

Data-driven model reconstruction for nonlinear wave dynamics

Ekaterina Smolina ¹, Lev Smirnov ¹, Daniel Leykam ², Franco Nori ^{3,4,5} and Daria Smirnova ^{3,6}

¹*Department of Control Theory, Nizhny Novgorod State University, Gagarin Av. 23, Nizhny Novgorod 603950, Russia*

²*Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, 8 Somapah Road, 487372, Singapore*

³*Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan*

⁴*Center for Quantum Computing (RQC), RIKEN, Wakoshi, Saitama 351-0198, Japan*

⁵*Physics Department, University of Michigan, Ann Arbor, Michigan 48109-1040, USA*

⁶*Research School of Physics, Australian National University, Canberra, ACT 2601, Australia*



(Received 24 November 2024; accepted 5 June 2025; published 26 June 2025)

The use of machine learning to predict wave dynamics is a topic of growing interest, but commonly used deep-learning approaches suffer from a lack of interpretability of the trained models. Here, we present an interpretable machine learning framework for analyzing the nonlinear evolution dynamics of optical wave packets in complex wave media. We use sparse regression to reduce microscopic discrete lattice models to simpler effective continuum models, which can accurately describe the dynamics of the wave packet envelope. We apply our approach to valley-Hall domain walls in honeycomb photonic lattices of laser-written waveguides with Kerr-type nonlinearity and different boundary shapes. The reconstructed equations accurately reproduce the linear dispersion and nonlinear effects, including self-steepening and self-focusing. This scheme is proven free of the *a priori* limitations imposed by the underlying hierarchy of scales traditionally employed in asymptotic analytical methods. It represents a powerful interpretable machine learning technique of interest for advancing design capabilities in photonics and framing the complex interaction-driven dynamics in various topological materials.

DOI: [10.1103/2jh8-p5y2](https://doi.org/10.1103/2jh8-p5y2)

I. INTRODUCTION

Machine learning (ML), a branch of artificial intelligence, is revolutionizing various scientific fields by enabling pattern extraction and prediction from large datasets [1–4]. In recent years, the use of ML to determine the governing equations of various dynamical systems and processes has shown remarkable potential.

One possible instrument for this purpose is the regression algorithm [5,6]. In particular, this tool has been successfully tested on famous fundamental physical models described by partial differential equations (PDEs), such as nonlinear Burgers' and Korteweg-de Vries equations [7], and the Belousov-Zhabotinsky reaction [7,8]. It was shown that the coefficients of governing PDEs of a known type can be recovered from numerical data generated by solving the equation within the same problem dimensionality.

Furthermore, the algorithms of symbolic regression, such as those implemented in the open-source library PySR [9], serve to automate the optimization problem for discovering empirical relationships from data. For the latter supervised learning task, search is performed in the model space spanned by analytic expressions, rather than fitting concrete pa-

rameters into overparameterized models. Further combining enhanced deep reinforcement learning and symbolic representations, the proposed DISCOVER framework explores open-form PDEs to reveal underlying physical laws [10].

The current challenge is to apply ML methods to real-world problems, moving beyond pre-expected basic models. Recently, beginning efforts in this direction have been reported. For instance, the authors of Ref. [11] used experimental data from degrading perovskite thin films subjected to environmental stressors to infer the underlying differential equation. Similarly, in nonlinear optics, data-driven approaches have been utilized to identify optimal conditions for four-wave mixing in an optical fiber [12].

A particularly striking form of wave dynamics arises in topological materials, where complicated microscopic lattice models give rise to robust guided modes at edges or interfaces between distinct topological phases. Such edge-based form of transport has sparked significant interest in photonics, as it shows promise for constructing scattering-resistant transmission waveguiding channels in high-speed photonic circuits and communication networks [13].

The analysis of electromagnetic wave propagation along topological domain walls traditionally relies on numerical simulations. Analytical methods are based on simplifying assumptions to achieve a quasi-one-dimensional reduction in the long-wavelength limit from higher-dimensional PDEs. In the case of two-dimensional lattices, this implies excluding the dimension transverse to the interface [14–17]. One can additionally use multiscale analysis to integrate out any

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

microscopic spin-like degrees of freedom to obtain simpler scalar wave equations [18].

More generally, dimensionality reduction is applicable to complex problems in other domains. For instance, the inherently nonlinear Navier-Stokes equations in hydrodynamics can be reduced to Lorentz-like ordinary differential equations (ODEs) to facilitate easier bifurcation analysis and identify conditions for chaotic dynamics corresponding to turbulence in the original system [19]. Another example is the nonlinear Schrödinger equation, analyzed using ODEs for moments, integral characteristics of the beam. ML regression was recently tested to reconstruct the coefficients of these ODEs, instead of analytical means [20].

However, the sophisticated geometry of topological lattices and the presence of optical nonlinearities often pose challenges in formulating reduced models that adequately capture the key properties needed to describe edge wave packet dynamics. The analytical approach based on perturbation theory involves a hierarchy of characteristic scales. Higher-order terms are typically omitted to obtain the simplest cubic nonlinear Schrödinger equation, regardless of the underlying lattice geometry or the form of modulation used to create the topological domain wall [14,15,21–23]. For example, to obtain an effective nonlinear Schrödinger equation for edge states of Floquet topological insulators, one needs to average over both spatial and temporal modulations. While the reduction to the cubic nonlinear Schrödinger equation provides an elegant way to understand the lowest-order nonlinear effects, it is only valid in the limit of weak nonlinearity. The question of which higher-order terms need to be taken into account and at what input power level remains unclear and cumbersome to address using asymptotic methods [18]. Data-driven ML approaches potentially offer an effective and systematic way to answer these questions. However, they generally require transitioning from multidimensional PDEs to simpler one-dimensional PDEs, necessitating the adaptation of existing regression schemes.

In this work, we apply a data-driven ML for the first time in the context of complex nonlinear photonic lattices with nontrivial topology, demonstrating how ML regression can be used to obtain simpler yet accurate PDE models for the edge state dynamics. The usefulness of such a model lies in its predictive power, which comes from identifying underlying physical effects. As a specific example, we consider the valley-Hall domain wall created by inversion in feasible honeycomb staggered arrays of waveguides laser-written in glass, with parameters comparable to those used in the experiments of Ref. [24]. In this context, a nonlinear optical response arises from the intensity-dependent refractive index of the ambient glass medium.

