldsymbol{Q}}$ is the output vector from the neural network and $\boldsymbol{Q}_{\text{desired}}$ is normalized to accommodate the behavior of the scattering efficiency on multiple energy scales. The error in (9) to be minimized measures the functional distance between neural-network generated scattering efficiency and the desired one. In general, there is no guarantee that a particular scattering configuration can be found to yield the desired function $\boldsymbol{Q}_{\text{desired}}$, so the error can be large. The goal is to find the optimal input vector $\boldsymbol{V}$ to minimize the error in (9). As we will describe below, this can be done in a computationally efficient manner. Note that the error in (9) is different from the training error in Eq. (8), which measures the difference between the neural-network output to the ground truth, i.e., the scattering efficiency function from a given set of input gate potentials.

To better describe the solutions of the optimization problem, we employ a concrete parameter setting. To generate either cloaking or superscattering, we use a small or a large

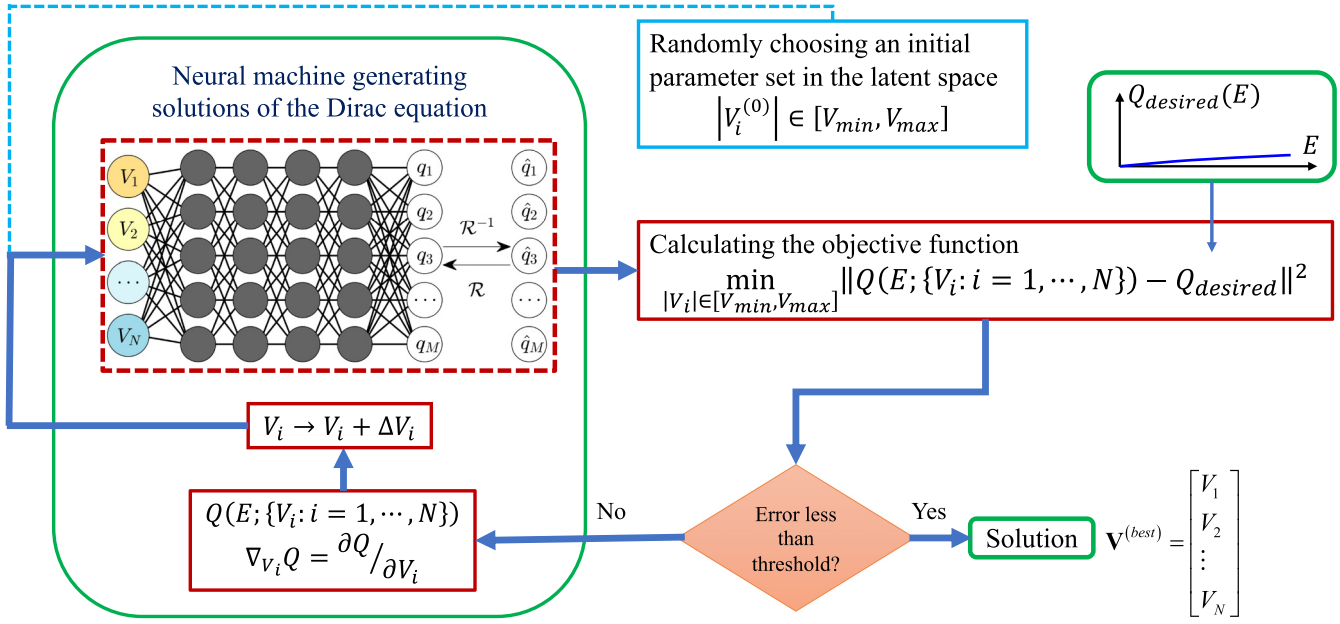

FIG. 3. Flow chart of machine-learning-based solution to the inverse-design problem of quantum scattering. A multilayer neural network is first trained using a number of functions $Q(E)$ of the scattering efficiency versus the electron energy for scattering from a multilayer graphene quantum dot subject to externally applied gate potentials, one to each graphene layer. For any given set of gate potentials, the corresponding function of the scattering efficiency is obtained from the exact solutions of the Dirac equation. A well-trained neural network will generate solutions of the Dirac equation. Solving the inverse-design problem requires the following steps: (1) choosing a random initial set of gate potentials as the input to the neural network to obtain the Q function, (2) calculating the difference between this and the desired Q function, (3) if the difference is sufficiently small (e.g., less than a predefined threshold), then a solution is deemed to be found, (4) otherwise calculating a change in each gate potential through the standard gradient approach applied to the neural network to yield a new set of input variables, and (5) repeat steps (3) and (4) until a solution is found.

quantum dot, respectively, as a reference to obtain the desired function $\boldsymbol{Q}_{\text{desired}}$, subject to the average gate potential 87.5 meV. Initially, the various gate potentials are chosen randomly from the two symmetric intervals defined by $|V_i| \in [75, 100]$ meV. Let the desired energy region be from 25 to 70 meV, which is discretized into 200 points, on which both the $\boldsymbol{Q}_{\text{desired}}$ and $\hat{\boldsymbol{Q}}(\mathbf{V})$ functions are evaluated. The number N of layers in the graphene quantum dot depends on the design imperative, i.e., whether cloaking or superscattering is to be realized. Given a task, once N has been determined, it is fixed so that training and inverse design are carried out on the same neural-network architecture. After training, the error function in (9) is a deterministic function, whose global minimum can be found by using, e.g., the standard interior point method [75,76]. We incorporate a gradient descent procedure into the interior point method as it was established previously that this can significantly improve the computation speed [28].

A detailed description of the steps involved in finding an optimal solution to the inverse-design problem is presented in Fig. 3.

Four remarks are in order. First, using the trained neural network in Fig. 3 as a Dirac-equation solver has the advantage of being extremely computational efficient. If the original Dirac equation were used for calculating the scattering-efficiency Q function and for finding the optimal set of gate potentials, a vast number of combinations of the potentials would typically be needed and, for each combination, the calculation of the Q involves calculating a large number of sophisticated mathematical functions. With the neural network, the standard gradient-descent method can be invoked to find the changes in the gate potentials, quickly yielding the optimal solution. For each given set of input variables, all needed for the neural network is matrix multiplications, which can be done extremely efficiently using well-developed packages.

Second, compared with designing a multilayer spherical scattering structure in optics where the input vector is chosen from a single interval of the physical parameter (e.g., dielectric constant), for graphene the input gate potentials are chosen from two symmetric intervals: One negative and another positive. There is then a gap in the input parameter space. To our knowledge, this type of inverse design has not been studied. Our solution is to test all possible combinations for $V_i \in [-100, -75] \cup [75, 100]$ meV. For each V_i, there are two possibilities: Positive or negative. For an N-layer circular quantum dot, altogether we have 2^N possible combinations. Take $N = 2$ as an example. If we wish to find the best $V_{(1,2)}$ such that $|V_{(1,2)}| \in [75, 100]$ meV, there are four combinations: $V_1 \in [75, 100]$ meV $\cap$ $V_2 \in [75, 100]$ meV, $V_1 \in [75, 100]$ meV $\cap$ $V_2 \in [-100, -75]$ meV, $V_1 \in [-100, -75]$ meV $\cap$ $V_2 \in [75, 100]$ meV, and $V_1 \in [-100, -75]$ meV $\cap$ $V_2 \in [-100, -75]$ meV. For each combination, an optimal solution can be found using one or two optimization steps. We find that, for a larger value of N, the best scattering structure can be found with one or two optimization steps for each combination of the gate potentials. As a result, for any reasonable value of N (e.g., $N = 10$), the required computational complexity is well manageable.

