PhD Dissertation

Quantum many-body physics with ultracold atoms and molecules:

exact dynamics and machine learning

Author:
Anna Dawid-Łękowska

Supervisors:
Dr. hab. Michał Tomza, prof. UW
Prof. Dr. hab. Maciej Lewenstein

March 4, 2022

Abstract

Quantum many-body physics poses a substantial computational challenge resulting from the exponential growth of the wave function complexity and many non-trivial correlations encoded in it. Studying many-body systems is thus a demanding quest that is being tackled via various methods. The research described within this thesis concerns two parallel approaches that are gaining the attention of the scientific community: quantum simulations with ultracold molecules and interpretable machine learning.

The first research path is a detailed analysis of the ultracold system of two interacting molecules in a one-dimensional trap. By comparing with the two-atom system in a harmonic trap, we identify differences in spectra and reactions to the external fields introduced by the molecular character of the system, i.e., rotational levels, anisotropic short-range interactions, and stronger dipolar interactions. Exactly these richer properties of molecules could allow for discovering new exotic phases of matter and simulating phenomena that are inaccessible for the physics of ultracold atoms. Inspired by materials with both electric and magnetic orders, in the next step, we focus on the interplay of the electric and magnetic properties of the two-body molecular system, analyze magnetization diagrams, and study the quench dynamics.

Alternatively, quantum many-body problems can be solved via numerical methods. Among them, machine learning algorithms are gaining significant momentum. However, so far, they have mostly enabled only the recovery of known results (but at much lower computational cost). Moreover, we usually lack the understanding of how the machine solves the problem at hand. Therefore, we propose a way to combine the efficiency of neural networks with Hessian-based interpretability and reliability methods like influence functions. In principle, these universal and model-independent tools allow to unravel the logic hidden in the machine and thus increase the chance to understand the physics of the problem. We show their power on the fundamental one-dimensional Fermi-Hubbard model and on the experimental data obtained from the Floquet realization of the topological two-dimensional Haldane model.

Streszczenie

Fizyka kwantowa wielu ciał stanowi duże wyzwanie obliczeniowe wynikające z wykładniczego wzrostu złożoności funkcji falowej i zakodowanych w niej wielu nietrywialnych korelacji. Na badania prowadzone w ramach tej rozprawy doktorskiej składa się eksploracja dwóch podejść do rozwiązywania problemów kwantowej fizyki wielu ciał. Pierwszym z nich są symulacje kwantowe z użyciem ultrazimnych cząsteczek, a drugim - uczenie maszynowe.

Pierwsza ścieżka badawcza to szczegółowa analiza ultrazimnego układu dwóch oddziałujących ze sobą cząsteczek w jednowymiarowej pułapce. Porównując wyniki z dwuatomowym układem w pułapce harmonicznej, identyfikujemy różnice w widmach i reakcjach na pola zewnętrzne wprowadzone przez molekularny charakter układu, tj. obecność poziomów rotacyjnych, anizotropowego oddziaływania krótkozasięgowego i silniejszych oddziaływań dipolarnych. Właśnie te bogatsze właściwości molekuł mogą pozwolić na odkrywanie nowych egzotycznych faz materii i symulowanie zjawisk niedostępnych dla fizyki ultrazimnych atomów. Zainspirowani materiałami o porządkach elektrycznych i magnetycznych, w kolejnym kroku skupiamy się na wzajemnym oddziaływaniu właściwości elektrycznych i magnetycznych dwuciałowego układu molekularnego, analizujemy diagramy magnetyzacji i badamy dynamikę po quenchu.

Alternatywnie, kwantowe problemy wielociałowe można rozwiązywać metodami numerycznymi. Wśród nich coraz większym zainteresowaniem cieszą się algorytmy uczenia maszynowego. Jednak do tej pory umożliwiały one głównie odtworzenie znanych wyników (ale przy znacznie niższych kosztach obliczeniowych). Co więcej, zwykle nie rozumiemy, w jaki sposób maszyna rozwiązuje dany problem. Dlatego proponujemy sposób na połączenie wydajności sieci neuronowych z bazującymi na Hesjanie metodami interpretacji i wiarygodności, takimi jak funkcje wpływu. Te uniwersalne i niezależne od modelu narzędzia pozwalają rozwikłać logikę ukrytą w maszynie i tym samym zwiększyć szansę na zrozumienie fizyki problemu. Pokazujemy ich moc na fundamentalnym jednowymiarowym modelu Fermi-Hubbarda oraz na danych eksperymentalnych uzyskanych z topologicznego dwuwymiarowego modelu Haldane'a w realizacji Floqueta.

Resumen

La física cuántica de muchos cuerpos plantea un desafío computacional sustancial debido al crecimiento exponencial de la complejidad de la función de onda y las múltiples correlaciones no triviales representadas en ella. Así pues, estudiar sistemas de muchos cuerpos supone una ardua tarea que se está abordando mediante varios métodos. La investigación descrita en esta tesis concierne dos enfoques paralelos que están ganando atención en la comunidad científica: las simulaciones cuánticas con moléculas ultra-frías y el aprendizaje automático interpretable.

La primera línea de investigación es un análisis detallado de un sistema ultra-frío de dos moléculas que interactúan en una trampa unidimensional. Al comparar con un sistema de dos átomos en una trampa armónica, identificamos diferencias en los espectros y reacciones a los campos externos inducidas por el carácter molecular del sistema. Es decir, aparecen niveles rotacionales, interacciones anisotrópicas de corto alcance, e interacciones dipolares más fuertes. Precisamente, estas propiedades más ricas de las moléculas podrían permitir descubrir nuevas fases exóticas de la materia y simular fenómenos inaccesibles para la física de átomos ultra-fríos. Inspirándonos en materiales con órdenes eléctricos y magnéticos, en el siguiente paso, nos centramos en la interacción de las propiedades eléctricas y magnéticas de sistemas de dos moléculas, analizamos diagramas de magnetización, y estudiamos la dinámica después de un cambio abrupto de los parámetros del sistema.

Alternativamente, los problemas cuánticos de muchos cuerpos se pueden resolver mediante métodos numéricos. Entre ellos, destacan los algoritmos de aprendizaje automático, que están ganando cada vez más importancia. Aún así, por el momento, solo han permitido recuperar resultados conocidos (aunque a un coste computacional mucho más bajo). Además, normalmente no entendemos cómo la máquina resuelve el problema en cuestión. Por tanto, proponemos una forma de combinar la eficiencia de las redes neuronales con métodos de interpretación basados en la matriz Hessiana y métodos de fiabilidad como la función de influencia. En principio, estas herramientas universales e independientes del modelo permiten desentrañar la lógica oculta en la máquina y, así, aumentar la probabilidad de entender la física del problema. Mostramos su poder en el fundamental modelo de Fermi-Hubbard en una dimensión y sobre datos experimentales obtenidos en realizaciones de Floquet del modelo topológico de Haldane en dos dimensiones.

Page of gratitude and appreciation

Wow, what a ride! I loved every minute of it (weeell, a large majority!). While I acknowledge my enthusiastic approach to life, this love would be much, much harder without terrific people that prepared the tracks, took care of both the wagon and its driver, and made the whole journey just wonderful.

Regarding people who prepared the tracks - I thank Michał Tomza for always making grand plans and teaching me basically everything: how to do science; how to be a scientist; how to find and use your strengths; how to understand and not be discouraged by own weaknesses. Remember, everyone, "panic and despair are forbidden"!

Of course, I also thank Maciek Lewenstein for opening me all the fantastic new doors I never even knew were closed. Being in your group and under your supervision gives so many opportunities that I wish everyone to have.

Alex Dauphin, thank you so much for your guidance in the machine learning and topology jungles. It is very easy to get lost there, and you never let it happen!

To Niklas Käming and Christof Weitenberg for making my first inter-institution collaboration extremely enjoyable and convincing me in the process that I would like to do it more often! Also, to the Quantum Optics Theory Group - thanks to you, I understood what a vibrant scientific community feels like. I will do my best to share the same vibes and build such an amazing space wherever my career will lead - you have undoubtedly been inspiring!

Last but definitely not least: Vindhiya Prakash, Ola Sierant, Piotrek Grochowski, Borja Requena Pozo, and Marcin Płodzień. Guys, you are incredible. Each of you is glorious and thought-provoking in a different way. Thanks for the laughs, scientific and not-so-scientific discussions. You made me a better scientist and a better person.

No i wreszcie najlepsi, najwspanialsi, najważniejsi. Mamo, Tato, dziękuję za Wasze wsparcie i odpowiednio celne kopniaki, gdy nie miałam siły. Chęć, żebyście byli ze mnie dumni i Wasza miłość to najlepsze motywacje na świecie.

A Tobie, Mirusiu - Ty już wiesz za co.

Contents

1

Introduction

The Hilbert space is a space of possibilities.
~Wolfgang Ketterle

1.1 Quantum many-body problems are difficult

Quantum many-body physics poses a substantial computational challenge for modern science for two reasons. Firstly, the complexity of the quantum many-body system description grows exponentially with the number of its constituents. It is an example of the so-called *curse of dimensionality* which is a phrase used across many scientific fields describing in general the apparent intractability of systematically searching through a high-dimensional space, the apparent intractability of accurately approximating a general high-dimensional function, or the apparent intractability of integrating a high-dimensional function [1].[1]

The second reason is of a much more fundamental nature. The behavior of large and complex aggregates of known particles rarely is a simple extrapolation of the properties of single bodies. In particular, the ground state of a many-body system may exhibit different symmetry than its governing Hamiltonian due to the spontaneous symmetry breaking [2]. This emergent behavior of many-body systems has been already noted by P. W. Anderson in 1972 [3]:

[1] Interestingly, there are also phenomena known as "blessings of dimensionality". For example, asymptotic methods widely used in statistical physics are so successful because the high-dimensional setting enables making some statements that largely simplify calculations [1].

> The reductionist hypothesis may still be a topic for controversy among philosophers, but among the great majority of active scientists I think it is accepted without question. The workings of our minds and bodies, and of all the animate or inanimate matter of which we have any detailed knowledge, are assumed to be controlled by the same set of fundamental laws, which except under certain extreme conditions we feel we know pretty well [...] the reductionist hypothesis does not by any means imply a "constructionist" one: *The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe.*

As a result, studying quantum many-body physics is a challenging quest that can be solved either by direct mathematical and numerical analysis or with devices that simulate its behavior, so-called quantum simulators.

This thesis explores both such directions. The first part of this introductory chapter (section 1.2) discusses the concept of quantum simulations and presents successes of ultracold atoms in optical lattices as simulators of many-body phenomena. Then, we explore the idea of replacing atoms with more complex particles like molecules and the promises it brings to the field of quantum simulations. In parallel, the quantum community has been harnessing the power of machine learning (ML) to solve conundrums of quantum many-body physics. We discuss this approach, its breakthroughs and challenges in section 1.3.

1.2 Quantum simulations at ultralow temperatures

A *simulator* is a well-founded concept in science. Prominent examples are astronomical clocks built in Asia and Europe in the period between 1000 and 1500 AD. They were used to predict the positions of the planets and the constellations as well as phases of the moon and its eclipses. On the other hand, a history of *quantum simulators* has only begun with the speech of Richard Feynman in 1981 [4] and his vision of "a quantum machine that could imitate any quantum system, including the physical world", known also as a universal (or digital) quantum simulator. This idea was further extended by Seth Lloyd in 1996 [5] and now is better known as a quantum computer. However, we are still at infancy of its experimental realization, i.e., in the so-called noisy intermediate-scale quantum (NISQ) era [6].

In turn, if instead of a machine solving any quantum problem, we go for more modest machines solving a selected class of quantum problems,

we enter the field of analog quantum simulations. This concept is about addressing an interesting, yet difficult or unsolvable quantum problem by building a simpler physical system that is described by the same Hamiltonian and performing measurements on it. Such simulators operate by individually manipulating its constituting quantum particles and their interactions. Therefore, those machines are inherently quantum which lifts the need of classically tracking all possible configurations that scale exponentially with the system size. Moreover, if we focus on certain physical properties, e.g., densities, magnetization per lattice site, or few-body correlations, instead of the full wave function, then results obtained with such simulators are to some degree robust against imperfections [7].

Therefore, the construction of such a device starts with a proper mapping of the Hamiltonian of the system being interesting to science or technology, but difficult to study. Next, the simulator needs to be prepared in a state relevant to the physical problem of interest, e.g., the ground state. Also, measurements must be performed on the simulator with the highest possible precision. Finally, Cirac & Zoller also indicated that there should be a way of increasing the confidence in the result, like benchmarking it first against the known solutions [7].

1.2.1 Ultracold atoms in lattices as quantum simulators

Quantum simulators need to be then highly controllable, well-isolated from the environment, allow for precise state detection, and be able to mimic interesting systems. All these conditions are met by trapped quantum atomic systems at ultralow temperatures [8].[2] As a result, ultracold atoms in traps are especially useful for quantum simulation of various models of many-body physics, and a plethora of quantum phenomena have been investigated and understood [13, 14].

Quantum simulations of strongly interacting systems with ultracold atoms loaded in optical lattice[3] have started [16] with experimental studies of the quantum dynamics of phase transition [17] and collapse and revival of the matter-wave field [18]. Soon after, Paredes *at al.* realized for the first time the Tonks–Girardeau gas where repulsive interactions between bosonic particles confined to one dimension dominate the physics of the system [19]. Highly tunable lattice also allowed for creation and observation of nearest-

[2] Other promising and interesting platforms for quantum simulations include, e.g., trapped ultracold ions [9], superconducting qubits [10, 11], and cold hybrid ion-atom systems [12].

[3] Of course, in parallel important research has been conducted with ultracold gases in the continuum, e.g., realization of the Bose-Einstein condensation–Bardeen-Cooper-Schrieffer crossover [15].

neighbor magnetic spin correlations in a thermalized Fermi gas [20]. At this point, the main source of the system tunability was provided by the Feshbach resonances controlling interatomic interactions and the *versatility of optical lattices*, schematically visualized in fig. 1.1(a). Experimentalists are able to control their geometry, dimensionality, depth, and disorder [21]. Moreover, they can engineer lattices to carry effective magnetic fields which are hundreds of times stronger than the strongest magnetic fields in solid-state laboratories [22].

The next breakthrough was the development of the *quantum gas microscopy* for bosons [23, 24] and fermions [25, 26]. It has enabled the single site-resolved detection of individual atoms (or more precisely, detection of the local parity of the on-site atom number) in a two-dimensional lattice, as presented in fig. 1.1(b). Finally, a simultaneous detection of two hyperfine spin states has also been demonstrated [27] leading to the single-atom level detection of spin and charge (density). Armed with such tools, Fukuhara *et al.* observed magnons for the first time [28], and studied the quantum dynamics of the spin impurity [29]. In the absence of reliable theoretical predictions, Choi *et al.* observed the two-dimensional many-body localization in the ultracold system [30]. Mazurenko *et al.* realized the antiferromagnetic repulsive gas of cold fermions [31]. A recent experimental realization of the two-dimensional Hubbard model with impurity showed that its accurate theoretical description is provided by geometric strings [32].

The last experimental breakthrough is the combination of the single-atom-level detection techniques with optical tweezers that enables construction of atomic systems *without defects* and with arbitrary geometries [33–35] as one can see in fig. 1.1(c). Since then, a fully controllable atomic quantum simulators only grow in size while preserving their fidelity. An impressive experimental progress can be presented also in terms of sizes of trapped atomic simulators. In 2017, Bernien *et al.* realized the programmable quantum spin model with tunable interactions and size up to 51 qubits [36]. In 2021, programmable quantum simulators had 219 and 256 atoms [37, 38]. In 2022, the ultracold arrays are made out of two atomic species [39, 40]! In this way the production of fully controllable synthetic quantum matter can be achieved using both top-down and bottom-up approaches.

With an increasing size of experimental ultracold toolbox, various regimes have been explored. For example, one-dimensional systems have attracted significant attention [19, 41–43], due to the important role played by quantum fluctuations [44]. Ultracold highly magnetic atoms were used to simulate the extended Bose-Hubbard models [45], the formation of self-bound micro droplets of a dilute magnetic quantum liquid [46–48], and the Fermi surface deformation in a dipolar quantum gas [49]. Another fascinating direction is study of topological systems that can be created by adding ar-

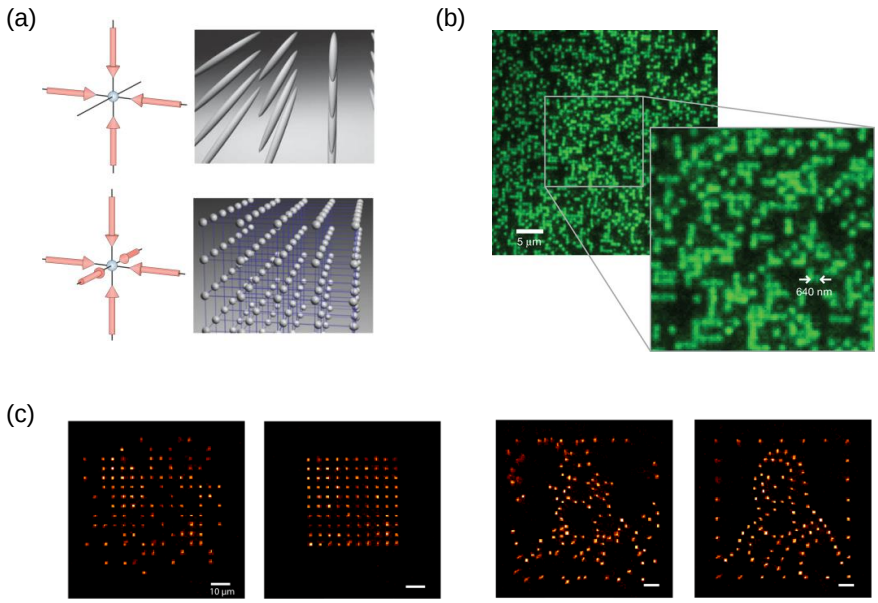

Figure 1.1: **Selected tools from the ultracold toolbox.** Next to the advances in cooling, there were breakthroughs regarding the control and detection of ultracold systems. (a) In 2000s, the main versatility of ultracold systems came from the control over the optical lattices and Feshbach resonances. (b) Quantum gas microscopy allowed for a single-site detection of ultracold systems. (c) When combined with optical tweezers, defect-free arrays of any geometry became possible. Taken (a-b) or adapted (c) with permission from Refs. [23, 54, 55], respectively.

tificial gauge fields [50, 51] using periodic driving, i.e., so-called Floquet engineering [52, 53].

Moreover, next to many-body studies, the deterministic preparation of a *tunable few-fermion systems* with complete control over the number of particles, and their quantum state has become possible [56], opening the way towards quantum simulation of strongly correlated few-body systems. The fermionization of two distinguishable fermions [57], formation of a Fermi sea [58], pairing in few-fermion systems [59], antiferromagnetic Heisenberg spin chain [60], and two fermions [61] or bosons [62, 63] in a double well have been experimentally investigated in one dimension. Such experimental possibilities have motivated also intensive theoretical studies of few-body atomic systems. The analytical solution is known for the general case of two atoms in a harmonic trap interacting *via* contact [64] or finite-range soft-core [65] potential. Energy spectra of harmonically trapped two-atom

systems with spin-orbit coupling have been investigated in one [66] and three dimensions [67, 68]. Low energy states of two atoms with a dipole moment in a three-dimensional harmonic trap have also been investigated [69–71]. Systems of several fermions [72–79], or several bosons [80–84] in a one-dimensional harmonic trap have been studied using various analytical and numerical approaches.

1.2.2 Replacing atoms with molecules

All these successes have been achieved, even though most atomic gases have a simple isotropic and short-range interaction that is well approximated by the contact interaction. Tackling more complex interactions becomes possible when working with magnetic [45] or Rydberg atoms [85] but, e.g., atoms with simultaneous significant electric and magnetic dipolar interactions are probably out of reach. In general, there is a limit to how complex physical phenomena we can simulate with such systems. We can push this limit further by replacing atoms with molecules.

Molecules have inherently more complex structure than atoms [86]. This includes *rotational and vibrational levels* together with possible *permanent electric dipole moment*. The energy scales connected to their degrees of freedom span several orders of magnitude as schematically presented in fig. 1.2. They also have more complex short-range interactions, and stronger dipole-dipole interactions of both magnetic and electric nature, being both long-range and anisotropic [87, 88].

The richer structure of the molecules is the reason for using them in the *measurements of fundamental constants*, i.e., in search for the upper limit for the electric dipole moment value of the electron [89–92] and the time dependence of fundamental constants such as the hyperfine structure constant [93, 94] and the mass ratio of the electron and proton [95, 96]. This search is necessary to eliminate potentially wrong theories aiming at unifying gravity with other interactions [97].

Ultracold molecules have also been employed in the ground-breaking experiments on *quantum-controlled chemistry* [98–104] enabled by the extensive control of internal states and relative motion of molecules with external electromagnetic fields [87, 105–113]. The intermolecular dipolar interactions promise exciting applications in *quantum information processing* [114–119].

Finally, higher complexity of molecules offers new possibilities in *quantum simulations*. Micheli *et al.* in 2006 were one of the first to propose a molecular simulator [120]. Their simplified system of molecules in the two-dimensional optical lattice showed topologically protected states. Molecular rotational states (in which pseudo-spins can be encoded with microwave-field dressing) combined with the dipolar interaction have allowed for several

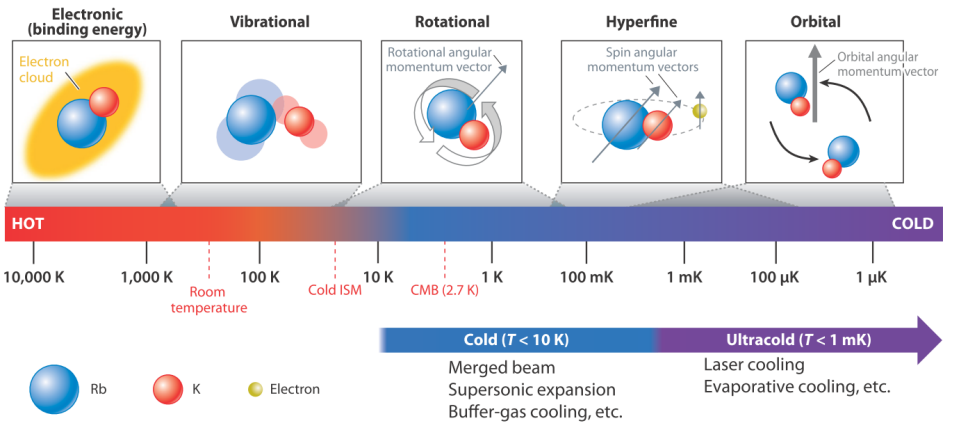

Figure 1.2: **Characteristic energy scales of various degrees of freedom of a diatomic molecule.** They are expressed in equivalent temperature values (T), according to $E = k_{\mathrm{B}}T$, where k_{B} is the Boltzmann constant. The energy scale for a given degree of freedom can span several orders of magnitude due to the wide range of masses, intramolecular bindings, and intermolecular interactions between different molecules. Cold temperatures ($T \lesssim 10\,\mathrm{K}$) can be achieved in molecular beams, but reaching the ultracold regime ($T \lesssim 1\,\mathrm{mK}$) requires techniques such as laser cooling and evaporative cooling. *Abbreviations:* CMB, cosmic microwave background; ISM, interstellar medium. Taken with permission from Ref. [142].

proposals to realize various models of quantum magnetism [121, 122, 122–126] or implement controllable spin-orbit coupling [127] being a key to understanding a variety of spin-transport and topological phenomena. Other suggestions proved the usefulness of ultracold molecules in modeling the extended Hubbard model for polarons [128, 129], magnetic Frenkel excitons [130], and exotic phases [131, 132] such as supersolid [133, 134] or topological ones [135–138]. Another notable example of quantum simulations are structures formed by molecules in a quantum atomic liquid, e.g., superfluid helium. They form angulons, i.e., quantum rotors dressed in field of many-body excitations (rotational analogs of polarons), which may be useful, among others, in simulations of problems of quantum field theory [139–141].

1.2.3 Molecules - experimental challenges

A more complex internal structure of molecules, as compared to atoms, is responsible for greater experimental challenges in molecular formation, cooling, and trapping. In particular, while *laser cooling* has been the indispensable first step of ultracold atomic experiments, in general it does not

work for molecules [142]. Effective laser cooling requires rapid and repeated scattering of photons between two levels without decaying to states not involved in the cooling scheme. Such photon cycles can be found in many atomic species, most notably alkali and alkaline earth atoms but not in molecules. An important exception is a class of molecules with a strongly diagonal Franck-Codon factors. While they are still more challenging to laser-cool than alkali atoms, those difficulties have been overcome for molecules like SrF [143], YO [144], and CaF [145], and recently even extended to polyatomics [146, 147]. On the other hand, evaporative and sympathetic cooling schemes remain important direct approaches also for molecules.

Due to direct laser cooling being limited to molecules with certain properties, in parallel *indirect methods* have been developed, consisting of first cooling atoms down to ultralow temperatures and then associating them into molecules. An example of such indirect cooling is using magnetoassociation via a Feshbach resonance [148] in combination with STIRAP [149]. This approach was pioneered in 2008 for Cs_2 [150] and fermionic $^{40}K^{87}Rb$ molecules [151], and it was later extended to bosonic $^{87}Rb^{133}Cs$ [152, 153], fermionic $^{23}Na^{40}K$ [154], and bosonic $^{23}Na^{87}Rb$ [155] molecules. While obtaining ground-state $^{40}K^{87}Rb$ molecules reaches 2000s, only recently a degenerate Fermi gas of those molecules was produced [156].

An unprecedented control over ultracold molecular collisions has been achieved by selecting molecules' internal state and by tuning dipolar collisions with an external electric field in a reduced dimensionality [101, 157–162]. Ultracold ground-state molecules have also been loaded into optical lattices [163], and dipolar spin-exchange interactions between lattice-confined polar molecules have been observed [164], opening the way towards quantum simulations with molecules. On the other hand, the methods of full quantum control, deterministic preparation, and detection at the single-particle level, developed for ultracold atoms in optical tweezers [33, 35], can readily be employed to molecules [165, 166], further extending the range of applications of ultracold molecules [167]. Recently, the first step towards atom-by-atom assembled few-body molecular systems has been taken [168] and optical tweezers have been used to assemble and control molecules at the single particle level [165, 169–171]. The defect-free molecular arrays are on the way [172].

There is still one challenge of both experimental and theoretical nature that stands in the way of obtaining dense ultracold samples of molecules as well as molecular quantum simulators with more than one molecule per site. Imagine cooling non-reactive molecules and loading them in a trap. Despite the lack of a chemical reaction, the universal two-body losses (i.e., taking place always once the colliding pair reaches the short range) are measured, including set-ups like RbCs [152, 173], NaK [154], and NaRb [155, 174].

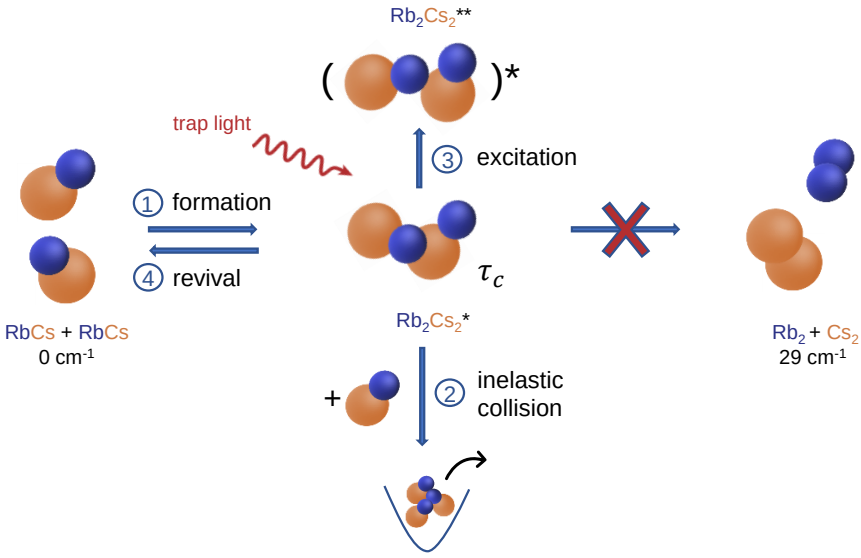

Figure 1.3: **Long-lived intermediate complexes.** Endothermic ultracold collision on the example of RbCs + RbCs. The intermediate complex with lifetime τ_c can either revive back into two molecules, undergo photoexcitation, or collide with the third molecule causing inelastic losses. Inspired by Ref. [142].

Mayle *et al.* [175, 176] hypothesized that the reason is so-called *sticky collisions*. In such a scenario, the colliding pair of molecules forms a long-lived four-atom intermediate complex as presented in fig. 1.3 with (1). When such complex meets the third molecule, the inelastic collision takes place, and molecules are lost as shown with (2). The longer lifetime of the complex, τ_c, the higher chance for the reaction and losses to occur. The Rice-Ramsperger-Kassel-Marcus (RRKM) statistical theory indicates that τ_c is proportional to the density of states of the complex at the incident energy and inversely proportional to the number of exit channels, i.e., quantum states available for the complex to dissociate [142]. For $Rb_2Cs_2^*$, τ_c was estimated to 45 ms [176].

Recently, however, Christiansen *et al.* [177] recalculated the lifetimes of such complexes, and they turned out to be two to three orders of magnitude shorter than in [176]. Therefore, three-body inelastic collisions cannot cause losses as rapid as ones observed in experiments. Instead, Christiansen *et al.* [178] proposed an alternative explanation: the losses are caused by excitations of the intermediate four-atom complexes *caused by the trapping light* as presented in fig. 1.3 with (3). This excitation is possible even if the trapping laser is detuned from all transitions in a molecule due to the

presence of plethora of possible transitions in a four-atom complex. This hypothesis got confirmed by the consequent experiments with RbCs [173, 179] where *revival* of $Rb_2Cs_2^*$ complexes back into RbCs was observed when the trapping light was off, as presented with arrow (4) in fig. 1.3.

However, the riddle is not yet fully solved. While the photoexcitation hypothesis from Ref. [178] and corrected lifetimes of complexes predicted from the RRKM theory from Ref. [177] agree with measurements performed for RbCs [173, 179], no evidence of revival in the absence of trapping light was found for other species. Moreover, the estimates for complexes' lifetimes resulting from measured losses are in a deep disagreement with RRKM theory for bosonic $^{23}Na^{87}Rb$ [180] and $^{23}Na^{39}K$ [181] molecules (by one-two orders of magnitude) as well as for a fermionic $^{23}Na^{40}K$ molecule [181] (by at least two orders of magnitude) and especially for collisions between $^{40}K^{87}Rb$ molecules and ^{87}Rb (five orders of magnitude difference!) [182]. There is an undergoing research addressing these discrepancies [183, 184]. In particular, Jachymski *et al.* [183] indicate that when hyperfine structure is taken into account, the RRKM theory recovers most of the experimentally measured lifetimes.[4]

Next to the efforts to understand the molecular losses, *shielding mechanisms* using electric [185–187] or microwave fields [188] have been developed. They aim at creating avoided crossings in the system to keep molecules at the distance, preventing them from short-range interactions and collisions, reducing the two-body losses by an order of magnitude. Another way of avoiding losses associated with the trap light is to study molecules with a magnetic dipole moment (such as fermionic $^{23}Na^6Li$ in their triplet ground state [189]) and use magnetic (rather than optical) traps.

These advancements pave the way towards so far unexplored regime of two and more molecules in a trap which constitute a missing building block of a molecular Hubbard model. Moreover, as we have explained in section 1.2.2, systems with magnetic and electric orders (like multiferroics [190]) are probably out of reach for atomic simulators, while being a natural system to mimic for highly magnetic and polar molecules. Guided by these motivations, within this thesis, we study the system of two ultracold molecules in a trap and focus on the interplay of their electric and magnetic properties. We devote to this research part I of the thesis, starting with chapter 2.

While quantum simulators may be the future of the quantum many-body physics, they cannot do without theory and numerics. They are needed both

[4] This correction explains especially the large difference for KRb + Rb as Rb has a strong hyperfine coupling. However, including hyperfine structure to the RRKM theory causes the discrepancy between the prediction of a $Rb_2Cs_2^*$ lifetime and the measured one. Ref. [183] argues that it is due the light absorption timescales used in those experiments.

to check the results obtained from quantum simulations in some regimes and to fully understand those results. Until now, the density-matrix renormalization group (DMRG) and tensor networks [191–193] have proved to be the most effective numerical methods for solving many-body problems. Besides, the quantum physics community has begun to explore a new approach - machine learning (ML).

1.3 Machine learning quantum many-body physics

Computers have been aiding in scientific discovery since 1950s. In 1955 a numerical simulation gave one of the first qualitative novel insights into a physical problem, namely into the famous Fermi–Pasta–Ulam–Tsingou problem [194].[5] It is considered as the first ever numerical experiment which is central in the solitons and chaos theories. Moreover, it was also one of the first out-of-equilibrium studies of the statistical mechanics. Since then, the role of computers in science has only grown, especially in the fields of many-particle dynamics, nonlinear dynamics, statistical physics, cosmology, climate science, quantum chemistry, and material science.

In parallel with increasing impact of numerical simulations, computer science has been developing entirely new data-driven paradigm of programming: machine learning (ML) [195, 196]. Nowadays, ML influences everyday life in multiple ways with applications like text and voice recognition software, fingerprint identification, self-driving cars, robotics, and many others. These versatile algorithms, dealing with big and high-dimensional data, are familiar with the curse of dimensionality and have an only growing influence on science. ML algorithms have already been encompassed by quantum chemistry, especially as an effective representation of potential energy surfaces [197–199] whose accuracy is a key element of efficient molecular dynamics or Monte Carlo simulations. ML has also significantly improved sampling for molecular dynamics [200] and promises generation of novel molecules with desired properties [201]. The material science uses ML to search huge databases for hidden relations between atomic and molecular structures and properties of interest [202] or to design new materials [203]. In medicine, ML was used, e.g., to analyze brain samples [204] and to classify tumors of the nervous system [205]. An example of ML groundbreaking impact on science is AlphaFold which continuously beats other numerical approaches in predicting the protein folding [206]. Another inspiring application of ML in science is the prediction of research trends [207].

[5] If a reader wonders why "Fermi–Pasta–Ulam–*Tsingou* problem", I highly recommend reading Ref. [194].

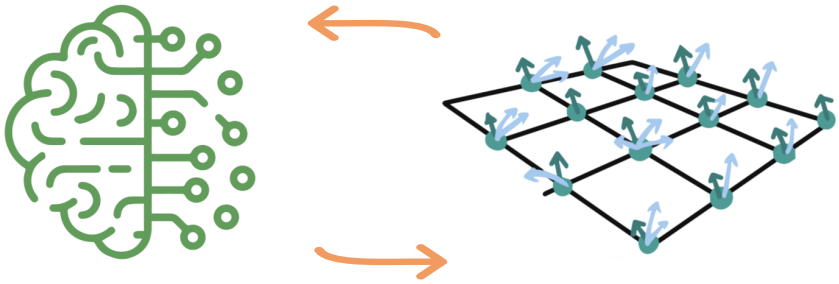

Figure 1.4: **Machine learning (ML) and quantum many-body physics.** While ML revolutionizes quantum many-body physics, there is also a less known reverse influence that includes physics-informed and quantum-inspired models as well as using statistical physics to solve riddles of ML.

ML methods are also becoming important numerical tools in quantum many-body physics [208, 209]. It is not surprising if one thinks about similar goals of physics and learning algorithms, i.e., collecting and analyzing data to find a model that predicts the behavior of a complex system. Interestingly, physics and ML seemingly use entirely different strategies for achieving the same objective. The driving force of physicists is an understanding of the mechanisms governing models created through knowledge, intelligence, and intuition. On the other hand, the logic building ML models is largely out of our reach, and "intelligence" is extracted from data. What is surprising in all this - this incomprehensible logic often achieves excellent results.

Neural networks (NNs), one of the ML models, are celebrating successes as *representations of quantum states*. It started with Carleo & Troyer who found the ground states of the Ising transverse model and the Heisenberg antiferromagnetic model with the variational method and simple NN as a wave function ansatz. In both one and two dimensions, this approach effectively overcame all known numerical methods [210]. Deng *et al.* also showed that the NN is an effective way of representing states with huge entanglement [211], and its deep version is even more effective [212]. Nowadays, one of the most efficient ansätze of many-electron wave functions is the PauliNet [213].

Another fruitful direction is the combination of ML and *quantum state tomography*, which aims to reconstruct the density matrix of an unknown quantum state by using experimentally available data [214–217].

ML techniques have also been encompassed by *quantum experiments*.

Here, ML can unfold its full potential by identifying the relevant information despite noise and other imperfections such as finite temperature or restricted access to relevant observables. ML methods have already been used to optimize preparation of experiments [218–222] and proposed to avoid the need for tedious human work in laboratories [223, 224]. An especially interesting direction is to use ML as a verification tool for preparation of quantum simulators via Hamiltonian learning [225]. Another fascinating idea is the comparison of experimental data to competing theory descriptions in an unbiased way [226] or generation of new snapshots of the experimental set-up at unobserved conditions using generative models [227]. Finally, ML methods have been proposed to design new quantum experiments [228, 229].

ML has also found its application in quantum information [230–232], astrophysics [233], and high-energy physics [234, 235]. ML techniques were also used to accelerate Monte Carlo simulations [236–238].

While the impact of ML on physical sciences is already well-known, there is also a fascinating reverse influence coming from physics to ML as presented schematically in fig. 1.4. Examples are the physics-informed ML [239, 240] and quantum-inspired ML [241]. Moreover, methods of statistical physics are used to solve conundrums of the ML theory including generalization abilities of overparametrized models [242].

1.3.1 Machine learning phases of matter

At the same time, especially abundant is the use of ML in phase classification. It is not surprising if one considers that determining the proper order parameters for unknown transitions is no trivial task, on the verge of being an art. It includes the search in the exponentially large Hilbert space and the examination of symmetries existing in the system, guided by the intuition and educated guess. The alternative route was shown, when NNs located the phase transitions for known models without *a priori* physical knowledge [243, 244]. Since then, deep fully-connected and convolutional neural networks (CNNs) have been applied to detect phase transitions in a variety of physical models, for classical [243, 245–249], quantum [244, 250–259], and topological [260–269] phase transitions with supervised [243, 246, 248, 256–259, 262, 270] and unsupervised [244, 247, 249–255, 266, 271, 272] approaches as well as for experimental data. In the last case, NNs were applied to scanning tunneling microscopy images of condensed matter systems [273, 274], neutron scattering data from spin ice systems [275] as well as real and momentum-space images of ultracold atom systems [226, 276–278]. Finally, NNs are not the only ML models used in phase classification problems [279, 280].

Next to all these successful applications, there are open problems, for instance, ones concerning topological models and many-body localization. They include the need for pre-engineered features [260, 264, 281], disagreement of predicted critical exponents [253], and high sensitivity to hyperparameters describing the training process [257]. Moreover, even in the models described by Landau's theory, so far, the ML approaches have mostly enabled only the recovery of known phase diagrams or the location of phase transitions in qualitative agreement with more conventional methods based, for instance, on order parameters or theory of finite-size scaling. Nonetheless, ML achieved this at a much lower computational cost, e.g., using fewer samples or smaller system sizes [253, 257].

Most importantly, however, the resulting models are mostly black boxes, i.e., systems with internal logic that is hidden from a user [282]. The reason is that in complex models, the relevant information is spread over multiple layers each containing a large number of neurons. Because of that, it is difficult to understand, e.g., which data features are the most important or what are reasons for a specific model prediction. As such, we can gain only limited new information from the ML model when applying it to unknown physical systems, as much as limited is our understanding of ML problems with capturing the topological or many-body localization signatures.

1.3.2 Understanding what machines learn

In general, the development of tools to build ML systems has outpaced [283] the growth and adoption of tools to understand what they learn (*interpretability* methods) and whether we can trust their predictions (*reliability* methods).[6] Given the ML presence in everyday life, it is no surprise that the European Parliament has already taken legal measures to assure that any individual can obtain meaningful explanations of the logic involved when automated decision-making takes place [285]. Besides the law, there is ethics. Songul *et al.* revealed that learning machines inherit biases from humans preparing data [286]. Also, deep NNs were shown to perfectly fit random labels [287], and that a group of local features can be their good approximation [288]. These studies prove that the learning process sometimes goes against our intuition, and indicate that the predictions should be accompanied by a justification understandable by humans to be trusted.

Interpretability. The lack of interpretability is now a widely recognized challenge in the computer science community [284, 289–292] especially when ML is applied to real-world problems like medical diagnosis, insurance cost

[6] Note, that the formal definitions of these terms are not agreed upon in either physics or computer science community [284]. Instead, in next paragraphs we provide intuitions behind these terms.

estimation, etc. In science, the lack of interpretability can be disturbing because the black-box behavior of the models prevents us from learning anything about novel physics.

Reliability. Another desired feature of ML models, which is intertwined with interpretability, is their reliability. A reliable ML model informs a user if its decisions are uncertain or result from pure extrapolation [293]. In computer science, the reliability is especially important in the context of safety-critical problems or adversarial attacks, i.e., careful perturbations of input samples that aims to mislead the ML model on the test set [294]. While data sets of physical interest are not endangered by such intentional attacks, adversarial ML tells us that tiny noise in the input may completely derail the model prediction. Moreover, while the reliability of ML in everyday problems can often be controlled by humans checking the predictions, it is improbable for, e.g., unknown phase diagrams. Nevertheless, the reliability of ML is not yet properly addressed in physics. While Bayesian ML, i.e., the most popular approach providing the uncertainty of predictions, proved to be a promising direction in molecular dynamics [295], it is generally difficult and needs a specific model architecture. Therefore, there is a great need for more model-agnostic tools to estimate the uncertainty of ML predictions.

1.3.3 Interpreting machines learning physics

While ML reliability and estimation of uncertainty have not yet been widely explored in physics, the need for interpretation of ML models and extraction of learned quantities has been noticed almost immediately in the community. In particular, Wetzel *et al.* [250, 296] in 2017 have introduced physicists to a new way of understanding ML models that can be viewed as *interpreting bottlenecks*.

The idea behind this approach is to identify possible bottlenecks in the information flow through the ML model and focus our attention there. While the entire NN architecture can be large and have many trainable parameters, the bottleneck itself is described only by few parameters. Because all the relevant information for the predictions of the NN must eventually flow through the bottleneck, we can limit our analysis to the small number of trainable parameters of the bottleneck as opposed to the entire NN. In particular, we can perform a regression on the output of such bottleneck neurons and extract the mapping between the input features and the activations of the bottleneck neurons. Such an approach was applied to the bottlenecks in autoencoders (AEs) [250, 297, 298] as well as shallow convolutional neural networks (CNNs) [296] and Siamese NNs [299]. It succeeded in extracting order parameters learned by ML models in the two-dimensional Ising model [250, 296] and the SU(2) lattice gauge theory [296] but also in

detecting learned symmetry invariants and conserved quantities in various physical problems [297–299]. While extremely successful, this approach has an obvious limitation of requiring bottlenecks that are absent in a majority of modern NNs.

Another fruitful direction has been *analysis of patterns present in filters* (or kernels) of shallow CNNs trained in phase classification problems [277, 300, 301]. When the CNN was trained on snapshots of ultracold atomic simulators, various patterns could be decoded from the filters of the CNN indicating various relevant correlations and therefore phases. In parallel, other studies took advantage of inherent higher interpretability of *kernel methods and decision trees* (as compared to NNs) [248, 249, 279, 302–306].

Therefore, physicists have already stressed the need for interpretation of ML models, and there has been a significant progress in application and development of such approaches. However, proposed methods are predominantly restricted to particular architectures of ML models. When it comes to the most modern and flexible models such as NNs, solutions developed so far are limited to AEs and shallow CNNs.

Within this thesis, we therefore address the lack of universality of known interpretability and reliability tools for machines learning physics. In part II of this thesis, we introduce an interpretability and reliability toolbox that is agnostic of a model architecture and any training details. We describe it in detail in chapter 4 and present its power on the example of NNs trained on quantum phase classification problems in chapter 5. In particular, in section 5.5 we show how this toolbox lead to the fully unsupervised discovery of the phase diagram from experimental topological data.

Part I

Two ultracold molecules in a trap

2

Description of the ultracold molecular system

Sometimes the smallest things
take up the most space in your heart.
~Alan Alexander Milne

In section 1.2.2 we have already introduced promises that are brought by ultracold molecules to the field of quantum simulations. Especially interesting and unexplored is the regime of ultracold molecules with both magnetic and electric dipole moments. In the same section, we have also discussed that the few-body atomic limit brings fascinating questions and challenges. These main motivations have guided us in studying an ultracold system of two interacting molecules in a one-dimensional harmonic trap. Before we describe our results in chapter 3, we devote this chapter to a detailed theoretical description of the system.

Therefore, section 2.1 presents an overview of the studied system and lists our assumptions put on the problem. In section 2.2 we describe the internal structure of our molecular model including rotation and spin-rotation coupling. Section 2.3 discusses the intermolecular interactions and their effective form under the one-dimensional trapping. Finally, in section 2.4, we sum up the theoretical description by providing a full Hamiltonian of a system.

Results presented in chapter 3 have been obtained using exact diagonalization. Therefore, in section 2.5, we discuss details of the exact diagonalization technique and describe the basis set used to build the Hamiltonian matrix. Apart from the spectrum, we also analyze properties like magnetization and quench dynamics of the system. We describe them in section 2.6. In

section 2.7 we briefly mention the convergence of our results with the basis set size. We conclude this chapter by discussing the experimental feasibility of the studied system in section 2.8.

2.1 Overview

Within this thesis, we consider two interacting distinguishable ultracold molecules bound to move along one dimension, chosen to be a z axis, due to the presence of a strong transverse confinement, as schematically shown in fig. 2.1. Along the axial direction (which is the z-direction), the molecules are further confined by a harmonic potential of frequency ω. The molecules are described within the rigid rotor approximation, and they interact by means of a multi-channel two-body potential. We approximate the interaction between molecules with the intermolecular isotropic and anisotropic contact potential. The molecules have the same mass m, magnetic spin s, and electric dipole moment d, and are in the same vibrational state.

We make a series of assumptions allowing to model this system. We assume that the strong transverse confinement does not affect the internal rotation of molecules. This is valid when the size of molecules R_e is much smaller than the transverse harmonic oscillator characteristic length $a_\perp = \sqrt{\hbar/(m\omega_\perp)}$, where m is the mass of molecules and $\omega_\perp$ is the transverse harmonic confinement frequency. At the same time, we assume that the transverse harmonic confinement is much stronger that the axial one ($\omega_\perp \gg \omega$), and no excitation of transverse motion is energetically allowed. Moreover, we assume that the trapping frequency does not depend on the rotational states of molecules. We also assume that the confinement does not affect the short-range intermolecular interaction and dynamics. This is valid when the range of chemical intermolecular interactions R_{vdW} is much smaller than the harmonic oscillator characteristic lengths $a_\perp$ and $a_{\mathrm{ho}} = \sqrt{\hbar/(m\omega)}$. For atomic systems, the effective one-dimensional behavior is observed for elongated harmonic traps with $\omega_\perp/\omega \geq 10$ [56, 307]. We neglect possible losses due to inelastic collisions (vibrational relaxation), or chemical reactions, but they can potentially be incorporated in our model within the complex contact interaction potential formalism [308, 309]. We also discuss such losses in more detail in section 2.8.

2.2 Internal structure

We consider molecules whose rotational behavior is approximated with a rigid quantum rotor model. Next to the rotation, molecules possess magnetic spin (and dipole moments which we discuss in section 2.4). Together

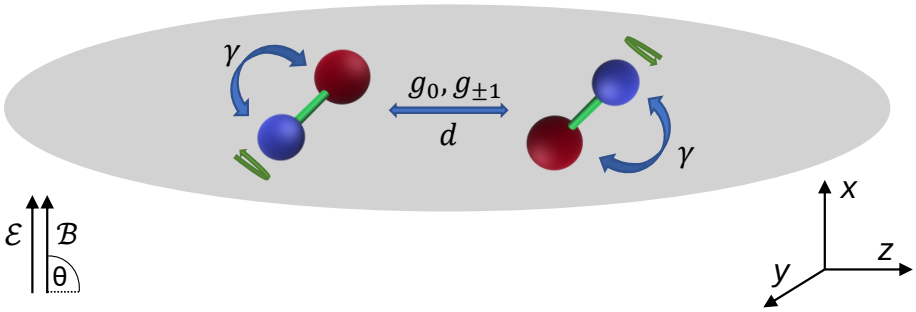

Figure 2.1: **Schematic representation of the investigated system.** Two rotating diatomic polar and magnetic molecules are trapped in a one-dimensional potential. They interact via the short-range isotropic and anisotropic interaction with strengths g_0 and $g_{\pm 1}$ or via an effective one-dimensional electric dipolar interaction of strength d (we neglect the magnetic one). Their internal rotational and magnetic momenta are coupled with the spin-rotation coupling of strength γ. Finally, external electric $\mathcal{E}$ and magnetic $\mathcal{B}$ fields act on the system with an angle θ with the x direction of the system.

they form the internal structure of the studied molecular model. Its Hamiltonian is

$$\hat{H}_{\mathrm{mol}} = \hat{H}_{\mathrm{rot}} + \hat{H}_{\mathrm{spin-rot}} \,, \tag{2.1}$$

where

$$\begin{aligned} \hat{H}_{\mathrm{rot}} &= \sum_{i=1}^{2} B\, \hat{\mathbf{j}}_i^2 \,, \\ \hat{H}_{\mathrm{spin-rot}} &= \sum_{i=1}^{2} \gamma\, \hat{\mathbf{s}}_i \cdot \hat{\mathbf{j}}_i \,. \end{aligned} \tag{2.2}$$

$\hat{H}_{\mathrm{rot}}$ stands for the rotational structure and $\hat{H}_{\mathrm{spin-rot}}$ stands for the spin-rotation coupling. $\hat{\mathbf{j}}_i$ is the i-th molecule's rotational angular momentum operator, B is the rotational constant, and $\hat{\mathbf{s}}_i$ is the i-th molecule's electronic spin angular momentum operator. The molecular spin-rotation interaction with the coupling constant γ is responsible for coupling the molecular intrinsic electric and magnetic dipole moments, because the permanent electric dipole moment is associated with the molecular rotation. We assume the same rotational constants and the same electric and magnetic dipole moments for both molecules. We also assume that the rotational and spin-rotation coupling constants do not depend on the rotational states of molecules. We neglect the nuclear spin in our description, although it can be used to control distinguishability of molecules.

The size of the ratio between the rotational constant, B, and harmonic trapping strength, ω, is an important choice determining how much molecular features dominate the system. Figure 2.2(a) shows the spectrum of the system for increasing ratio between B and ω. In the regime of small B/ω, the trapping levels are separated by a number of rotational levels as presented schematically in panel (b). In such a scenario, the molecular features of the system are exceedingly pronounced, especially when it comes to the interaction anisotropy and reaction to external electric field. Therefore, it poses an informative toy model for analyzing differences between atomic [64] and molecular ultracold trapped two-body systems. We focus on this regime in section 3.1 by choosing $B/\omega = 0.3$. We mark this dense part of the spectrum with a dashed line in fig. 2.2(a).

While the regime of small B/ω is fascinating, we note that typical ultracold molecular experimental set-ups [151, 153, 154] use three-dimensional trap frequencies ranging from around 1 kHz to at most 1 MHz [310], while studied ground-state molecules have rotational constants reaching hundreds of MHz or more [311]. This combination amounts to the ratio of $B/\omega \gg 1$ and results in the rotational levels separated by many harmonic trap states. In such a scenario, the spectra approach atomic solutions and molecular features related to the molecular rotational structure are less important (unless coupled by, e.g., microwave fields).

Therefore, in section 3.2 we turn to the intermediate regime which is closer to the experimental conditions while retaining the predominantly molecular character of the system. We therefore choose the regime of $B/\omega \approx 3$, which amounts to rotational levels separated by a few harmonic levels as presented in fig. 2.2(c). We devote a separate section to discuss the experimental feasibility of this system (see section 2.8).

2.3 Interactions

The Hamiltonian describing the interaction between molecules is

$$\hat{H}_{\text{int}} = \hat{H}_{\text{iso}} + \hat{H}_{\text{aniso}}\,, \tag{2.3}$$

where we distinguish two parts: the isotropic one $\hat{H}_{\text{iso}}$ and anisotropic one $\hat{H}_{\text{aniso}}$. The isotropic part of the interaction potential is of the same nature as the spherically symmetric interaction potential between alkali-metal atoms in the electronic ground state. The anisotropic part of the interaction potential results from the existence of molecular internal structure and orientation dependence of intermolecular interactions [160, 161]. It is responsible for the transfer of the internal rotational angular momenta between

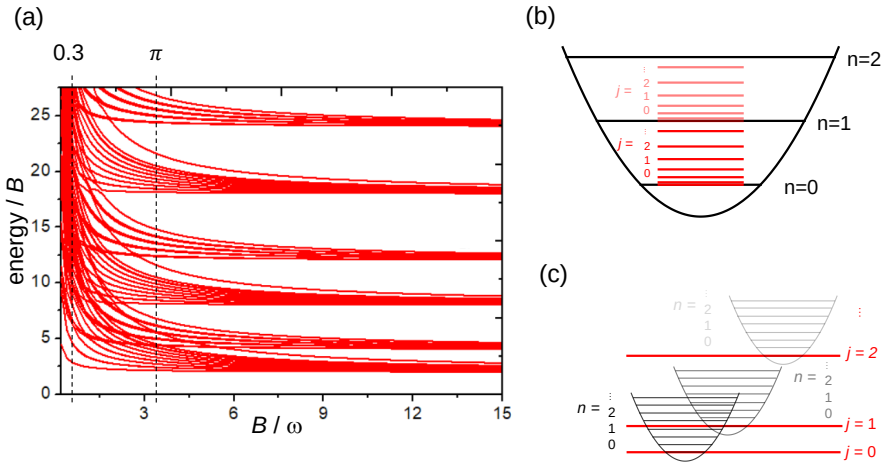

Figure 2.2: **Importance of the rotational constant** that determines the molecular character of the system. (a) Spectrum of a discussed system of two trapped molecules vs the ratio between the rotational constant, B, and the harmonic trapping frequency, ω. Within this thesis, we explore $B/\omega = 0.3$ and $B/\omega = \pi$. Corresponding rotational and trapping levels are plotted schematically in (b) and (c), respectively.

interacting molecules. Neglecting the anisotropic part of the interaction potential restores results known for two interacting atoms in a harmonic trap (apart from the additional states coming from the rotational structure of molecules).

The intermolecular interaction potential may in general depend on all the internal degrees of freedom and the relative orientation of interacting molecules. This dependence is responsible for effective mixing and exchanging different angular momenta present in interacting molecules and the relative motion during molecular collisions [160, 161, 312]. Within this thesis we propose and employ a model multichannel two-body contact interaction potential to account effectively for the coupling of molecular rotational angular momenta during ultracold molecular collisions. The contact interaction potential is commonly used in ultracold physics as a successful approximation to describe atom-atom interactions of the short-range van der Waals character [313]. Similar performance of this approximation can be expected while applied to anisotropic intermolecular interactions of the short-range van der Waals nature. Additionally, in reduced dimensionality, even long-range interactions can be approximated effectively by short-range ones [314].

The isotropic part of the intermolecular interaction potential is

$$\hat{H}_{\text{iso}} = \sum_{\alpha} g_{\alpha} \delta(z_1 - z_2) \hat{P}_{\alpha} \,, \tag{2.4}$$

where g_α is the strength of the isotropic interaction for channel α, $\delta(z)$ is the Dirac delta function imposing assumed contact-type interaction, and $\hat{P}_\alpha$ is the projection operator

$$\hat{P}_{\alpha} = |\alpha\rangle\langle\alpha| \,, \tag{2.5}$$

where $|\alpha\rangle$ is a basis set function (channel) describing all degrees of freedom of the system except for the intermolecular distance.

The anisotropic part of the intermolecular interaction potential is

$$\hat{H}_{\text{aniso}} = \sum_{\alpha \neq \alpha'} g_{\alpha\alpha'} \delta(z_1 - z_2) \hat{P}_{\alpha\alpha'} \,, \tag{2.6}$$

where $g_{\alpha\alpha'}$ is the strength of the anisotropic interaction between channels α and α', and $\hat{P}_{\alpha\alpha'}$ is of the form

$$\hat{P}_{\alpha\alpha'} = |\alpha\rangle\langle\alpha'| + |\alpha'\rangle\langle\alpha| \,. \tag{2.7}$$

Different models of intermolecular interactions can be represented by imposing different forms of g_α and $g_{\alpha\alpha'}$.

The isotropic and anisotropic interaction strengths g_α and $g_{\alpha\alpha'}$ result from specific values of scattering lengths and chemical short-range intermolecular interaction potentials. In principle, they can be controlled by means of magnetic or optical Feshbach resonances, as well as by changing chemical composition or vibrational states of molecules.

The general form of the interaction potential of eq. (2.3) can capture all types of intermolecular interactions. Nevertheless, due to the special importance of the intermolecular dipole-dipole interaction in ultracold physics, we consider this interaction separately. In three dimensions, it is described by the following Hamiltonian

$$\hat{H}^{\text{3D}}_{\text{dip}} = \frac{\hat{\mathbf{d}}_1 \cdot \hat{\mathbf{d}}_2 - 3(\hat{\mathbf{d}}_1 \cdot \mathbf{e}_r)(\hat{\mathbf{d}}_2 \cdot \mathbf{e}_r)}{r^3} \,, \tag{2.8}$$

where r and $\mathbf{e}_r$ are the distance and versor connecting two molecules, respectively. By restricting the motion to one dimension, $\mathbf{e}_r = \mathbf{e}_z$, and assuming the contact form of the interaction, $1/z^3 \to \delta(z)$, which is the exact result for polarized dipoles in one dimension [315, 316], we arrive at the effective Hamiltonian, which we use in the thesis

$$\hat{H}_{\text{dip}} = -\delta(z_1 - z_2) \left(2\hat{d}_{1,0}\hat{d}_{2,0} + \hat{d}_{1,1}\hat{d}_{2,-1} + \hat{d}_{1,-1}\hat{d}_{2,1} \right) \,, \tag{2.9}$$

where $\hat{d}_{i,q} \equiv \mathbf{e}_q \cdot \hat{\mathbf{d}}_i$ are the spherical components of the projection of the dipole operator of the i-th molecule onto the space-fixed frame $\mathbf{e}_q$ in spherical coordinates. We approximate the long-range character of the dipole-dipole interaction by the contact potential to simplify our model and analysis. This interaction limited to one dimension does not conserve the total rotational angular momentum.

2.4 Hamiltonians

We here provide the Hamiltonian describing the studied system as well as remind some of the assumptions listed in section 2.1. The generic Hamiltonian describing two interacting polar and paramagnetic molecules in a one-dimensional trap is of the form

$$\hat{H} = \hat{H}_{\text{trap}} + \hat{H}_{\text{mol}} + \hat{H}_{\text{field}} + \hat{H}_{\text{int}} \,, \tag{2.10}$$

where $\hat{H}_{\text{trap}}$ describes the motion of molecules in a trap, $\hat{H}_{\text{mol}}$ describes the internal (rotational and spin) structure of molecules discussed in section 2.2, $\hat{H}_{\text{field}}$ describes the interaction of molecules with external fields, and $\hat{H}_{\text{int}}$ describes the interaction between molecules discussed in section 2.3.

The Hamiltonian describing two structureless particles in a one-dimensional harmonic trap is

$$\hat{H}_{\text{trap}} = \sum_{i=1}^{2} \frac{\hat{p}_i^2}{2m} + \sum_{i=1}^{2} \frac{1}{2} m\omega z_i^2 \,, \tag{2.11}$$

where $\hat{p}_i$ and z_i are the linear momentum and position of the i-th particle, respectively, m is their mass, and ω is the trapping frequency. We assume that two molecules have the same mass (consisting of the same atoms) and are in the same vibrational state. We also assume that the trapping frequency does not depend on the rotational states of molecules.

We have already discussed the Hamiltonian describing the internal structure of two studied molecules in section 2.2 but for the sake of completeness we put the Hamiltonian here again:

$$\hat{H}_{\text{mol}} = \hat{H}_{\text{rot}} + \hat{H}_{\text{spin-rot}} \,, \tag{2.12}$$

where

$$\begin{aligned} \hat{H}_{\text{rot}} &= \sum_{i=1}^{2} B\, \hat{\mathbf{j}}_i^2 \,, \\ \hat{H}_{\text{spin-rot}} &= \sum_{i=1}^{2} \gamma\, \hat{\mathbf{s}}_i \cdot \hat{\mathbf{j}}_i \,. \end{aligned} \tag{2.13}$$

Finally, the Hamiltonian describing the interaction with external electric and magnetic fields is

$$\hat{H}_{\text{field}} = \hat{H}_{\text{Stark}} + \hat{H}_{\text{Zeeman}}\,, \tag{2.14}$$

where

$$\begin{aligned} \hat{H}_{\text{Stark}} &= -\sum_{i=1}^{2} \hat{\mathbf{d}}_i \cdot \mathcal{E}\,, \\ \hat{H}_{\text{Zeeman}} &= 2\mu_B \sum_{i=1}^{2} \hat{\mathbf{s}}_i \cdot \mathcal{B}\,. \end{aligned} \tag{2.15}$$

$\hat{\mathbf{d}}_i$ is the i-th molecule's electric dipole moment operator and $\mathcal{E}$ and $\mathcal{B}$ are the electric and magnetic fields, which couple with the molecules' electric and magnetic dipole moments, respectively. $\hat{H}_{\text{Stark}}$ results in the Stark effect and $\hat{H}_{\text{Zeeman}}$ results in the Zeeman effect. We assume that the electric and magnetic fields are parallel to each other and parallel or perpendicular to the motion of molecules in the trap.