Starting from simulations within the standard paraxial equation governing the electric field $\mathcal{E}(x, y, z)$ and light propagation in optical lattices [14,16–18,24,25], $i\partial_z\mathcal{E}(x, y, z) = \hat{H}_{\text{par}}(x, y, |\mathcal{E}(x, y, z)|^2)\mathcal{E}(x, y, z)$, we aim to obtain an equation for the slowly varying envelope $\mathcal{A}(x, z)$ of the edge wave packet, localized on the domain wall. Here, $\hat{H}_{\text{par}}(x, y, |\mathcal{E}|^2)$ represents a complex paraxial operator, incorporating diffraction, the spatially varying refractive index profile [as depicted in Fig. 1(a)], and the nonlinear Kerr effect in the medium. Our objective is to obtain an equation of the form $i\partial_z\mathcal{A}(x, z) =$

$\hat{H}_{\text{PDE}}(\mathcal{A}(x, z), |\mathcal{A}(x, z)|^2, \dots)\mathcal{A}(x, z)$ [17,18], where \hat{H}_{PDE} describes the underlying PDE model, with its constituent terms schematically illustrated in Fig. 1(c). This dimensional reduction, aided by an ML-assisted approach, simplifies the analysis of wave processes in complex systems, particularly phenomena at interfaces and topological boundaries.

We determine the PDE model governing the evolution of the edge wave packet envelope as it propagates along the domain wall using a sparse-regression method based on data obtained from direct modeling. Our approach is summarized in Fig. 1. The lattice geometry in the xy plane, transverse to the waveguide axis z , is depicted on the left in Fig. 1. While the wave packet propagates in xz plane, it remains localized along the y axis, forming an edge wave packet bound to the interface. We seek a continuum model that is universally applicable to relatively broad wave packets, with widths significantly exceeding the lattice period. The envelope function profile is extracted from the waveguides at the interface. In view of discreteness, more complete data are accumulated by sweeping the position of the input beam between the lattice sites.

We use a split-step strategy to differentiate between linear and nonlinear scenarios and switch between them by tailoring the input power magnitude. Accordingly, for convenience, we represent the complex-valued function \mathcal{A} either in terms of its real and imaginary parts or in terms of its intensity and phase. The regression is used to identify a subset of relevant terms from a large library of potential functions that best replicate the system's dynamics [26]. The (x, z) -dependent datasets \mathcal{A} , whose derivatives $\partial_z\mathcal{A}$ are denoted as \mathbf{Y} in Fig. 1, are used to create a library of candidate functions represented by the matrix \mathbf{X} . We then iteratively solve an optimization problem to obtain a sparse vector of coefficients \mathbf{c} , representing the unknown dynamical equations (see also the Supplemental Material [27]).

While the initial library contains an extended set of possible functions, the scheme allows filtering out absent or inessential contributions and retaining only the physical effects relevant for the studied propagation distances. The nonlinear terms in the PDE model at higher intensities are obtained as a refinement of the linear differential operator, initially reconstructed for low intensities in the linear regime. This approach is somewhat analogous to the split-step solution for evolutionary problems and ML boosting, which iteratively improves the predictive power of the model by refining its weaker predictive versions [28].

II. DATA COLLECTION

To model the paraxial evolution of light through the optical lattice, with the lattice profile in the xy plane and extended along the z axis, we employ the paraxial wave equation for the field \mathcal{E} :

$$i\partial_z\mathcal{E} + \frac{1}{2k_0}(\partial_{xx} + \partial_{yy})\mathcal{E} + \frac{k_0 n_L(x, y)}{n_0}\mathcal{E} + n_2|\mathcal{E}|^2\mathcal{E} = 0, \quad (1)$$

where z is aligned parallel to the waveguides axis, $k_0 = 2\pi n_0/\lambda$ is the wave number, n_0 is the background refractive index, $n_L(x, y)$ is the perturbation of the refractive index forming the geometry of the lattice, $|n_L(x, y)| \ll n_0$. Equation (1)