Third, to find the global minimum in (9), including derivatives can increase the computational efficiency [28]. In particular, to optimize $\mathbf{V}$ from (9), we calculate the derivative $\partial\,\mathrm{Loss}/\partial\mathbf{V}$. Since Q_{desired} is fixed, it is only necessary to evaluate the derivative of $\hat{Q}(\mathbf{V})$ with respect to $\mathbf{V}$. Once the neural network has been trained, such derivative can be evaluated through back propagation.

Fourth, there is no guarantee that the "best" scattering configuration can be found to make the scattering efficiency function match the desired function. Empirically, we find that increasing the number N of graphene layers typically results in a smaller optimization loss. This issue will be addressed in detail in Sec. IV.

III. RESULTS

A. Cloaking

The goal of designing a graphene-based cloaking system is to generate near-zero scattering cross section or efficiency from a relatively large quantum-dot structure. Since a small structure tends to generate minuscule scattering efficiency, the desired vector $\boldsymbol{Q}_{\mathrm{desired}}$ can be found by solving the relativistic quantum scattering problem over a small single-layer structure, e.g., a circular cavity of radius $R_C = 3$ nm subject to a gate potential of $V = 87.5$ meV, which produces near-zero scattering efficiency, as shown in Fig. 1(b). Now consider a nine-layer ($N = 9$) structure of size 10 times larger: $R_N = 30$ nm. The question can be stated as follows: Is it possible to find a set of suitable gate potentials represented by input vector $\boldsymbol{V}$ such that the resulting scattering efficiency function is arbitrarily close to $\boldsymbol{Q}_{\mathrm{desired}}$, subject to the constraint $\sum |V_i|/N \approx 87.5$ meV?

Figure 4(a) presents three curves of the resulting scattering efficiency function: The desired function (dashed blue), the function from the optimal multilayer graphene structure found by our physics-constrained neural network (solid blue curve), and a training curve that is the nearest to the target curve from the training data set (dotted orange curve). The optimal physical structure of the multilayer graphene quantum dot found by our physics-constrained neural network is shown in the lower right inset, where the gate potential values applied to different layers are specified by the color bar with $\sum |V_i|/N = 85.7$ meV. We see that the optimal structure produces a scattering efficiency function that well approximates the desired function. The purpose of the dotted orange curve is to demonstrate that the optimal structure is not a trivial interpolation of some of the structures in the training data set. Apparently, this "best" training function deviates from the desired function as the energy increases. The structure found by the neural network is thus one that produces real cloaking in that the scattering efficiency is as small as that produced by a structure of 10 times smaller in the entire energy range of interest.

Intuitively, the emergence of cloaking from a large scattering structure requires destructive quantum interference. If the gate potentials applied to nearby layers have the same sign, it will be difficult to produce such interference. However, alternating signs of the gate potentials between neighboring layers can facilitate the occurrence of destructive interference. The pattern of gate potentials of the optimal multilayer structure predicted by the neural network indeed has this feature, as demonstrated in Fig. 4(a).

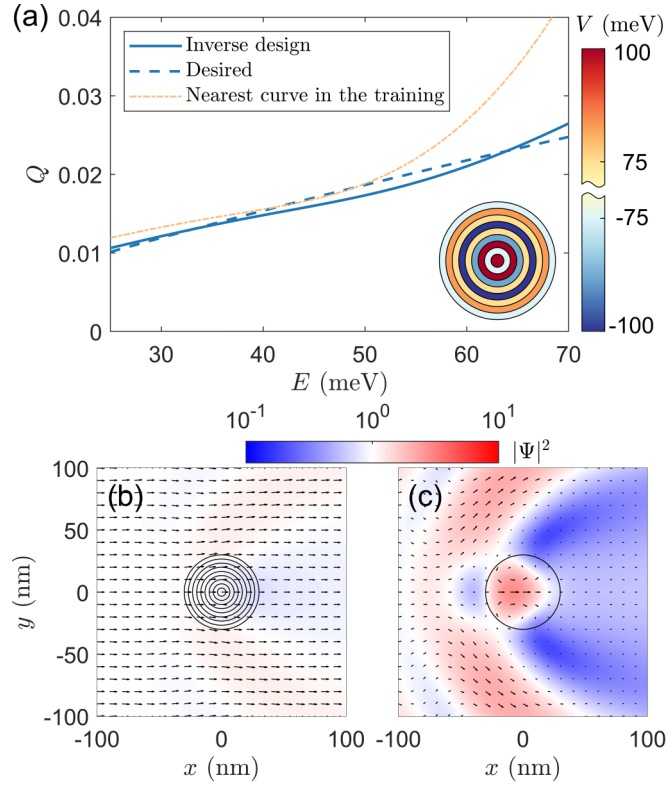

FIG. 4. Optimal multilayer graphene structure exhibiting cloaking found by machine learning. The parameter requirements of inverse design are $R_N = 30$ nm and $\sum |V_i|/N \approx 87.5$ meV. (a) Comparison of three curves of scattering efficiency: The desired function from a single-layer scatterer of 10 times smaller (dashed blue curve), the function underlying cloaking from the neural network (solid blue curve), and a curve closest to the target curve from the training data set (dotted orange curve). The inset illustrates the optimal quantum-dot structure predicted by machine learning. That the two blue curves stay close to each other in the entire energy range is indicative of cloaking, due to the much larger size of the neural-network produced multilayer structure. (b) Spatial patterns of the spinor wave function and the local currents associated with cloaking, with the quantum-dot structure at the center and an incident plane wave from the left. The wave-function intensity is nearly constant (unity) in the relevant spatial domain and the current vectors hardly change their directions, signifying cloaking. (c) Conventional scattering patterns from a single-layer quantum dot of the same size. There is no cloaking due to the significant changes in both the wave-function intensity and the current direction across the domain.

To visualize the cloaking phenomenon, we show the spatial distribution of the scattering wave function. In particular, we compare two cases: The optimal nine-layer structure found by the neural network and a single-layer quantum dot of the same size ($R = 30$ nm), in terms of the wave-function intensity defined as $|\psi_1|^2 + |\psi_2|^2$, where ψ_1 and ψ_2 are the two spinor components and the directions of the local current vectors are given by $[\langle\sigma_x\rangle, \langle\sigma_y\rangle]$. For a planar incident wave of unit intensity from the left, if cloaking has indeed occurred, the wave-function intensity should be nearly constant

in the spatial domain of interest and the current vectors should not change their directions, i.e., to maintain their horizontal directions. Such behaviors characteristic of cloaking are exemplified in Fig. 4(b) for the optimal multilayer structure. In contrast, for the single-layer structure where no cloaking occurs, the wave-function intensity varies across the domain and there are significant changes in the directions of the local current vectors, as shown in Fig. 4(c).

For a graphene quantum-dot scatter with a weak scattering strength as measured by $\bar{V}R$, where $\bar{V}$ is the average potential and R is the dot size, the scattering efficiency with the strength exhibits the following scaling relation [56]: $Q \sim (\bar{V}R)^3$. The target quantum dot used for generating the cloaking behavior has the size of 3 nm. According to this scaling law, the scattering efficiency should be quite small. For a relatively large circular structure of size, e.g., 10 nm, the scattering efficiency saturates and the scaling law no longer holds, but it can still be used to obtain an order-of-magnitude estimate of the scattering efficiency. We find that for a conventional structure of size 30 nm without cloaking, the scattering efficiency is typically two orders of magnitude larger than that from a structure of size 10 times smaller. It is remarkable that our physics-constrained machine-learning approach is able to predict a large quantum-dot structure but with scattering efficiency two orders of magnitude smaller than it "should be" in the conventional sense, effectively realizing cloaking.