2.5 Exact diagonalization and the basis states

It is possible to separate the center-of-mass and relative motions in the Hamiltonian of eq. (2.10) by introducing new coordinates $Z = \frac{1}{\sqrt{2}}(z_1 + z_2)$ and $z = \frac{1}{\sqrt{2}}(z_1 - z_2)$, and related momenta $\hat{P}$ and $\hat{p}$. Thanks to unconventional factors of $\sqrt{2}$, the effective masses for both types of motion are the same ($M = \mu = m$). The total wave function in new coordinates $|\Phi(Z,z)\rangle = |\varphi_{\text{CM}}(Z)\rangle|\Psi_{\text{rel}}(z)\rangle$ is a product of the wave function for the center-of-mass motion $|\varphi_{\text{CM}}(Z)\rangle$, which is an eigenstate of the quantum harmonic oscillator Hamiltonian

$$\hat{H}_{\text{CM}} = \frac{\hat{P}^2}{2m} + \frac{1}{2}m\omega Z^2\,, \tag{2.16}$$

and the wave function for the relative motion $|\Psi_{\text{rel}}(z)\rangle$, which is a solution of the Schrödinger equation with the following Hamiltonian

$$\hat{H}_{\text{rel}} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega z^2 + \hat{H}_{\text{mol}} + \hat{H}_{\text{field}} + \frac{1}{\sqrt{2}}\hat{H}_{\text{int}}\,. \tag{2.17}$$

In the rest of this thesis we focus on finding eigenstates of the above Hamiltonian, therefore whenever we refer to spectra or wave functions of the system we mean properties related to the relative motion. For convenience, we use

units of energy and interaction strength that correspond to $\omega = m = \hbar = 1$. This amounts to measuring energies E in units of $\hbar\omega$, lengths in units of the harmonic oscillator characteristic length $a_{\mathrm{ho}} = \sqrt{\hbar/(m\omega)}$, and the interaction strengths g_α and $g_{\alpha\alpha'}$ in units of $\hbar\omega a_{\mathrm{ho}}$.

For two polar molecules without spin, we represent the relative motion part of the total wave function in the basis of eigenstates of the one-dimensional harmonic oscillator $|n\rangle$ with known behavior when acted upon with $\hat{H}_{\mathrm{trap}}$:

$$(\frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega z^2)|n\rangle = (n + \frac{1}{2})\hbar\omega|n\rangle\,, \tag{2.18}$$

The basis set also contains eigenstates of the total rotational angular momentum operator $|J, M, j_1, j_2\rangle$, which fulfil the following conditions:

$$\begin{aligned} \hat{\mathbf{J}}^2|J, M, j_1, j_2\rangle &= J(J+1)|J, M, j_1, j_2\rangle\,, \\ \hat{\mathbf{J}}_z|J, M, j_1, j_2\rangle &= M|J, M, j_1, j_2\rangle\,, \hat{\mathbf{J}} = \hat{\mathbf{j}}_1 + \hat{\mathbf{j}}_2\,. \end{aligned} \tag{2.19}$$

Therefore, the full wave function is:

$$|\Psi_k\rangle = \sum_{n,J,M,j_1,j_2} C^k_{n,J,M,j_1,j_2}|n\rangle|J, M, j_1, j_2\rangle\,, \tag{2.20}$$

where

$$|J, M, j_1, j_2\rangle = \sum_{m_1,m_2} \langle j_1, m_1, j_2, m_2|J, M\rangle|j_1, m_1\rangle|j_2, m_2\rangle\,, \tag{2.21}$$

where $\langle j_1, m_1, j_2, m_2|J, M\rangle$ are Clebsch-Gordan coefficients and $|j_i, m_i\rangle$ are eigenstates of $\hat{\mathbf{j}}_i$. The symmetries of the Hamiltonian of eq. (2.17) resulting in the conservation of J or M quantum numbers are used to restrict properly the size of the employed basis set. All possible combinations of basis functions with $n \leq n_{\max}$, $j_1 \leq j_{\max}$, and $j_2 \leq j_{\max}$ are employed in calculations. Numerical coefficients C^k_{n,J,M,j_1,j_2} for the k-th state are calculated by means of the exact diagonalization method.

Having the basis set discussed, we can look in more detail into the multichannel interaction model. In the computational basis set (channels) introduced in eqs. (2.20) and (2.21), the projection operator $\hat{P}_\alpha$ in the isotropic part of the intermolecular interaction potential of eq. (2.4) takes the form

$$\hat{P}_\alpha = |J, M, j_1, j_2\rangle\langle J, M, j_1, j_2|\,. \tag{2.22}$$

We assume that the corresponding strengths of the isotropic interaction are independent of molecular internal states and the same for all channels

$$g_\alpha = g_0\,. \tag{2.23}$$

The coupling operator $\hat{P}_{\alpha\alpha'}$ in the anisotropic part of the intermolecular interaction potential of eq. (2.5) is of the form

$$\hat{P}_{\alpha\alpha'} = |J, M, j_1, j_2\rangle\langle J, M, j_1', j_2'| + \text{H.c.}\,, \tag{2.24}$$

where we assume that the anisotropic part of the intermolecular interaction potential does conserve the total rotational angular momentum of two molecules J, M. The corresponding strength of the anisotropic interaction can be written as

$$g_{\alpha\alpha'} \equiv g_{JMj_1j_2,JMj_1'j_2'}\,. \tag{2.25}$$

We assume that the anisotropic interaction strengths do not depend on the total rotational angular momentum J, M

$$g_{JMj_1j_2,JMj_1'j_2'} = g_{j_1j_2,j_1'j_2'}\,, \tag{2.26}$$

and consider two types of the anisotropic interaction. The first one couples molecular states which differ by k quanta of molecular rotational angular momentum

$$g^{\pm k}_{j_1j_2,j_1'j_2'} = \delta_{j_1,j_1'\pm k}\delta_{j_2,j_2'\mp k}\, g_{\pm k}\,. \tag{2.27}$$

In this notation, the dipole-dipole interaction has non-zero terms only related to g_0 and $g_{\pm 2}$. The second, simplified type allows us only to exchange the rotational angular momenta between two interacting molecules if they differ by $k = |j_1 - j_2|$

$$g^{\pm k,\text{ex}}_{j_1j_2,j_1'j_2'} = \delta_{j_1,j_2'}\delta_{j_2,j_1'}\delta_{j_1\pm k,j_2}\, g^{\text{ex}}_{\pm k}\,. \tag{2.28}$$

The coefficients $g_{\pm k}$ and $g^{\text{ex}}_{\pm k}$ can depend on k. In our model calculations we assume that these coefficients are the largest for $k = 1$ and we neglect them for $k > 1$ or assume geometric decay with k, $g_{\pm k} = g_{\pm 1}/A^{k-1}$.

When we consider molecules with spin, the basis set used in eq. (2.20) is augmented by eigenstates of the total electronic spin angular momentum operator $\hat{\mathbf{S}}$ denoted as $|S, M_S, s_1, s_2\rangle$ resulting in the final computational basis of the form

$$|n\rangle|J, M_J, j_1, j_2\rangle|S, M_S, s_1, s_2\rangle\,. \tag{2.29}$$

All spin configurations allowed by the symmetry are included in the basis set. We assume that the intermolecular interaction potential of eq. (2.3) does not depend on the electronic spin.

The mentioned throughout this section total angular momenta are the sums of the angular momenta of individual molecules, $\hat{\mathbf{J}} = \hat{\mathbf{j}}_1 + \hat{\mathbf{j}}_2$ and $\hat{\mathbf{S}} = \hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2$. The total angular momentum of the system is then the sum of

the total rotational and spin angular momenta $\hat{\mathbf{J}}_{\text{tot}} = \hat{\mathbf{J}} + \hat{\mathbf{S}}$, and its projection $M_{\text{tot}} = M_J + M_S$ is a sum of projections of the total rotational M_J and spin M_S angular momenta. Matrix elements of the system's Hamiltonian in the employed basis set are provided in appendix A. We discuss the convergence of results with the basis set size in section 2.7. The diagonalization of molecular Hamiltonian matrices has been performed within this thesis with Intel Math Kernel Library and supported by PL-Grid Infrastructure.

2.6 Magnetization and quench dynamics

In section 3.2, instead of spectrum, we focus on the system magnetization and its quench dynamics. The magnetization $\langle \hat{S}_z \rangle$ is an expectation value of the z-component of the total electronic spin operator. We calculate it for the several lowest eigenstates. We also analyze the nonequilibrium dynamics of the system after the quench.

The quench dynamics experiments [317–319] may allow to probe the internal parameters of the Hamiltonian governing the analyzed system. The quench dynamics has been thoroughly studied for two [320–323] and three [317] ultracold atoms in a trap. Quantum quenches also allow one to study nonequilibrium dynamics [324–326]. In a quench scenario, the system prepared initially in the chosen state $|\Psi\rangle$ (e.g. the ground state) of a Hamiltonian $\hat{H}_{\text{ini}}$, evolves unitarily in time following the sudden change (quench) of the parameters to a final Hamiltonian $\hat{H}_{\text{fin}}$ [327]. This dynamics can be expressed in terms of overlaps of the initial eigenstate $|\Psi\rangle$ of $\hat{H}_{\text{ini}}$ with the eigenstates $\left|\tilde{\Psi}_j\right\rangle$ of $\hat{H}_{\text{fin}}$:

$$|\Psi(t)\rangle = e^{-i\hat{H}_{\text{fin}}t} |\Psi(0)\rangle = \sum_j \left\langle \tilde{\Psi}_j \middle| \Psi(0) \right\rangle e^{-iE_j t} \left| \tilde{\Psi}_j \right\rangle . \tag{2.30}$$

The time evolution of any observable $\hat{O}$ can be then described as:

$$\begin{aligned}
&\langle \Psi(t)| \hat{O} |\Psi(t)\rangle = \\
&= \sum_{j,j'} \left\langle \tilde{\Psi}_j \middle| \Psi(0) \right\rangle \left\langle \tilde{\Psi}'_j \middle| \Psi(0) \right\rangle e^{-i(E_j - E_{j'})t} \left\langle \tilde{\Psi}_{j'} \middle| \hat{O} \middle| \tilde{\Psi}_j \right\rangle = \\
&= \sum_j \left| \left\langle \tilde{\Psi}_j \middle| \Psi(0) \right\rangle \right|^2 \left\langle \tilde{\Psi}_j \middle| \hat{O} \middle| \tilde{\Psi}_j \right\rangle + \\
&+ 2 \sum_{j<j'} \left\langle \tilde{\Psi}_j \middle| \Psi(0) \right\rangle \left\langle \tilde{\Psi}_{j'} \middle| \Psi(0) \right\rangle \cos\left[(E_j - E_{j'})t\right] \left\langle \tilde{\Psi}_{j'} \middle| \hat{O} \middle| \tilde{\Psi}_j \right\rangle .
\end{aligned}$$

We choose to study the evolution of two observables: magnetization $\langle \hat{S}_z \rangle$ of the system and the cloud size $\langle \hat{r}^2 \rangle$. The formula for the cloud size is provided in appendix B. The dynamics is calculated till time $t = 10\,000 \frac{2\pi}{\omega}$ with

a time step of $0.1\frac{2\pi}{\omega}$. Elongating the dynamics calculations provides no additional frequency peaks of amplitudes larger than 10^{-4} in the corresponding discrete Fourier transforms (DFT). We perform the Fourier transform of the observables' evolution using SciPy package [328].

2.7 Convergance with the basis set size

In chapter 3 we focus on three main properties of the described system. In section 3.1 we analyze spectra as functions of various internal or external Hamiltonian parameters. There, in a majority of calculations, we calculate spectra using the exact diagonalization method with the basis set composed of quantum harmonic oscillator eigenfunctions up to $n_{\max} = 30$, and quantum rigid rotor eigenfunctions up to $j_{\max} = 8$. The exceptions is diagonalization of those Hamiltonians that include the spin momenta of the molecules. Such spectra are studied in section 3.1.4. There the basis set is $n_{\max} = 30$ and $j_{\max} = 5$. Without using mentioned symmetries resulting in momenta conservation, the size of the Hamiltonian matrices would be $10^5 - 10^6$. If symmetries are employed, the size of the Hamiltonian matrices to be diagonalized is between around 10^3 and $2 \cdot 10^4$, depending on the angular momentum and the presence of external fields and spin structure. Grining *et al.* [75, 76] showed that the convergence of correlated energy calculations for interacting particles in one dimension with the number of used single-particle harmonic oscillator functions is relatively slow and asymptotically proportional to $\frac{1}{\sqrt{n_{\max}}}$ Nevertheless, for the range of weak and intermediate interaction strengths, physically meaningful results can be obtained with used $n_{\max}$. Additionally, in the present case, we have checked that the convergence of energy calculations with the size of the rotational basis set is much faster and obtained results are close to converged with respect to $j_{\max}$ [329].

In section 3.2 we focus on the system magnetization and its quench dynamics. Following the results from chapter 3, we limit the basis set size for efficient calculations of the system's dynamics to $n_{\max} = 20$ and $j_{\max} = 4$. Again, we acknowledge the slow convergence with $n_{\max}$ but in the analyzed system $n_{\max} = 20$ provides a satisfying convergence of the lowest-energy states, which are the main focus of section 3.2. The convergence of the cloud-size time-evolution calculations with $n_{\max}$, however, is problematic. On the one hand, the mean value of the cloud size, $\langle n| \hat{r}^2 |n'\rangle$, calculated for two harmonic oscillator eigenfunctions, increases rapidly with the harmonic oscillator levels, n, becoming divergent for large n (see appendix B). In the quench dynamics, this divergence is faster than the decrease of the overlap between the ground state and the highly excited states leading to the

divergence of the cloud-size excitation. However, this nonphysical behavior can be neglected by restricting the basis sets size, knowing that all realistic traps have a finite size, and all realistic quenches have a finite time. On the other hand, the highest-energy eigenstates in a finite basis set are not converged [64, 76] and a nonphysically larger occupation of the highest-energy eigenstate can be observed. We solve this problem by neglecting unconverged part of spectrum from the quench dynamics calculations, namely by removing eigenfunctions $\left|\tilde{\Psi}_j\right\rangle$ of $\hat{H}_{\text{fin}}$ with the contribution from any basis state with $|n_{\max}\rangle$ larger than 10%. The removed part of eigenfunctions is up to a few percent of the whole spectrum.

According to our previous convergence analysis [329], selected $j_{\max}$ already enables convergence for the significant part of the spectrum as well as good convergence of the quench dynamics.

2.8 Experimental feasibility

We have already discussed in section 2.2 that the smaller the ratio of the molecular rotational constants to the harmonic trap frequency, B/ω, the more pronounced the analyzed system's molecular character [330]. However, typical ultracold molecular experimental set-ups [151, 153, 154] use three-dimensional trap frequencies ranging from around 1 kHz to at most 1 MHz [310], while studied ground-state molecules have rotational constants reaching hundreds of MHz or more [311]. This combination amounts to the ratio of $B/\omega \gg 1$ and results in the rotational levels separated by many harmonic trap states. In such a scenario, the molecular features related to the molecular rotational structure are less important.

Within this thesis, we study two regimes of B/ω equal to 0.3 and ≈ 3. While the first regime has been chosen to deliberately enhance molecular features in the system and can be treated as an informative toy model, the regime of $B/\omega \approx 3$ can be reached with tight traps (e.g., a nanoplasmonic one [332]) and weakly-bound molecules (e.g., Feshbach molecules [148]) which have rotational constants up to few MHz.

Such a proposed realization needs an additional discussion. Within this work, we also assume that the studied molecules possess electric dipole moment, while electric dipole moments of Feshbach molecules may be vanishingly small, as they scale asymptotically as R^{-7} with the internuclear separation R [333]. Figure 2.3 presents the dependence of the rotational constant and permanent electric dipole moment on the mean distance between atoms in the ^{87}Rb^{133}Cs molecule in the lowest electronic state with non-zero spin, i.e., $a^3\Sigma$. The choice of this species is just exemplary as we expect other classes of weakly-bound molecules to have similar charac-

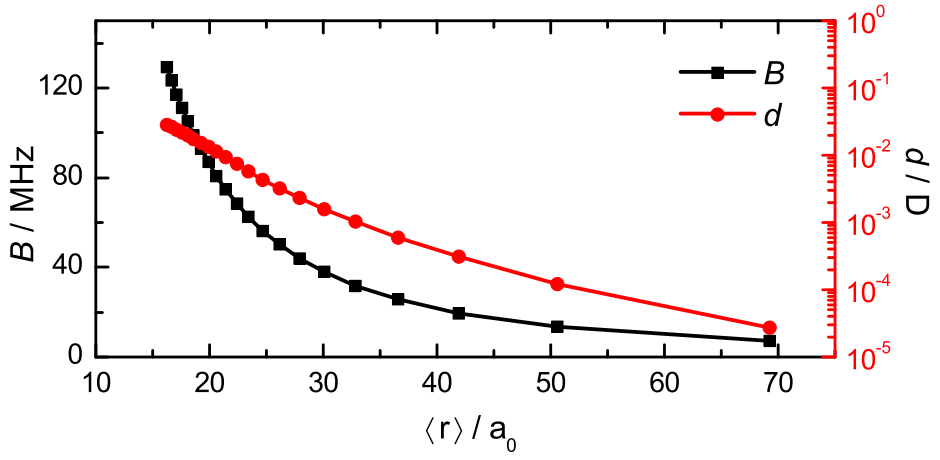

Figure 2.3: **Scaling of molecular properties with the interatomic distance.** The rotational constant, B, and estimated electric dipole moment, d, of the exemplary $^{87}Rb^{133}Cs$ molecule in the lowest triplet state, as functions of the mean distance between atoms $\langle r \rangle$, corresponding to different vibrational levels. Similar behavior may be expected for other weakly-bound molecules. From Ref. [331].

teristics. One of the highest vibrational levels of $a^3\Sigma\,^{87}Rb^{133}Cs$, $v = 39$, corresponds to the mean distance between atoms equal to 51 bohr. We estimate the corresponding electric dipole moment of 10^{-4} D and rotational constant of 13.5 MHz. To reach the B/ω ratio of the order of the selected one, the trap frequency needs to be at least around 0.5 MHz. Note that the upper limit for ω is set also by the size of two molecules, which increases for weakly-bound states. The choice of the trap frequency determines the time scale used within this work. For example, the time evolution plotted in figs. 3.18 and 3.19 takes $200\,(2\pi/\omega) \approx 2.5$ ms for $\omega = 0.5$ MHz. Finally, to reach the effective one-dimensional behavior, elongated harmonic traps with $\omega_\perp/\omega \geq 10$ could be used [56, 307].

Within our work, we also analyze the time evolution of the cloud size after the quench of the intermolecular interaction. The sudden change of the intermolecular interaction strength can be achieved via the change of the magnetic field strength and related Feshbach resonances, the change of the molecular vibrational state, or the change of the trapping frequency. However, the selective quench of intermolecular interaction is impossible to achieve, as, in reality, all molecular characteristics are intimately connected that leads to emergent behavior and challenging description. For example, changing the vibrational state impacts the interaction but also the polarization of molecules that modifies their response to external fields. The

quench of the trapping frequency should correspond to the most uncorrelated change of the intermolecular interaction [317]. Regarding the observables whose time evolution we study, the internal state of molecules can be probed with quantum gas microscopy [23, 24, 27], and the cloud size can be measured via destructive time-of-flight experiments as realized for ultracold atoms [334–336]. We describe the time-of-flight imaging in more detail in section 5.4.2.

Non-reactive trapped alkali dimers in the lowest rovibrational states have been recently shown to form four-atom complexes that are long-lived in the dark but are prone to decay under the trapping field that results in losses [179, 337, 338] as we have discussed in more detail in section 1.2.3. The lifetime of such complexes is proportional to the density of states at the collision threshold. The density of states of weakly-bound molecules, compared to deeply-bound, can be up to eight orders of magnitude smaller [177]. Therefore, the weakly-bound molecules may be less prone to such losses. Additionally, vibrationally-excited molecules may undergo reactive collisions and other decoherence or loss processes may occur, whose detailed characterization is, however, out of the scope of the present study. Search for the molecules which exhibit low losses in a trap is a challenging and impactful task. They probably should be lighter and have a less dense spectrum of electronic states such as AlF [339]. Regarding the non-zero or large electronic spin, it may be realized with alkaline-earth-metal fluoride molecules in the doublet $^2\Sigma^+$ electronic state [340], alkali-metal molecules in the triplet $^3\Sigma^+$ electronic state [189, 341], or molecules containing highly-magnetic atom [342, 343], respectively. Thus, the considered system may potentially be realized in state-of-the-art experiments on ultracold molecules trapped in optical tweezers [165, 166]. However, exact experimental conditions have to be yet carefully researched and designed.

We have already mentioned that weakly-bound molecules may be less prone to the inelastic losses. Additionally, there are shielding mechanisms that may be employed to prevent such losses which we have described in section 1.2.3. They result in keeping molecules at the safe distance preventing complex formation. With regards to our work we note that while this shielding is extremely promising in terms of creating dense ultracold polar molecular gases, as it keeps long-distance interactions mainly intact, the shielding prevents interactions at the short-range distance which are the main focus of our one-dimensional study.

Finally, the studied system of two molecules in a trap has recently found its close experimental realization in Ref. [188] which probed three-dimensional collisions by merging two optical tweezers, each containing a single molecule CaF. Moreover, the same experiment involved the microwave shielding that suppresses the inelastic loss rate by a factor of six.

More ground-breaking experimental tools and realizations are surely on the way!

3

Molecular path - results

The yellow Lego was brick-shaped again.
Pretending innocence.
~William Gibson

While ultracold atoms in lattices has already proven to be powerful and highly-tunable quantum simulators, we expect much richer physics if we replace atoms with molecules. This motivation guides us in choosing the theoretical system of interest that we study in this chapter: two ultracold interacting molecules in a one-dimensional harmonic trap. In particular, we consider it as a building block of a molecular Hubbard Hamiltonian. The studied system is also interesting as a toy model highlighting features introduced by molecules to ultracold systems, therefore we often compare our results against the seminal work of Busch *et al.* [64] regarding two ultracold atoms in a harmonic trap. Another motivation, guiding the second part of our work, has been the unexplored regime of highly magnetic and polar ultracold molecules like DyK or ErLi. Finally, we have already seen in section 1.2 that few-body limits is full of interesting phenomena on its own.

Therefore, in this chapter, we present a detailed analysis of the properties of two interacting ultracold polar molecules in a one-dimensional trap. In particular, section 3.1 contains a thorough study of the molecular spectra as functions of various internal and external parameters. We study the system in the regime of small B/ω ratio, resulting in an exaggerated impact of molecular features on the system. On the other hand, in section 3.2, we focus on magnetic properties and quench dynamics of the system. Here, we study a B/ω ratio that preserves pronounced molecular character of the

system while being experimentally realizable.

3.1 Two ultracold molecules in a one-dimensional harmonic trap

This section is based on our work presented in Ref. [330], where the Ph.D. candidate was responsible for: writing down the Hamiltonian, full numeric implementation (construction of Hamiltonian matrices and exact diagonalization), analysis and proposal of result interpretation, preparation of plots, participation in writing the manuscript. We present here the unexplored so far regime of two trapped molecules with chemically-inspired short-range interactions and compare it against atomic results to highlight what features can we expect in molecular quantum simulators. The code and data that enable the recovery of results in this section are provided in Ref. [344].

Here we investigate a fundamental building block of a molecular quantum simulator, that is, two interacting polar molecules effectively trapped in a one-dimensional harmonic potential. We have described our model in detail in chapter 2. We only remind the reader that in this section we study the regime of small B/ω ratio, resulting in an exaggerated impact of molecular features on the system as presented schematically in fig. 3.1.

We analyze in detail the properties of such a system including the interplay of the molecular rotational structure, anisotropic interactions, spin-rotation coupling, external electric and magnetic fields, and harmonic trapping potential. Energy spectra and eigenstates are calculated by means of the exact diagonalization. Our calculations may be considered as a microscopic model for the on-site interaction of the molecular multichannel Hubbard Hamiltonian [345–347] and may provide underlying parameters for effective molecular many-body Hamiltonians.

We show that the anisotropic intermolecular interaction brings states with higher total rotational angular momenta to lower energies such that the absolute ground state of the molecular system can have total angular momentum larger than zero and be degenerate. Such systems may potentially be useful for realizing quantum simulators of exotic spin models. A strong anisotropic intermolecular interaction can induce the emergence of the molecular equivalent of the atomic super-Tonks-Girardeau limit but, at the same time, the importance of the anisotropic intermolecular interaction is reduced in the limit of a very strong isotropic interaction. Magnetic and electric fields induce a high density of states and a large number of avoided

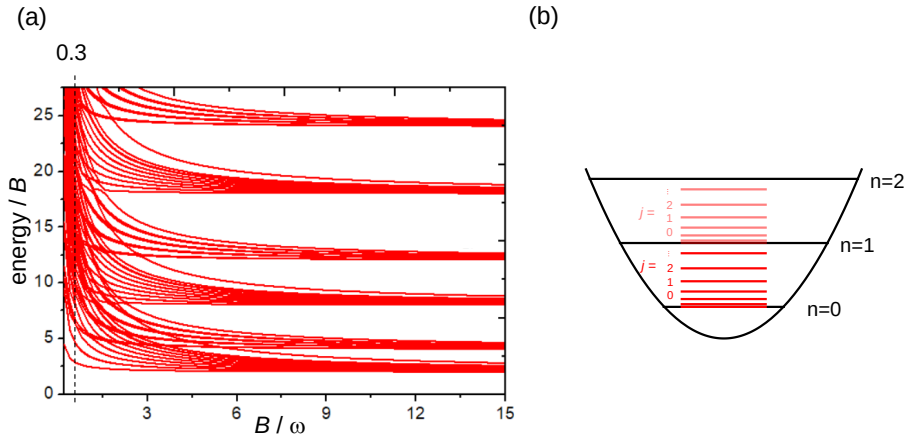

Figure 3.1: **Small rotational constant** as compared to the trapping frequency, $B/\omega = 0.3$, results in trapping levels being separated by several rotational levels and the enhanced impact of molecular features on the system.

crossings, which can be used to control a system's properties. No signatures of quantum chaotic behavior in energy spectra are, however, found.

We begin in section 3.1.1 by studying the impact of short-range isotropic and anisotropic interaction on spectra. Next, in section 3.1.2, we study the impact of external electric field and its interplay with the anisotropic short-range interaction. Section 3.1.3 discusses molecular spectra as functions of effective one-dimensional electric dipole-dipole interaction strength. Finally, in section 3.1.4 we discuss the interplay of the spin-rotation coupling and external magnetic field. We summarize our results and discuss the outlook in section 3.1.5.

3.1.1 Short-range anisotropy of intermolecular interaction

Before we focus on systems with small rotational constants $B \leq \omega$, which are the main subject of this section, we analyze the impact of the anisotropic interaction on systems with $B \gg \omega$, where the effect of the anisotropic interaction is relatively smaller, but easier to interpret. Figure 3.2 presents energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 10\,\hbar\omega$ in a one-dimensional harmonic trap. Results for three total angular momenta $J = 1, 2$, and 3 are presented as a function of the isotropic interaction strength g_0 with the anisotropic interaction strength $g_{\pm 1} = 2$ in panel (a) and as a function of the anisotropic inter-

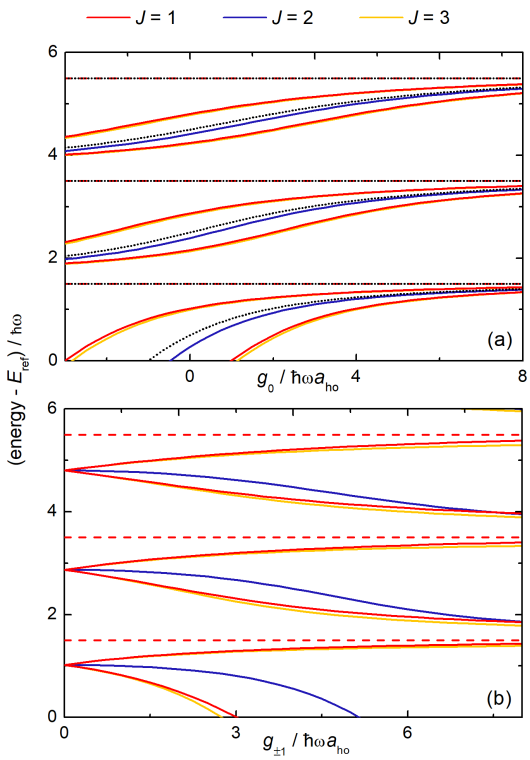

Figure 3.2: **Rotational splitting of the molecular spectra compared to atoms.** Energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 10\,\hbar\omega$ in a one-dimensional harmonic trap: (a) as a function of the isotropic interaction strength g_0 with the anisotropic interaction strength $g_{\pm1} = 2$ and (b) as a function of the anisotropic interaction strength $g_{\pm1}$ with the isotropic interaction strength $g_0 = 2$. The spectra for different total angular momentum J are shifted by the energy of non-interacting systems with this total angular momentum. Solid and dashed lines are for states of even and odd spatial symmetries, respectively. Dotted lines in panel (a) are the result for two interacting atoms. From Ref. [330].

action strength $g_{\pm1}$ with the isotropic interaction strength $g_0 = 2$ in panel (b). Energy spectra are compared with the known result for two interacting atoms [64], that is equivalent to the energy spectrum for interacting molecules with $g_{\pm1} = 0$ or with $J = 0$. In panel (a), energies of states with $J = 2$ are only slightly shifted as compared with the atomic case due to the coupling by the anisotropic interaction with higher-energy states. Instead, there are two energy states for each branch for $J = 1$ and 3. They

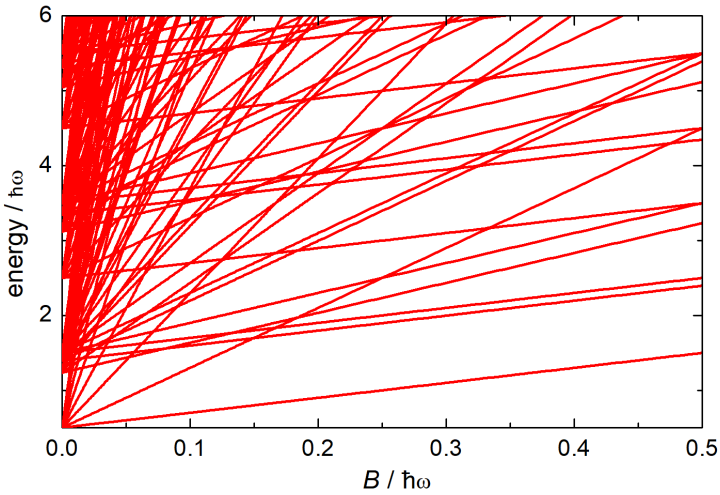

Figure 3.3: **Rotational constant vs spectrum.** Energy spectrum of the relative motion for two interacting molecules in a one-dimensional harmonic trap as a function of their rotational constant B. The isotropic and anisotropic interaction strengths are set at $g_0 = g_{\pm 1} = 4$. From Ref. [330].

originate from the fact that, in the presented energy range, those total angular momenta can be constructed from two rotational configurations with $j_1 = 1, j_2 = 0$ and $j_1 = 0, j_2 = 1$ for $J = 1$, and $j_1 = 2, j_2 = 1$ and $j_1 = 1, j_2 = 2$ for $J = 3$, which are coupled by the anisotropic interaction, whereas the lowest state with $J = 2$ originates from a single rotational configuration with $j_1 = 1, j_2 = 1$. The emergence of the splitting between two states for $J = 1$ and 3, and the shift for $J = 2$ as a function of the anisotropic interaction strength, are presented in panel (b). The energy spectra for $J = 1$ and 3 are very similar to each other, because in our model the anisotropy of the intermolecular interaction is assumed to be independent of the total rotational angular momentum.

To choose a rotational constant for further investigations, fig. 3.3 presents the energy spectrum of the relative motion for two interacting molecules in a one-dimensional harmonic trap as a function of the molecules' rotational constant B with the isotropic and anisotropic interaction strengths set at $g_0 = g_{\pm 1} = 4$. For the unphysical regime of a tiny B/ω ratio, the energy spectrum becomes very dense. In the rest of the present section 3.1 we assume a value of the rotational constant $B = 0.3\,\hbar\omega$.

Figure 3.4 shows the dependence of the energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in

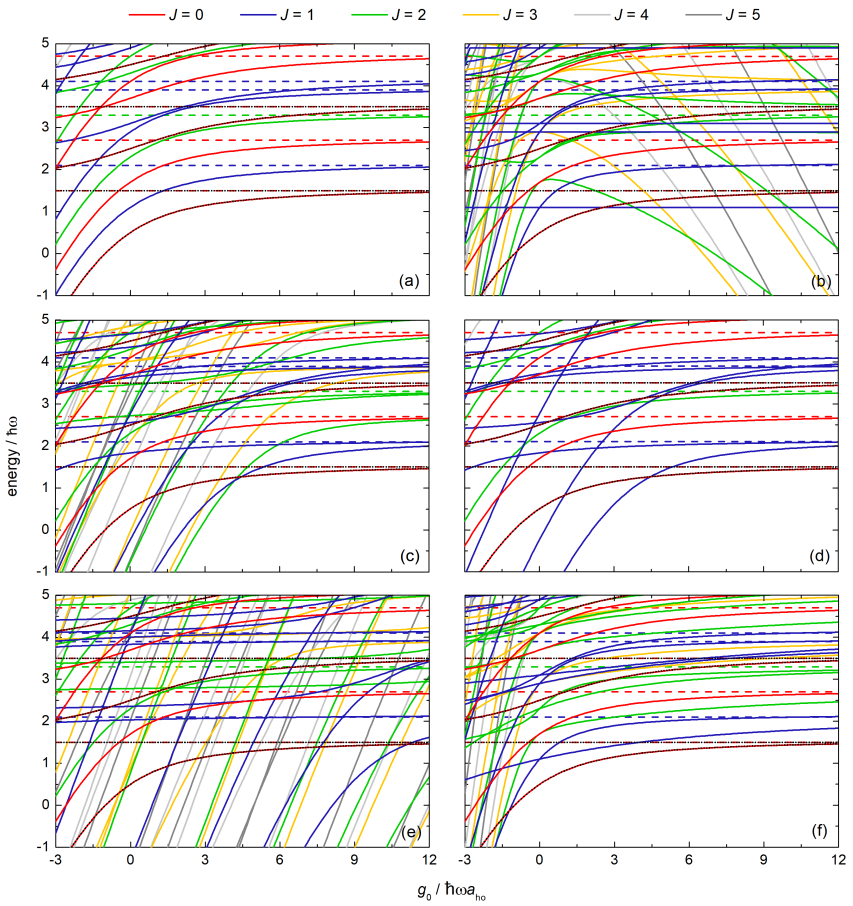

Figure 3.4: **Intermolecular interaction vs spectrum.** Energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap as a function of the isotropic interaction strength g_0 for different models of the short-range anisotropy of intermolecular interaction: (a) $g_{\pm k} = 0$, (b) $g_{\pm 1} = g_0$, (c) $g_{\pm 1} = 4$, (d) $g_{\pm 1}^{\mathrm{ex}} = 4$, (e) $g_{\pm 1} = 10$, and (f) $g_{\pm k} = g_0/1.5^k$. Solid and dashed lines are for states of even and odd spatial symmetries. Different colors represent states with different total rotational angular momenta. Dotted lines are for the result for two interacting atoms. From Ref. [330].

a one-dimensional harmonic trap on the isotropic interaction strength g_0 for different models of the anisotropic interaction. The states with the even and odd spatial symmetries are denoted by solid and dashed lines, respectively, and they are compared with the known result for two atoms denoted by the dotted lines. Because of the assumed contact-type interaction potential, the

energy of states with odd symmetry does not depend on the intermolecular interactions, similarly as in the atomic case [64]. Additionally, for systems without any rotational angular momentum, $J = 0$, $j_1 = 0$, and $j_2 = 0$, the assumed form of the anisotropic intermolecular interaction does not affect the system, and energies reduce to the atomic spectrum, in agreement with collisional results obtained with a complete description of intermolecular interactions [312]. As a reference, fig. 3.4(a) presents the energy spectrum for the molecular system without any anisotropic interaction, thus it corresponds to the atomic result multiplied and shifted by rotational energies only and it reveals a complex nature of investigated systems resulting from the richer internal structure of molecules as compared to atoms.

Figure 3.4(b)-(f) present the energy spectra with non-zero anisotropic intermolecular interactions in different scenarios. In fig. 3.4(b) we assume that the anisotropic interaction strength is the same as the isotropic one, $g_{\pm 1} = g_0$. Interestingly, in such a case, some energy levels for higher total rotational angular momenta diverge to minus infinity with increasing g_0. This indicates that molecules form clusters deeply bound by a strong anisotropic interaction, while other levels, which converge to constant energies, can be interpreted as metastable gas-like super-Tonks states [42, 348, 349]. Thus, a strong anisotropic intermolecular interaction can induce the existence of the molecular equivalent of the atomic super-Tonks-Girardeau limit in investigated systems. Specifically, in our numerical tests, we have observed such a behavior for interaction models with $g_{\pm 1}/g_0 \geq 1$.

Figure 3.4(c) and fig. 3.4(e) present the energy spectra for the system with the anisotropic interaction strength set at $g_{\pm 1} = 4$ and 10, respectively. A larger anisotropy leads to a larger distortion of the spectrum as compared to the atomic case and brings states with higher total rotational angular momentum to lower energies. Figure 3.4(d) presents the energy spectrum for the simplified version of the anisotropic interaction $g_{\pm 1}^{\mathrm{ex}} = 4$ allowing for exchanging angular momentum only. As expected, the spectrum in this case is distorted only for total angular momenta for which the assumed form of the intermolecular anisotropic interaction affects the system. Finally, fig. 3.4(f) presents the energy spectrum for the anisotropic interaction, which is proportional to the isotropic interaction strength, is non-zero for higher k, but decays geometrically with k, $g_{\pm k} = g_0/1.5^k$. Since the anisotropic interaction strength in this model is always smaller than the isotropic one, the molecular features in the spectrum are less pronounced, despite the presence of couplings for higher k. In general, the anisotropic interactions related to larger k in our model are less important than the leading coupling term $g_{\pm 1}$, because they couple states with increasing energy differences.

Figure 3.5 shows the dependence of the energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$

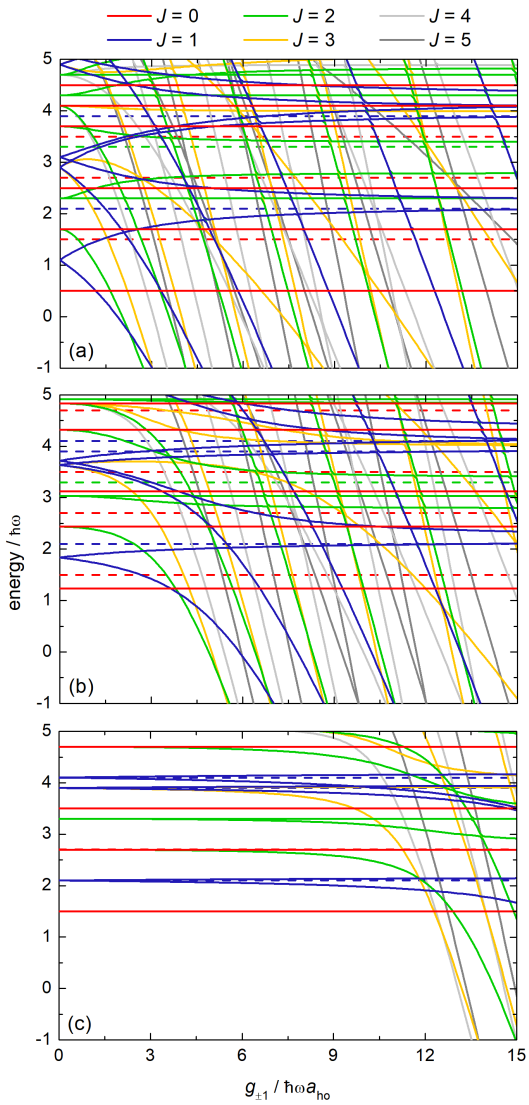

Figure 3.5: **Anisotropic interaction vs spectrum.** Energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap as a function of the anisotropic interaction strength $g_{\pm 1}$ for different isotropic interaction strengths: (a) $g_0 = 0$, (b) $g_0 = 4$, and (c) $g_0 = 17.3$ ($g_0^{\text{eff}} = \infty$ for a calculation in a finite basis set). Solid and dashed lines are for states of even and odd spatial symmetries. Different colors represent states with different total rotational angular momenta. From Ref. [330].

in a one-dimensional harmonic trap on the anisotropic interaction strength $g_{\pm k}$ for different values of the isotropic interaction strength g_0. Panels (a), (b), and (c) present results for very small ($g_0 = 0$), intermediate ($g_0 = 4$), and very large ($g_0 \to \infty$) isotropic interaction strengths, respectively. The impact of the anisotropic intermolecular interaction decreases with increasing the isotropic interaction strength. Especially, in the limit of a very large isotropic interaction strength ($g_0 \to \infty$), corresponding to the Tonks-Girardeau limit in atomic systems, the strength of the anisotropic intermolecular interaction has to be tuned to very large values to induce observable effects. It is not surprising, since for large positive (repulsive) values of the isotropic interaction, that is, in the Tonks-Girardeau limit, the interacting particles avoid each other, decreasing their wave functions' overlap and thus decreasing the effect of the short-range anisotropic interaction. For a negative strength of the isotropic interaction interaction, $g_0 < 0$, the anisotropic interaction affects systems more easily because molecules are attracted to each other. The energy spectra for simplified versions of the anisotropic interaction allowing for exchanging angular momentum only are distorted only for total angular momenta for which the assumed form of the intermolecular anisotropic interaction affects the system.

A common and interesting feature for all investigated models of intermolecular interactions analyzed in figs. 3.4 and 3.5 is that, in the presence of the anisotropic interaction, the absolute ground state of the system can have total angular momentum larger than zero, $J > 0$. Such a ground state has a $(2J + 1)$ degeneracy that can allow for the realization of interesting many-body Hamiltonians in the limit of many optical lattice sites each occupied by two molecules. The ground state and its degeneracy in such a scenario can be controlled by tuning the anisotropy of the intermolecular interaction.

Figure 3.6 presents the dependence of the total rotational angular momentum J of the ground state for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap on the isotropic g_0 and anisotropic $g_{\pm 1}$ interaction strengths. This plot clearly shows the interplay of the isotropic and anisotropic intermolecular interactions observed already in figs. 3.4 and 3.5. In the absence of anisotropy or for weak anisotropic interactions, the ground state has $J = 0$. With increasing strength of the anisotropic interaction, the ground state has increasingly higher total rotational angular momentum. For small isotropic interaction strengths it is easier to induce incrementally higher total rotational angular momentum in the ground state. For large isotropic interaction strengths, much larger anisotropic interaction strengths are needed to induce the ground state with higher J, and ground states with higher J neighbor one with $J = 0$.

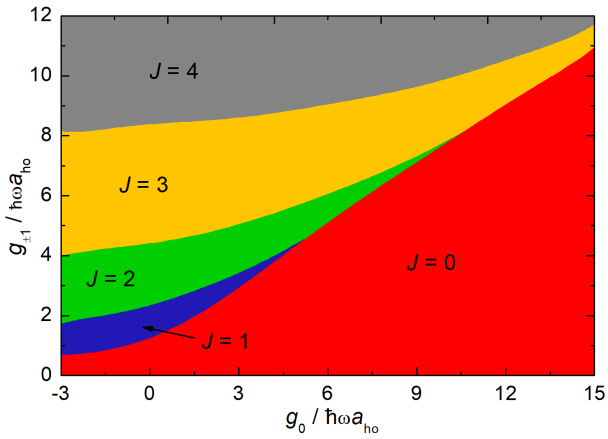

Figure 3.6: **Ground states with non-zero** J. Total rotational angular momentum J of the ground state for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap as a function of the isotropic g_0 and anisotropic $g_{\pm1}$ interaction strengths. From Ref. [330].

In the absence of external electric or magnetic fields, the total rotational angular momentum is a conserved quantity (J is a good quantum number). Therefore, it is not possible to drive the system between ground states with different total rotational angular momenta by simply tuning systems' parameters. However, transitions involving photon absorption or emission can potentially be used to reach the ground state with higher total rotational angular momentum after changing the system's parameters. Adiabatic evolution between different ground states can, however, be possible if an external electric field is applied to couple states with different J.

3.1.2 Impact of external electric field

If considered molecules are heteronuclear and posses a permanent electric dipole moment, then a static electric field can be used as a knob to control their interactions and dynamics in a trap via Stark effect [87]. An electric field couples and mixes states with different total rotational angular momenta J and removes the degeneracy of states with different $|M|$.

Figure 3.7 presents energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external static electric field as a function of the isotropic interaction strength g_0 with the anisotropic interaction strength $g_{\pm1} = 4$ and electric field strength $d\mathcal{E} = 0.5\,\hbar\omega$ and $2.5\,\hbar\omega$. These spectra result from the

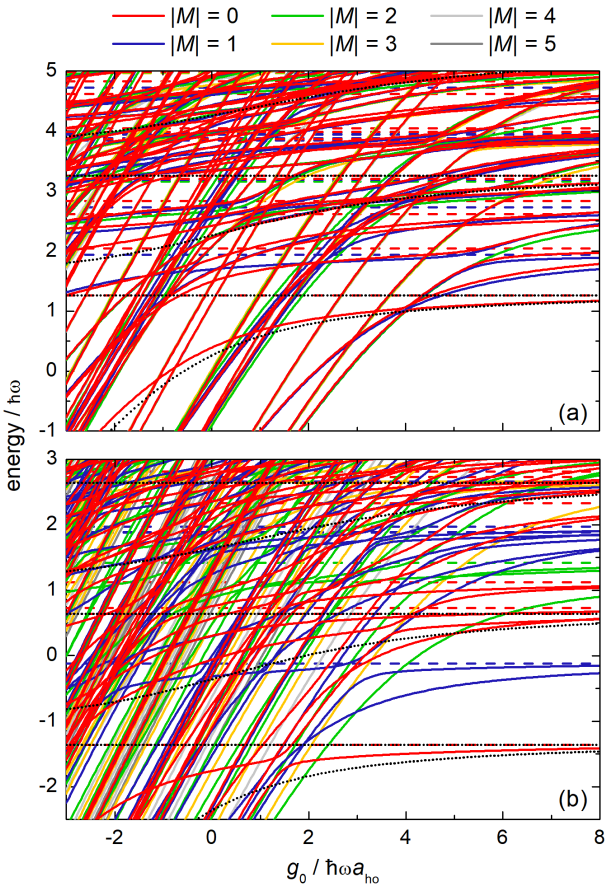

Figure 3.7: **Interplay of electric field and interaction.** Energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external static electric field as a function of the isotropic interaction strength g_0 with the anisotropic interaction strength $g_{\pm 1} = 4$ and the electric field strengths. (a) $d\mathcal{E} = 0.5\,\hbar\omega$ and (b) $d\mathcal{E} = 2.5\,\hbar\omega$. Solid and dashed lines are for states of even and odd spatial symmetries. Different colors represent states with different projections $|M|$ of the total rotational angular momentum along the field. Dotted lines are for the result for two interacting atoms shifted by the energy of two non-interacting polar molecules. From Ref. [330].

field-free spectrum shown in fig. 3.4(c). The assumed model of intermolecular interactions affects all states in the presence of a static electric field, because all eigenstates in this field are mixtures of field-free states with different total rotational angular momenta. For this reason, in fig. 3.7, there is no state overlapping with the atomic result and the deviation from the

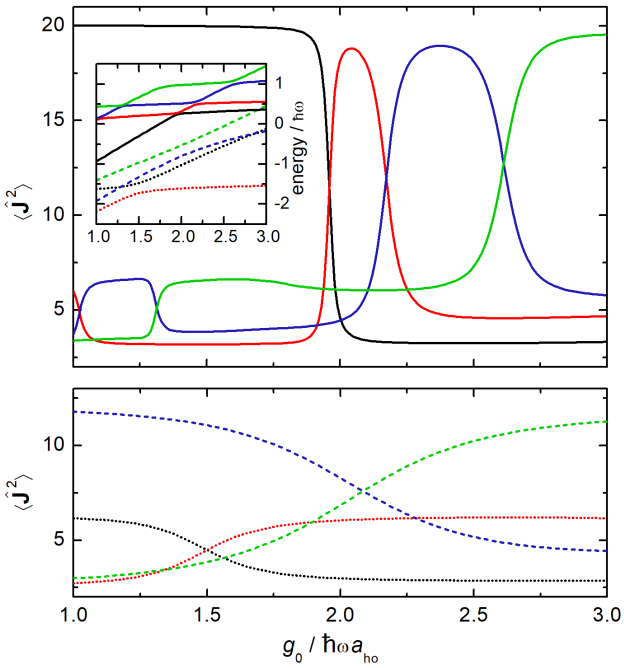

Figure 3.8: **Interaction- and electric field-induced avoided crossings.** Mean values of the square of the total rotational angular momentum operator $\hat{\mathbf{J}}^2$ for selected eigenstates of two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external static electric field as a function of the isotropic interaction strength g_0 with the anisotropic interaction strength $g_{\pm 1} = 4$ and electric field strength $d\mathcal{E} = 2.5\,\hbar\omega$. Solid, dashed, and dotted lines represent states with different projections $|M|$ of the total rotational angular momentum along the field. The inset shows the energy spectrum of analyzed eigenstates using the same color code. From Ref. [330].

atomic case is increasing with increasing the electric field strength. The electric field splits previously degenerate states into a larger number of states leading to a high density of states, especially in the strong electric field as plotted in fig. 3.7(b). The coupling between states originating from different total angular momenta results in the emergence of a large number of avoided crossings between these states when the electric field is applied. They are visible for all projections of the total rotational angular momentum along the field presented in fig. 3.7 and they are more pronounced for the larger electric field strength.

Following adiabatically eigenstates across an avoided crossing can lead

to a drastic (ex)change of eigenstates' properties. Figure 3.8 shows mean values of the square of the total rotational angular momentum operator $\hat{\mathbf{J}}^2$ for selected eigenstates presented in fig. 3.7(b). Selected eigenstates originate from states with $J = 0 - 4$. It is apparent that each avoided crossing is associated with the exchange of the total rotational angular momenta between eigenstates. The widths of the avoided crossings depend on parameters of original states and the coupling strengths between them, but all observed transitions are smooth. Interestingly, such avoided crossing can be used to control and pump the total rotational angular momentum of the system by tuning intermolecular interactions or external electric field. Such a control with an electric field would be a loose electric equivalent of using a magnetic field to pump the rotational angular momentum in the quantum variant of the Einstein–de Haas effect [350].

Figure 3.9 shows energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external static electric field as a function of the electric field strength $d\mathcal{E}$ for different models of the isotropic and anisotropic short-range interactions. As a reference, panel (a) presents the spectrum of the non-interacting molecules, thus it corresponds to the doubled result for a single polar molecule multiplied and shifted by trap vibrational energies only. Panels (b)-(d) show the spectra for the systems with the anisotropic interaction strength fixed at an intermediate value of $g_{\pm 1} = 4$, whereas the isotropic interaction strength is zero in panel (b), intermediate in panel (c), and effectively infinite in panel (d). The electric field removes the degeneracy of states with different $|M|$ for a given J, leading to a high density of states especially when both the isotropic and anisotropic interaction strengths have intermediate values. Interestingly, for the system dominated by the anisotropic interaction presented in panel (b), the lowest states have high total rotational angular momentum, low energies, and relatively low density of states. This is in agreement with fig. 3.5(a). As a result, tthe effect of the electric field is weaker. In the limit of a very large isotropic interaction strength ($g_0 \to \infty$), corresponding to the molecular Tonks-Girardeau limit, presented in panel (d), the importance of the anisotropic intermolecular interaction is reduced and the energy spectrum of the systems in the electric field simplifies to the spectrum of a single polar molecule in the field combed with the trap vibrational energies (spectra for odd and even spatial symmetries are the same).

The complex spectra with numerous avoided crossings that are presented in figs. 3.7 to 3.9 raise a question whether investigated systems show a quantum chaotic behavior [351]. To verify this hypothesis we have calculated the nearest-neighbor spacing distributions of energy levels for investigated systems in a broad range of the intermolecular interactions and external

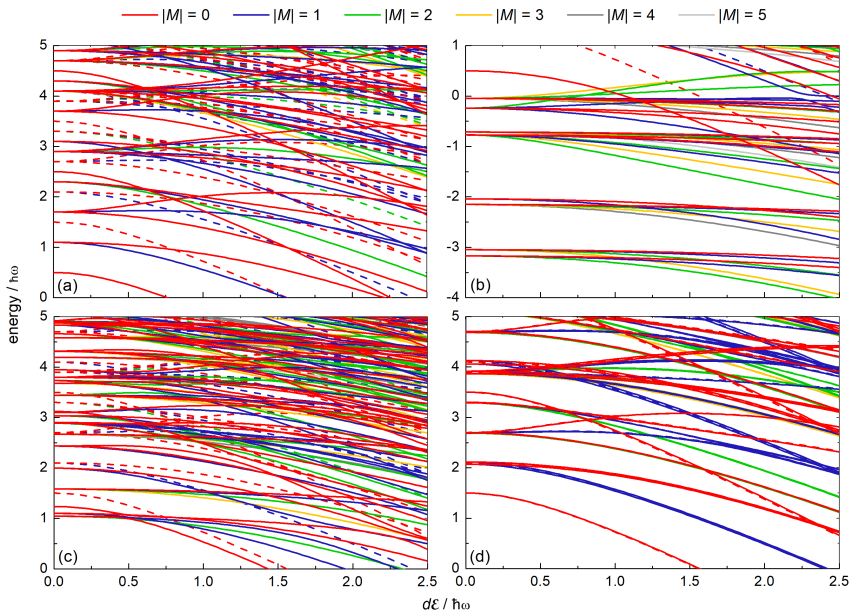

Figure 3.9: **Electric field vs spectrum.** Energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external static electric field as a function of the electric field strength $d\mathcal{E}$ for different models of the isotropic and anisotropic short-range interactions: (a) $g_0 = 0$, $g_{\pm1} = 0$; (b) $g_0 = 0$, $g_{\pm1} = 4$; (c) $g_0 = 4$, $g_{\pm1} = 4$; and (d) $g_0 = 17.3$ ($g_0^{\text{eff}} = \infty$ for a calculation in a finite basis set), $g_{\pm1} = 4$. Solid and dashed lines are for states of even and odd spatial symmetries. Different colors represent states with different projections $|M|$ of the total rotational angular momentum along the field. From Ref. [330].

fields strengths. In all calculations we have found level spacing distributions in much better agreement with the Poisson distribution than with the Wigner-Dyson one. This observation strongly suggests that the statistical properties of calculated energy spectra do not follow the predictions of the Gaussian orthogonal ensemble of random matrices, and neither quantum chaotic behavior nor level repulsion is observed. The investigated systems of two interacting molecules with short-range intermolecular interactions in the electric field behave thus rather like quantum integrable systems.

3.1.3 Dipole-dipole interaction

The dipole-dipole interaction plays an important role in physics of ultracold molecules because heteronuclear molecules can possess a permanent

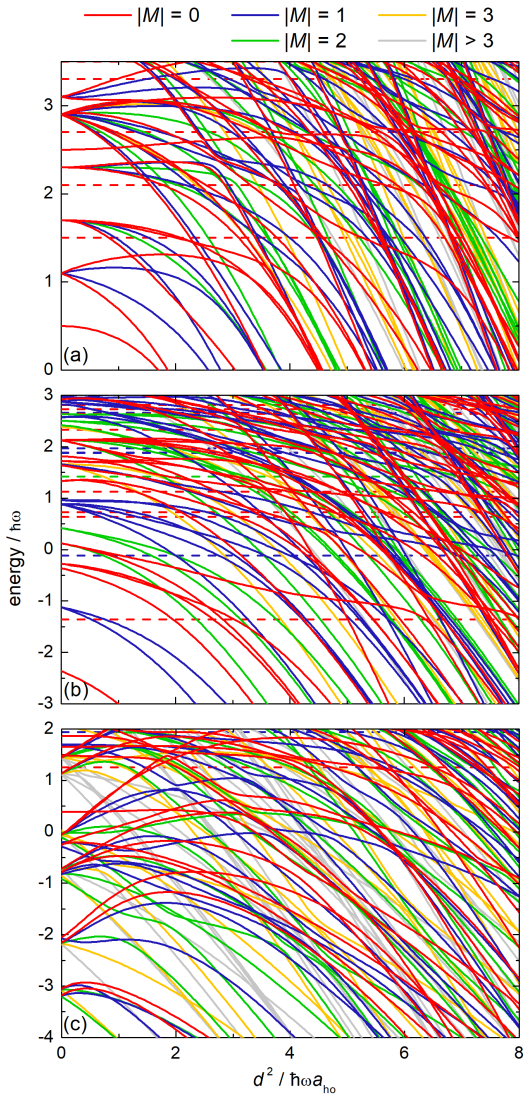

Figure 3.10: **Effective dipole interaction vs spectrum.** Energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external transverse or axial electric field as a function of the dipole-dipole interaction strength d^2: (a) $\mathcal{E} = 0$, $g_0 = 0$, $g_{\pm1} = 0$, (b) $\mathcal{E}_z = 2.5$, $g_0 = 0$, $g_{\pm1} = 0$, (c) $\mathcal{E}_z = 0.5$, $g_0 = 0$, $g_{\pm1} = 4$. Solid and dashed lines are for states of even and odd spatial symmetries. Different colors represent states with different projections $|M|$ of the total rotational angular momentum along the field. From Ref. [330].

electric dipole moment. At the same time, this interaction is of the long-range nature and can be controlled with external electric field [87]. In the present work, we neglect its long-range character and focus on its impact on coupling and mixing molecular rotational angular momenta and interplay with an external static electric field.

Figure 3.10 presents the dependence of energy spectra of the relative motion for two interacting molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap in an external electric field as a function of the dipole-dipole interaction strength. Here, we define the dipole-dipole interaction strength as a square of the molecules' permanent electric dipole moment d^2. Panel (a) shows the energy spectrum for the field-free case with the dipole-dipole interaction only. Panel (b) presents the energy spectrum with the dipole-dipole interaction in a strong external electric field parallel to the trap axis. Panel (c) presents the energy spectrum with the dipole-dipole interaction and additionally with the anisotropic interaction of intermediate strength also in an intermediate external electric field parallel to the trap axis.

Numerous avoided crossings and high density of states are visible for all cases. The dipole-dipole interaction combines effects of the isotropic g_0 and anisotropic $g_{\pm 2}$ intermolecular interactions, thus it both shifts and splits energy levels. The dipole-dipole interaction restricted to one dimension does not conserve the total rotational angular momentum J, but its projection M onto the trap axis is conserved, also in the presence of an electric field parallel to the trap axis. If an electric field is not parallel to the trap axis, then M is not a good quantum number anymore. If the electric field is perpendicular to the trap axis, then it tends to align molecules perpendicularly to the trap axis leading to repulsive interaction between molecules and increasing their energy. No signature of quantum chaos is found in these spectra.

3.1.4 Spin-rotation interaction and external magnetic field

If considered molecules posses a non-zero electronic spin angular momentum, then a static magnetic field can be used as a knob to control their interactions and dynamics in a trap via Zeeman effect [87]. In the present model, we assume that the intermolecular interaction potential does not depend on the electronic spin. Therefore, the magnetic field can couple with intermolecular dynamics through the molecular spin-rotation coupling only.

Panel (a) in fig. 3.11 presents the dependence of the energy spectrum of the relative motion for two interacting spin-1/2 molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap on the spin-rotation coupling constant γ. Intermediate strengths of the isotropic

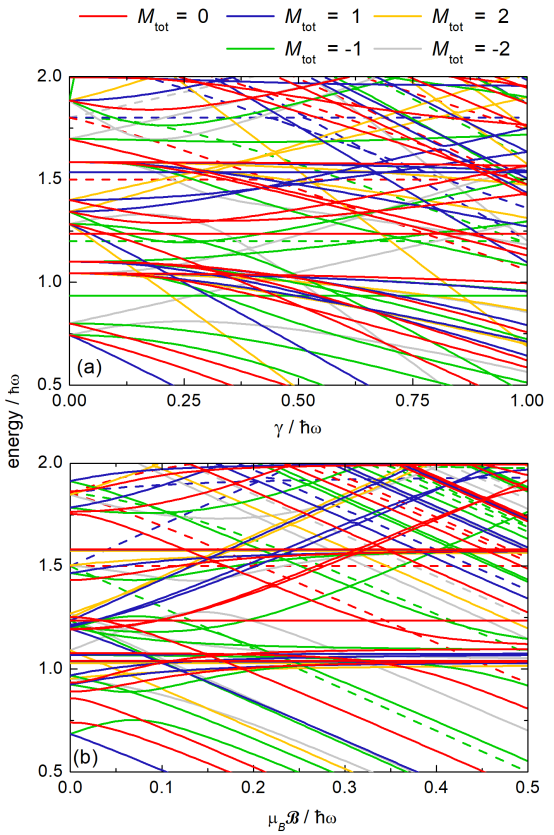

Figure 3.11: **Interplay of the spin-rotation coupling and magnetic field.** Energy spectra of the relative motion for two interacting spin-1/2 molecules with the rotational constant $B = 0.3\,\hbar\omega$ in a one-dimensional harmonic trap: (a) as a function of the spin-rotation coupling strength γ with $g_0 = 4$, $g_{\pm1} = 4$, $\mathcal{B} = 0.15\,\hbar\omega/\mu_B$ and (b) as a function of the magnetic field strength $\mathcal{B}$ with $g_0 = 4$, $g_{\pm1} = 4$, $\gamma = 0.3\,\hbar\omega$. Basis set: $n_{\max} = 30$, $j_{\max} = 5$. Solid and dashed lines are for states of even and odd spatial symmetries. Different colors represent states with different projections M of the total rotational angular momentum along the field. From Ref. [330].

and anisotropic interactions are assumed, $g_0 = g_{\pm1} = 4$ together with the magnetic field of intermediate value $\mathcal{B} = 0.15\,\hbar\omega/\mu_B$. Without magnetic field the total angular momentum $J_{\text{tot}}, M_{\text{tot}}$ is conserved and real crossings between states with different $J_{\text{tot}}, M_{\text{tot}}$ are expected. The energy spectrum gets more complex with increasing the spin-rotation coupling strength because the degeneracy of states related to the spin configuration is removed by this coupling. The magnetic field couples and splits states with differ-

ent J_{tot}, while its projection M_{tot} is conserved. Additionally, some crossings from the field-free case become avoided crossings in the magnetic field.