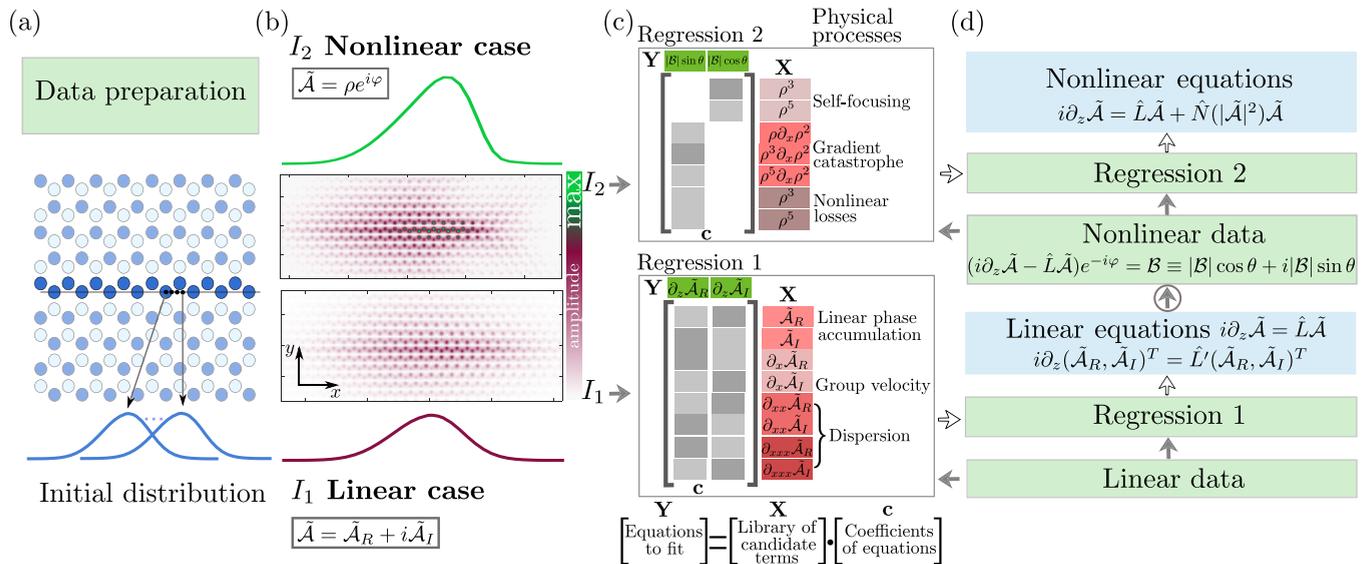


FIG. 1. Schematic of the methodology for extracting the partial differential equation (PDE) model with machine learning using regression and numerically calculated datasets. (a) Refractive index distribution in the (x, y) plane. The perturbations n_A and n_B [see Eq. (2)] are shown in different shades of blue. Waveguides forming the domain wall at $y = 0$ (i.e., the interface) are highlighted in dark blue (n_B). (b) Edge wave packet profiles for $z > 0$ in the xy plane, localized at the interface $y = 0$, obtained by numerically solving Eq. (1). Their envelopes along x at the domain wall are visualized as line plots. Bottom: linear propagation (low initial intensity), with the convenient representation for the complex-valued envelope function $\tilde{A} = \tilde{A}_R + i\tilde{A}_I$. Top: nonlinear propagation (high initial intensity), with the convenient intensity-phase representation $\tilde{A} = \rho e^{i\varphi}$. (c) Schematic of the regression algorithms for the linear (bottom) and nonlinear (top) regimes. The schematic illustrates the library of candidate functions (X) corresponding to possible physical processes, target functions (Y), and fitted equation coefficients (c), which constitute the unknown dynamical equations. (d) Pipeline of the procedure. The sequence of steps includes: data collection through paraxial modeling at low input powers (linear data), reconstruction of the linear operator terms \hat{L} (Regression 1), analysis of the data calculated at higher input powers (nonlinear data), and solving the regression problem for the expression $(i\partial_z \tilde{A} - \hat{L}\tilde{A})e^{-i\varphi}$ (Regression 2), refining the PDE by including nonlinear terms. Filled arrows indicate data loading, empty arrows represent coefficient determination (i.e., identifying the appropriate library of functions), and an encircled arrow denotes the transition from the linear to the nonlinear case, where the coefficients describing the linear operator are utilized during the fitting of the nonlinear equations.

also includes a nonlinear term $\propto n_2 |\mathcal{E}|^2 \mathcal{E}$, which is responsible for the focusing cubic Kerr-type nonlinearity.

The lattice geometry is imprinted into a refractive index distribution

$$n_L(x, y) = n_A \sum_{n,m} F(x - x_{nm}, y - y_m) + n_B \sum_{n,m} F(x - x_{nm}, y - y_m - a_0). \quad (2)$$

Here, A and B are the indices labeling the two triangular sublattices of the graphene lattice, which has a spatial period $a = \sqrt{3}a_0$, and n_A and n_B refer to the perturbations of the refractive index in waveguides. The Gaussian-shaped elliptical waveguides have semiaxes L_x and L_y , described by the function $F(x, y) = e^{-x^2/L_x^2 - y^2/L_y^2}$. The parameters corresponding to different degrees of parity breaking are listed in Table I.

The interface between two domains is created by swapping the refractive index perturbations between two sublattices, $n_A \rightarrow n_B, n_B \rightarrow n_A$. This results in a domain wall where the refractive index perturbation is the same across the interfacing neighboring elements, as highlighted by the gray rectangle in Figs. 2(a) and 2(b). In a dimerized honeycomb lattice, this interface can take two distinct shapes: zigzag and bearded shapes.

To prepare datasets, we solve the paraxial Eq. (1) numerically. First, the plane wave expansion method is employed to obtain the edge mode profile transverse to the interface, $u(x, y)$, with the substitution $\mathcal{E} = u(x, y)e^{-i\beta z + ikx}$ and setting $n_2 = 0$. The band structure $\beta(k)$ is obtained by solving the stationary form of the paraxial equation in the linear regime, which leads to the following eigenvalue problem for $u(x, y)$ and β : $\beta u = -\frac{1}{2k_0} [(\partial_x + ik)^2 + \partial_y^2] u - \frac{k_0}{n_0} n_L(x, y) u$ within a supercell geometry [indicated in Figs. 2(a) and 2(b) by the violet rectangles]. According to Bloch's theorem, applied to the supercell, $u(x, y) = u(x + a, y)$ is a periodic function. The

TABLE I. Two sets of lattice parameters simulated using paraxial modeling.

Parameter	Set I	Set II
L_x	3.2 μm	4.9 μm
L_y	4.9 μm	3.2 μm
a_0	20 μm	18.5 μm
n_0	1.47	1.47
n_A	2.6×10^{-3}	7.5×10^{-4}
n_B	2.8×10^{-3}	12.4×10^{-4}
n_2	$3 \times 10^{-20} \text{ m}^2/\text{W}$	$3 \times 10^{-20} \text{ m}^2/\text{W}$
λ	1650 nm	1045 nm
I_0	10^{16} W/m^2	10^{16} W/m^2