B. Superscattering

We now address the opposite (to cloaking) problem: Is it possible to design a multilayer scattering structure of certain size to generate scattering efficiency that conventionally would be achieved only with a much larger scatterer? To be concrete, we use a nine-layer structure of size $R_N = 30$ nm with gate potentials subject to $\sum |V_i|/N \approx 87.5$ meV and generate the target scattering efficiency function $\boldsymbol{Q}_{\text{desired}}$ using a single-layer scatterer of size $R_A = 63$ nm, as illustrated in Fig. 1(b). The actual curve of $\boldsymbol{Q}_{\text{desired}}$ is shown by the dashed blue curve in Fig. 5(a), where a scattering resonance arises at energy about $E = 30$ meV. Applying our machine-learning-based inverse-design algorithm, we obtain the quantum-dot structure as illustrated in the upper right corner of Fig. 5(a) with $\sum |V_i|/N = 89.8$ meV, where the resulting scattering efficiency function is represented by the solid blue curve that stays near the target curve in the entire energy range. This means that the optimal structure of size 30 nm is able to produce scattering efficiency as large as that from a structure of much larger size. For reference, the "best" scattering efficiency function from the training data set is shown (the dotted orange curve). The optimal quantum-dot structure discovered by the neural network for superscattering requires a negative gate potential for majority of the layers. The spatial distribution of the wave-function intensity and the local current vectors associated with superscattering are shown in Fig. 5(b), and the same quantities from a single-layer scatterer of the same size are shown in Fig. 5(c). It can be seen that scattering from the optimal multilayer structure is markedly enhanced. An advantage of our machine-learning design is that no additional training of the neural network is required here, insofar as it has been trained for cloaking

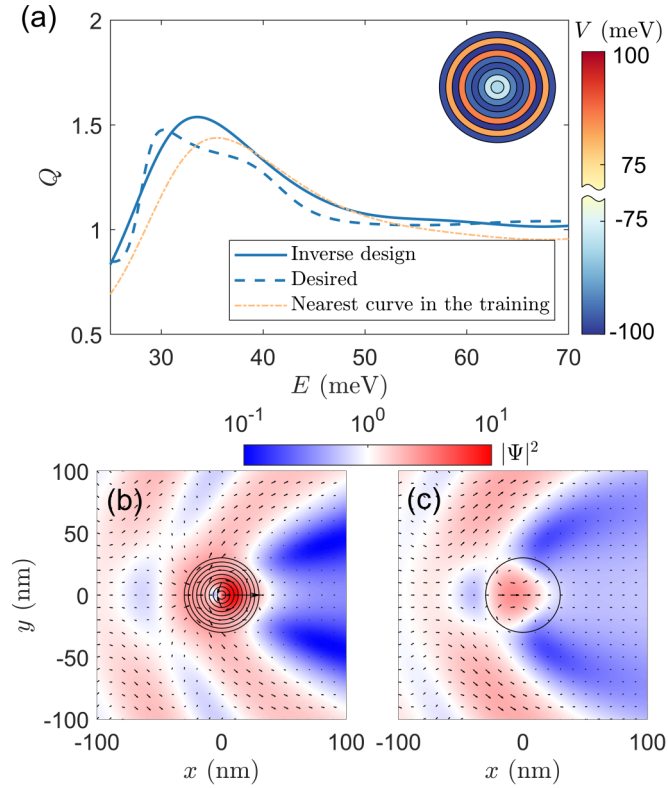

FIG. 5. Optimal multilayer graphene structure exhibiting superscattering produced by machine learning. As for the case of cloaking in Fig. 4, the structure has the size $R_N = 30$ nm and the gate potentials satisfy the constraint $\sum |V_i|/N \approx 87.5$ meV. (a) Comparison of three curves of scattering efficiency: The desired function from a single-layer scatterer of twice larger size (dashed blue curve), the function associated with superscattering from machine learning (solid blue curve), and a training curve closest to the target curve from the training data set (dotted orange curve). The upper right inset illustrates the optimal quantum-dot structure predicted by the neural network. (b) Spatial patterns of the spinor wave function and the local currents associated with superscattering at the energy $E = 30$ meV, with the quantum-dot structure at the center and an incident plane wave from the left. There are significant variations in the wave-function intensity and in the directions of the local current vectors across the spatial domain of interest. (c) Conventional scattering patterns from a single-layer quantum dot of the same size.

with a predefined set of scattering configurations, each with a distinct scattering efficiency curve. That is, the same neural network used to realize cloaking can be used to find a system that exhibits superscattering.

Comparing Fig. 5 with Fig. 4, we see that the effect of scattering enhancement is not as pronounced as that associated with scattering suppression in cloaking. The physical reason is Klein tunneling. In particular, the median Fermi energy is about half of the maximum gate potential: $E \approx |V|/2$, for which Klein tunneling is significant in graphene [55]. As a result, a spinor wave tends to change little in its wave vector after entering the scattering region. To generate strong cloaking in graphene is thus much more likely than to produce superscattering. We note that Klein tunneling is a unique feature of Dirac waves.

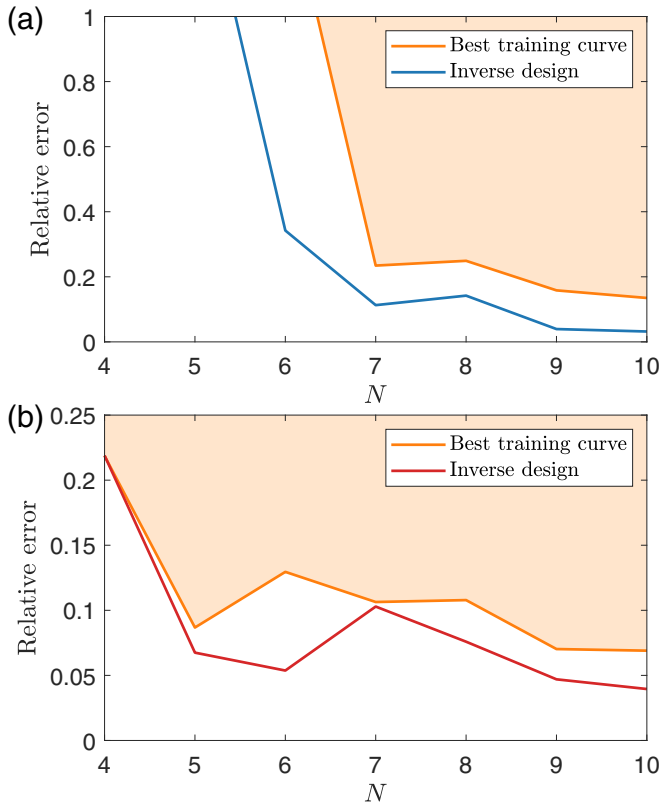

FIG. 6. Effect of the number of quantum-dot layers on training and inverse-design errors. (a) For designing a cloaking device, the errors versus N, where the shaded region indicates the normalized mean-square error from the training data set and the solid blue curve represents the error between the scattering function from the optimal structure and the desired function. As the number of quantum-dot layers increases, the errors decrease. (b) Similar behaviors arise when designing a superscattering device.

