

PAPER • OPEN ACCESS

Morphology of three-body quantum states from machine learning

To cite this article: David Huber *et al* 2021 *New J. Phys.* **23** 065009

View the [article online](#) for updates and enhancements.



PAPER




Morphology of three-body quantum states from machine learning

OPEN ACCESS

RECEIVED
3 March 2021REVISED
21 May 2021ACCEPTED FOR PUBLICATION
26 May 2021PUBLISHED
23 June 2021

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

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

David Huber¹, Oleksandr V Marchukov² , Hans-Werner Hammer^{1,3,*}  and
Artem G Volosniev⁴ ¹ Technische Universität Darmstadt, Department of Physics, Institut für Kernphysik, 64289 Darmstadt, Germany² Technische Universität Darmstadt, Institut für Angewandte Physik, Hochschulstraße 4a, 64289 Darmstadt, Germany³ ExtreMe Matter Institute EMMI and Helmholtz Forschungsakademie Hessen für FAIR (HFHF), GSI Helmholtzzentrum für Schwerionenforschung GmbH, 64291 Darmstadt, Germany⁴ Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

* Author to whom any correspondence should be addressed.

E-mail: hans-werner.hammer@physik.tu-darmstadt.de

Keywords: quantum billiards, machine learning, impurity systems, quantum chaos

Abstract

The relative motion of three impenetrable particles on a ring, in our case two identical fermions and one impurity, is isomorphic to a triangular quantum billiard. Depending on the ratio κ of the impurity and fermion masses, the billiards can be integrable or non-integrable (also referred to in the main text as chaotic). To set the stage, we first investigate the energy level distributions of the billiards as a function of $1/\kappa \in [0, 1]$ and find no evidence of integrable cases beyond the limiting values $1/\kappa = 1$ and $1/\kappa = 0$. Then, we use machine learning tools to analyze properties of probability distributions of individual quantum states. We find that convolutional neural networks can correctly classify integrable and non-integrable states. The decisive features of the wave functions are the normalization and a large number of zero elements, corresponding to the existence of a nodal line. The network achieves typical accuracies of 97%, suggesting that machine learning tools can be used to analyze and classify the morphology of probability densities obtained in theory or experiment.

1. Introduction

The correspondence principle conjectures that highly excited states of a quantum system carry information about the classical limit [1]. In particular, it implies that there must be means to tell a difference between a ‘typical’ high-energy quantum state that corresponds to an integrable classical system from a ‘typical’ high-energy state that corresponds to a chaotic system. A discovery of such means is a complicated task that requires a coherent effort of physicists, mathematicians, and philosophers [2–4]. Currently, there are two main approaches to study chaotic features in quantum mechanics. One approach relies on the statistical analysis of the energy levels of a quantum-mechanical system. Another focuses on the morphology of wave functions. These approaches led to a few celebrated conjectures that postulate features of energy spectra and properties of eigenstates [5–7]. The postulates are widely accepted now, thanks to numerical as well as experimental data [8, 9].

Numerical and experimental data sets produced to confirm the proposed conjectures are so large that it is difficult, if not hopeless, for the human eye to find universal patterns beyond what has been conjectured. Therefore, it is logical to look for computational tools that can learn (with or without supervision) universal patterns from large datasets. One such tool is deep learning (DL) [10], which is a machine learning method that uses artificial neural networks (NNs) with multiple layers for a progressive learning of features from the input data. It requires very little engineering by hand, and can easily be used to analyze big data across disciplines, in particular in physics [11]. DL tools present an opportunity to go beyond the standard approaches of quantum chaologists [12]. For example, in this paper, NN built upon many states are used to

analyze the morphology of individual wave functions. Therefore, DL provides us with means to connect and extend tools already used to understand ‘chaos’ in quantum mechanics.

Recent work [13] has already opened an exciting possibility to study the quantum–classical correspondence in integrable and chaotic systems using DL. In particular, it has been suggested that a NN can learn the difference between wave functions that correspond to integrable and chaotic systems. It is important to pursue this research direction further, and to understand and interpret how an NN distinguishes the two situations. This information can be used in the future to formulate new conjectures on the role of classical chaos in quantum mechanics. The main challenge here is the extraction of this information from an NN, which often resembles a black box. Ongoing research on interpretability of NNs suggests certain routes to understand the black box [14–16] (see also recent works that discuss this question for applications in physical sciences [17–19]). However, there is no standard approach to this problem. In part, this is connected to the fact that DL relies on general-purpose learning procedures, therefore, one does not expect that there can be a unique way to analyze a NN at hand. For example, as we will see, the training of a network for the ‘integrable’ vs ‘chaotic’ state recognition is very similar to the classic dog-or-cat classification⁵. It is not clear, however, that the tools that can be used to interpret the latter (e.g., based on stable spatial relationships [20]) are also useful for the former. In particular, a training set for the ‘integrable’-or-‘chaotic’ problem contains information about vastly different length scales (determined by the energy), whereas a training set for cats vs dogs has only length scales given by the size of the animal. Therefore, it is imperative to study interpretability of NNs used in physics separately from that in other applications.

In this paper we analyze a NN, which has been trained using highly excited states of a triangular billiard, and attempt to extract the learned features. Billiards are conceptually simple systems, yet it is expected that they contain all necessary ingredients for studying the role of chaos in quantum mechanics [8]. Furthermore, eigenstates of quantum billiards are equivalent to the eigenstates of the Helmholtz equation with the corresponding boundary conditions, which connects quantum billiards and the wave chaos in microwave resonators [8, 21]. The triangular billiard is one of the most-studied models in quantum chaology [22–26], and therefore it is well-suited for our study focused on analyzing neural networks as a possible tool for quantum chaology.

In our analysis, we rely on convolutional neural networks (ConvNets) for image classification [27], which have recently been successfully applied to categorize numerical and experimental data in physical sciences [13, 28–35]. These advances motivate us to apply ConvNets to categorize quantum states as integrable and non-integrable. Our goal can be stated as follows: given a set of highly excited states, build a network that can classify any input state as integrable or not, and, moreover, study features of this network. One comment is in order here. There are various definitions of quantum integrability [36], so we need to be more specific. In this work, we call a quantum system integrable, if it is Bethe-ansatz integrable, i.e., if one can write any eigenstate as a finite superposition of plane waves. We shall also sometimes use the word chaotic instead of non-integrable. Finally, we note that the properties ‘integrable’ and ‘non-integrable’ are usually attached to a given physical system, e.g., following an analysis of global properties like the distribution of energy levels. However, the correspondence principle implies that these labels can also be applied to individual states of a quantum system. In this paper, we use both notions and show that they are compatible. We employ NNs to analyze the wave functions of individual quantum states.

We show that a trained network accurately classifies a state as being ‘integrable’ or ‘non-integrable’, which implies that a ConvNet learns certain universal features of highly-excited states. We argue that a trained NN considers almost any random state generated by a Gaussian, Laplace or other distribution as ‘chaotic’, as long as the state includes a sufficient amount of zero values. This observation agrees with our intuition that a non-integrable state has only weak correlations. We discuss the effect of the noise and coarse graining in our classification scheme, which sets limitations on the applicability of NN to analyze experimental and numerical data.

Our motivation for this work is thus threefold: first, we want to demonstrate that NNs can classify the morphology of the three-body states correctly. Therefore, we investigate a known model system with two identical fermions and an impurity as a function of the impurity mass in order to be able to verify the NN analysis. Second, we want to analyze the classifying network and understand the way it operates. Our third goal is to use the network analysis to clarify the situation regarding suggested new integrable cases for other parameter values than the established ones [37, 38]. However, we do not find any evidence of such cases.

The paper is organized as follows. In section 2 we introduce the system at hand: a triangular quantum billiard that is isomorphic to three impenetrable particles on a ring. Its properties are discussed in section 3

⁵ This figure is generated by adapting the code from https://github.com/gwding/draw_convnet.

using standard methods. In section 4, we present our NN approach and use it in section 5 to classify the states of the system. Moreover, we analyze the properties of the network. In section 6, we conclude. Some technical details are presented in the [appendix](#).

2. Formulation

We study billiards isomorphic to the relative motion of three impenetrable particles in a ring: two fermions and one impurity. Characteristics of these triangular billiards are presented below, see also reference [39, 40]. Our choice provides us with a simple parametrization of triangles in terms of the mass ratio, $\kappa = m_I/m$, where m_I (m) is the mass of the impurity (fermions). Furthermore, it allows us to shed light on the problem of three particles in a ring with broken integrability [41–43].

For simplicity, we always assume that the impurity is heavier than (or as heavy as) the fermions, corresponding to $1/\kappa \in [0, 1]$. As we show below, this leads to a family of isosceles triangles with the limiting cases $(90^\circ, 45^\circ, 45^\circ)$ for $1/\kappa = 0$ and $(60^\circ, 60^\circ, 60^\circ)$ for $\kappa = 1$. These limiting triangles correspond to two identical hard-core particles in a square well and a 2 + 1 Gaudin–Yang model on a ring [44], respectively. Both limits are Bethe-ansatz integrable, see references [23, 45] for a more detailed discussion. Note that certain extensions to the Bethe ansatz suggest that additional solvable cases exist [37, 38]. However, our numerical analysis does not find any traces of solvability beyond the two limiting cases, and supports the widely accepted idea that almost any one-dimensional problem with mass imbalance is non-integrable (notable exceptions are discussed in references [46–50]). Therefore, in this work we refer to systems with $1/\kappa = 0$ and 1 as integrable, in the sense that they can be analytically solved using the Bethe ansatz (cf reference [36]). Systems with other mass ratios are called non-integrable.

2.1. Hamiltonian

The Hamiltonian of a three-particle system with zero-range interactions reads as

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_2^2} - \frac{\hbar^2}{2\kappa m} \frac{\partial^2}{\partial y^2} + g \sum_i \delta(x_i - y). \quad (1)$$

Everywhere below we focus on the limit $g \rightarrow \infty$. In equation (1), $0 < x_i < L$ ($0 < y < L$) is the coordinate of the i th fermion (impurity), while L is the length of the ring, see figure 1(a). The eigenstates (ϕ) of H are periodic functions in each variable. They are antisymmetric with respect to the exchange of fermions, i.e., $\phi(x_1, x_2, y) = -\phi(x_2, x_1, y)$. Furthermore, the limit $g \rightarrow \infty$ demands that ϕ vanishes when the fermion approaches the impurity, i.e., $\phi(x_i \rightarrow y) \rightarrow 0$. For convenience, we use the system of units in which $\hbar = 1$ and $m = 1$ in the following. For our numerical analysis, we choose units such that $L = \pi$.

The Hamiltonian H can be written as a sum of the relative and center-of-mass parts. To show this, we expand ϕ using a basis of non-interacting states, i.e.,

$$\phi(x_1, x_2, y) = \sum_{n_1, n_2, n_3} a_{n_1, n_2}^{(n_3)} e^{-\frac{2\pi i}{L}(n_1 x_1 + n_2 x_2 + n_3 y)}, \quad (2)$$

where $a_{n_1, n_2}^{(n_3)} = -a_{n_2, n_1}^{(n_3)}$ to satisfy antisymmetric condition on the wave function. It is straightforward to see that the Hamiltonian does not couple states with different values of the ‘total momentum’, $P = 2\pi \frac{n_{\text{tot}}}{L}$; $n_{\text{tot}} = n_1 + n_2 + n_3$ because of translational invariance. For example, the operator $\delta(x_1 - y)$ couples two states via the matrix element:

$$\int dx_1 dx_2 dy \delta(x_1 - y) e^{-\frac{2\pi i}{L}(n_1 - n'_1)x_1 + (n_2 - n'_2)x_2 + (n_3 - n'_3)y}, \quad (3)$$