Panel (b) in fig. 3.11 shows the dependence of the energy spectrum of the relative motion for two interacting spin-1/2 molecules in a one-dimensional harmonic trap on the magnetic field strength $\mathcal{B}$. The intermolecular interactions are the same as in panel (a) and an intermediate value of the spin-rotation coupling constant $\gamma = 0.3\,\hbar\omega$ is assumed. If there is no spin-rotation coupling present in the system, then the magnetic field only simply splits and shifts states with different projections of the total electronic spin angular momentum on the magnetic field $M_S = m_{s_1} + m_{s_2}$, but the energy spectra for given M_S look the same. However, when the rotational and spin angular momenta are coupled and mixed by the molecular spin-rotation coupling, then the magnetic field affects the system's dynamics and can be used to control it. The magnetic field induces numerous avoided crossings similarly as the electric field in previous sections. No signature of quantum chaos is found for spectra in the magnetic field, either.

3.1.5 Conclusions

Motivated by experimental possibilities and ongoing efforts aiming at the production and application of fully controllable systems of few ultracold molecules trapped in optical tweezers or optical lattices, we have developed the model description of two interacting ultracold polar molecules effectively trapped in a one-dimensional harmonic potential. Molecules are described as distinguishable quantum rigid rotors interacting via multichannel two-body contact potential incorporating the short-range anisotropy of intermolecular interactions including dipole-dipole interaction. The form of the employed multichannel potential is motivated by the known nature of short-range chemical intermolecular interactions. We have also included interactions with external electric and magnetic fields via Stark and Zeeman effects, respectively.

We have carefully applied several approximations needed to simplify calculations, and to separate the impact and importance of different features of the molecular structure and intermolecular interactions on the system's dynamics. Thus, we have attempted to understand two interacting ultracold polar molecules trapped in a one-dimensional harmonic potential in a step-by-step manner. We have investigated the properties of such a system in a broad range of system parameters and external field strengths. We have analyzed in detail the interplay of the molecular rotational structure, anisotropic interactions, spin-rotation coupling, electric and magnetic fields, and harmonic trapping potential.

Our most important findings can be summarized as follows.

1. The anisotropic intermolecular interaction brings states with higher total rotational angular momenta to lower energies such that the absolute ground state of the molecular system can have the total angular momentum larger than zero and be degenerate.

2. If the anisotropic interaction strength is larger than the isotropic one, then some energy levels for higher total rotational angular momenta diverge to minus infinity with increasing the intermolecular interaction strength. This indicates the emergence of the molecular equivalent of the atomic super-Tonks-Girardeau limit with clustered ground state and excited gas-like super-Tonks states.

3. In the limit of a very large isotropic interaction strength, corresponding to the Tonks-Girardeau limit in the atomic system, the molecular character of the system is less pronounced and impact of the anisotropic interaction and electric field is smaller.

4. The electric and magnetic fields efficiently couple and mix states with different total angular momenta and result in complex energy spectra with a high density of states.

5. The electric and magnetic fields as well as dipole-dipole interaction induce a large number of avoided crossings.

6. Driving adiabatically the system across above avoided crossings can be used to control its properties. Especially, the total rotational angular momentum can be pumped to the system in a loose electric equivalent of the quantum Einstein–de Haas effect.

7. We have not found signatures of quantum chaotic behavior in energy spectra of investigated systems which suggests their quantum integrability.

Replacing atoms with molecules in ultracold quantum few- and many-body systems opens up new possibilities stemming from molecules' rich internal structure and anisotropic intermolecular interactions, including long-range ones. Therefore, the present model and results may provide understanding and microscopic parameters for effective molecular many-body Hamiltonians. For example, our calculations may be considered as a microscopic model for the on-site interaction of the molecular multichannel Hubbard Hamiltonian. Thus, our results may be useful for the development of bottom-up molecule-by-molecule assembled molecular quantum simulators.

The results presented here may be followed by many applications of the proposed model and numerical approach to investigate interesting physics in different molecular systems, geometries, and dimensions. There are several possible extensions of the present work. The most straightforward direction is moving to two- or three-dimensional arrangements. In contrast to the atomic case, no direct correspondence between energy spectra in one dimension and three dimensions is expected for the molecular system. The bosonic or fermionic nature of interacting molecules can also be addressed. On the other hand, the long-range character of the dipole-dipole interaction can be included and its interplay with isotropic and anisotropic short-range van der Waals interactions can be investigated. Polar and paramagnetic molecules possessing at the same time both permanent electric dipole moment and spin structure can be studied together with their control with external electric and magnetic fields. Such systems may show an interesting interplay of magnetic and electric properties coupled by the molecular internal structure. Hyperfine structure of molecules and chemical reactivity can also be included. Molecules possessing different masses, rotational constants, and trapping frequencies, and resulting coupling between the center-of-mass and relative motions, can be considered. Emergence of quantum chaotic properties with increasing complexity of the system is another intriguing question. Time-dependent dynamics in few-body molecular systems, especially after a quench of system parameters, is another not explored but potentially interesting direction of research. Finally, two interacting molecules can be trapped in two sites of an optical lattice or optical tweezer and such a double-well configuration can be investigated as a fundamental building block for the implementation of quantum gates and quantum computation with molecular systems.

3.2 Magnetic properties and quench dynamics

This section is based on our work presented in Ref. [331], where the Ph.D. candidate was responsible for: full numeric implementation (construction of Hamiltonian matrices, exact diagonalization, and quench dynamics), analysis and interpretation of results, preparation of plots, writing the manuscript. To our knowledge, this work is the first step towards the incorporation of highly magnetic and polar molecules into the field of ultracold quantum simulators. The code and data that enable the recovery of results in this section are provided in Ref. [344].

In this section, we theoretically investigate the magnetic properties and nonequilibrium dynamics of two interacting ultracold polar and paramagnetic molecules in a one-dimensional harmonic trap in external electric and magnetic fields (see fig. 3.12). Intriguing features of the system arise from the interplay of the molecular electronic spins, dipole moments, rotational structures, external electric and magnetic fields, and spin-rotation coupling. We present rich diagrams of the system's magnetization and explain the mechanisms allowing its control on the example of molecules with spins 1/2 and 3/2. We identify the anisotropic part of the intermolecular interaction and the spin-rotation coupling as crucial for observing the system's nontrivial magnetic behavior. We propose the quench dynamics experiments to probe and reconstruct the system's molecular characteristics from observing its time evolution. We show that the strong anisotropic interaction leaves a clear mark on the system's time evolution after the quench of the interaction strength. On the other hand, the time evolution of the system's magnetization, after the electric field quench, depends significantly on the spin-rotation coupling strength. The results show an intimate coupling between the electric and magnetic properties of the system and indicate the possibility of controlling the molecular few-body magnetization with the external electric field. In this way, we complement the previous studies on the coupling between the molecular electronic spins and external electric field in the free-space collisions [107–109]. The investigated model system paves the way towards studying the controlled magnetization of ultracold molecules trapped in optical tweezers or optical lattices and their application in quantum simulation of molecular multichannel many-body Hamiltonians and quantum information storing.

The rotational constant B in this study is set to $\pi\hbar\omega$ as presented in fig. 3.13. It corresponds to a less pronounced molecular character of the system as compared to Ref. [330], where tiny rotational constants were selected to enhance the impact of molecular features. Now, trap levels are more dense than rotational ones, which is a regime closer to experimental conditions (in which molecular rotational constants are usually much larger than a trap frequency). Still, we select the relatively small rotational constant to reveal an important role played by the rotational degree of freedom. With a choice of an irrational value, we additionally avoid accidental degeneracies of energy levels. We also limit the set of $|n\rangle$ to even functions due to the trivial behavior of odd states as showed in the previous section and Ref. [330].

We begin by studying in section 3.2.1 the magnetic properties of two interacting ultracold molecules in a one-dimensional harmonic trap. We focus on the system's magnetization. We analyze how it depends on the intermolecular interaction between molecules and the coupling between the electronic spins and the molecular rotational momenta. We present how the

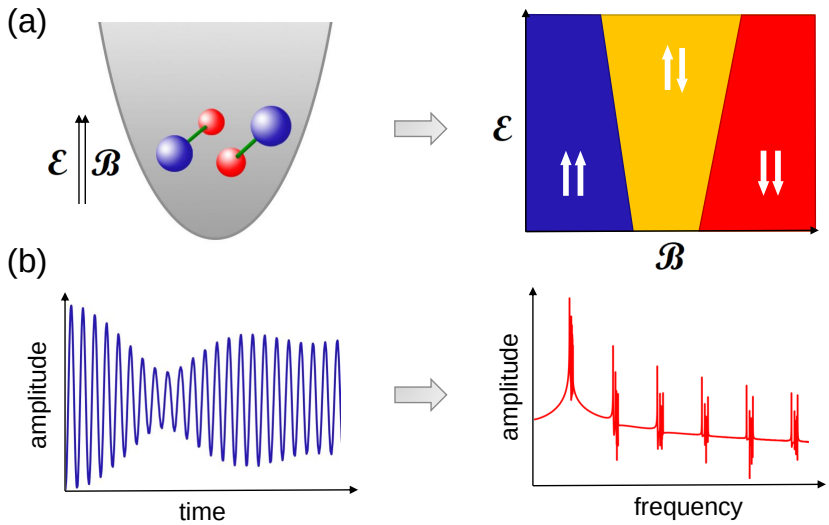

Figure 3.12: **Schematic representation of the investigated system and its features.** (a) Two interacting molecules is a one-dimensional harmonic trap under the influence of the electric and magnetic fields can be described by a magnetization diagram depending on the field strengths. The fields are parallel to each other and to the direction of molecular motion. (b) Time evolution of the system's observable after the quench can reveal information on the underlying molecular characteristics by using the Fourier transform. From Ref. [331].

system's magnetization can be controlled and how this control depends on the molecular properties. Next, in section 3.2.2, we study the quench dynamics designed to extract the strengths of the spin-rotation coupling as well as the isotropic and anisotropic interaction strengths between the molecules. We summarize our results and discuss the outlook in section 3.2.3.

3.2.1 Magnetic properties and its control

Mechanisms that allow control over the system's magnetization can be observed from the energy spectra. Figure 3.14 shows calculated energy spectra as functions of the isotropic interaction strength g_0. We select a set of internal and external Hamiltonian parameters to present the interplay between the system's magnetic properties and the external fields.

Figure 3.14(a) presents the energy spectrum of two molecules in a one-dimensional harmonic trap with a strongly anisotropic intermolecular interaction $g_{\pm 1} = 10$ without external fields or spin-rotation coupling. The comparison with the neighboring panel (b) shows the impact of the exter-

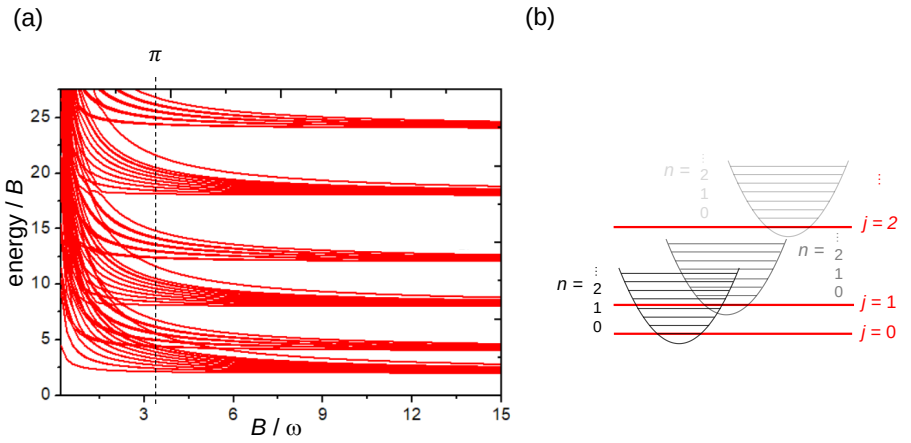

Figure 3.13: **Larger ratio between the rotational constant and trapping frequency.** $B/\omega \approx 3$ preserves the importance of molecular features and is realizable with weakly-bound Feshbach molecules in a tight trap. The rotational levels are separated by several trap levels.

nal electric field on the system. The electric field moves the energy levels by a Stark shift and removes their degeneracy with respect to the total rotational angular momentum, J. The electric field splits states with $J = 1$ and $J = 2$ into two and three states, respectively, according to the different number of possible projections of total rotational angular momentum, M_J. Shifted states then often anticross due to the coupling between different total rotational momentum states.

A comparison of panels (a) and (c) in fig. 3.14 shows the impact of the medium magnetic field on the spectrum of two molecules with a small spin-rotation coupling ($\gamma = 0.3\,\hbar\omega$). The only conserved quantum number is M_{tot}, i.e., the sum of projections of total rotational, M_J, and spin, M_S, angular momenta. States split accordingly to the Zeeman shift. States with the positive projection of the total spin angular momentum M_S are high-field seekers, while energies of states with negative M_S decrease with the magnetic field strength. Shifted states then often anticross due to a non-zero spin-rotation coupling, which mixes states with different J, S, M_J, and M_S (see the corresponding Hamiltonian elements in appendix A). The impact of the spin-rotation coupling on the system increases with the absolute values of states' total rotational and spin angular momenta' projections, M_J and M_S. The larger $|M_J|$ and $|M_S|$, the more numerous are possible combinations of the projections of individual rotational and spin angular

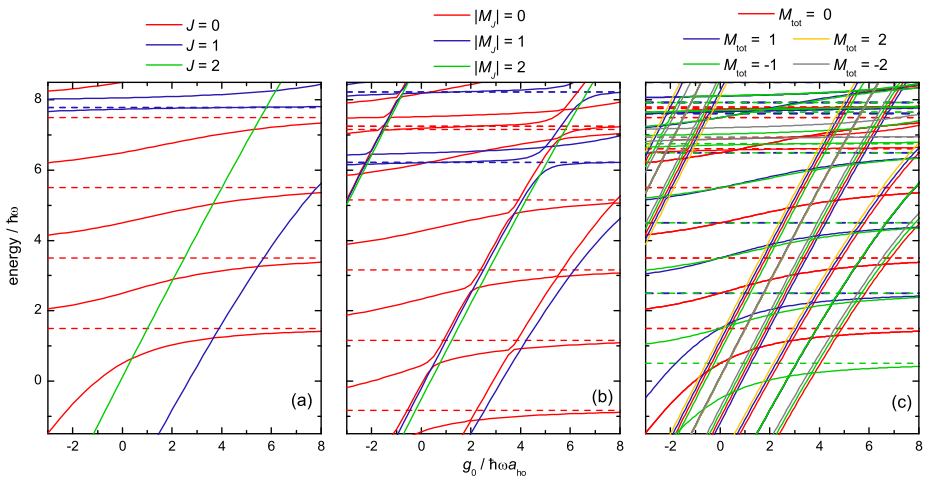

Figure 3.14: **Spectrum vs molecular properties and external fields.** Energy spectra of the relative motion for two interacting molecules with the spin $s = 1/2$ and rotational constant $B = \pi\,\hbar\omega$ in a one-dimensional harmonic trap as a function of the isotropic interaction strength g_0 with the anisotropic interaction strength $g_{\pm 1} = 10$ and (a) no external fields, (b) electric field strength $d\mathcal{E} = 5\,\hbar\omega$, and (c) the magnetic field strength $\mathcal{B} = 0.5\,\hbar\omega/\mu_B$ and spin-rotation coupling constant $\gamma = 0.3\,\hbar\omega$. Solid and dashed lines are for states of even and odd spatial symmetries, respectively. From Ref. [331].

momenta, m_i, m_{s_i}, which are mixed by the spin-rotation coupling.

Figure 3.15 presents rich magnetization diagrams of the studied system in the ground state as functions of the magnetic and electric fields. The upper row shows results for two molecules with electronic spins 1/2 and $M_{\rm tot} = 0$, while the bottom one - with spins 3/2 and $M_{\rm tot} = 2$. The primary observation is that the number of possible magnetization values of the ground state is limited by the total electronic spin momenta of molecules and the selected $M_{\rm tot}$ value. This restriction results from the definition of $M_{\rm tot} = M_J + M_S$, but also because low-energy states are characterized by small values of rotational angular momenta, j_1 and j_2, resulting in M_J values, which are close to zero.

The main reason for all magnetization changes in fig. 3.15 is the interplay between the Zeeman and Stark effects. The magnetic field linearly brings down energies of the states with negative M_S, with speed depending on the M_S value. The Stark effect lowers the ground state's energy, composed mostly of the basis state with $J = M = j_1 = j_2 = 0$, faster than the lowest state with negative magnetization, composed mostly of the basis state with $J = M = 1$. Therefore, larger external electric field strengths effectively

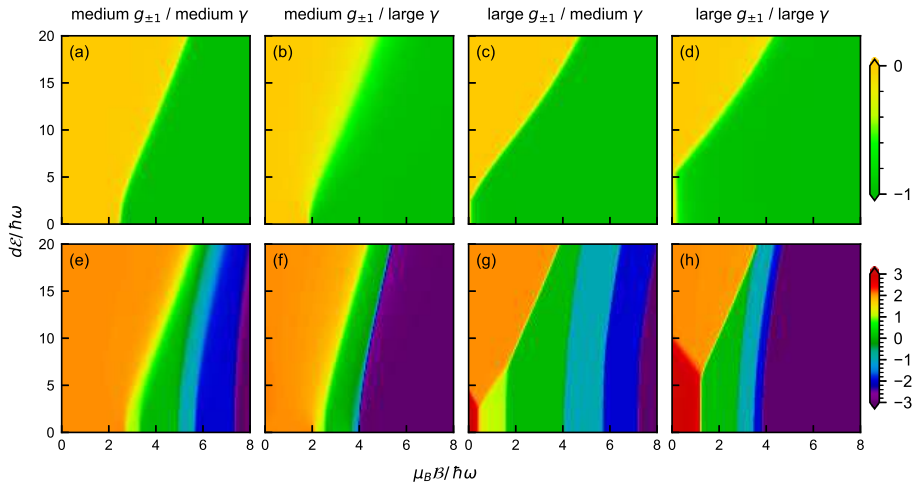

Figure 3.15: **Magnetization vs molecular properties and external fields.** Ground-state magnetization diagrams as functions of the electric $d\mathcal{E}$ and magnetic $\mu_B\mathcal{B}$ field strengths for two interacting molecules with the rotational constants $B = \pi\,\hbar\omega$ in a one-dimensional harmonic trap for medium isotropic interaction strength $g_0 = 4$. Panels (a)-(d) and (e)-(h) present results for spins $s = 1/2$ ($M_{\rm tot} = 0$) and $3/2$ ($M_{\rm tot} = 2$), respectively. The first two and last two columns show results for medium $g_{\pm1} = 4$ and large $g_{\pm1} = 10$ anisotropic interaction strengths, respectively. Moreover, the first and third columns present results for a medium spin-rotation coupling $\gamma = 1\,\hbar\omega$, and the second and the fourth columns - for strong spin-rotation coupling $\gamma = 3\,\hbar\omega$, respectively. From Ref. [331].

force larger magnetic field strengths for the magnetization change to happen, when the lowest states exchange their order. Such underlying interplay is visible in all panels of fig. 3.15.

Magnetization for spins 1/2

Panels (a)-(d) of fig. 3.15 present the ground-state magnetization diagrams for the system composed of two interacting molecules with spins 1/2 and $M_{\rm tot} = 0$. The choice of $M_{\rm tot}$ limits the number of possible $\langle\hat{S}_z\rangle$ values. In this case, two states are mainly responsible for the magnetization change, namely the ground state dominated by the $J = M_J = j_1 = j_2 = 0$ and $M_S = 0$ basis state and the excited state dominated by the $J = M_J = 1$ and $M_S = -1$ basis state.

Panels (a)-(b) of fig. 3.15 show the magnetization diagrams for medium strength of the anisotropic interaction ($g_{\pm1} = 4$). In the absence of the spin-rotation coupling, the magnetization change results simply from the Zeeman

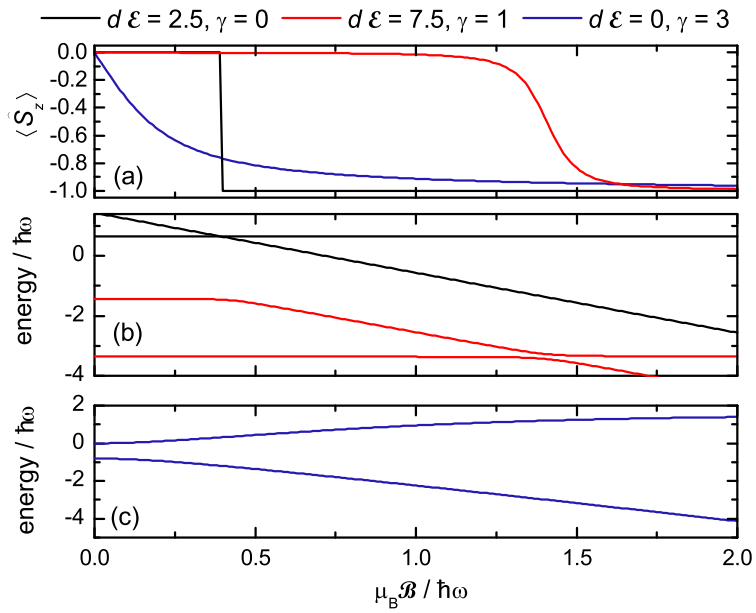

Figure 3.16: **Magnetic field and magnetization exchange of spin−1/2 molecules.** (a) Ground-state magnetization of two interacting molecules with the rotational constants $B = \pi\,\hbar\omega$, spins $s = 1/2$, and $M_{\mathrm{tot}} = 0$ in a one-dimensional harmonic trap with medium isotropic and strong anisotropic intermolecular interaction strengths ($g_0 = 4$ and $g_{\pm 1} = 10$) as a function of the magnetic field strength $\mu_{\mathrm{B}}\mathcal{B}$ for different electric field strengths $d\mathcal{E}$ and spin-rotation coupling constants γ. (b)-(c) Energies of the analyzed ground states and coupled lowest-energy excited states as functions of the magnetic field strength $\mu_{\mathrm{B}}\mathcal{B}$. Color code indicates the strengths of the electric field $d\mathcal{E}$ and the spin-rotation coupling γ. From Ref. [331].

and Stark effects' interplay. Figure 3.16(b) depicts in black the energies of states taking part in such a change. The black line in the panel (a) of the same figure presents the resulting $\langle \hat{S}_z \rangle$ of the ground state. The electric field can control the magnetic field's strength at which the change takes place.

The sharply crossing states are not coupled either by the electric field (which conserves M_J) or the spin-rotation coupling (which conserves j_1 and j_2). However, if both couplings are present, the intermediate state dominated by the $J = 1, M_J = 0$ basis state, provides the second-order coupling between the discussed states, which results in their anticrossing visible in red in fig. 3.16(a)-(b). The repulsion between states grows with the spin-rotation coupling strength, as seen when comparing the magnetization diagrams in panels (a) and (b) of fig. 3.15. The spin-rotation coupling also lowers the energy of states with the largest absolute values of M_J and M_S, allowing

for the magnetization change in the weaker magnetic field.

The importance of the intermolecular interaction anisotropy is visible when comparing the first two columns ($g_{\pm 1} = 4$) with the last two ones ($g_{\pm 1} = 10$) of fig. 3.15, i.e., panels (a)-(b) with (c)-(d). Firstly, the systems with the larger anisotropic interaction strength require smaller external fields for the ground-state magnetization change related to the crossing of the lowest-energy states. For electric field strengths larger than a few $\hbar\omega$, the $\langle \hat{S}_z \rangle$ change takes place within the same two mechanisms described above. States with different magnetization either cross in the absence of the spin-rotation coupling or anticross when both the electric field and spin-rotation coupling are present. However, the intermolecular interaction's large anisotropy allows an additional mechanism for small electric field strengths. It brings down the states with higher total rotational momenta, including $J = 1$ (as seen in fig. 3.14 (b)), and non-zero M_J and M_S. So when the magnetic field, through the Zeeman effect, lifts the degeneracy of M_S, the two states strongly repel each other thanks to the spin-rotation coupling, as seen in blue in fig. 3.16(a),(c). The larger the spin-rotation coupling strength, the larger electric field strength is needed to push down the state with $J = 0$ and reproduce the mechanisms described for smaller $g_{\pm 1}$.

Magnetization for spins 3/2

Panels (e)-(h) of fig. 3.15 show the ground-state magnetization diagrams of two interacting molecules with spins 3/2 and $M_{\text{tot}} = 2$. This choice results in a larger number of possible $\langle \hat{S}_z \rangle$ values, ranging from -3 to 3. Also, the states taking part in the magnetization changes have much higher rotational angular momenta than ones in the previous section.

In the case of the medium anisotropic interaction strength, the choice of $M_{\text{tot}} = 2$ in the absence of external fields results in a ground state with $M_J = 0, M_S = 2$. In the absence of the electric field or the spin-rotation coupling, the lowest states cross each other due to the Zeeman effect, resulting in abrupt ground-state magnetization changes. Blue lines in fig. 3.17(a)-(b) present an example of the magnetization and energies of such states as a function of the magnetic field strength.

When the spin-rotation coupling is present, the spectra become dense and exhibit multiple anticrossings, due to mixing of J, S, and projections of individual molecular rotational and spin angular momenta $m_1, m_2, m_{s_1}, m_{s_2}$. The anticrossing strength depends on two factors. The strongest anticrossings occur between states belonging to the same harmonic level, as both the spin-rotation coupling and electric field conserve n. Another factor is the anticrossing states' composition of individual rotational angular momenta, j_1 and j_2. The larger difference between them, the smaller is the coupling

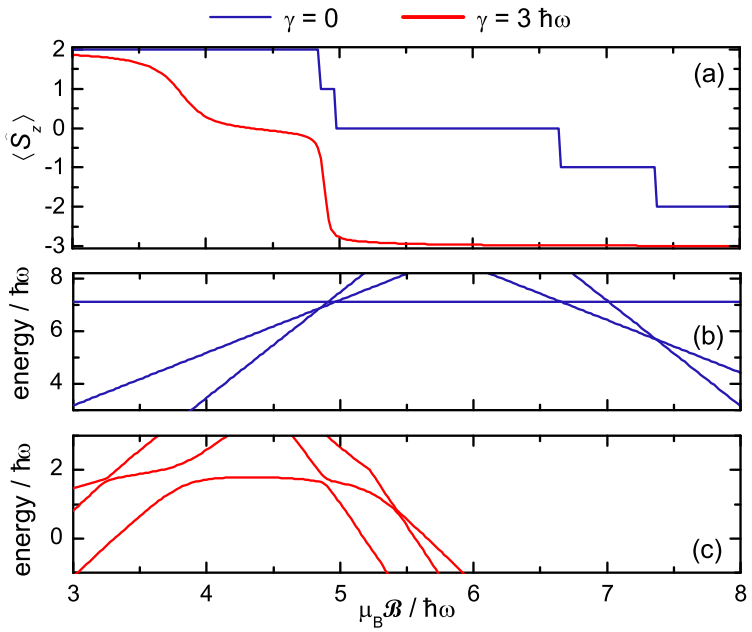

Figure 3.17: **Magnetic field and magnetization exchange of spin−3/2 molecules.** (a) Ground-state magnetization of two interacting molecules with the rotational constants $B = \pi\,\hbar\omega$, spins $s = 3/2$, and $M_{\mathrm{tot}} = 2$ in a one-dimensional harmonic trap with medium isotropic and anisotropic intermolecular interaction strengths ($g_0 = g_{\pm 1} = 4$) as a function of the magnetic field strength $\mu_{\mathrm{B}}\mathcal{B}$ for the electric field strength $d\mathcal{E} = 15\,\hbar\omega$ and different spin-rotation coupling constants γ. (b)-(c) Energies of the analyzed ground states and coupled lowest-energy excited states as functions of the magnetic field strength $\mu_{\mathrm{B}}\mathcal{B}$. Color code indicates the spin-rotation coupling constants γ. From Ref. [331].

induced by the electric field. This is why the anticrossing strength decreases with the difference of states' magnetization, as presented in fig. 3.17, where states depicted in red change the magnetization from 2 to 0 and 0 to -3. Due to conserved M_{tot}, the change of M_S is compensated by the increase of M_J. Larger M_J forces higher rotational momenta, j_1 and j_2. Therefore, the larger magnetization change, the larger difference between the states' rotational momenta, and the smaller coupling between anticrossing states.

The spin-rotation coupling's stronger impact on states with large absolute values of M_S and M_J is more prominent for spins 3/2 than for spins 1/2. A comparison between panels (e) and (f) as well as (g) and (h) of fig. 3.15 shows that the larger spin-rotation coupling can not only bring the magnetization change to lower magnetic field strengths, but also effectively

limits the number of accessible magnetization values, as in the case of panels (f) and (h).

The same dependencies determine the magnetization diagrams for the systems with large anisotropy of the intermolecular interaction ($g_{\pm 1} = 10$), as seen in panels (g) and (h) of fig. 3.15. The main difference comes from an additional $\langle \hat{S}_z \rangle$ value accessible for weaker electric fields due to bringing down the state with higher rotational angular momentum with $\langle \hat{S}_z \rangle = 3$ by the anisotropy of the intermolecular interaction.

3.2.2 Quench dynamics

The analyzed system's magnetic properties depend strongly on the anisotropic part of the intermolecular interaction and the spin-rotation coupling. These molecular properties are challenging to calculate or measure. However, they can be extracted through the analysis of the quench dynamics of observables that they influence.

Therefore, first, we study the nonequilibrium dynamics of the cloud size, $\langle \hat{r}^2 \rangle$, after the quench of the intermolecular interaction which can be achieved via the change of the trapping frequency [317] (see the discussion in section 2.8). We aim at identifying dynamical signatures of isotropic and anisotropic part of the interaction between the molecules. Next, we analyze the time evolution of the magnetization, $\langle \hat{S}_z \rangle$, after the quench of the external electric or magnetic field. It provides insight into the strength of the spin-rotation coupling present in the molecular system.

To reconstruct couplings governing the dynamics, we perform the discrete Fourier transform (DFT) of the corresponding time evolution. The resulting function indicates the frequencies dictating the time evolution. These frequencies can then be transformed into energy differences between states whose coupling causes the system's nontrivial dynamics. The strength of the coupling is related to the peak's amplitude at the corresponding frequency.

The isotropic and anisotropic interactions

Figure 3.18 presents the nonequilibrium dynamics of the analyzed molecular system with total rotational angular momentum, $J = 1$, after the quench of the interaction, starting from the noninteracting case, with the initial state $|\Psi_0\rangle = |n = 0,\, J = 1,\, j_1 = 0,\, j_2 = 1\rangle$.

Panel (a) shows the time evolution of the cloud size after quenching the isotropic part of the intermolecular interaction from zero to medium strength $g_0 = 4$, which couples states composed of different harmonic trap levels. The inset of panel (a) presents the DFT of the studied time evolution, which can be used to unravel couplings between states governing

the quench dynamics. The molecular system's dynamics in these conditions is almost identical to two ultracold atoms in a one-dimensional harmonic trap, even though the rotational structure is present, and J is nonzero. The reason is the lack of the anisotropic part of the intermolecular interaction and the absence of couplings between the rotational states. As a result, the multiple peaks visible in the DFT correspond to the couplings with different harmonic states only, which are almost exactly evenly separated by $2\,\omega$ (we ignore odd states in this work). The divergence from the single ladder of frequencies comes from the slightly different influence that the isotropic interaction has on the system's ground state, as compared to excited ones, as known for both atomic [64] and molecular [330] cases. The additional ladder of frequencies slightly below $2\,\omega$ comes from the couplings of higher-energy harmonic states to the ground state. If the isotropic intermolecular interaction is quenched to negative strengths, the multiple peaks would become more detached. In the case of the quench to the strongly repulsive interaction, entering the Tonks-Girardeau regime, the DFT would show a single ladder of peaks evenly separated by $2\,\omega$. The analysis above is independent of the total rotational angular momentum value, J as long as the anisotropic part of the interaction is zero.

Panel (b) of fig. 3.18 presents the dynamics after the quench of the anisotropic part of the intermolecular interaction from zero to medium strength $g_{\pm 1} = 4$, keeping the isotropic part equal to zero. The anisotropic interaction couples not only different harmonic trap states but also rotational ones, preserving J. It impacts the spectrum in two ways. Firstly, as showed in Ref. [330], the anisotropic interaction splits each harmonic trap state with $J = 1$ into two states, the antisymmetric and symmetric one. The splitting depends slightly on the harmonic level and is largely similar across the spectrum, except for the lowest-energy state. The antisymmetric ground state, resulting from the splitting of the lowest-energy harmonic state, is brought down rapidly by the anisotropic part of the interaction. Both effects can be seen in the DFT of the time evolution of the cloud size in fig. 3.18(b) as well as in the system's spectrum presented in fig. 3.14(a). The described splitting results in two close ladders of excited symmetric and antisymmetric eigenstates with frequencies close to $2\,\omega$. The ladder of frequencies separated by $\approx 1.9\,\omega$ comes from the couplings between the symmetric states, while the neighboring ladder starting from $\approx 2.1\,\omega$ results from the couplings between the antisymmetric states. The couplings with the separated ground state cause an additional ladder starting for a frequency equal to the energy difference between the ground energy and the nearest excited antisymmetric state (here around $4.4\,\omega$). The ground state's sensitivity to the anisotropic interaction strength allows for using this frequency as a quite precise signature of this molecular property. Invisible for

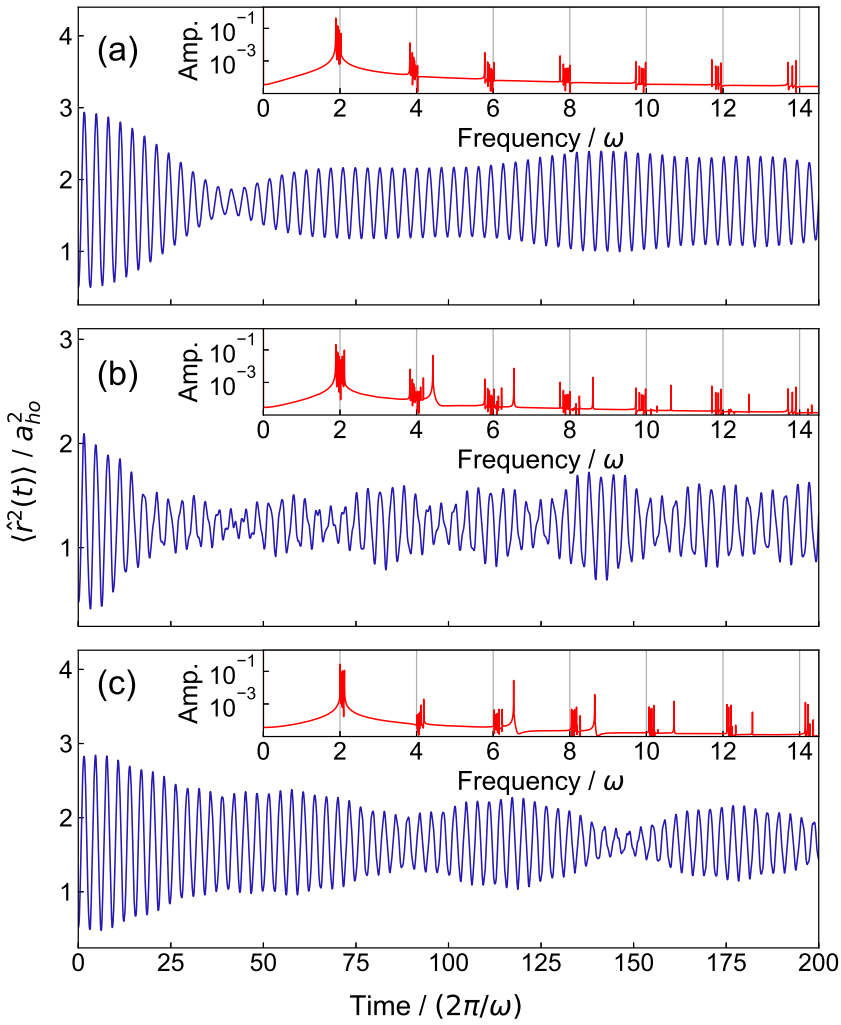

Figure 3.18: **Quench of interaction strength.** The time evolution of the cloud size, $\langle \hat{r}^2 \rangle$, after the quench of the intermolecular interaction strength between two interacting molecules without spin and with the rotational constants $B = \pi\hbar\omega$ in a one-dimensional harmonic trap with the total rotational angular momentum $J = 1$. The isotropic and anisotropic interaction is quenched from zero $g_0 = g_{\pm 1} = 0$ to (a) $g_0 = 4$, $g_{\pm 1} = 0$, (b) $g_0 = 0$, $g_{\pm 1} = 4$, and (c) $g_0 = 4$, $g_{\pm 1} = 10$. Insets present the discrete Fourier transforms of the studied time evolutions. From Ref. [331].

the quench dynamics are states with $j_1 = j_2$, as nothing couples them to the system's initial state. If the initial state, $|\Psi_0\rangle$, is antisymmetric, instead of composed solely by $|n = 0, J = 1, j_1 = 0, j_2 = 1\rangle$, the only couplings governing the system are the ones between antisymmetric states. The time evolution of $\langle \hat{r}^2 \rangle$ is then simpler, and the DFT contains no splittings of the main frequency ladder. If $|\Psi_0\rangle$ is chosen to be symmetric, the couplings to the antisymmetric ground state disappear, rendering this quench scenario less sensitive to the anisotropic interaction strength.

Panel (c) of fig. 3.18 shows the time evolution of the cloud size after quenching both parts of the intermolecular interaction, i.e., the isotropic part from zero to medium strength $g_0 = 4$ and anisotropic part from 0 to large strength $g_{\pm 1} = 10$. In this case, the DFT shows a frequency ladder similar to the one observed in panel (a) and the additional ladder starting from the frequency around $6.5\,\omega$ analogous to the panel (b). However, the result is not a simple sum of two interaction parts' effects, but it rather comes from the competition between them. Firstly, the ground state, detached from the rest of the spectrum by the anisotropic part of the interaction, is pushed back up by the isotropic part, as seen in fig. 3.14(a). For a molecular system with the dominantly isotropic interactions or with an anisotropic part equal to isotropic, the separated ground state is gone and the additional ladder in the corresponding DFTs. On the other hand, the larger frequency at which the additional ladder starts, the larger is the dominance of the anisotropy over the isotropy of the intermolecular interaction. The splittings between ladders starting around $2\,\omega$ are another signature of the intermolecular interaction. They get negligible small, resulting in a single ladder, in two cases. The first case is entering the Tonks-Girardeau regime with a very strong isotropic part of the interaction. Second, when the anisotropic part of the interaction's strength is equal to the isotropic one. This regime corresponds to the metastable gas-like super-Tonks states [42, 330, 348, 349], being a molecular equivalent of the Tonks-Girardeau regime.

The results showed in fig. 3.18 and discussed above concern the system with the total rotational angular momentum $J = 1$. They present the possibility of extracting the relative strength of the anisotropic part of the intermolecular interaction, $g_{\pm 1}$, compared to the isotropic part, g_0. On the other hand, the quench analysis of the system with a zero total rotational angular momentum allows determining g_0. The reason is the lack of the dependence of eigenstates with $J = 0$ on the anisotropic part of the intermolecular interaction. Therefore, to extract the full information about intermolecular interactions, one should start with the interaction quench performed in the system with $J = 0$, interpreted as in fig. 3.18(a), followed by the investigation of the quench dynamics of the system with $J = 1$ or higher.

The spin-rotation coupling

Figure 3.19 presents the nonequilibrium dynamics of the magnetization, $\langle \hat{S}_z \rangle$, of the analyzed molecular system with molecular electric spins 1/2 with zero projection of the total angular momentum, $M_{\mathrm{tot}} = 0$, after the quench of external electric or magnetic field with different spin-rotation coupling strengths, γ. To see a nontrivial time evolution of $\langle \hat{S}_z \rangle$, i.e., its value changing in time with a significant amplitude, the initial state, $|\Psi_0\rangle$, must be coupled to a subset of the final eigenfunctions, $\left|\tilde{\Psi}_j\right\rangle$, with $\left\langle \tilde{\Psi}_j \right| \hat{S}_z \left| \tilde{\Psi}_{j'} \right\rangle \neq 0$. These conditions are not met in the system without the spin-rotation coupling, as it is the only part of the Hamiltonian, which mixes states with different projections of the total spin angular momentum, M_S. Therefore, in general, the smaller γ, the smaller amplitude of $\langle \hat{S}_z(t) \rangle$ after a quench of any external field. However, the presence of the spin-rotation coupling is not a sufficient condition for the nontrivial dynamics. The quench of the external fields must also be performed in the vicinity of the magnetization change, discussed in section 3.2.1, otherwise the value of $\left\langle \tilde{\Psi}_j \right| \hat{S}_z \left| \tilde{\Psi}_{j'} \right\rangle$ becomes negligibly small.

Panels (a) and (b) of fig. 3.19 present the time evolution of the magnetization, $\langle \hat{S}_z \rangle$, for the system under the constant impact of the external magnetic field of $3\,\hbar\omega$, after the quench of the electric field from 0 to 7.5 $\hbar\omega$, for the medium ($\gamma = 1\,\hbar\omega$) and large ($\gamma = 3\,\hbar\omega$) spin-rotation coupling strengths, respectively. The corresponding magnetization diagrams are presented in panels (a) and (b) of fig. 3.15. The initial state of the system, $|\Psi_0\rangle$, is antisymmetric with $J = 1$, $M_J = 1$, and $M_S = -1$. It is composed predominantly of the $|0\rangle$ harmonic trap state. Firstly, we see that the amplitude of $\langle \hat{S}_z(t) \rangle$ variation increases with the spin-rotation strength. For a medium γ, the $\langle \hat{S}_z \rangle$ changes by around 20% of its value, while for a large γ – by 100%. As already stated, the $\langle \hat{S}_z \rangle$ value would be constant without the spin-rotation coupling.

Secondly, the DFT in insets of panels (a) and (b) of fig. 3.19 indicates that the dynamics is governed by a manifold of couplings. They result from the intermolecular interaction mixing the harmonic levels, the magnetic field lifting the degeneracy with respect to the projection of the total spin angular momentum, M_S, the electric field mixing the rotational states, and finally the spin-rotation coupling, which mixes states with different J, S, M_J, and M_S. The quench aims to assess the spin-rotation coupling strength, γ. Therefore we compare results from panels (a) and (b). While the number of present couplings is vast, most of them are negligible, especially in panel (a). Their impact on the dynamics grows with the spin-rotation coupling strength. Therefore, while just two couplings dominate the time evolution of $\langle \hat{S}_z \rangle$ for a medium γ, they are joined by many new ones for

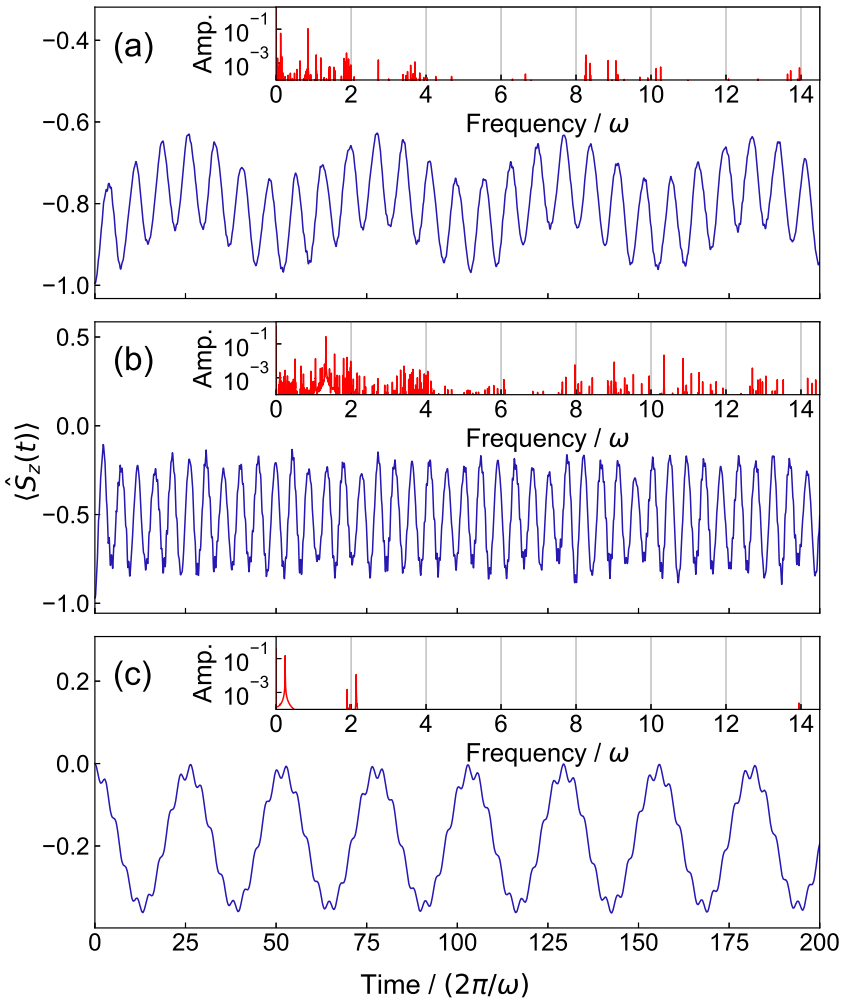

Figure 3.19: **Quench of external fields.** The time evolution of the magnetization, $\langle \hat{S}_z \rangle$, of the system of two interacting molecules with the spins 1/2 and rotational constants $B = \pi\hbar\omega$ in a one-dimensional harmonic trap, described by the medium isotropic and anisotropic interaction strengths, $g_0 = g_{\pm 1} = 4$, and with zero projection of the total angular momentum, $M_{\text{tot}} = 0$, after the quench of (a) the electric field from 0 to $d\mathcal{E} = 7.5\,\hbar\omega$ for a system with spin-rotation coupling $\gamma = 1\,\hbar\omega$ and with the constant magnetic field, $\mu_B\mathcal{B} = 3\,\hbar\omega$, (b) the electric field from 0 to $d\mathcal{E} = 7.5\,\hbar\omega$ for a system with $\gamma = 3\,\hbar\omega$ and constant magnetic field, $\mu_B\mathcal{B} = 3\,\hbar\omega$, and (c) the magnetic field from 0 to $\mu_B\mathcal{B} = 4\,\hbar\omega$ with $\gamma = 1\,\hbar\omega$ and the constant electric field $d\mathcal{E} = 5\,\hbar\omega$. Insets present the discrete Fourier transforms of the studied time evolutions. From Ref. [331].

a large γ. In both cases, the dominant coupling is between the states taking part in the magnetization change, described in section 3.2.1. The strength of this coupling grows with the spin-rotation coupling strength γ.

Summing up, in the case of the quench of the electric field, there are two signatures of the spin-rotation coupling strength in the time evolution of the magnetization, $\langle \hat{S}_z \rangle$. First is the size of $\langle \hat{S}_z(t) \rangle$ amplitude, which increases with γ. Second is the number of couplings present in the system and the amplitude of the dominant coupling, increasing with γ.

Instead of quenching the electric field strength, the magnetic field can also be suddenly turned on. Panel (c) of fig. 3.19 shows the time evolution of $\langle \hat{S}_z \rangle$ of the studied molecular system under the influence of the constant electric field of $5\,\hbar\omega$, after the quench of the magnetic field from 0 to $4\,\hbar\omega$. The initial state, $|\Psi_0\rangle$, is already impacted by the constant electric field. It is predominantly $|\Psi_0\rangle = |n = 0,\, J = 0,\, j_1 = j_2 = 0,\, S = 0,\, M_S = 0\rangle$, but mixed with the symmetric rotational state with j_1 and j_2 equal to 0 and 1. It also has a significant contribution from the higher harmonic states ($n = 2, 4$) due to the intermolecular interaction. The selected quench of the magnetic field does not modify the initial state significantly, so in the end we probe only couplings between $\left|\tilde{\Psi}\right\rangle \approx |\Psi_0\rangle$ and other eigenstates of the final Hamiltonian. This significantly limits the number of couplings influencing the dynamics, what is visible when comparing the corresponding DFTs in fig. 3.19. Moreover, the initial state has $\langle \hat{S}_z \rangle = 0$, therefore the only significant couplings are between $\left|\tilde{\Psi}\right\rangle$ and the eigenstates with $\langle \hat{S}_z \rangle \neq 0$, which further limits the number of visible peaks in the DFT. However, the remaining peaks are related to the spin-rotation coupling as it is the only part of the Hamiltonian mixing states with different M_S.

In the magnetic field's quench, the initial state $|\Psi_0\rangle$ has the largest overlap with the fourth excited state of the final Hamiltonian, instead of the ground state, as it is in the electric field's quench. This means that the couplings governing the dynamics are not between the states taking part in the system's magnetization change. Moreover, the amplitude of $\langle \hat{S}_z(t) \rangle$ variation is not anymore linearly dependent on the spin-rotation coupling strength. It grows with the spin-rotation coupling strength till γ reaches medium values, and then remains almost constant. Different sets of couplings govern these two regimes. The first regime (small γ, $\langle \hat{S}_z(t) \rangle$ amplitude $\propto \gamma$) is dominated by a single coupling between $\left|\tilde{\Psi}\right\rangle \approx |\Psi_0\rangle$ and the nearest antisymmetric eigenstate with $J = M = 1$, $M_S = -1$, and $n = 2$. This coupling strength grows with γ till γ reaches medium values. In the second regime (larger γ, constant $\langle \hat{S}_z(t) \rangle$ amplitude), the mentioned coupling strength decreases when γ grows and a new coupling, with the ground state of the quenched system, grows. Competition between two couplings re-

sults in the $\langle \hat{S}_z(t) \rangle$ amplitude being nearly independent of the spin-rotation strength. This quench scenario is thus less straightforward to determine the spin-rotation coupling strength value than the quench of the electric field. It still allows us to determine γ by fitting the theoretical model to the experimental data.

The quench of the magnetic field strength proves useful also in determining whether the anisotropic part of the intermolecular interaction dominates the system's properties. As discussed in section 3.2.1, in the system with medium interaction anisotropy, the main source of coupling between states with different M_S values is the combination of the electric field and the spin-rotation coupling. When the electric field is missing, the main source of such a coupling is the large anisotropic part of the intermolecular interaction which brings states with higher rotational angular momenta to lower energies (see fig. 3.16(c)). It means that if the quench scenario from fig. 3.19(c) is performed without the constant electric field, the nontrivial time evolution of $\langle \hat{S}_z \rangle$ indicates that the anisotropic part dominates the intermolecular interaction.

3.2.3 Conclusions

Within this work, we have studied the magnetic properties of two interacting ultracold polar and paramagnetic molecules in a one-dimensional harmonic trap. We have focused on the interplay of the molecular electronic spins, electric dipole moments, rotational structures, external electric and magnetic fields, and spin-rotation coupling. We have shown that control over the molecular system's magnetization could be achieved using an external electric field. This result is a complementary extension of the analogous studies focused on the free-space collisions. We have also presented the resulting magnetization diagrams depending strongly on two molecular properties of the system, namely the spin-rotation coupling and the anisotropic part of the intermolecular interaction. Motivated by the theoretical and experimental challenges in determining such molecular properties of few-body systems, we have employed the quench dynamics to find signatures of the anisotropic intermolecular interaction strength and the electronic spin-rotation coupling.

Our findings can be summarized as follows:

- The magnetization of the system can be controlled via external fields. The main underlying mechanism is the competition between the Zeeman and Stark effects. The spin-rotation coupling strength affects the smoothness of the transition between possible magnetization values.

- The number of accessible magnetization values depends on selected M_{tot} of the system, the electronic spins of the molecules, and the strength of the anisotropic part of the intermolecular interaction as it brings the states with higher total rotational momenta to lower energies.

- The time evolution of the system's cloud size after the quench of the intermolecular interaction has clear signatures of the ratio between the anisotropic, $g_{\pm 1}$, and isotropic, g_0, part of the interaction. In the regime of large g_0 or $g_0 = g_{\pm 1}$, the dynamics is governed by couplings between evenly separated harmonic states of the system. For $g_{\pm 1} > g_0$, the ladder of additional couplings becomes visible in the Fourier transform of the time evolution, coming from the antisymmetric ground state of the system. This ground state is highly sensitive to $g_{\pm 1}$ and may be used to determine its strength compared to g_0.

- The time evolution of the magnetization after the electric field's quench depends strongly on the spin-rotation coupling strength. The larger spin-rotation coupling, the larger is the amplitude of the magnetization variation and the larger number of couplings governing the dynamics. It can thus be used to assess the strength of the spin-rotation coupling in the molecular system.

- The time evolution of the magnetization after the magnetic field's quench is governed by a smaller number of couplings than after the electric field's quench. In the studied example, it is caused by a large similarity of the initial state to one of the eigenstates of the system after the quench. The dynamics probes then only couplings to this single eigenstate. While it may allow to assess the spin-rotation coupling strength, this scenario serves better for probing the anisotropic part of the intermolecular interaction.

The presented intrinsic coupling between the electric and magnetic properties of the studied model system paves the way towards studying the controlled magnetization of the ultracold many-body molecular systems trapped in optical tweezers or optical lattices. The results provide also the first step in studying dynamical magnetic properties of a few-body molecular systems with varied geometries. The potential applications range from quantum simulations of molecular multichannel many-body Hamiltonians to quantum information storing.

The studied model can be extended by including the fermionic or bosonic statistics of indistinguishable molecules or allowing dimers to be different.

Another direction is to incorporate the state dependence of molecular characteristics and trapping potential. The interaction potential with more realistic dependence on the relative distance between molecules may capture the physics of four-atom complexes that are now of central interest for ultracold molecular experiments. Another extension is the more realistic quench dynamics taking into account all correlations and dependencies between molecular characteristics. A natural extension to the many-body limit is the double molecular Mott insulator in an optical lattice with two molecules per site. The present system constitutes exotic monomers for such a system with large total rotational angular momenta in the ground state and magnetization controllable with the electric field.

Part II

Interpretable machine learning for phase detection

4

Machine learning methods

Artificial intelligence is the new electricity.
~Andrew Ng

In section 1.3 we have already hinted that ML methods are powerful numerical tools that soon can join the physicists' toolbox for scientific discovery. Before it can happen, we need to develop ways to interpret and increase the reliability of ML models and incorporate them into our everyday work. Within this thesis, we present the first steps towards this ambitious goal. In particular, in chapter 5, we show how we can enhance phase-detection schemes by applying interpretability and reliability tools based on the Hessian of the training loss.

Before we dive into the results, we devote this chapter to familiarize the reader with the ML concepts and methods that we then apply to tackle phase classification problems in chapter 5. Therefore, we do not provide here a thorough and deep introduction to ML or deep learning (DL). For this, we refer the reader to excellent textbooks [352–354] or tailored for physicists Refs. [195, 196].

Here, we start by listing the ingredients of the ML problem (task, data, model) in section 4.1. We discuss briefly types of ML tasks in section 4.2 with a special focus on those addressed within this thesis. In section 4.3 we introduce the supervised and unsupervised learning that are most commonly used for phase classification and detection. In sections 4.4 to 4.6, we describe the learning process in general and used optimization methods in particular. Then, in section 4.7 we discuss the selected ML models that we have employed in our work, namely neural networks (NNs), with a focus on convolutional neural networks (CNNs) and autoencoders (AEs). In

particular, we explicitly present NN architectures used in chapter 5.

In section 4.8 we discuss the loss landscape of DL models and how a usual minimum reached within the optimization looks like. Finally, in section 4.9 we dive into the Hessian-based toolbox of interpretability and reliability methods, namely influence functions ($\mathcal{I}$s) in section 4.9.1, relative influence functions (Relat-$\mathcal{I}$s) in section 4.9.2, resampling uncertainty estimation (RUE) in section 4.9.3, and local ensembles (LEs) in section 4.9.4.

4.1 How to make computers learn

Making machines think is a long-lived dream of human civilization. Artificial creations becoming alive have been a part of our mythology since ancient times. The more we know about intelligence, the more challenging it seems. But on the way to true artificial intelligence (AI), humans have developed impressive machinery which is already revolutionizing our everyday life, industry, and science.

We want to make computers think so they solve problems that are out of our reach. A broad definition of AI is a problem-solving computer. In general, we want machines to solve any problem, but we noticed quite early that humans and computers have a disjoint set of tasks in which they naturally excel. This observation was made in the 1980s and called *Moravec's paradox*. As Moravec wrote in 1988, "it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility" [355].

There are tasks that are easy for computers but difficult for humans. These are problems that can be described by a list of formal, mathematical rules. Therefore, computers excel at solving logic problems, algebra, geometry, and finding an ideal move in chess. To solve these kinds of problems, we can use a knowledge-based AI with hard-coded solutions.

But we also want a computer to tackle problems that are not easy to present in a formal mathematical way (such as face recognition) or whose exact mathematical formulation is not known (detection of new quantum phases). The solution is to allow a computer to learn from experience (or data). This data-based idea gave birth to *machine learning (ML)*. Therefore, the extraction of patterns from data is at the core of ML. This field can be thought of as the applied statistics with increased emphasis on the use of computers to estimate complicated functions and decreased emphasis on proving confidence intervals around these functions [353]. If these patterns are extracted by detecting easier patterns first and then building more complex ones via the hierarchy of concepts, we enter the *deep learn-*

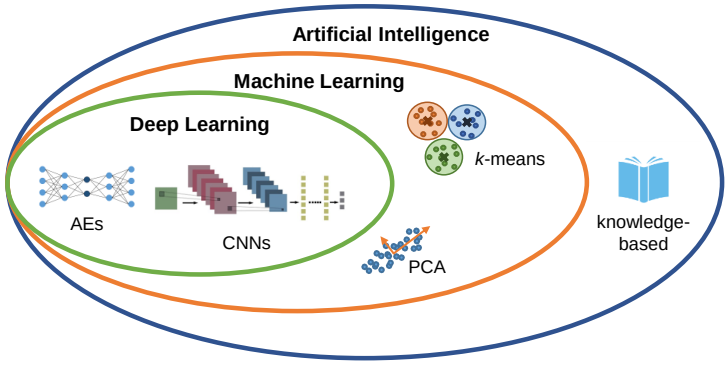

Figure 4.1: **Sketch of the relation between AI, ML, and DL** with examples from each field including k-means, principal component analysis (PCA), autoencoders (AEs), and convolutional neural networks (CNNs).

ing (DL) regime. The relation between these three fields (AI, ML, and DL) is presented schematically in fig. 4.1.

To make a computer learn, we need three main ingredients:

1. a *task* to solve (section 4.2),
2. *data* which gives an equivalent of *experience* and allows for solving the task (section 4.3),
3. a *model* which learns how to solve the task (section 4.7).

To check whether a computer successfully learned how to solve a task, we need also to define a performance measure, e.g., an error. Effectively, the learning process can be described as minimizing an error of a data-guided model performing a task.

4.2 Types of tasks

Let us look at the first learning ingredient which is the task. The first large type of tasks is *regression*. In this setting, we typically assume an immediate relation between the two variables of interest. Thus, we seek to express one of the real-valued variables $\boldsymbol{y}$ (which we call the output, also known as target) in terms of the other one $\boldsymbol{x}$ (the input or sometimes also referred to as *features*). The regression objective is to find the function f that ideally yields the mapping $\boldsymbol{y} = f(\boldsymbol{x})$ for all possible tuples of $(\boldsymbol{x}, \boldsymbol{y})$.

The example of a regression task within this thesis is the anomaly detection scheme described in section 4.7.3 and applied to topological experimental data in section 5.3.2. Other examples from quantum sciences include the prediction of potential energy surfaces in quantum chemistry [356, 357] or parameter estimation of Hamiltonians given measurement data [225].

The second large common ML task is *classification*. Its general idea is to use an algorithm to assign (discrete) class labels to examples. In contrast to regression, we are optimizing a model to find a mapping from an input vector $\boldsymbol{x}$ to a target $\boldsymbol{y}$ which now is a string of binary bits instead of real numbers. Each bit of the vector $\boldsymbol{y}$ corresponds to a specific pre-defined class. This encoding of the output vector is called *one-hot encoding* and can be used for an arbitrary number of classes. The values of the output vector are 0 except for one element that corresponds to a respective class which has the value 1. For example, in a K-class problem, instead of having a label with K possible values such as $y_i = 1, 2, \ldots, K$, each label is encoded as a K-element vector with all-zero elements except for one at the index corresponding to the class. For example, $\boldsymbol{y}_i = [0, 0, 1, \ldots, 0]$, means a sample i belongs to the third class as only $y_{i,3}$ is non-zero. The easiest example of classification is the binary classification where the algorithm has to distinguish between two classes, e.g., true or false. If the task includes more than two classes, we enter the multi-class classification, e.g., distinguishing between different species of iris of the Iris [358] data set or handwritten numbers of the famous MNIST [359] data set. A popular example from physics is the classification of classical and quantum phases of matter, including also those of a topological nature, discussed already in section 1.3.1. Other examples are classification subroutines in the automatization of quantum experiments [223].

Another common example of an ML task is *clustering*. It consists of grouping examples based on their similarity in the input space. In other words, clustering algorithms aim at grouping data points in such a way that examples in the same groups are more similar to other data points in the same group than those in other groups. Such groups are called *clusters*. We perform such clustering in section 5.5.4. Clustering can be thought of as an example of *dimensionality reduction* that aims at finding a simpler representation of input data that preserves maximum information contained in the data.