IV. ISSUES PERTINENT TO INVERSE DESIGN

A. Effect of the number of quantum-dot layers

Physically, the objective of machine-learning-based inverse design is to find a multilayer quantum-dot structure to generate cloaking or superscattering by tuning a set of experimentally feasible, external parameters. From a mathematical point of view, the task is to find a complicated multilayer scattering structure to achieve a desired function. While increasing the number of layers in general allows more complicated scattering functions, the computational complexity increases as well. From the point of view of machine learning, a more sophisticated scattering structure requires more training data.

Figure 6(a) shows, when designing a cloaking device, the relative training and inverse-design errors versus N, where the shaded area denotes all possible error values from training, with the solid orange curve denoting the minimally possible training error. Physically, a circular quantum dot of too few layers is unable to lead to cloaking, so the errors are large for small-N values. As N increases, it is more likely to generate destructive interference to realize cloaking, leading to a decrease in the errors. The solid blue curve in Fig. 6(a) represents the error in the scattering function from the optimal dot structure found by inverse design, which is below the minimum training-error curve. That is, the neural network is able to find a quantum-dot structure whose scattering function is closer to the desired one than any function in the training data set. A similar behavior arises for the task of generating superscattering, as shown in Fig. 6(b).

The results in Fig. 6 indicate that there is no guarantee that an increase in the number of layers can lead to better inverse design. In our work, the set of radii of the graphene quantum dot is fixed to be $R_i = iR_N/N$. Suppose N is increased by some integer factor. The resulting quantum dot would lead to a better chance of realizing scattering possibilities but the structure becomes more sophisticated. In general, the performance of the new structure should be improved, but empirically this is true only when the numbers of layers of the two structures are commensurate. From another angle, a perturbation on the gate potential of a graphene layer will affect the resulting error in the scattering efficiency but, as will be shown below in Fig. 7, and modifying the gate potential of an outer layer has a stronger effect on augmenting the error than disturbing the potential of an inner layer. This suggests that the nonmonotonous behavior in the error be an intrinsic physical property of the scattering system that has little dependence on the machine-learning procedure.

Due to the physical nature of the error behavior, choosing the "best" number of graphene layers is difficult. In fact, the same problem arises in optical inverse-design problems. For example, in Refs. [28,39], it was necessary to preassign the number of neural-network layers. A mathematical difficulty is that this number is not a continuous variable. In Refs. [77,78], reinforcement learning was used to find the best device whose dimension is only allowed to take on some discrete value. Another difficulty is that the multilayer perceptron employed in our work is suitable for inputs with a fixed dimension. In the field of computer vision, a method called spatial pyramid pooling [79] was proposed to address this difficulty. The idea was extended to adaptive pooling layers in PYTORCH [80]. While for our quantum-dot inverse-design problem, increasing the number of layers can lead to better performance in most cases, to the best of our knowledge to predetermine the optimal number of layers before training remains to be an open problem.

B. Robustness of inverse design

The main idea underlying machine-learning-based inverse design is that a properly trained neural network is able to generate an approximation of a desired mathematical function. In general, the optimal physical structure found by machine learning is close to but not necessarily the global minimum of the loss function (9). It is useful to examine how "far away" the loss associated with the optimal structure is from the global minimum. Consider the "best" quantum-dot structures for cloaking and superscattering design, as shown in Figs. 4 and 5, respectively, where the structure contains nine layers so the loss function depends on nine gate-potential variables and is thus nine dimensional. Suppose the optimal structure is described by the following set of N potential values: $[V_1, V_2, \ldots, V_N]$. We apply a small perturbation, e.g.,

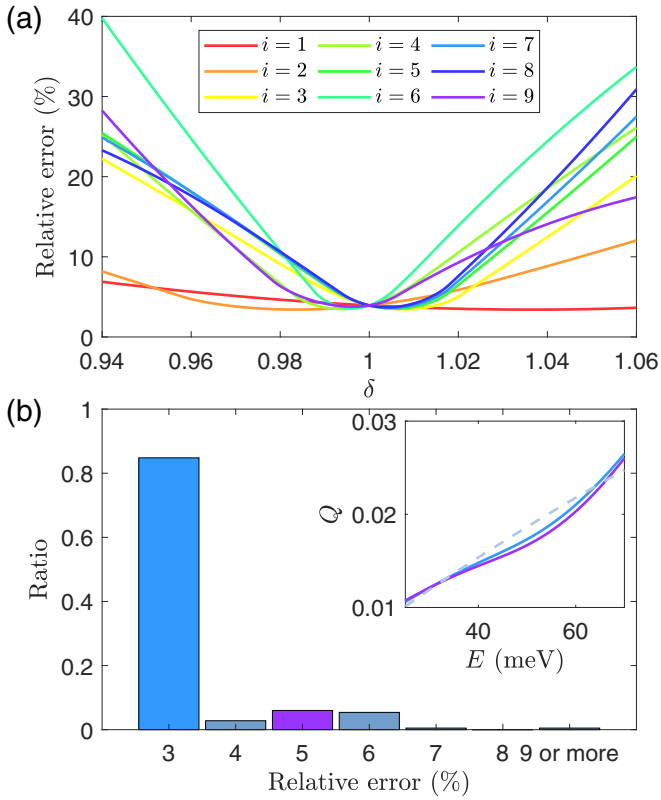

FIG. 7. Accuracy and robustness of inverse design. (a) For a graphene quantum dot of $N = 9$ layers to realize cloaking, normalized mean-square error in the machine-generated scattering efficiency function versus the perturbation parameter δ in the gate potential as applied to different layers of the graphene quantum dot, where $\delta = 1$ corresponds to zero perturbation. The error is calculated from the exact solution of the Dirac equation. The ideal global minimum of the loss function should occur at $\delta = 1$. The actual minima are quite close to $\delta = 1$. (b) Histogram of the relative error in the machine-generated scattering efficiency function from an ensemble of 10^4 combinations of the input parameters. More than 80% of times, the global minima are within 3% of $\delta = 1$, indicating robustness of the inverse-design scheme. The inset shows two examples of the scattering efficiency function from the machine-predicted optimal dot structure, with the relative error of 3% and 5%, respectively, where the desired function is indicated by the dashed curve.

V_1, which leads to a new set of potentials: $[V_1\delta, V_2, \ldots, V_N]$, and examine the dependence of the loss function on δ. If the original solution $[V_1, V_2, \ldots, V_N]$ corresponds to the global minimum, the loss function versus δ should exhibit a minimum at $\delta = 1$. Figure 7(a) shows the loss $\mathcal{L}^{(i)}$ versus δ for different layer index i. The functions are approximately parabolic with its minimum close to $\delta = 1$ but not exactly at $\delta = 1$, because both the neural network and the optimization algorithm are approximate solvers and the resulting scattering structure is not the absolute best. When the potential change occurs on one of the innermost layers (corresponding to small layer indices), the parabolic functions are rather shallow, but they are steep for large layer indices, indicating that modifying the gate potential of an outer layer has a stronger effect on augmenting the error. A plausible reason is that an outer layer has a larger area than that of an inner layer for a fixed layer width.