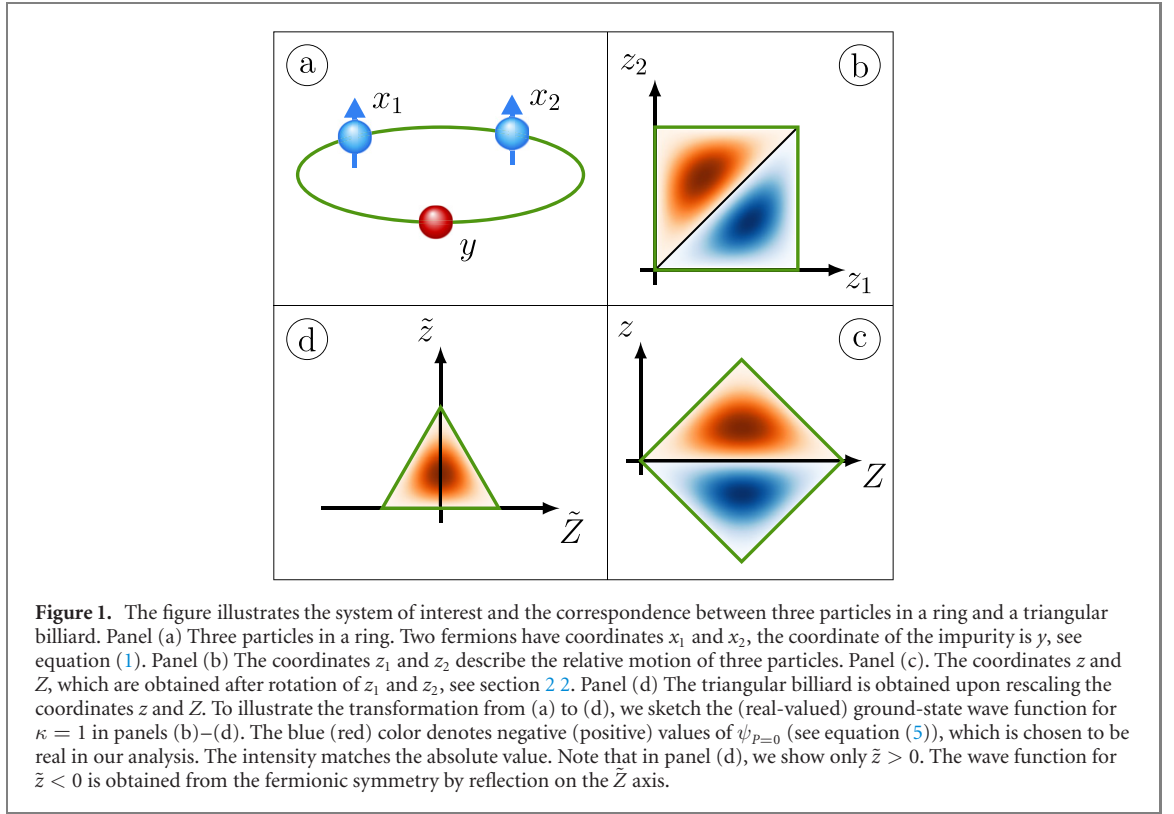
which equals $\delta_{n_2, n'_2} \delta_{n_1 + n_3, n'_1 + n'_3}$, and, hence, conserves P . The integral of motion, P , allows us to write the wave function as

$$\phi = e^{-iPy} \sum_{n_1, n_2} a_{n_1, n_2}^{(n_{\text{tot}} - n_1 - n_2)} e^{-\frac{2\pi i}{L}(n_1(x_1 - y) + n_2(x_2 - y))}, \quad (4)$$

and define the function, which depends only on the relative coordinates:

$$\psi_P(z_1, z_2) = e^{iPy} \phi(x_1, x_2, y), \quad (5)$$

where $z_i = L\theta(y - x_i) + x_i - y$, with the Heaviside step function: $\theta(x > 0) = 1, \theta(x < 0) = 0$. The coordinates z_i are chosen such that the function $\psi_P(z_1, z_2)$ takes values on $z_i \in [0, L]$, see figure 1(b).



The function ψ_p is an eigenstate of the Hamiltonian

$$H_p = -\frac{1}{2} \sum_{i=1}^2 \frac{\partial^2}{\partial z_i^2} - \frac{1}{2\kappa} \left(\sum_{i=1}^2 \frac{\partial}{\partial z_i} \right)^2 + i \frac{P}{\kappa} \sum_{i=1}^2 \frac{\partial}{\partial z_i}, \quad (6)$$

which will be the cornerstone of our analysis. As we show below, it is enough to consider only $H_{p=0}$ for our purposes. To diagonalize H_0 , we resort to exact diagonalization in a suitable basis. As a basis element, we use the real functions $\sin\left(\frac{n_1 \pi z_1}{L}\right) \sin\left(\frac{n_2 \pi z_2}{L}\right) - \sin\left(\frac{n_1 \pi z_2}{L}\right) \sin\left(\frac{n_2 \pi z_1}{L}\right)$, where n_1 and n_2 are integers with $n_{\max} > n_1 > n_2 > 0$, which is a standard choice for this type of problems, see, e.g., [51, 52]. Our choice of the basis ensures that $\psi_{p=0}$ is real for the ground and all excited states. The parameter n_{\max} defines the maximum element beyond which the basis is truncated. Note that the basis element is the eigenstate of the system for $1/\kappa = 0$. Therefore, we expect exact diagonalization to perform best for large values of κ and more poorly for $\kappa = 1$. To estimate the accuracy of our results, we benchmark against the exact solution for an equilateral triangle ($\kappa = 1$), see the discussion in the appendix. Using $n_{\max} = 130$, we calculate about 4000 states whose energies have relative accuracy of the order of 10^{-3} . This set of 4000 states is an input for our analysis in the next section.

To summarize this subsection: we perform the transformation from H, ϕ to H_p, ψ_p to eliminate the coordinate of the impurity from the consideration. Our procedure can be considered as the Lee–Low–Pines transformation [53] in coordinate space, which is a known tool for studying many-body systems with impurities in a ring [54–57]. Below we argue that H_p can be further mapped onto a triangular billiard. Note however that we are going to work with H_p everywhere. Its eigenfunctions are defined on a square (see figure 1(b)), allowing us to use them directly as an input for ConvNets.

2.2. Mapping onto a triangular billiard

It is known that three particles in a ring can be mapped onto a triangular billiard [39, 40]. Here we show this mapping starting with H_p . First of all we rotate the system of coordinates to eliminate the mixed derivative $\frac{\partial}{\partial z_1} \frac{\partial}{\partial z_2}$; see figure 1(c). To this end, we introduce the system of coordinates $z = (z_2 - z_1)/\sqrt{2}$ and $Z = (z_2 + z_1)/\sqrt{2}$, where the Hamiltonian reads as

$$H_p(z, Z) = -\frac{1}{2} \frac{\partial^2}{\partial z^2} - \frac{1}{2} \frac{\partial^2}{\partial Z^2} - \frac{1}{\kappa} \frac{\partial^2}{\partial Z^2} + i \frac{\sqrt{2}P}{\kappa} \frac{\partial}{\partial Z}. \quad (7)$$

The last term here can be eliminated by a gauge transformation $\psi_P \rightarrow \exp\left(\frac{i\sqrt{2}P}{\kappa+2}Z\right)\psi_P$. Therefore, in what follows we only consider $P = 0$ without loss of generality. We shall omit the subscript, i.e., we write ψ . Note that it is enough to study only $z \geq 0$, because of the symmetry of the problem.

To derive the standard Hamiltonian for quantum billiards:

$$h = -\frac{1}{2}\frac{\partial^2}{\partial \tilde{z}^2} - \frac{1}{2}\frac{\partial^2}{\partial \tilde{Z}^2}, \quad (8)$$

we rescale and shift the coordinates as $\tilde{z} = z$, and $\tilde{Z} = \sqrt{\kappa/(\kappa+2)}(Z - L/\sqrt{2})$, see figure 1(d). The Hamiltonian h is defined on an isosceles triangle with the base angle obtained from $\tan(\alpha) = \sqrt{(\kappa+2)/\kappa}$. For systems with more particles the corresponding transformations $H \rightarrow H_P \rightarrow h$ lead to quantum billiards in polytopes, allowing one to connect an N -body quantum mechanical problem to a quantum billiard in $N - 1$ dimensions. This can be a route for finding new applications of solvable models, see references [39, 46, 48].

Finally, we note that if the interaction term in equation (1) was an impenetrable wall with some radius R instead of the delta function, then the considerations above would also lead to a mapping of the system onto a triangle. (See reference [58] for an illustration with an equilateral triangle.)

3. Properties of the system

A discussion of highly excited states of triangular billiards can be found in the literature [22–26]. However, we find it necessary to review some known results and calculate some new quantities in order to explain our current understanding of the difference between integrable and non-integrable states. In principle, highly excited states of a quantum system can be simulated using microwave resonators (see, e.g., [59, 60]), or generated by means of Floquet engineering—by choosing the driving frequency to match the energy difference between the initial and the desired final state (see, e.g., reference [61]). Therefore the results of this section are not of purely theoretical interest, as they can be observed in a laboratory.

As we outlined in the introduction, there are two main approaches for analyzing a connection between highly-excited states and classical integrability. The first one relies on statistical properties of the energy spectra, while the second one focuses on the morphology of individual quantum states. This section sets the stage for our further study by discussing these approaches in more detail.

3.1. Energy

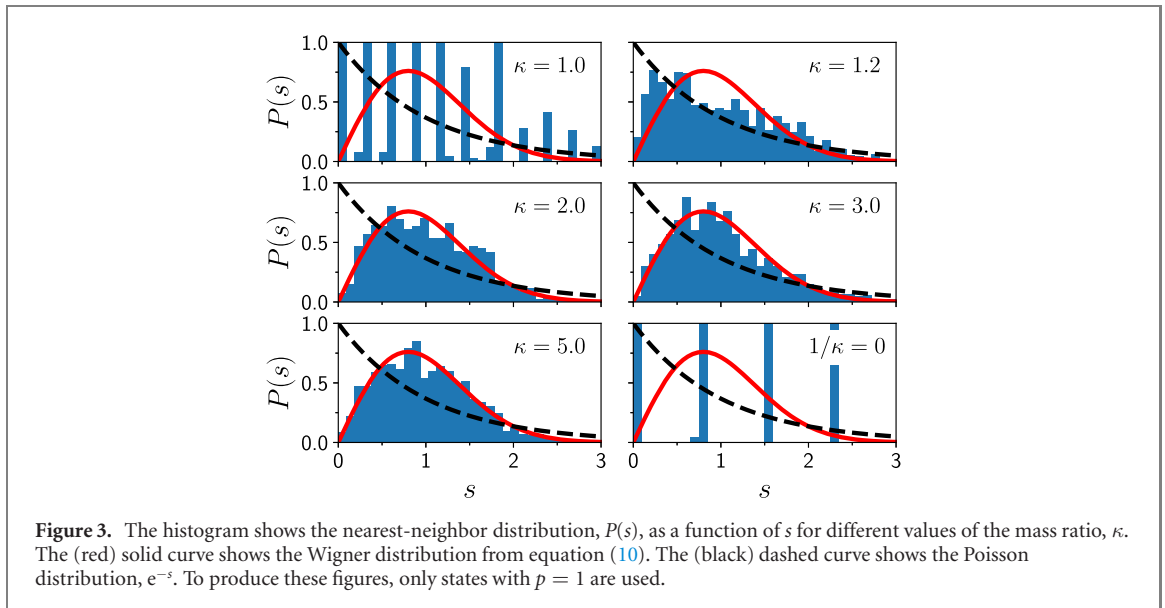
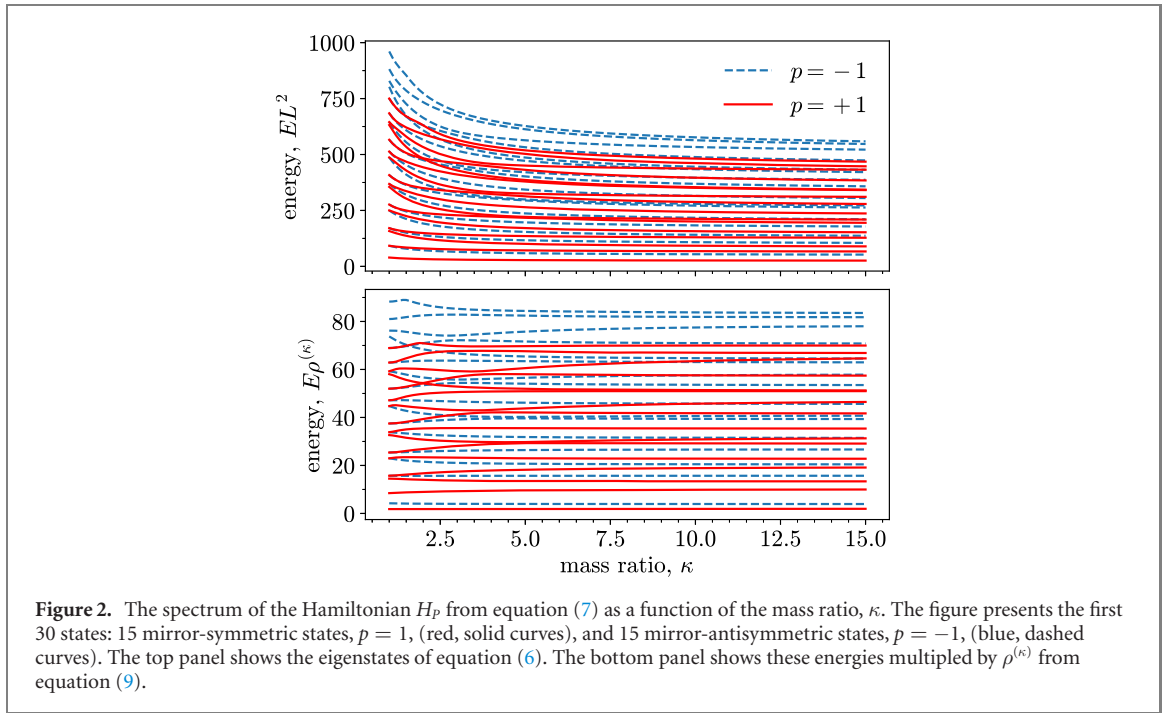
We start by calculating the energy spectrum. It provides a basic understanding of the evolution from an ‘integrable’ to a ‘chaotic’ system in our work as a function of κ . We present the first 30 states of H_0 in figure 2 (top). Note that an isosceles triangle has a symmetry axis ($\tilde{Z} \rightarrow -\tilde{Z}$), which corresponds to a mirror transformation (in the particle picture this symmetry corresponds to $z_i \rightarrow L - z_i$). The wave function can be symmetric or antisymmetric with respect to the mirror transformation and we consider these cases separately. The former states are denoted as having $p = 1$, and the latter have $p = -1$.