Finally, ML is also used for *anomaly detection*. In this type of task, the algorithm goes through examples and flags some of them as being unusual or atypical. A real-life example of an anomaly detection task is credit card fraud detection. In the case of quantum sciences, it can be the detection of anomalous or especially noisy quantum measurements. Within this thesis, we have two instances of anomaly detections. One is performed with influ-

ence functions ($\mathcal{I}$s) in section 5.5.6, the other with autoencoders (AEs) in section 4.7.3. The latter anomaly detection scheme with AEs is also briefly described in section 4.7.3.

4.3 Types of learning

So far, we have discussed examples of ML tasks. The second learning ingredient is the *data* whose accessibility also determines the type of learning we have to consider. These two ingredients (task and data) are intertwined: certain tasks can only be solved if sufficient data is available and, in turn, a richer data set allows to tackle different tasks. We refer to data in terms of a data set $\mathcal{D}$ containing a finite amount of (tuples of) data instances, i.e., $\mathcal{D} = \{\boldsymbol{z}_i\}_{i=0}^{n}$, where $\boldsymbol{z}_i$ is either $\boldsymbol{x}_i$ or $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ and n is a training data set size. The input data is coming from some input space $x_i \in \mathcal{X}$. While the notation is clear, there is much less convention on how the data should be *represented*. We refer to every element in each $\boldsymbol{x}_i$ as a *feature*. The central question in ML is how to represent the data and its features? This is the core of the representation learning on which we touch upon by means of autoencoders (AEs) in section 4.7.2 and especially by using experimental data pre-processing in section 5.5.2. Lastly, data can be identified with experience and a model can have various access to it. This access defines the type of learning. Within this thesis, we discuss only two of three main ML paradigms: supervised learning in section 4.3.1 and unsupervised learning in section 4.3.2. However, we acknowledge a powerful and successful approach that is reinforcement learning, applied, e.g., to gaming [360, 361], quantum control [362–364], quantum circuit optimization [365], or quantum error correction [366].

4.3.1 Supervised learning

Supervised learning can be performed when an ML model has access to labeled data, i.e., $\mathcal{D} = \{\boldsymbol{z}_i\}_{i=0}^{n}$, where $\boldsymbol{z}_i = (\boldsymbol{x}_i, \boldsymbol{y}_i)$. The input data comes from some input space $\boldsymbol{x}_i \in \mathcal{X}$, and the labels - from an output space, $\boldsymbol{y}_i \in \mathcal{Y}$. This type of learning has an obvious drawback that is its need for data labeling which is usually done by a human. On the other hand, as we show in section 5.5 it is also a more informed type of learning.

Both regression and classification are tackled with supervised approaches. Anomaly detection can also be performed in a supervised way. The supervised learning is especially straightforward in phase classification [243, 246, 248, 256–259, 262, 270, 276], where the inputs $\boldsymbol{x}_i$ are some representations of states of a given physical system, and $\boldsymbol{y}_i$ are the corresponding phase

labels. In our set-ups in sections 5.2 and 5.3, the inputs $\boldsymbol{x}_i$ are the ground-state wave functions of the one-dimensional spinless Fermi-Hubbard model at half-filling, obtained with an exact diagonalization of 12-site or 14-site Hamiltonian and expressed in the Fock basis. On the other hand, in section 5.5, the input data are time-of-flight images of ultracold atoms.

4.3.2 Unsupervised learning

When we do not have labeled data, we can turn to unsupervised learning. In this case, we have unlabeled training data $\mathcal{D} = \{\boldsymbol{z}_i\}_{i=0}^{n}$, with $\boldsymbol{z}_i = (\boldsymbol{x}_i)$. The unsupervised approach is especially tempting in case of unknown phase diagrams and so far has been widely used for known phase classification problems. Unsupervised learning of phase transitions can roughly be divided in two categories: clustering-based methods [250, 252, 265, 279, 367–371] and learning-success-based methods [244, 254, 266, 271, 372]. Moreover, if ML is to guide the scientific discovery, it needs to have unsupervised elements.

What underlies the unsupervised learning is finding the "best" representation of the data, where "best" can mean different things. In general, the best representation of data preserves as much information about $\boldsymbol{x}$ as possible while obeying constraints aimed at keeping the representation simpler than original $\boldsymbol{x}$. Typical tasks tackled in an unsupervised way are dimensionality reduction such as clustering and anomaly detection.

Clustering aims at finding a sparse simple representation of the data set, where each data point belongs to a single group or cluster based on their similarity. We can therefore think of this method as searching for optimal one-hot encoding representation of $\boldsymbol{x}$. A popular clustering algorithm is the (naive) k-means that groups examples into k clusters whose number is pre-defined by a user. It starts by random placement of k centroids in the input space. The centroids are centers of to-be-formed clusters. Then, the algorithm assigns each data point to its nearest centroid (in an Euclidean sense) forming k clusters. Then, the centroids are moved to the center of mass of their respective cluster. Now, the examples may have another centroid that is the nearest to them, so they get reassigned to their closest centroids again. The procedure gets repeated until the position of centroids converges as presented in fig. 4.2. We apply k-means in section 5.5.4.

Another way of reducing the dimensionality of data is the *principal component analysis (PCA)*. PCA identifies mutually orthogonal directions in the data space along which the linear correlation in the data vanishes, called principal components. Each principal component can be then ranked based on the variance of the data in the corresponding direction. To reduce the dimensionality of the space and preserve maximum information, we can discard the principal components along which the data shows the least

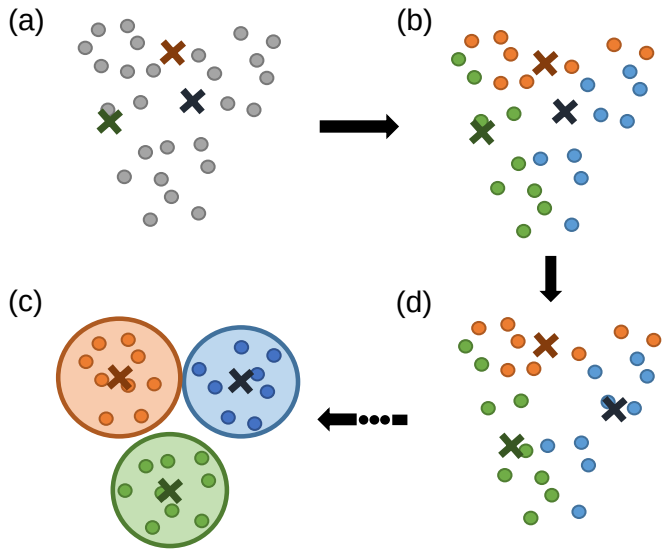

Figure 4.2: **Scheme of the k-means algorithm.** (a) Random initialization of centroids. (b) Assigning data points to the nearest centroid. (c) Moving centroids to the centers of their respective clusters. (d) Repeating the procedure till convergence of centroids' positions.

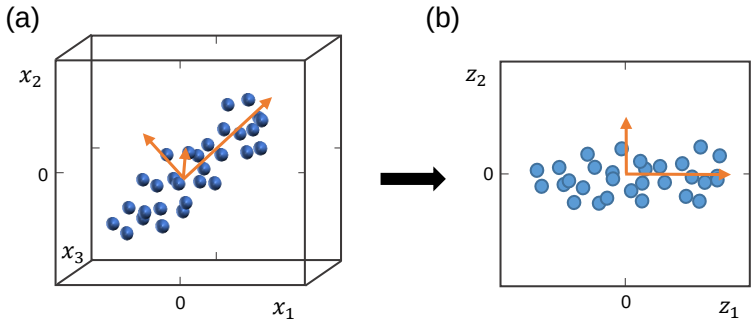

Figure 4.3: **Scheme of the principal component analysis (PCA).** (a) Finding orthogonal principal components along which data variance is different. (b) Projecting three-dimensional data onto two-dimensional space following their simpler representation which includes only two principal components with the largest variance.

variance. As such, in PCA, directions along which the data exhibits the *largest variance* are considered to contain the most important information. After reducing the dimensionality, in the new representation space, data may cluster which may be further analyzed with clustering methods. We apply PCA in section 5.5.5.

4.4 Learning is an optimization problem

We have already discussed that ML solves various tasks (e.g., classification or regression) and that the learning process can be either supervised or unsupervised. The final ingredient is a model which learns how to solve the task. In general, it is a function of the input data, $f(\boldsymbol{x})$, whose output is interpreted as a prediction made for the input data. The form of the output (prediction) depends on the task. It can be, e.g., a discrete-valued class in the classification task or a real-valued output in the regression task. Finding the function which provides the best mapping between data and the desired outcome for a specific task is at the heart of ML. We start with declaring a certain parametrization of a model (function), e.g., $f(\boldsymbol{x}) = \sum_i^m \theta_i x_i$. Then, all possible parametrizations of this function form the set of functions, i.e., the hypothesis class. Section 4.7 presents specific examples of the hypothesis classes (or spaces), namely neural networks (NNs), but for now, we focus on the learning process itself.

The underlying process of learning consists in finding an optimal model $\hat{f} = f_{\tilde{\boldsymbol{\theta}}}$ in the hypothesis space with optimal parameters, $\tilde{\boldsymbol{\theta}}$, that minimizes the target loss function.[1] The training starts with randomly initialized parameters.[2] Then, machines "learn" by varying parameters in such a way to minimize the loss function of the training data, i.e., all the data accessible to the ML model during the learning process. The loss function formula varies between tasks and there is a certain freedom of how it can be chosen. In general, the loss function compares model predictions or a developed solution against reality or expectations. Therefore, learning becomes an optimization problem. We show the learning process schematically in fig. 4.4 on the example of the supervised paradigm (so with labeled training data).

[1] In this thesis, we use terms of loss, error, and cost functions interchangeably following Ref. [353]. Their definitions are not strict. Following Goodfellow: "The function we want to minimize or maximize is called the objective function, or criterion. When we are minimizing it, we may also call it the cost function, loss function, or error function. (...) some machine learning publications assign special meaning to some of these terms". For example, a loss function may be defined for a single data point, and a cost/error function may be a sum of loss functions.

[2] It is not entirely random. They are initialized with zero mean and constant variance across layers, otherwise we may face problems with vanishing or exploding gradients [373].

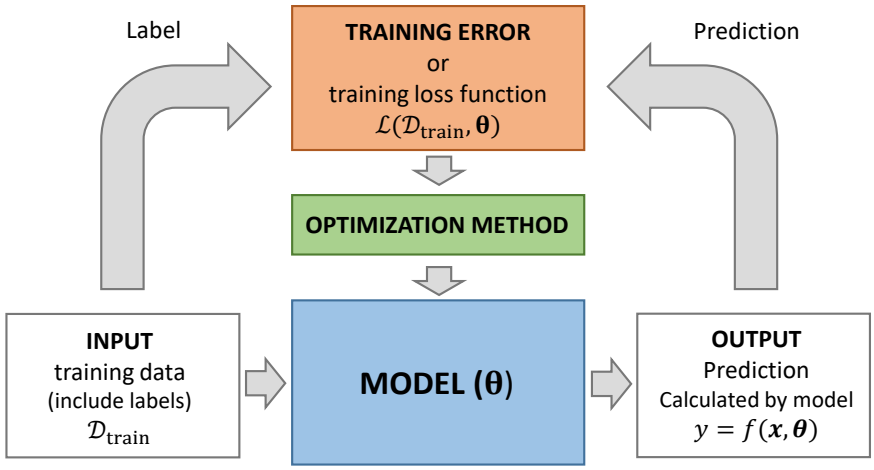

Figure 4.4: **Supervised training of an ML model.** The training starts with randomly initialized model parameters, $\boldsymbol{\theta}$. Such a model makes predictions at training points, and its predictions are compared against the ground-truth labels within the training loss function. Then, the optimizer varies $\boldsymbol{\theta} \to \boldsymbol{\theta}'$ to minimize the training error. The model with $\boldsymbol{\theta}'$ makes predictions again, and the training procedure repeats till the training error converges.

Popular examples of loss functions include mean-squared error (MSE) and categorical cross-entropy (CCE), used for real-valued and classification problems, respectively:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i))^2\,, \tag{4.1}$$

$$\mathcal{L}_{\text{CCE}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{K} y_{i,c}\cdot\log\left(f(\boldsymbol{x}_i)\right)\,, \tag{4.2}$$

where K is the number of classes. Note that eq. (4.2) requires one-hot encoding of labels. Values of every loss function depend on the model (which enters into formulas via predictions) and the data set. They are also normalized by the number of data points to compare their values between problems with different data set sizes.

Once we choose a loss function, we can minimize it by varying the parameters of the ML model using any optimization method. In general, we can divide optimization methods into analytical (rarely useful in DL), gradient-based, and gradient-free approaches. A popular example of a gradient-based

method used also within this thesis is *gradient descent*. The optimization starts usually in a random place within the loss landscape (meaning with a model with randomly initialized parameters, $\boldsymbol{\theta} = \boldsymbol{\theta}_0$). Using the model with $\boldsymbol{\theta}_0$, one can compute the loss function on training data. The next step is to compute gradients of the loss function with respect to each model parameter, θ. The final step is to update the parameters by subtracting the respective gradients multiplied by a learning rate, η, i.e., $\theta_j := \theta_j - \eta \frac{\partial}{\partial \theta_j} \mathcal{L}$. These steps need to be repeated till the minimum is reached, and each repetition is called an *epoch*. The intuition is that the gradient descent updates model parameters by making steps towards the minimum of the function (so in the opposite direction than the gradient which indicates where the function value grows).

The learning rate controls the size of these steps. Figure 4.5 presents in a simplified way the importance of the η choice. Both too large and too small η's make optimization more challenging, and only optimal η promises efficient convergence to a minimum. There is rarely an obvious way of choosing η which, therefore, has to be found, e.g., by trial and error. As such, the learning rate is one of the so-called *hyperparameters* of the learning process. Hyperparameters are parameters whose values control the learning process (especially speed and quality) and are chosen by a user (in contrast to model parameters, which are derived via training). The number of epochs or the choice of the loss function are hyperparameters too. We encounter more examples of hyperparameters in this introductory section. To find optimal hyperparameters, a good practice is to form (next to the training data set) a separate *validation data set*. Then, we can set various hyperparameters and choose them in such a way as to minimize the error on the validation set. You can even use optimization methods to find optimal hyperparameters which minimize the validation error as we do in section 5.5.2 with the optuna library [374] but the choice of hyperparameters guided by a proper intuition may prove to be a faster and cheaper approach.

The optimization picture that we drew in the previous paragraph along with fig. 4.5 may suggest that the loss landscape in ML problems is convex. Especially for DL, it cannot be further from the truth. The loss landscapes in DL are highly non-convex to the degree that makes the scientific community wonder how can we even find minima in this hyperspace? Two immediate problems arise from the non-convexity. How not get stuck in local minima corresponding to large loss function values or in saddle points of such landscapes? Secondly, are some minima better than others? The learning dynamics is a still ongoing research direction. For now, we can make a popular modification to the gradient descent algorithm and discuss so-called *mini-batch stochastic gradient descent (SGD)*. This optimization method

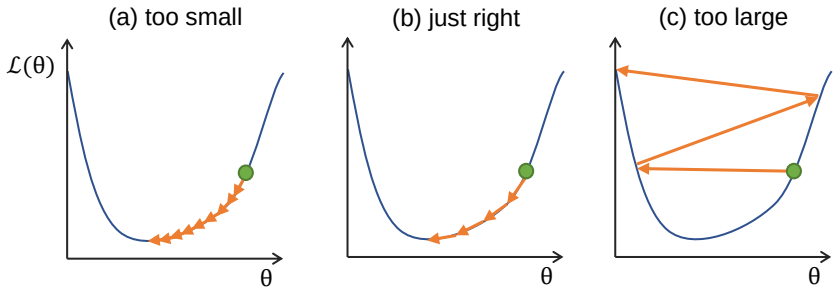

Figure 4.5: **Choosing a learning rate has an impact on convergence to the minimum.** (a) If η is too small, the training needs many epochs. (b) The right η allows for fast convergence to a minimum and needs to be found. (c) If η is too large, optimization can take you away from the minimum (you "overshoot"). This figure suggests that the loss function is convex which is rarely true.

consists in computing the loss function at each epoch on randomly selected mini-batches of training data. It means that at each epoch the gradients may point in (slightly) different directions. The resulting stochasticity helps in escaping saddle points and narrow local minima.[3] Moreover, computing loss function and gradients only for a mini-batch of data instead for the whole data set provides a nice computational speed-up for large data sets. Finally, we can add so-called *momentum* to the mini-batch SGD. The idea is closely related to momentum in physics, and it consists in remembering the update of parameters at each iteration and determining the next update as a linear combination of the gradient and the previous update [376, 377]. As a result, SGD with momentum is less prone to getting stuck in narrow minima and has less noisy estimates of parameter updates. Within this thesis, when performing supervised learning of CNNs, we use mini-batch SGD with momentum as an optimization method.

So far the only gradient-based optimization method we have described is mini-batch SGD with momentum. Other popular examples include Adam [378, 379] or L-BFGS algorithm [380]. There are also gradient-free optimization approaches that are used especially when the gradients or loss function itself are expensive or impossible to compute, e.g., when optimizing experiments. Examples include genetic algorithms, particle swarm optimization, random search and simulated annealing [381].

[3] In practice, stochasticity is helpful in avoiding saddle points but there are theoretical works showing it is not a necessary condition for a proper convergence [375].

4.5 Generalization and regularization

So far, ML may seem like a cleverly named function fitting. However, the heart of ML lies in the *generalization* which is the ability to make correct predictions on new data, i.e., those not seen during the training. Checking whether your model generalizes well requires an additional data set which is the *test set* composed of data points which are used neither for optimization of model parameters nor for searching for the best hyperparameters describing the learning process. Therefore, the original full data set needs to be separated into the training, validation, and test sets (a usual ratio is around 8:1:1).

Lower model performance on the test set compared to the training set (so higher test error than training error - the difference is called the generalization error) is a common problem in ML, even when all data points are generated by an identical probability distribution.[4] The main reason is the large capacity of ML models.[5] The *capacity* can be loosely understood as the ability of a model to fit a variety of functions. When the model capacity is much higher than one needed to solve the task, the model tends to *overfit*, i.e., memorize all possible properties of the training set which may not be true for the general distribution (and in particular, the test set). In particular, the model can even fit the noise in the training data. As a result, overfitting increases the test error while keeping training error low (or even decreasing it). Optimal capacity allows for the lowest gap between the test and training error, so the lowest generalization error. However, the capacity which is too low results in an overly constrained model which can *underfit*, i.e., have a high training error. The intuition behind the under- and overfitting is schematically shown in fig. 4.6.

Therefore, we can improve the generalization of the model by controlling its capacity. Every modification of the model aiming to improve the generalization of a model (possibly at the cost of the increased training error) is called a *regularization* technique. A straightforward way of restricting the model capacity is to limit the magnitude of its trainable parameters which effectively limits the hypothesis space of a parametrized model. It can be done by adding a penalizing term to the training loss function which increases with the parameters' values. Such an approach is used within the two popular regularization techniques, i.e., L_1 and L_2 regularization [353]. We use L_2 regularization on ML models used within this thesis.

[4] To mathematically study the relation between training error and test error, you usually need to start by making so-called *i.i.d. assumptions*, i.e., that the examples in each data set are independent of each other, and that the train set and test set are identically distributed, drawn from the same probability distribution as each other [353]

[5] DL models are able to fit large datasets with random labels [287]!

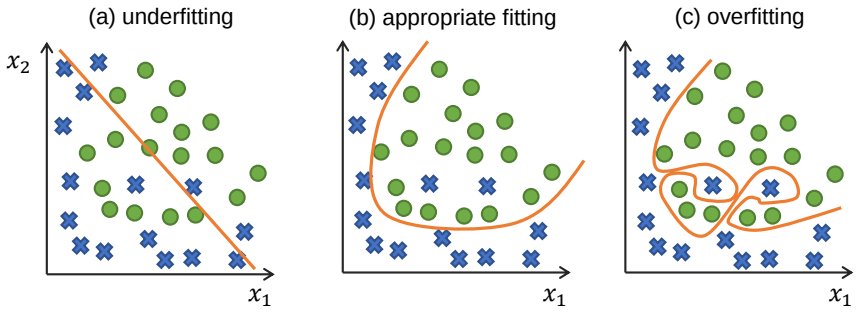

Figure 4.6: **Schematical picture of under- and overfitting.** (a) When the model capacity is too low, the model cannot fit the training data properly. (b) With the model capacity corresponding to the task complexity, the fitting is optimal. (c) When the model capacity exceeds the task complexity, the model tends to overfit and the generalization error increases.

4.6 Test loss

We have already encountered a concept of the training loss function. Finally, we can introduce the test loss. An ML model, f, is determined by the set of d parameters $\boldsymbol{\theta} = \{\theta_0, \ldots, \theta_{d-1}\}$. For a given input $\boldsymbol{x}$, the model outputs a real-valued K-dimensional vector, $\boldsymbol{f_x} = f(\boldsymbol{x}; \boldsymbol{\theta})$. The output encodes the prediction of the model, $y' = \text{argmax}(\boldsymbol{f_x})$. For example, for a two-class problem, $\boldsymbol{f_x}$ could be $[0.1, 0.9]$ which would correspond to predicting a label $y' = 1$. In a supervised scheme, the loss function, $\mathcal{L} = (\boldsymbol{x}, \boldsymbol{\theta})$, compares a model's output, $\boldsymbol{f_x}$ with a ground-truth label y of the corresponding input $\boldsymbol{x}$. $\mathcal{L}$ is small when the predicted label y' is the same as the ground-truth label y. Moreover, $\mathcal{L}$ gets smaller, the larger are differences between the value of the element corresponding to y and other elements' values in the model's output, $\boldsymbol{f_x}$. For example, the $\mathcal{L}$ would be smaller for $\boldsymbol{f_x} = [0.1, 0.9]$ than for $[0.45, 0.55]$, even though the predicted label is the same in both cases. For this reason, the elements of the output vector $\boldsymbol{f_x}$ tend to be connected to probabilities of the input belonging to the corresponding classes. However, this interpretation can be misleading in the presence of data set shift [382, 383] or non-uniformity of errors [384]. Training ends when a minimum of $\mathcal{L}$ is found, and parameters at this minimum are $\tilde{\boldsymbol{\theta}}$.

After training, the model can make a prediction at an unseen test point $y'_{\text{test}} = \text{argmax}(\boldsymbol{f}_{\text{test}})$, where $\boldsymbol{f}_{\text{test}} = f(\boldsymbol{x}_{\text{test}}, \tilde{\boldsymbol{\theta}})$ with a test loss function $\mathcal{L}(\boldsymbol{x}_{\text{test}}, \tilde{\boldsymbol{\theta}})$ as presented in fig. 4.7. Within this work, we use two versions of the test loss. The "ground-truth" version is the standard test loss defined

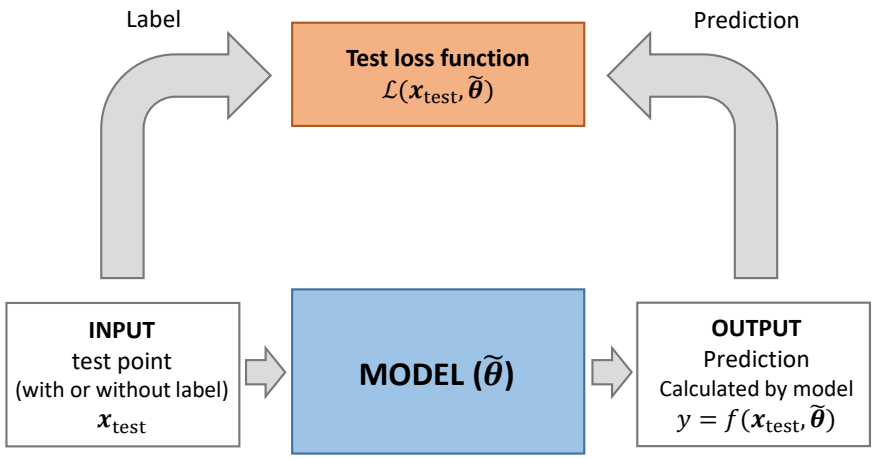

Figure 4.7: **Testing of an ML model.** A trained ML model determined by $\tilde{\boldsymbol{\theta}}$ makes a prediction at a test point, $\boldsymbol{x}_{\text{test}}$, which has not been seen during training. The prediction is compared against the ground-truth label (resp. against the predicted label) via the ground-truth (resp. minimal) test loss function.

in supervised problems and compares the output of the model, $\boldsymbol{f}_{\text{test}}$ with the ground-truth label y_{test} of $\boldsymbol{x}_{\text{test}}$. When the ground-truth label of a test point is unavailable, one can use a "minimal" version of the test loss. It compares the model's output $\boldsymbol{f}_{\text{test}}$ to the model's predicted label y'_{test}. We stress we use it only during the test stage to imitate the real-life situation when we ask an ML model for predictions at test points we do not know ground-truth labels for.

4.7 Neural networks

The loss function $\mathcal{L}(\mathcal{D}, \boldsymbol{\theta})$ is deeply connected to the problem type and the chosen learning approach. The choice of a model $f(\boldsymbol{\theta})$ however is more of a methodological nature, and different models can be combined with different types of learning. One of the most popular and flexibe models are NNs. A basic unit of an NN is a *neuron* taking inputs being features of a single data point $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ with some weights $\boldsymbol{w}$ and modified by neuron-specified bias b and producing a scalar output $y(\boldsymbol{x}, \boldsymbol{w}, b)$. Such a neuron is presented in fig. 4.8. The exact value of the output y depends

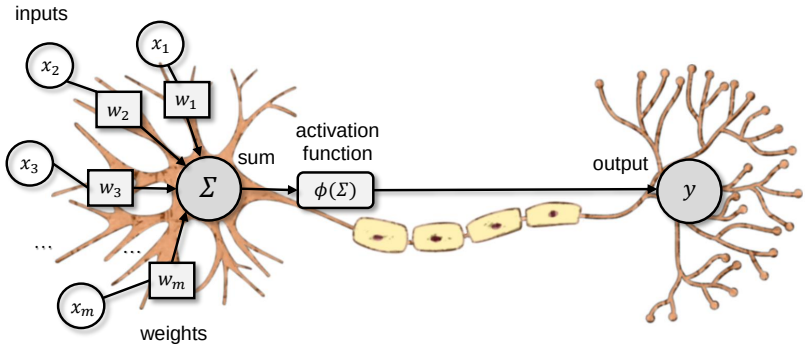

Figure 4.8: **Scheme of an artificial neuron** composed of inputs, their weights, a corresponding sum, an activation function, and an output against the simplified sketch of a biological nerve cell.

also on a chosen non-linear activation function ϕ:

$$y(\boldsymbol{x}, \boldsymbol{w}, b) = \phi\left(\sum_{i}^{m} w_i x_i + b\right) . \tag{4.3}$$

In the past, common choices of activation functions included step functions (like in perceptron) and sigmoids. However, the discontinuous behavior of the step-functions derivative in perceptrons makes it impossible to train them using gradient descent, and the sigmoid functions saturate quickly for big input values. To avoid these drawbacks, rectified linear units (ReLUs) are often used, where $\phi(x) = \max(0, x)$.

The basic idea of all NNs is to layer neurons in a hierarchical way. The first layer in an NN is called the input layer, the middle layers are often called *hidden layers*, and the final layer is called the output layer [195]. The outputs of the input layer are then treated as the inputs to the next hidden layer. Deep NNs have multiple hidden layers, and this procedure is repeated then several times until the top or output layer is reached. The whole NN can be thus thought of as a complicated non-linear transformation of the inputs $\boldsymbol{x}$ into an output y that depends on the weights and biases of all the neurons in the input, hidden, and output layers. Learning consists in minimizing the loss function by varying the parameters $\boldsymbol{\theta}$ of the model, here the weights $\boldsymbol{w}$ and biases $\boldsymbol{b}$. If every output of one layer is connected to every input of the following layer, as presented in fig. 4.9, the NN is called to be *fully connected* (or dense).

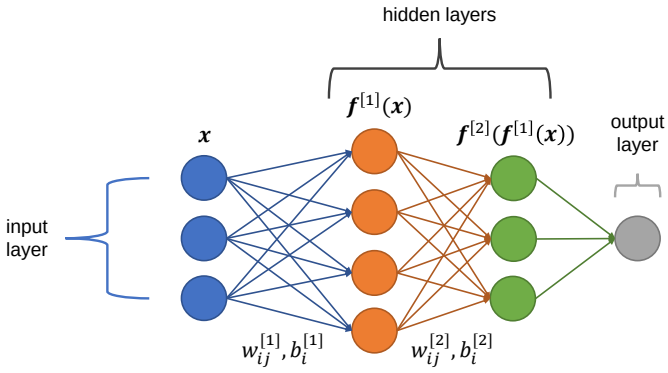

Figure 4.9: **Illustration of a typical fully-connected NN.** Basic components of NN are neurons consisting of a linear transformation that weights the importance of various inputs, followed by a non-linear activation function. Neurons are arranged into layers with the output of one layer serving as the input to the next layer.

The use of hidden layers greatly expands the representational power of NNs. As the universal approximation theorem states, an NN with a single hidden layer and arbitrary activation functions can approximate any continuous, multi-input/multi-output function with arbitrary accuracy [385] limited by the number of hidden units. The basic idea behind the proof is that hidden neurons allow NNs to generate step functions with arbitrary offsets and heights. These can then be added together to approximate arbitrary functions. In physics, a good analogy is matrix product states, which can approximate any quantum many-body state to arbitrary accuracy, provided the bond dimension can be increased arbitrarily. While it is understandable why the larger number of hidden units in a single layer may be preferable, it is still not clear why the larger number of hidden layers is favorable for learning [195], however, it is thought to allow NNs to learn more complex features from data.

NNs are just one out of many schemes used in ML, however, thanks to their hierarchical nature, they are believed to be one of the most powerful and flexible architectures. Other architectures include linear classifiers (e.g., logistic regression), support vector machines, random forest, nearest neighbors, and decision trees [302, 386]. Within this thesis, we use a type of NNs, namely CNNs.

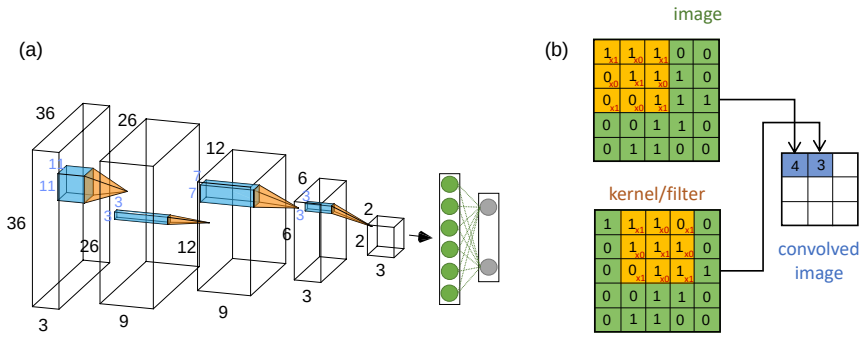

Figure 4.10: **Convolutional neural network (CNN).** (a) CNN scheme with convolutional, pooling, and fully-connected layers for an RGB image of 36x36 pixels. (b) Kernel (or filter) in a convolutional layer makes a convolution out of the scanned image. The convolved image is fed to next convolutional layers.

4.7.1 Convolutional neural networks

Fully-connected NNs, as presented in fig. 4.9, are powerful models, but the number of weights and biases (parameters), that during learning has to be varied in order to minimize the loss function, grows quickly with the size of the input. It is a problem especially in computer vision, as even an ordinary grey-scale image of a handwritten digit, being an element of MNIST dataset - analog of "Hello, world!" for the ML community, is encoded in 28x28 pixels, resulting in the input size of 784. If a colored image of 256x256 pixels is considered, the network needs to have an input layer containing 196 608 values!

Convolutional neural networks (CNNs), presented in fig. 4.10, are models used primarily in the image recognition, and their characteristics are convolutional and pooling layers that noticeably reduce the number of parameters, at the same time taking advantage of a hierarchical composition of the network as well as the recognition of the spatial relation of the data. In a *convolutional layer*, a small matrix (called kernel or filter) is passed over the input matrix (image) to create a feature map for the next layer [387]. The dimensions of the kernel are hyperparameters that have to be set before the training. The values of the feature map are computed either by taking the sum of the result of an element-wise multiplication of the kernel and an appropriately sized section of the input matrix or more often, the element-wise multiplication is replaced by a dot product. In con-

volutional layers, parameters that are be varied during the training, are the filters' values, and kernels found in such a way may extract meaningful information from images. One of the techniques to improve CNNs consists in using multiple kernels in a given convolutional layer and concatenating results to create the feature map.

Next to the convolutions, so-called pooling layers are used. They reduce the number of parameters in each layer. The most common pooling procedure, Max pooling, consists simply in taking the largest number from a chosen region of data, and the shape and size of this region are also the hyperparameters. Another pooling technique that we use within this thesis is global average pooling (GAP) consisting in taking only an average over the whole feature map. CNNs usually have also at least one fully-connected layer, as described in the previous section, that performs a traditional classification (or regression) based on the features extracted by the previous layers. It is important to note, that even with the described methods designed to reduce the parameters number, modern CNNs can have millions of them, ranging from 7 millions for GoogLeNet (Inception v1) [388] to 138 millions for VGGNet [389].

Within this thesis, we use three CNN architectures. Figure 4.11 shows them all schematically. Panel (a) presents a CNN consisting of three one-dimensional convolutional layers with five filters on the input vector, eight filters on the first hidden layer, and ten filters for the last convolution layer. After the first two convolutions, we apply a max-pooling layer to reduce the dimension, and the last convolutional layer is followed by an average pooling layer. Finally, we have one fully connected layer with two output neurons that predict the labels. When designing the architecture, we make sure that the convolutional part contains a large part of the NN's parameters. We use the architecture presented in panel (a) of fig. 4.11 for phase classification in the one-dimensional spinless 12-site Fermi-Hubbard model in section 5.2 of this thesis as well as in sections 5.3.2 and 5.3.3. Thanks to its relatively small size (only 720 parameters), its Hessian-based analysis is quick and efficient. However, it is designed for inputs of size 924, which is the size of eigenvectors of the 12-site Fermi-Hubbard model. To have an ML model which is invariant to the input size, we design a CNN architecture with a global average pooling (GAP) layer, which reduces the size of each input filter to one. In fig. 4.11(b), we list the sizes of convoluted data passing through the model for two input sizes, 924 and 3432, corresponding to 12- and 14-site eigenfunctions of the Fermi-Hubbard model. After the GAP layer, the sizes of convoluted data are the same, which shows how the size-invariance is reached. We use this model in sections 5.3.1 and 5.3.4. Finally, to analyze two-dimensional time-of-flight images in sections 5.5.3 and 5.5.7, we employ a two-dimensional CNN as presented in fig. 4.11(c).

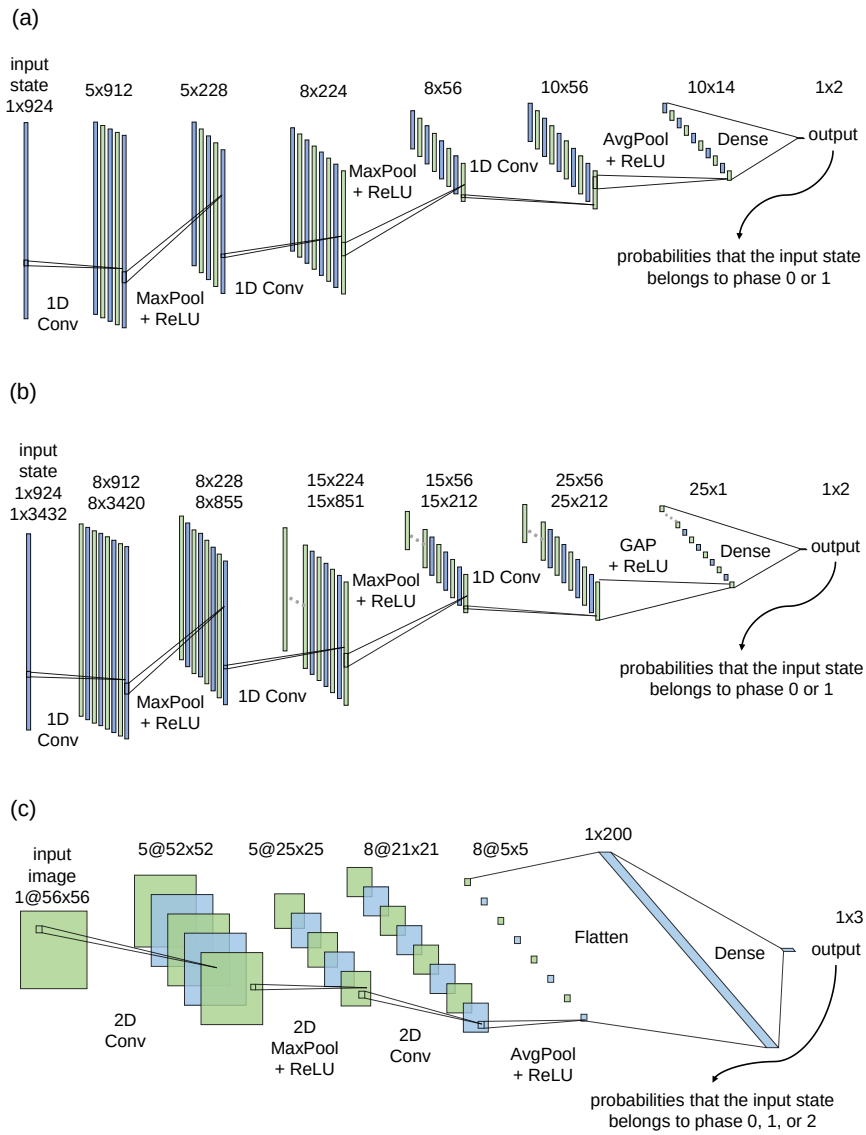

Figure 4.11: **Architectures of CNNs used in this work.** (a) A one-dimensional CNN has been applied to the classification of eigenstates of the 12-site one-dimensional Fermi-Hubbard model. CNN has 720 parameters. (b) A one-dimensional CNN with global average pooling (GAP) is largely invariant to the input size and has been applied to the classification of eigenstates of the 12-site and 14-site one-dimensional Fermi-Hubbard model. CNN with GAP has 1675 parameters, regardless of the input size. (c) A two-dimensional CNN has been employed to study time-of-flight images of 56x56 pixels.

4.7.2 Autoencoders

Autoencoders (AEs) are NNs with special architecture containing a bottleneck in the middle [390, 391] that are trained to reconstruct the input at the output. The architecture of such an AE is depicted in fig. 4.12. This bottleneck architecture is based on two NNs called encoder and decoder. Such a NN is trained by minimizing the so-called error reconstruction loss. As such, the optimal setup for such encoder-decoder pair is the one for which the output $\boldsymbol{x}_{\text{rec}}$ is reconstructed as similar as possible to the original input data $\boldsymbol{x}$. The possible loss function here is MSE. Due to the bottleneck, the information passing through the network needs to get compressed at the bottleneck, and then decompressed to recover the input. As a consequence of the compression, some information is lost. However, the retained information in the bottleneck should ideally contain everything relevant for the reconstruction of the input. Therefore, the bottleneck forms a *latent space*[6] which contains a compressed representation of the input data. We can think of it in terms of dimensionality reduction schemes that preserve the most important features for the reconstruction. As a result, we can analyze the latent representation of the input data in a similar unsupervised way as the lower-dimensional representation of input data obtained by a PCA in section 4.3.2. We do it in sections 5.5.1 and 5.5.4 with a convolutional AE, i.e., where next to fully-connected layers as shown in fig. 4.12, there are also convolutional layers described in the previous section.

However, an AE can also be trained in a supervised way given a series of inputs associated with the corresponding output. Such a supervised method has applications in image denoising or colorization [392–394].

4.7.3 Anomaly detection with autoencoders

Analysis of the AE latent representation of the input data is not the only way of an AE-based unsupervised phase classification. Another successful and more robust scheme called anomaly detection has been presented in Ref. [254]. The idea is based on the following intuition. Imagine training an AE to reconstruct states coming from one phase. Then ask it to reconstruct states coming from the rest of the phase diagram. Such a task is difficult as the training data is limited only to one phase, and the AE is bound to make reconstruction errors on other phases. Moreover, we expect that the error is lower for phases that are similar to the "training" phase and higher for different phases. Finally, the transition regimes are usually distinctive in the phase diagram in terms of how quantum states look like. Altogether,

[6] A latent variable is a random variable that we cannot observe directly. In this case, we call variables latent because we do not observe them in the data.

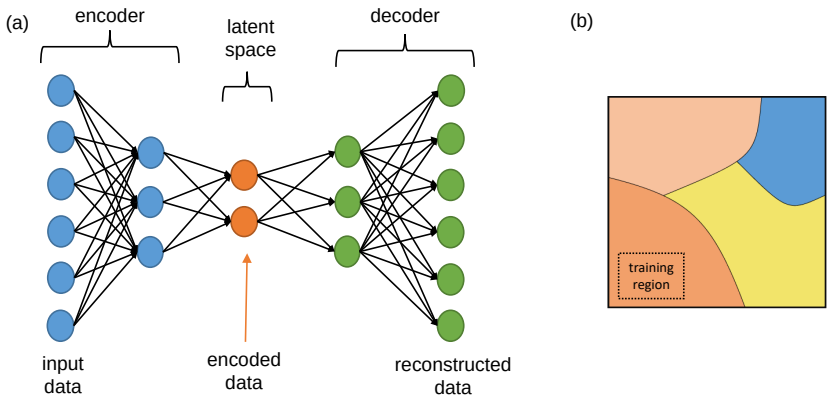

Figure 4.12: **Autoencoder (AE) and AE-based anomaly detection scheme.** (a) Scheme of an AE architecture with a two-neuron bottleneck. (b) A phase diagram with a color-coded similarity between phases. Imagine training an AE to reconstruct the states from the orange phase (dashed box). Then ask the trained AE to reconstruct all other states. Its error will be small in the same orange phase, a little higher in the light orange and yellow phases, and highest in the blue phase.

the reconstruction error across the phase diagram, made by an AE trained to reproduce states from one phase, is expected to vary according to the phases in the system. This scheme leads to the discovery of phases in a fully unsupervised way. The authors of Ref. [254] used the anomaly detection approach to recover a full phase diagram of the extended Bose Hubbard model in one dimension at exact integer filling. Interestingly, their work also revealed a phase-separated region (between supersolid and superfluid parts) with unexpected properties which may be one of the first fully unsupervised discoveries in the ML-guided phase classification.

A similar approach was proposed to detect anomalous events within data produced by the Large Hadron Collider [235]. We use the AE-based anomaly detection scheme in section 5.5.6.

4.7.4 Variational autoencoders and a question neuron

The architecture of an AE presented in section 4.7.2 can be further modified. One of the approaches is to make probabilistic observations on the latent space. Therefore, rather than building an encoder that for an input $\boldsymbol{x}$ outputs single latent features $z_i(\boldsymbol{x})$, the encoder can output a probability distribution for each latent attribute, $p(z_i|\boldsymbol{x})$. Then, the decoder starts the reconstruction by sampling z_i from the latent probability distributions. Such AEs are called variational autoencoders (VAEs) [395, 396].

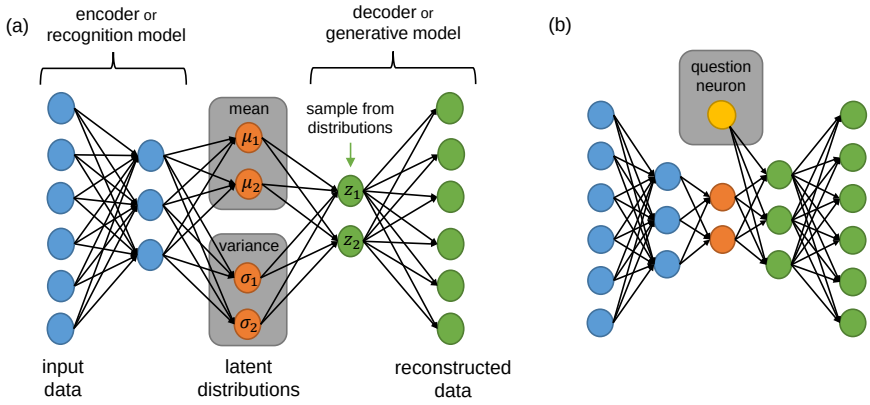

Figure 4.13: **Variational autoencoder (VAE) and an AE with a question neuron.** (a) Scheme of a VAE architecture with a bottleneck encoding two probability distributions of two latent variables. (b) A question neuron can be added to an AE.

Their architecture is presented in fig. 4.13(a). Change of the encoding results also in a modified loss function. In the case of regular AEs, the training consists in minimizing the reconstruction loss which may be formulated as the mean-squared error (MSE). In VAEs, next to the reconstruction loss, training aims to minimize also the difference between the latent probability distribution (which we usually assume is Gaussian) and the true probability distribution (which can have any form).

Such a modification to the architecture introduces three main advantages. Firstly, the data reconstruction is usually more stable. Secondly, the feature space is regularized such that neuron activations at the bottleneck are more interpretable [397]. Finally, this way we can generate new data after training by sampling at the bottleneck. Therefore, in VAEs, the decoder part is sometimes called a generative model, while the encoder part is a recognition model.

Another modification that can be done to an AE architecture is adding a *question neuron*, which is an extra input neuron feeding directly into the bottleneck as presented in fig. 4.13(b). The information provided by the question neuron can be, e.g., a physical parameter corresponding to the image we provide. In section 5.5.2, we use VAEs with a question neuron to post-process the data to a fixed micromotion phase.

4.8 Hessian and curvature

As described in section 4.4, the training loss landscape of deep NNs is highly non-convex. This renders the optimization problem difficult, e.g., due to the presence of many local minima as presented in fig. 4.14(b). Moreover, these minima may not have equally good generalization properties which can be connected to the curvature around a minimum. For example, there is a general consensus that wide, flat minima generalize better than sharp minima [398–401]. Keep in mind that flatness is not a well-developed concept in non-convex landscapes of deep models [402]. An analysis of how the generalization ability depends on the local curvature is an example of a case where the shape of the reached minimum can tell us something useful about trained ML models.

Let us look into the minimum reached during the optimization of DL models with more detail. To do that and to describe the curvature around such a minimum, we use a Hessian of the training loss function, $H_{\tilde{\boldsymbol{\theta}}}$, i.e., a matrix of second derivatives of $\mathcal{L}$ with respect to the model parameters, calculated at the minimum, $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$:

$$H_{\tilde{\boldsymbol{\theta}},ij} = \frac{\partial^2}{\partial\theta_i\theta_j}\mathcal{L}_{\text{train}}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}\,. \tag{4.4}$$

The eigenvectors of $H_{\tilde{\boldsymbol{\theta}}}$ corresponding to the largest positive eigenvalues indicate directions with the steepest ascent around the minimum as presented in fig. 4.14(a).

In contrary perhaps to the common intuition, the training of an ML model leads to a *local minimum* or a *saddle point* [403–405]: the vast majority of the eigenvalues is close to zero, indicating various flat directions and some small negative eigenvalues are also present, indicating directions with negative curvature. One can wonder why we should trust a model which does not land in the global minimum. A series of empirical results as well as applying spin-glass theory to DL [406] indicate, among others, that for large-size networks, most local minima are equivalent and yield similar performance on a test set. Also, the probability of finding a "bad" (high value) local minimum is non-zero for small-size networks and decreases quickly with network size. Finally, struggling to find the global minimum on the training set (as opposed to one of the many good local ones) is not useful in practice and may lead to overfitting, i.e., much better performance on the training set than the test set, which is equivalent to bad generalization.

The directions with high curvature, i.e., those aligned with eigenvectors corresponding to large eigenvalues, are the directions along which the training data strongly determines the model parameters. In other words, if the model parameters are varied in directions of the steepest ascent around the

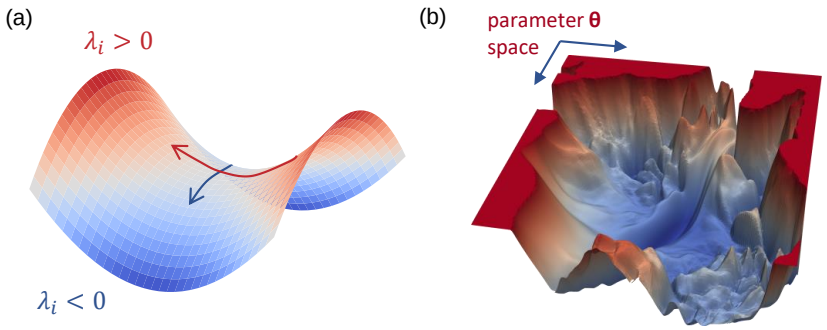

Figure 4.14: **Hessian, curvature, and loss landscape.** (a) Saddle point with directions of positive (ascending) and negative (descending) curvature. The corresponding Hessian eigenvalues, λ_i, are positive or negative. (b) Low-dimensional visualization of a non-convex loss landscape of a deep NN called VGG-56 trained on CIFAR-10 [410].

minimum, i.e., along the eigenvectors corresponding to the largest positive eigenvalues of the Hessian, the value of the training loss function changes the most. It means that these directions are bounded the most by the training data. There are also two empirical observations supporting this claim. Firstly, numerical simulations on the example of a one-dimensional nonlinear regression problem show that gradients of training examples lie in the directions of the highest curvature [407]. Secondly, the analysis of the Hessian spectrum shows that in DL problems the number of directions with a significant ascent around the minimum is equal roughly to the number of classes in the problem minus one [404, 408, 409].

Therefore, there are intimate relations between curvature around the minimum, model generalization, and training data. The knowledge of the curvature around the minimum also allows us to approximate how our ML model (and as a result, its predictions) would change upon some action. Possible actions could be the removal of a single training point or a slight modification of $\tilde{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}^*$, resulting in a shift to an adjacent minimum with identical training error. The study of how a model reacts to such actions is at the heart of the Hessian-based toolbox, which contains influence functions ($\mathcal{I}$s) [411], the resampling uncertainty estimation (RUE) [407], and local ensembles (LEs) [412] whose conceptual ideas we introduce in the following.

Finally, note that a non-positive curvature around the minimum, while it does not affect the quality of the model predictions in general, may cause numerical problems when working with the Hessian. We discuss these challenges in section 4.9.5.

4.9 Hessian-based toolbox

The four interpretability and reliability methods discussed in this thesis are all based on local perturbations of the loss function. They study how a particular action, e.g., removal of a training point, changes the model parameters. This change, in turn, impacts the prediction of the model at a test point, y'_{test}. The change of the model parameters caused by some action can be approximated using the Hessian matrix of the empirical risk (training loss) calculated at the minimum of the loss landscape at $\tilde{\boldsymbol{\theta}}$, namely $(H_{\tilde{\boldsymbol{\theta}}})_{ij} = \partial^2_{\theta_i\theta_j}\mathcal{L}(\mathcal{D}, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$.

All the methods we study in this work approximate how the change of parameters impacts the model predictions, but the reason for the change of parameters is different for each method. Influence functions ($\mathcal{I}$s) and relative influence functions (Relat-$\mathcal{I}$s) study the removal of a single training point from a training data set, resampling uncertainty estimation (RUE) analyzes training on various samples of the training data set, while local ensembles (LEs) modify model parameters in the flat directions of the Hessian. These methods are presented schematically in fig. 4.15 and aim to answer different questions regarding the reliability and interpretability of the model. We discuss them in detail in the following sections.

4.9.1 Influence functions and similarity

Leave-one-out (LOO) training. Let us consider a model trained on n training points and making a prediction at a test point. Now, we remove a single training point $\boldsymbol{z}_{\text{r}}$ from the training set $\mathcal{D}$, $\mathcal{D} \to \mathcal{D}_{\setminus \boldsymbol{z}_{\text{r}}}$, retrain the model, and check the influence of this removal on the test loss. If the prediction is now worse (resp. better), i.e., the test loss is higher (resp. lower), then $\boldsymbol{z}_{\text{r}}$ is a helpful (resp. harmful) training example for this specific test point. If the prediction stays the same, $\boldsymbol{z}_{\text{r}}$ is not influential to this prediction. With such an analysis, called leave-one-out (LOO) training, we can therefore judge how influential a certain training point is for a test prediction. Figure 4.16 presents schematically this procedure.

Influence functions ($\mathcal{I}$s). Retraining the model is, however, expensive, and an approximation of the LOO training was proposed and named influence functions ($\mathcal{I}$s) [415–417]. It was then ported to ML applications by Koh & Liang [411, 418]. The $\mathcal{I}$ reads

$$\mathcal{I}(\boldsymbol{z}_{\text{r}}, \boldsymbol{z}_{\text{test}}) = \frac{1}{n}\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{z}_{\text{test}}, \tilde{\boldsymbol{\theta}})^T H^{-1}_{\tilde{\boldsymbol{\theta}}} \nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{z}_{\text{r}}, \tilde{\boldsymbol{\theta}}) \equiv \frac{1}{n}\nabla\mathcal{L}^T_{\text{test}} H^{-1}_{\tilde{\boldsymbol{\theta}}} \nabla\mathcal{L}_{\text{r}}\,, \quad (4.5)$$

and it estimates the change of the test loss for a chosen test point $\boldsymbol{z}_{\text{test}}$ after the removal of a chosen training point $\boldsymbol{z}_{\text{r}}$. $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{z}_{\text{test}}, \tilde{\boldsymbol{\theta}})$ is the gradient of

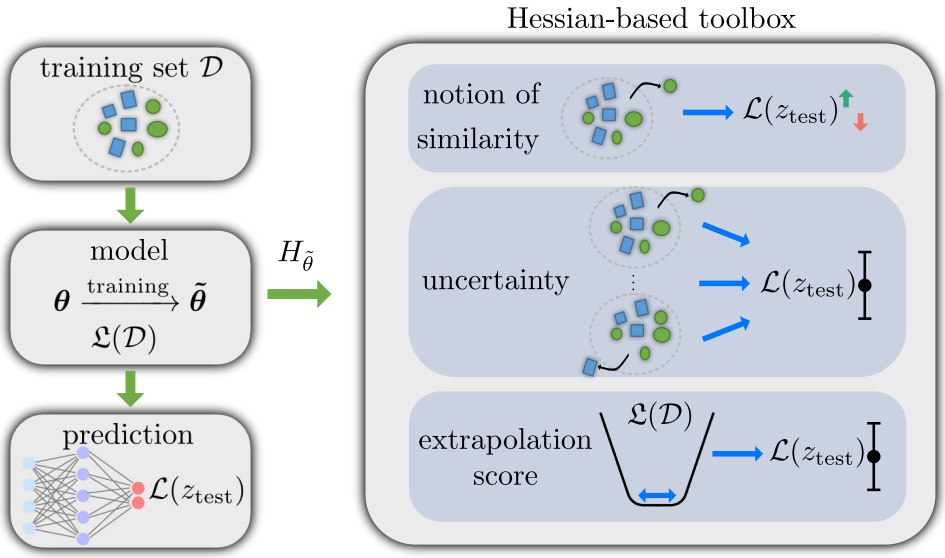

Figure 4.15: **Hessian-based toolbox.** The scheme of this study's scope. The ML problem starts with a model depending on parameters $\boldsymbol{\theta}$. Training a model consists in finding optimal parameters $\tilde{\boldsymbol{\theta}}$ which minimize a training loss function, $\mathcal{L}(\mathcal{D})$, calculated for the training data set, $\mathcal{D}$. The Hessian of the training loss at the minimum, $H_{\tilde{\boldsymbol{\theta}}}$, describes the curvature around the minimum and is a basis for four methods which provide the notion of similarity, i.e., influence functions ($\mathcal{I}$s) and relative influence functions (Relat-$\mathcal{I}$s); estimation of uncertainty, i.e., resampling uncertainty estimation (RUE); and extrapolation score, i.e., local ensembles (LEs), of the model prediction. They give insight into the reliability and interpretability of the model after its training. From Ref. [413].

the loss function of the single test point, and $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{z}_{\text{r}}, \tilde{\boldsymbol{\theta}})$ is the gradient of the loss function of the single training point whose removal's impact is being approximated. Both are calculated at the minimum $\tilde{\boldsymbol{\theta}}$ of the training loss landscape.

Geometrical interpretation. The $\mathcal{I}$ in eq. (4.5) can be written as the inner product of $-\nabla\mathcal{L}_{\text{test}}$ and $-H_{\tilde{\boldsymbol{\theta}}}^{-1}\nabla\mathcal{L}_{\text{r}}$ [419], where the term $-H_{\tilde{\boldsymbol{\theta}}}^{-1}\nabla\mathcal{L}_{\text{r}}$ describes an approximated change in parameters $\tilde{\boldsymbol{\theta}} \to \boldsymbol{\theta}'$ due to the removal of the training point $\boldsymbol{z}_{\text{r}}$ (for a derivation, see appendix D). This formulation emphasizes the geometric interpretation of $\mathcal{I}$s, which is a projection of the approximated change in parameters due to the removal of a training point onto the test sample's negative loss gradient, see fig. 4.18(b). The term $-H_{\tilde{\boldsymbol{\theta}}}^{-1}\nabla\mathcal{L}_{\text{r}}$ can also be understood as a Newton step [353] towards a new minimum resulting from the removal of $\boldsymbol{z}_{\text{r}}$. Note that the same term involves scaling by the inverse of eigenvalues of $H_{\tilde{\boldsymbol{\theta}}}^{-1}$. In other words, we see that the

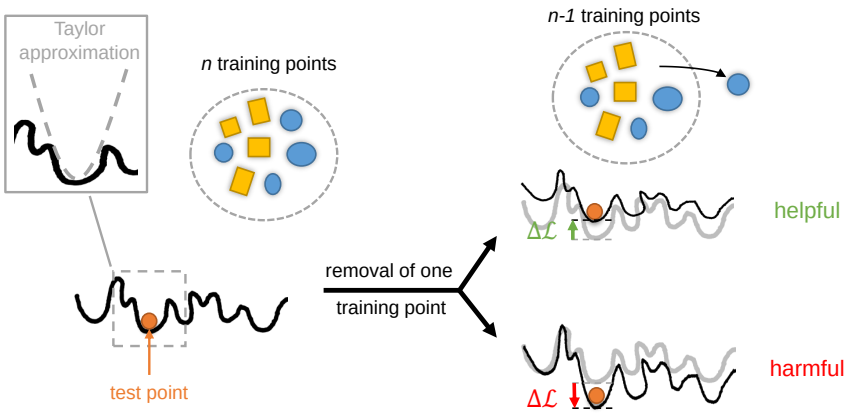

Figure 4.16: **Leave-one-out (LOO) training.** The scheme of LOO training whose approximation are influence functions illustrated using sketches of the trained loss function versus test samples. Consider removing a single training example from the training set and retraining the model. If the loss of the test point, X_{test}, increases (decreases), the removed training example is helpful (harmful) for making a prediction on X_{test}. Adapted from Ref. [414].

$\mathcal{I}$ is a scalar product of the gradients $\nabla\mathcal{L}_{\text{test}}$ and $\nabla\mathcal{L}_{\text{r}}$ accounting for a local curvature of the loss landscape described by $H_{\tilde{\boldsymbol{\theta}}}$. The resulting value of $\mathcal{I}$s depends on two factors: how similar are the test and the removed training point and how representative they are in the data set.

Similarity measure. Firstly, the more similar the test point and the removed training point are, the larger is the value of the $\mathcal{I}$ between them. More specifically, the largest influence is for the change in parameters which is in the direction of $\nabla\mathcal{L}_{\text{test}}$. It happens when the gradients $\nabla\mathcal{L}_{\text{test}}$ and $\nabla\mathcal{L}_{\text{r}}$ are aligned in the parameter space, corrected for the local curvature of the loss landscape, so when the test point, $\boldsymbol{z}_{\text{test}}$, is similar to the removed training point, $\boldsymbol{z}_{\text{r}}$. Note that by similarity here we understand the distance in the model's internal representation, so in the model's parameter space, corrected by the local curvature described by the Hessian. This similarity is different than, e.g., similarity as a distance of input vectors $\boldsymbol{z}_{\text{r}}$ and $\boldsymbol{z}_{\text{test}}$ in the input space $\mathcal{X}$ or the similarity in the Euclidean parameter space. In particular, the predictive model and especially neural networks can be highly nonlinear and may use an internal representation in which similar (close) points are far away in both the input and the Euclidean parameter space. We can then define a model's similarity measure between data points

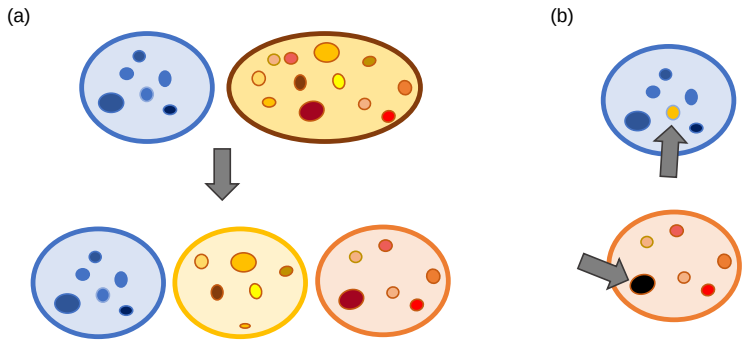

Figure 4.17: **Gains from the model-viewed data similarity.** (a) Detection of non-uniformity within the data. (b) Detection of anomalies and outliers.

$\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ equal to [407]

$$\begin{aligned} S(\boldsymbol{z}_i, \boldsymbol{z}_j) &= [\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{z}_i, \tilde{\boldsymbol{\theta}})^T H_{\tilde{\boldsymbol{\theta}}}^{-1} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{z}_j, \tilde{\boldsymbol{\theta}})]^2 \\ &\equiv [\nabla \mathcal{L}_i^T H_{\tilde{\boldsymbol{\theta}}}^{-1} \nabla \mathcal{L}_j]^2 \propto \mathcal{I}(\boldsymbol{z}_i, \boldsymbol{z}_j)^2 \,. \end{aligned} \tag{4.6}$$

Note that when having access to such a similarity concept, we can use it to our advantage (see fig. 4.17). For example, a study of similarity between data can indicate non-uniformity within classes or even plateaus of similarities. Finally, we can detect outliers and anomalies within both the training and test set.