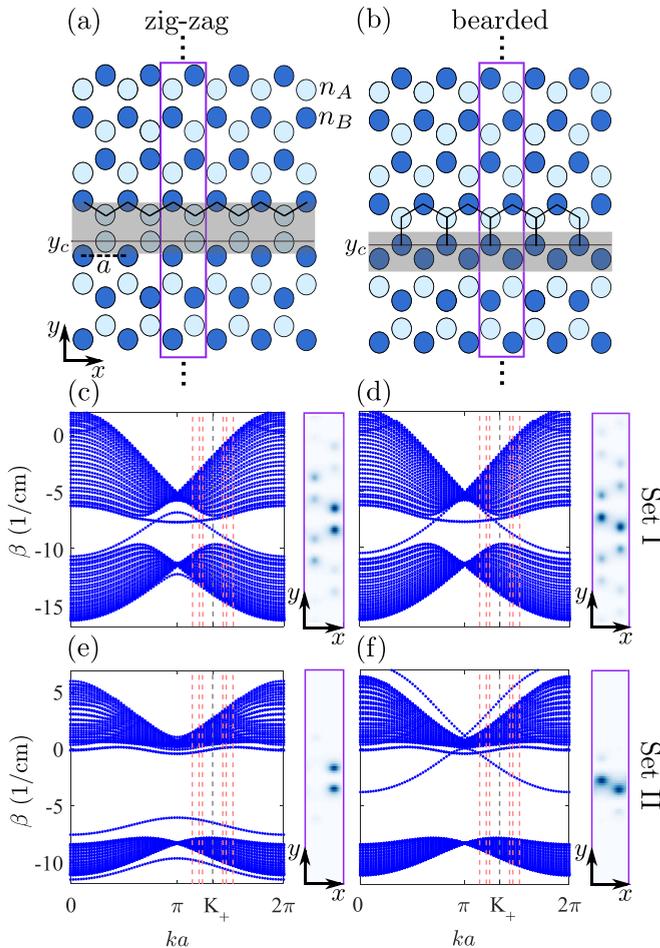


FIG. 2. (a), (b) Two distinct shapes of valley-Hall domain walls in a honeycomb lattice of laser-written waveguides: (a) zigzag and (b) bearded. Parameter a denotes the spatial lattice period. Supercells of these configurations are framed in violet rectangles, with the interface area of the domain wall shaded in gray. Panels (c)–(f) show the band structures, $\beta(k)$, for two parameter sets: Set I (upper row) and Set II (lower row). Calculations are performed for supercells of the staggered graphene lattice composed of 64 waveguides, each having the same elliptical shape (see parameters for Set I and Set II in Table I), but differing in the perturbations of the refractive indices: for 32 waveguides in sublattice A, the refractive index is n_A (light blue), and for the remaining 32 waveguides in sublattice B, it is n_B (dark blue). The distinct domain walls have zigzag (c), (e) and bearded (b), (d) shapes, with open boundary conditions along the y -direction. The gray dashed line marks the position of wavenumber $K_+ = 4\pi/(3a)$. The red dashed vertical lines indicate the three scanning beam widths in the momentum space, centered at the wavenumbers $(K_+, a \pm \delta_{1,2,3})$, where $\delta_{1,2,3} = a/\mathcal{L}_{1,2,3}$ with $\mathcal{L}_1 = 3a$, $\mathcal{L}_2 = 3a/2$, and $\mathcal{L}_3 = 3a/2 \times 1.5$. The corresponding transverse intensity profiles of the plane-wave-like edge state $|u(x, y)|^2$ at the K_+ point are shown in the violet frames on the left. These spatial distributions along the y axis are associated with several waveguides positioned adjacent to domain walls, with the corresponding supercell configurations schematically depicted in panels (a) and (b).

supercell consists of 64 waveguides in total, ensuring that the computational domain is sufficiently wide to prevent interactions between the interface mode and the mode localized at the outer boundary of the lattice during edge wave packet

propagation for both sets of parameters. Periodic boundary conditions are applied in the x direction, and we consider open boundary conditions in the y direction, that is, the termination of the periodic modulation of the refractive index at the ends of the lattice. The electric field satisfies Dirichlet boundary conditions, $u(y = y_{\max}) = u(y = y_{\min}) = 0$, at the edges of the computational domain, which extends beyond the lattice to ensure proper field decay in the cladding region. In Figs. 2(c)–2(f), we plot the band structure $\beta(k)$ for the projected Brillouin zone over the interval $ka \in [0, 2\pi]$, which is mathematically equivalent to the symmetric interval $ka \in [-\pi, \pi]$. This range contains two valleys near the points $K_- = 2\pi/(3a)$ and $K_+ = 4\pi/(3a)$. Note that away from the bandgap, the spectrum forms nearly a continuum, a feature characteristic of bulk modes.

Then, the beam propagation method is used to simulate the evolution of the beams as they propagate along the waveguides' axis z . The beams are confined to the interface and localized along the interface with a Gaussian profile $f(x, \bar{x}_0) = f_0 e^{-(x-\bar{x}_0)^2/(2\mathcal{L}^2)}$ centered at \bar{x}_0 . Accordingly, the initial condition is set as $\mathcal{E}(z = 0, x, y, \bar{x}_0) = f(x, \bar{x}_0)u(x, y)e^{ik_0 x}$. To capture general features, nonspecific to initial conditions, and to deduce the corresponding effective model within reasonable propagation distances, simulations are performed for several different beam widths, \mathcal{L} , in the linear case, and then for several values of the initial intensity in the nonlinear case. Here \mathcal{L} is chosen to be $\mathcal{L} \gg a$ to ensure the applicability of the slowly varying envelope approximation.

To describe the evolution in terms of the envelope function, we extract the values of the envelope at the centers of the waveguides forming the domain wall (to ensure the correct mapping to the continuum limit), $\tilde{\mathcal{A}}(z, x_m, \bar{x}_0) = \mathcal{E}(z, x \equiv x_m, y \equiv y_c, \bar{x}_0)$, where y_c is the vertical coordinate of the waveguides comprising the interface [see Figs. 2(a) and 2(b)]. The sparse nature of the function $\tilde{\mathcal{A}}(z, x_m, \bar{x}_0)$ along the x -axis, defined within the waveguides at points $x = x_m$, presents a potential challenge in accurately determining the governing continuum equation, where fine discretization is typically assumed. The latter is particularly important for providing an accurate approximation of derivatives in the PDE model. To circumvent this challenge, we set a range of initial conditions and calculate multiple envelopes to generate more complete data. Specifically, we mesh the intervals $[x_m, x_{m+1}] = [x_m, x_m + \Delta_x q]$, where the number of steps $q = 32$ and step size $\Delta_x = a/q$, and sweep \bar{x}_0 such that $\bar{x}_0(n) = \bar{x}_0 + \Delta_x n$, $n = [1, \dots, q]$. This enables us to perform several calculations of $\tilde{\mathcal{A}}(z, x_m, \bar{x}_0(n))$, which can then be combined into a smooth function $\tilde{\mathcal{A}}(z, x)$.