The energy spectrum features inflation of the spacing between levels, which can be understood as a repulsion of levels in the Pechukas gas [62, 63]. According to Weyl’s law, it can also be interpreted as a change of the density of states, $\rho^{(\kappa)}(E) = dN/dE$, where $N(E)$ is the number of states with the energy less than E . The function $\rho^{(\kappa)}(E)$ can be easily calculated using Weyl’s law [64] for the triangular billiard described by the Hamiltonian h :

$$\rho^{(\kappa)}(E \rightarrow \infty) \rightarrow \frac{L^2}{4\pi} \sqrt{\frac{\kappa}{\kappa+2}}. \quad (9)$$

The density of states is independent of the energy in this equation because we work with a two-dimensional object. Equation (9) is derived assuming large values of E , however, in practice, it also describes well the density of states in a lower part of spectrum (cf reference [24]). If we multiply the energies presented in figure 2 (top) by $\rho^{(\kappa)}$ then we obtain a spectrum without inflation, i.e., all levels are equally spaced on average, see figure 2 (bottom).

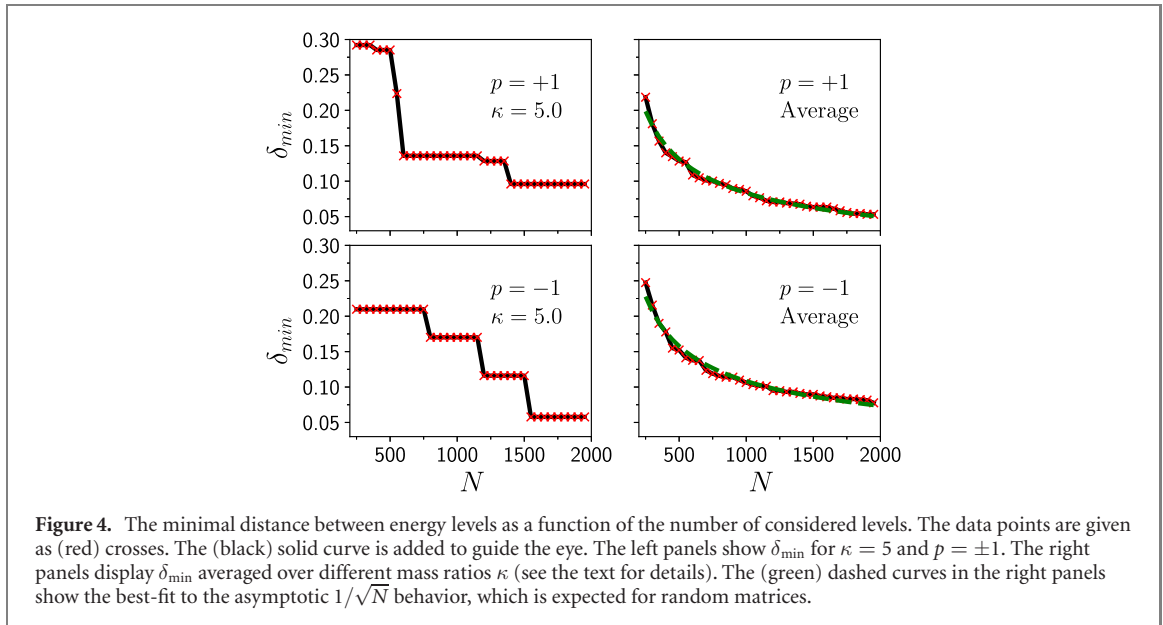
Multiplication of E by $\rho^{(\kappa)}$ is a simple example of unfolding, which allows us to directly compare features of the energy spectrum for different values of κ . The goal of the unfolding is to extract the ‘average’ properties of the levels distribution and, thus, diminish the effect of local level density fluctuations in the spectrum. While there are many possible ways to implement the unfolding procedure, which depend on the properties of the energy spectrum (for further information see, e.g., references [9, 65, 66]), the ultimate goal is to obtain rescaled levels with unit mean spacing. Below, we rescale all of the energy levels by the mean distance between them, thus, obtaining the unit mean spacing. We benchmarked results of this unfolding against more complicated approaches, and found qualitatively equivalent outcomes.



We use unfolded spectra to analyze the distribution of nearest neighbors, $P(s)$, which shows the probability that the distances between a random pair of two neighboring energy levels is s . The function $P(s)$ is presented in figure 3, see also [24, 25, 45], where some limiting cases are analyzed. For the sake of discussion, we only study the states with $p = 1$, however, we have checked that the case with $p = -1$ leads to qualitatively identical results. The size of bins in the histograms in figure 3 is virtually arbitrary [66]. For visual convenience, we have followed a rule of thumb that the number of bins should be taken at approximately a square root of the number of the considered levels.

Figure 3 shows a transition from regular to chaotic when the mass ratio changes from $\kappa = 1$ to larger values. The degeneracies in the energy spectrum for $\kappa = 1$ and $1/\kappa = 0$ lead to well spaced bins in the figure. This behavior is however rather unique, and it is immediately broken for other mass ratios. For example, already for $\kappa = 1.2$ the levels start to repel each other, and the distribution $P(s)$ can be approximated by the Wigner distribution [8]

$$P_{\text{GOE}}(s) = \frac{\pi s}{2} e^{-\frac{\pi s^2}{4}}. \quad (10)$$



Note that it is important to use only one value of p for this conclusion. Levels that correspond to different values of p do not repel each other, and the Wigner distribution cannot be realized [24].

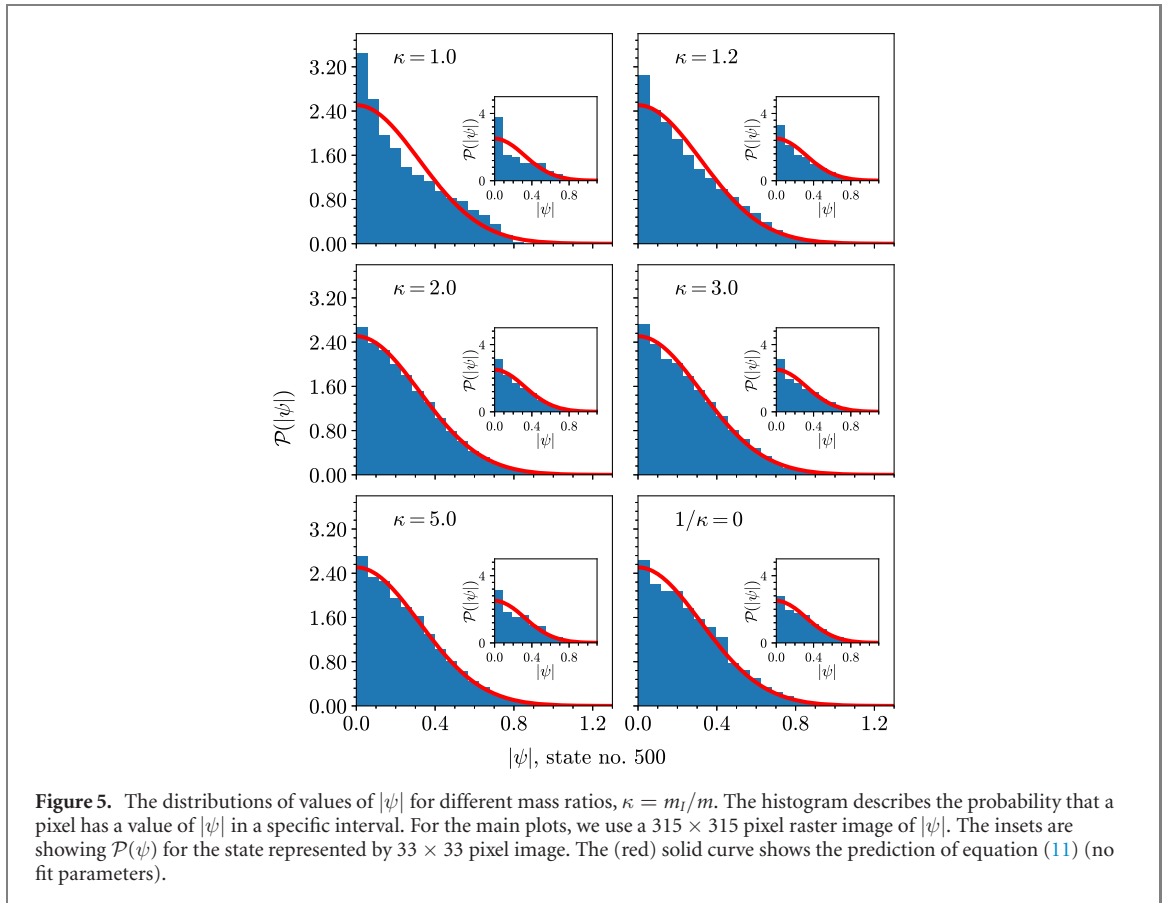
It is impossible to analyze every value of κ . However, we can also say something on average about our system. To that end, we calculate the dependence of the minimal distance between levels as a function of the number of considered levels, $\delta_{\min}(N) = \inf E_n - E_{n-1, n < N}$, here E_n is the energy of the n th state with a given value of p . For a random matrix, δ_{\min} is expected to scale as $1/\sqrt{N}$ [67, 68]. To the best of our knowledge, this result is not strictly proven, but at the intuitive level it can be understood from equation (10). The probability that the distance between two energy levels is smaller than δ_{\min} is given by $\int_0^{\delta_{\min}} P_{\text{GOE}} ds = 1 - e^{-\pi\delta_{\min}^2/4}$. In the limit $\delta_{\min} \rightarrow 0$, this expression can be approximated by $\pi\delta_{\min}^2/4$. If we consider N lowest states, then the probability that all nearest neighbors are separated by $\delta > \delta_{\min}$ is given by $(1 - \pi\delta_{\min}^2/4)^N$. To keep this probability independent of N , the parameter δ_{\min} must be proportional to $1/\sqrt{N}$.

We show δ_{\min} for our system for $\kappa = 5$ in figure 4 (left panels). We see that for a given value of κ it is impossible to verify $1/\sqrt{N}$ scaling, at least for the considered amount of eigenstates. However, the randomness present in a mass-imbalanced system can be recovered. To show this, we average δ_{\min} over different masses, i.e., $\delta_{\min}^{\text{average}} = \frac{1}{\mathcal{M}} \sum \delta_{\min}(\kappa_i)$, where \mathcal{M} determines how many values of κ appear in the sum. To produce figure 4 (right panels), we sum over the following values of the mass ratios: $\kappa = 1.1, 1.2, \dots, 5$. The parameter $\delta_{\min}^{\text{average}}$ has approximately $1/\sqrt{N}$ behavior at large values of N , which confirms our expectation that systems with $1/\kappa \in (0, 1)$ are not integrable.

3.2. Wave function

The analysis above shows a drastic change of properties of the system when moving from integrable to non-integrable regimes. Information about this transition is extracted by analyzing the energy levels as in figure 3, although, the correspondence principle conjectures properties at the level of individual wave functions. The wave function of a highly excited state contains too much information for the human eye, and one has to rely on a few existing conjectures that allow one to connect classical chaos to quantum states. For example, the chaotic states are expected to be similar to a random superposition of plane waves [6], since the underlying classical phase space has no structure, i.e., the classical motion is not associated with motion on a two-dimensional torus. This expectation applies to a typical random state (not to atypical, e.g., scared states [69]). In contrast, the wave functions of integrable states are expected to have some non-trivial morphology, since classical phase space of integrable systems has some structure. Below, we illustrate these ideas for our problem. We focus on a distribution of wave-function amplitudes, although, other signatures of ‘chaos’ in eigenstates connected to local currents and nodal lines⁶ [70–72] will also be important when we analyze our NN.