Representative data and outliers. The second factor impacting the value of the $\mathcal{I}$ (and therefore the similarity measure) is the direction in which $\nabla \mathcal{L}_{\text{test}}$ or $\nabla \mathcal{L}_i$ lie. For example, the gradient may be aligned with the eigenvectors of $H_{\tilde{\boldsymbol{\theta}}}$ corresponding to the largest eigenvalues, which are the directions where the training data strongly determines the model parameters. Such an alignment happens for the most common or representative data points. The gradient also can point in the direction of one of many eigenvectors with almost zero eigenvalues, which may happen for distinct or unrepresentative data points called outliers. Due to projection onto the inverse of $H_{\tilde{\boldsymbol{\theta}}}$ and scaling by the inverse of corresponding eigenvalues, the $\mathcal{I}$ is larger for gradients pointing in the flat curvature of $H_{\tilde{\boldsymbol{\theta}}}$ than for gradients pointing to the high curvature. Therefore, the values of $\mathcal{I}$s between two data points are determined by how similar the two data points are from the model's perspective and how representative these data points are in the data set.

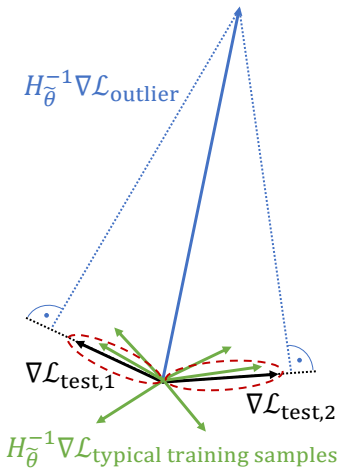

Figure 4.18: **Similarity and gradients of training and test losses.** All four methods address the change in the model's predictions due to various actions. This change can be approximated by analyzing the projection of the gradient of the test loss (black arrows) and the gradients of training points (green arrows) corrected for the local curvature described by the Hessian. An outlier (blue arrow) is here a training point being exceedingly different from an average one in a data set. Blue dashed lines are large projections of the outlier gradient corrected by the local curvature onto the gradients of the test points. From Ref. [413].

Sensitivity to outliers. A careful reader can notice that $\mathcal{I}$s as well as the similarity measure $S(\boldsymbol{z}_i, \boldsymbol{z}_j)$ may then be sensitive to outliers, i.e., data points with extreme values that significantly deviate from the majority of data points [420]. The removal of such an outlier can cause a large change in parameters. Therefore, the outlier is likely to have a large influence on a wide range of test samples, having a global effect on the test set. This global effect is visualized in fig. 4.18(b), where the blue gradient of the outlier projected onto the inverted Hessian space has large projection lengths with the gradients of two distinct test points. Conversely, the removal of a typical training example $\boldsymbol{z}_i$ whose gradient points towards high curvature of the Hessian, e.g., any of the green arrows in fig. 4.18(b), causes a small change in parameters and has a significant influence only for similar test points (circled in red in the figure). For more intuition on a simple example of Gaussian mixtures, we refer to fig. 5.3 in Ref. [421].

4.9.2 Relative influence functions

Barshan *et al.* [419] proposed a variant of $\mathcal{I}$s that takes advantage of the similarity measure but eradicates the influence of unrepresentative data points and outliers. The method is called Relat-$\mathcal{I}$ and restricts the pool of influential points to the most similar ones - see the gradients circled in red in fig. 4.18(b). Mathematically, it amounts to introducing a normalization to the $\mathcal{I}$'s formula

$$\mathcal{I}_{\text{Relat}}(\boldsymbol{z}_{\text{r}}, \boldsymbol{z}_{\text{test}}) = \frac{\mathcal{I}(\boldsymbol{z}_{\text{r}}, \boldsymbol{z}_{\text{test}})}{||H_{\tilde{\boldsymbol{\theta}}}^{-1} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{z}_{\text{r}}, \tilde{\boldsymbol{\theta}})||} = \frac{\frac{1}{n} \nabla \mathcal{L}_{\text{test}}^{T} H_{\tilde{\boldsymbol{\theta}}}^{-1} \nabla \mathcal{L}_{\text{r}}}{||H_{\tilde{\boldsymbol{\theta}}}^{-1} \nabla \mathcal{L}_{\text{r}}||} . \tag{4.7}$$

Both tools increase the interpretability of the ML model by indicating what the model regards as similar. The fact that $\mathcal{I}$s focus on the unrepresentative data points also allows one to judge the model's reliability by finding outliers in the training data set. The model's reliability is the central issue addressed by two other methods discussed in this work, namely resampling uncertainty estimation (RUE) and local ensembles (LEs).

4.9.3 Resampling uncertainty estimation

The resampling uncertainty estimation (RUE) [407] aims at assessing the uncertainty of predictions of the ML model. It can be applied to the already trained model and requires no specific architecture or learning scheme in contrast to, e.g., Bayesian methods, which are the most common approach in ML for assessing uncertainty [422, 423]. The RUE method makes use of two important criteria to judge whether a prediction is reliable: the density criterion and local fit criterion [407]. The density criterion states that a prediction at the input $\boldsymbol{z}_{\text{test}}$ is reliable if there are samples in the training data that are similar to $\boldsymbol{z}_{\text{test}}$. The local fit criterion states that a prediction at the $\boldsymbol{z}_{\text{test}}$ is reliable if the model has a small error on samples in the training data similar to $\boldsymbol{z}_{\text{test}}$. Both criteria hinge upon a measure of similarity defined in eq. (4.6) and can be addressed with bootstrap sampling.

To quantify uncertainty, the RUE algorithm makes b 'bootstrap' samples created by sampling with replacement from the uniform distribution over the original training data set. Let us start from the original data set $\mathcal{D}$ containing each training data point once, indicated with the short-hand notation by $\mathcal{D}[1, 1, \ldots, 1]$. We can then create a bootstrap sample by drawing the same point more than once and omitting others, e.g., $\mathcal{D}_b[2, 0, 3 \ldots, 1, 0]$, which stands for taking twice the first training example, omitting the second training example, etc. If an ML model trained on $\mathcal{D}[1, 1, \ldots, 1]$ converges to parameters $\tilde{\boldsymbol{\theta}}$, b ML models trained on b different bootstrap samples converge to similar model parameters $\boldsymbol{\theta}'_b$. Now we can make b predictions at

the same test point $\boldsymbol{z}_{\text{test}}$ with b ML models and calculate the variance of the test loss across these b models. Small variance means that a prediction of the original model can be trusted, while large variance means that it is not reliable. Intuitively, this variance estimates how much the model prediction would change if we fitted the model on different data sets drawn from the same distribution as the original training data. This intuition is connected to the idea of the classical bootstrap [424].

As for the LOO training, such a retraining procedure is prohibitively expensive. Therefore, one can make a similar approximation of the change of the model parameters due to the removal of some training examples and adding copies of others to the training set within a bootstrap sample. We can approximate the new parameters via

$$\boldsymbol{\theta}'_b \approx \tilde{\boldsymbol{\theta}} - H_{\tilde{\boldsymbol{\theta}}}^{-1} \cdot L \cdot w_{\Delta_b} \,, \tag{4.8}$$

where w_{Δ_b} is a vector of differences between the composition of the original $\mathcal{D}[1, 1, \ldots, 1]$ and $\mathcal{D}_b$. For example, for $\mathcal{D}_b[2, 0, 3, \ldots, 1, 0]$, the vector of differences w_{Δ_b} is $[1, -1, 2, \ldots, 0, 1]$. L is a matrix of all the n single training loss gradients w.r.t. every model parameter, so it has the size d x n, and it takes the form

$$L = [\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{z}_0, \tilde{\boldsymbol{\theta}}), \ldots, \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{z}_{n-1}, \tilde{\boldsymbol{\theta}})]^T \,. \tag{4.9}$$

In the next step, one generates predictions at a test point with b models with approximated parameters $\tilde{\boldsymbol{\theta}}_b$, obtaining b test losses based on $\mathbf{f}_b = f(\tilde{\boldsymbol{\theta}}_b, x_{\text{test}})$. Finally, we calculate the variance of b test losses, i.e., the average of the squared deviations from the original test loss.

4.9.4 Extrapolation score with local ensembles

We say the prediction at a test point is underdetermined if many different predictions are equally consistent with the constraints posed by the training data and the learning problem specification (i.e., the model architecture and the loss function). An example of such behavior is when a model trained on the same training data arrives at different predictions depending on the choice of a random seed and, therefore, relies on arbitrary choices outside the learning problem specification. Intuitively, underdetermination can be understood as the model converging during the optimization process to various distinct points in a flat basin forming a minimum. As discussed in section 4.8, the training data puts limited constraints on the flat directions around the minimum. However, changing the model parameters in these directions still can impact predictions at test points drawn from a different distribution than the training set, i.e., so-called out-of-distribution (OOD)

test points [425]. A reliable model should warn the user when making a prediction at such a test point.

Local ensembles (LEs) are a method to detect the underdetermination at test time in a pre-trained model [412]. LE consists of local perturbations of the parameters of the trained model that fit the training data equally well, i.e., have the same value of training loss. In other words, we perturb the parameters of the model only in the directions of the Hessian eigenvectors corresponding to close to zero eigenvalues, meaning we explore only flat basins around the minimum. Analogously to RUE, the next step is to make predictions at the same test point $\boldsymbol{z}_{\text{test}}$ with LE models and calculate the variance of the test loss within the LE. Madras *et al.* [412] found an even simpler approximation of this variance for a test point, $\boldsymbol{z}_{\text{test}}$, and named it a LE extrapolation score

$$\mathcal{E}_m(\boldsymbol{z}_{\text{test}}) = ||U_m^\top \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{z}_{\text{test}}, \tilde{\boldsymbol{\theta}})||_2 = ||U_m^\top \nabla \mathcal{L}_{\text{test}}||_2 \,. \tag{4.10}$$

U_m is a matrix of $(d - m)$ Hessian eigenvectors spanning a subspace of low curvature, i.e., after removing m eigenvectors corresponding to largest eigenvalues and, therefore, to directions with the highest curvature, which are well-constrained by training data. The authors of the method admit that choosing m is not a trivial task, with the danger of omitting under-constraint directions if m is set too high or including well-constraint directions if m is set too low. In this work, we choose the smallest possible m for which $\mathcal{E}_m$ starts to converge for all test points ($m = 12$ in fig. 5.9).

4.9.5 Practical aspects of the Hessian computation

A careful reader could have noticed two numerical challenges resulting from eqs. (4.5) to (4.10). Firstly, the calculation of $\mathcal{I}$s, Relat-$\mathcal{I}$, and RUE requires inverting the Hessian of a training loss which in DL is known to be highly non-convex [406]. As we have pointed out in section 4.8, optimization usually leads to a critical point with a majority of flat or almost flat directions (corresponding to eigenvectors with zero or close to zero eigenvalues) and a small number of directions of negative curvature (corresponding to eigenvectors with negative eigenvalues). The inverse of a matrix exists only if it is positive-definite (has only positive eigenvalues). Therefore, Koh & Liang proposed to add a damping term to the Hessian [411], λI, with I being the identity matrix and λ being larger than the absolute value of the largest negative Hessian eigenvalue, $|E_0|$. It is equivalent to L_2 regularization [419] and amounts to shifting all eigenvalues by λ, guaranteeing the existence of the Hessian inverse. Regardless of the exact value of the damping, the Hessian-based toolbox keeps giving meaningful results [411].

We usually choose $\lambda \approx |E_0| + 0.01$, except for RUE where the authors explicitly state that the smallest eigenvalue of the damped Hessian needs to be around one, rendering $\lambda_{\mathrm{RUE}} = \lambda + 1$.

Secondly, the calculation of the Hessian matrix for a model with a large number of parameters can be highly demanding. Fortunately, we do not need to calculate the full Hessian, only Hessian-vector products [e.g., $H_{\tilde{\boldsymbol{\theta}}}^{-1} \nabla \mathcal{L}_{\mathrm{r}}$ in eqs. (4.5) and (4.7)] or the top part of the Hessian spectrum, which significantly reduces the computational complexity of the problem [426]. The inverse of the Hessian for $\mathcal{I}$s, RUE, and Relat-$\mathcal{I}$ can be approximated with so-called stochastic approximation with LiSSA [426]. Additionally, the authors of Relat-$\mathcal{I}$ approximated the normalization factor in eq. (4.7) with a method related to K-FAC [427]. On the other hand, LE needs no inverse but an ensemble subspace of eigenvectors with zero or close to zero eigenvalues. The authors of LE proposed to use the Lanczos iteration [428] to calculate m eigenvectors with the largest eigenvalues, build an m-dimensional subspace (of highest curvatures), and create its orthogonal complement, namely the ensemble subspace (of flat directions). There is also a Python library called PYHESSIAN designed to tackle Hessian-based problems [429]. Within this paper, however, we calculate the Hessian explicitly, due to the limited size of our neural network.

5

Machine learning path - results

My advice is to go for the messes —
that's where the action is.
~Steven Weinberg

In sections 1.3.3 and 4.9, we have already shown that while DL models are usually black boxes, there are methods that can peek inside. In particular, interpretation methods may extract additional information acquired by the model. Additionally, reliability methods can estimate the uncertainty of ML model predictions, which is highly desired when applying ML to novel and unknown physical problems. This chapter aims to show the power of these approaches in phase classification problems. In particular, we use the techniques contained in the Hessian-based toolbox described in detail in section 4.9.

We apply the interpretability and reliability methods to CNNs trained on data coming from two physical models. Firstly, we show a series of results for the numerically simulated data, namely quantum states of the extended one-dimensional half-filled spinless Fermi-Hubbard model. We describe this model and preparation of data in section 5.1. Then, in section 5.2 we present how influence functions ($\mathcal{I}$s) interpret what an NN learns in such a phase classification problem, particularly how $\mathcal{I}$s can detect unknown phases in the data [414]. To complete the picture, in section 5.3, we use $\mathcal{I}$s, relative influence functions (Relat-$\mathcal{I}$s), resampling uncertainty estimations (RUEs), and local ensembles (LEs) to CNNs trained on the same physical model [413]. We show how $\mathcal{I}$s detect anomalies in the data, how RUE sees the width of the transition, and gains from detecting extrapolation by LEs.

The results in sections 5.2 to 5.3 can be thought of as the proofs of

concept. Finally, we present results of the ultimate test for the influence functions ($\mathcal{I}$s): applying them to NNs trained on experimental topological data [278]. We start by discussing the experimental Floquet realization of the two-dimensional Haldane model and features of the time-of-flight images in section 5.4. Then, in section 5.5 we show how $\mathcal{I}$s proved crucial in the fully unsupervised discovery of the topological phase diagram based on the experimental time-of-flight images.

5.1 Numerically simulated data: Fermi-Hubbard model

In sections 5.2 to 5.3, we interpret and study CNNs trained on numerically simulated data that is quantum states coming from various phases of the extended one-dimensional spinless half-filled Fermi-Hubbard model. The Hubbard models are of fundamental importance to condensed-matter physics, with the two-dimensional Fermi-Hubbard model believed to describe the high-temperature superconductivity of cuprates [430]. The selected one-dimensional system has the advantage of being within the power of efficient numerical simulations. As a result, it has a rich and well-studied phase diagram [431, 432] and is a promising candidate to be simulated in a quantum simulator [430]. Therefore, it is suitable to benchmark the $\mathcal{I}$s (or any interpretability method) in phase classification problems. In this model, fermions hop between neighboring sites with amplitudes J and interact with nearest neighbors with strength V_1 and next-nearest neighbors with strength V_2:

$$\hat{H} = -J \sum_{\langle i,j \rangle} c_i^\dagger c_j + V_1 \sum_{\langle i,j \rangle} n_i n_j + V_2 \sum_{\langle\langle i,j \rangle\rangle} n_i n_j \,, \tag{5.1}$$

where $c_i^\dagger$ and c_i are fermionic creation and annihilation operators at site i, respectively, and $n_i = c_i^\dagger c_i$ is the number operator.

Phase diagram. The model exhibits four different phases, two of which co-exist in a limited range of parameters. Without the next-nearest-neighbor interaction, V_2, the system can follow only patterns of the gapless Luttinger liquid (LL) phase or the charge-density wave (CDW) of type I with the degenerated density pattern 101010. The CDW-I order parameter describing this transition reads $O_{\text{CDW-I}} = \frac{1}{L}\sum_{\langle i,j \rangle} |n_i - n_j|$, where $\langle\rangle$ symbolizes nearest neighbors. The next-nearest-neighbor interaction, V_2 competes with V_1, so for non-zero V_2 but still smaller than V_1 the transition between LL and CDW-I shifts towards larger V_1. For sufficiently strong V_2 the bond-order (BO) phase emerges with the order parameter $O_{\text{BO}} = \frac{1}{L}\sum_i (-1)^i B_i$, where

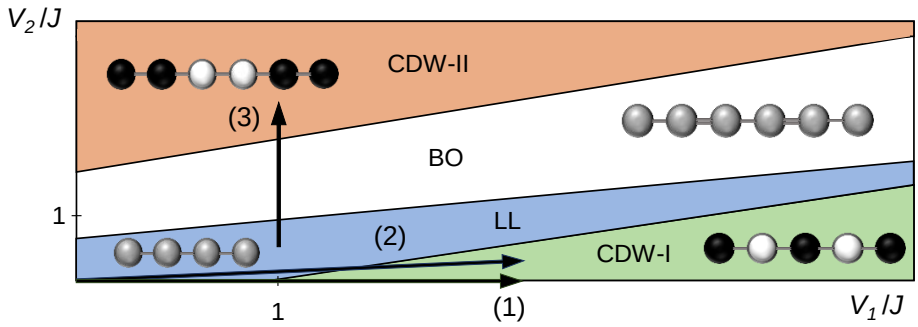

Figure 5.1: **Phase diagram of the extended one-dimensional spinless Fermi-Hubbard model at half-filling.** The phase diagram contains schemes of the corresponding states: Luttinger liquid (LL), bond-order (BO), charge-density wave (CDW) type I and II. The arrows indicate the transitions studied in this thesis. Adapted from Ref. [414].

$B_i = \left\langle c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i \right\rangle$. It turns into the charge-density wave (CDW) of type II with the degenerated density pattern 11001100 for large V_2 values, with $O_{\text{CDW-II}} = \frac{1}{L}\sum_{\langle\langle i,j\rangle\rangle} |n_i - n_j|$, where $\langle\langle\rangle\rangle$ symbolizes next-nearest neighbors. Figure 5.1 presents the scheme of the described phase diagram.

Exact diagonalization. CNNs within this thesis are fed with ground states of the described Fermi-Hubbard model, expressed in the Fock basis, and labeled with their appropriate phases. To calculate the ground states and order parameters of the model, we use the QuSpin package [433] to write the Hamiltonian in the Fock basis. In the majority of cases, we consider a 12-site system with periodic boundary conditions. Such a size results in 924 basis states, and therefore input vectors of 924 elements. We perform the exact diagonalization with the SciPy package [328]. The hopping amplitude, J, is set to 1 throughout the thesis.

Three transition lines. Within this thesis, in various set-ups, instead of working on the whole phase diagram, we train the CNNs on three selected transition lines indicated with arrows (1)-(3) in fig. 5.1. The first transition line leads from the LL to the CDW-I phase. We calculate it for a constant $V_2 = 0$ and $V_1/J = \langle 0, 40\rangle$. It is a source of training data for both figs. 5.3 and 5.4, and test data for fig. 5.3. It is also the main source of data for numerical studies in section 5.3. It is symbolized in fig. 5.1 with the arrow (1), and the values of corresponding order parameter $O_{\text{CDW-I}}$ are plotted in fig. 5.2(a). The transition, defined as above, occurs for $V_1/J = 1$. The second transition line is calculated for $V_2 = 0.25\,V_1$ and $V_1/J = \langle 0, 80\rangle$. Indicated with the arrow (2), it is the source of test data for fig. 5.4. We plot the

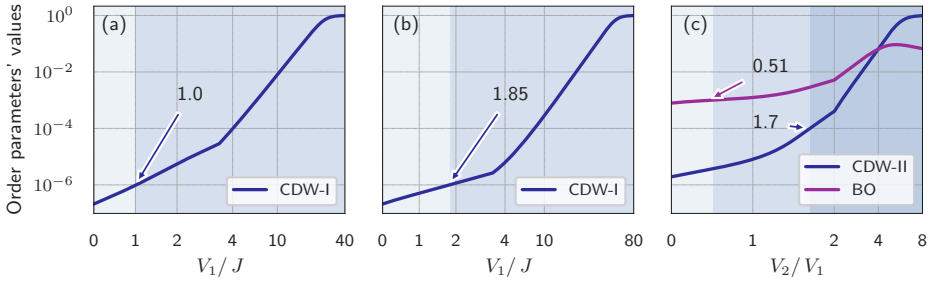

Figure 5.2: **Corresponding order parameters' values for three transition lines studied within this work**, indicated with arrows (1)-(3) in fig. 5.1. (a)-(b) CDW-I order parameter for the transition line between the LL and the CDW-I phase for $V_2 = 0$ and $0.25V_1$, respectively. (c) CDW-II and BO order parameters for the transition line between LL, BO, and CDW-II for $V_1 = 1/J$. Note the logarithmic scale of y-axis, and the symmetric log scale of x-axis with threshold points selected to be 3, 3, and 2, respectively. Cusps in the lines are artificial and result from the symmetric log scale of x-axis. From Ref. [414].

corresponding order parameter CDW-I in fig. 5.2(b), and the transition takes place for $V_1/J = 1.85$. The final transition line cuts three phases: LL, BO, and CDW-II. It is marked with the arrow (3) and provides both training and test data for fig. 5.6. It is calculated for constant $V_1/J = 1$ and $V_2 = \langle 0, 8\rangle\, V_1$. The transition between LL and BO occurs for $V_2 = 0.51\, V_1$, and between BO and CDW-II for $V_2 = 1.7\, V_1$. It is important to notice that for the selected range of parameters $V_2 = \langle 1.7, 8\rangle\, V_1$, two phases co-exist which can be seen in fig. 5.2(c).

Degeneracy of ground states and finite-size effect. Lastly, before diving into results, let us note that the ground states belonging to BO, CDW-I, and II phases are degenerate. To lift the degeneracy of the ground state, we apply symmetry breaking (guiding) fields favoring one of the patterns. This approach results in the order parameters in the LL being not exactly constant and equal to zero. Instead, their values are growing very slowly when approaching the transition points. Therefore, there is no exact transition point, so we define it as such parameters of the system that correspond to the order parameter being ten times larger than the corresponding symmetry breaking fields. Due to the guiding fields of values 10^{-7}, 10^{-5}, and 10^{-4} for 101010 and 11001100 density patterns and 1010 hopping pattern, respectively, the order parameters of values 10^{-6}, 10^{-4}, and 10^{-3} signal the transition to the CDW-I, CDW-II, and BO phase, respectively.

5.2 Phase detection with neural networks and influence functions

This section is based on our work presented in Ref. [414], where the Ph.D. candidate was responsible for: study and implementation of influence functions ($\mathcal{I}$s), testing the method on simpler examples (Iris and MNIST database), interpretation of results, drawing conclusions, preparation of figures, and writing the manuscript. To our knowledge, this work is the first application of the universal and fully agnostic interpretation method to NNs trained to solve a physical problem. The code and data that enable the recovery of results in this section are provided in Ref. [434].

NNs usually hinder any insight into the reasoning behind their predictions. In this section, we demonstrate how $\mathcal{I}$s can unravel the black box of NNs when trained to predict the phases of the one-dimensional extended spinless Fermi-Hubbard model at half-filling. Results provide strong evidence that the NN correctly learns an order parameter describing the quantum transition in this model. We demonstrate that $\mathcal{I}$s allow to check that the network, trained to recognize known quantum phases, can predict new unknown ones within the data set. Moreover, we show they can guide physicists in understanding patterns responsible for the phase transition. We follow a paradigm without relying on the *a priori* knowledge on the order parameter or the system itself, with an approach that is straightforwardly applicable to any physical model or experimental data with no dependence on the architecture of the ML model.

5.2.1 From Luttinger liquid to charge-density wave-I

In this section, we use a CNN with an architecture shown in fig. 4.11(a). We train this CNN to classify ground states of the one-dimensional Fermi-Hubbard model into two phases: LL and CDW-I based on the transition line marked with the arrow (1) in fig. 5.1 for $V_2 = 0$. We plot the $\mathcal{I}$s of all training examples for a selected test point (marked with orange line) in fig. 5.3. The order parameter describing the transition here is the average difference between nearest-site densities, which is zero in the LL phase and non-zero (growing to one) in the CDW-I phase.

Influence of training points on the test point in the Luttinger liquid (LL) phase. The panels (a)-(b) of fig. 5.3 present how influential training points are for test points from the LL phase. The test state (a) is the ground state located deeply in the LL phase, while (b) is closer to the

transition. If the CNN learns an order parameter, all training points, i.e., ground states from the LL phase exhibiting a zero order parameter, should be similarly positively influential, and that is precisely what we observe. They form an almost flat line in panels (a) and (b). For both test points (a)-(b) from the LL phase, the most harmful training points are the ones closest to the transition, but on the CDW-I side. These states are the most similar (with the smallest order parameter value), but already labeled differently.

A slight divergence from the expected behavior. A careful reader can notice that if the CNN learns an order parameter, the training points from the LL phase, all exhibiting a zero order parameter, should be similarly influential and form a flat line in all the panels (a)-(d). However, we see that in reality, their influence changes linearly, which panel (c) shows especially well. This divergence from expected behavior is because, in our exact diagonalization calculations, the order parameter in the LL phase is not exactly constant and equal to zero. Instead, it is growing very slowly, that is why finally the most helpful points are the ones near the transition - they are also the most unique from the training points labeled as LL, and the information they provide is the most valuable. The nonzero order parameter is caused by three phenomena: the finite-size effect, use of the guiding fields, and the numerical arbitrariness of choosing the transition point. In the perfect scenario (observed, for example, for training on states obtained from mean-field calculations), the five most influential points should be randomly distributed over the whole LL phase. It is interesting to note that the results presented in this work stay the same without the symmetry-breaking fields and do not depend on the size of the system.

Influence of training points on the test point in the charge-density wave (CDW)-I phase. On the side of the CDW-I phase, the influence pattern is significantly different. The curvature of influential points corresponds to the growth of the order parameter, and the most influential helpful points are the ones closest to the test point in the order parameter space, slightly shifted towards the transition point, as they provide more information. Panel (c) shows the $\mathcal{I}$s of training points for the test states on the CDW-I side, close to the transition. The most harmful examples are, as in the previous test points, the ones closest to the transition, but on its other side. However, panel (d) presents a distinct behavior of the most harmful examples being in the same phase. All the training points are similarly influential with small values of $\mathcal{I}$s resulting in the almost flat line. It is a signature of the CNN's high certainty regarding the prediction made in panel (d) manifesting with a small test loss function $\mathcal{L}(z_{\text{test}}, \hat{\theta})$. Also, the analyzed test point is deeply in the CDW-I phase, with all neighboring states being almost identical with the order parameter close to 1. The most harmful examples are the ones we label as the CDW-I phase, but very

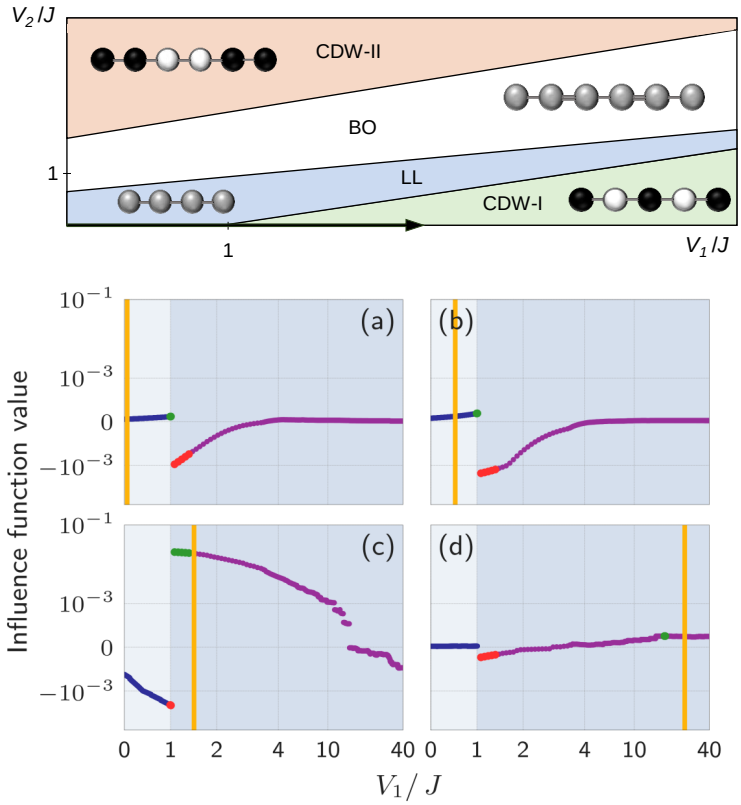

Figure 5.3: **Transition line between LL and CDW-I** presented with an arrow in the phase diagram. (a)-(d) Influence functions ($\mathcal{I}$s) of training examples for selected test points marked with an orange line. Blue (purple) dots are $\mathcal{I}$ for training states from the LL (CDW-I) phase. Larger green (red) dots are five the most influential helpful (harmful) examples. Different background shades indicate two phases. (a)-(b) Blue training points from the LL phase are similarly influential to the classification of the test point from the same phase. They all are characterized by a zero order parameter. (c)-(d) The most helpful training examples for the classification of the test points from the CDW-I phase are the ones with the most similar order parameter. Note the use of symmetric log scale both in x and y axis with 3 and $|10^{-3}|$ selected as threshold points, respectively. From Ref. [414].

different, so the ones closest to the transition.

While analyzing the figures, it is vital to keep in mind that we do not explicitly provide any information on the nearest-neighbor interaction, V_1/J, present on the x-axis (or any physical parameters, in general). We provide the input states in random order. Therefore, the smooth patterns created by the $\mathcal{I}$s and resulting ordering of training points, especially on the CDW-I phase's side, is the sole consequence of the internal analysis of the states by the machine.

5.2.2 Testing the model on another transition line

With a similar approach, we validate the model performance on another transition line. We take the trained CNN from fig. 5.3, and in fig. 5.4 we apply it to test states coming from the transition line for $V_2 = 0.25\,V_1$, where the next-nearest-neighbor interaction shifts phase transition to higher values of V_1/J. Therefore the training and test states come from different transition lines, $V_2 = 0$ and $0.25\,V_1$, marked in fig. 5.1 with the arrows (1) and (2), respectively. Notice the shift of the panels' backgrounds as compared to fig. 5.3. They mark two phases of the test transition line, having a different transition point ($V_1/J = 1.85$) than the training transition line ($V_1/J = 1$).

Panels (a) and (b) of fig. 5.4 show the $\mathcal{I}$ values of training data for test states from the LL phase, while (c) and (d) - from the CDW-I phase. Panel (a) is identical to panel (a) of fig. 5.3, but already panel (b) shows an interesting divergence from fig. 5.3(b), which is a result of a shifted transition point of the test line compared to the training line. No longer the same value of V_1/J, for which test and training states were calculated, yields the same order parameter for both of them. For example, the test state being in the LL phase, close to the transition point for $V_2 = 0.25\,V_1$ should be the most similar to the training points from the LL phase, close to the transition point $V_2 = 0$, and have the most similar order parameter. The ML algorithm follows this similarity with regards to the order parameter, which implies a successful generalization of the model. We see similar behavior in panel (c), where the most helpful points are also shifted as compared to fig. 5.3(c).

5.2.3 Inferring the existence of the third phase

This time we analyze the transition line crossing three phases, LL, BO, and CDW-II, which is indicated by the arrow (3) in fig. 5.1. Two order parameters describe this transition. One is the average difference of the next-nearest neighbor density, which equals zero in the LL and BO phases, and grows to 1 in the CDW-II phase. The other is the staggering of effective

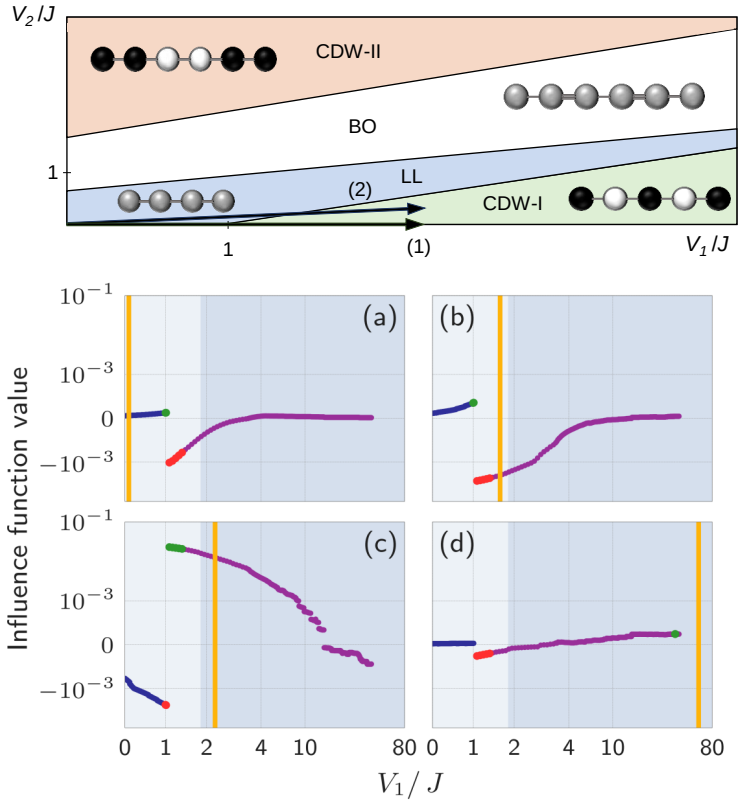

Figure 5.4: **Testing a model on a different transition line.** The model has been trained on training data coming from the arrow (1) at $V_2 = 0$. The test points come from the arrow (2) at $V_2 = 0.25\,V_1$. (a)-(d) Influence functions ($\mathcal{I}$s) of all training examples, marked with dots, for selected test states marked with an orange line. Blue (purple) dots are $\mathcal{I}$ values for training examples from the LL (CDW-I) phase for $V_2 = 0$. Larger green (red) dots are five the most influential helpful (harmful) training examples. Different background shades indicate the phase transition for $V_2 = 0.25\,V_1$ line, from which test states come from. Panels (a)-(d) show very similar patterns to the ones in fig. 5.3, but shifted. It indicates that the similarity of test and training points is connected to their order parameters, as the order parameter of test points is shifted towards larger V_1/J values, compared to training points, due to coming from the $V_2 = 0.25\,V_1$ transition line. Note the use of symmetric log scale both in x and y axis with 3 and $|10^{-3}|$ selected as threshold points, respectively. From Ref. [414].

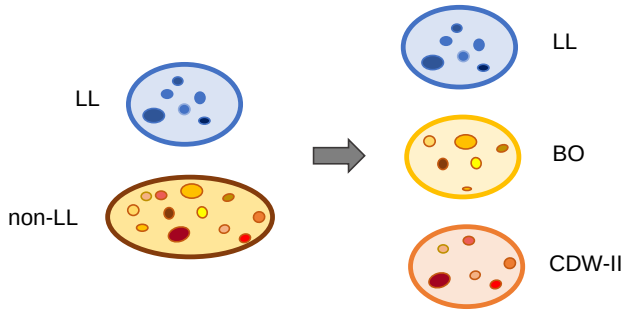

Figure 5.5: **Similarity and detection of additional phases.** We mimic the situation when we have incomplete knowledge of the phase diagram and incorrectly label the data. We label states as either belonging to LL or not when in fact they belong to three distinct phases: LL, BO, and CDW-I. $\mathcal{I}$-based similarity detects the additional phase!

nearest-neighbor hoppings, being 0 in the LL phase, non-zero in the BO phase, and slowly decaying to 0 in the CDW-II phase. In the studied range of parameters, two phases (BO and CDW-II) co-exist (see section 5.1 for the details). It is crucial to note that in this section, we train on the mentioned transition line crossing three phases, but we label ground states only as belonging to one out of two phases.

The first set-up for clarity is presented in fig. 5.5. We show the results for this set-up in the panels (a)-(b) of fig. 5.6. We label ground states as belonging to the LL (blue dots, label 0) or the BO and CDW-II phases (purple dots, label 1). Independently on the test point location, notice two similarity regions within purple training points. Apparently, the NN learns two different patterns (order parameters) to classify the data correctly. Therefore, it notices the existence of the third phase within the incorrectly labeled data. Inferring the third phase would be impossible without interpretability methods, which in this sense pave the way towards unknown phases detection.

The second set-up consists of labeling the same data as belonging to the LL and BO phases (blue dots, label 0) or the CDW-II phase (purple dots, label 1). The $\mathcal{I}$s' values, resulting from this classification, are in panels (c)-(d) of fig. 5.6. The pattern they form is starkly different. First of all, there is no additional similarity region within training points from the LL and BO. The behavior is then more similar to the one seen in fig. 5.3 with the transition between LL and CDW-I. It is not identical, though, as in the phase LL+BO the most helpful training points are always distributed

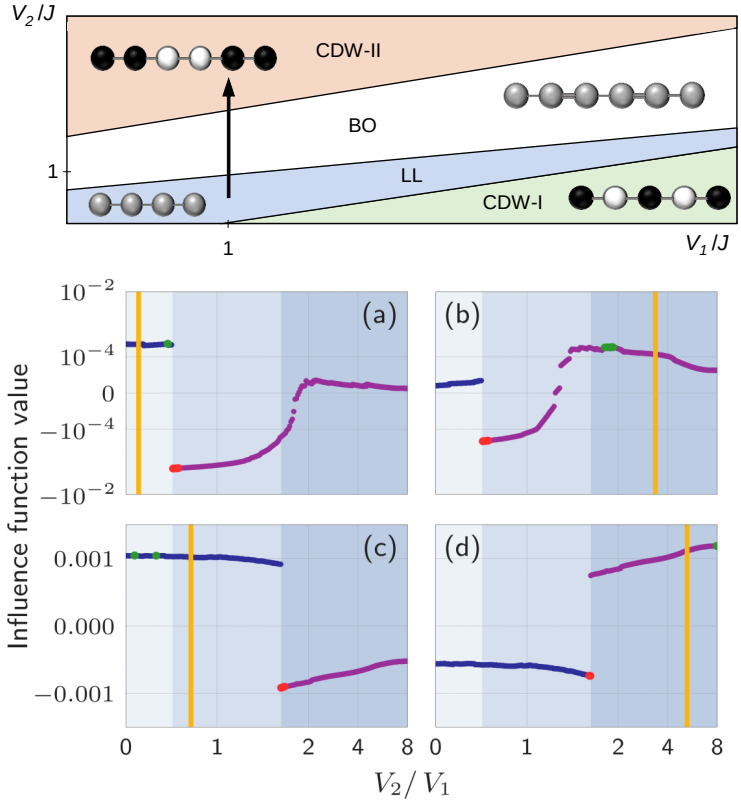

Figure 5.6: **Transition between LL, BO, and CDW-II.** Data comes from the arrow at $V_1/J = 1$. (a)-(d) Influence functions ($\mathcal{I}$s) of all training examples for selected test points marked with an orange line. Training examples are marked with dots and labeled differently within two rows, i.e., as (a)-(b) LL - not LL and (c)-(d) CDWII - not CDWII. Blue dots are $\mathcal{I}$ values for training examples from (a)-(b) the LL phase, (c)-(d) the LL and BO phases, while purple ones (a)-(b) from the BO and CDW-II phases, (c)-(d) from the CDW-II phase. Larger green (red) dots are five the most influential helpful (harmful) training examples. Different background shades indicate phase transitions. (a)-(b) CNN classifies states as belonging to the LL phase or not and detects two similarity regions in the 'not-LL phase'. It effectively indicates the existence of an additional phase. (c)-(d) The model exhibits overfitting. Note the use of a symmetric log scale, except for the linear y axis in panels (c)-(d). In all subplots, a symmetric log scale is used in x axis with 2 selected as a threshold point. A symmetric log scale with $|10^{-4}|$ as a threshold is used in y axis in panels (a)-(b), while in panels (c)-(d) the scale is linear. From Ref. [414].

randomly, but deep in the LL phase, avoiding the BO phase. The most helpful points on the CDW-II side are deep in the CDW-II phase in contrast to fig. 5.3, where they mostly follow the test point. Consider, that the deeper the CDW-II phase, the smaller the BO order parameter, which makes CDW-II predictions easier. The observed pattern is the example of NN not learning correctly the order parameter and potentially overfitting.

Finally, we trained a CNN on the same data, but with three labels correctly corresponding to all three phases. The influence patterns resemble those seen in fig. 5.3 and panels (c)-(d) of fig. 5.6, indicating that CNN correctly learns both appropriate order parameters.

5.2.4 Conclusions

We used the interpretability method called $\mathcal{I}$s on the CNN trained in a supervised way to classify ground states of the extended one-dimensional half-filled spinless Fermi-Hubbard model. We provided strong evidence that the ML algorithm learned a relevant order parameter describing the quantum phase transition. If no knowledge on the actual order parameter were available, $\mathcal{I}$s' values would guide the search for patterns responsible for phase transition and help extract a relevant order parameter, however not providing it explicitly. We have shown that the $\mathcal{I}$s, applied to the trained NN, are able to detect an unknown phase. Two aspects have impacted which training points are the most important for a given test point: how similar they are to the test state and how unique within the training data set. Together they have given a notion of distance or similarity used by the CNN in the phase classification problem and indicated that the patterns relevant for the predictions coincided with the order parameters.

Our approach may be used to address open problems of topological models and many-body localization with NNs, whose logic can be finally discovered by $\mathcal{I}$s. They may be easily applied to any physical model in general. $\mathcal{I}$s should be very successful at distinguishing between types of phase transitions. In particular, the curvature of the line drawn by $\mathcal{I}$s' values should be different for the transitions characterized by continuous and discontinuous change of the order parameter. Moreover, this tool proved to be very sensitive to outliers existing in the data set and may serve for anomaly detection which we present in section 5.3.2. Finally, along with unsupervised learning techniques, it can serve as the first search for unknown phases and order parameters in experimental data which we show in practice in section 5.5.7.

5.3 Hessian-based toolbox for reliable and interpretable machines

This section is based on our work in Ref. [413], where the Ph.D. candidate was responsible for: proposing a problem, finding and developing all necessary numerical methods, implementation of all methods, interpretation of results, drawing conclusions, preparation of charts, writing the manuscript. To our knowledge, this work introduces for the first time universal and fully agnostic ML reliability methods to physics. It also discusses the meaning of feature invariance of an ML model. The code and data that enable the recovery of results in this section are provided in Ref. [435].

ML techniques applied to quantum many-body physics have emerged as a new research field. While the numerical power of this approach is undeniable, the most expressive ML algorithms, such as neural networks, are black boxes: The user does neither know the logic behind the model predictions nor the uncertainty of the model predictions. In this section, we present a toolbox for interpretability and reliability, agnostic of the model architecture. To better understand what an ML model learns, we extract the concept of similarity between input data from a machine. As a result, we can find out what is the relation between data according to the ML model and deduce what features are important for the classification. To this end, we employ and compare $\mathcal{I}$s [411, 418] and relative $\mathcal{I}$s (RelatIFs) [419]. Moreover, we address the need for model-agnostic assessment of the uncertainty of model predictions. We present resampling uncertainty estimation (RUE) [407], which allows for generating analogues of error bars for ML model predictions. Finally, we apply a tool called local ensembles (LEs) [412], which warns a user if an ML model makes predictions with a high level of extrapolation. The four methods require a single calculation of the Hessian of the training loss. Together, they form a Hessian-based toolbox that can be applied to any ML model and any learning scheme that relies on the calculation of the test and training loss functions and therefore finds application outside of physics and the detection of phase transitions. We believe this work opens the road to the systematic use of interpretability and reliability methods in ML applied to physics and, more generally, science.

We present these methods on the CNN trained to detect the quantum phase transition in the one-dimensional spinless Fermi-Hubbard model described in section 5.1. We focus on a single transition line marked with an arrow (1) in fig. 5.1, resulting from a competition between hopping and

nearest-neighbor interaction. It starts in the gapless LL phase, and the increase of V_1 leads to a phase transition to a gapped CDW-I with density patterns 1010 [431, 432]. The order parameter describing the transition, $O_{\mathrm{CDW-I}}$, is the average difference between the nearest-neighbor densities. It is zero in the LL phase and grows to 1 in the CDW-I phase.

Similarity learned by an ML model. We feed the CNN with ground states expressed in the Fock basis, labeled with their appropriate phases, calculated for a 12-site or 14-site system. The architectures of the used CNNs are presented in section 4.7.1. Intuitively, we could expect that the most similar quantum states, according to the ML model, are those generated for the most similar V_1. Instead, as we have shown in the previous section and in Ref. [414], the similarity learned by the ML model is based on the order parameter (or something related to it), as it is a much better discriminator between the phases. Therefore, a well-trained model sees all LL data points as very similar (as they all have a zero order parameter), while the similarity of CDW-I data points depends on V_1 (on this side, the order parameter continuously goes up from 0 to 1).

5.3.1 Influence and relative influence functions

While in the previous section and in Ref. [414] we have studied in detail the potential of $\mathcal{I}$s for interpretability, we here focus on the differences between $\mathcal{I}$s and relative $\mathcal{I}$s (RelatIFs). To this end, we train a CNN on the eigenvectors of the 12-site one-dimensional Fermi-Hubbard model with the labels indicating phases they belong to, i.e., LL or CDW-I. Note that this time we use CNNs with an architecture drawn in fig. 4.11(b), so with a global average pooling (GAP) layer. The reason becomes apparent below.

We then analyze the trained model with $\mathcal{I}$s [eq. (4.5)] and Relat-$\mathcal{I}$s [eq. (4.7)] and present results in fig. 5.7. Each column of fig. 5.7 presents both methods calculated for the same test points, indicated by the orange vertical lines. In the case of $\mathcal{I}$s (first row of fig. 5.7), the most helpful training points are the most similar to the test point and the most unrepresentative in the data set. Relat-$\mathcal{I}$s, presented in the second row of fig. 5.7, are expected instead to ignore how representative data is and to indicate the most similar training points to the test point, paying less attention to outliers.

Behavior in the CDW-I phase. The usefulness of Relat-$\mathcal{I}$ can be seen in the last column of fig. 5.7 when comparing $\mathcal{I}$s and Relat-$\mathcal{I}$s calculated for the whole training set and the test point located deeply in the CDW-I phase [panels (c) and (f), respectively]. According to $\mathcal{I}$s, the location of the most helpful points balances between those being the most unrepresentative (close to the phase transition) and most similar (with the most similar order parameter). As a result, the most helpful points do not follow the test point

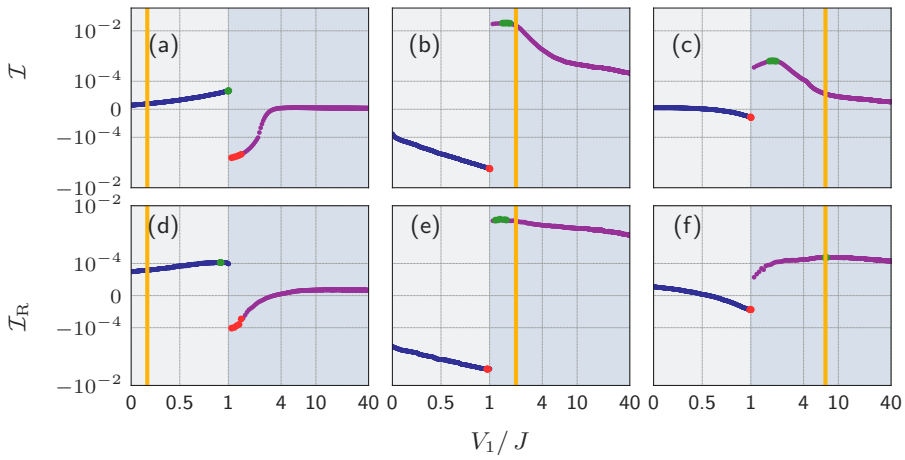

Figure 5.7: **Influence functions ($\mathcal{I}$) vs. relative influence functions ($\mathcal{I}_R$).** (a)-(c) $\mathcal{I}$s and (d)-(f) Relat-$\mathcal{I}$s between the training set and the single test point indicated by the orange vertical line. Every column analyzes the same test point. We mark the five most helpful (harmful) training points with green (red) color. Blue (purple) training points belong to the LL (CDW-I) phase. We mark the transition point with the change of the background color. From Ref. [413].

into the deep region of the CDW-I phase but get stuck instead. Relat-$\mathcal{I}$ ignores how representative data is, and the five most helpful training points follow the test point much deeper to the CDW-I phase because they have the most similar order parameter. Let us now compare the results for the test point in the transition regime in panels (b) and (e). $\mathcal{I}$s and Relat-$\mathcal{I}$s yield here nearly the same results. The test point is in the unrepresentative regime, so the Relat-$\mathcal{I}$'s correction is not needed to observe which data is, in reality, most similar. Still, values of Relat-$\mathcal{I}$s for CDW-I points are much closer to each other than values of $\mathcal{I}$s due to a smaller focus on distinctive points close to the phase transition.

Model-agnostic Relat-$\mathcal{I}$s. A careful reader may notice that in the previous section, in fig. 5.3(d) $\mathcal{I}$s for test points located deeply in the CDW-I phase have a different pattern than in fig. 5.7(c). The reason has been foreshadowed three paragraphs before. This time we use a CNN with a different architecture that underwent also a different training. Therefore, we see that the similarity measure learned by ML models may depend on their architecture and the hyperparameters determining their training. In particular, models may differ in the magnitude to which they regard the transition points as outliers. In our case, both models (correctly) see them as unrep-

resentative in the data set, and the influence of the transition points consistently varies from LL and CDW-I points deep in the phases. However, the CNN with GAP (see fig. 4.11(b)) treats transition points as less representative than the smaller ML model without GAP (see fig. 4.11(a)). This is the interpretation of the mathematical fact that the Hessian of the CNN with GAP has larger eigenvalues, i.e., describes a more curved minimum of a GAP model. As a result, representativeness dominates influence, and the normalization brought by Relat-$\mathcal{I}$ (see eq. (4.7)) is necessary to overcome this effect and focus on the similarity. In this sense, Relat-$\mathcal{I}$s, which apply a correction for the model's focus on unrepresentative data, are more model-agnostic. We have chosen to show this feature on the example of the CNN with GAP simply because when we attempted to do it for the CNN from fig. 4.11(a), values of $\mathcal{I}$s and Relat-$\mathcal{I}$s are very similar. In particular, as seen in fig. 5.3(d), most influential points, according to the $\mathcal{I}$s, follow the test point much further into the CDW-I phase, in the same way as Relat-$\mathcal{I}$s. Only the analysis for the CNN with GAP rendered differences between $\mathcal{I}$s and Relat-$\mathcal{I}$s described in section 5.3.1.

Behavior in the LL phase. Finally, let us compare panels (a) and (d) of fig. 5.7 with the test point located deeply in the LL phase. RelatIFs' values of the LL training points in panel (d) form a more flat line (are less varied) than corresponding influence values in panel (a). In other words, Relat-$\mathcal{I}$s indicate that the LL points are more similar to each other, which agrees with the zero order parameter of the LL. Now let us compare the most influential training points between the methods. The five most helpful training points (marked in green), according to $\mathcal{I}$s in panel (a), are LL data points closest to the transition, while the five most harmful training points (marked in red) are CDW-I data points closest to the transition. They are the most similar to the test point but labeled oppositely, so they confuse the model. Relat-$\mathcal{I}$s indicate the same training points as the most harmful, as the logic behind it is independent of the data representativeness. However, the most helpful points in panel (d) are shifted deeper towards the LL phase than in panel (a). With unrepresentative data being less important, this is the desired direction, i.e., taking the most helpful points further away from the unrepresentative transition regime. The persistent non-zero slope of the LL points may result from the imperfect removal of the impact of unrepresentative data by the Relat-$\mathcal{I}$ normalizer and the finite-size effect present in the 12-site Fermi-Hubbard model.

Therefore, we can study the similarity with $\mathcal{I}$s unless the model focuses predominantly on outliers. However, as we present in the next section, we can use $\mathcal{I}$s' property to focus on outliers to our advantage for anomaly detection. When we need to study similarity, Relat-$\mathcal{I}$s provide a needed correction to ignore how representative data points are.

5.3.2 Influence functions for anomaly detection

We come back to the CNN architecture without GAP as presented in fig. 4.11(a) and used in section 5.2. We continue working with the data set of the eigenvectors of the 12-site one-dimensional Fermi-Hubbard model with labels. The global sign of such an eigenvector is not a physical observable, and therefore a well-generalizing model may ignore this property. To challenge this intuition, we prepare two data sets differing only in the distribution of the global sign. The first data set is composed of a large majority of positive-sign eigenvectors and several negative-sign eigenvectors. In the second, half of the eigenvectors have a positive global sign and the other half - a negative global sign. We call them the 'sign-imbalanced' and 'sign-balanced' data sets, respectively.

Detection of outliers. The CNN trained on the sign-imbalanced data set has high accuracy on positive-sign test points. Regarding negative-sign test points, the CNN correctly classifies them on the CDW-I side but has lower accuracy on the LL side. This suggests that the ML model grasped the global-sign invariance only to a limited level. Figure 5.8(a) presents the $\mathcal{I}$s between the training data and the single positive-sign test point near the transition (marked with the orange vertical line). The pattern is dominated by two smooth lines on both sides of the phase transition, formed by positive-sign training points, as in fig. 5.7(b) in the previous section. However, there are several single training points that are outside the continuous patterns. These 'outsiders' are negative-sign training points that the model regards as distinct from positive training points, in the sense of the similarity as described in section 4.9.1. $\mathcal{I}$s then immediately pinpoint anomalies in the training data and improve the reliability of the model. The non-zero influence of the outliers and the decent test accuracy on negative-sign points indicate that the model gains some information from a minority of negative-sign training points. However, a model can develop different similarity measures depending on the training process and its architecture. For example, all anomalies (or outliers) can have zero influence regardless of the test point. It shows that the model ignores them during the training (which may be encouraged by large regularization), and we could further use such knowledge to improve the model.

Global sign (in)variance of the ML model. We also train a CNN on the sign-balanced data set. The model achieves high accuracy on both positive- and negative-sign test data, suggesting that it learns the global-sign invariance and ignores this property in the decision-making process. This would mean the model is genuinely invariant to the global sign. We check this interpretation with $\mathcal{I}$s plotted in fig. 5.8(b) and calculated for the whole sign-balanced training set and the single negative-sign test point

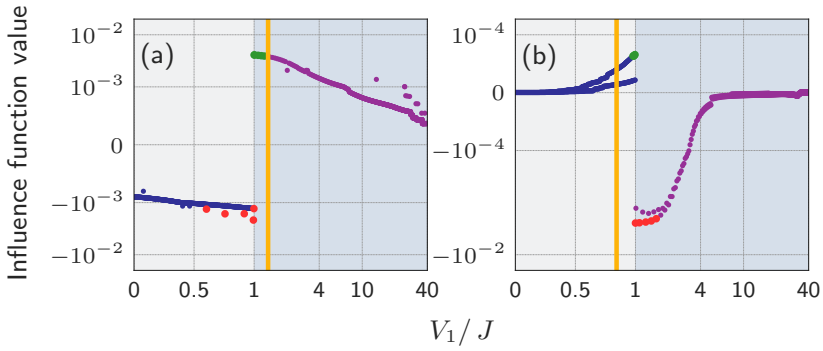

Figure 5.8: **Anomaly detection with influence functions.** Influence functions ($\mathcal{I}$s) between the training set and the single test point indicated by the orange vertical line. We mark the five most helpful (harmful) training points with green (red) color. Blue (purple) training points belong to the LL (CDW-I) phase. We indicate the transition point by the change of the background color. (a) $\mathcal{I}$s show anomalies in the training data. Here, in the sign-imbalanced data set, training points diverging from the smooth lines are negative-sign points. (b) In the sign-balanced data set, $\mathcal{I}$s form two subgroups. The first one is created by the training points with the same global sign as the negative-sign test point, the other - with the opposite sign. The ML model is not truly invariant with respect to this property, even if the classification accuracy is high. Note the use of the symmetric-log scale. From Ref. [413].

marked again with the orange line. If a model is sign-invariant, we should reproduce the smooth patterns of fig. 5.7(a)-(c). Surprisingly, the training points form two subgroups of influence, following their global sign. The subset of training points with the same global sign as the test point reproduces the smooth shape of fig. 5.7(a)-(c), while the subgroup with opposite sign separates. The separation is the strongest near the transition. In the end, the most influential training points are always the ones with the same global sign. Thus, if a feature, to the best of our knowledge, is irrelevant for the classification, the model may still note the property. Again, this behavior and developed concept of similarity depend on the architecture and training process. Still, in our numerical experiments, we always arrived at a model that recognized the training points' global sign.

Strategies to make an ML model invariant. Therefore, we see that having an ML model that is truly invariant to some properties may be a challenging, yet highly rewarding task. A convincing example is provided in section 5.5 and Ref. [278], where we have been studying experimental Floquet data, where a property called micromotion has had no impact on

physical phases in the system but has been consistently recognized by an ML model, rendering unsupervised phase classification ineffective. In the end, we have processed the experimental data with a VAE, effectively removing this property. Another approach would be using domain adaptation neural networks, which could be trained towards ignoring a selected property [436].

The results presented so far concern two methods: $\mathcal{I}$s and Relat-$\mathcal{I}$s. Both analyze the relation between a test point and training point and are especially useful for analyzing the training data or the similarity learned by a model. When assessing the reliability of ML model predictions, more appropriate tools are LEs and RUE.

5.3.3 Local ensembles for out-of-distribution test points

We now challenge the concept of a test loss as a reliability measure. As we will show, the extrapolation score with local ensembles (LEs) highlights better out-of-distribution (OOD) test points. To do so, we introduce in our test set a percentage of test elements whose components are randomly permuted. A well-trained ML model, combined with a reliability method, should be able to inform us of the OOD test points.

Minimal test loss. Let us start by looking at the loss function for every test point, plotted in fig. 5.9 with a purple line. We here use the minimal version of the loss function (i.e., assuming all predicted labels are correct, more details in section 4.6) to mimic a real-life scenario: we ask a trained ML model to make predictions at test points whose labels we do not know. However, here we have access to the ground-truth labels, and we know the model misclassifies the test points generated for $V_1/J \in [1, 2.1]$. The model wrongly predicts that the points in this interval belong to the LL phase. We see that the minimum test loss in fig. 5.9 is primarily smooth and reaches a maximum around $V_1/J = 2.1$. This is the predicted transition point of the model, i.e., for this test point, the model outputs values corresponding to the LL and CDW-I class, which are very close to each other.

LE-based extrapolation score. We then calculate the extrapolation score of the CNN's predictions at the same test set. We plot the LE-based extrapolation score values as blue triangles in fig. 5.9. Let us start with the analysis of the extrapolation score for the regular test points, ignoring the OOD test points. A first observation is that the extrapolation score mostly follows the test loss, which we expect as by its definition the score is proportional to the gradient of the test loss. Secondly, the extrapolation score approaches zero for the predictions at the test points deep in the LL and CDW-I phases. These examples are well-constrained by the typical training examples, and the model needs no extrapolation to make its predictions. Thirdly, we also see that, while the extrapolation score is close to zero for

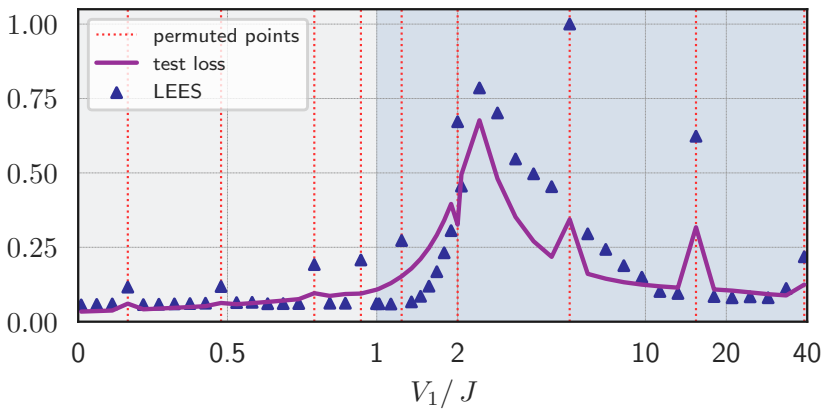

Figure 5.9: **Detection of the CNN extrapolation.** We plot the minimal test loss (i.e., assuming all predicted labels are correct) with a purple line. We mark the LE-based extrapolation score (LEES) with blue triangles. Red vertical lines symbolize OOD test points. All OOD test points have a higher extrapolation score, but not always they are detected by a minimal test loss. We mark the transition point with the change of the background color. Note the symmetric log scale in the x-axis. From Ref. [413].

all the LL test points, it is not for the CDW-I side. This lack of symmetry is contrary to the test loss, which is non-zero close to the transition, regardless of the phase. Strictly speaking, it means that the gradients of all LL test points are parallel to the gradients of the most typical LL training examples, corrected for the local curvature of the loss landscape, i.e., parallel to some of the Hessian eigenvectors with the largest eigenvalues, which we removed to build a flat LE. The ML model sees them as very similar, needs no extrapolation to make predictions at them, and changes of parameters within the LE make no difference. CDW-I test points, on the other hand, exhibit larger diversity, and the model reaches the same low extrapolation level of its predictions as in the LL only deeply in the phase. This makes perfect sense if we assume that two typical training examples representing two classes are ones with $O_{\mathrm{CDW-I}} = 0$ and 1, respectively. Then we can explain the extrapolation score's large values by combining two facts: firstly, it approximately follows the test loss; secondly, it is larger for test points being far from the representative training example from the appropriate class.