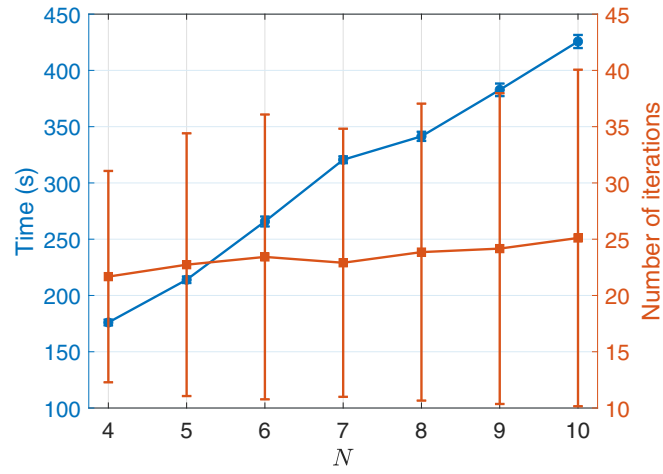

FIG. 8. Computational time required for training and convergence. The data points associated with the left and right y axis represent, respectively, the average time (with variance) of the training time and the number of iterations required for the neural network to converge as the number of layers constituting the quantum-dot structure increases from 4 to 10. The training time increases linearly with N but the number of iterations changes little due to the fact that the number of parameters in the neural network is approximately invariant as N increases in this range.

As discussed in Sec. II C, in comparison with the inverse problem of electromagnetic wave scattering, a difficulty in solving the inverse problem of quantum scattering in graphene is that the input parameters can be selected from a positive and a negative interval, leading to a gap in-between the two allowed parameter intervals. Implementing an optimization algorithm, e.g., the interior point method, requires examining a large number of parameter combinations from the two intervals. For an N-layer graphene quantum dot, the optimization needs to be done 2^N times, requiring that the algorithm be computationally efficient. In general, even based on a fixed nonlinear function, the optimization result would depend on the initial condition. Figure 7(b) shows, for $N = 9$ and cloaking design, the statistical distribution of the relative error of the optimization algorithm from 10^4 different initial conditions. It can be seen that for more than 80% of the initial conditions, the optimization errors are about 3%, indicating the robustness of our machine-learning-based inverse-design scheme.

C. Computational time required for training and convergence

Two main factors determine the time required for training: The amount of training data and the number of training parameters. Empirically, the amount of training data is proportional to the number N of the layers constituting the quantum-dot structure. With regard to the training parameters, our neural network contains four hidden layers. For different values of N, the number of neural-network parameters remains approximately constant. Figure 8 exemplifies how the average training time and variance depend on N. With CPU i7-6850k without GPU acceleration, the average training time ranges from 100 to 400 s as the number of layers of the scattering structure increases from 4 to 10.

About the convergence of the training process, we observe that the training error depends on the initial condition for the neural network and data sampling, as demonstrated in Fig. 2(c). However, the variance in the error is small, indicating that convergence is always granted for training.

The computational speed for the inverse design depends on two factors: Strong the number of iterations required to achieve convergence and the time cost of each iteration. Because the number of parameters involved in the neural network does not change significantly with N, for each iteration approximately a similar amount of time is required for calculating the gradient for different values of N. As N increases, while more iterations are needed for optimizing the neural network, we find that the required increase in the number of iterations is insignificant. For example, as N increases from 4 to 10, the average number of iterations required increases only from about 21 to about 23, as shown in Fig. 8 as well as in Fig. 7(b) where more than 80% of the initial conditions lead to satisfactory convergence.

V. DISCUSSION

To develop quantum systems to realize drastically distinct functions under a different set of externally imposed parameter values is a challenging problem. For example, for a multilayer graphene quantum-dot system in which each layer is subject to a gate potential, one may wish to make the system "invisible" by generating cloaking with weak scattering under one set of gate potentials, but as the environment changes the opposite extreme may be desired: Strong scattering (or superscattering defined in a broad sense). Would it be possible to simply change the set of external gate potentials, while keeping the physical structure of the system intact, so as to induce a metamorphic transition in the quantum scattering dynamics from one extreme to the opposite extreme? In optics, if the goal is to realize one of the two extremes, then the method of scattering cancellation can be effective [33,34,36,46]. The challenge is whether it is possible to create a system that is capable of the two extremes of scattering, superscattering and cloaking, in certain energy range by tuning a set of experimentally adjustable parameters. In principle, by exploiting a multilayer scattering structure to induce unusually strong destructive or constructive quantum interference, it should be possible to meet the challenge. However, the computational complexity associated with a brute-force search of the parameter space, aided by the exact solution of the Dirac equation, tends to grow exponentially with the number of layers constituting the scatterer, making the problem NP hard.

The main idea underlying this work is that machine learning provides an effective and computationally efficient approach to solving the challenging inverse-design problem, i.e., to generate the two opposite extremes of quantum scattering by tuning a set of experimentally accessible parameters. For the concrete setting of a multiple-ring graphene quantum-dot structure, the basic component of the machine-learning architecture is a conventional multilayer, feed-forward neural network that takes a set of gate potentials (one applied to each ring layer of the quantum dot) as input and generates the scattering cross section or efficiency versus the Fermi energy (scattering function) as the output. Straightforwardly training this baseline neural-network architecture using an adequate number of predetermined scattering functions obtained from directly solving the Dirac equation reveals that a nonnegligible fraction of the output functions can be nonphysical in that they contain negative values of the scattering efficiency. Imposing the physical constraint that the scattering efficiency must be positive through an additional output layer leads to a class of physics-constrained neural networks. The main accomplishment of this paper is a successful demonstration that physics-constrained machine learning can effectively solve the inverse design problem of generating quantum scattering dynamics at the opposite extremes in the same system.

The key advantage and appealing feature of our machine-learning approach to the extreme inverse-design problem can be appreciated using the paradigmatic multilayer graphene quantum-dot structure studied in this paper. Given a desired scattering function, the starting point is a random set of gate potentials. It is practically impossible for this initial set of input parameter values to produce a scattering function close to the desired one, so adjustments to the inputs are necessary. Solving the Dirac equation for random parameter perturbations is computationally infeasible not only because the solutions of the Dirac equation require evaluating a large number of sophisticated mathematical functions but, more importantly, because of the NP-hard nature of the search problem. These fundamental difficulties can be overcome by exploiting machine learning for the following reasons: (1) all mathematical operations needed is matrix multiplication that can be done extremely efficiently and (2) the standard gradient descent can guild the solution in the input-parameter space to approach the optimal one in only a few steps. We have demonstrated that, typically, the optimal solution is close to the global minimum of the physics-constrained loss function.

In this work, we have studied quantum-dot systems in the weak scattering regime, where the wave vector inside the cavity has the same order of magnitude as the wave vector outside the cavity. In this case, the scattering efficiency does not exhibit sharp resonant peaks, in contrast to the case of strong resonances where the wave vector inside the cavity is much larger than the wave vector outside [81]. Empirically, a change in the potential profile requires that the neural network be retrained.

In general, inverse problems in physical sciences are challenging, but to design a system that can generate physical behaviors of two opposite extremes is even harder. We emphasize that no feasible conventional methods exist which can be used to solve this type of *extreme inverse problems* due to the following difficulties: The need to search through a high-dimensional parameter space, the complex relation between the system structure and functions, and the uncertainty about the existence of a configuration that can deliver the desired system functions. Our machine-learning-based solution represents a step forward in this area.

ACKNOWLEDGMENTS

We thank Dr. H.-Y. Xu for providing great inputs during the initial stage of this project. This work was supported by AFOSR under Grant No. FA9550-21-1-0438.

APPENDIX: SCATTERING CROSS SECTION FOR CIRCULAR GRAPHENE QUANTUM DOT

1. Single-layer circular graphene quantum dot

Consider a single-layer graphene quantum dot of radius R_1, subject to gate potential V_1, where the circular boundary separates the whole space into two regions: Inside ($r < R_1$) and outside ($r > R_1$) the dot, denoted as regions 1 and 2, respectively. The energies inside and outside the dot are $E - V$ and E, respectively, giving the corresponding wave vectors as $k_1 = |E - V|/v_g$ and $k_2 = |E|/v_g$. For convenience, we define two sign quantities: $\tau_1 \equiv \text{Sign}(E - V)$ and $\tau_2 \equiv \text{Sign}(E)$.