⁶ The dog-or-cat classifier in this case is a network with one output label for a dog and one for a cat, which has been trained using a set of a few thousand pictures.



A celebrated result of the random-wave conjecture is a Gaussian distribution of wave-function amplitudes, see examples in references [73–75]:

$$\mathcal{P}(\psi) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{|\psi|^2}{2v}}, \quad (11)$$

where the variance $v = 1/L$ fixes the normalization of the wave function⁷. We present our numerical calculations of $\mathcal{P}(\psi)$ in figure 5. For this figure, we discretize the wave function for the 500th state using either a 315×315 pixel grid or a 33×33 pixel grid, and assign to each unit of the grid a value that corresponds to ψ in the center of the unit. The distribution of these central values for a given value of κ is presented as a histogram in figure 5. For $\kappa = 1$ the $\mathcal{P}(\psi)$ resembles an exponential function (cf reference [75]). For larger values of κ , a Gaussian profile is developed. The distinction between the histogram and equation (11) is clear for $\kappa = 1$. For $1/\kappa = 0$ the difference is less evident. Note that the peak at $\psi = 0$ is enhanced in comparison to the prediction of equation (11) for all values of κ . This is due to the evanescence of the wave function at the boundaries, which is a finite-size effect beyond equation (11). Finally, the characteristics of the states are also visible in a low-resolution images, see the insets of figure 5. This feature will be used in the design of our NN discussed below.

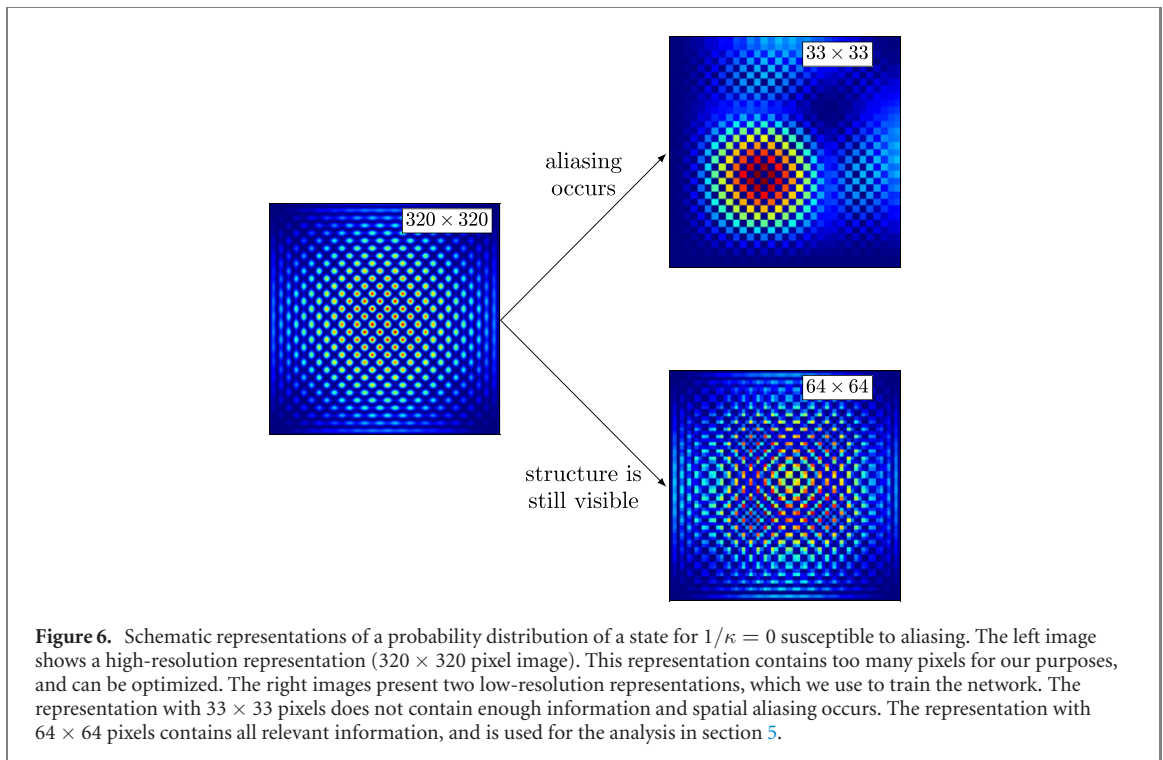
4. Neural network

To construct a NN that can distinguish integrable states from non-integrable ones, we need to

- Prepare a data set for training the network
- Choose a suitable architecture and a training algorithm

In this section, we discuss these two items in detail.

⁷ A nodal line is a set of points $\{X, Y\}$ that satisfy $\psi(X, Y) = 0$.



4.1. A data set

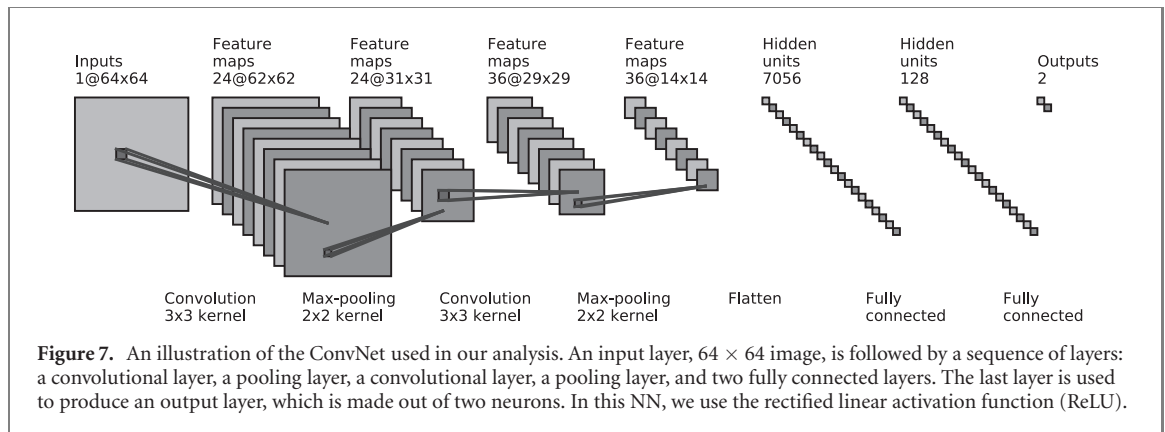
As data set we use the set \mathcal{A} made of two-dimensional images that represent highly excited states. We can use images of (real) wave functions, $\psi(z_1, z_2)$ or probability densities, $|\psi(z_1, z_2)|^2$. We have checked that these two representations lead to similar results. In the paper, we present only our findings for $|\psi(z_1, z_2)|^2$. To produce \mathcal{A} , we diagonalize the Hamiltonian $H_{p=0}$ of equation (6) for $\kappa = 1, 2, 5$ and $1/\kappa = 0$. Each image has a label—integrable (for $\kappa = 1$ and $1/\kappa = 0$) or non-integrable (for $\kappa = 2$ and 5)⁸. We do not include information about the mirror symmetry, i.e., states with different values of p are treated on the same footing, since we do not expect that this information is relevant for a coarse-grained (see below) image of $|\psi(z_1, z_2)|^2$. This allows us to work with twice as large datasets compared to figure (3). Each mass ratio contributes 1000 states to \mathcal{A} , which therefore contains 4000 images in total. It is reasonable to not use data sets that contain states with very different energies: very different energies lead to very different length scales, and hence different information content that should be learned. We choose to include all states from the 50th to 1050th excited states. Not much should be deducible about the low-lying states (with $N \sim 10$) from the correspondence principle, therefore, we do not use them in our study.

A wave function $\psi(z_1, z_2)$ is a continuous function of the variables z_1 and z_2 , see figure 1(b). To use it as an input for a network, we need to discretize and coarse-grain it. To this end, we represent ψ as a 64×64 pixel image, and as the value of the pixel we use the value of the wave function at the center of the pixel⁹. The resolution is important for this discretization. Low resolution might not be able to capture oscillations present in highly excited states, leading to a loss of important physical information. For example, the approximately N th state in the spectrum for $1/\kappa = 0$ will have about \sqrt{N} oscillations in each direction, and it is important therefore to use a $2\sqrt{N} \times 2\sqrt{N}$ representation of the wave function (similar to the Nyquist–Shannon sampling theorem)¹⁰. For a lower resolution, the oscillations are not faithfully reproduced in the low resolution image and spatial aliasing occurs. We illustrate this using the 33×33 resolution in figure 6 for an integrable state that is susceptible to spatial aliasing.

Note that out of curiosity, we have also used images with 33×33 pixel resolution to train our network. The network could reach relatively high accuracy (higher than 90%). However, not all integrable states were detected properly. For example, the one in figure 6 was classified as non-integrable by the network. In general, spatial aliasing is more damaging for integrable states, which have symmetries that should be respected; non-integrable states are more random, and some noise does not change the classification of the

⁸ The value of v is calculated by using the average value of ψ^2 , i.e., $\overline{\psi^2} = \int x^2 P(x) dx$, in the normalization condition, i.e., $\int \psi^2 dz_1 dz_2 = \overline{\psi^2} \int dz_1 dz_2 = 1$, which leads to the condition on v as $\int x^2 P(x) dx = 1 / \int dz_1 dz_2$.

⁹ To avoid any bias toward non-integrable states, we use non-integrable states for only two values of κ . However, we have checked that our conclusions hold also true if we include other values of κ into the data set, in particular, if we add 1000 states to \mathcal{A} from a system with $\kappa = 15$.



network. Everywhere below we use the 64×64 pixel representation, which gives a sufficiently accurate representation of the state, so that we do not need to worry that the network learns unphysical properties. Note that certain local features (e.g., $\psi(z_1 = z_2) = 0$) of the wave function may disappear at this resolution. The overall high accuracy of our network suggests that such features are not important for our analysis.

The set \mathcal{A} seems somewhat small. For example, the well-known Asirra dataset [76] for the cat–dog classification contains 25 000 images that are commonly used to test and compare different image recognition techniques. However, we will see that \mathcal{A} is large enough to train a network that can accurately classify integrable and non-integrable states. The dataset \mathcal{A} is further divided into two parts. We draw randomly 85% of all states and use them as a training set. The remaining 15% is used for testing. We fix the random seed used for drawing to avoid discrepancies between different realizations of the network. It is worthwhile noting that in image-recognition applications, the dataset \mathcal{A} may be divided into more than two parts. For example, in addition to the training set and testing set, one can introduce a validation (or development) set [77], which is used to fine-tune parameters of the model. We do not use this additional set here. The focus of this work is on understanding features of our general image classifier, and not on improving its accuracy.

4.2. Architecture

The NN in our problem is a map that acts in the space \mathcal{X} made of all 64×64 pixel representations of $|\psi|^2$. By analogy to the standard dog-vs-cat classifier, the output of the network is a vector with two elements \mathbf{b} . Note that n output neurons are usual for classifying n classes. However, it is possible to use a single output neuron for a binary classification, since we know that $b_1 + b_2$ must be equal to one. We use two neurons, since such an architecture can be straightforwardly extended to more output neurons, which may be useful for the future studies, as we discuss in section 6.

The first element of the output layer, $0 \leq b_1 \leq 1$, determines the probability that the input state is integrable, whereas the second element $b_2 = 1 - b_1$ is the probability that the input state is non-integrable. An input state is classified as integrable (non-integrable) if $b_1 > b_2$ ($b_1 < b_2$).

Mathematically, the network is a map \mathbf{f} , which acts on the element a of \mathcal{X} as

$$\mathbf{f}(a; \theta, \theta_{\text{hyp}}) = \mathbf{b}. \quad (12)$$

The \mathbf{f} is determined by the set of parameters θ , which are optimized by training. Since our problem is similar to image recognition (in particular dog-vs-cat classification) [10], which is one of the standard applications in machine learning, we can use the already known training routines (SGD, ADAM, Adadelta, ...) for optimizing θ . The outcome for the parameters θ may vary between different trainings, and we use this variability to check the universality of our results. Specifications of \mathbf{f} that are not trained but specified by the user are called hyperparameters (θ_{hyp}). Examples of them include the loss function, optimization algorithm, learning rate, network architecture, size of batches that the data is split into for training (batch size) and the length of training (epochs). We find hyperparameters by trial-and-error.