The fate of the OOD test points. We now focus on the OOD test points, marked in fig. 5.9 with red vertical lines. On the one hand, if one interprets the test loss as the uncertainty, the test loss should be enough

to highlight the OOD test points as the corresponding predictions should be less confident than those of the neighboring test points. Partially, this intuition holds as we see significant jumps in test loss for some of the OOD test points, especially for those deep in the CDW-I phase. However, the jumps corresponding to the OOD points are much less prominent on the LL side and disappear completely when getting closer to the transition regime. In particular, prediction at the OOD test point at $V_1/J = 2$ is recognized as more confident than on the neighbors, which further challenges the concept of the test loss as the uncertainty measure. On the other hand, the extrapolation score based on LE perfectly highlights each of them, being significantly larger for all OOD test points than its neighbors. The CNN needs to extrapolate on these unseen, atypical test points that the LEs is detecting. With this method, we can then have a neural network built and trained without any constraints, which informs us of predictions made with a high extrapolation level. It is crucial for the predictions at OOD test points, which are not detected at all by the test loss, here, e.g., those between $V_1/J = 0.75$ and 2.

5.3.4 Uncertainty estimation and transition width

We finally use the resampling uncertainty estimation (RUE) method to analyze the uncertainty of predictions of the ML model and therefore identify phase transitions regions. In particular, it is known that finite size effects play a great role in the analysis of phases and should be taken into account when predicting phase transitions. We, therefore, train two copies of the same CNN architecture on two data sets, so on the eigenvectors of the 12-site and 14-site one-dimensional Fermi-Hubbard model with labels corresponding to the LL or CDW-I phase. We choose a CNN architecture that is invariant to the input size (see section 4.7.1).

Visualization of RUE. To detect the quantum phase transition region, we calculate RUE for the whole test set uniformly spread across the transition line. Panels fig. 5.10(a)-(b) show the results for the CNNs trained on the 12- and 14-site systems, respectively. We represent RUE as error bars as they are the variance of the test loss' change under the different sampling of the original training data[1]. It is important to note that we calculate RUE using the minimal version of the test loss (see the discussion in section 4.6), assuming that predicted labels are the correct ones.

[1]We could plot RUE as the error bars of the test loss along with the values of the test loss. However, in our case, various sampling of training data set leads to the test loss' change of the order of $|0.001|$, so around 0.1% of the test loss value, and plotting them together would be infeasible. Secondly, we focus on the width of the regime where the error bars are non-zero, which is better visualized without plotting the test loss values.

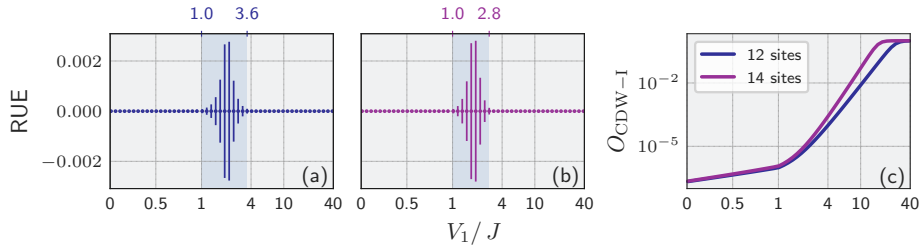

Figure 5.10: **Resampling uncertainty estimation (RUE) vs the quantum phase transition width.** (a)-(b) RUE plotted as error bars of predictions of two CNNs with the same architecture trained on the 12- and 14-site systems. The uncertainty regime is narrower for the 14-site system than for the 12-site one. (c) Order parameters across the same transition for the 12- and 14-site systems. Due to the finite-size effect, the transition is sharper in the 14-site system than the 12-site. From Ref. [413].

Transition regime and error bars. By construction, we know that the RUE indicates uncertainty caused by the limited number of training examples being similar to the test point or due to the ML model making mistakes on training examples that are similar to the test point. As a result, non-zero error bars cover the whole transition regime where the model has troubles with classification. Note we use here a minimal loss function, so RUE has no information on the ground-truth labels. Nevertheless, the error bars are the largest for the incorrect model predictions (i.e., for V_1/J in the interval $[1, 1.6]$ for 12 sites and $[1, 1.45]$ for 14 sites). RUE then manages to detect misclassification of test points. Error bars are non-zero also for correct predictions (i.e., in the interval $[1.6, 3.6]$ for 12 sites and $[1.45, 2.8]$ for 14 sites), but here RUE warns a user that the ML model made these decisions based on the limited number of training data. Moreover, note that our choice of the phase transition point (set to $V_1/J = 1$) is to some level arbitrary for numerical reasons discussed in appendix A of Ref. [414].

Error bars for 12- and 14-site systems. More importantly, we see that these uncertainty regimes have different widths for two different system sizes (see panels (a) and (b) of fig. 5.10). If we set a threshold for RUE's value to 5×10^{-5}, the uncertainty regime spans between 1 and $3.6\,V_1/J$ for 12 sites and 1 and $2.8\,V_1/J$ for 14 sites. It is a direct consequence of the finite-size effect because of which the transition is sharper for the 14-site system than for the 12-site one. Figure 5.10(c) depicts the order parameter $O_{\mathrm{CDW-I}}$ for both system sizes, and one can clearly see a sharper phase transition for a larger system size. Due to the sharper transition and

smaller number of test data with low representation in the training data, the non-zero RUE regime is always narrower in the 14-site case, regardless of the selected threshold for the RUE's value.

Therefore, RUE is a way of providing similarity-based confidence in the ML model predictions.

5.3.5 Conclusions

In this section, we have presented four interpretability and reliability methods that are independent of the architecture and the training procedure of the ML models. They rely on the computation of the Hessian of the training loss describing the curvature around the local minimum. We have shown how these methods could be applied to ML models that classify many-body physics phase diagrams, here the phase classification of the one-dimensional spinless Fermi-Hubbard model. Our findings are summarized in the following:

- We have compared $\mathcal{I}$s and Relat-$\mathcal{I}$s. According to $\mathcal{I}$s, the most influential training points are the most similar to the test point and the most unrepresentative in the data set. Relat-$\mathcal{I}$ has ignored the second aspect and focused on similarity. Here, we mean the similarity as a distance in the model's learned internal representation space. The analysis of this learned similarity, enabled by $\mathcal{I}$s and Relat-$\mathcal{I}$s, increases the interpretability of the ML model.

- Thanks to the focus on unrepresentative data, $\mathcal{I}$s have immediately pinpointed anomalies in the training set and improved the model's reliability. The model can be better understood when the influence of the training outliers is known. For example, an outlier training point can have zero influence on all the test points. This shows that the model ignores such outliers during the training. In phase-detection tasks, in general, the model should be prone against outliers, and therefore, one can further use such knowledge to improve the model and its training.

- With the help of $\mathcal{I}$s, we have also shown that, even if a feature is irrelevant for the phase classification (like a global sign of the wavefunction), the model may still note such features. This finding is consistent with the results found in section 5.5 and Ref. [278]. These findings challenge our intuition about what really is a feature-invariant model.

- The test loss calculated by comparing the output of the ML model and the ground-truth label tends to be interpreted as an uncertainty

of the ML model. This interpretation is tempting due to the simplicity of the test loss calculation but has limited use and can fail miserably on OOD test points. Therefore, we need other tools to increase our trust in the ML model.

- We have shown how LEs are able to identify predictions that are made by an ML model with a high level of extrapolation. In other words, it shows how sensitive the prediction is to the arbitrary choices outside the learning problem, e.g., the random seed. Thanks to this property, the extrapolation score based on LEs has perfectly highlighted all out-of-distribution (OOD) points in our test set.

- We have shown that RUE provided error bars for the ML model's predictions. RUE is large when the training set lacks data similar to the test point or when the ML model makes mistakes on training data similar to the test point. Analysis of RUE values across the test set allowed to assess the phase transition region, which is smaller in the case of the CNN trained on the 14-site system's data than for the 12-site, accordingly with the finite-size effect.

The presented functionalities of the four methods do not exhaust the possible applications. For example, $\mathcal{I}$s and Relat-$\mathcal{I}$s may be used for building more physics-informed ML models. If we know a proper similarity measure, e.g., based on the order parameter in some solvable regime, we can select an ML model which learned the desired similarity and apply the model later to unknown regimes. Moreover, LEs are helpful for active learning in which the model informs the user which additional data points would be the most informative for the training [412]. This approach can prove extremely useful for ML based on expensive experimental measurements. LEs and RUE also can be used to detect additional phases in the data by informing the user about the part of test data on which predictions are highly extrapolated or uncertain. The idea is analogous to the use of $\mathcal{I}$s in Ref. [278] and to the anomaly detection scheme in Ref. [254]. Finally, it would be interesting to apply the same toolbox in the context of quantum ML with variational quantum circuits, where the Hessian of the loss function can also be computed [437, 438].

5.4 Experimental topological data: Haldane model

In sections 5.2 to 5.3, we have interpreted and studied CNNs trained on numerically simulated data that is quantum states coming from various

phases of the extended one-dimensional spinless half-filled Fermi-Hubbard model. In particular, we have shown that interpretation of NNs may lead to the detection of additional phases. We have also shown that we can identify influential properties of the data.

While numerical data has provided a clean environment to present the power of interpretability methods, the final test is their performance when applied to ML models trained on experimental data which may be much noisier and have many natural outliers. We do exactly such a test in section 5.5. Before that, this section aims at describing the studied experimental data. In section 5.4.1 we present theoretical foundations of the two-dimensional Haldane model and we describe its topological invariant - the Chern number. In section 5.4.2, we describe briefly the Floquet realization of the two-dimensional Haldane model that was performed with ultracold atoms by the experimental team at the University of Hamburg and presented in Refs. [276, 439–442]. In section 5.4.3 we discuss a feature of the data, called a micromotion, and its connection to the center-of-mass motion, foreshadowing its importance in section 5.5. Once we discuss the numerical preparation of the experimental images in section 5.4.4, in the next section we present how to extract from such data the full topological phase diagram in a completely unsupervised way.

5.4.1 Two-dimensional Haldane model

The Haldane model has been proposed as a toy model for a quantum Hall effect with no net magnetic flux through the unit cell [443]. In an archetypal scenario, in a two-dimensional electronic system, a strong magnetic field leads to the quantization of the allowed energies of the electrons into discrete Landau levels. As the field strength increases, the spacing between these levels also increases, resulting in a smaller number of filled levels. The number of filled Landau levels determines the Hall resistance which forms plateaus with the value h/le^2, where l is an integer depending on the number of filled levels. Moreover, each filled Landau level gives rise to a chiral edge mode. This can be understood semiclassically in terms of the skipping orbit of the cyclotron motion around the edge of the sample [51] as presented in fig. 5.11(a).

Interestingly, the quantization of Hall conductance can be understood in more general terms and takes place for two-dimensional systems with a broken time-reversal symmetry that host edge states. In such systems, only such conducting edge states contribute to the Hall conductivity, whereas the bulk is insulating. Moreover, these edge states turn out to be connected to a topological invariant called the *Chern number* (eq. (5.6)) and therefore are topologically protected. As a result, the conductance has nearly

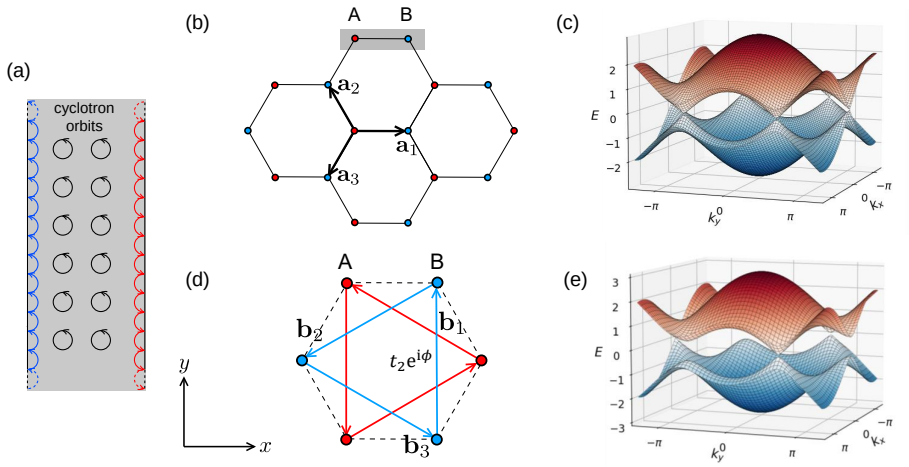

Figure 5.11: **Quantum Hall effect and Haldane model.** (a) Semiclassical picture of edge states arising from cyclotron orbits. (b) Spinless fermions on a hexagonal lattice and (c) its band structure with six Dirac cones touching at Dirac points. (d) Haldane model with chiral next-nearest-neighbor hoppings and staggered potential breaking time-reversal and inversion symmetries resulting in (e) gap opening between Dirac cones. This gives rise to topologically-protected edge states. Adapted from Ref. [444].

no dissipation even in the presence of disorder, which renders these systems highly interesting for various applications, e.g., metrology or quantum computing [51].

The Haldane model describes, therefore, the so-called anomalous quantum Hall effect where the time-reversal symmetry is broken by means other than external magnetic field. It is realizable with ultracold atomic simulators and creates a playground for studying topological phases. The Haldane model describes a tight-binding Hamiltonian of spinless fermions on a two-dimensional hexagonal lattice with a real nearest-neighbor hopping t_1, an on-site staggering potential Δ_{AB}, and a complex next-to-nearest-neighbor hopping $t_2' = t_2 e^{i\phi}$.

Let us construct this Hamiltonian, starting only with nearest-neighbor hoppings across the hexagonal lattice (so with a model of the graphene). Taking a simple tight-binding model where electrons can hop between neighboring sites with hopping strength t_1 and introducing Pauli matrices σ, one obtains the Bloch Hamiltonian:

$$H_0(\mathbf{k}) = \begin{pmatrix} 0 & h(\mathbf{k}) \\ h^\dagger(\mathbf{k}) & 0 \end{pmatrix} = t_1 \sum_i \left[\sigma_x \cos\left(\mathbf{k}\cdot\mathbf{a}_i\right) - \sigma_y \sin\left(\mathbf{k}\cdot\mathbf{a}_i\right)\right] . \quad (5.2)$$

Here $\mathbf{a}_i$ are the three vectors in fig. 5.11(b), connecting nearest neighbors of the lattice (whose spacing we set to one). If we stop at that, the energy spectrum $E(\mathbf{k}) = \pm|h(\mathbf{k})|$ gives rise to the famous band structure of graphene, with the two bands touching at the six corners of the Brillouin zone as presented in fig. 5.11(c). These six cones are called the Dirac cones, and they touch at the so-called Dirac points.

To generate edge states in the system we need to create a gap between these cones. The Dirac points are protected, however, by both inversion and time-reversal symmetry (note that the inversion centers are between sites A and B and in the plaquette centre). We need to break them. To this end, we introduce imaginary next-nearest-neighbor hoppings as well as a staggering potential as shown in fig. 5.11(d). The imaginary hopping is chiral, meaning we broke time-reversal symmetry. The staggering potential on the other hand breaks the inversion symmetry. We arrive to the Haldane Hamiltonian:

$$H(\mathbf{k}) = H_0(\mathbf{k}) + \Delta_{\mathrm{AB}}\sigma_z + 2t_2' \sum_i \sigma_z \sin\left(\mathbf{k}\cdot\mathbf{b}_i\right) . \tag{5.3}$$

Here $\mathbf{b}_i$ are the three vectors in fig. 5.11(d), connecting next-nearest neighbors within the lattice. With appropriate t_2, ϕ, and Δ_{AB}, Dirac cones become gapped. This gives rise to edge states which, as we have already discussed, are connected to the Chern number and therefore topologically protected.

To understand the source of this protection, let us look more closely at the Chern number. The Chern number is intimately connected to the Berry curvature that can be regarded as a magnetic field in the momentum space [444]. See the following argument: when an electron moves on a closed loop enclosing a magnetic field, it acquires a phase shift which is known as the Aharonov-Bohm phase [445]. It is an example of a geometric phase (or Berry phase) that can in general be acquired by a quantum system when transported adiabatically on a closed loop in its parameter space [51]. Imagine now making an adiabatic evolution of the wave function $\psi(\mathbf{k})$ of Hamiltonian in eq. (5.3) on a closed loop L in the momentum space. This wave function also acquires a Berry phase:

$$\gamma(L) = \oint_L \mathbf{A}(\mathbf{k}) \cdot d\mathbf{k} \tag{5.4}$$

Here, $\mathbf{A}(\mathbf{k}) = i\left\langle\psi(\mathbf{k}) \mid \nabla_{\mathbf{k}}\psi(\mathbf{k})\right\rangle$ is a vector called the Berry connection and has two complex entries being the derivatives of $|\psi(\mathbf{k})\rangle$ with respect to k_x and k_y and then taking the inner product with $\langle\psi(\mathbf{k})|$. The Berry connection suffers from the same gauge dependence as the magnetic vector potential.

As the magnetic field is the curl of magnetic vector potential, the *Berry curvature* is the curl of the Berry connection:

$$\mathbf{\Omega}(\mathbf{k}) = \nabla_{\mathbf{k}} \times \mathbf{A}(\mathbf{k}) = i \left[\left\langle \frac{\partial \psi(\mathbf{k})}{\partial k_x} \mid \frac{\partial \psi(\mathbf{k})}{\partial k_y} \right\rangle - \left\langle \frac{\partial \psi(\mathbf{k})}{\partial k_y} \mid \frac{\partial \psi(\mathbf{k})}{\partial k_x} \right\rangle \right] . \quad (5.5)$$

Finally, using the Stokes theorem, we can calculate what is the phase acquired by $\psi(\mathbf{k})$, in particular, after going around the entire Brillouin zone. In the Haldane model, the Brillouin zone has a hexagonal shape, so the curve needs to go along its six borders encompassing its surface. Therefore, the acquired phase is

$$2\pi C = \iint_{\mathrm{BZ}} \mathbf{\Omega}(\mathbf{k}) \cdot d\mathbf{S} \, . \quad (5.6)$$

This way we recover the definition of the Chern number, C. It is the phase acquired when going adiabatically around the entire Brillouin zone. In other words, it is the area integral of the Berry curvature over the whole Brillouin zone. Now the protection of edge states becomes apparent. They are connected to the Chern number and the Chern number is protected by the topology of the Brillouin zone surface that requires significant perturbations to be modified.

5.4.2 Experimental set-up

Challenges of topological models. The Chern number has an interpretation only in the momentum space and is therefore a global quantity. In real space, it becomes apparent, e.g., in the system's linear response, like the quantization of Hall conductance. Because of this global nature, topological invariants are highly non-trivial to detect [51]. Therefore, the classification of topological phases receives particular attention with ML techniques [260–269]. With cold atoms, many detection methods have been demonstrated including the transverse Hall drift [446–449], Berry phase measurements [450], quantized circular dichroism [442, 451] as well as Bloch state tomography [440, 441, 452–454]. The latter is based on momentum-space images after quench dynamics, which also form the basis of the ML analysis in this thesis.

Experimental realization. The data, which is a subject of the unsupervised ML study in section 5.5, has been taken in experiments performed with ultracold atoms in optical lattices [13], which are established as a well-controllable system for studying solid-state physics in general and topological phases in particular [50, 51]. The topological Haldane model [443] is realized by Floquet-driving of a honeycomb lattice [440, 448, 455, 456]. In this specific configuration, the experiments start with a hexagonal lattice

with a large offset $\Delta_{\mathrm{AB}} = 2\pi \cdot 6.1\,\mathrm{kHz}$ between the two sublattices, realized by a suitable polarization of the three interfering laser beams that form the optical lattice [440] as presented in fig. 5.12(a). The lattice is then accelerated on elliptical trajectories by phase modulation of the lattice beams characterized by the shaking phase φ between the modulation along the x and y-direction. The resulting effective Floquet Hamiltonian features non-trivial Chern numbers and gives rise to a topological phase diagram closely related to the original Haldane model and presented in fig. 5.12(d) [276, 441]. The control parameters are the shaking phase φ, which gives rise to the time-reversal-symmetry-breaking, and the shaking frequency f_{sh}, which gives rise to non-trivial Chern numbers $C = \pm 1$ for near-resonant shaking with the sublattice offset $f_{\mathrm{sh}} \approx \Delta_{\mathrm{AB}}/2\pi$.

Theoretically obtained phase boundaries. The numerical prediction for the phase boundary (see fig. 5.12[d]) results from a Floquet calculation for a tight-binding model of the hexagonal lattice based on the shaking parameters and the calibrated parameters of the static lattice. These calculations and predicted theoretical boundaries have been provided by the collaborators in Hamburg. The prediction has been shown to agree well with previous measurements of topological properties in the system [276, 440–442], except for a slight shift of the topological region towards higher frequencies for the experimental data. This shift might have two reasons. The first is the heating introduced in this Floquet realization of the system which leads to contributions of higher bands, which have been neglected in the two-band tight-binding model. The second is the uncertainty in the calibration of the static lattice. Note that the calibration uncertainty of the polarization of the lattice beams of 0.2° leads to an uncertainty of the expected phase transition points of around 200 Hz [276].

Experimental protocol. The experiments have been performed with ultracold spin-polarized fermionic atoms of ^{40}K with mass $m = 40\mathrm{u}$ prepared in the lowest band of the optical lattice formed by laser beams with a wavelength of $\lambda = 1064\,\mathrm{nm}$ as in the earlier work [276, 440]. In the transverse direction, the cloud is weakly harmonically confined. In order to adiabatically prepare the lowest band of the Floquet system, the Floquet drive is gradually ramped up in two steps as presented in fig. 5.12(b):

1. The shaking amplitude is ramped up to 1 kHz within 5 ms at the far off-resonant shaking frequency of $f_{\mathrm{sh}}^{\mathrm{ini}} = 4.5\,\mathrm{kHz}$,

2. The shaking frequency is ramped to the final values $f_{\mathrm{sh}}^{\mathrm{fin}}$ within $t_{\mathrm{ramp}} = 2\,\mathrm{ms}$ at the fixed shaking amplitude.

This ramping protocol strategy aims to keep the band gaps as large as possible and is well-established via earlier experimental and theoretical

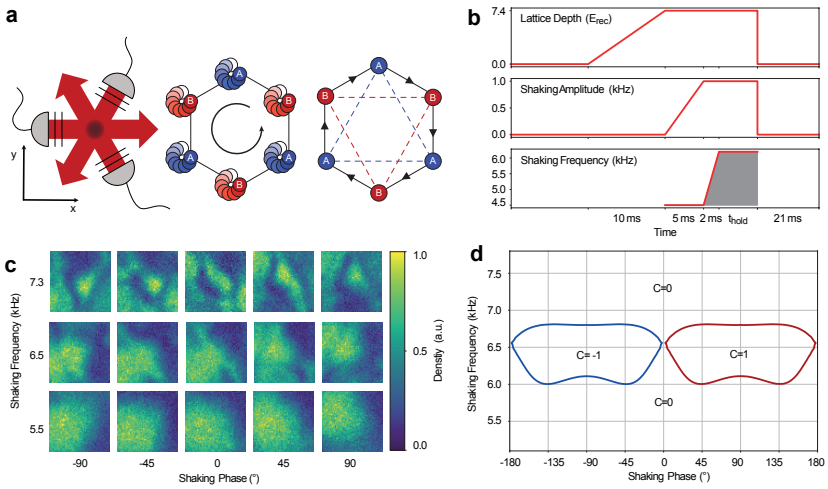

Figure 5.12: **Experimental setup.** (a) Three laser beams form a hexagonal optical lattice by interfering under 120° (left). The polarization and thus the geometry can be tuned via the two waveplates indicated by the black lines. The hexagonal lattice consists of A and B sites with an energy offset, Δ_{AB}. We periodically drive the lattice on an elliptical trajectory (center) in order to obtain an effective Floquet Hamiltonian with Peierls phases on the nearest-neighbor tunneling elements (right) giving rise to topological bands. (b) The atoms are adiabatically prepared in the lowest Floquet band by ramping the lattice depth, the shaking amplitude, and the shaking frequency to different final values. Different hold times allow sampling different micromotion phases given by the grey area under the curve. The atoms are imaged after a time-of-flight expansion of 21 ms. (c) Typical momentum-space images for different shaking frequencies and shaking phases. The images are centered around zero momentum and have a width of one reciprocal lattice vector. (d) The topological phase diagram for the lowest band as predicted by a numerical Floquet calculation featuring the two lobes of non-trivial Chern number, $C = \pm 1$, characteristic of the Haldane model. From Ref. [278].

work [440, 441, 454, 457]. Due to the Floquet heating, this procedure leads to a population of the lowest band of typically 50-75%. Previous work on supervised ML has shown that the Chern number of the lowest band can be faithfully obtained despite the non-zero temperature [276].

Detection and time-of-flight images. For the detection of the state, all potentials are switched off and atoms are allowed to fall due to gravitation for a certain time, t_{tof} (here, $t_{\mathrm{tof}} = 21$ ms). If we neglect interactions between atoms (e.g., because the sample is not dense, and the system is far from a Feshbach resonance), they can be considered to fall freely [13]. Therefore, after t_{tof}, the position of an atom is given by $\boldsymbol{r}(t_{\mathrm{tof}}) = \frac{\hbar t_{\mathrm{tof}} \boldsymbol{q}}{m}$, where $\boldsymbol{q}$ is the

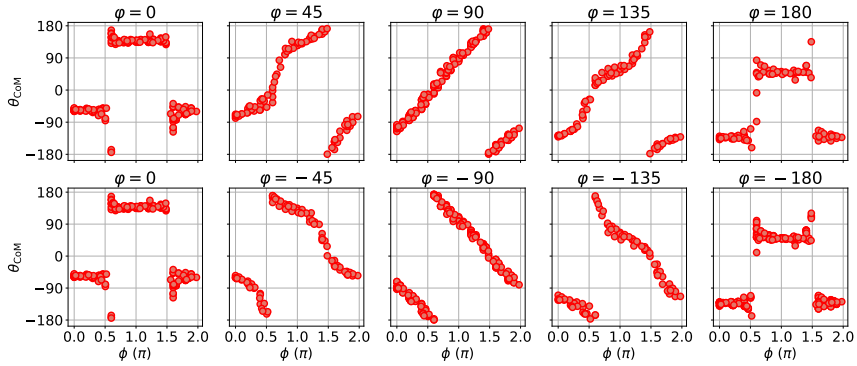

Figure 5.13: **Dependence of the azimuthal coordinate θ_{CoM} of the center of mass and the micromotion phase ϕ.** The dependence can be explained by the elliptical shaking. For the shaking phases $\varphi = 0, \pm 180$ the shaking is linear, thus the cloud can only be displaced in k-space along the direction of the shaking. For a shaking phase of $\varphi = \pm 90$ the shaking is circular and thus the dependence is linear. The sign of φ decides on the direction of shaking which is encoded in the phase jump and the direction of the slope. From Ref. [278].

momentum of an atom in the trap and m is its mass. Thus, such a system expansion maps the original momentum distribution onto the real-space density, which is then imaged by absorption imaging leading to *time-of-flight images*. In section 5.5 they are the data that is analyzed by ML models.

5.4.3 Micromotion

In the experimental protocol, the atoms in the Floquet system are held for different hold times t_{hold} at the final shaking frequency in steps smaller than the Floquet period, in order to sample different instances of the Floquet micromotion ϕ. The micromotion phase is then given by $\phi = f_{\mathrm{sh}}^{\mathrm{fin}}(t_{\mathrm{ramp}}/2 + t_{\mathrm{hold}}) + f_{\mathrm{sh}}^{\mathrm{ini}} t_{\mathrm{ramp}}$. This convention traces the micromotion back to the start of the driving with a kick in a fixed direction and allow relating the micromotion phases of data with different shaking frequencies. The micromotion is an intrinsic property of Floquet systems and while it can give rise to new physics [458, 459], it is often a nuisance when studying the effective Floquet Hamiltonian [457, 460].

For a better intuition behind the micromotion, in fig. 5.13 we show the dependence of the azimuthal coordinate of the center of mass on the micromotion phase. For circular shaking, i.e., a shaking phase of $\pm 90°$, the center of mass moves in a circular fashion yielding a linear dependence between the azimuthal coordinate of the center of mass and the micromotion

phase. For linear shaking, i.e., a shaking phase of 0° and 180°, the center of mass moves along a diagonal line yielding a constant azimuthal coordinate of the center of mass at ±45°, with a phase jump of 180° due to small disturbances in the experiment forcing an exchange of the long and short axis in the fitting scheme of the ellipse. Other shaking phases interpolate between these two behaviors.

5.4.4 Data preparation

For the analysis, we restrict the images to a square region of 56x56 pixels centered around zero momentum, $k = 0$, where 56 pixels correspond to the length of a reciprocal lattice vector - see fig. 5.12(c). The images are furthermore individually normalized to the interval [0,1]. In total, we use 10,436 images with varying shaking phase, shaking frequency, and micromotion phase with just a few images per parameter. While supervised learning often requires an additional large training data set at parameters corresponding to known parts of the phase diagram, which allows labeling, the unsupervised methods discussed below can identify the phase transitions with the data homogeneously sampled across the parameter space alone.

5.5 Unsupervised learning of topological experimental data

This section is based on our work in Ref. [278], where the Ph.D. candidate was responsible for: analysis of the dominant features of the experimental data, the discovery of the final full phase diagram, participation in the analysis and interpretation of results, preparation of figs. 5.15 and 5.22, participation in writing the manuscript. To our knowledge, this work is the first fully unsupervised discovery of a topological phase diagram from experimental data. The code and data that enable the recovery of the results in this section are provided in Ref. [461].

In this section, we apply unsupervised ML techniques to experimental data from topological phases of a Haldane-like [443] model realized in an ultracold-atom quantum simulator. Our aim is to obtain the full topological phase diagram in a fully unsupervised way. It proves extremely challenging with raw data (section 5.5.1), and the problematic feature turns out to be the micromotion. Therefore, we address the problem of dealing with the micromotion inherently arising in any Floquet system using ML for data

post-processing, which allows fixing the micromotion phase of all data to the desired value (section 5.5.2). We show that removal is successful by checking the influence of this feature on ML model predictions (section 5.5.3). Fixing the micromotion phase has proved vital for a successful application of unsupervised ML methods, apparently dominated by a micromotion-induced change of graphic patterns (sections 5.5.4 and 5.5.6). However, it has not been enough to distinguish between topological phases present in the model. The final ingredient has been influence functions ($\mathcal{I}$s) used to gauge similarity within the data preliminary labeled with previous unsupervised approaches (section 5.5.7). By carefully combining the information from these techniques we can uncover the full phase diagram from noisy experimental data in a completely unsupervised way. As expected, unsupervised ML is more challenging than supervised ML, where the identification of the phase transitions was successful even in the presence of micromotion [276]. Our results provide an important benchmark for unsupervised ML of phases of matter and evaluate methods that might be useful for revealing new exotic order in complex systems.

We have implemented all ML techniques using NumPy, PyTorch, and Tensorflow [462–464]. The specifics of the architectures with reproducible code for all performed tasks can be found in our notebooks [461].

5.5.1 Latent-space interpretation of autoencoders

As a first step, we produce and analyze a low-dimensional representation of the time-of-flight images in the latent space of an AE formed by the activations of the bottleneck neurons. We hope to see their clustering corresponding to phases present in the system. AEs are important tools for unsupervised learning [465], and we have described them in section 4.7.2. The AE consists of several convolutional layers and a fully connected bottleneck formed by two neurons visible in fig. 5.14(a). We have chosen a two-dimensional latent space to create an easy-to-understand visual representation of the given samples. The complete implementation details can be found in our notebooks [461]. We have checked that choosing more dimensions in latent space does not lead to an improvement in separation between phases. The AE here is trained on the complete data set.

Limited clustering. The two-dimensional latent-space representation of all images yields a dense cloud of data points without an apparent clustering in fig. 5.14(b). The picture becomes clearer, when we restrict the data to fixed shaking phases, i.e., vertical cuts through the phase diagram (see fig. 5.14[c-d]). The data then lies on elliptical structures with the radius related to the shaking frequency. For further analysis, we fit an ellipse using direct least square fitting [466] and perform a coordinate transformation to

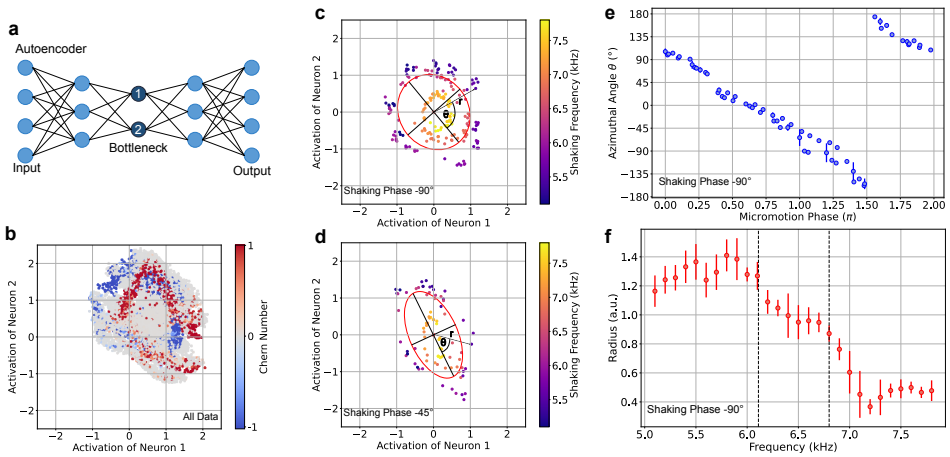

Figure 5.14: **Bottleneck analysis.** (a) Sketch of the AE with two bottleneck neurons, which is trained to reproduce the entire data set. (b) Activations of the two bottleneck neurons form the latent space, in which each point corresponds to one image. In the scatter plot of all data, the Chern number is color coded according to the numerics, illustrating that the data does not cluster according to the Chern number in latent space. (c-d) The analysis of single cuts through the phase diagram, i.e., of data with a fixed shaking phase of (c) $\varphi = -90°$ and (d) $\varphi = -45°$. The data forms rings in latent space and is fitted by an ellipse (red line), which forms a new coordinate system with the azimuthal angle θ measured relative to the longer half axis and the radius r as sketched with the arrow. (e-f) Analysis of latent space in elliptical coordinates. (e) The azimuthal angle θ is linearly related to the micromotion phase for $\varphi = 90°$ of the individual images independent of their shaking frequency. (f) The mean radial coordinate for a given shaking frequency traces out a monotonously decreasing curve with no clear signature of the phase transition and indicates three plateaus in accordance with the phase boundaries. The error bar is the standard deviation from averaging over the images with a given shaking frequency. The plots for other shaking frequencies look similar. This association with the micromotion means that latent space can be interpreted, but also that micromotion is the dominant feature hiding possible signatures of the topological phase transitions. From Ref. [278].

extract the elliptical coordinates radius r and azimuthal angle θ measured from the major axis of the fitted ellipse.

The model learns micromotion. The azimuthal angle can be clearly connected to the micromotion phase showing a linear dependence as seen in fig. 5.14(e) for a shaking phase of $\varphi = 90°$. The same dependence can also be seen with the azimuthal coordinate of the center of mass of the raw images, which provides a direct connection between the time-of-flight images and latent space as we have discussed in section 5.4.3. We furthermore explore possible information hidden in the radial coordinate, presented in fig. 5.14(f). The mean radius decreases with shaking frequency with some signs of plateaus, but without a sufficiently clear separation with shaking frequency for making a prediction of phase transitions. The latent space representation can thus be physically interpreted via the micromotion, but it cannot provide an identification of the topological phases. We attribute this to the dominance of the micromotion, which we aim to eliminate in the following section 5.5.2.

5.5.2 Data post-processing to the desired micromotion phase

In Floquet systems, micromotion poses an additional challenge for identifying phase transitions. We have found that all attempts to apply unsupervised ML methods failed for data with varying micromotion phases. In contrast, as we show in this thesis, unsupervised ML is successful if all images have the same micromotion phase, i.e., the center-of-mass displacement is in the same direction. Changes in the micromotion phase induce different graphic patterns, including the moving center of mass, which seems to dominate the models' predictions. The studied experimental data has been taken by sampling various micromotion phases. Therefore, we need to post-process it to the desired single value of the micromotion phase.

A VAE with a question neuron. We show that this can be accomplished by ML techniques based on VAEs, which are powerful tools for data transformation and generation [395, 396] and have been described in section 4.7.4. Our architecture here uses an additional question neuron in the bottleneck, which has also been introduced in section 4.7.4 and which previously has been proven successful in identifying relevant physical properties [298]. Here we use the additional question neuron for the supervised training of the VAE. The given challenge is related to tasks such as fringe removal in absorption imaging [467] or removal of timing jitter in pump-probe experiments [468], but we believe that our method based on VAE is very broadly applicable to post-processing to the desired sampling in different experimental scenarios.

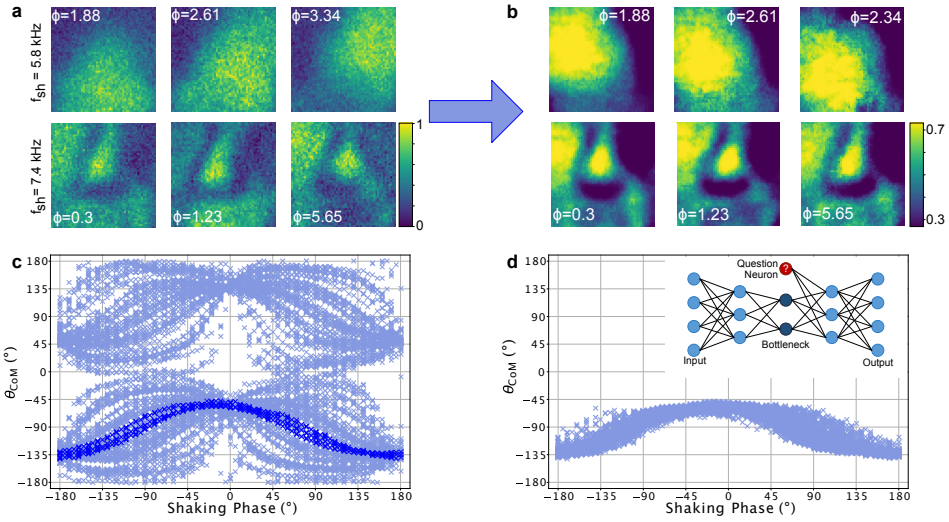

Figure 5.15: **Post-processing data to the desired micromotion phase using a VAE with an additional question neuron.** (a) Example images at shaking frequencies of 5.8 kHz and 7.4 kHz and shaking phase $\varphi = 90°$ for different micromotion phases ϕ illustrating the motion of the center of mass with the micromotion phase. (b) The same example images as in (a) rephased by the VAE to a micromotion phase of $\phi_0 = 0.0$ show a center of mass position only weakly dependent on the original micromotion phase ϕ. (c) The azimuthal coordinate of the center of mass of the original images as a function of the shaking phase shows a complicated dependence. (d) The azimuthal coordinate of the center of mass (θ_{CoM}) of all images forms a narrow band and is only dependent on the shaking phase, but not the original micromotion phase. This dependence on the shaking phase is similar to the dependence of the data for actual micromotion phase $\phi = 0.0$ before rephasing presented with dark blue points in panel (c). The comparison of the distribution of azimuthal coordinates of the center of mass in (c) and (d) illustrates that the rephasing of the micromotion has been successful. The micromotion phase of the processed data $\phi_0 = 0.0$ is controlled by an additional question neuron in the bottleneck of the VAE presented as inset in (d). From Ref. [278].

Used architecture and training details. The encoder of the VAE consists of several convolutional stages and several layers of fully connected neurons. The last layer of the encoder has 26 fully connected neurons thus the latent space covers 13 uncorrelated Gaussians distributions, each defined by its mean and log variance. The decoder has again several fully connected layers followed by a few transposed convolutional stages. The first fully connected layer is also attached to the input of the question neuron. In total, the AE has over 3 million trainable parameters. The complete implementation details can be found in our notebooks [461]. To optimize the hyperparameters of our AE, we use the hyperparameter optimization library optuna [374] and test over 60 000 different network architectures. To identify the best working network we use the structural similarity index [469] as a measurement for performance. For each point in the phase diagram with a fixed shaking frequency and shaking phase, several images with different micromotion phases were taken by varying the hold time. As a new data set, we select all combinations and permutations of images for a fixed shaking frequency and shaking phase and calculated their micromotion phase difference $\Delta\phi = \phi_{\text{output}} - \phi_{\text{input}}$. Thus the data set includes 63,050 image pairs with a given $\Delta\phi$. Thus in contrast to the other AEs we employed in the context of this work, the input and output are different for the VAE. We randomly choose 10% of the data set for validation purposes and hide them from the network during training. We train with one image as input, which we refer to as input image, and one image with the same shaking frequency and shaking phase but a different micromotion phase as output and the micromotion phase difference as input for the question neuron. Samples of original images are given in fig. 5.15(a).

Postprocessing of time-of-flight images. After training, we use the VAE to transform all original images in the data set to a micromotion phase of $\phi_0 = 0.0$ by choosing their micromotion phase with an opposite sign as input for the question neuron. The post-processed images with a single micromotion phase are similar to the original data except for some noise removal and a squeezing of the distribution of pixel values to a range of 0.3 to 0.7, which we attribute to the non-linear activations in the network - see fig. 5.15(b).

Confirming successful fixing of the micromotion phase. Because the micromotion is directly related to the center of mass of the images, one can see the success of the post-processing in the example images in fig. 5.15(b): after micromotion is set to a fixed value, the images with different micromotion phases look very similar and have the same direction of the displacement of the center of mass. This is further confirmed by a comparison of the distributions of the azimuthal center of mass coordinates for all images before and after rephasing in fig. 5.15(c)-(d). The highlighted blue

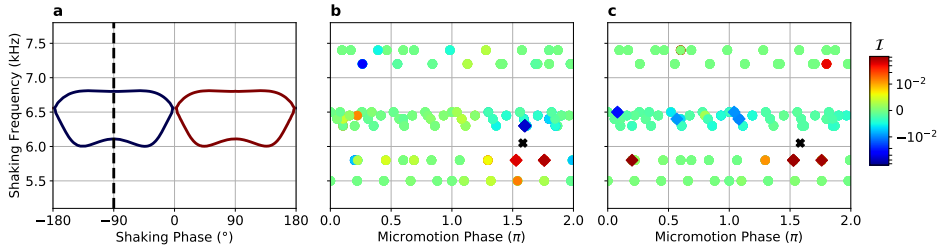

Figure 5.16: **Confirming the post-processing to the desired micromotion phase with influence functions.** (a) Phase diagram with theoretical boundaries between phases and with the cut at a shaking phase of $-90°$ marked with a black line, which is analyzed on the right. Panels (b)-(c) show the color-coded influence function ($\mathcal{I}$) values for all training points marked with dots and for the same test point (marked with a black cross) for a single cut at phase $-90°$ across the shaking frequency before and after the post-processing. The five most helpful (harmful) training points are marked with red (blue) diamonds and they can overlap. (b) The most influential (both helpful and harmful) training points are the points with the shaking frequency and the micromotion phase closest to the test point. (c) After fixing the micromotion phase, the most influential points are distributed across various original micromotion phase values indicating the successful removal of this property which now is ignored by the CNN. From Ref. [278].

data in in fig. 5.15(c) are the center of masses for the micromotion phase $\phi = 0.0$. The narrow distribution in fig. 5.15(d) shows that the rephasing has been successful. In the next section, we further confirm the success of rephasing with $\mathcal{I}$s.

5.5.3 Confirming rephasing with influence functions

To get further evidence that the post-processing described in section 5.5.2 has been successful, we confirm it with $\mathcal{I}$s introduced in section 4.9.1. $\mathcal{I}$s provide an interpretation of the ML model by indicating which training points are influential for a selected prediction. Analysis of the most influential examples can reveal the characteristics which impact ML predictions.

Supervised training of a CNN. Firstly, we train a CNN in a supervised way to classify original images with the varied micromotion phase. Instead of analyzing the whole two-dimensional diagram, we consider only a single cut at the fixed shaking phase $-90°$ which simplifies the visualization of the results without changing them. Within this single cut, the Chern number of the system changes from 0 to -1 and back to 0 with increasing shaking frequency. The labeled training data contains then only two phases

($C = 0$ and $C = -1$). To avoid influence from experimental imperfections, we exclude data close to the theoretically-predicted phase transitions. With the trained CNN, we calculate the $\mathcal{I}$s determining how influential the whole training data set is for the prediction selected to be in the transition region. The results are presented in fig. 5.16(b-c) with the black cross indicating the test point and dots representing the training data and their color-coded $\mathcal{I}$ values. Colors vary from red for helpful training points, through green for least influential (ignored), to blue for harmful.

Micromotion is influential in raw images. We see in panel (b) that the most influential (both helpful and harmful) data for the selected prediction are those with both the most similar shaking frequency and micromotion phase. Learning the shaking frequency is expected as it is the parameter governing the phase transition. However, the CNN also regards the micromotion phase as influential when making a prediction, while we know that this property is physically irrelevant for the transition. The micromotion phase is an intrinsic property of the Floquet realization of the topological Hamiltonian but does not change the topology of the effective Floquet Hamiltonian.

The influence of micromotion is removed. We do the analogous analysis for post-processed training data, i.e., with the removed micromotion phase. Panel (c) in fig. 5.16 shows that the most influential points are now randomly distributed along the original micromotion phase axis. It tells us that the CNN no longer sees this parameter as influential and confirms that the data has been successfully post-processed to a constant micromotion phase.

Let us note that when training on data with or without the micromotion, in both cases the validation and test accuracy of the trained CNN are similar. It means, interestingly, that in this set-up the predictive power of the network is not impacted by learning the quantity which is physically irrelevant.

5.5.4 Clustering in latent space

In the following sections, we apply different unsupervised ML methods to the post-processed data with a constant micromotion phase and compare their performance. We start with a method of the clustering category, which identifies clusters in some low-dimensional representation of the data as different phases. Specifically, we employ the same convolutional AE as in section 5.5.1 but now applied to the post-processed data.

Clustering in the AE latent space. The latent space representation of the complete data set is shown in fig. 5.17(b). The data is color encoded by the theoretical predictions for the Chern number. It appears that the dif-

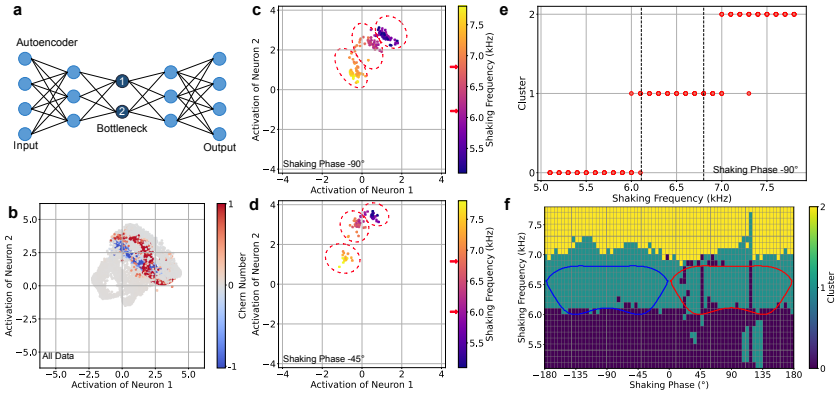

Figure 5.17: **Bottleneck analysis of rephased data.** The bottleneck analysis with a simple AE, sketched in (a), for the data after rephasing to a fixed micromotion phase. (c)-(d) The analysis of single cuts through the phase diagram, i.e., of data with a fixed shaking phase of $\varphi = -90°$ in (b) and $\varphi = -45°$ in (d), shows a significant clustering according to the shaking frequency (color coded). The dotted lines guide the eye to separate the different clusters found by the k-means algorithm. The arrows next to the color bar mark the theoretically expected phase transitions. (e) The cluster analysis of latent space. k-means clustering analysis for $k = 3$ clusters shows that the data is indeed clustered according to shaking frequency. There are several values for each shaking frequency for the different original micromotion phases, which mostly lie on top of each other in the plot. (f) Naming the three clusters increasing with frequency and combining the data from all cuts yields a topological phase diagram, which is reasonable agreement with the numerical prediction (blue and red lines), but does not distinguish between $C = 1$ and $C = -1$. From Ref. [278].

ferent topological classes tend to form ring structures in the two-dimensional latent space. As in section 5.5.1, we now restrict ourselves to single shaking phase cuts. Figure 5.17(c) and (d) show the latent space for two such cuts and we observe three main clusters related to three frequency regimes.

***k*-means clustering.** Choosing k-means clustering, a standard method to solve cluster problems described in section 4.3, it is possible to automate the clustering process of the different latent space representations of the observed image data. We use the k-means functionality of Scikit-learn [470] and set the number of clusters to 3 and the maximal number of iterations to 500. All other parameters are set to standard according to the documentation. We tried different random seeds to prove stability. The results of the cluster analysis are shown in fig. 5.17(e) and (f).

Indistinguishable topological phases. This allows one to predict phase boundaries in good agreement with the theoretical predictions given

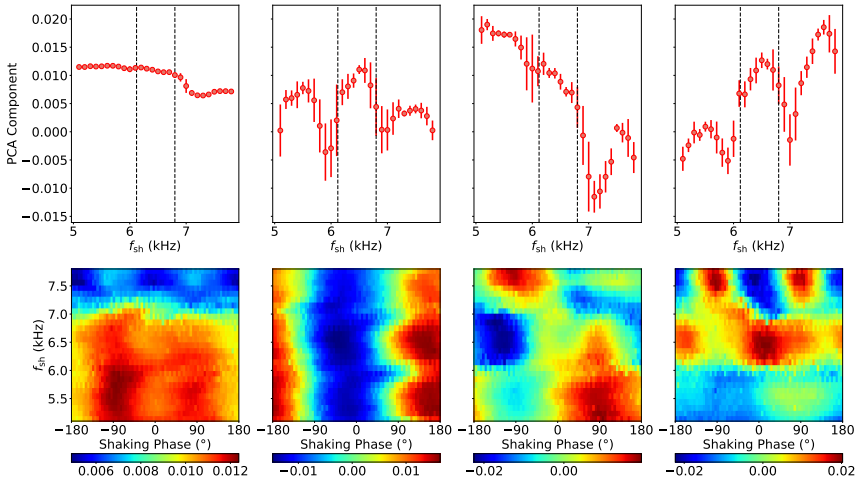

Figure 5.18: Principal component analysis (PCA) of the time-of-flight images rephased to a fixed micromotion phase. From left to right: principal components 1, 2, 6, and 8 that have been selected arbitrarily. For all other components, see the notebooks in [461]. In the upper row, a cut along a shaking phase of 90° is plotted. The error is the standard deviation of the different values for the component for the given shaking frequency. The theoretical predictions for the phase transitions are given by the dashed lines. In the lower row, the averaged components are plotted in the Haldane phase diagram fashion. From Ref. [278].

by the dashed lines. The slight shift to higher frequencies is in accordance with all other methods and has been explained in section 5.4.2. Evaluating the data for all shaking phases allows reconstructing the complete two-dimensional Haldane phase diagram shown in fig. 5.17(f). The procedure of separately analyzing vertical cuts through the phase diagram can fundamentally not distinguish between the $C = 1$ and $C = -1$ phases at positive and negative shaking phases. Furthermore, a similar analysis of horizontal cuts along constant shaking frequencies through the phase diagram does not produce clustering. Therefore further methods are required to fully identify the topological phases.

5.5.5 Principal component analysis of rephased data

In addition to k-means applied to the AE latent representation of the data, we also use principal component analysis (PCA) on the processed data. We have described PCA in section 4.3.2 as an alternative way of dimensionality reduction. In fig. 5.18, we select four components of this analysis. The different principle components clearly show features that cor-

respond well with the theoretical predictions, such as sharp local minima at the expected phase transitions. However, the data also shows features that are not related to phase transitions, such as strong dependence on the shaking phase in the trivial regions. Therefore the data does not provide a clear recipe for identifying the topological phase transitions in a completely unsupervised way. This is true in particular for the choice of the components to be analyzed.

5.5.6 Anomaly detection scheme

So far, we have analyzed the lower-dimensional representation of the rephased data using an AE and PCA. Finally, we follow the approach in [254] and perform an unsupervised learning scheme called anomaly detection (described in section 4.7.3) to map out the phase diagram in a few training iterations. We use a convolutional AE, similar to the network described in section 5.5.2. The network consists of an encoder and decoder made of two convolutional layers each, with a fully-connected bottleneck of 50 units. For full details about the model and process, we refer the interested reader to [461], where all steps described in this section can be exactly reproduced.

Anomaly detection on rephased time-of-flight images. The idea is the following: We started by defining a region of the phase diagram in which we trained the AE to encode and decode the images with low mean-squared error $\mathcal{L}_{\mathrm{MSE}}(\boldsymbol{x}^{\mathrm{in}}, \boldsymbol{x}^{\mathrm{out}})$ between input and output images $\boldsymbol{x}^{\mathrm{in}}, \boldsymbol{x}^{\mathrm{out}} \in \mathbb{R}^{56\times 56}$. The network learns the characteristic features of the phase that the images were taken from and fails to reproduce images from the other phases. By looking at the loss for all images after training in only part of the phase diagram, we distinguish between the phase it has been trained on and the remaining phases via different plateaus of the loss function. Furthermore, by fitting a sum of tanh functions to the loss curve as a function of shaking frequency, we obtain a phase boundary. We then repeat this process by training in the region of high loss from the previous training round until we find all boundaries.

Training in the low-frequency regime. We show this process in fig. 5.19 where we have started by training in the low-frequency regime (Training 1). We use all values of the shaking phase from $[-180°, 180°]$ and small shaking frequencies up to 5.5 kHz as indicated by the cyan rectangle. We identify three different plateaus in the loss, from which we obtain two boundaries. As seen in panel 1(b), where we take a single cut of the phase diagram at $-90°$, the different phases make up plateaus. Therefore, we fit a tanh function in a relevant parameter region to estimate the boundaries as indicated by the grey lines.

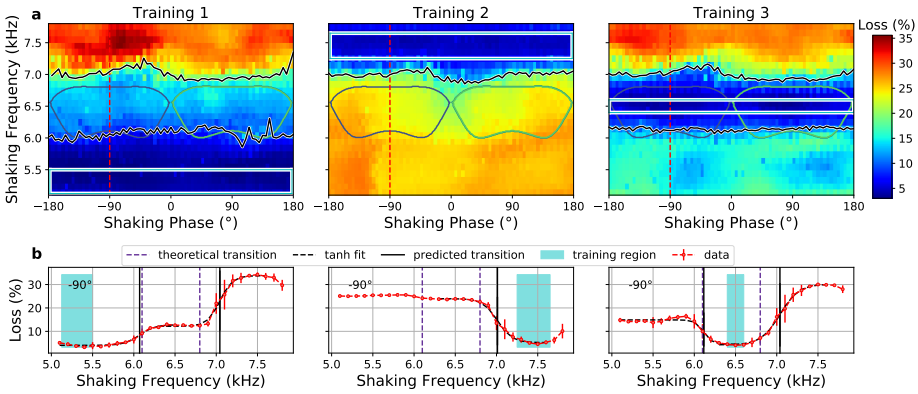

Figure 5.19: **Anomaly detection scheme.** We start by training in the topologically trivial phase (Training 1). Due to the symmetry in the shaking phase we use all angles from $[-180°, 180°]$ and small frequencies up to 5.5 kHz as indicated by the light blue box in **a**. Panel **b** shows a single cut at shaking phase $-90°$ where we can see two plateaus which we identify with the topological non-trivial and trivial phases, respectively. Training 2: We continue by training in the region of highest loss in the first iteration for high frequencies. From these two training iterations, we can already narrow down the two boundaries in the phase diagram. Training 3 completes the overall picture and confirms the phases mapped in Training 1 and 2. From Ref. [278].

Training in two other regimes. We continue the process and train in the high-frequency regime, where the first iterations have yielded the highest loss. As seen in Training 2 (see fig. 5.19), we find the reverse picture with a clear boundary slightly above the theoretically predicted transition. This boundary from Training 2 nicely coincides with the second boundary from Training 1. To complete the picture, we also train in the intermediate-frequency regime that has yielded higher loss in the previous training iterations. Here the training region is significantly smaller, yet the results still match nicely with the previous training iterations. We stress that the three training procedures are independent. Therefore, the good agreement of the phase boundaries in fig. 5.19 is a strong indication of the method's validity.

Topological phases remain indistinguishable. The images provide sufficient information to separate the different phases and map out the phase diagram with this method. We present the predicted diagram in fig. 5.20. We notice that with this method, it is not possible to differentiate the non-trivial topological phases with Chern numbers 1 and -1 because the trained compression does not generalize well in the shaking phase parameter. For further details, see appendix C in [278].

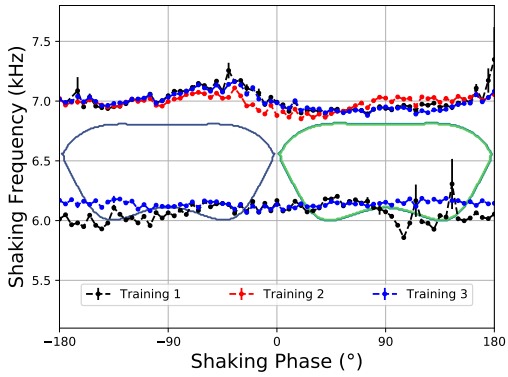

Figure 5.20: **Phase boundaries, concluded from four training iterations of anomaly detection** in fig. 5.19, separate three different phases, which we identify with the topologically trivial and non-trivial phases. The obtained boundaries from the different training iterations are consistent with each other. From Ref. [278].

In the following section, we overcome this shortcoming and complete the phase diagram, i.e., separating the two topological regions, employing the $\mathcal{I}$s. We see that the transition between intermediate and high frequencies for all three training rounds is slightly above the theoretically predicted transition, which is due to a mismatch between theory and experiment as discussed in section 5.4.2.

5.5.7 Analysis of similarity with influence functions

After obtaining the phase boundaries from the anomaly detection scheme, as described in the previous section, we analyze how similar data are within the three distinguished phases. Such an analysis not only can confirm the predictions of unsupervised ML schemes but also reveal the existence of additional phase transitions. Following the proof of concept presented in fig. 5.6 on the example of the Fermi-Hubbard model, we train a CNN on the post-processed experimental data, i.e., with the single micromotion phase, with labels assigned by the anomaly detection scheme. Therefore, we have three labels corresponding to the low-, intermediate-, and high-frequency phases as presented schematically in fig. 5.21. We employ $\mathcal{I}$s, described in section 4.9.1, to analyze which training data are influential for a given prediction. Similarly influential training data, $\{\boldsymbol{x}_{\text{train}}\}$, with a similar $\mathcal{I}$s' values, $\mathcal{I}(\boldsymbol{x}_{\text{train}}, \boldsymbol{x}_{\text{test}})$, for a particular test point, $\boldsymbol{x}_{\text{test}}$, can be then interpreted as similar from the ML model's point of view in a given problem.

$\mathcal{I}$s for three test points from different phases. To analyze similar-

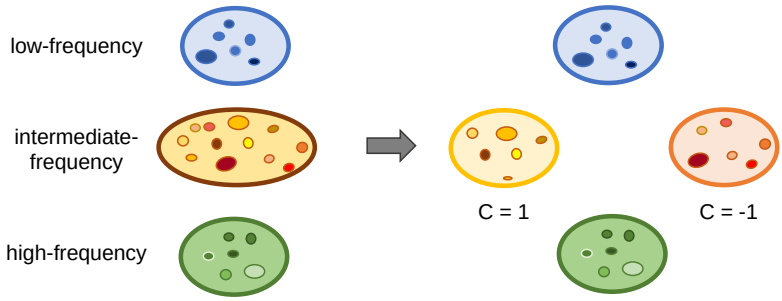

Figure 5.21: **Similarity and detection of topological phases.** We label time-of-flight images as belonging to low-, intermediate-, and high-frequency phases. $\mathcal{I}$-based similarity detects that there are two phases within the intermediate-frequency phase!

ity of training data, we need to compare $\mathcal{I}(\boldsymbol{x}_{\text{train}}, \boldsymbol{x}_{\text{test}})$ and therefore fix the test point for which $\mathcal{I}$ is calculated. In fig. 5.22, we plot three sets of $\mathcal{I}$s calculated for all training data and three different test points: one is located in the low-frequency regime, the second in the intermediate-frequency regime, and final in the high-frequency regime that is in panels (a), (b), and (c), respectively. Each element of the phase diagrams in the upper row indicates color-coded $\mathcal{I}$ value for a corresponding test point marked with a black cross. If an element corresponds to more than one training point (if more measurements were performed for a given frequency and shaking phase), then we plot the mean of $\mathcal{I}$s. Yellow (dark blue)[2] color indicates the most helpful (harmful) training points for a given prediction. White dots correspond to the lack of available training data. The lower row of fig. 5.22 contains the mean of $\mathcal{I}$s for a single cut across the phase diagram for a fixed shaking frequency, f_{sh}. The error bars represent the standard deviation, being non-zero for all shaking phases for which multiple measurements were taken.

Uniform low-frequency phase. In panel (a), we see that the low-frequency training data are all quite similarly influential for the model while predicting that the black-cross test point belongs to the same low-frequency phase. The uniformity in question is also well visible in the cut through the phase diagram in the lower panel of fig. 5.22. Apart from single $\mathcal{I}(\boldsymbol{x}_{\text{train}}, \boldsymbol{x}_{\text{test}})$ with large variation indicating experimental outliers in the

[2]If a reader is familiar with our Ref. [278], they may note that fig. 5.22 features a different colormap (*viridis*) as compared to the mentioned reference (where *jet* has been used). The Ph.D. candidate has learned about the importance of perceptually uniform maps [471] too late, and only now she is correcting her mistake (thank you, Patrick Emonts!).