In Ref. [18], we used an analytical asymptotic procedure to derive the evolution equation for the slowly varying amplitude $\tilde{\mathcal{A}}(z, x)$ of edge wave packets in the continuum limit, applicable regardless of the shape of the domain wall, to both zigzag and bearded cuts,

$$i \frac{\partial \tilde{\mathcal{A}}}{\partial z} = iv \frac{\partial \tilde{\mathcal{A}}}{\partial x} - G |\tilde{\mathcal{A}}|^2 \tilde{\mathcal{A}} - iv_g |\tilde{\mathcal{A}}|^2 \frac{\partial |\tilde{\mathcal{A}}|^2}{\partial x} \tilde{\mathcal{A}} - \eta \frac{\partial^2 \tilde{\mathcal{A}}}{\partial x^2} + \beta_0 \tilde{\mathcal{A}}. \quad (3)$$

To derive this equation, one can introduce a small parameter μ and assume that $\mu \propto |G/\beta_0| \times \max(|\tilde{\mathcal{A}}|^2)$, and that

$|\eta\beta_0| \propto \mu^2$. This implies that the asymptotic scope is always limited by *a priori* assumptions for the lattice parameters, which narrow the range of applicability to specific conditions. Instead of performing series expansions at different orders, the aim of the present study is to determine the structure of the equation and its coefficients using ML techniques and data for the numerically extracted envelope $\mathcal{A}(z, x)$. For this, we developed custom code implementing the linear regression algorithm.

We mix data prepared using the evolution profiles of wave packets with all beam widths (all intensities), randomly select points (z_j, x_j) , and compute the matrices \mathbf{Y}, \mathbf{X} by utilizing $\mathcal{A}(z_j, x_j)$ and its numerically obtained derivatives. When collecting data, we use only the points where the intensity exceeds a certain threshold, which is fixed for the entire procedure [27]. For our purposes, it is essential to focus on and accurately describe the dynamics of the field within the beam localization region. Including all data points, especially those with very low values, can introduce uncertainty and ambiguity in phase calculations and lead to inaccurate derivative computations. Randomly shuffled data points are divided into 80% for training and 20% for testing. Coefficients \mathbf{c} are determined from the larger training set by minimizing $\|\mathbf{X}\mathbf{c} - \mathbf{Y}\|^2$ (see the Supplemental Material [27]), and then these coefficients are tested on the test set. Additionally, the coefficients are validated by comparing numerical solutions of the full paraxial and the PDE models [27] for datasets not used in training.

III. LINEAR LOW-INTENSITY REGIME

We begin with the low-intensity linear case. The envelope equation can be formulated as a system for the real and imaginary parts in the representation $\tilde{\mathcal{A}} = \tilde{\mathcal{A}}_R + i\tilde{\mathcal{A}}_I$. We assume this system has the following form:

$$\begin{aligned} \frac{\partial \tilde{\mathcal{A}}_R}{\partial z} &= \beta_0 \tilde{\mathcal{A}}_I + \beta_{0I} \tilde{\mathcal{A}}_R + v \frac{\partial \tilde{\mathcal{A}}_R}{\partial x} - v_I \frac{\partial \tilde{\mathcal{A}}_I}{\partial x} \\ &\quad - \eta \frac{\partial^2 \tilde{\mathcal{A}}_I}{\partial x^2} - \eta_I \frac{\partial^2 \tilde{\mathcal{A}}_R}{\partial x^2} + \eta' \frac{\partial^3 \tilde{\mathcal{A}}_I}{\partial x^3} + \eta'_I \frac{\partial^3 \tilde{\mathcal{A}}_R}{\partial x^3}, \end{aligned} \quad (4a)$$

$$\begin{aligned} \frac{\partial \tilde{\mathcal{A}}_I}{\partial z} &= -\beta_0 \tilde{\mathcal{A}}_R + \beta_{0I} \tilde{\mathcal{A}}_I + v \frac{\partial \tilde{\mathcal{A}}_I}{\partial x} + v_I \frac{\partial \tilde{\mathcal{A}}_R}{\partial x} \\ &\quad + \eta \frac{\partial^2 \tilde{\mathcal{A}}_R}{\partial x^2} - \eta_I \frac{\partial^2 \tilde{\mathcal{A}}_I}{\partial x^2} - \eta' \frac{\partial^3 \tilde{\mathcal{A}}_R}{\partial x^3} + \eta'_I \frac{\partial^3 \tilde{\mathcal{A}}_I}{\partial x^3}. \end{aligned} \quad (4b)$$

The library of functions consists of differential operators of various orders for recovering spatial dispersion. This essentially addresses the problem of reconstructing the dispersion of the edge state by scanning it (setting different k_0) with a beam of finite spectral width in the vicinity of the specified wave vector k_0 (see Fig. 3). The real and imaginary parts representation is optimal for this problem, as it enables a more compact function library for differential operators compared with the intensity-phase representation.

We generated a dataset by sequentially launching pulses with varying beam widths: specifically, $\mathcal{L}_1 = 3a$, $\mathcal{L}_2 = 3a/2$, and $\mathcal{L}_3 = (3a/2) \times 1.5$ (their spectral widths in momentum space are shown in Fig. 2). This approach was designed to capture different aspects of the system's response. Narrower pulses provide higher resolution for estimating dispersion and