The incident wave is

$$\psi_{\text{in}}(\mathbf{r}) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \tau_2\end{pmatrix}e^{ik_2x}. \tag{A1}$$

Using the Jacobi-Anger identity

$$e^{iz\cos\theta} \equiv \sum_{l=-\infty}^{\infty} i^l J_l(z)e^{il\theta}, \tag{A2}$$

we expand the incident wave in the polar coordinates as

$$\psi_{\text{in}}(\mathbf{r}) = \sum_{l=-\infty}^{\infty}\frac{i^l}{\sqrt{2}}\begin{pmatrix}J_l(k_2r)\\ i\tau_2 J_{l+1}(k_2r)\end{pmatrix}e^{il\theta}. \tag{A3}$$

Inside the quantum dot, we have

$$\psi_1(\mathbf{r}) = \sum_{l=-\infty}^{\infty} A_l^{(1)}\begin{pmatrix}J_l(k_1r)\\ i\tau_1 J_{l+1}(k_1r)\end{pmatrix}e^{il\theta}, \tag{A4}$$

and the wave function outside the dot is given by

$$\psi_2(\mathbf{r}) = \psi_{\text{in}} + \sum_{l=-\infty}^{\infty} A_l^{(2)}\begin{pmatrix}H_l^{(1)}(k_2r)\\ i\tau_2 H_{l+1}^{(1)}(k_2r)\end{pmatrix}e^{il\theta}. \tag{A5}$$

Matching the boundary conditions for each angular momentum l, we get

$$\begin{bmatrix}J_l(k_1R_1) & -H_l^{(1)}(k_2R_1)\\ \tau_1 J_{l+1}(k_1R_1) & -\tau_2 H_{l+1}^{(1)}(k_2R_1)\end{bmatrix}\begin{pmatrix}A_l^{(1)}\\ A_l^{(2)}\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}i^l J_l(k_2R_2)\\ \tau_2 i^l J_{l+1}(k_2R_2)\end{pmatrix}. \tag{A6}$$

2. Multilayer circular graphene quantum dot

Consider an N-layer circular graphene quantum dot, where all the layers are concentric and have radii R_1, R_2, ..., and R_N with the respective gate potentials $V_1, V_2, \ldots$, and V_N. Let $k_i = |E - V_i|/v_g$ and $\tau_i \equiv \text{Sign}\,(E - V_i)$. In region 2, we have $k_{N+1} = |E|/v_g$ and $\tau_{N+1} = \text{Sign}(E)$. To represent the wave function in each region, we use $A^{(1)}$ for $r < R_1$ and

$$\psi_1(\mathbf{r}) = \sum_{l=-\infty}^{\infty} A_l^{(1)}\begin{pmatrix}J_l(k_1r)\\ i\tau_1 J_{l+1}(k_1r)\end{pmatrix}e^{il\theta}, \tag{A7}$$

where $r < R_1$. For $R_i < r < R_{i+1}$ we have

$$\psi_{i+1}(\mathbf{r}) = \sum_{l=-\infty}^{\infty} A_l^{(i+1)}\begin{pmatrix}H_l^{(1)}(k_{i+1}r)\\ i\tau_{i+1}H_{l+1}^{(1)}(k_{i+1}r)\end{pmatrix}e^{il\theta} + \sum_{l=-\infty}^{\infty} B_l^{(i+1)}\begin{pmatrix}H_l^{(2)}(k_{i+1}r)\\ i\tau_{i+1}H_{l+1}^{(2)}(k_{i+1}r)\end{pmatrix}e^{il\theta}. \tag{A8}$$

For $r > R_N$ we have

$$\psi_{N+1}(\mathbf{r}) = \sum_{l=-\infty}^{\infty}\frac{i^l}{\sqrt{2}}\begin{pmatrix}J_l(k_{N+1}r)\\ i\tau_{N+1}J_{l+1}(k_{N+1}r)\end{pmatrix}e^{il\theta} + \sum_{l=-\infty}^{\infty} A_l^{(N+1)}\begin{pmatrix}H_l^{(1)}(k_{N+1}r)\\ i\tau_{N+1}H_{l+1}^{(1)}(k_{N+1}r)\end{pmatrix}e^{il\theta}. \tag{A9}$$

The unknown coefficients are $A_l^{(1)}, A_l^{(2)}, \ldots, A_l^N$ and $A_l^{(N+1)}$ as well as $B_l^{(2)}, B_l^{(3)}, \ldots$, and $B_l^{(N)}$. Altogether, there are $2N$ unknown parameters to be determined. We use the N boundary conditions to generate $2N$ equations, as follows.

For the boundary at $r = R_1$, we have

$$\begin{aligned}A_l^{(1)}J_l(k_1R_1) - A_l^{(2)}H_l^{(1)}(k_2R_1) - B_l^{(2)}H_l^{(2)}(k_2R_1) &= 0,\\ A_l^{(1)}\tau_1 J_{l+1}(k_1R_1) - A_l^{(2)}\tau_2 H_{l+1}^{(1)}(k_2R_1) - B_l^{(2)}\tau_2 H_{l+1}^{(2)}(k_2R_1) &= 0.\end{aligned} \tag{A10}$$

For the boundary at $r = R_i$, where $1 < i < N$, we have

$$\begin{aligned}A_l^{(i)}H_l^{(1)}(k_iR_i) + B_l^{(i)}H_l^{(2)}(k_iR_i) - A_l^{(i+1)}H_l^{(1)}(k_{i+1}R_i) - B_l^{(i+1)}H_l^{(2)}(k_{i+1}R_i) &= 0,\\ A_l^{(i)}\tau_i H_{l+1}^{(1)}(k_iR_i) + B_l^{(i)}\tau_i H_{l+1}^{(2)}(k_iR_i) - A_l^{(i+1)}\tau_{i+1}H_{l+1}^{(1)}(k_{i+1}R_i) - B_l^{(i+1)}\tau_{i+1}H_{l+1}^{(2)}(k_{i+1}R_i) &= 0.\end{aligned} \tag{A11}$$

For the boundary at $r = R_N$, we have

$$\begin{aligned}A_l^{(N)}H_l^{(1)}(k_NR_N) + B_l^{(N)}H_l^{(2)}(k_NR_N) - A_l^{(N+1)}H_l^{(1)}(k_{N+1}R_N) &= \frac{i^l}{\sqrt{2}}J_l(k_{N+1}R_N),\\ A_l^{(N)}\tau_N H_{l+1}^{(1)}(k_NR_N) + B_l^{(N)}\tau_N H_{l+1}^{(2)}(k_NR_N) - A_l^{(N+1)}\tau_{N+1}H_{l+1}^{(1)}(k_{N+1}R_N) &= \frac{i^l}{\sqrt{2}}\tau_{N+1}J_{l+1}(k_{N+1}R_N).\end{aligned} \tag{A12}$$

Solving this set of linear equations, we get the coefficients $A^{(N+1)}$. The scattering efficiency is given by

$$Q = \frac{2}{k_{N+1}R_N} \sum_{l=-\infty}^{\infty} \left(A_l^{(N+1)}\right)^2. \tag{A13}$$

In numerical computations, we truncate the summation at max $|l| = 7$ to achieve the desired accuracy.

[1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Electric field effect in atomically thin carbon films, Science **306**, 666 (2004).