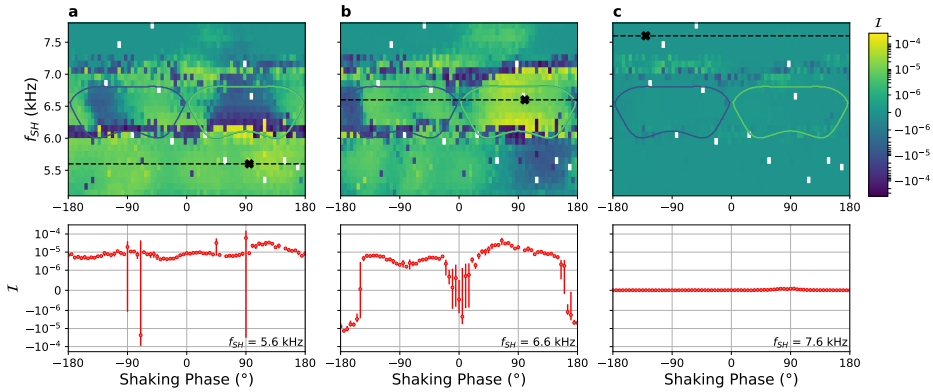

Figure 5.22: **Analysis of data similarity within three anomaly-detected phases with influence functions ($\mathcal{I}$s).** The upper row shows the phase diagrams with the color-coded $\mathcal{I}$ values for all training data and the prediction for the test point marked with a black cross. The test point belongs to (a) low-frequency, (b) intermediate-frequency, and (c) high-frequency phase. The lower row presents the $\mathcal{I}$ values from the single cut through a corresponding phase diagram above at the fixed shaking frequency, f_{sh}. Panels (a) and (c) show quite uniform similarity patterns, while (b) suggests the existence of an additional phase. Note the use of a symmetric logarithmic scale. We mark theoretically-predicted boundaries with continuous blue and green lines. White pixels correspond to the lack of available training data. From Ref. [278].

training set, the formed similarity pattern is quite uniform. Notice, however, the symmetric logarithmic scale for $\mathcal{I}$s. When ignoring the outliers, $\mathcal{I}$ values span almost one order of magnitude. The lowest $\mathcal{I}$ values of around $5 \cdot 10^{-6}$ are located in the negative shaking phase, and the largest $\mathcal{I}$ values around $3 \cdot 10^{-5}$ are for training points which have a positive shaking phase similar to $\boldsymbol{x}_{\mathrm{test}}$. It tells us that the shaking phase is an influential factor in the predictions in the low-frequency phase. However, it is not a determining one. Otherwise, the largest $\mathcal{I}$ values would be much more localized in the shaking phase axis. We also note that the $\mathcal{I}$ values always highlight the boundaries between phases for two reasons. Firstly, data around the phase transitions are usually the most confusing for the model. They are labeled as belonging to either of the phases, being at the same time non-representative of any phase. The second reason is of a purely numerical nature. Regardless if boundaries are placed in accordance with physical ones, the data around the boundaries plays a unique role in the training, containing the most important information for the model. In general, we expect that the confusing phase transition regions, indicated by large $\mathcal{I}$ values, in experimental data

should be broader as compared to numerical studies [414]. It is due to the fact that the experimental system is finite and inhomogeneous and therefore the phase transition is intrinsically broadened.

Uniform high-frequency phase. Panel (c) shows even more uniform behavior. It contains $\mathcal{I}$ values for the test point localized in the high-frequency regime. What may seem surprising is that almost all $\mathcal{I}$ values are practically zero. It means that none of the training points is of significant influence when making the selected prediction. The reason is that the prediction on the test point from panel (c) has an extremely high certainty and it has an impact on the $\mathcal{I}$ values. In fact, $|\mathcal{I}(\boldsymbol{x}_{\text{train}}, \boldsymbol{x}_{\text{test}})|$ values are proportional to the uncertainty of the prediction made on $\boldsymbol{x}_{\text{test}}$. When the prediction's uncertainty is very low, the $\mathcal{I}$ values are also very small.

Two plateaus of similarity in the intermediate-frequency regime. Panel (b) is analogous to the previous panels, but this time the test point, for which $\mathcal{I}$ values are computed, is localized in the intermediate-frequency regime (which we know contains two topological phases). The striking feature of the panel (b) is the lack of uniformity in the intermediate-frequency regime which is well visible in the lower plot of the panel (b), which contains $\mathcal{I}$ values for the single cut through the phase diagram for the fixed frequency of 6.6 kHz. In between two plateaus, i.e., around the shaking phase of 0 and 180°, there are significant dips in the $\mathcal{I}$s' values reaching negative values. They show that the training data in this part of the diagram is different enough to be harmful for the analyzed prediction. They are misleading for the model as they are labeled the same (as belonging to the topological phase) while they are actually quite different. It is analogous to the reason why $\mathcal{I}$s always highlight boundaries between phases. This leads to the conclusion that within the anomaly-detected intermediate-frequency regime there is an additional boundary separating two more phases, which we know to be the $C = +1$ and $C = -1$ phases. Another observation supporting this conclusion is two similarity plateaus on the negative and positive sides of the shaking phase, separated by the detected boundary. They are well visible in the lower plot of the panel (b). The average values of two plateaus differ by almost order of magnitude, indicating two distinct patterns. Simultaneously, these patterns are more similar to each other than to the low- or high-frequency phase, which suggests the similar character of two phases detected in the intermediate-frequency regime.

The similarity analysis described above reveals the existence of two phases within the anomaly-detected topological phase. We note that this analysis is vastly simplified by removing the micromotion phase from the time-of-flight images. Results from section 5.5.3 show that the micromotion phase has been a very influential factor for the trained CNN before the post-processing. Therefore, the analysis would need to include the impact

of the micromotion phase on the CNN's predictions.

5.5.8 Conclusions

In this section, we have applied different unsupervised ML methods for identifying topological phase transitions in experimental data of the Haldane-like model realized with ultracold atoms. The topological phase diagram of the elliptically shaken hexagonal lattice hosts topologically non-trivial phases at an intermediate shaking frequency and trivial phases both for low and high shaking frequencies. Furthermore, the sign of the Chern number changes with the sign of the shaking phase, i.e., the orientation of shaking, giving rise to two distinct non-trivial phases.

A necessary step for successful unsupervised learning has been fixing the micromotion phase inherent to the Floquet realization of the topological phases via a VAE with a question neuron. This post-processing of the experimental data to the desired sampling demonstrated here is an exciting tool on its own. When generalizing such a post-processing procedure to data sets, where known physical observables are invariant under the parameter which is removed, it would be interesting to test if these observables stay the same when evaluated on the pre- and post-processed images.

Both a clustering analysis in an appropriate low-dimensional representation of the data and anomaly detection in the loss function correctly identified the three regions as a function of shaking frequency. The correct identification of the two regions with the opposite sign of the Chern number has only been possible by combining this information with the insights from an $\mathcal{I}$ on supervised training on the incomplete phase diagram. In total, the full phase diagram, which can also be identified via supervised ML on labeled data, have been obtained in a fully unsupervised way by combining the different methods.

The successful identification of the phase diagram demonstrates that unsupervised ML can correctly identify phases even for noisy data and despite the finite temperature of the system. In the future, these methods can be applied to strongly correlated systems to determine corrections to numerical predictions or to exotic quantum many-body systems with unknown phase diagrams or hidden order [277, 300] to support the interpretation of the data and to guide the experimental exploration of the parameter space.

6

Conclusions

Va'esse deireádh aep eigean, va'esse eigh faidh'ar.
~Andrzej Sapkowski

This chapter concludes the thesis and four years of our work. Our results belong to two parallel research paths. Here we summarize the findings described in chapters 3 and 5 and present the outlook and planned research directions for next stages of the scientific journey.

Molecular path

Within the molecular path, we have studied a building block of the extended molecular Hubbard model, i.e., two ultracold interacting molecules in a one-dimensional harmonic trap. The system is presented schematically in fig. 6.1. Two diatomic molecules rotate and interact via isotropic and anisotropic short-range interaction. They also interact via electric dipolar interaction. Within each molecule, there is a coupling between its rotational and spin angular momenta. In the same figure, there is also a scheme of mechanisms underlying results obtained within this thesis and discussed in chapter 3. We have focused on the interplay of the rotational structure and magnetization of the molecular system. They are intertwined via the spin-rotation coupling. The rotational angular momentum can be controlled with the external electric field and anisotropic short-range interaction strength. The magnetization is controlled with the external magnetic field via the Zeeman effect. This scheme clearly shows that electric and magnetic properties of molecules are naturally coupled, and that external magnetic and electric fields can control molecular electric and magnetic properties, respectively.

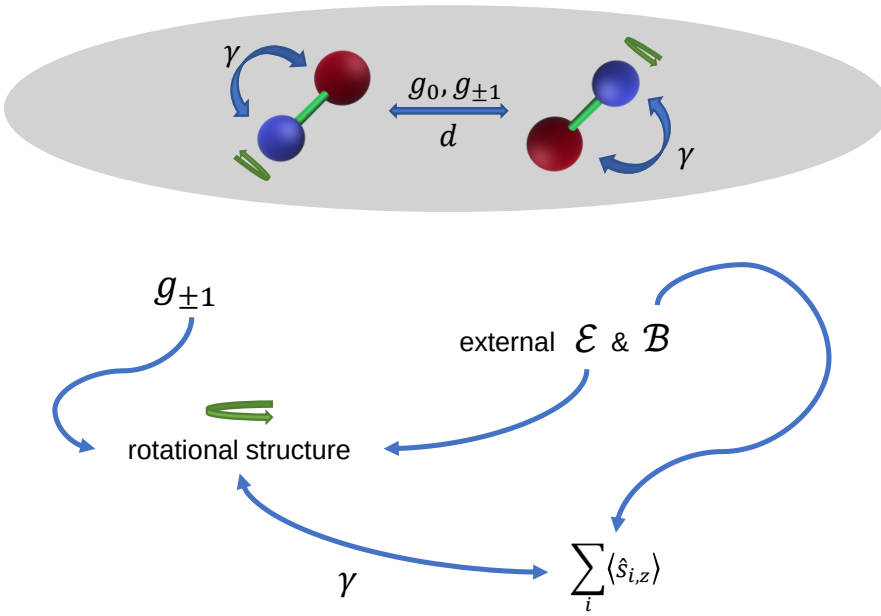

Figure 6.1: **Two interacting molecules in a trap.** The upper row presents the studied system. Iso- and anisotropic short-range interaction strengths are g_0 and $g_{\pm 1}$. γ is the spin-rotation coupling, $\sum_i \langle \hat{s}_{i,z} \rangle$ is the system magnetization. The bottom row shows the interplay between the molecular properties and external fields.

The research within the molecular path has increased our knowledge about the properties, interactions, dynamics, and control of ultracold molecular systems. The results should be especially promising for quantum simulations. Successes of highly magnetic atoms in this field [45–49] promise exciting results for ultracold molecular systems with large magnetic and electric dipole moments. Hopefully, they can simulate new quantum phases, spin models, and other problems of condensed-matter physics. Another interesting example that could be addressed with molecular simulators is multiferroics. They are multifunctional materials that exhibit both electric and magnetic long-range ordered phases. Such materials are candidates for key components of future technology such as memories (fast hard disk drive technologies) and logic devices.

This work will be followed by the analysis of a few highly magnetic and polar molecules in optical tweezers with varied geometry as presented in fig. 6.2. We will focus on the system's potential magnetic frustration and its control with external fields. We will also study the chain and the ladder

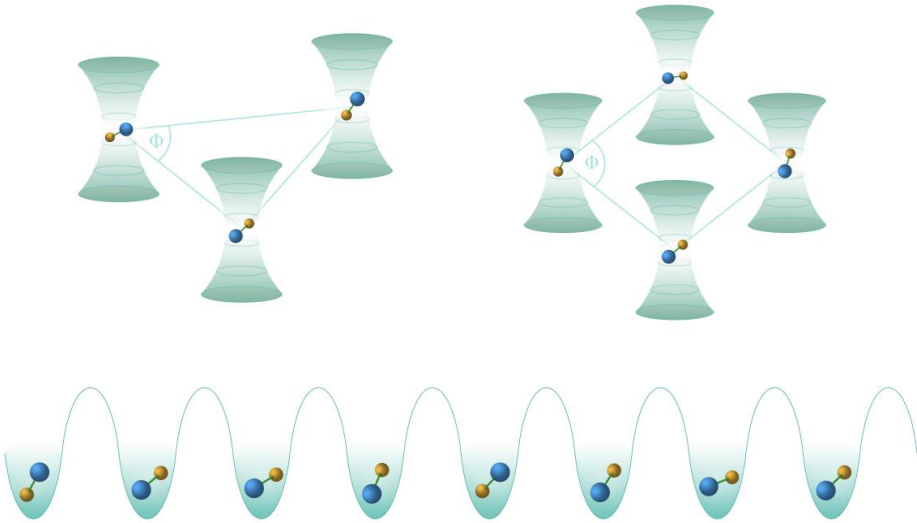

Figure 6.2: **Outlook for the molecular path.** Few- and many-body ultracold systems of interest composed of highly magnetic and polar molecules.

of ultracold molecules with large electric and magnetic dipole moments. We expect the emergence of interesting quasi-particles and topological phases.

Machine learning path

The ML path has concerned the Hessian-based toolbox increasing interpretability and reliability of any ML model. The toolbox requires a single computation of the Hessian of the training loss function at the minimum and provides a deeper understanding of the trained ML model. In particular, it gives a notion of data similarity developed by an ML model thanks to influence functions. Moreover, it provides the estimation of uncertainty of model predictions as well as the estimate of the extrapolation score. As schematically presented in fig. 6.3, we have shown the usefulness of this toolbox on two examples of quantum phase classification problems. Proofs of concept have been presented with data obtained from the exact diagonalization of the one-dimensional extended spinless Fermi-Hubbard Hamiltonian at half-filling. In particular, we have shown how to detect additional phases, anomalies, and out-of-distribution (OOD) points. We have also presented

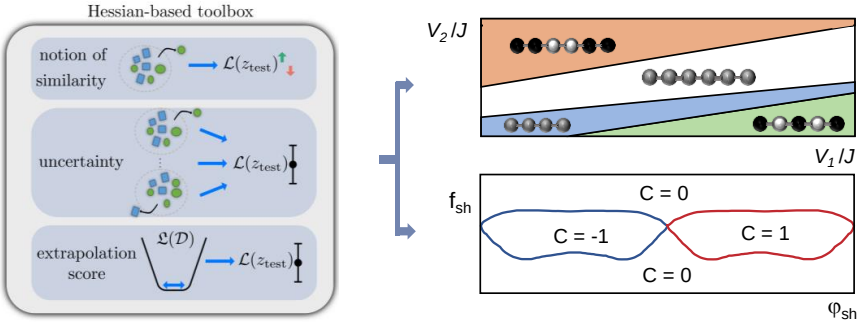

Figure 6.3: **Hessian-based toolbox for phase classification.** We have successfully applied interpretablity and reliability tools to CNNs trained to classify phases of the Fermi-Hubbard and Haldane-like models. Adapted from Refs. [413, 414].

how to check the phase transition width. Finally, on the experimental data obtained from the Floquet realization of the two-dimensional Haldane-like model, we have shown how to obtain a complete topological phase diagram in a fully unsupervised way. Influence functions have been an indispensable element of this discovery.

One of the most fundamental conclusions that this path offers is the following. ML models sometimes pick up additional signals from training data. These signals are additional in a sense that for us (users, physicists, humans) they do not seem relevant for the task at hand. This ability may be connected to the overparametrization of NNs, so their exaggerated capacity as compared to the problem. It can also come from the regularization of ML models that explicitly forbids them from focusing on single data features and forces them to develop parallel information flows. This crucial characteristics of ML is a double-edged sword. For example, when we have trained a CNN to classify three phases (LL, BO, and CDW-II) into two (LL, no-LL), the model developed such a similarity notion that corresponded to reality and allowed for detection of the third phase. But in another example, it may happen that an ML model picks up irrelevant or harmful features which may destroy generalization in some regime. Finally, what can happen is that ML models notice physically irrelevant quantities (like global sign or micromotion) and their predictive powers stay intact. It begs a question what it means for an ML model to be invariant to some property. Fortunately, it seems that learning "sensible" features is more likely than learning noise from the data set, solely due to the nature of the learning process which promotes signals leading in the common direction than random ones which usually cancel.

The ML path has already created various future research directions as

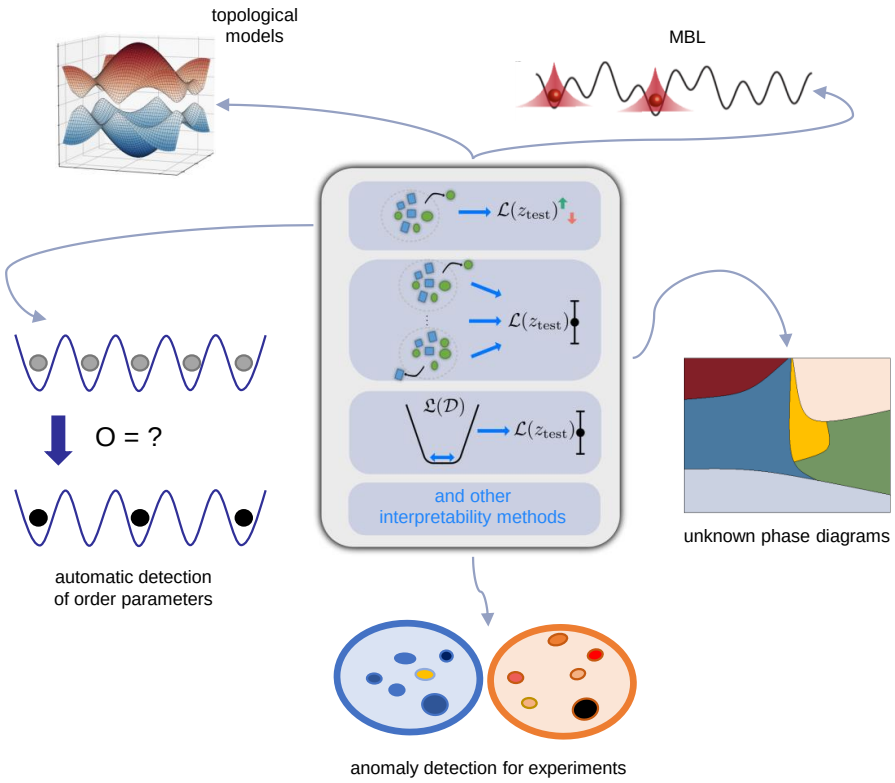

Figure 6.4: **Future research directions** that will be addressed with the Hessian-based toolbox and other interpretability methods.

presented in fig. 6.4. In particular, the Hessian-based toolbox and other interpretability methods can shed a light onto difficulties of ML models with detecting signatures of many-body localization and topological phases. Various interpretability and reliability methods along with different types of learning could be combined for an automatic detection of order parameters governing the data as well as for the complete discovery of unknown phase diagrams. Finally, the anomaly detection schemes can be useful at the intersection of quantum experiments and quantum tomography with ML or Hamiltonian learning. Detection of the most noisy and spurious measurements would allow for their removal from the training set that would improve the ML model. Other ideas include understanding how the notion of similarity evolves during the training of an ML model or how it varies with the increasing overparametrization of a model.

Within this research path, we advocate the use of interpretable and reli-

able ML in science. Without these properties, we cannot learn much from the ML model when applying it to unknown physical systems, nor understand its problems with capturing the topological or many-body localization signatures. Moreover, interpretation and reliability methods indicate whether the ML model has learned the problem accurately and can be trusted. We hope that our work opens the road to the systematic use of interpretability and reliability methods in ML applied to physics and, more generally, science.

List of publications

Molecular path:

1. **Anna Dawid**, Maciej Lewenstein, Michał Tomza. 2018. *Two ultracold interacting molecules in a one-dimensional harmonic trap.* **Phys. Rev. A** 97, 063618.

2. **Anna Dawid**, Michał Tomza. 2020. *Magnetic properties and quench dynamics of two interacting ultracold molecules.* **Phys. Chem. Chem. Phys.** 22, 28140-28153.

Machine learning path:

3. **Anna Dawid**, Patrick Huembeli, Michał Tomza, Maciej Lewenstein, Alexandre Dauphin. 2020. *Phase detection with neural networks: interpreting the black box.* **New J. Phys.** 22, 115001.

4. Niklas Käming, **Anna Dawid**, Korbinian Kottmann, Maciej Lewenstein, Klaus Sengstock, Alexandre Dauphin, Christof Weitenberg. 2021. *Unsupervised machine learning of topological phase transitions from experimental data.* **Mach. Learn.: Sci. Technol.** 2, 035037.

5. **Anna Dawid**, Patrick Huembeli, Michał Tomza, Maciej Lewenstein, Alexandre Dauphin. 2021. *Hessian-based toolbox for more reliable and interpretable machine learning in physics.* **Mach. Learn.: Sci. Technol.** 3, 015002.

Other works:

6. **Anna Dawid** *et al.* 2022. *Modern applications of machine learning in quantum sciences*. In preparation.

7. Michał Suchorowski, **Anna Dawid**, Michał Tomza. 2022. *Two highly magnetic atoms in a trap*. In preparation.

8. Marta Sroczyńska, **Anna Dawid**, Michał Tomza, Tomasso Calarco, Zbigniew Idziaszek, Krzysztof Jachymski. 2021. *Controlling the dynamics of ultracold polar molecules in optical tweezers*. **New J. Phys.** 24, 015001.

9. Leszek Dawid, Michał Tomza, **Anna Dawid**. 2019. *Estimation of usable area of flat-roof residential buildings using topographic data with machine learning methods*. **Remote Sens.** 11(20), 2382.

10. Yinghua Qiu, **Anna Dawid**, Zuzanna S. Siwy. 2017. *Experimental investigation of dynamic deprotonation/protonation of highly charged particles*. **J. Phys. Chem. C** 121(11), 6255-6263.

Code developed within this thesis:

- **Anna Dawid**, Michał Tomza. 2020. *Two-Molecules-1D-Trap*. DOI:10.5281/zenodo.3985911 (C++)

- **Anna Dawid**, Patrick Huembeli, Michał Tomza, Maciej Lewenstein, Alexandre Dauphin. 2020. *Interpretable-Phase-Classification*. DOI:10.5281/zenodo.3759432 (Python & PyTorch)

- Niklas Käming, **Anna Dawid**, Korbinian Kottmann, Maciej Lewenstein, Klaus Sengstock, Alexandre Dauphin, Christof Weitenberg. 2021. *Unsupervised-ML-of-Topological-Experimental-Data*. DOI:10.5281/zenodo.4459311 (Python, PyTorch, & Tensorflow)

- **Anna Dawid**, Patrick Huembeli, Michał Tomza, Maciej Lewenstein, Alexandre Dauphin. 2021. *Hessian-Based-Toolbox* DOI:10.5281/zenodo.5148870(Python & PyTorch)

Nomenclature

Physics quantities

δ_{ij}	Kronecker delta
$\hat{H}$	Hamiltonian
$\langle \hat{r}^2 \rangle$	cloud size
$\langle \hat{S}_z \rangle$	magnetization
$\mathcal{H}$	Hilbert space
k_B	Boltzmann constant

Machine learning quantities

$\boldsymbol{b}$	model biases
K	number of classes in a classification problem
$\tilde{\boldsymbol{\theta}}$	converged $\boldsymbol{\theta}$
$\mathcal{D}$	data set
n	size of $\mathcal{D}$, number of training examples
η	learning rate
m	dimensionality of data point $\boldsymbol{x}$, number of data features
$\hat{f}$	model with converged $\boldsymbol{\theta}$
$\mathcal{I}$	influence function
$\mathcal{L}$	loss (or cost, or error) function
d	size of $\boldsymbol{\theta}$, number of model parameters
$\boldsymbol{\theta}$	model parameters or weights
H	Hessian matrix

List of acronyms

AE	autoencoder
AI	artificial intelligence
BO	bond-order
CDW	charge-density wave
CNN	convolutional neural network
DFT	discrete Fourier transform
DL	deep learning
GAP	global average pooling
$\mathcal{I}$	influence function
LE	local ensemble
LL	Luttinger liquid
LOO	leave-one-out
ML	machine learning
MSE	mean-squared error
NISQ	noisy intermediate-scale quantum
NN	neural network
OOD	out-of-distribution
PCA	principal component analysis
Relat-$\mathcal{I}$	relative influence function
RRKM	Rice-Ramsperger-Kassel-Marcus
RUE	resampling uncertainty estimation
SGD	stochastic gradient descent
VAE	variational autoencoder

List of figures

Appendices

A Hamiltonian matrix elements of molecular systems

Here, we provide matrix elements of the components of the Hamiltonian given by eq. (2.17) defined in eqs. (2.2), (2.4), (2.5) and (2.9) in the computation basis of $|n, J, M_J, j_1, j_2, S, M_S, s_1, s_2\rangle \equiv |n\rangle|J, M_J, j_1, j_2\rangle|S, M_S, s_1, s_2\rangle$ as described in section 2.5. For the remainder of this appendix, for conciseness by $\langle \hat{H}_X \rangle$ we mean:

$$\langle n, J, M_J, j_1, j_2, S, M_S, s_1, s_2|\hat{H}_X|n', J', M', j_1', j_2', S', M_S', s_1', s_2'\rangle. \tag{1}$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{trap}} = \sum_{i=1}^{2} \frac{\hat{p}_i^2}{2m} + \sum_{i=1}^{2} \frac{1}{2} m\omega z_i^2$ describing the motion of molecules in a trap, are:

$$\langle \hat{H}_{\text{trap}} \rangle = \delta_{nn'}\delta_{JJ'}\delta_{MM'}\delta_{j_1 j_1'}\delta_{j_2 j_2'}\delta_{SS'}\delta_{M_S M_S'}\delta_{s_1 s_1'}\delta_{s_2 s_2'}\hbar\omega\left(n + \frac{1}{2}\right). \tag{2}$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{rot}} = \sum_{i=1}^{2} B\,\hat{\mathbf{j}}_i^2$ describing the internal rotational structure of molecules (section 2.2), are:

$$\begin{aligned}\langle \hat{H}_{\text{rot}} \rangle = \delta_{nn'}\delta_{JJ'}\delta_{MM'}\delta_{j_1 j_1'}\delta_{j_2 j_2'}\delta_{SS'}\delta_{M_S M_S'}\delta_{s_1 s_1'}\delta_{s_2 s_2'} \\ \times B\left(j_1(j_1+1) + j_2(j_2+1)\right).\end{aligned} \tag{3}$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{spin–rot}} = \sum_{i=1}^{2} \gamma\, \hat{\mathbf{s}}_i \cdot \hat{\mathbf{j}}_i$ describing the spin-rotation coupling within molecules (section 2.2), are:

$$\begin{aligned}
\langle \hat{H}_{\text{spin–rot}} \rangle &= \delta_{nn'} \delta_{M_{\text{tot}}, M'_{\text{tot}}} \delta_{j_1 j'_1} \delta_{j_2 j'_2} \delta_{s_1 s'_1} \delta_{s_2 s'_2} \\
&\times \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} \sum_{m'_1=-j_1}^{j_1} \sum_{m'_2=-j_2}^{j_2} \langle j_1 m'_1 j_2 m'_2 | J M_J \rangle \langle j_1 m'_1 j_2 m'_2 | J' M'_J \rangle \\
&\times \sum_{m_{s_1}=-s_1}^{s_1} \sum_{m_{s_2}=-s_2}^{s_2} \sum_{m'_{s_1}=-s_1}^{s_1} \sum_{m'_{s_2}=-s_2}^{s_2} \langle s_1 m'_{s_1} j_2 m'_2 | S M_S \rangle \langle s_1 m'_{s_1} s_2 m'_{s_2} | S' M'_S \rangle \\
&\times \Big(\delta_{m_1 m'_1} \delta_{m_2 m'_2} \delta_{m_{s_1} m'_{s_1}} \delta_{m_{s_2} m'_{s_2}} \gamma (m_1 m_{s_1} + m_2 m_{s_2}) \\
&+ \frac{\gamma}{2} \Big(\delta_{m_{s_1}+1, m'_{s_1}} \delta_{m_{s_2}, m'_{s_2}} \delta_{m_1 - 1, m'_1} \delta_{m_2 m'_2} \\
&+ \delta_{m_{s_1}, m'_{s_1}} \delta_{m_{s_2}+1, m'_{s_2}} \delta_{m_1 m'_1} \delta_{m_2 - 1, m'_2} \\
&+ \delta_{m_{s_1}-1, m'_{s_1}} \delta_{m_{s_2}, m'_{s_2}} \delta_{m_1+1, m'_1} \delta_{m_2, m'_2} \\
&+ \delta_{m_{s_1}, m'_{s_1}} \delta_{m_{s_2}-1, m'_{s_2}} \delta_{m_1 m'_1} \delta_{m_2+1, m'_2} \Big) \Big) \,.
\end{aligned} \tag{4}$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{Stark}} = -\sum_{i=1}^{2} \hat{\mathbf{d}}_i \cdot \mathcal{E}$ describing the Stark effect when the electric field is aligned with the system quantization axis ($\mathcal{E}_z \neq 0$) and when it is not ($\mathcal{E}_x \neq 0$), (section 2.4), are:

$$\begin{aligned}
\langle \hat{H}_{\text{Stark}}(\mathcal{E}_z \neq 0) \rangle &= -d\mathcal{E}_z \delta_{nn'} \delta_{SS'} \delta_{M_S M'_S} \delta_{s_1 s'_1} \delta_{s_2 s'_2} \\
&\times \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} \sum_{m'_1=-j'_1}^{j'_1} \sum_{m'_2=-j'_2}^{j'_2} \langle j_1 m_1 j_2 m_2 | J M \rangle \langle j'_1 m'_1 j'_2 m'_2 | J' M' \rangle \\
&\times \delta_{m_1 m'_1} \delta_{m_2 m'_2} \left(\delta_{j_1 \pm 1, j'_1} \delta_{j_2 j'_2} \sqrt{\frac{2j_1+1}{2j'_1+1}} \langle j_1 0 1 0 | j'_1 0 \rangle \langle j_1 m_1 1 0 | j'_1 m_1 \rangle + \right. \\
&\left. + \delta_{j_1 j'_1} \delta_{j_2 \pm 1, j'_2} \sqrt{\frac{2j_2+1}{2j'_2+1}} \langle j_2 0 1 0 | j'_2 0 \rangle \langle j_2 m_2 1 0 | j'_2 m_2 \rangle \right)
\end{aligned} \tag{5}$$

$$\begin{aligned}
\langle \hat{H}_{\text{Stark}}(\mathcal{E}_x \neq 0) \rangle &= \frac{d\mathcal{E}_x}{\sqrt{2}} \delta_{nn'} \delta_{SS'} \delta_{M_S M'_S} \delta_{s_1 s'_1} \delta_{s_2 s'_2} \\
&\times \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} \sum_{m'_1=-j'_1}^{j'_1} \sum_{m'_2=-j'_2}^{j'_2} \langle j_1 m_1 j_2 m_2 | J M \rangle \langle j'_1 m'_1 j'_2 m'_2 | J' M' \rangle \\
&\times \left(\delta_{j_1 \pm 1, j'_1} \delta_{j_2 j'_2} \delta_{m_2 m'_2} \sqrt{\frac{2j_1+1}{2j'_1+1}} \langle j_1 0 1 0 | j'_1 0 \rangle \right. \\
&\left. \times \left(\delta_{m_1+1, m'_1} \langle j_1 m_1 1 1 | j'_1 m'_1 \rangle - \delta_{m_1-1, m'_1} \langle j_1 m_1 1 (-1) | j'_1 m'_1 \rangle \right) \right.
\end{aligned} \tag{6}$$

$$+\delta_{j_1 j_1'}\delta_{j_2\pm1,j_2'}\delta_{m_1 m_1'}\sqrt{\frac{2j_2+1}{2j_2'+1}}\langle j_2 010|j_2' 0\rangle$$
$$\times\left(\delta_{m_2+1,m_2'}\langle j_2 m_2 11|j_2' m_2'\rangle-\delta_{m_2-1,m_2'}\langle j_2 m_2 1(-1)|j_2' m_2'\rangle\right),$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{Zeeman}}=2\mu_B\sum_{i=1}^{2}\hat{\mathbf{s}}_i\cdot\mathcal{B}$ describing the Zeeman effect (section 2.4), are:

$$\langle\hat{H}_{\text{Zeeman}}\rangle=\delta_{nn'}\delta_{JJ'}\delta_{MM'}\delta_{j_1 j_1'}\delta_{j_2 j_2'}\delta_{SS'}\delta_{M_S M_S'}\delta_{s_1 s_1'}\delta_{s_2 s_2'}2\mu_B M_S\mathcal{B}, \tag{7}$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{iso}}=\sum_\alpha g_\alpha\delta(z_1-z_2)|\alpha\rangle\langle\alpha|$ modeling the isotropic short-range interaction between molecules (section 2.3), are:

$$\langle\hat{H}_{\text{iso}}\rangle=\delta_{JJ'}\delta_{MM'}\delta_{j_1 j_1'}\delta_{j_2 j_2'}\delta_{SS'}\delta_{M_S M_S'}\delta_{s_1 s_1'}\delta_{s_2 s_2'}g_0\varphi_n(0)\varphi_{n'}(0), \tag{8}$$

The elements of the Hamiltonian matrix, $\hat{H}_{\text{aniso}}=\sum_{\alpha\neq\alpha'}g_{\alpha\alpha'}\delta(z_1-z_2)$ $(|\alpha\rangle\langle\alpha'|+\text{h.c})$ modeling the anisotropic short-range interaction between molecules (section 2.3), are:

$$\begin{aligned}\langle\hat{H}_{\text{aniso}}\rangle=\delta_{JJ'}\delta_{MM'}\delta_{j_1\pm k,j_1'}\delta_{j_2\mp k,j_2'}\delta_{SS'}\delta_{M_S M_S'}\delta_{s_1 s_1'}\delta_{s_2 s_2'}\\ g_{\pm k}\varphi_n(0)\varphi_{n'}(0),\end{aligned} \tag{9}$$

Finally, the elements of the Hamiltonian matrix, $\hat{H}_{\text{dip}}=-\delta(z_1-z_2)\left(2\hat{d}_{1,0}\hat{d}_{2,0}\right.$ $\left.+\hat{d}_{1,1}\hat{d}_{2,-1}+\hat{d}_{1,-1}\hat{d}_{2,1}\right)$ modeling the electric dipole-dipole interaction between molecules (section 2.3), are:

$$\begin{aligned}\langle\hat{H}_{\text{dip}}\rangle=-d^2\varphi_n(0)\varphi_{n'}(0)\delta_{SS'}\delta_{M_S M_S'}\delta_{s_1 s_1'}\delta_{s_2 s_2'}\delta_{j_1\pm1,j_1'}\delta_{j_2\pm1,j_2'}\\ \times\sqrt{\left(\frac{2j_1+1}{2j_1'+1}\right)\left(\frac{2j_2+1}{2j_2'+1}\right)}\langle j_1 010|j_1' 0\rangle\langle j_2 010|j_2' 0\rangle\\ \times\sum_{m_1=-j_1}^{j_1}\sum_{m_2=-j_2}^{j_2}\sum_{m_1'=-j_1'}^{j_1'}\sum_{m_2'=-j_2'}^{j_2'}\\ \left(\delta_{m_1+1,m_1'}\delta_{m_2-1,m_2'}\langle j_1 m_1 11|j_1' m_1'\rangle\langle j_2 m_2 1(-1)|j_2' m_2'\rangle\right.\\ +\delta_{m_1-1,m_1'}\delta_{m_2+1,m_2'}\langle j_1 m_1 1(-1)|j_1' m_1'\rangle\langle j_2 m_2 11|j_2' m_2'\rangle+\\ \left.+2\delta_{m_1 m_1'}\delta_{m_2 m_2'}\langle j_1 m_1 10|j_1' m_1\rangle\langle j_2 m_2 10|j_2' m_2\rangle\right),\end{aligned} \tag{10}$$

where δ_{ij} is the Kronecker delta and $\varphi_n(0)$ is a harmonic oscillator wave function for state n at point $z=0$.

B Derivation of the molecular cloud size

The wave function of the studied molecular system is following:

$$|\Psi_k\rangle = \sum_{\substack{n,J,M,j_1,j_2,\\ S,M_S,s_1,s_2}} C^k_{n,J,M,j_1,j_2,S,M_S,s_1,s_2} |n\rangle |J,M,j_1,j_2\rangle |S,M_S,s_1,s_2\rangle \,, \quad (11)$$

where

$$\begin{aligned} |n\rangle &= \frac{1}{\sqrt{2^n n!}} \pi^{-1/4} \exp\left\{ \frac{-z^2}{2} H_n(z) \right\}, \\ |J,M,j_1,j_2\rangle &= \sum_{m_1,m_2} \langle j_1,m_1,j_2,m_2|J,M\rangle |j_1,m_1\rangle |j_2,m_2\rangle \,, \\ |S,M_S,s_1,s_2\rangle &= \sum_{m_{s_1},m_{s_2}} \langle s_1,m_{s_1},s_2,m_{s_2}|S,M_S\rangle |s_1,m_{s_1}\rangle |s_2,m_{s_2}\rangle \,. \end{aligned} \quad (12)$$

where H_n are the Hermite polynomials, $\langle j_1,m_1,j_2,m_2|J,M\rangle$ and $\langle s_1,m_{s_1},s_2,m_{s_2}|S,M_S\rangle$ are the Clebsch-Gordan coefficients, while $|j_i,m_i\rangle$ and $|s_i,m_{s_i}\rangle$ are the eigenfunctions of the rotational and spin angular momenta of the molecule i. The size of the molecular cloud is then:

$$\begin{aligned} \langle \hat{r}^2 \rangle &= \langle \Psi_k | \hat{r}^2 | \Psi_k \rangle = \sum_{n,n'} \langle n | \hat{r}^2 | n' \rangle = \\ &= \sum_{n,n'} \frac{1}{\sqrt{2^{n+n'} n! n'!}} \pi^{-\frac{1}{2}} \int_0^\infty dr\, r^2 e^{-r^2} H_n(r) H_{n'}(r) \end{aligned} \quad (13)$$

$$\begin{aligned} &\int_0^\infty dr\, r^2 e^{-r^2} H_n(r) H_{n'}(r) \overset{①}{=} \\ &= n!\, n'! \sum_{N=0}^{\min n,n'} \frac{2^N}{(n-N)!(n'-N)!N!} \int_0^\infty dr\, r^2 e^{-r^2} H_{n+n'-2N}(r) \\ &\overset{②}{=} n!\, n'! \sum_{N=0}^{\min n,n'} \frac{2^N (n+n'-2N)!}{(n-N)!(n'-N)!N!} \\ &\times \sum_{k=0}^{\mathrm{floor}(\frac{n+n'}{2}-N)} \frac{(-1)^k 2^{n+n'-2N-2k-1} \Gamma(\frac{n+n'-2N+3}{2}-k)}{k!(n+n'-2N-2k)!} \end{aligned} \quad (14)$$

Hermite polynomials' properties used in calculations:

① $H_m(z)H_n(z) = m!n! \sum_{N=0}^{\min m,n} \frac{2^N H_{m+n-2N}(z)}{(m-N)!(n-N)!N!}$ [472],

② $\int_0^\infty t^2 e^{-t^2} H_n(t) dt = n! \sum_{k=0}^{\mathrm{floor}(\frac{n}{2})} \frac{(-1)^k 2^{n-2k-1} \Gamma(\frac{n+3}{2}-k)}{k!(n-2k)!}$ [473].

C Leave-one-out training

Influence functions ($\mathcal{I}$s) are an approximation of the leave-one-out (LOO) training [415]. Hence, to understand the idea behind influence functions we should first start with the concept of LOO.

Its name speaks for itself: for an LOO training one retrains an ML algorithm (in our case a CNN) while leaving one training example $z_\mathrm{r} = (x_\mathrm{r}, y_\mathrm{r})$ away. The aim of such a retraining is to find out how the loss $\mathcal{L}(z_\mathrm{test})$ for a given test example $z_\mathrm{test} = (x_\mathrm{test}, y_\mathrm{test})$ changes compared to the training on the whole training set $\mathcal{D} = \{z_i\}_{i=1}^n$ which also contains z_r. More concretely, after training an NN on the full data set $\mathcal{D}$, one obtains the parameters $\tilde{\boldsymbol{\theta}}$ that minimize the training error. When retraining the NN on the training set without z_r, i.e., $\mathcal{D} \setminus \{z_\mathrm{r}\}$ one obtains a new minimum of the training error with the parameters $\tilde{\boldsymbol{\theta}}_{\mathcal{D}\setminus\{z_\mathrm{r}\}}$.

Because of the highly non-convex nature of NNs, it is important that the retraining does not start from a new random initialization of the parameters, but from the minimum $\tilde{\boldsymbol{\theta}}$. Moreover, to make sure that the shift in the minimum comes only (or at least mostly) from the removal of a single training poine, it is useful to remove all sources of stochasticity. For example, disable all Dropout layers during LOO training. In particular, keep all random seeds fixed. If possible, load all data in one batch. Also, before starting the LOO training, do a sanity check with a selected (small!) learning rate that you intend to use in the LOO. If you see any changes of the training loss function at this point, think whether you have for sure converged yet to the minimum or of stochasticity that you have not removed yet or rethink the selection of the learning rate. Maybe take the average of these jumps and treat it as a baseline for LOO results. Last but not least, remember to use regularization. In our numerical experiments with an Iris data set, we have seen limited correlation between LOO results and influence functions when the NN had zero regularization.

After the LOO training, make a prediction with your new model on given test point and check the change of the test loss, $\Delta\mathcal{L} \equiv \mathcal{L}(z_\mathrm{test}, \tilde{\boldsymbol{\theta}}) - \mathcal{L}(z_\mathrm{test}, \tilde{\boldsymbol{\theta}}_{\mathcal{D}\setminus\{z_\mathrm{r}\}})$ before and after the retraining. The influence of a training point can be both helpful ($\Delta\mathcal{L} > 0$) and harmful ($\Delta\mathcal{L} < 0$). Why certain training examples are harmful or helpful for the prediction of a test example is mostly due to the similarity of features of z_r and x_test. If the two examples have very similar features and the same label, the NN can learn by example and therefore will be able to predict the label reliably. If they have very similar features but different labels it is likely that the NN gets confused by the training example. In general training points being the closest in the $\Delta\mathcal{L}$ space can be understood as the most similar.

Once the most influential points are identified, we can analyze what

characteristics and features are relevant to the NN to see them as most similar and influential to each other. This can be especially useful in phase classification problems where the analysis of $\Delta\mathcal{L}$ enables the recovery of patterns being crucial for distinguishing the phases. The use of this technique to check the influence of every training point in $\mathcal{D}$ on a given test point is, however, prohibitively expensive, as the model has to be retrained for each removed z.

D Derivation of influence functions

This appendix mostly follows appendix A of Ref. [411]. The derivation of influence functions ($\mathcal{I}$s) starts with an observation that removal of a training point z is mathematically the same as upweighting (or changing the weight) a training point by $\epsilon = -\frac{1}{n}$, where n is the size of the training set:

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_{-z} &= \arg\min_{\theta} \frac{1}{n} \sum_{z_i \neq z} L(z_i, \theta) \equiv \\ \equiv \tilde{\boldsymbol{\theta}}_{\epsilon,z} &= \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} L(z_i, \theta) + \epsilon L(z, \theta) \mid_{\epsilon=-\frac{1}{n}} . \end{aligned} \tag{15}$$

Firstly, let us derive the change of the model parameters after upweighting a training point.

How upweighting a training point changes model parameters

$$\mathrm{I}_{\text{up, params}}(z) \stackrel{\text{def}}{=} \left. \frac{d\tilde{\boldsymbol{\theta}}_{\epsilon,z}}{d\epsilon} \right|_{\epsilon=0} = ? \tag{16}$$

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_{\epsilon,z} &= \arg\min_{\boldsymbol{\theta}} \{ \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(z_i, \boldsymbol{\theta}) + \epsilon \mathcal{L}(z, \boldsymbol{\theta}) \} = \\ &= \arg\min_{\boldsymbol{\theta}} \{ \mathcal{R}(\boldsymbol{\theta}) + \epsilon \mathcal{L}(z, \boldsymbol{\theta}) \} \end{aligned} \tag{17}$$

Since $\tilde{\boldsymbol{\theta}}_{\epsilon,z}$ minimizes $\{\mathcal{R}(\boldsymbol{\theta}) + \epsilon\mathcal{L}(z, \boldsymbol{\theta})\}$, it meets the following condition:

$$\nabla\mathcal{R}(\tilde{\boldsymbol{\theta}}_{\epsilon,z}) + \epsilon\nabla\mathcal{L}(z, \boldsymbol{\theta}) = 0$$

Now we can use Taylor expansion of $\{\mathcal{R}(\boldsymbol{\theta}) + \epsilon\mathcal{L}(z, \boldsymbol{\theta})\}$ and expand it around $\tilde{\boldsymbol{\theta}}$, using the fact that with $\epsilon \to 0$, $\tilde{\boldsymbol{\theta}}_{\epsilon,z} \to \tilde{\boldsymbol{\theta}}$:

$$\{\nabla\mathcal{R}(\tilde{\boldsymbol{\theta}}) + \epsilon\nabla\mathcal{L}(z,\tilde{\boldsymbol{\theta}})\} + \overbrace{(\tilde{\boldsymbol{\theta}}_{\epsilon,z} - \tilde{\boldsymbol{\theta}})}^{\equiv\Delta_\epsilon}\{\nabla^2\mathcal{R}(\tilde{\boldsymbol{\theta}}) + \epsilon\nabla^2\mathcal{L}(z,\tilde{\boldsymbol{\theta}})\} \approx 0 \tag{18}$$

$$\Delta_\epsilon \approx -\frac{\nabla\mathcal{R}(\tilde{\boldsymbol{\theta}}) + \epsilon\nabla\mathcal{L}(z,\tilde{\boldsymbol{\theta}})}{\nabla^2\mathcal{R}(\tilde{\boldsymbol{\theta}}) + \epsilon\nabla^2\mathcal{L}(z,\tilde{\boldsymbol{\theta}})} \tag{19}$$

We note that $\nabla\mathcal{R}(\tilde{\boldsymbol{\theta}}) = 0$ as $\tilde{\boldsymbol{\theta}}$ minimizes $\mathcal{R}$ (by definition provided in eq. (17)). We keep only $O(\epsilon)$ terms. As $\nabla^2\mathcal{L}(z,\tilde{\boldsymbol{\theta}})$ is proportional to ϵ (probably you can check it rigorously using Schwartz norm), we ignore $\epsilon\nabla^2\mathcal{L}(z,\tilde{\boldsymbol{\theta}})$ as it is of order $O(\epsilon^2)$, and we have:

$$\Delta_\epsilon \approx -\epsilon\frac{\nabla\mathcal{L}(z,\tilde{\boldsymbol{\theta}})}{\nabla^2\mathcal{R}(\tilde{\boldsymbol{\theta}})}\,.$$

$\nabla^2\mathcal{R}(\tilde{\boldsymbol{\theta}}) = -\frac{1}{n}\sum_{i=1}^{n}\nabla^2\mathcal{L}(z_i,\tilde{\boldsymbol{\theta}})$ is nothing else but a Hessian calculated at the minimum/optimum model's parameters. In eq. (18) we made a silent, yet very non-trivial to meet, assumption, that $\mathcal{R}$ is twice-differentiable and strictly convex in $\tilde{\boldsymbol{\theta}}$ what provides that Hessian exists, is positive definite (PD), and therefore has an inverse. It means that it has only positive eigenvalues. Noting that $\frac{d\Delta_\epsilon}{d\epsilon} = \frac{d(\tilde{\boldsymbol{\theta}}_{\epsilon,z}-\tilde{\boldsymbol{\theta}})}{d\epsilon} = \frac{d\tilde{\boldsymbol{\theta}}_{\epsilon,z}}{d\epsilon}$, as $\tilde{\boldsymbol{\theta}}$ does not depend on ϵ allows us to write the final result:

$$\mathcal{I}_{\text{up, params}}(z) \stackrel{\text{def}}{=} \left.\frac{d\tilde{\boldsymbol{\theta}}_{\epsilon,z}}{d\epsilon}\right|_{\epsilon=0} = -H_{\tilde{\boldsymbol{\theta}}}^{-1}\nabla\mathcal{L}(z,\tilde{\boldsymbol{\theta}})\,. \tag{20}$$

We made therefore two assumptions on the way.

1. **Hessian is PD $\Leftrightarrow$ the loss function is strictly convex**. If it has eigenvalues equal to zero, its inverse does not exist.

2. Small additional assumption: we assumed that the loss function derivatives are continuous around the minimum, because then, from Schwartz theorem, one can change the order of derivatives in Hessian without consequences, and Hessian is equal to its transpose.

How removing a training point z impacts the test loss of z_{test}

In the previous section, we have obtained that the influence of upweighting z on the model parameters $\tilde{\boldsymbol{\theta}}$ is:

$$\mathcal{I}_{\text{up, params}}(z) = \left.\frac{d\tilde{\boldsymbol{\theta}}_{\epsilon,z}}{d\epsilon}\right|_{\epsilon=0} = -H_{\tilde{\boldsymbol{\theta}}}^{-1}\nabla_\theta\mathcal{L}(z,\tilde{\boldsymbol{\theta}})\,, \tag{21}$$

where $H_{\tilde{\boldsymbol{\theta}}} = \frac{1}{n}\sum_i \nabla^2_\theta L(z_i, \tilde{\boldsymbol{\theta}})$ is the Hessian of the empirical risk calculated for parameters corresponding to the minimum, $\tilde{\boldsymbol{\theta}}$. In the last step we can find how upweighting z affects the loss of a test point z_{test} by applying the chain rule.

$$\begin{aligned}\mathcal{I}_{\text{up, loss}}(z, z_{\text{test}}) &= \frac{d\mathcal{L}(z_{\text{test}}, \tilde{\boldsymbol{\theta}}_{\epsilon,z})}{d\epsilon}\Big|_{\epsilon=0} = \\ &= -\nabla_\theta \mathcal{L}(z_{\text{test}}, \tilde{\boldsymbol{\theta}})^T H^{-1}_{\tilde{\boldsymbol{\theta}}} \nabla_\theta \mathcal{L}(z, \tilde{\boldsymbol{\theta}})\end{aligned} \tag{22}$$

Finally, since removing a point from the training set is the same as upweighting it by $\epsilon = -1/n$ we can linearly approximate:

$$\mathcal{I} \approx -\frac{1}{n}\mathcal{I}_{\text{up,loss}}(z, z_{\text{test}}) = \frac{1}{n}\nabla_\theta \mathcal{L}(z_{\text{test}}, \tilde{\boldsymbol{\theta}})^T H^{-1}_{\tilde{\boldsymbol{\theta}}} \nabla_\theta \mathcal{L}(z, \tilde{\boldsymbol{\theta}})\,. \tag{23}$$

E Stochastic estimation of the Hessian's inverse

The stochastic estimation of the inversed Hessian is based on the so-called von Neumann series and was proposed by Ref. [426].

Matrix inversion with von Neumann series

Von Neumann series can be written for any squared matrix, T:

$$S_n = \sum_{k=0}^{n-1} T^k = I + T + \ldots + T^{n-1}, \quad S_0 = I\,. \tag{24}$$

S_n is called the geometric series generated by T.

Note the following relation (analogous to geometric series in $\mathbb{R}$):

$$\begin{aligned}(I-T)S_n = (I-T)\sum_{k=0}^{n-1} T^k = (I-T)(I + T + T^2 + \ldots + T^{n-1} = \\ = I - T + T - T^2 + T^2 - T^3 + \ldots + T^{n-1} - T^n = \\ = I - T^n\end{aligned} \tag{25}$$

Series $S_n = \sum_{k=0}^{n-1} T^k$ converges for $n \to \infty$ if and only if $||T|| < 1 \Leftrightarrow |\lambda_i| < 1$ for each eigenvalue λ_i of T. If this condition holds, then T^n with $n \to \infty$ is vanishingly small, $I - T$ is invertible, and we have:

$$\sum_{k=0}^{\infty} T^k = (I-T)^{-1} \tag{26}$$

After plugging $M = I - T$:

$$M^{-1} = \sum_{k=0}^{\infty} (I-M)^k \tag{27}$$

This gives an approximation for the matrix inversion by truncating eq. (27) at finite k.

Stochastic estimation of the inversed Hessian

Following the derivation from the previous section, we use the expansion from eq. (27) to recursively calculate the Hessian. We define the finite series for H as

$$H_j^{-1} := \sum_{i=0}^{j} (I - H)^i \tag{28}$$

and can rewrite it recursively as

$$H_j^{-1} := I + (I - H)H_{j-1}^{-1} \tag{29}$$

It can be validated by replugging H_{j-1}^{-1} into the equation or simply by expanding the sum and reindexing the elements. As long as the eigenvalues of the Hessian are smaller than 1, the series converge, and in our calculations we have needed up to 1000 iterations (the number of iterations is called in Ref. [411] and in our code [434, 435] *the recursion depth, t*).

What to do if the eigenvalues of the Hessian are not smaller than 1? Then before making an approximation, you need to scale down your Hessian (our code [434, 435], the scale is called *scale*), and it should be a little larger than the largest absolute eigenvalue. Usually, if not always, it is a positive one, of course. Then, after getting an approximate inverse of the Hessian, you need to scale it back.

We can make additional approximation to make computation even more feasible. Instead of calculating the Hessian for the whole training set, we can approximate it by calculating it only on a batch of training points sampled from the whole set at every step. To reduce the variance of the results, the series should be then calculated a few times (this number is called in in Ref. [411] and in our code [434, 435]), and the results should be averaged. Koh & Liang [411] showed that for $tr < n$, training set size, i.e. when not all training points were taken into account, the approximation for the inverse Hessian was already good enough to get close to 1:1 ratio between LOO trainings and $\mathcal{I}$s for a CNN trained on the MNIST database.

We substitute therefore the full Hessian H by an estimator $\mathbb{E}[H]$. We sample test points z_i and calculate the estimators $\nabla_\theta^2 L(z_i, \theta)$. In Ref. [426], they are referred to as samples X_i. Starting with the identity $H_0^{-1} v = v$ and using the recursion from before, we arrive to:

$$\tilde{H}_t^{-1} v := v + (I - \nabla_\theta^2 L(z_t, \theta))\tilde{H}_{t-1}^{-1} v \tag{30}$$

This estimation is efficient due to two main reasons. Firstly, it takes advantage of results by Pearlmutter (1994) [474] who rigorously proved that the time complexity of calculating the so-called Hessian-vector product is approximately same as time complexity of gradients computations which is linear in number of parameters $\mathcal{O}(\left[\nabla^2_\theta \mathcal{L}\left(z_i, \tilde{\boldsymbol{\theta}}\right)\right] v) \sim \mathcal{O}(\nabla_\theta \mathcal{L}\left(z_i, \tilde{\boldsymbol{\theta}}\right)) \sim \mathcal{O}(d)$.

Bibliography

[1] D. L. Donoho, *Aide-memoire. High-dimensional data analysis: the curses and blessings of dimensionality*, American Math. Society Lecture – Math Challenges of the 21st Century , 1 (2000).

[2] M. L. Wall, *Quantum many-body physics of ultracold molecules in optical lattices: Models and simulation methods*, Ph.D. thesis, Colorado School of Mines (2012).

[3] P. W. Anderson, *More is different*, Science **177**, 393 (1972).

[4] R. P. Feynman, *Simulating physics with computers*, Int. J. Theor. Phys. **21** (1982), 10.1007/bf02650179.

[5] S. Lloyd, *Universal quantum simulators*, Science **273**, 1073 (1996).

[6] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, *Noisy intermediate-scale quantum algorithms*, Rev. Mod. Phys. **94**, 015004 (2022).

[7] J. I. Cirac and P. Zoller, *Goals and opportunities in quantum simulation*, Nat. Phys. **8**, 264 (2012).

[8] I. Bloch, J. Dalibard, and W. Zwerger, *Many-body physics with ultracold gases*, Rev. Mod. Phys. **80**, 885 (2008).

[9] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. W. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, P. Richerme, C. Senko, and N. Y. Yao, *Programmable quantum simulations of spin systems with trapped ions*, Rev. Mod. Phys. **93**, 025001 (2021).

[10] L. Lamata, A. Parra-Rodriguez, M. Sanz, and E. Solano, *Digital-analog quantum simulations with superconducting circuits*, Adv. Phys.: X **3**, 1457981 (2018).

[11] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, *Superconducting qubits: Current state of play*, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[12] M. Tomza, K. Jachymski, R. Gerritsma, A. Negretti, T. Calarco, Z. Idziaszek, and P. S. Julienne, *Cold hybrid ion-atom systems*, Rev. Mod. Phys. **91**, 035001 (2019).

[13] M. Lewenstein, A. Sanpera, and V. Ahufinger, *Ultracold atoms in optical lattices: Simulating quantum many-body systems* (Oxford University Press, 2012).

[14] I. Bloch, J. Dalibard, and S. Nascimbène, *Quantum simulations with ultracold quantum gases*, Nat. Phys. **8**, 267 (2012).

[15] M. Randeria and E. Taylor, *Crossover from Bardeen-Cooper-Schrieffer to Bose-Einstein condensation and the unitary Fermi gas*, Annu. Rev. Condens. Matter Phys. **5**, 209 (2014).

[16] M. Lewenstein, A. Sanpera, V. Ahufinger, B. Damski, A. Sen De, and U. Sen, *Ultracold atomic gases in optical lattices: Mimicking condensed matter physics and beyond*, Adv. Phys. **56**, 243 (2007).

[17] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch, *Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms*, Nature **415**, 39 (2002).

[18] M. Greiner, O. Mandel, T. W. Hänsch, and I. Bloch, *Collapse and revival of the matter wave field of a Bose-Einstein condensate*, Nature **419**, 51 (2002).

[19] B. Paredes, A. Widera, V. Murg, O. Mandel, S. Fölling, I. Cirac, G. V. Shlyapnikov, T. W. Hänsch, and I. Bloch, *Tonks-Girardeau gas of ultracold atoms in an optical lattice*, Nature **429**, 277 (2004).

[20] D. Greif, T. Uehlinger, G. Jotzu, L. Tarruell, and T. Esslinger, *Short-range quantum magnetism of ultracold fermions in an optical lattice*, Science **340**, 1307 (2013).

[21] C. Gross and I. Bloch, *Quantum simulations with ultracold atoms in optical lattices*, Science **357**, 995 (2017).

[22] N. Goldman, J. C. Budich, and P. Zoller, *Topological quantum matter with ultracold gases in optical lattices*, Nat. Phys. **12**, 639 (2016).

[23] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, *A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice*, Nature **462**, 74–77 (2009).

[24] J. F. Sherson, C. Weitenberg, M. Endres, M. Cheneau, I. Bloch, and S. Kuhr, *Single-atom-resolved fluorescence imaging of an atomic Mott insulator*, Nature **467**, 68–72 (2010).

[25] L. W. Cheuk, M. A. Nichols, M. Okan, T. Gersdorf, V. V. Ramasesh, W. S. Bakr, T. Lompe, and M. W. Zwierlein, *Quantum-gas microscope for fermionic atoms*, Phys. Rev. Lett. **114**, 193001 (2015).

[26] E. Haller, J. Hudson, A. Kelly, D. A. Cotta, B. Peaudecerf, G. D. Bruce, and S. Kuhr, *Single-atom imaging of fermions in a quantum-gas microscope*, Nat. Phys. **11**, 738–742 (2015).

[27] M. Boll, T. A. Hilker, G. Salomon, A. Omran, J. Nespolo, L. Pollet, I. Bloch, and C. Gross, *Spin- and density-resolved microscopy of antiferromagnetic correlations in Fermi-Hubbard chains*, Science **353**, 1257 (2016).

[28] T. Fukuhara, P. Schauß, M. Endres, S. Hild, M. Cheneau, I. Bloch, and C. Gross, *Microscopic observation of magnon bound states and their dynamics*, Nature **502**, 76–79 (2013).

[29] T. Fukuhara, A. Kantian, M. Endres, M. Cheneau, P. Schauß, S. Hild, D. Bellem, U. Schollwöck, T. Giamarchi, C. Gross, I. Bloch, and S. Kuhr, *Quantum dynamics of a mobile spin impurity*, Nat. Phys. **9**, 235–241 (2013).

[30] J.-y. Choi, S. Hild, J. Zeiher, P. Schauss, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, *Exploring the many-body localization transition in two dimensions*, Science **352**, 1547 (2016).

[31] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, ..., and M. Greiner, *A cold-atom Fermi-Hubbard antiferromagnet*, Nature **545**, 462 (2017).

[32] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, *String patterns in the doped Hubbard model*, Science **365**, 251 (2019).

[33] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, *Atom-by-atom assembly of defect-free one-dimensional cold atom arrays*, Science **354**, 1024 (2016).

[34] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, *Cold matter assembled atom-by-atom*, arXiv:1607.03044 (2016).

[35] D. Barredo, S. de Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys, *An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays*, Science **354**, 1021 (2016).

[36] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, *Probing many-body dynamics on a 51-atom quantum simulator*, Nature **551**, 579–584 (2017).

[37] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, *Quantum phases of matter on a 256-atom programmable quantum simulator*, Nature **595**, 227 (2021).

[38] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletić, and M. D. Lukin, *Probing topological spin liquids on a programmable quantum simulator*, Science **374**, 1242 (2021).

[39] C. Sheng, J. Hou, X. He, K. Wang, R. Guo, J. Zhuang, B. Mamat, P. Xu, M. Liu, J. Wang, and M. Zhan, *Defect-free arbitrary-geometry assembly of mixed-species atom arrays*, Phys. Rev. Lett. **128**, 083202 (2022).

[40] K. Singha, S. Anand, A. Pocklington, J. T. Kemp, and H. Bernien, *A dual-element, two-dimensional atom array with continuous-mode operation*, arXiv:2110.05515 (2022).

[41] T. Kinoshita, T. Wenger, and D. S. Weiss, *Observation of a one-dimensional Tonks-Girardeau gas*, Science **305**, 1125 (2004).

[42] E. Haller, M. Gustavsson, M. J. Mark, J. G. Danzl, R. Hart, G. Pupillo, and H.-C. Nägerl, *Realization of an excited, strongly correlated quantum gas phase*, Science **325**, 1224 (2009).

[43] G. Pagano, M. Mancini, G. Cappellini, P. Lombardi, F. Schäfer, H. Hu, X.-J. Liu, J. Catani, C. Sias, M. Inguscio, and L. Fallani, *A one-dimensional liquid of fermions with tunable spin*, Nat. Phys. **10**, 198 (2014).

[44] T. Giamarchi, *Quantum physics in one dimension* (Clarendon Press, Oxford, 2004).