The simplest form of a network is called a dense network in which all input neurons are connected to all output neurons. However in most cases of image detection, this architecture does not lead to accurate results. This also happens in our case. Instead, we resort to a standard architecture based on ConvNets for image recognition, see figure 7¹⁰. Our network consists of two convolutional layers and two max-pooling

¹⁰ The color depth (i.e., how many colors are available) of a pixel is effectively given by the numerical precision used to produce the input data. If experimental data is used as an input, then their accuracy will determine the color depth of a pixel.

layers. The former use a set of filters and apply them in parts to the image to produce a new smaller image. This is somewhat analogous to a renormalization group transformation [78]. A set of images that are produced by a convolutional layer is called feature map. Each convolutional layer is followed by a max-pooling layer which reduces the size of an image. The size of max-pooling layers is a hyperparameter. In our implementation, max-pooling layers take the largest pixel out of groups of 2 by 2.

One could use architectures different from the one presented in figure 7. However, we checked that they do not lead to noticeably different results. Therefore, we do not investigate this possibility further.

5. Numerical experiments

Following the discussion in section 4, we train and test the NN. We observe that a typical accuracy of the trained network (which we refer later to as \mathcal{N}) is $\sim 97\%$ ¹¹. This means that about 18 states out of 600 used for testing are given the wrong label. Out of these 18 states, roughly one half is integrable. We illustrate typical wave functions that are classified correctly and wrongly in figure 8. It does not mean that these states are in anyway special—another implementation (e.g., another random seed for weights) will lead to other states that are given the wrong label. Non-integrable states with some structure (e.g., states with scars) in general confuse the network and might be classified as integrable.

In general, it is hard to interpret predictions of the NN. This becomes clear after noticing that some images can be changed so that a human eye can hardly detect any variation. At the same time, this change completely modifies the prediction of the network. Such a change can be accomplished especially easily for integrable states¹², thus, DL confirms our intuition that integrable states are a small subset in the space of all possible states. However, such a situation can also occur for non-integrable (in particular scarred) states. We illustrate this in figure 9, which is obtained by slightly modifying states from \mathcal{A} using tools of adversarial machine learning, see reference [79].

One simple way to extract features of the network is to look at feature maps, which should contain information about what features are important. For example, the first layer might represent edges at particular parts of the image, the second might detect specific arrangements of edges, etc. However, we could not extract any meaningful information from this analysis. This is expected: the features of integrable and non-integrable states are more abstract and not as intuitive as the features of cats and dogs or images of other objects we encounter in everyday life.

Other approaches to analyze a network rely on estimating the effect of removing a single (or a group) of elements on a model. For large data sets, this can be done by introducing influence functions [80, 81]. Here, we work with a small data set, and, therefore, we can calculate directly the actual effect of leaving states out of training on a given prediction. Our goal is to understand correlations between states of different energies. In our implementation we compare the prediction of \mathbf{f} from equation (12) for $|\psi|^2$ to a prediction of $\mathbf{f}_{-\beta}$ for the same state. Here $\mathbf{f}_{-\beta}$ is obtained by training a NN after leaving out the set β from \mathcal{A} . The comparison of the two predictions ($\mathbf{f} - \mathbf{f}_{-\beta}$) allows us to estimate the importance of the set β for the classification of a test state ψ ¹³. We present a typical example in figure 10, where one observes no clear energy correlations, which suggests that the network learns different energies simultaneously, at least in the energy interval we choose to work with. This observation is consistent with our discussion below that the network does not learn specific features of non-integrable states, and only overall randomness, which does not depend on a specific energy window. Finally, we note that our result in figure 10 is an example of cross-validation. It suggests weak dependence of the output of the network on a particular state, which is a necessary condition for a good performance of our NN.

Below, we explore the network \mathcal{N} further. To this end, we resort to numerical experiments. We employ \mathcal{N} to analyze states outside of the set \mathcal{A} . First, we study physical states, and then non-physical ones.

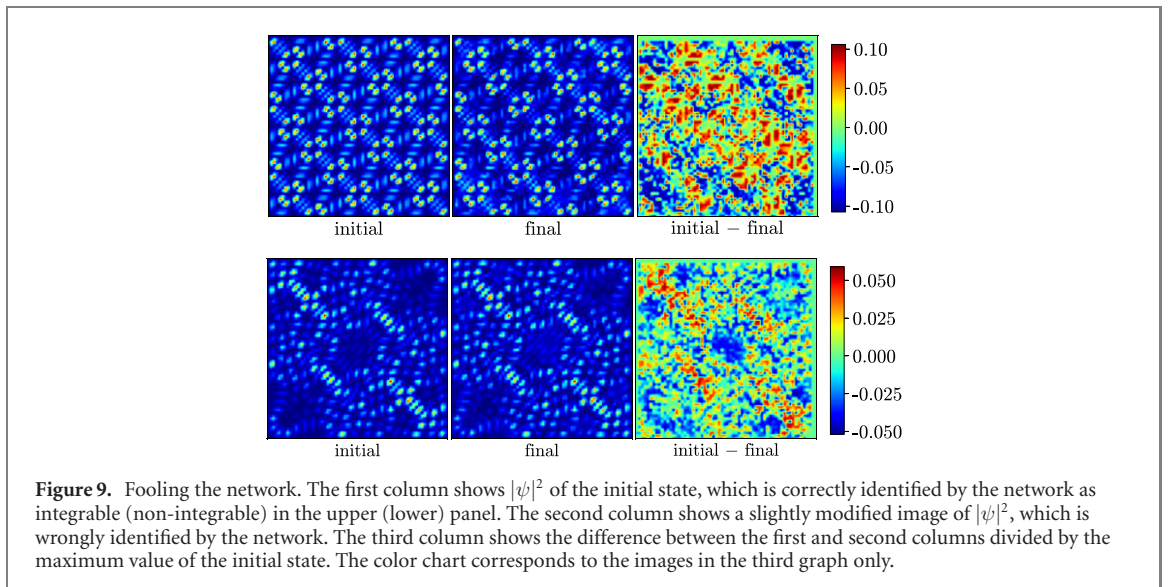
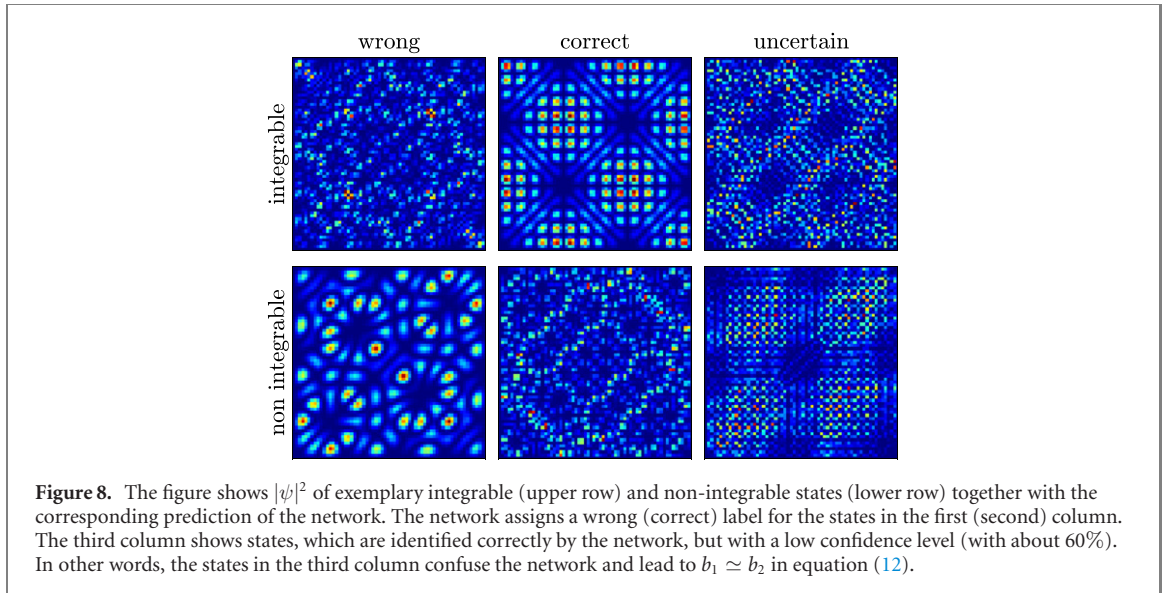
5.1. Classification of physical states outside of \mathcal{A}

As a first application of \mathcal{N} , we use it to classify eigen-states of $H_{P=0}$ not used in the training, i.e., for $\kappa \neq 1, 2, 5$ and $1/\kappa \neq 0$. These states are non-integrable (cf figure 4), and we observe that \mathcal{N} accurately classifies them as such as long as κ is far enough from $\kappa = 1$, see figure 11. The figure shows that

¹¹ We use the word ‘typical’ to emphasize that a trained network depends on hyperparameters and random seeds. Even for a given set of hyperparameters, each set of random parameters leads to a slightly different network \mathcal{N} . We can tune hyperparameters to reach higher accuracies. We do not discuss this possibility here, since high accuracy is not the main purpose of our study.

¹² Tools of adversarial machine learning can make a NN classify an arbitrary integrable state as ‘non-integrable’ using a small number (~ 100) of iterations. This is not possible for an arbitrary non-integrable state.

¹³ Note that it is important to choose a test state ψ for which the network gives an accurate prediction with high confidence level, i.e., $b_i \rightarrow 1$. For other states, an intrinsic randomness of ConvNets can lead to a drastic change in the classification of the network.



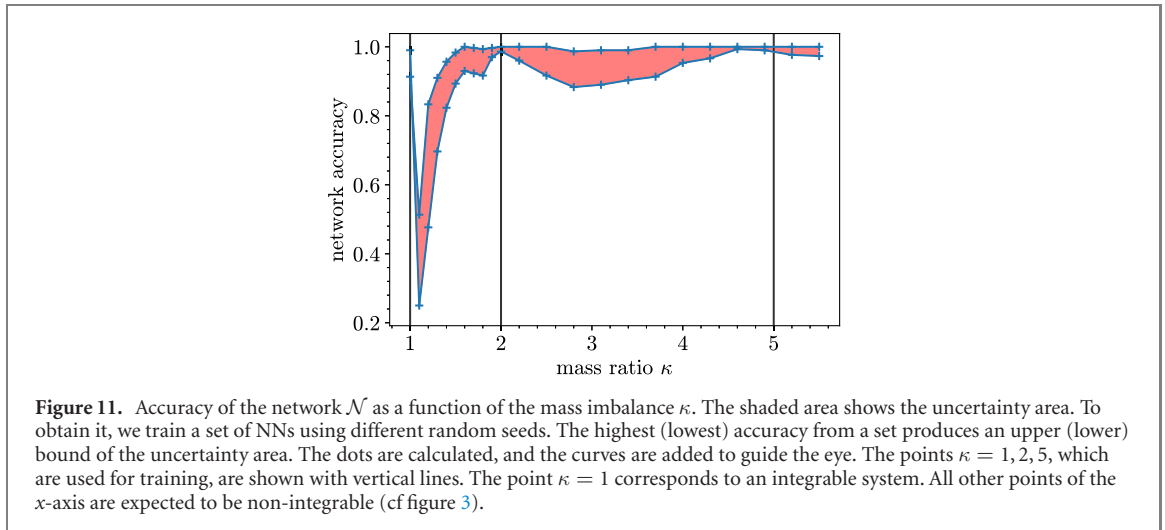
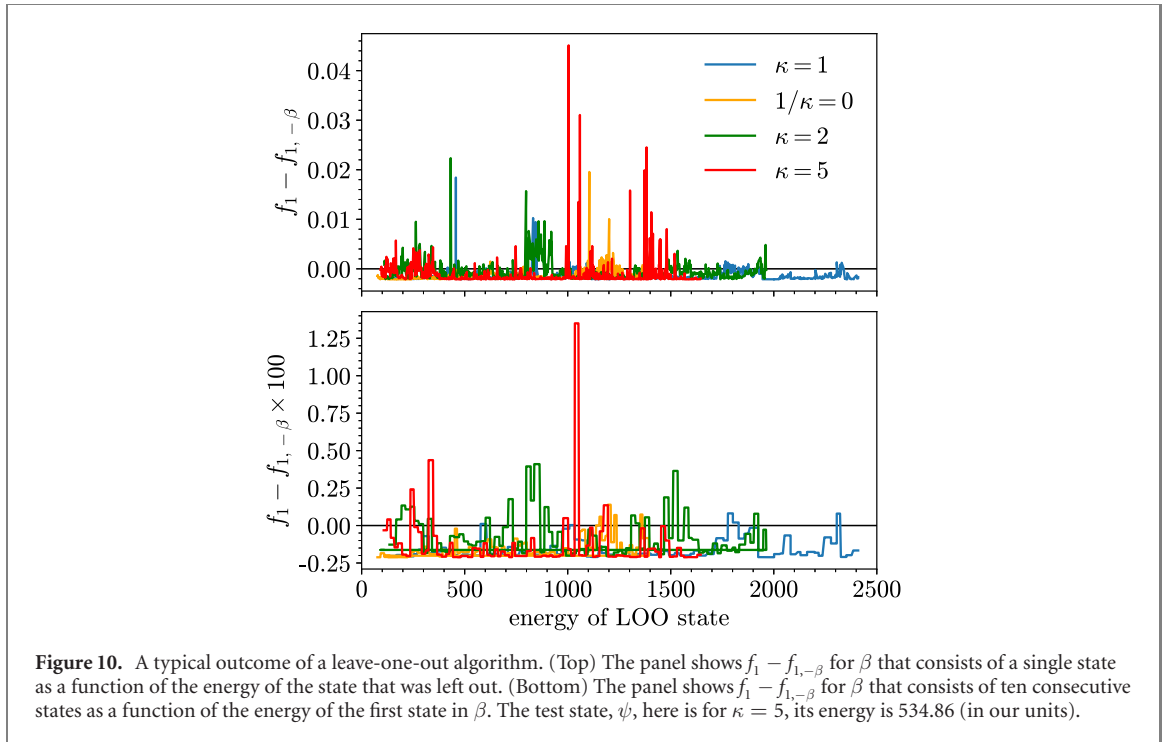
predictions of \mathcal{N} are not accurate only for systems with $\kappa = 1 + \epsilon$, where ϵ is a small parameter. These systems are non-integrable, however, the morphology of their eigenstates is very similar to the integrable ones at $\kappa = 1$. The network classifies them wrongly because of this. Already for $\kappa \simeq 1.5$, the accuracy of the network is close to one, and it stays high for larger values of κ . The region between 1 and 1.5 can be interpreted as a transition of the network classification from integrable to chaotic [13]. We do not expect this region to be universal—it should depend on hyperparameters, and the states used for training of \mathcal{N} . Therefore, we do not investigate it further.