[2] K. S. Novoselov, A. K. Geim, S. Morozov, D. Jiang, M. I. Katsnelson, I. Grigorieva, S. Dubonos, and A. Firsov, Two-dimensional gas of massless Dirac fermions in graphene, Nature (London) **438**, 197 (2005).

[3] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, The electronic properties of graphene, Rev. Mod. Phys. **81**, 109 (2009).

[4] T. Wehling, A. M. Black-Schaffer, and A. V. Balatsky, Dirac materials, Adv. Phys. **63**, 1 (2014).

[5] J. Wang, S. Deng, Z. Liu, and Z. Liu, The rare two-dimensional materials with Dirac cones, Nat. Sci. Rev. **2**, 22 (2015).

[6] W. Han, R. K. Kawakami, M. Gmitra, and J. Fabian, Graphene spintronics, Nat. Nanotechnol. **9**, 794 (2014).

[7] D. Khokhriakov, A. W. Cummings, K. Song, M. Vila, B. Karpiak, A. Dankert, S. Roche, and S. P. Dash, Tailoring emergent spin phenomena in Dirac material heterostructures, Sci. Adv. **4**, eaat9349 (2018).

[8] A. R. Kirmani, Dirac material heterostructures lead to next-generation spintronics, MRS Bull. **44**, 86 (2019).

[9] A. Avsar, H. Ochoa, F. Guinea, B. Özyilmaz, B. J. van Wees, and I. J. Vera-Marun, Colloquium: Spintronics in graphene and other two-dimensional materials, Rev. Mod. Phys. **92**, 021003 (2020).

[10] L. Huang, Y.-C. Lai, D. K. Ferry, S. M. Goodnick, and R. Akis, Relativistic Quantum Scars, Phys. Rev. Lett. **103**, 054101 (2009).

[11] H. Xu, L. Huang, Y.-C. Lai, and C. Grebogi, Chiral Scars in Chaotic Dirac Fermion Systems, Phys. Rev. Lett. **110**, 064102 (2013).

[12] L. Huang, H.-Y. Xu, C. Grebogi, and Y.-C. Lai, Relativistic quantum chaos, Phys. Rep. **753**, 1 (2018).

[13] Y.-C. Lai, H.-Y. Xu, L. Huang, and C. Grebogi, Relativistic quantum chaos - an emergent interdisciplinary field, Chaos **28**, 052101 (2018).

[14] Y. Zhao, J. Wyrick, F. D. Natterer, J. F. Rodriguez-Nieva, C. Lewandowski, K. Watanabe, T. Taniguchi, L. S. Levitov, N. B. Zhitenev, and J. A. Stroscio, Creating and probing electron whispering-gallery modes in graphene, Science **348**, 672 (2015).

[15] J. Lee, D. Wong, J. Velasco Jr, J. F. Rodriguez-Nieva, S. Kahn, H.-Z. Tsai, T. Taniguchi, K. Watanabe, A. Zettl, F. Wang *et al.*, Imaging electrostatically confined Dirac fermions in graphene quantum dots, Nat. Phys. **12**, 1032 (2016).

[16] C. Gutiérrez, L. Brown, C.-J. Kim, J. Park, and A. N. Pasupathy, Klein tunnelling and electron trapping in nanometre-scale graphene quantum dots, Nat. Phys. **12**, 1069 (2016).

[17] J. Velasco, Jr., L. Ju, D. Wong, S. Kahn, J. Lee, H.-Z. Tsai, C. Germany, S. Wickenburg, J. Lu, T. Taniguchi *et al.*, Nanoscale control of rewriteable doping patterns in pristine graphene/boron nitride heterostructures, Nano Lett. **16**, 1620 (2016).

[18] F. Ghahari, D. Walkup, C. Gutiérrez, J. F. Rodriguez-Nieva, Y. Zhao, J. Wyrick, F. D. Natterer, W. G. Cullen, K. Watanabe, T. Taniguchi *et al.*, An on/off Berry phase switch in circular graphene resonators, Science **356**, 845 (2017).

[19] Z. Ge, D. Wong, J. Lee, F. Joucken, E. A. Quezada-Lopez, S. Kahn, H.-Z. Tsai, T. Taniguchi, K. Watanabe, F. Wang *et al.*, Imaging quantum interference in stadium-shaped monolayer and bilayer graphene quantum dots, Nano Lett. **21**, 8993 (2021).

[20] L. Brey and H. A. Fertig, Electronic states of graphene nanoribbons studied with the Dirac equation, Phys. Rev. B **73**, 235411 (2006).

[21] Y.-W. Son, M. L. Cohen, and S. G. Louie, Energy Gaps in Graphene Nanoribbons, Phys. Rev. Lett. **97**, 216803 (2006).

[22] A. Rycerz, J. Tworzydło, and C. Beenakker, Valley filter and valley valve in graphene, Nat. Phys. **3**, 172 (2007).

[23] B. Apagyi, G. Endrédi, and P. Lévay, *Inverse and Algebraic Quantum Scattering Theory* (Springer, Berlin, 1996), Vol. 488.

[24] L. Banchi, N. Pancotti, and S. Bose, Quantum gate learning in qubit networks: Toffoli gate without time-dependent control, npj Quantum Inf. **2**, 16019 (2016).

[25] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, Defining network topologies that can achieve biochemical adaptation, Cell **138**, 760 (2009).

[26] B. Sanchez-Lengeling and A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, Science **361**, 360 (2018).

[27] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, Inverse design in nanophotonics, Nat. Photonics **12**, 659 (2018).

[28] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, Nanophotonic particle simulation and inverse design using artificial neural networks, Sci. Adv. **4**, eaar4206 (2018).

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

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

[31] S. An, B. Zheng, H. Tang, M. Y. Shalaginov, L. Zhou, H. Li, M. Kang, K. A. Richardson, T. Gu, J. Hu *et al.*, Multifunctional metasurface design with a generative adversarial network, Adv. Optical Mater. **9**, 2001433 (2021).

[32] D. Zhu, Z. Liu, L. Raju, A. S. Kim, and W. Cai, Building multifunctional metasystems via algorithmic construction, ACS Nano **15**, 2318 (2021).

[33] A. Alù and N. Engheta, Achieving transparency with plasmonic and metamaterial coatings, Phys. Rev. E **72**, 016623 (2005).

[34] C. Qian, X. Lin, Y. Yang, X. Xiong, H. Wang, E. Li, I. Kaminer, B. Zhang, and H. Chen, Experimental Observation of Superscattering, Phys. Rev. Lett. **122**, 063901 (2019).

[35] D. Schurig, J. J. Mock, B. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith, Metamaterial electromagnetic cloak at microwave frequencies, Science **314**, 977 (2006).

[36] B. Edwards, A. Alù, M. G. Silveirinha, and N. Engheta, Experimental Verification of Plasmonic Cloaking at Microwave Frequencies with Metamaterials, Phys. Rev. Lett. **103**, 153901 (2009).

[37] F. Monticone and A. Alu, Do Cloaked Objects Really Scatter Less?, Phys. Rev. X **3**, 041005 (2013).

[38] R. Fleury, F. Monticone, and A. Alù, Invisibility and Cloaking: Origins, Present, and Future Perspectives, Phys. Rev. Appl. **4**, 037001 (2015).

[39] A. Sheverdin, F. Monticone, and C. Valagiannopoulos, Photonic Inverse Design with Neural Networks: The Case of Invisibility in the Visible, Phys. Rev. Appl. **14**, 024054 (2020).