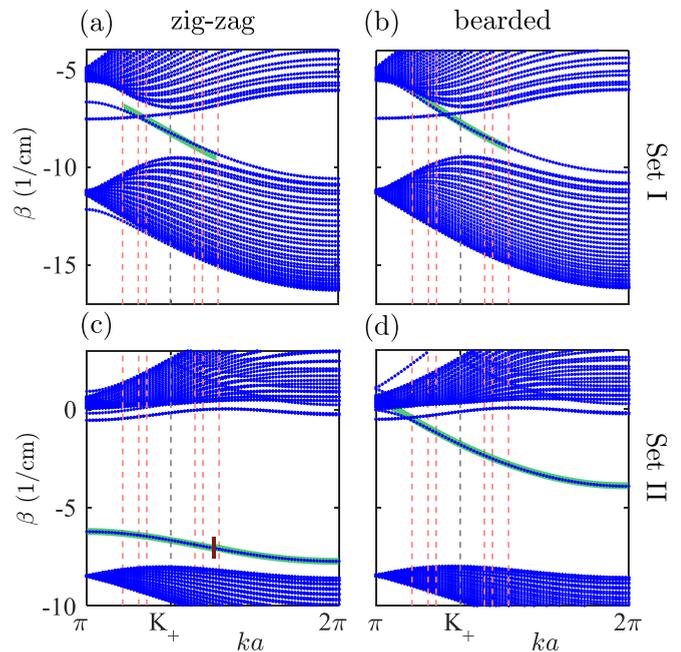


FIG. 3. Comparison between the numerical (blue dots, as in Fig. 2) edge-state dispersion curve in the band structures and the dispersion plotted using machine learning (ML)-determined coefficients (green lines). Parameters for panels (a)–(d) correspond to Figs. 2(c)–2(f), respectively. Panels (a), (b), and (d) show the ML-based approximation recovered near $k_0 = K_+$. In plot (c), the approximations to the left and right of the brown vertical line segment are recovered near $k_0 = K_+$ and $k_0 = 2\pi$, respectively. Dashed vertical lines visualize the three scanning beam widths, similar to Fig. 2.

pulse broadening, while the wider pulse allows for better extraction of the linear velocity parameter. By combining data from different excitation beam widths, we can obtain equations that apply universally to various beam widths.

The schematic organization of the library is partially visualized in Fig. 1. Even when we assume all the coefficients in the equations for $\tilde{\mathcal{A}}_R$ and $\tilde{\mathcal{A}}_I$ are different, some of these coefficients group according to physical processes, while others are minor and can be omitted, following the principles of sparse regression [27]. The coefficients of the dominant terms are then further refined. The leading terms at short propagation distances stem from linear phase accumulation, drift with group velocity, and second-order dispersion, which causes gradual symmetric broadening. These are described by the coefficients $\langle \mathbf{c}_i \rangle = \langle \beta_0 \rangle, \langle v \rangle, \langle \eta \rangle$ and their standard deviations $\delta \mathbf{c}_i$ calculated on 100 validation folds of the training dataset [27].

The obtained nonzero terms are the coefficients that best describe our system. We may verify these coefficients by reconstructing the edge-state dispersion relation near different k_0 , as depicted by the green lines in Fig. 3:

$$\beta_e = \beta_0 - v(k - k_0) + \eta(k - k_0)^2 + \eta'_I(k - k_0)^3. \quad (5)$$

Notice that in Fig. 3, we show an enlarged view of Figs. 2(c)–2(f) over the range $ka \in [\pi, 2\pi]$, which encompasses a single valley in the vicinity of the point $k_0 \equiv K_+$.

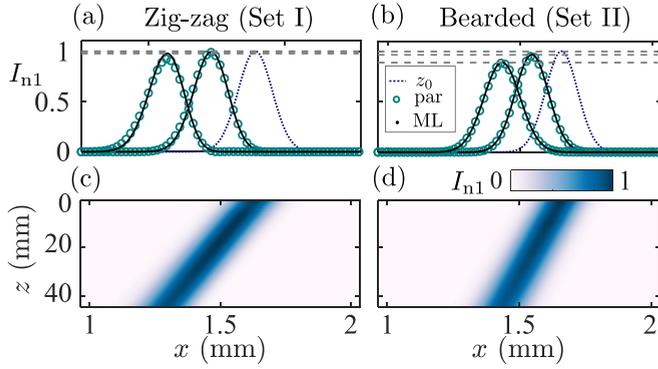


FIG. 4. Evolution of the beam envelope at zigzag (left column) and bearded (right column) domain walls with $\mathcal{L}_t = 3a/2 \times 1.8$ at low input intensity, $I(z=0) < 0.04 \times I_0$. Plotted is the intensity, normalized to its maximum, I_{n1} . (a), (b) Snapshots taken at three propagation distances: at the input ($z = z_0 = 0$, blue dotted line), at $z = 22$ mm, and at $z = 44$ mm. Green circles: paraxial modeling results for the wave packet's envelope obtained using Eq. (1); black solid lines: numerical solution of PDE Eq. (4) with machine learning-determined coefficients. (c), (d) Spatial profile mapped along the z -directed propagation. For the domain wall of the zigzag and bearded shapes, we operate at $k_0 = K_+ = 4\pi/(3a)$ and $k_0 = (4\pi/3 - 0.4)/a$, respectively.

The determined coefficients can additionally be validated by analyzing the spatial dynamics of the system. To do so, we numerically solve PDE Eq. (4) with recovered nonzero terms, setting the Gaussian profile $f(x, \bar{x}_0) = f_0 e^{-(x-\bar{x}_0)^2/(2\mathcal{L}_t^2)}$ as the initial condition at $z = 0$, where \mathcal{L}_t is the beam width different from $\mathcal{L}_{1,2,3}$. The regimes of undistorted propagation and broadening, accompanied by a decrease in amplitude, are illustrated in Fig. 4. The results obtained from the full paraxial modeling show strong agreement with those derived from the simplified envelope dynamics, as described by the recovered PDE. At longer propagation distances, the model can be further refined by incorporating third-derivative terms responsible for the asymmetric distortions.

IV. NONLINEAR HIGH-INTENSITY REGIME

As the optical power increases, we anticipate the manifestations of nonlinear effects such as self-focusing or self-steepening originating from the nonlinear group velocity. Therefore, it is natural to focus on these effects and evaluate whether they suffice to accurately describe the beam's transformation at certain distances. The library of functions (see also Ref. [27]) is assembled to fit the following equation:

$$\begin{aligned}
 i \frac{\partial \tilde{\mathcal{A}}}{\partial z} = & iv \frac{\partial \tilde{\mathcal{A}}}{\partial x} - \eta \frac{\partial^2 \tilde{\mathcal{A}}}{\partial x^2} + i\eta_l \frac{\partial^3 \tilde{\mathcal{A}}}{\partial x^3} + \beta_0 \tilde{\mathcal{A}} \\
 & - i(v_{g1} \tilde{\mathcal{A}} + v_{g2} |\tilde{\mathcal{A}}|^2 \tilde{\mathcal{A}} + v_{g3} |\tilde{\mathcal{A}}|^4 \tilde{\mathcal{A}}) \frac{\partial |\tilde{\mathcal{A}}|^2}{\partial x} \\
 & - (G_1 |\tilde{\mathcal{A}}|^2 + G_2 |\tilde{\mathcal{A}}|^4) \tilde{\mathcal{A}} + i(\gamma_1 |\tilde{\mathcal{A}}|^2 + \gamma_2 |\tilde{\mathcal{A}}|^4) \tilde{\mathcal{A}}.
 \end{aligned} \tag{6}$$