To test the network on integrable states, we use wave functions of two non-interacting bosons in a box potential of size L :

$$\Psi_B = \frac{N_{k_1, k_2}}{L} [\sin(k_1 z_1) \sin(k_2 z_2) + \sin(k_2 z_1) \sin(k_1 z_2)], \quad (13)$$

where $k_1 \leq k_2$, and N_{k_1, k_2} is a normalization constant, $N_{k_1 = k_2} = 1$ and $N_{k_1 < k_2} = \sqrt{2}$. The order of the states is determined by the energy $(k_1^2 + k_2^2)/2$. The set of functions $\{\Psi_B\}$ is complementary to the $1/\kappa = 0$ case studied above for fermions. The bosonic symmetry yields states that are orthogonal to the training set, and therefore, the training routine can have no microscopic information about the wave function, Ψ_B . We use 1000 states of the bosonic type (from the 50th to 1050th) as an input for \mathcal{N} . We observe that \mathcal{N} accurately (accuracy is $\simeq 96\%$) classifies states Ψ_B as integrable.

To connect the analysis of Ψ_B to studies of quantum billiards, we note that two bosonic impurities in an infinite square well can be mapped on a right triangle with two impenetrable boundaries. At the third boundary a zero Neumann boundary condition should be satisfied— $\Psi'_B|_{z_1 = z_2} = 0$. The mapping follows



from the mapping discussed for fermions (see figure 1) assuming that the impurity is infinitely heavy. In particular, figure 1(b) shows the geometry of the problem in this case. Note that the bosonic symmetry requires that the derivative of the wave function vanishes on the diagonal of the square in figure 1(b). The high accuracy of the classification of the bosonic states suggests that a network trained using the Dirichlet boundary condition can also be used to classify states with the Neumann boundary condition. In other words, the network is mainly concerned with the ‘bulk’ properties of the wave function, the boundary is not important.

5.2. Classification of non-physical states

The network \mathcal{N} can classify any 64×64 pixel input image, and it is interesting to explore the outcome of the network for images that have no direct physical meaning. We start by considering non-normalized eigenstates of $H_{P=0}$. The normalization coefficient does not change the physics behind the states. However, since the function \mathbf{f} is non-linear, i.e., $\mathbf{f}(\alpha x) \neq \alpha \mathbf{f}(x)$, input states must have the same normalization as the states in the training set for a meaningful interpretation of the network. To illustrate this statement, we use states from \mathcal{A} multiplied by a factor, i.e., we use $\alpha|\psi|^2$ instead of $|\psi|^2$. Figure 12 shows the accuracy¹⁴ of the

¹⁴ Here we still can talk about the accuracy, since the states $\alpha\psi^2$ correspond to integrable or non-integrable situations.

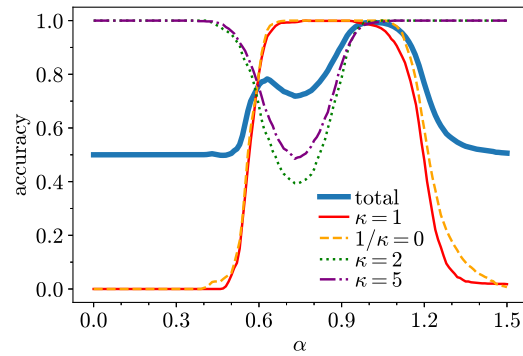


Figure 12. Prediction of the network for the states from \mathcal{A} multiplied by a factor α . The curves show the accuracies for four values of the mass ratio κ . The average of these four curves is shown as a thick solid curve.

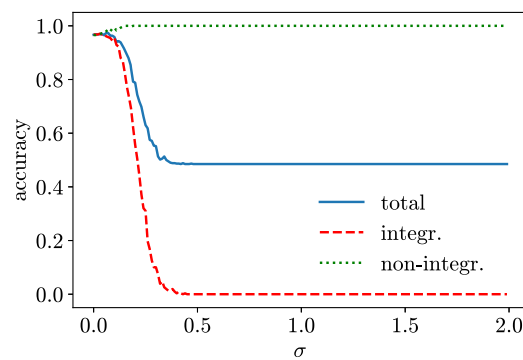
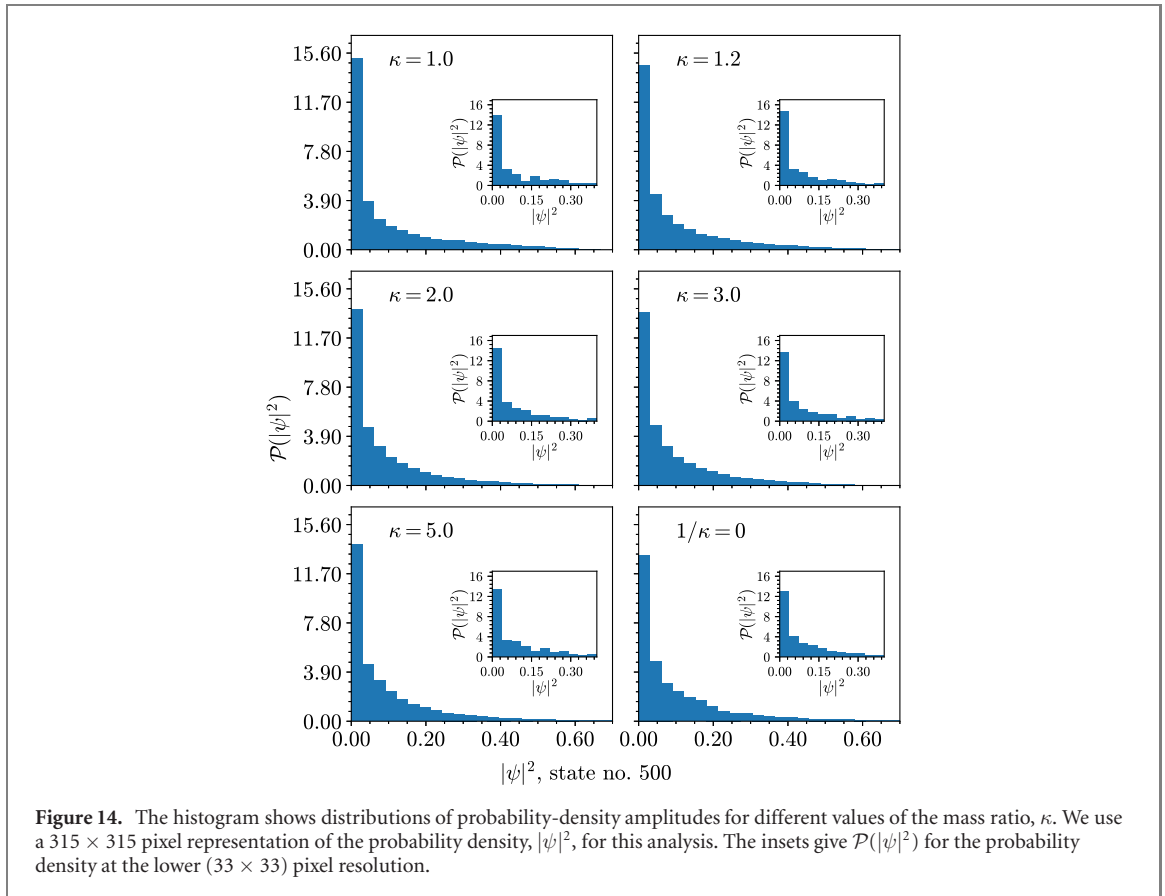


Figure 13. Predictions of the network for the states from \mathcal{A} with noise. The (red) dashed curve shows the accuracy for states which are integrable for $\sigma = 0$ (the accuracy here is defined as the percentage of the states identified as integrable). The (green) dotted curve shows the accuracy of the network for non-integrable states. The (blue) solid curve shows the average of the first two curves.

network as a function of α . The maximum accuracy of the network is reached at $\alpha = 1$, i.e., for the states used for the training. Integrable states are classified as non-integrable almost everywhere except close to $\alpha = 1$. A different situation happens for non-integrable states. They are classified correctly almost everywhere, and we conclude that they are less susceptible to the factor α . The shape of the curves in figure 12 is not universal, it depends on hyperparameters of the network. However, a general conclusion holds—the normalization is important, and we use normalized input functions in the further analysis. In principle, it is possible to reduce the sensitivity of our NN to α . To that end, one needs to add states $\alpha|\psi|^2$ to a training set (see data augmentation in DL [82]). We do not pursue this possibility here, to demonstrate that the sensitivity of integrable states on α is different from that for non-integrable states.

As a next step, we add noise to the images from \mathcal{A} , i.e., we build a new data set using wave functions $\tilde{\psi} = a_\sigma \psi(1 + r_\sigma)$, where r_σ is a noise function whose values are drawn from the normal distribution with zero mean and the standard deviation σ ; a_σ is a normalization factor, which is determined for each input state depending on the function r_σ . We assume that r_σ possesses basic symmetries of the problem: fermionic and mirror. Functions $\tilde{\psi}$ naturally appear in applications, and therefore, it is interesting to investigate the resilience of the network to random noise. We use 4000 states of \mathcal{A} with noise to make a relevant statistical statement, see figure 13. Small values of σ lead to weak noise and the network correctly classifies almost all input states. However, larger values of σ lead to confusing input states, and the network fails. It actually fails for integrable states where the noise destroys correlations. The accuracy for non-integrable states is always high. The resilience of the network to noise suggests it as a tool to analyze experimental data (e.g., obtained using microwave billiards). These experiments [21] can produce a large amount of data, however, there are limited variety of tools to analyze the simulated states. In particular, NNs can be used to identify atypical states, which do not fit the overall pattern, e.g., scars.

Our choice of $\tilde{\psi}$ to represent a noisy state is not unique. One could, for example, use instead of $\tilde{\psi}$ the function $\tilde{\psi} = a_\sigma(\psi + \mathcal{G}r_\sigma)$, where the parameter \mathcal{G} determines the relative weight of the function r_σ , which is defined as above. Note that \mathcal{G} cannot be absorbed in r_σ , since for a given value of σ , the average



amplitude of r_σ^2 is well defined. For $\mathcal{G} = 0$, the function $\bar{\psi}$ is a physical state, whereas for $\mathcal{G} \rightarrow \infty$, the function $\bar{\psi}$ is completely random. In contrast to $\bar{\psi}$, the function ψ can become completely independent of the function ψ . This happens if the parameter \mathcal{G} is large.