[40] L. Verslegers, Z. Yu, Z. Ruan, P. B. Catrysse, and S. Fan, From Electromagnetically Induced Transparency to Superscattering with a Single Structure: A Coupled-Mode Theory for Doubly Resonant Structures, Phys. Rev. Lett. **108**, 083902 (2012).

[41] Y. Huang, Y. Shen, C. Min, and G. Veronis, Switching photonic nanostructures between cloaking and superscattering regimes using phase-change materials, Opt. Mater. Express **8**, 1672 (2018).

[42] S. Lepeshov, A. Krasnok, and A. Alù, Nonscattering-to-superscattering switch with phase-change materials, ACS Photonics **6**, 2126 (2019).

[43] J. Luo, X. Li, X. Zhang, J. Guo, W. Liu, Y. Lai, Y. Zhan, and M. Huang, Deep-learning-enabled inverse engineering of multi-wavelength invisibility-to-superscattering switching with phase-change materials, Opt. Express **29**, 10527 (2021).

[44] M. Farhat, C. Rockstuhl, and H. Bağcı, A 3d tunable and multi-frequency graphene plasmonic cloak, Opt. Express **21**, 12592 (2013).

[45] R. Li, X. Lin, S. Lin, X. Liu, and H. Chen, Atomically thin spherical shell-shaped superscatterers based on a bohr model, Nanotechnology **26**, 505201 (2015).

[46] R. Li, B. Zheng, X. Lin, R. Hao, S. Lin, W. Yin, E. Li, and H. Chen, Design of ultracompact graphene-based superscatterers, IEEE J. Select. Topics Quantum Electron. **23**, 130 (2016).

[47] S. Zhang, D. A. Genov, C. Sun, and X. Zhang, Cloaking of Matter Waves, Phys. Rev. Lett. **100**, 123002 (2008).

[48] B. Liao, M. Zebarjadi, K. Esfarjani, and G. Chen, Cloaking Core-Shell Nanoparticles from Conducting Electrons in Solids, Phys. Rev. Lett. **109**, 126806 (2012).

[49] R. Fleury and A. Alù, Quantum cloaking based on scattering cancellation, Phys. Rev. B **87**, 045423 (2013).

[50] B. Liao, M. Zebarjadi, K. Esfarjani, and G. Chen, Isotropic and energy-selective electron cloaks on graphene, Phys. Rev. B **88**, 155432 (2013).

[51] D. Oliver, J. H. Garcia, T. G. Rappoport, N. M. R. Peres, and F. A. Pinheiro, Cloaking resonant scatterers and tuning electron flow in graphene, Phys. Rev. B **91**, 155416 (2015).

[52] M. Sadrara and M. Miri, Collective cloaking of a cluster of electrostatically defined core–shell quantum dots in graphene, J. Phys.: Condens. Matter **34**, 115703 (2022).

[53] P.-Y. Chen and A. Alu, Atomically thin surface cloak using graphene monolayers, ACS Nano **5**, 5855 (2011).

[54] D. S. Novikov, Elastic scattering theory and transport in graphene, Phys. Rev. B **76**, 245435 (2007).

[55] M. I. Katsnelson, K. S. Novoselov, and A. K. Geim, Chiral tunnelling and the Klein paradox in graphene, Nat. Phys. **2**, 620 (2006).

[56] C.-Z. Wang, H.-Y. Xu, and Y.-C. Lai, Super skew scattering in two-dimensional Dirac material systems with a flat band, Phys. Rev. B **103**, 195439 (2021).

[57] H. C. Nguyen, N. T. Nguyen, and V. L. Nguyen, The transfer matrix approach to circular graphene quantum dots, J. Phys.: Condens. Matter **28**, 275301 (2016).

[58] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signal Syst. **2**, 303 (1989).

[59] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, Tensorflow: A system for large-scale machine learning, in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (USENIX (The Advanced Computing Systems Association), Savannah, GA, USA, 2016), pp. 265–283.

[60] F. Chollet *et al.*, KERAS, https://keras.io (2015).

[61] P. De Wilde, Class of Hamiltonian neural networks, Phys. Rev. E **47**, 1392 (1993).

[62] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, in *NIPS'19 Proceedings of the 33rd International Conference on Neural Information Processing Systems, December 2019, Whistler, British Columbia, Canada* (The MIT Press, Cambridge, Massachusetts, 2019), pp. 15379–15389.

[63] P. Toth, D. J. Rezende, A. Jaegle, S. Racanière, A. Botev, and I. Higgins, Hamiltonian generative networks, arXiv:1909.13789.

[64] T. Bertalan, F. Dietrich, I. Mezić, and I. G. Kevrekidis, On learning Hamiltonian systems from data, Chaos **29**, 121107 (2019).

[65] A. Choudhary, J. F. Lindner, E. G. Holliday, S. T. Miller, S. Sinha, and W. L. Ditto, Physics-enhanced neural networks learn order and chaos, Phys. Rev. E **101**, 062207 (2020).

[66] Z. Chen, J. Zhang, M. Arjovsky, and L. Bottou, Symplectic recurrent neural networks, arXiv:1909.13334.

[67] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho, Lagrangian neural networks, arXiv:2003.04630.

[68] A. Choudhary, J. F. Lindner, E. G. Holliday, S. T. Miller, S. Sinha, and W. L. Ditto, Forecasting Hamiltonian dynamics without canonical coordinates, Nonlinear Dyn. **103**, 1553 (2021).

[69] C.-D. Han, B. Glaz, M. Haile, and Y.-C. Lai, Adaptable Hamiltonian neural networks, Phys. Rev. Res. **3**, 023156 (2021).

[70] S. Xiong, X. He, Y. Tong, and B. Zhu, Neural vortex method: From finite Lagrangian particles to infinite dimensional Eulerian dynamics, arXiv:2006.04178.

[71] H. Gao, L. Sun, and J.-X. Wang, Super-resolution and denoising of fluid flow using physics-informed convolutional neural net-

works without high-resolution labels, Phys. Fluids **33**, 073603 (2021).

[72] J. Jiang and J. A. Fan, Global optimization of dielectric metasurfaces using a physics-driven neural network, Nano Lett. **19**, 5366 (2019).

[73] A. Sehanobish, H. H. Corzo, O. Kara, and D. van Dijk, Learning potentials of quantum systems using deep neural networks, arXiv:2006.13297.

[74] C.-D. Han, B. Glaz, M. Haile, and Y.-C. Lai, Tomography of time-dependent quantum Hamiltonians with machine learning, Phys. Rev. A **104**, 062404 (2021).

[75] R. H. Byrd, M. E. Hribar, and J. Nocedal, An interior point algorithm for large-scale nonlinear programming, SIAM J. Optim. **9**, 877 (1999).

[76] R. H. Byrd, J. C. Gilbert, and J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, Math. Program. **89**, 149 (2000).

[77] I. Sajedian, T. Badloe, and J. Rho, Optimisation of colour generation from dielectric nanostructures using reinforcement learning, Opt. Express **27**, 5874 (2019).

[78] S. So, T. Badloe, J. Noh, J. Bravo-Abad, and J. Rho, Deep learning enabled inverse design in nanophotonics, Nanophotonics **9**, 1041 (2020).

[79] K. He, X. Zhang, S. Ren, and J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. **37**, 1904 (2015).

[80] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems* (The MIT Press, Cambridge, Massachusetts, 2019), pp. 8026–8037.

[81] C.-D. Han, H.-Y. Xu, and Y.-C. Lai, Electrical confinement in a spectrum of two-dimensional dirac materials with classically integrable, mixed, and chaotic dynamics, Phys. Rev. Res. **2**, 013116 (2020).