Within the intensity-phase representation, $\tilde{\mathcal{A}} = \rho e^{i\varphi}$, it is rewritten as a system of equations for ρ and φ . The last two terms can be interpreted as nonlinear losses, which may occur

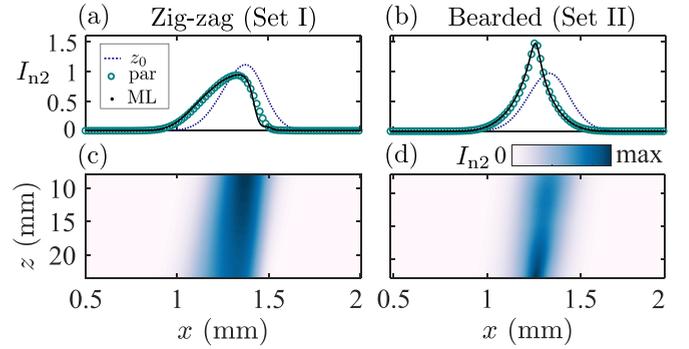


FIG. 5. The nonlinear scenarios of beam propagation of the width $\mathcal{L} = 5a$ at zigzag (left column) and bearded (right column) domain walls. Similar to Fig. 4, panels (a), (b) show snapshots taken at the input ($z = z_0 = 7.5$ mm, blue dotted line) and at $z = 23$ mm, with the central wavenumbers the same as in Fig. 4. The initial intensity is $I(z=0) = (0.45)^2 \times I_0$, plotted is $I_{n2} = I(z)/[(0.45)^2 \times I_0]$. (c), (d) Nonlinear beam transformation mapped along z .

due to interactions with bulk modes in the nonlinear regime at increasing intensity [18], causing possible energy dissipation from the edge state into the bulk [27]. In other words, even though the paraxial model Eq. (1) is lossless, energy may escape from the domain wall via nonlinearity-induced coupling into bulk states, corresponding to the effective nonlinear loss terms in Eq. (6). Remarkably, nonlinear effects appear separated in this framework: self-focusing contributes to the phase, while loss and self-steepening effects are incorporated into the intensity equation. We assume that the coefficients of the linear operator have already been determined from the low-intensity analysis. Therefore, we now solve the regression problem for the difference between the evolution operator and the linear operator, multiplied by the phase (see also the scheme in Fig. 1), $(i\partial_z \tilde{\mathcal{A}} - \hat{\mathcal{L}} \tilde{\mathcal{A}}) e^{-i\varphi} \equiv |\mathcal{B}| \cos \theta + i|\mathcal{B}| \sin \theta$, which can be attributed to the nonlinear correction $\hat{N} \tilde{\mathcal{A}}$ showing up at higher intensities,

$$-|\mathcal{B}| \cos \theta = -\beta_0 \rho + G_1 \rho^3 + G_2 \rho^5, \tag{7a}$$

$$|\mathcal{B}| \sin \theta = -(v_{g1} \rho + v_{g2} \rho^3 + v_{g3} \rho^5) \frac{\partial \rho^2}{\partial x} + \gamma_1 \rho^3 + \gamma_2 \rho^5. \tag{7b}$$

In our datasets, we include evolution profiles of beams with three different intensity values at the input ($z = 0$): $I_1(z=0) = (0.5)^2 \times I_0$, $I_2(z=0) = (0.425)^2 \times I_0$, and $I_3(z=0) = (0.35)^2 \times I_0$, while keeping the beam width fixed at $\mathcal{L} = 5a$.

The determined coefficients (see exact values in the Supplemental Material [27]) confirm the presence of two competing major nonlinear effects, namely self-steepening and self-focusing. Figure 5 shows representative cases of the nonlinear dynamics, similar to Fig. 4, which showcases the linear regime. Specifically, Fig. 5 presents the evolution of the wave packet with an initial intensity of $I_1(z=0) = (0.45)^2 \times I_0$. This input field intensity, distinct from the dataset intensities, allows us to validate our analysis. A growing asymmetry of the wave packet is visible in Figs. 5(a) and 5(c). This self-steepening deformation occurs due to the prevailing nonlinear velocity term v_{g2} [18,25]. In contrast, gradual self-compression, a hallmark of self-focusing [14,17,18],

is evident in Figs. 5(b) and 5(d). The latter behavior is typical of the nonlinear Schrödinger equation with attractive nonlinearity and dominant quadratic spatial dispersion. Notice that while our presumed model allows for the presence of nonlinear losses with coefficients $\gamma_{1,2}$, their effect remains negligible in our system due to their small magnitude (see Table SII in the Supplemental Material [27]).

Which nonlinear effect is dominant depends on the microscopic lattice parameters. Conventionally, the corresponding nonlinear coefficients would be computed in terms of integrals of the edge states' spatial profiles transverse to the interface when applying a conventional asymptotic multiscale analysis [16,18]. The present data-driven approach allows one to bypass this step and infer the corresponding nonlinear coefficients purely using beam propagation simulations.

V. CONCLUSION

In this study, we demonstrated how ML can effectively unveil the continuum PDE model governing edge waves confined to domain walls in photonic lattices. It enables the revelation of various effects, including spatial dispersion and higher-order nonlinearities, and the resultant corrections to the nonlinear Schrödinger equation, such as the nonlinear velocity term, for domain walls of different cuts. Thus, our approach

provides a valuable alternative to traditional fully analytical asymptotic methods for addressing wave dynamics in nonlinear and crystalline systems, allowing for the exploration of distinct topological phases and being applicable beyond photonics to various wave media. It can readily be extended to other types of nonlinearities and interparticle interactions, for example, in polaritonic platforms.