For the data set based upon $\bar{\psi}$ with small values of \mathcal{G} , the accuracy of the network is similar to that presented in figure 13. For large values of \mathcal{G} , the network is confused and classifies states in a random manner. This behavior should be compared with the data set $\tilde{\psi}$ for which the states are classified as non-integrable when the noise is large. To understand this difference, note that $\tilde{\psi}$ retains information about the nodal lines of the physical state. It turns out that it is important for the input state to have enough pixels with small (zero) values (note that the number of zero pixels is very large for $|\psi|^2$, see figure 14). Only such states have a direct meaning for the network, all other states confuse the network and do not allow for extraction of any meaningful information. It is worthwhile noting that the network does not learn the physical nodal lines. We checked that almost any random state with a large number ($\simeq 30\%$ – 40%) of zero pixels is classified as non-integrable.

Our discussion above suggests that the network classifies all states with nodal lines but without clear spatial correlations as non-integrable. In other words, the network perceives almost all states as non-integrable, and there are only a few islands recognized as integrable states. Such asymmetry of learning is not expected to happen for labels with equal significance (cat and dog for example). We speculate that the observed asymmetry is related to the fact that non-integrable states that correspond to different energies do not correlate in space [83], and, therefore, a NN cannot learn any significant patterns when trained with those states.

All in all, the state classifier validates our expectation that non-integrable states are random and abundant, whereas integrable states are rather unique and occur rarely. For example, in our analysis, integrable states are a set of measure zero in κ , and perturbation in κ or noise change the prediction of \mathcal{N} for these states. The asymmetry also suggests that the standard implementation of DL presented here should be modified to reveal the physics behind non-integrable states. The present network does not distinguish a non-physical random state with a large number of zero values from a physical non-integrable state since it was not trained for that question. A possible modification is the addition of an extra label (b_3 in equation (12)) for non-physical states. Since there are many possible non-physical states, one should frame the problem having an experimental or numerical set-up in mind, where such states have some origin and interpretation. We leave an investigation of this possibility for future studies.

Finally, we note that the conclusion that the network classifies almost all random images as non-integrable is general—it does not depend on the values of hyperparameters, initial seed or the distribution (Laplace, Gaussian, etc) that we use for generation of the random states. We check this by performing a number of numerical experiments. In particular, this means that the network does not learn $\mathcal{P}(\psi)$ from equation (11).

To summarize: a trained network can accurately classify integrable and non-integrable states. The network can even classify input states from an orthogonal (bosonic) Hilbert space, on which it can have no microscopic information. Thus network identifies generic features of the ‘integrable’ and ‘non-integrable’ states, although this might be not that surprising provided that we work with low-resolution 64×64 pixel images. The network classifies almost all random images with nodal lines as non-integrable, which suggests that useful information for the network is mostly contained in integrable states.

6. Summary and outlook

We used convolutional NNs to analyze states of a quantum triangular billiard illustrated in figure 1. We argued that NNs can correctly classify integrable and non-integrable states. The important features of the states for the network are the normalization of the wave functions and a large number of zero elements, corresponding to the existence of a nodal line. Almost any random image that satisfies these criteria is classified as non-integrable. All in all, the NN supports our expectation that non-integrable states are resilient to noise as discussed in subsection 5.2, and have a ‘random’ structure, unlike integrable states whose structure can be revealed by considering, for example, nodal lines.

Our results suggest that machine learning tools can be used to analyze the morphology of wave functions of highly excited states obtained numerically or experimentally, to solve problems like: find exceptional states (e.g., scars or integrable states) in the spectra, investigate the transition from chaotic to integrable dynamics, etc. However, further investigations are needed to set the limits of applicability for DL tools. For example, our network considers all states without clear correlations as non-integrable. This means that it must be modified for the analysis of noisy data, where a noisy image without any physical meaning could be classified as non-integrable. To circumvent this, one could introduce additional labels for training the network. For example, one could consider three classes—‘integrable’, ‘non-integrable’, and ‘noise’. This classification might allow for a more precise description of data sets, and may help one to extract more information about the physics behind the problem.

We speculate that a network memorizes integrable states, and all other states are classified as non-integrable, provided that an image has a large number of vanishing values. This would explain our findings and it would align nicely with the observation that we observe no overfitting even after many epochs of working through the training set. In the future, it will be interesting to use other integrable systems to test this idea. In particular, one could use non-triangular billiards, or systems without an impenetrable boundary. For example, one can consider a two-dimensional harmonic oscillator with cold atoms. At a single body level, the integrability in this system can be broken by potential bumps [84, 85] or spin–orbit coupling [86].

In the present work, we focus on data related to the spatial representation of quantum states. However, our approach can also be used to analyze other data. For example, for few-body cold atom systems, correlation functions in momentum space are of interest in theory and experiment (see, e.g., [87]). Therefore, in the future, it will be interesting to train NNs using experimental/numerical data that correspond to a momentum-space representation of quantum states, and study the corresponding features of ‘quantum chaos’. Although, in the present work we focused on two-dimensional data (images of probability densities), it might be interesting to consider higher-dimensional analogues (e.g., four-body systems) as well. Further applications include the identification of phase transitions, the representation of quantum many-body states, and the solution of many-body problems [88]. Thus machine learning techniques may not only be used to classify quantum states but also to obtain them by solving the corresponding many-body problem.

To extract further information about the map \mathbf{f} , one could investigate its geometry close to its maximum (minimum) values. For example, in the vicinity of some accurately determined integrable state ($f_1(x_0) = 1$), we can write

$$f_1(x_0 + \delta x) \simeq 1 + \frac{1}{2} \delta x^T G \delta x, \quad (14)$$

where the position of the maximum x_0 is understood as a vector and G is the Hessian matrix. The first derivative of f_1 vanishes since the function f_1 is analytic and bounded. Eigenstates of the Hessian G provide us with the most important correlations. A preliminary study shows that there are only a handful of eigenvalues of G for our network, which suggests the next step in the analysis of our image classifier.

A deeper understanding of the way a NN works may be obtained by combining different techniques to interpret its operation. The potential of this approach was illustrated in reference [15], for instance, by analyzing how a network looking at images of a labrador retriever detects floppy ears and how that influences its classification.

Acknowledgments

We thank Aidan Tracy for his input during the initial stages of this project. We thank Nathan Harshman, Achim Richter, Wojciech Rzadkowski, and Dane Hudson Smith for helpful discussions and comments on the manuscript. This work has been supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 754411 (AGV); by the German Aeronautics and Space Administration (DLR) through Grant No. 50 WM 1957 (OVM); by the Deutsche Forschungsgemeinschaft through Project VO 2437/1-1 (Project No. 413495248) (AGV and HWH); by the Deutsche Forschungsgemeinschaft through Collaborative Research Center SFB 1245 (Project No. 279384907) and by the Bundesministerium für Bildung und Forschung under Contract 05P18RDFN1 (HWH). HWH also thanks the ECT* for hospitality during the workshop 'Universal physics in Many-Body Quantum Systems—From Atoms to Quarks'. This infrastructure is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 824093. We acknowledge support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of Technische Universität Darmstadt.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix

To generate the input for a NN, we diagonalize the Hamiltonian $H_{p=0}$ from equation (6) in a truncated Hilbert space whose basis element is

$$\xi_{n_1, n_2}(x_1, x_2) = N \left[\sin\left(\frac{n_1 \pi x_1}{L}\right) \sin\left(\frac{n_2 \pi x_2}{L}\right) - \sin\left(\frac{n_1 \pi x_2}{L}\right) \sin\left(\frac{n_2 \pi x_1}{L}\right) \right], \quad (15)$$

where $c > n_1 > n_2 > 0$, c is the cutoff parameter, and N is the normalization constant. In this basis, matrix elements of the Hamiltonian read as (in our numerical analysis, we use units in which $L = \pi$)

$$\begin{aligned} \int_0^\pi \int_0^\pi \xi_{m_1, m_2}(x_1, x_2)^* H_0 \xi_{n_1, n_2}(x_1, x_2) dx_1 dx_2 &= \frac{1}{2} (n_1^2 + n_2^2) \left(1 + \frac{1}{\kappa} \right) (\delta_{m_1, n_1} \delta_{m_2, n_2} - \delta_{m_1, n_2} \delta_{m_2, n_1}) \\ &+ \frac{n_1 n_2}{\pi^2 \kappa} [I(m_1 + n_2, m_2 + n_1) + I(m_1 + n_2, m_2 - n_1) \\ &+ I(m_1 - n_2, m_2 + n_1) + I(m_1 - n_2, m_2 - n_1) \\ &- I(m_1 + n_1, m_2 - n_2) - I(m_1 + n_1, m_2 - n_2) \\ &- I(m_1 - n_1, m_2 + n_2) - I(m_1 - n_1, m_2 - n_2)], \quad (16) \end{aligned}$$

where $I(s, t) = \frac{1}{st} [(-1)^s - 1] [(-1)^t - 1]$ if $s, t \neq 0$, and $I(s, t) = 0$ otherwise. To write these matrix elements in a matrix form, we use the index

$$n = n_2 - n_1 + c(n_1 - 1) - \frac{(n_1 - 1)n_1}{2}. \quad (17)$$

The parameters n_1 and n_2 can be uniquely defined as:

$$n_1 = \frac{1 + 2c}{2} - \sqrt{c^2 - c - 2n + \frac{9}{4}}, \quad n_2 = n + n_1 - c(n_1 - 1) + \frac{(n_1 - 1)n_1}{2}. \quad (18)$$

To choose the cutoff parameter, c , we should find a good balance between the calculation time and the accuracy of our results. To quantify the accuracy, we compute energies for $\kappa = 1$ by diagonalizing H_0 , and

compare them to the exact ones obtained with the Bethe ansatz:

$$E_{\text{BA}} = \begin{cases} n_1^2 + n_2^2 + n_3^2, & \sum n_i = 0, \\ \left(n_1 + \frac{1}{3}\right)^2 + \left(n_2 + \frac{1}{3}\right)^2 + \left(n_3 + \frac{1}{3}\right)^2, & \sum n_i = -1, \\ \left(n_1 + \frac{2}{3}\right)^2 + \left(n_2 + \frac{2}{3}\right)^2 + \left(n_3 + \frac{2}{3}\right)^2, & \sum n_i = -2. \end{cases} \quad (19)$$

This solution assumes that the total momentum is zero. The relative difference


$$\epsilon = \frac{E_{\text{BA}} - E(c)}{E_{\text{BA}}} \quad (20)$$

provides us a measure for the accuracy. Note that our exact diagonalization method is expected to work better for $\kappa > 1$, since ξ is the eigenstate of a system with $1/\kappa = 0$. The input for a NN is obtained using $c = 130$, for which we obtain 726 (3795) states with $\epsilon < 10^{-4}$ (10^{-3}) within a short enough computation time.

ORCID iDs

Oleksandr V Marchukov  <https://orcid.org/0000-0003-0331-0322>

Hans-Werner Hammer  <https://orcid.org/0000-0002-2318-0644>

Artem G Volosniev  <https://orcid.org/0000-0003-0393-5525>

References

- [1] Born M 1957 *Atomic Physics* 6th edn (New York: Hafner Publishing Co.) p 103
- [2] Berry M V, Percival I C and Weiss N O 1987 The Bakerian lecture, 1987 quantum chaology *Proc. R. Soc. A* **413** 183–98
- [3] Belot G and Earman J 1997 Chaos out of order: quantum mechanics, the correspondence principle and chaos *Stud. Hist. Phil. Sci. B* **28** 147–82
- [4] Zurek W H 2003 Decoherence, einselection, and the quantum origins of the classical *Rev. Mod. Phys.* **75** 715–75
- [5] Berry M V, Tabor M and Ziman J M 1977 Level clustering in the regular spectrum *Proc. R. Soc. A* **356** 375–94
- [6] Berry M V 1977 Regular and irregular semiclassical wavefunctions *J. Phys. A: Math. Gen.* **10** 2083–91
- [7] Bohigas O, Giannoni M J and Schmit C 1984 Characterization of chaotic quantum spectra and universality of level fluctuation laws *Phys. Rev. Lett.* **52** 1–4
- [8] Stöckmann H-J 1999 *Quantum Chaos—an Introduction* (Cambridge: Cambridge University Press)
- [9] Haake F 2001 *Quantum Signatures of Chaos* 2nd edn (Berlin: Springer)
- [10] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
- [11] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 Machine learning and the physical sciences *Rev. Mod. Phys.* **91** 045002
- [12] Berry M 1989 Quantum chaology, not quantum chaos *Phys. Scr.* **40** 335–6
- [13] Kharkov Y A, Sotskov V E, Karazeev A A, Kiktenko E O and Fedorov A K 2020 Revealing quantum chaos with machine learning *Phys. Rev. B* **101** 064406
- [14] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F and Pedreschi D 2019 A survey of methods for explaining black box models *ACM Comput. Surv.* **51** 1
- [15] Olah C, Satyanarayan A, Johnson I, Carter S, Schubert L, Ye K and Mordvintsev A 2018 The building blocks of interpretability *Distill* **3** e10
- [16] Zhang Q-s and Zhu S-c 2018 Visual interpretability for deep learning: a survey *Front. Inf. Technol. Electron. Eng.* **19** 27–39
- [17] Kaspchak B and Meißner U-G 2020 How machine learning conquers the unitary limit (arXiv:2003.09137 [physics.comp-ph])
- [18] Kaspchak B and Meißner U-G 2020 A neural network perturbation theory based on the born series (arXiv:2009.03192 [cs.LG])
- [19] Dawid A, Huembeli P, Tomza M, Lewenstein M and Dauphin A 2020 Phase detection with neural networks: interpreting the black box *New J. Phys.* **22** 115001
- [20] Zhang Q, Wang X, Cao R, Wu Y N, Shi F and Zhu S-C 2020 Extracting an explanatory graph to interpret a CNN *IEEE Trans. Pattern Anal. Mach. Intell.* **1**
- [21] Richter A 1999 Playing billiards with microwaves—quantum manifestations of classical chaos *Emerging Applications of Number Theory* ed D A Hejhal, J Friedman, M C Gutzwiller and A M Odlyzko (Berlin: Springer) pp 479–523
- [22] Berry M V and Wilkinson M 1984 Diabolical points in the spectra of triangles *Proc. R. Soc. A* **392** 15–43
- [23] Li W K and Blinder S M 1985 Solution of the Schrödinger equation for a particle in an equilateral triangle *J. Math. Phys.* **26** 2784–6
- [24] Kaufman D L, Kosztin I and Schulten K 1999 Expansion method for stationary states of quantum billiards *Am. J. Phys.* **67** 133–41
- [25] de Aguiar F M 2008 Quantum properties of irrational triangular billiards *Phys. Rev. E* **77** 036201
- [26] Araújo Lima T, Rodríguez-Pérez S and de Aguiar F M 2013 Ergodicity and quantum correlations in irrational triangular billiards *Phys. Rev. E* **87** 062902
- [27] Rawat W and Wang Z 2017 Deep convolutional neural networks for image classification: a comprehensive review *Neural Comput.* **29** 2352–449
- [28] Wang L 2016 Discovering phase transitions with unsupervised learning *Phys. Rev. B* **94** 195105
- [29] Broecker P, Carrasquilla J, Melko R G and Trebst S 2017 Machine learning quantum phases of matter beyond the fermion sign problem *Sci. Rep.* **7** 8823

- [30] Hu W, Singh R R P and Scalettar R T 2017 Discovering phases, phase transitions, and crossovers through unsupervised machine learning: a critical examination *Phys. Rev. E* **95** 062122
- [31] Zhang Y et al 2019 Machine learning in electronic-quantum-matter imaging experiments *Nature* **570** 484–90
- [32] Rem B S, Käming N, Tarnowski M, Asteria L, Fläschner N, Becker C, Sengstock K and Weitenberg C 2019 Identifying quantum phase transitions using artificial neural networks on experimental data *Nat. Phys.* **15** 917–20
- [33] Bohrdt A, Chiu C S, Ji G, Xu M, Greif D, Greiner M, Demler E, Grusdt F and Knap M 2019 Classifying snapshots of the doped Hubbard model with machine learning *Nat. Phys.* **15** 921–4
- [34] Pekalski J, Rządowski W and Panagiotopoulos A Z 2020 Shear-induced ordering in systems with competing interactions: a machine learning study *J. Chem. Phys.* **152** 204905
- [35] Rządowski W, Defenu N, Chiacchiera S, Trombettoni A and Bighin G 2020 Detecting composite orders in layered models via machine learning *New J. Phys.* **22** 093026
- [36] Caux J-S and Mossel J 2011 Remarks on the notion of quantum integrability *J. Stat. Mech.* P02023
- [37] Li Y-Q and Ma Z-S 1995 Exact results of a hard-core interacting system with a single impurity *Phys. Rev. B* **52** R13071–4
- [38] McGuire J B and Dirk C 2001 Extending the Bethe ansatz: the quantum three-particle ring *J. Stat. Phys.* **102** 971
- [39] Krishnamurthy H R, Mani H S and Verma H C 1982 Exact solution of the Schrödinger equation for a particle in a tetrahedral box *J. Phys. A: Math. Gen.* **15** 2131–7
- [40] Glashow S L and Mittag L 1997 Three rods on a ring and the triangular billiard *J. Stat. Phys.* **87** 937–41
- [41] Lamcraft A 2013 Diffractive scattering of three particles in one dimension: a simple result for weak violations of the Yang–Baxter equation *Phys. Rev. A* **87** 012707
- [42] Barfknecht R E, Brouzos I and Foerster A 2015 Contact and static structure factor for bosonic and fermionic mixtures *Phys. Rev. A* **91** 043640
- [43] Joseph S and Sanjuán M 2016 Entanglement entropy in a triangular billiard *Entropy* **18** 79
- [44] Guan X-W, Batchelor M T and Lee C 2013 Fermi gases in one dimension: from Bethe ansatz to experiments *Rev. Mod. Phys.* **85** 1633–91
- [45] Schachner H C and Obermair G M 1994 Quantum billiards in the shape of right triangles *Z. Phys. B* **95** 113–9
- [46] Olshanii M and Jackson S G 2015 An exactly solvable quantum four-body problem associated with the symmetries of an octacube *New J. Phys.* **17** 105005
- [47] Loft N J S, Dehkharghani A S, Mehta N P, Volosniev A G and Zinner N T 2015 A variational approach to repulsively interacting three-fermion systems in a one-dimensional harmonic trap *Eur. Phys. J. D* **69** 65
- [48] Scoquart T, Seaward J J, Jackson S G and Olshanii M 2016 Exactly solvable quantum few-body systems associated with the symmetries of the three-dimensional and four-dimensional icosahedra *SciPost Phys.* **1** 005
- [49] Harshman N L, Olshanii M, Dehkharghani A S, Volosniev A G, Jackson S G and Zinner N T 2017 Integrable families of hard-core particles with unequal masses in a one-dimensional harmonic trap *Phys. Rev. X* **7** 041001
- [50] Liu Y, Qi F, Zhang Y and Chen S 2019 Mass-imbalanced atoms in a hard-wall trap: an exactly solvable model associated with D_6 symmetry *iScience* **22** 181–94
- [51] Miltenburg A G and Ruijgrok T h W 1994 Quantum aspects of triangular billiards *Physica A* **210** 476–88
- [52] Dehkharghani A S, Volosniev A G and Zinner N T 2016 Impenetrable mass-imbalanced particles in one-dimensional harmonic traps *J. Phys. B: At. Mol. Opt. Phys.* **49** 085301
- [53] Lee T D, Low F E and Pines D 1953 The motion of slow electrons in a polar crystal *Phys. Rev.* **90** 297–302
- [54] Volosniev A G and Hammer H-W 2017 Analytical approach to the Bose-polaron problem in one dimension *Phys. Rev. A* **96** 031601
- [55] Panochko G and Pastukhov V 2019 Mean-field construction for spectrum of one-dimensional Bose polaron *Ann. Phys., NY* **409** 167933
- [56] Mistakidis S I, Volosniev A G, Zinner N T and Schmelcher P 2019 Effective approach to impurity dynamics in one-dimensional trapped Bose gases *Phys. Rev. A* **100** 013619
- [57] Jäger J, Barnett R, Will M and Fleischhauer M 2020 Strong-coupling Bose polarons in one dimension: condensate deformation and modified Bogoliubov phonons *Phys. Rev. Res.* **2** 033142
- [58] Hill R N 1980 An exactly solvable one-dimensional three-body problem with hard cores *J. Math. Phys.* **21** 1083–5
- [59] Stöckmann H-J and Stein J 1990 ‘Quantum’ chaos in billiards studied by microwave absorption *Phys. Rev. Lett.* **64** 2215–8
- [60] Sridhar S and Heller E J 1992 Physical and numerical experiments on the wave mechanics of classically chaotic systems *Phys. Rev. A* **46** R1728–31
- [61] Lenz F, Liebchen B, Diakonov F K and Schmelcher P 2011 Resonant population transfer in the time-dependent quantum elliptical billiard *New J. Phys.* **13** 103019
- [62] Pechukas P 1983 Distribution of energy eigenvalues in the irregular spectrum *Phys. Rev. Lett.* **51** 943–6
- [63] Yukawa T 1985 New approach to the statistical properties of energy levels *Phys. Rev. Lett.* **54** 1883–6
- [64] Ivrii V 2016 100 years of Weyl’s law *Bull. Math. Sci.* **6** 379–452
- [65] Bohigas O and Giannoni M-J 1984 *Mathematical and Computational Methods in Nuclear Physics* ed J S Dehesa, J M G Gomez and A Polls (Berlin: Springer) pp 1–99
- [66] Prosen T and Robnik M 1993 Energy level statistics in the transition region between integrability and chaos *J. Phys. A: Math. Gen.* **26** 2371–87
- [67] Ben Arous G and Bourgade P 2013 Extreme gaps between eigenvalues of random matrices *Ann. Probab.* **41** 2648–81
- [68] Blomer V, Bourgain J, Radziwiłł M and Rudnick Z 2017 Small gaps in the spectrum of the rectangular billiard *Ann. Sci. Éc. Norm. Sup.* **50** 1283–300
- [69] Kaplan L and Heller E J 1998 Linear and nonlinear theory of eigenfunction scars *Ann. Phys., NY* **264** 171–206
- [70] Evans J R and Stockman M I 1998 Turbulence and spatial correlation of currents in quantum chaos *Phys. Rev. Lett.* **81** 4624–7
- [71] Berggren K-F, Pichugin K N, Sadreev A F and Starikov A 1999 Signatures of quantum chaos in the nodal points and streamlines in electron transport through billiards *JETP Lett.* **70** 403–9
- [72] Jain S R and Samajdar R 2017 Nodal portraits of quantum billiards: domains, lines, and statistics *Rev. Mod. Phys.* **89** 045005
- [73] Shapiro M and Goelman G 1984 Onset of chaos in an isolated energy eigenstate *Phys. Rev. Lett.* **53** 1714–7
- [74] McDonald S W and Kaufman A N 1988 Wave chaos in the stadium: statistical properties of short-wave solutions of the Helmholtz equation *Phys. Rev. A* **37** 3067–86
- [75] Samajdar R and Jain S R 2018 Exact eigenfunction amplitude distributions of integrable quantum billiards *J. Math. Phys.* **59** 012103

- [76] Elson J, Douceur J, Howell J and Saul J J 2007 A captcha that exploits interest-aligned manual image categorization *Proc. of 14th ACM Conf. on Computer and Communications Security (CCS)* (Association for Computing Machinery, Inc.)
- [77] James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning* (Berlin: Springer)
- [78] Wilson K G 1983 The renormalization group and critical phenomena *Rev. Mod. Phys.* **55** 583–600
- [79] Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I and Fergus R 2014 Intriguing properties of neural networks (arXiv:1312.6199 [cs.CV])
- [80] Hampel F R 1974 The influence curve and its role in robust estimation *J. Am. Stat. Assoc.* **69** 383–93
- [81] Koh P W and Liang P 2017 Understanding black-box predictions via influence functions *Int. Conf. on Machine Learning, ICML'17* vol 70 (JMLR Org.) pp 1885–94
- [82] Shorten C and Khoshgoftaar T M 2019 A survey on image data augmentation for deep learning *J. Big Data* **6** 60
- [83] Prigodin V N 1995 Spatial structure of chaotic wave functions *Phys. Rev. Lett.* **74** 1566–9
- [84] Harshman N L 2017 Infinite barriers and symmetries for a few trapped particles in one dimension *Phys. Rev. A* **95** 053616
- [85] Keski-Rahkonen J, Ruhanen A, Heller E J and Räsänen E 2019 Quantum Lissajous scars *Phys. Rev. Lett.* **123** 214101
- [86] Marchukov O V, Volosniev A G, Fedorov D V, Jensen A S and Zinner N T 2014 Statistical properties of spectra in harmonically trapped spin-orbit coupled systems *J. Phys. B: At. Mol. Opt. Phys.* **47** 195303
- [87] Bergschneider A, Klinkhamer V M, Becher J H, Klemt R, Palm L, Zürn G, Jochim S and Preiss P M 2019 Experimental characterization of two-particle entanglement through position and momentum correlations *Nat. Phys.* **15** 640–4
- [88] Das Sarma S, Deng D-L and Duan L-M 2019 Machine learning meets quantum physics *Phys. Today* **72** 48
- [89] Courant Richard Hilbert David 1989 *Methods of Mathematical Physics* vol 1 (New York: Wiley)