ACKNOWLEDGMENTS

E.S. and L.S. are supported in part by the Ministry of Science and Higher Education of the Russian Federation (Contract FSWR-2024-0005). E.S. thanks the Foundation for the Advancement of Theoretical Physics and Mathematics, "BASIS" (Grant No. 22-1-5-80-1). D.L. acknowledges support from the Ministry of Education, Singapore, under its SUTD Kickstarter Initiative (Grant No. SKI 20210501). F.N. is supported in part by: the Japan Science and Technology Agency (JST) [via the CREST Quantum Frontiers program Grant No. JPMJCR24I2, the Quantum Leap Flagship Program (Q-LEAP), and the Moonshot R&D Grant No. JPMJMS2061], and the Office of Naval Research (ONR) Global (via Grant No. N62909-23-1-2074). D.S. acknowledges support from the Australian Research Council (FT230100058).

-
- [1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [2] Z. Chen, Y. Liu, and H. Sun, Physics-informed learning of governing equations from scarce data, *Nat. Commun.* **12**, 6136 (2021).
- [3] L. Pilozzi, F. A. Farrelly, G. Marcucci, and C. Conti, Machine learning inverse problem for topological photonics, *Commun. Phys.* **1**, 57 (2018).
- [4] J. Yun, S. Kim, S. So, M. Kim, and J. Rho, Deep learning for topological photonics, *Adv. Phys.: X* **7**, 2046156 (2022).
- [5] A. A. Kaptanoglu, B. M. de Silva, U. Fasel, K. Kaheman, A. J. Goldschmidt, J. Callahan, C. B. Delahunt, Z. G. Nicolaou, K. Champion, J.-C. Loiseau, J. N. Kutz, and S. L. Brunton, PySINDy: A comprehensive Python package for robust sparse system identification, *J. Open Source Softw.* **7**, 3994 (2022).
- [6] B. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Kutz, and S. Brunton, PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data, *J. Open Source Softw.* **5**, 2104 (2020).
- [7] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.* **3**, e1602614 (2017).
- [8] K. J. N. Kaheman Kadierdan and B. S. L., SINDy-Pi: A robust algorithm for parallel implicit sparse identification of nonlinear dynamics, *J. Open Source Softw.* **476**, 20200279 (2020).
- [9] M. Cranmer, Interpretable machine learning for science with PySR and SymbolicRegression.jl, [arXiv:2305.01582](https://arxiv.org/abs/2305.01582).
- [10] M. Du, Y. Chen, and D. Zhang, DISCOVER: Deep identification of symbolically concise open-form partial differential equations via enhanced reinforcement learning, *Phys. Rev. Res.* **6**, 013182 (2024).
- [11] R. R. Naik, A. Tiuhonen, J. Thapa, C. Batali, Z. Liu, S. Sun, and T. Buonassisi, Discovering equations that govern experimental materials stability under environmental stress using scientific machine learning, *npj Comput. Mater.* **8**, 72 (2022).
- [12] A. V. Ermolaev, A. Sheveleva, G. Genty, C. Finot, and J. M. Dudley, Data-driven model discovery of ideal four-wave mixing in nonlinear fibre optics, *Sci. Rep.* **12**, 12711 (2022).
- [13] T. Ozawa, H. M. Price, A. Amo, N. Goldman, M. Hafezi, L. Lu, M. C. Rechtsman, D. Schuster, J. Simon, O. Zilberberg, and I. Carusotto, Topological photonics, *Rev. Mod. Phys.* **91**, 015006 (2019).
- [14] Y. V. Kartashov and D. V. Skryabin, Modulational instability and solitary waves in polariton topological insulators, *Optica* **3**, 1228 (2016).
- [15] Y. Lumer, M. C. Rechtsman, Y. Plotnik, and M. Segev, Instability of bosonic topological edge states in the presence of interactions, *Phys. Rev. A* **94**, 021801 (2016).
- [16] D. A. Smirnova, L. A. Smirnov, D. Leykam, and Y. S. Kivshar, Topological edge states and gap solitons in the nonlinear Dirac model, *Laser Photonics Rev.* **13**, 1900223 (2019).
- [17] Q. Tang, B. Ren, V. O. Kompanets, Y. V. Kartashov, Y. Li, and Y. Zhang, Valley Hall edge solitons in a photonic graphene, *Opt. Express* **29**, 39755 (2021).
- [18] D. A. Smirnova, L. A. Smirnov, E. O. Smolina, D. G. Angelakis, and D. Leykam, Gradient catastrophe of nonlinear photonic valley-Hall edge pulses, *Phys. Rev. Res.* **3**, 043027 (2021).

- [19] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci.* **113**, 3932 (2016).
- [20] S. Yang, S. Chen, W. Zhu, and P. G. Kevrekidis, Identification of moment equations via data-driven approaches in nonlinear Schrödinger models, *Front. Photonics* **5**, 1444993 (2024).
- [21] M. J. Ablowitz and J. T. Cole, Tight-binding methods for general longitudinally driven photonic lattices: Edge states and solitons, *Phys. Rev. A* **96**, 043868 (2017).
- [22] S. K. Ivanov, Y. V. Kartashov, A. Szameit, L. Torner, and V. V. Konotop, Vector Topological Edge Solitons in Floquet Insulators, *ACS Photon.* **7**, 735 (2020).
- [23] S. K. Ivanov, Y. V. Kartashov, L. J. Maczewsky, A. Szameit, and V. V. Konotop, Edge solitons in Lieb topological Floquet insulator, *Opt. Lett.* **45**, 1459 (2020).
- [24] J. Noh, S. Huang, K. P. Chen, and M. C. Rechtsman, Observation of photonic topological valley Hall edge states, *Phys. Rev. Lett.* **120**, 063902 (2018).
- [25] E. O. Smolina, L. A. Smirnov, D. Leykam, and D. A. Smirnova, Self-steepening-induced stabilization of nonlinear edge waves at photonic valley-Hall interfaces, *Phys. Rev. A* **108**, L061501 (2023).
- [26] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B* **58**, 267 (1996).
- [27] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/2jh8-p5y2> for more detailed information on data preparation and analysis across both linear and nonlinear regimes.
- [28] R. E. Schapire, *The Boosting Approach to Machine Learning: An Overview* (Springer, New York, 2003), pp. 149–171.