[45] S. Baier, M. J. Mark, D. Petter, K. Aikawa, L. Chomaz, Z. Cai, M. Baranov, P. Zoller, and F. Ferlaino, *Extended Bose-Hubbard models with ultracold magnetic atoms*, Science **352**, 201 (2016).

[46] I. Ferrier-Barbut, H. Kadau, M. Schmitt, M. Wenzel, and T. Pfau, *Observation of quantum droplets in a strongly dipolar Bose gas*, Phys. Rev. Lett. **116**, 215301 (2016).

[47] M. Schmitt, M. Wenzel, F. Böttcher, I. Ferrier-Barbut, and T. Pfau, *Self-bound droplets of a dilute magnetic quantum liquid*, Nature **59**, 259–262 (2016).

[48] L. Chomaz, S. Baier, D. Petter, M. J. Mark, F. Wächtler, L. Santos, and F. Ferlaino, *Quantum-Fluctuation-Driven Crossover from a Dilute Bose-Einstein Condensate to a Macrodroplet in a Dipolar Quantum Fluid*, Phys. Rev. X **6**, 041039 (2016).

[49] K. Aikawa, S. Baier, A. Frisch, M. Mark, C. Ravensbergen, and F. Ferlaino, *Observation of Fermi surface deformation in a dipolar quantum gas*, Science **345**, 1484 (2014).

[50] J. Dalibard, F. Gerbier, G. Juzeliunas, and P. Öhberg, *Colloquium: Artificial gauge potentials for neutral atoms*, Rev. Mod. Phys. **83**, 1523 (2011).

[51] N. R. Cooper, J. Dalibard, and I. B. Spielman, *Topological bands for ultracold atoms*, Rev. Mod. Phys. **91**, 015005 (2019).

[52] M. Bukov, L. D'Alessio, and A. Polkovnikov, *Universal high-frequency behavior of periodically driven Systems: from dynamical stabilization to Floquet engineering*, Adv. Phys. **64**, 139 (2015).

[53] A. Eckardt, *Colloquium: Atomic quantum gases in periodically driven optical lattices*, Rev. Mod. Phys. **89**, 011004 (2017).

[54] I. Bloch, *Ultracold quantum gases in optical lattices*, Nat. Phys. **1**, 23 (2005).

[55] K.-N. Schymik, V. Lienhard, D. Barredo, P. Scholl, H. Williams, A. Browaeys, and T. Lahaye, *Enhanced atom-by-atom assembly of arbitrary tweezer arrays*, Phys. Rev. A **102**, 063107 (2020).

[56] F. Serwane, G. Zürn, T. Lompe, T. B. Ottenstein, A. N. Wenz, and S. Jochim, *Deterministic preparation of a tunable few-fermion system*, Science **332**, 336 (2011).

[57] G. Zürn, F. Serwane, T. Lompe, A. Wenz, M. G. Ries, J. E. Bohn, and S. Jochim, *Fermionization of two distinguishable fermions*, Phys. Rev. Lett. **108**, 075303 (2012).

[58] A. Wenz, G. Zürn, S. Murmann, I. Brouzos, T. Lompe, and S. Jochim, *From few to many: Observing the formation of a Fermi sea one atom at a time*, Science **342**, 457 (2013).

[59] G. Zürn, A. N. Wenz, S. Murmann, A. Bergschneider, T. Lompe, and S. Jochim, *Pairing in few-fermion systems with attractive interactions*, Phys. Rev. Lett. **111**, 175302 (2013).

[60] S. Murmann, F. Deuretzbacher, G. Zürn, J. Bjerlin, S. M. Reimann, L. Santos, T. Lompe, and S. Jochim, *Antiferromagnetic Heisenberg spin chain of a few cold atoms in a one-dimensional trap*, Phys. Rev. Lett. **115**, 215301 (2015).

[61] S. Murmann, A. Bergschneider, V. M. Klinkhamer, G. Zürn, T. Lompe, and S. Jochim, *Two fermions in a double Well: Exploring a fundamental building block of the Hubbard model*, Phys. Rev. Lett. **114**, 080402 (2015).

[62] A. M. Kaufman, B. J. Lester, C. M. Reynolds, M. L. Wall, M. Foss-Feig, K. R. A. Hazzard, A. M. Rey, and C. A. Regal, *Two-particle quantum interference in tunnel-coupled optical tweezers*, Science **345**, 306 (2014).

[63] A. Kaufman, B. Lester, M. Foss-Feig, M. Wall, A. Rey, and C. Regal, *Entangling two transportable neutral atoms via local spin exchange*, Nature **527**, 208 (2015).

[64] T. Busch, B.-G. Englert, K. Rzażewski, and M. Wilkens, *Two cold atoms in a harmonic trap*, Found. Phys. **28**, 549 (1998).

[65] P. Kościk and T. Sowiński, *Exactly solvable model of two trapped quantum particles interacting via finite-range soft-core interactions*, Sci. Rep. **8**, 48 (2018).

[66] Q. Guan, X. Y. Yin, S. E. Gharashi, and D. Blume, *Energy spectrum of a harmonically trapped two-atom system with spin–orbit coupling*, J. Phys. B: At. Mol. Opt. Phys **47**, 161001 (2014).

[67] C. D. Schillaci and T. C. Luu, *Energy spectra of two interacting fermions with spin-orbit coupling in a harmonic trap*, Phys. Rev. A **91**, 043606 (2015).

[68] X. Y. Yin, S. Gopalakrishnan, and D. Blume, *Harmonically trapped two-atom systems: Interplay of short-range s-wave interaction and spin-orbit coupling*, Phys. Rev. A **89**, 033606 (2014).

[69] R. Ołdziejewski, W. Górecki, and K. Rzażewski, *Two dipolar atoms in a harmonic trap*, EPL **114**, 46003 (2016).

[70] W. Górecki and K. Rzażewski, *Making two dysprosium atoms rotate–Einstein-de Haas effect revisited*, EPL **116**, 26004 (2016).

[71] W. Górecki and K. Rzażewski, *Electric dipoles vs. magnetic dipoles - For two molecules in a harmonic trap*, EPL **118**, 66002 (2017).

[72] T. Sowiński, T. Grass, O. Dutta, and M. Lewenstein, *Few interacting fermions in a one-dimensional harmonic trap*, Phys. Rev. A **88**, 033607 (2013).

[73] A. G. Volosniev, D. V. Fedorov, A. S. Jensen, M. Valiente, and N. T. Zinner, *Strongly interacting confined quantum systems in one dimension*, Nat. Commun. **5**, 5300 (2014).

[74] F. Deuretzbacher, D. Becker, J. Bjerlin, S. M. Reimann, and L. Santos, *Quantum magnetism without lattices in strongly interacting one-dimensional spinor gases*, Phys. Rev. A **90**, 013611 (2014).

[75] T. Grining, M. Tomza, M. Lesiuk, M. Przybytek, M. Musiał, R. Moszynski, M. Lewenstein, and P. Massignan, *Crossover between few and many fermions in a harmonic trap*, Phys. Rev. A **92**, 061601 (2015).

[76] T. Grining, M. Tomza, M. Lesiuk, M. Przybytek, M. Musiał, P. Massignan, M. Lewenstein, and R. Moszynski, *Many interacting fermions in a one-dimensional harmonic trap: a quantum-chemical treatment*, New J. Phys. **17**, 115001 (2015).

[77] J. Levinsen, P. Massignan, G. M. Bruun, and M. M. Parish, *Strong-coupling ansatz for the one-dimensional Fermi gas in a harmonic potential*, Sci. Adv. **1**, e1500197 (2015).

[78] D. Pecak and T. Sowiński, *Few strongly interacting ultracold fermions in one-dimensional traps of different shapes*, Phys. Rev. A **94**, 042118 (2016).

[79] D. Pecak, M. Gajda, and T. Sowiński, *Two-flavour mixture of a few fermions of different mass in a one-dimensional harmonic trap*, New J. Phys. **18**, 013030 (2016).

[80] F. Deuretzbacher, K. Fredenhagen, D. Becker, K. Bongs, K. Sengstock, and D. Pfannkuche, *Exact solution of strongly interacting quasi-one-dimensional spinor Bose gases*, Phys. Rev. Lett. **100**, 160405 (2008).

[81] M. A. Garcia-March, B. Julia-Diaz, G. E. Astrakharchik, T. Busch, J. Boronat, and A. Polls, *Quantum correlations and spatial localization in one-dimensional ultracold bosonic mixtures*, New J. Phys. **16**, 103004 (2014).

[82] P. Massignan, J. Levinsen, and M. M. Parish, *Magnetism in strongly interacting one-dimensional quantum mixtures*, Phys. Rev. Lett. **115**, 247202 (2015).

[83] A. Dehkharghani, A. Volosniev, J. Lindgren, J. Rotureau, C. Forssén, D. Fedorov, A. Jensen, and N. Zinner, *Quantum magnetism in strongly interacting one-dimensional spinor Bose systems*, Sci. Rep. **5**, 10675 (2015).

[84] L. Yang and X. Cui, *Effective spin-chain model for strongly interacting one-dimensional atomic gases with an arbitrary spin*, Phys. Rev. A **93**, 013617 (2016).

[85] A. Browaeys and T. Lahaye, *Many-body physics with individually controlled Rydberg atoms*, Nat. Phys. **16**, 132 (2020).

[86] L. D. Carr, D. DeMille, R. V. Krems, and J. Ye, *Cold and ultracold molecules: science, technology and applications*, New J. Phys. **11**, 055049 (2009).

[87] M. Lemeshko, R. V. Krems, J. M. Doyle, and S. Kais, *Manipulation of molecules with electromagnetic fields*, Mol. Phys. **111**, 1648 (2013).

[88] J. L. Bohn, A. M. Rey, and J. Ye, *Cold molecules: Progress in quantum engineering of chemistry and quantum matter*, Science **357**, 1002 (2017).

[89] M. G. Kozlov and D. DeMille, *Enhancement of the Electric Dipole Moment of the electron in PbO*, Phys. Rev. Lett. **89**, 133001 (2002).

[90] J. J. Hudson, B. E. Sauer, M. R. Tarbutt, and E. A. Hinds, *Measurement of the Electron Electric Dipole Moment Using YbF Molecules*, Phys. Rev. Lett. **89**, 023003 (2002).

[91] J. J. Hudson, D. M. Kara, I. J. Smallman, B. E. Sauer, M. R. Tarbutt, and E. A. Hinds, *Improved measurement of the shape of the electron*, Nature **473**, 493 (2011).

[92] J. Baron, W. C. Campbell, D. DeMille, J. M. Doyle, G. Gabrielse, Y. V. Gurevich, P. W. Hess, N. R. Hutzler, E. Kirilov, I. Kozyryev, B. R. O'Leary, C. D. Panda, M. F. Parsons, E. S. Petrik, B. Spaun, A. C. Vutha, and A. D. West, *Order of Magnitude Smaller Limit on the Electric Dipole Moment of the Electron*, Science **343**, 269 (2014).

[93] J. van Veldhoven, J. Küpper, H. L. Bethlem, B. Sartakov, A. J. A. van Roij, and G. Meijer, *Decelerated molecular beams for high-resolution spectroscopy*, Eur. Phys. J. D **31**, 337 (2004).

[94] N. R. Hutzler, H. I. Lu, and J. M. Doyle, *The Buffer Gas Beam: An Intense, Cold and Slow Source for Atoms and Molecules*, Chem. Rev. **112**, 4803 (2012).

[95] T. Zelevinsky, S. Kotochigova, and J. Ye, *Precision Test of Mass-Ratio Variations with Lattice-Confined Ultracold Molecules*, Phys. Rev. Lett. **100**, 043201 (2008).

[96] H. L. Bethlem and W. Ubachs, *Testing the time-invariance of fundamental constants using microwave spectroscopy on cold diatomic radicals*, Faraday Discuss. **142**, 25 (2009).

[97] V. V. Flambaum, *Variation of fundamental constants in space and time: theory and observations*, Eur. Phys. J. ST **163**, 159 (2008).

[98] S. Ospelkaus, K.-K. Ni, D. Wang, M. H. G. de Miranda, B. Neyenhuis, G. Quemener, P. S. Julienne, J. L. Bohn, D. S. Jin, and J. Ye, *Quantum-state controlled chemical reactions of ultracold potassium-rubidium molecules*, Science **327**, 853 (2010).

[99] K.-K. Ni, S. Ospelkaus, D. Wang, G. Quéméner, B. Neyenhuis, M. H. G. de Miranda, J. L. Bohn, J. Ye, and D. S. Jin, *Dipolar collisions of polar molecules in the quantum regime*, Nature **464**, 1324 (2010).

[100] M. H. G. de Miranda, A. Chotia, B. Neyenhuis, D. Wang, G. Quéméner, S. Ospelkaus, J. L. Bohn, J. Ye, and D. S. Jin, *Controlling the quantum stereodynamics of ultracold bimolecular reactions*, Nat. Phys. **7**, 502 (2011).

[101] M. Tomza, *Energetics and Control of Ultracold Isotope-Exchange Reactions between Heteronuclear Dimers in External Fields*, Phys. Rev. Lett. **115**, 063201 (2015).

[102] A. Klein, Y. Shagam, W. Skomorowski, P. S. Żuchowski, M. Pawlak, L. M. C. Janssen, N. Moiseyev, S. Y. T. van de Meerakker, A. van der Avoird, C. P. Koch, and E. Narevicius, *Directly probing anisotropy in atom–molecule collisions through quantum scattering resonances*, Nat. Phys. **13**, 35–38 (2016).

[103] P. Puri, M. Mills, C. Schneider, I. Simbotin, J. A. Montgomery, R. Côté, A. G. Suits, and E. R. Hudson, *Synthesis of mixed hypermetallic oxide BaOCa+ from laser-cooled reagents in an atom-ion hybrid trap*, Science **357**, 1370 (2017).

[104] T. de Jongh, M. Besemer, Q. Shuai, T. Karman, A. van der Avoird, G. C. Groenenboom, and S. Y. T. van de Meerakker, *Imaging the onset of the resonance regime in low-energy NO-He collisions*, Science **368**, 626 (2020).

[105] B. Friedrich and D. Herschbach, *Steric proficiency of polar $^2\Sigma$ molecules in congruent electric and magnetic fields*, Phys. Chem. Chem. Phys. **2**, 419 (2000).

[106] R. V. Krems, *Molecules near absolute zero and external field control of atomic and molecular dynamics*, Int. Rev. Phys. Chem. **24**, 99 (2005).

[107] T. V. Tscherbul and R. V. Krems, *Controlling electronic spin relaxation of cold molecules with electric fields*, Phys. Rev. Lett. **97**, 083201 (2006).

[108] T. V. Tscherbul and R. V. Krems, *Manipulating spin-dependent interactions in rotationally excited cold molecules with electric fields*, J. Chem. Phys. **125**, 194311 (2006).

[109] E. Abrahamsson, T. V. Tscherbul, and R. V. Krems, *Inelastic collisions of cold polar molecules in nonparallel electric and magnetic fields*, J. Chem. Phys. **127**, 044302 (2007).

[110] R. V. Krems, *Cold controlled chemistry*, Phys. Chem. Chem. Phys. **10**, 4079 (2008).

[111] T. V. Tscherbul, Y. V. Suleimanov, V. Aquilanti, and R. V. Krems, *Magnetic field modification of ultracold molecule-molecule collisions*, New J. Phys. **11**, 55021 (2009).

[112] S. V. Alyabyshev and R. V. Krems, *Controlling collisional spin relaxation of cold molecules with microwave laser fields*, Phys. Rev. A **80**, 033419 (2009).

[113] L. M. Janssen, A. Van Der Avoird, and G. C. Groenenboom, *Quantum Reactive Scattering of Ultracold NH(X Σ-3) Radicals in a Magnetic Trap*, Phys. Rev. Lett. **110**, 063201 (2013).

[114] K.-A. Brickman Soderberg, N. Gemelke, and C. Chin, *Ultracold molecules: Vehicles to scalable quantum information processing*, New J. Phys. **11**, 055022 (2009).

[115] J. Zhu, S. Kais, Q. Wei, D. Herschbach, and B. Friedrich, *Implementation of quantum logic gates using polar molecules in pendular states*, J. Chem. Phys. **138**, 024104 (2013).

[116] J. W. Park, Z. Z. Yan, H. Loh, S. A. Will, and M. W. Zwierlein, *Second-scale nuclear spin coherence time of ultracold ^{23}Na^{40}K molecules*, Science **357**, 372 (2017).

[117] K.-K. Ni, T. Rosenband, and D. D. Grimes, *Dipolar exchange quantum logic gate with polar molecules*, Chem. Sci. **9**, 6830 (2018).

[118] M. Hughes, M. D. Frye, R. Sawant, G. Bhole, J. A. Jones, S. L. Cornish, M. R. Tarbutt, J. M. Hutson, D. Jaksch, and J. Mur-Petit, *Robust entangling gate for polar molecules using magnetic and microwave fields*, Phys. Rev. A **101**, 062308 (2020).

[119] M. Sroczyńska, A. Dawid, M. Tomza, Z. Idziaszek, T. Calarco, and K. Jachymski, *Controlling the dynamics of ultracold polar molecules in optical tweezers*, New J. Phys. **24**, 1 (2022).

[120] A. Micheli, G. K. Brennen, and P. Zoller, *A toolbox for lattice-spin models with polar molecules*, Nat. Phys. **2**, 341 (2006).

[121] R. Barnett, D. Petrov, M. Lukin, and E. Demler, *Quantum Magnetism with Multicomponent Dipolar Molecules in an Optical Lattice*, Phys. Rev. Lett. **96**, 190401 (2006).

[122] A. V. Gorshkov, S. R. Manmana, G. Chen, E. Demler, M. D. Lukin, and A. M. Rey, *Quantum magnetism with polar alkali-metal dimers*, Phys. Rev. A **84**, 033619 (2011).

[123] Y. L. Zhou, M. Ortner, and P. Rabl, *Long-range and frustrated spin-spin interactions in crystals of cold polar molecules*, Phys. Rev. A **84**, 052332 (2011).

[124] B. Yan, S. A. Moses, B. Gadway, J. P. Covey, K. R. Hazzard, A. M. Rey, D. S. Jin, and J. Ye, *Observation of dipolar spin-exchange interactions with lattice-confined polar molecules*, Nature **501**, 521 (2013).

[125] K. R. A. Hazzard, S. R. Manmana, M. Foss-Feig, and A. M. Rey, *Far-from-Equilibrium Quantum Magnetism with Ultracold Polar Molecules*, Phys. Rev. Lett. **110**, 075301 (2013).

[126] S. R. Manmana, E. M. Stoudenmire, K. R. A. Hazzard, A. M. Rey, and A. V. Gorshkov, *Topological phases in ultracold polar-molecule quantum magnets*, Phys. Rev. B **87**, 081106 (2013).

[127] S. V. Syzranov, M. L. Wall, V. Gurarie, and A. M. Rey, *Spin-orbital dynamics in a system of polar molecules*, Nat. Commun. **5**, 1 (2014).

[128] M. Ortner, A. Micheli, G. Pupillo, and P. Zoller, *Quantum simulations of extended Hubbard models with dipolar crystals*, New J. Phys. **11**, 055045 (2009).

[129] F. Herrera, K. W. Madison, R. V. Krems, and M. Berciu, *Investigating Polaron Transitions with Polar Molecules*, Phys. Rev. Lett. **110**, 223002 (2013).

[130] J. Pérez-Ríos, F. Herrera, and R. V. Krems, *External field control of collective spin excitations in an optical lattice of 2Σ molecules*, New J. Phys. **12**, 103007 (2010).

[131] K. Osterloh, N. Barberán, and M. Lewenstein, *Strongly Correlated States of Ultracold Rotating Dipolar Fermi Gases*, Phys. Rev. Lett. **99**, 160403 (2007).

[132] H. P. Büchler, E. Demler, M. Lukin, A. Micheli, N. Prokof'ev, G. Pupillo, and P. Zoller, *Strongly Correlated 2D Quantum Phases with Cold Polar Molecules: Controlling the Shape of the Interaction Potential*, Phys. Rev. Lett. **98**, 060404 (2007).

[133] K. Góral, L. Santos, and M. Lewenstein, *Quantum phases of dipolar bosons in optical lattices*, Phys. Rev. Lett. **88**, 170406 (2002).

[134] A. O. G. Wallis, S. A. Gardiner, and J. M. Hutson, *Conical Intersections in Laboratory Coordinates with Ultracold Molecules*, Phys. Rev. Lett. **103**, 083201 (2009).

[135] N. R. Cooper and G. V. Shlyapnikov, *Stable Topological Superfluid Phase of Ultracold Polar Fermionic Molecules*, Phys. Rev. Lett. **103**, 155302 (2009).

[136] N. Y. Yao, C. R. Laumann, A. V. Gorshkov, S. D. Bennett, E. A. Demler, P. Zoller, and M. D. Lukin, *Topological flat bands from dipolar spin systems*, Phys. Rev. Lett. **109**, 266804 (2012).

[137] N. Y. Yao, A. V. Gorshkov, C. R. Laumann, A. M. Läuchli, J. Ye, and M. D. Lukin, *Realizing Fractional Chern Insulators in Dipolar Spin Systems*, Phys. Rev. Lett. **110**, 185302 (2013).

[138] D. Peter, N. Y. Yao, N. Lang, S. D. Huber, M. D. Lukin, and H. P. Büchler, *Topological bands with a Chern number $C = 2$ by dipolar exchange interactions*, Phys. Rev. A **91**, 053617 (2015).

[139] R. Schmidt and M. Lemeshko, *Rotation of Quantum Impurities in the Presence of a Many-Body Environment*, Phys. Rev. Lett. **114**, 203001 (2015).

[140] R. Schmidt and M. Lemeshko, *Deformation of a Quantum Many-Particle System by a Rotating Impurity*, Phys. Rev. X **6**, 011012 (2016).

[141] M. Lemeshko, *Quasiparticle Approach to Molecules Interacting with Quantum Solvents*, Phys. Rev. Lett. **118**, 095301 (2017).

[142] Y. Liu and K.-K. Ni, *Bimolecular Chemistry in the Ultracold Regime*, Annu. Rev. Phys. Chem. **73**, 1 (2021).

[143] E. Shuman, J. Barry, and D. DeMille, *Laser cooling of a diatomic molecule*, Nature **467**, 820 (2010).

[144] A. L. Collopy, S. Ding, Y. Wu, I. A. Finneran, L. Anderegg, B. L. Augenbraun, J. M. Doyle, and J. Ye, *3D magneto-optical trap of yttrium monoxide*, Phys. Rev. Lett. **121**, 213201 (2018).

[145] L. Anderegg, B. L. Augenbraun, Y. Bao, S. Burchesky, L. W. Cheuk, W. Ketterle, and J. M. Doyle, *Laser cooling of optically trapped molecules*, Nat. Phys. **14**, 890 (2018).

[146] I. Kozyryev, L. Baum, K. Matsuda, B. L. Augenbraun, L. Anderegg, A. P. Sedlack, and J. M. Doyle, *Sisyphus Laser Cooling of a Polyatomic Molecule*, Phys. Rev. Lett. **118**, 173201 (2017).

[147] D. Mitra, N. B. Vilas, C. Hallas, L. Anderegg, B. L. Augenbraun, L. Baum, C. Miller, S. Raval, and J. M. Doyle, *Direct laser cooling of a symmetric top molecule*, Science **369**, 1366 (2020).

[148] C. Chin, R. Grimm, P. Julienne, and E. Tiesinga, *Feshbach resonances in ultracold gases*, Rev. Mod. Phys. **82**, 1225 (2010).

[149] N. V. Vitanov, A. A. Rangelov, B. W. Shore, and K. Bergmann, *Stimulated Raman adiabatic passage in physics, chemistry, and beyond*, Rev. Mod. Phys. **89**, 015006 (2017).

[150] J. G. Danzl, E. Haller, M. Gustavsson, M. J. Mark, R. Hart, N. Bouloufa, O. Dulieu, H. Ritsch, and H.-C. Nägerl, *Quantum Gas of Deeply Bound Ground State Molecules*, Science **321**, 1062 (2008).

[151] K.-K. Ni, S. Ospelkaus, M. H. G. de Miranda, A. Pe'er, B. Neyenhuis, J. J. Zirbel, S. Kotochigova, P. S. Julienne, D. S. Jin, and J. Ye, *A high phase-space-density gas of polar molecules*, Science **322**, 231 (2008).

[152] T. Takekoshi, L. Reichsöllner, A. Schindewolf, J. M. Hutson, C. R. Le Sueur, O. Dulieu, F. Ferlaino, R. Grimm, and H.-C. Nägerl, *Ultracold Dense Samples of Dipolar RbCs Molecules in the Rovibrational and Hyperfine Ground State*, Phys. Rev. Lett. **113**, 205301 (2014).

[153] P. K. Molony, P. D. Gregory, Z. Ji, B. Lu, M. P. Köppinger, C. R. Le Sueur, C. L. Blackley, J. M. Hutson, and S. L. Cornish, *Creation of Ultracold* ^{87}Rb^{133}Cs *Molecules in the Rovibrational Ground State*, Phys. Rev. Lett. **113**, 255301 (2014).

[154] J. W. Park, S. A. Will, and M. W. Zwierlein, *Ultracold Dipolar Gas of Fermionic* ^{23}Na^{40}K *Molecules in Their Absolute Ground State*, Phys. Rev. Lett. **114**, 205302 (2015).

[155] M. Guo, B. Zhu, B. Lu, X. Ye, F. Wang, R. Vexiau, N. Bouloufa-Maafa, G. Quéméner, O. Dulieu, and D. Wang, *Creation of an ultracold gas of ground-state dipolar* $^{23}Na^{87}Rb$ *molecules*, Phys. Rev. Lett. **116**, 205303 (2016).

[156] L. De Marco, G. Valtolina, K. Matsuda, W. G. Tobias, J. P. Covey, and J. Ye, *A degenerate Fermi gas of polar molecules*, Science **363**, 853 (2019).

[157] G. Quéméner and P. S. Julienne, *Ultracold Molecules under Control!*, Chem. Rev. **112**, 4949 (2012).

[158] Z. Li, S. V. Alyabyshev, and R. V. Krems, *Ultracold Inelastic Collisions in Two Dimensions*, Phys. Rev. Lett. **100**, 073202 (2008).

[159] A. Micheli, Z. Idziaszek, G. Pupillo, M. A. Baranov, P. Zoller, and P. S. Julienne, *Universal Rates for Reactive Ultracold Polar Molecules in Reduced Dimensions*, Phys. Rev. Lett. **105**, 073202 (2010).

[160] G. Quéméner and J. L. Bohn, *Dynamics of ultracold molecules in confined geometry and electric field*, Phys. Rev. A **83**, 012705 (2011).

[161] G. Quéméner, M. Lepers, and O. Dulieu, *Dynamics of ultracold dipolar particles in a confined geometry and tilted fields*, Phys. Rev. A **92**, 042706 (2015).

[162] Z. Idziaszek, K. Jachymski, and P. S. Julienne, *Reactive collisions in confined geometries*, New J. Phys. **17**, 035007 (2015).

[163] A. Chotia, B. Neyenhuis, S. A. Moses, B. Yan, J. P. Covey, M. Foss-Feig, A. M. Rey, D. S. Jin, and J. Ye, *Long-Lived Dipolar Molecules and Feshbach Molecules in a 3D Optical Lattice*, Phys. Rev. Lett. **108**, 080405 (2012).

[164] B. Yan, S. A. Moses, B. Gadway, J. P. Covey, K. R. Hazzard, A. M. Rey, D. S. Jin, and J. Ye, *Observation of dipolar spin-exchange interactions with lattice-confined polar molecules*, Nature **501**, 521 (2013).

[165] L. R. Liu, J. D. Hood, Y. Yu, J. T. Zhang, N. R. Hutzler, T. Rosenband, and K.-K. Ni, *Building one molecule from a reservoir of two atoms*, Science **360**, 900 (2018).

[166] L. Anderegg, L. W. Cheuk, Y. Bao, S. Burchesky, W. Ketterle, K.-K. Ni, and J. M. Doyle, *An optical tweezer array of ultracold molecules*, Science **365**, 1156 (2019).

[167] L. W. Cheuk, L. Anderegg, Y. Bao, S. Burchesky, S. S. Yu, W. Ketterle, K.-K. Ni, and J. M. Doyle, *Observation of Collisions between Two Ultracold Ground-State CaF Molecules*, Phys. Rev. Lett. **125**, 043401 (2020).

[168] L. R. Liu, J. T. Zhang, Y. Yu, N. R. Hutzler, Y. Liu, T. Rosenband, and K.-K. Ni, *Ultracold Molecular Assembly*, arXiv:1701.03121 (2017).

[169] X. He, K. Wang, J. Zhuang, P. Xu, X. Gao, R. Guo, C. Sheng, M. Liu, J. Wang, J. Li, G. V. Shlyapnikov, and M. Zhan, *Coherently forming a single molecule in an optical trap*, Science **370**, 331 (2020).

[170] J. T. Zhang, Y. Yu, W. B. Cairncross, K. Wang, L. R. B. Picard, J. D. Hood, Y.-W. Lin, J. M. Hutson, and K.-K. Ni, *Forming a Single Molecule by Magnetoassociation in an Optical Tweezer*, Phys. Rev. Lett. **124**, 253401 (2020).

[171] W. B. Cairncross, J. T. Zhang, L. R. B. Picard, Y. Yu, K. Wang, and K.-K. Ni, *Assembly of a Rovibrational Ground State Molecule in an Optical Tweezer*, Phys. Rev. Lett. **126**, 123402 (2021).

[172] J. T. Zhang, L. R. B. Picard, W. B. Cairncross, K. Wang, Y. Yu, F. Fang, and K.-K. Ni, *An optical tweezer array of ground-state polar molecules*, arXiv:2112.00991 (2021).

[173] P. D. Gregory, M. D. Frye, J. A. Blackmore, E. M. Bridge, R. Sawant, J. M. Hutson, and S. L. Cornish, *Sticky collisions of ultracold RbCs molecules*, Nat. Commun. **10**, 1 (2019).

[174] X. Ye, M. Guo, M. L. González-Martínez, G. Quéméner, and D. Wang, *Collisions of ultracold 23Na87Rb molecules with controlled chemical reactivities*, Sci. Adv. **4**, eaaq0083 (2018).

[175] M. Mayle, B. P. Ruzic, and J. L. Bohn, *Statistical aspects of ultracold resonant scattering*, Phys. Rev. A **85**, 062712 (2012).

[176] M. Mayle, G. Quéméner, B. P. Ruzic, and J. L. Bohn, *Scattering of ultracold molecules in the highly resonant regime*, Phys. Rev. A **87**, 012709 (2013).

[177] A. Christianen, T. Karman, and G. C. Groenenboom, *Quasiclassical method for calculating the density of states of ultracold collision complexes*, Phys. Rev. A **100**, 32708 (2019).

[178] A. Christianen, M. W. Zwierlein, G. C. Groenenboom, and T. Karman, *Photoinduced two-body loss of ultracold molecules*, Phys. Rev. Lett. **123**, 123402 (2019).

[179] P. D. Gregory, J. A. Blackmore, S. L. Bromley, and S. L. Cornish, *Loss of ultracold $Rb^{87}Cs^{133}$ molecules via optical excitation of long-lived two-body collision complexes*, Phys. Rev. Lett. **124**, 163402 (2020).

[180] P. Gersema, K. K. Voges, M. Meyer zum Alten Borgloh, L. Koch, T. Hartmann, A. Zenesini, S. Ospelkaus, J. Lin, J. He, and D. Wang, *Probing photoinduced two-body loss of ultracold nonreactive bosonic $^{23}Na^{87}Rb$ and $^{23}Na^{39}K$ molecules*, Phys. Rev. Lett. **127**, 163401 (2021).

[181] R. Bause, A. Schindewolf, R. Tao, M. Duda, X.-Y. Chen, G. Quéméner, T. Karman, A. Christianen, I. Bloch, and X.-Y. Luo, *Collisions of ultracold molecules in bright and dark optical dipole traps*, Phys. Rev. Res. **3**, 033013 (2021).

[182] M. A. Nichols, Y.-X. Liu, L. Zhu, M.-G. Hu, Y. Liu, and K.-K. Ni, *Detection of Long-Lived Complexes in Ultracold Atom-Molecule Collisions*, arXiv:2105.14960 (2021).

[183] K. Jachymski, M. Gronowski, and M. Tomza, *Collisional losses of ultracold molecules due to intermediate complex formation*, arXiv:2110.07501 (2021).

[184] J. F. E. Croft, J. L. Bohn, and G. Quéméner, *Anomalous Lifetimes of Ultracold Complexes Decaying into a Single Channel: What's Taking So Long in There?*, arXiv:2111.09956 (2021).

[185] G. Quéméner and J. L. Bohn, *Shielding $^2\Sigma$ ultracold dipolar molecular collisions with electric fields*, Phys. Rev. A **93**, 012704 (2016).

[186] K. Matsuda, L. D. Marco, J.-R. Li, W. G. Tobias, G. Valtolina, G. Quéméner, and J. Ye, *Resonant collisional shielding of reactive molecules using electric fields*, Science **370**, 1324 (2020).

[187] J.-R. Li, W. G. Tobias, K. Matsuda, C. Miller, G. Valtolina, L. D. Marco, R. R. W. Wang, L. Lassablière, G. Quéméner, J. L. Bohn, and J. Ye, *Tuning of dipolar interactions and evaporative cooling in a three-dimensional molecular quantum gas*, Nat. Phys. **17**, 1144 (2021).

[188] L. Anderegg, S. Burchesky, S. S. Yu, T. Karman, E. Chae, K.-K. Ni, W. Ketterle, and J. M. Doyle, *Observation of microwave shielding of ultracold molecules*, Science **373**, 779 (2021).

[189] T. M. Rvachov, H. Son, A. T. Sommer, S. Ebadi, J. J. Park, M. W. Zwierlein, W. Ketterle, and A. O. Jamison, *Long-Lived Ultracold Molecules with Electric and Magnetic Dipole Moments*, Phys. Rev. Lett. **119**, 143001 (2017).

[190] N. A. Spaldin and R. Ramesh, *Advances in magnetoelectric multiferroics*, Nature Materials **18**, 203 (2019).

[191] U. Schollwöck, *The density-matrix renormalization group*, Rev. Mod. Phys. **77**, 259 (2005).

[192] R. Orùs, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Ann. Phys. **349**, 117 (2014).

[193] S.-J. Ran, E. Tirrito, C. Peng, X. Chen, G. Su, and M. Lewenstein, *Review of tensor network contraction approaches*, arXiv:1708.09213 (2018).

[194] T. Dauxois, *Fermi, Pasta, Ulam, and a mysterious lady*, Phys. Today **61**, 55 (2008).

[195] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, *A high-bias, low-variance introduction to machine learning for physicists*, Phys. Rep. **810**, 1 (2019).

[196] T. Neupert, M. H. Fischer, E. Greplova, K. Choo, and M. Denner, *Introduction to machine learning for the sciences*, arXiv:2102.04883 (2021).

[197] C. M. Handley and P. L. A. Popelier, *Potential energy surfaces fitted by artificial neural networks*, J. Phys. Chem. A **114**, 3371 (2010).

[198] J. Behler, *Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations*, Phys. Chem. Chem. Phys. **13**, 17930 (2011).

[199] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, and T. Carrington Jr., *Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy*, J. Chem. Phys. **148**, 241702 (2018).

[200] F. Noé, S. Olsson, J. Köhler, and H. Wu, *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning*, Science **365**, eaaw1147 (2019).

[201] C. Shen, M. Krenn, S. Eppel, and A. Aspuru-Guzik, *Deep molecular dreaming: Inverse machine learning for de-novo molecular design and interpretability with surjective representations*, Mach. Learn.: Sci. Technol. **2**, 03LT02 (2021).

[202] L. Ward and C. Wolverton, *Atomistic calculations and materials informatics: A review*, Curr. Opin. Solid State Mater. Sci. **21**, 167 (2016).

[203] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao, and A. Aspuru-Guzik, *Data-Driven Strategies for Accelerated Materials Design*, Acc. Chem. Res. **54**, 849 (2021).

[204] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, *In silico labeling: Predicting fluorescent labels in unlabeled images*, Cell **173**, 792 (2018).

[205] D. Wong and S. Yip, *Machine learning classifies cancer*, Nature **555**, 446 (2018).

[206] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, *Highly accurate protein structure prediction with AlphaFold*, Nature **596**, 583 (2021).

[207] M. Krenn and A. Zeilinger, *Predicting research trends with semantic and neural networks with an application in quantum physics*, Proc. Natl. Acad. Sci. U.S.A. **117**, 1910 (2020).

[208] 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).

[209] J. Carrasquilla, *Machine learning for quantum matter*, Adv. Phys.: X **5**, 1797528 (2020).

[210] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science **355**, 602 (2017).

[211] D.-L. Deng, X. Li, and S. Das Sarma, *Quantum Entanglement in Neural Network States*, Phys. Rev. X **7**, 021021 (2017).

[212] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, *Quantum Entanglement in Deep Learning Architectures*, Phys. Rev. Lett. **122**, 065301 (2019).

[213] J. Hermann, Z. Schätzle, and F. Noé, *Deep-neural-network solution of the electronic Schrödinger equation*, Nat. Chem. **12**, 891 (2020).

[214] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Neural-network quantum state tomography*, Nat. Phys. **14**, 447–450 (2018).

[215] J. F. Carrasquilla, G. Torlai, and R. G. Melko, *Latent Space Purification via Neural Density Operators*, Nat. Mach. Intell. **1**, 155 (2019).

[216] G. Torlai, B. Timar, E. P. Van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, *Integrating Neural Networks with a Quantum Simulator for State Reconstruction*, Phys. Rev. Lett. **123**, 19 (2019).

[217] M. Neugebauer, L. Fischer, A. Jäger, S. Czischek, S. Jochim, M. Weidemüller, and M. Gärttner, *Neural network quantum state tomography in a two-qubit experiment*, Phys. Rev. A **102**, 042604 (2020).

[218] P. B. Wigley, P. J. Everitt, A. Van Den Hengel, J. W. Bastian, M. A. Sooriyabandara, G. D. Mcdonald, K. S. Hardman, C. D. Quinlivan, P. Manju, C. C. Kuhn, I. R. Petersen, A. N. Luiten, J. J. Hope, N. P. Robins, and M. R. Hush, *Fast machine-learning online optimization of ultra-cold-atom experiments*, Sci. Rep. **6**, 25890 (2016).

[219] X.-C. Yang, M.-H. Yung, and X. Wang, *Neural-network-designed pulse sequences for robust control of singlet-triplet qubits*, Phys. Rev. A **97**, 042324 (2018).

[220] A. D. Tranter, H. J. Slatyer, M. R. Hush, A. C. Leung, J. L. Everett, K. V. Paul, P. Vernaz-Gris, P. K. Lam, B. C. Buchler, and G. T. Campbell, *Multiparameter optimisation of a magneto-optical trap using deep learning*, Nat. Comm. **9**, 4360 (2018).

[221] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, *Reinforcement Learning in Different Phases of Quantum Control*, Phys. Rev. X **8**, 031086 (2018).

[222] E. T. Davletov, V. V. Tsyganok, V. A. Khlebnikov, D. A. Pershin, D. V. Shaykin, and A. V. Akimov, *Machine learning for achieving Bose-Einstein condensation of thulium atoms*, Phys. Rev. A **102**, 011302(R) (2020).

[223] E. Greplova, C. Gold, B. Kratochwil, T. Davatz, R. Pisoni, A. Kurzmann, P. Rickhaus, M. H. Fischer, T. Ihn, and S. D. Huber, *Fully automated identification of two-dimensional material samples*, Phys. Rev. Appl. **13**, 064017 (2020).

[224] R. Durrer, B. Kratochwil, J. V. Koski, A. J. Landig, C. Reichl, W. Wegscheider, T. Ihn, and E. Greplova, *Automated tuning of double quantum dots into specific charge states using neural networks*, Phys. Rev. Appl. **13**, 054019 (2020).

[225] A. Valenti, G. Jin, J. Léonard, S. D. Huber, and E. Greplova, *Scalable Hamiltonian learning for large-scale out-of-equilibrium quantum dynamics*, Phys. Rev. A **105**, 023302 (2022).

[226] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, *Classifying snapshots of the doped Hubbard model with machine learning*, Nat. Phys. **15**, 921 (2019).

[227] C. Casert, K. Mills, T. Vieijra, J. Ryckebusch, and I. Tamblyn, *Optical lattice experiments at unobserved conditions with generative adversarial deep learning*, Phys. Rev. Research **3**, 033267 (2021).

[228] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, *Active learning machine learns to create new quantum experiments*, Proc. Natl. Acad. Sci. U.S.A. **115**, 1221 (2018).

[229] M. Krenn, M. Erhard, and A. Zeilinger, *Computer-inspired quantum experiments*, Nat. Rev. Phys. **2**, 649 (2020).

[230] M. Swaddle, L. Noakes, H. Smallbone, L. Salter, and J. Wang, *Generating three-qubit quantum circuits with neural networks*, Phys. Lett. A **381**, 3391 (2017).

[231] S. Krastanov and L. Jiang, *Deep Neural Network Probabilistic Decoder for Stabilizer Codes*, Sci. Rep. **7**, 11003 (2017).

[232] G. Torlai and R. G. Melko, *Latent Space Purification via Neural Density Operators*, Phys. Rev. Lett. **120**, 240503 (2018).

[233] B. Naul, J. S. Bloom, F. Pérez, and S. van der Walt, *A recurrent neural network for classification of unevenly sampled variable stars*, Nat. Astron. **2**, 151 (2018).

[234] P. Baldi, P. Sadowski, and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, Nat. Commun. **5**, 4308 (2014).

[235] M. Farina, Y. Nakai, and D. Shih, *Searching for new physics with deep autoencoders*, Phys. Rev. D **101**, 075021 (2020).

[236] J. Liu, H. Shen, Y. Qi, Z. Y. Meng, and L. Fu, *Self-learning Monte Carlo method and cumulative update in fermion systems*, Phys. Rev. B **95**, 241104 (2017).

[237] L. Huang and L. Wang, *Accelerated Monte Carlo simulations with restricted Boltzmann machines*, Phys. Rev. B **95**, 35105 (2017).

[238] L. Wang, *Exploring cluster Monte Carlo updates with Boltzmann machines*, Phys. Rev. E **96**, 51301 (2017).

[239] A. Valenti, E. Greplova, N. H. Lindner, and S. D. Huber, *Correlation-enhanced neural networks as interpretable variational quantum states*, Phys. Rev. Research **4**, L012010 (2022).

[240] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, *Physics-informed machine learning*, Nat. Rev. Phys. **3**, 422 (2021).

[241] V. Dunjko and P. Wittek, *A non-review of Quantum Machine Learning: trends and explorations*, Quantum Views **4**, 32 (2020).

[242] L. Zdeborová, *Understanding deep learning is also a job for physicists*, Nat. Phys. **16**, 602 (2020).

[243] J. Carrasquilla and R. G. Melko, *Machine learning phases of matter*, Nat. Phys. **13**, 431 (2017).

[244] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Learning phase transitions by confusion*, Nat. Phys. **13**, 435–439 (2017).

[245] A. Tanaka and A. Tomiya, *Detection of phase transition via convolutional neural networks*, J. Phys. Soc. Jpn. **86**, 63001 (2017).

[246] C.-D. Li, D.-R. Tan, and F.-J. Jiang, *Applications of neural networks to the studies of phase transitions of two-dimensional Potts models*, Ann. Phys. **391**, 312 (2018).

[247] F. Schäfer and N. Lörch, *Vector field divergence of predictive model output as indication of phase transitions*, Phys. Rev. E **99**, 062107 (2019).

[248] A. Cole, G. J. Loges, and G. Shiu, *Quantitative and interpretable order parameters for phase transitions from persistent homology*, Phys. Rev. B **104**, 104426 (2021).

[249] Y. Liu, S. Arunachalam, and K. Temme, *A rigorous and robust quantum speed-up in supervised machine learning*, Nat. Phys. **17**, 1013 (2021).

[250] S. J. Wetzel, *Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders*, Phys. Rev. E **96**, 022140 (2017).

[251] Y.-H. Liu and E. P. L. van Nieuwenburg, *Discriminative cooperative networks for detecting phase transitions*, Phys. Rev. Lett. **120**, 176401 (2018).

[252] K. Ch'ng, N. Vazquez, and E. Khatami, *Unsupervised machine learning account of magnetic transitions in the Hubbard model*, Phys. Rev. E **97**, 13306 (2018).

[253] P. Huembeli, A. Dauphin, P. Wittek, and C. Gogolin, *Automated discovery of characteristic features of phase transitions in many-body localization*, Phys. Rev. B **99**, 104106 (2019).

[254] K. Kottmann, P. Huembeli, M. Lewenstein, and A. Acín, *Unsupervised Phase Discovery with Deep Anomaly Detection*, Phys. Rev. Lett. **125**, 170603 (2020).

[255] J. Arnold, F. Schäfer, M. Žonda, and A. U. J. Lode, *Interpretable and unsupervised phase classification*, Phys. Rev. Res. **3**, 033052 (2021).

[256] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Machine learning quantum phases of matter beyond the fermion sign problem*, Sci. Rep. **7**, 8823 (2017).

[257] H. Théveniaut and F. Alet, *Neural network setups for a precise detection of the many-body localization transition: Finite-size scaling and limitations*, Phys. Rev. B **100**, 224202 (2019).

[258] X.-y. Y. Dong, F. Pollmann, and X.-F. F. Zhang, *Machine learning of quantum phase transitions*, Phys. Rev. B **99**, 121104 (2019).

[259] S. Blücher, L. Kades, J. M. Pawlowski, N. Strodthoff, and J. M. Urban, *Towards novel insights in lattice field theory with explainable machine learning*, Phys. Rev. D **101**, 094507 (2020).

[260] Y. Zhang and E.-A. Kim, *Quantum loop topography for machine learning*, Phys. Rev. Lett. **118**, 216401 (2017).

[261] D.-L. Deng, X. Li, and S. Das Sarma, *Machine learning topological states*, Phys. Rev. B **96**, 195145 (2017).

[262] P. Zhang, H. Shen, and H. Zhai, *Machine Learning Topological Invariants with Neural Networks*, Phys. Rev. Lett. **120**, 066401 (2018).

[263] D. Carvalho, N. A. García-Martínez, J. L. Lado, and J. Fernández-Rossier, *Real-space mapping of topological invariants using artificial neural networks*, Phys. Rev. B **97**, 115453 (2018).

[264] M. J. S. Beach, A. Golubeva, and R. G. Melko, *Machine learning vortices at the Kosterlitz-Thouless transition*, Phys. Rev. B **97**, 045207 (2018).

[265] J. F. Rodriguez-Nieva and M. S. Scheurer, *Identifying topological order through unsupervised machine learning*, Nat. Phys. **15**, 790 (2019).

[266] E. Greplova, A. Valenti, G. Boschung, F. Schäfer, N. Lörch, and S. D. Huber, *Unsupervised identification of topological phase transitions using predictive models*, New J. Phys. **22**, 045003 (2020).

[267] N. L. Holanda and M. A. R. Griffith, *Machine learning topological phases in real space*, Phys. Rev. B **102**, 054107 (2020).

[268] Y. Long, J. Ren, and H. Chen, *Unsupervised Manifold Clustering of Topological Phononics*, Phys. Rev. Lett. **124**, 185501 (2020).

[269] M. S. Scheurer and R.-J. Slager, *Unsupervised Machine Learning and Band Topology*, Phys. Rev. Lett. **124**, 226401 (2020).

[270] Y.-H. Tsai, M.-Z. Yu, Y.-H. Hsu, and M.-C. Chung, *Deep learning of topological phase transitions from entanglement aspects*, Phys. Rev. B **102**, 054512 (2020).

[271] P. Huembeli, A. Dauphin, and P. Wittek, *Identifying quantum phase transitions with adversarial neural networks*, Phys. Rev. B **97**, 134109 (2018).

[272] O. Balabanov and M. Granath, *Unsupervised interpretable learning of topological indices invariant under permutations of atomic bands*, Mach. Learn.: Sci. Technol. **2**, 025008 (2021).

[273] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. Davis, E. Khatami, and E. A. Kim, *Machine learning in electronic-quantum-matter imaging experiments*, Nature **570**, 484 (2019).

[274] M. Ziatdinov, A. Maksov, L. Li, A. S. Sefat, P. Maksymovych, and S. V. Kalinin, *Deep data mining in a real space: Separation of intertwined electronic responses in a lightly doped BaFe2As2*, Nanotechnology **27**, 475706 (2016).

[275] A. M. Samarakoon, K. Barros, Y. W. Li, M. Eisenbach, Q. Zhang, F. Ye, V. Sharma, Z. L. Dun, H. Zhou, S. A. Grigera, C. D. Batista, and D. A. Tennant, *Machine-learning-assisted insight into spin ice Dy2Ti2O7*, Nat. Comm. **11**, 1 (2020).

[276] B. S. Rem, N. Käming, M. Tarnowski, L. Asteria, N. Fläschner, C. Becker, K. Sengstock, and C. Weitenberg, *Identifying quantum phase transitions using artificial neural networks on experimental data*, Nat. Phys. **15**, 917 (2019).

[277] E. Khatami, E. Guardado-Sanchez, B. M. Spar, J. F. Carrasquilla, W. S. Bakr, and R. T. Scalettar, *Visualizing strange metallic correlations in the two-dimensional Fermi-Hubbard model with artificial intelligence*, Phys. Rev. A **102**, 33326 (2020).

[278] N. Käming, A. Dawid, K. Kottmann, M. Lewenstein, K. Sengstock, A. Dauphin, and C. Weitenberg, *Unsupervised machine learning of topological phase transitions from experimental data*, Mach. Learn.: Sci. Technol. **2**, 035037 (2021).

[279] L. Wang, *Discovering phase transitions with unsupervised learning*, Phys. Rev. B **94**, 195105 (2016).

[280] R. A. Vargas-Hernández, J. Sous, M. Berciu, and R. V. Krems, *Extrapolating quantum observables with machine learning: Inferring multiple phase transitions from properties of a single phase*, Phys. Rev. Lett. **121**, 255702 (2018).

[281] M. Richter-Laskowska, H. Khan, N. Trivedi, and M. M. Maśka, *A machine learning approach to the Berezinskii-Kosterlitz-Thouless transition in classical and quantum models*, Condens. Matter Phys. **21**, 33602 (2018).

[282] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, and F. Giannotti, *A Survey Of Methods For Explaining Black Box Models*, arXiv:1802.01933 (2018).

[283] L. Breiman, *Reflections After Refereeing Papers for NIPS*, in *The Mathematics of Generalization* (CRC Press, 2018) pp. 11–15.

[284] Z. C. Lipton, *The Mythos of Model Interpretability*, Commun. ACM **61**, 35 (2018).

[285] EU General Data Protection Regulation (GDPR), *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, OJ **L 119/1** (2016).

[286] T. Songül, *Fair and Unbiased Algorithmic Decision Making*, JRC Tech. Rep. **10** (2018), 10.48550/arXiv.1901.04730.

[287] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Understanding deep learning requires rethinking generalization*, in *ICLR 2017 - Int. Conf. Learn. Represent.* (2017).

[288] W. Brendel and M. Bethge, *Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet*, in *ICLR 2019 - Int. Conf. Learn. Represent.* (2019).

[289] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, *Machine learning interpretability: a survey on methods and metrics*, Electronics **8**, 832 (2019).

[290] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, *Definitions, methods, and applications in interpretable machine learning*, PNAS **116**, 22071 (2019).

[291] C. Molnar, *Interpretable Machine Learning* (2019) https://christophm.github.io/interpretable-ml-book/.

[292] M. Du, N. Liu, and X. Hu, *Techniques for interpretable machine learning*, Communications of the ACM **63**, 68 (2020).

[293] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, *A review of uncertainty quantification in deep learning: Techniques, applications and challenges*, Inf. Fusion **76**, 243 (2021).

[294] B. Biggio and F. Roli, *Wild patterns: Ten years after the rise of adversarial machine learning*, Pattern Recognit. **84**, 317 (2018).

[295] R. V. Krems, *Bayesian machine learning for quantum molecular dynamics*, Phys. Chem. Chem. Phys. **21**, 13392 (2019).

[296] S. J. Wetzel and M. Scherzer, *Machine learning of explicit order parameters: From the Ising model to SU(2) lattice gauge theory*, Phys. Rev. B **96**, 184410 (2017).

[297] P. Y. Lu, S. Kim, and M. Soljačić, *Extracting Interpretable Physical Parameters from Spatiotemporal Systems Using Unsupervised Learning*, Phys. Rev. X **10**, 1 (2020).

[298] R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, *Discovering physical concepts with neural networks*, Phys. Rev. Lett. **124**, 010508 (2020).

[299] S. J. Wetzel, R. G. Melko, J. Scott, M. Panju, and V. Ganesh, *Discovering Symmetry Invariants and Conserved Quantities by Interpreting Siamese Neural Networks*, Phys. Rev. Res. **2**, 033499 (2020).

[300] C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, *Correlator Convolutional Neural Networks: An Interpretable Architecture for Image-like Quantum Matter Data*, Nat. Commun. **12**, 1 (2021).

[301] C. Miles, R. Samajdar, S. Ebadi, T. T. Wang, H. Pichler, S. Sachdev, M. D. Lukin, M. Greiner, K. Q. Weinberger, and E.-A. Kim, *Machine learning discovery of new phases in programmable quantum simulator snapshots*, arXiv:2112.10789 (2021).

[302] P. Ponte and R. G. Melko, *Kernel methods for interpretable machine learning of order parameters*, Phys. Rev. B **96**, 205146 (2017).

[303] W. Zhang, L. Wang, and Z. Wang, *Interpretable machine learning study of the many-body localization transition in disordered quantum Ising spin chains*, Phys. Rev. B **99**, 054208 (2019).

[304] J. Greitemann, K. Liu, L. D. C. Jaubert, H. Yan, N. Shannon, and L. Pollet, *Identification of emergent constraints and hidden order in frustrated magnets using tensorial kernel methods of machine learning*, Phys. Rev. B **100**, 174408 (2019).

[305] J. Greitemann, K. Liu, and L. Pollet, *The view of TK-SVM on the phase hierarchy in the classical kagome Heisenberg antiferromagnet*, J. Phys.: Condens. Matter **33**, 054002 (2021).

[306] P. Friederich, M. Krenn, I. Tamblyn, and A. Aspuru-Guzik, *Scientific intuition inspired by machine learning-generated hypotheses*, Mach. Learn.: Sci. Technol. **2**, 025027 (2021).

[307] Z. Idziaszek and T. Calarco, *Analytical solutions for the dynamics of two trapped interacting ultracold atoms*, Phys. Rev. A **74**, 022712 (2006).

[308] K. Jachymski, M. Krych, P. S. Julienne, and Z. Idziaszek, *Quantum Theory of Reactive Collisions for* $1/r^n$ *Potentials*, Phys. Rev. Lett. **110**, 213202 (2013).

[309] J. Jankowska and Z. Idziaszek, *Reactive collisions of two ultracold particles in a harmonic trap*, Phys. Rev. A **93**, 052714 (2016).

[310] E. L. Bolda, E. Tiesinga, and P. S. Julienne, *Effective-scattering-length model of ultracold atomic collisions and Feshbach resonances in tight harmonic traps*, Phys. Rev. A **66**, 013403 (2002).

[311] J. Deiglmayr, M. Aymar, R. Wester, M. Weidemüller, and O. Dulieu, *Calculations of static dipole polarizabilities of alkali dimers: Prospects for alignment of ultracold molecules*, J. Chem. Phys. **129**, 064309 (2008).

[312] J. M. Hutson, *Intermolecular forces from the spectroscopy of van der Waals molecules*, Annu. Rev. Phys. Chem. **41**, 123 (1990).

[313] M. Olshanii, *Atomic Scattering in the Presence of an External Confinement and a Gas of Impenetrable Bosons*, Phys. Rev. Lett. **81**, 938 (1998).

[314] D. R. Herrick and F. H. Stillinger, *Variable dimensionality in atoms and its effect on the ground state of the helium isoelectronic sequence*, Phys. Rev. A **11**, 42 (1975).

[315] S. Sinha and L. Santos, *Cold dipolar gases in quasi-one-dimensional geometries*, Phys. Rev. Lett. **99**, 140406 (2007).

[316] F. Deuretzbacher, J. C. Cremon, and S. M. Reimann, *Ground-state properties of few dipolar bosons in a quasi-one-dimensional harmonic trap*, Phys. Rev. A **81**, 063616 (2010).

[317] M. A. Garcìa-March, T. Fogarty, S. Campbell, T. Busch, and M. Paternostro, *Non-equilibrium thermodynamics of harmonically trapped bosons*, New J. Phys. **18**, 103035 (2016).

[318] A. Usui, B. Buča, and J. Mur-Petit, *Quantum probe spectroscopy for cold atomic systems*, New J. Phys. **20**, 103006 (2018).

[319] Z. Li, L. Zou, and T. H. Hsieh, *Hamiltonian Tomography via Quantum Quench*, Phys. Rev. Lett. **124**, 160502 (2020).

[320] L. M. Kehrberger, V. J. Bolsinger, and P. Schmelcher, *Quantum dynamics of two trapped bosons following infinite interaction quenches*, Phys. Rev. A **97**, 013606 (2018).

[321] L. Budewig, S. I. Mistakidis, and P. Schmelcher, *Quench dynamics of two one-dimensional harmonically trapped bosons bridging attraction and repulsion*, Mol. Phys. **117**, 2043 (2019).

[322] G. Bougas, S. I. Mistakidis, and P. Schmelcher, *Analytical treatment of the interaction quench dynamics of two bosons in a two-dimensional harmonic trap*, Phys. Rev. A **100**, 053602 (2019).

[323] Q. Guan, V. Klinkhamer, R. Klemt, J. H. Becher, A. Bergschneider, P. M. Preiss, S. Jochim, and D. Blume, *Density oscillations induced by individual ultracold two-body collisions*, Phys. Rev. Lett. **122**, 083401 (2019).

[324] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, *Colloquium: Nonequilibrium dynamics of closed interacting quantum systems*, Rev. Mod. Phys. **83**, 863 (2011).

[325] M. Gring, M. Kuhnert, T. Langen, T. Kitagawa, B. Rauer, M. Schreitl, I. Mazets, D. A. Smith, E. Demler, and J. Schmiedmayer, *Relaxation and Prethermalization in an Isolated Quantum System*, Science **337**, 1318 (2012).

[326] S. Erne, R. Bücker, T. Gasenzer, J. Berges, and J. Schmiedmayer, *Universal dynamics in an isolated one-dimensional Bose gas far from equilibrium*, Nature **563**, 225 (2018).

[327] A. Mitra, *Quantum Quench Dynamics*, Annu. Rev. Condens. Matter Phys. **9**, 245 (2018).

[328] P. Virtanen, R. Gommers, and T. t. Oliphant, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nat. Methods **17**, 261 (2020).

[329] A. Dawid, *Two interacting molecules in a one-dimensional harmonic trap*, Master's thesis, University of Warsaw, Warsaw, Poland (2017).

[330] A. Dawid, M. Lewenstein, and M. Tomza, *Two interacting ultracold molecules in a one-dimensional harmonic trap*, Phys. Rev. A **97**, 063618 (2018).

[331] A. Dawid and M. Tomza, *Magnetic properties and quench dynamics of two interacting ultracold molecules in a trap*, Phys. Chem. Chem. Phys. **22**, 28140 (2020).

[332] J. D. Thompson, T. G. Tiecke, N. P. de Leon, J. Feist, A. V. Akimov, M. Gullans, A. S. Zibrov, V. Vuletić, and M. D. Lukin, *Coupling a Single Trapped Atom to a Nanoscale Optical Cavity*, Science **340**, 1202 (2013).

[333] T. Lahaye, C. Menotti, L. Santos, M. Lewenstein, and T. Pfau, *The physics of dipolar bosonic quantum gases*, Rep. Prog. Phys. **72**, 126401 (2009).

[334] R. Bücker, A. Perrin, S. Manz, T. Betz, C. Koller, T. Plisson, J. Rottmann, T. Schumm, and J. Schmiedmayer, *Single-particle-sensitive imaging of freely propagating ultracold atoms*, New J. Phys. **11**, 103039 (2009).

[335] A. Fuhrmanek, A. M. Lance, C. Tuchendler, P. Grangier, Y. R. Sortais, and A. Browaeys, *Imaging a single atom in a time-of-flight experiment*, New J. Phys. **12**, 053028 (2010).

[336] A. Bergschneider, V. M. Klinkhamer, J. H. Becher, R. Klemt, G. Zürn, P. M. Preiss, and S. Jochim, *Spin-resolved single-atom imaging of Li 6 in free space*, Phys. Rev. A **97**, 63613 (2018).

[337] M. G. Hu, Y. Liu, D. D. Grimes, Y. W. Lin, A. H. Gheorghe, R. Vexiau, N. Bouloufa-Maafa, O. Dulieu, T. Rosenband, and K. K. Ni, *Direct observation of bimolecular reactions of ultracold KRb molecules*, Science **366**, 1111 (2019).

[338] Y. Liu, M. G. Hu, M. A. Nichols, D. D. Grimes, T. Karman, H. Guo, and K. K. Ni, *Photo-excitation of long-lived transient intermediates in ultracold reactions*, Nat. Phys. **16**, 1132 (2020).

[339] S. Truppe, S. Marx, S. Kray, M. Doppelbauer, S. Hofsäss, H. C. Schewe, N. Walter, J. Pérez-Riós, B. G. Sartakov, and G. Meijer, *Spectroscopic characterization of aluminum monofluoride with relevance to laser cooling and trapping*, Phys. Rev. A **100**, 052513 (2019).

[340] E. Shuman, J. Barry, and D. DeMille, *Laser cooling of a diatomic molecule*, Nature **467**, 820 (2010).

[341] M. Tomza, K. W. Madison, R. Moszynski, and R. V. Krems, *Chemical reactions of ultracold alkali-metal dimers in the lowest-energy* $^3\Sigma$ *state*, Phys. Rev. A **88**, 050701 (2013).

[342] K. Zaremba-Kopczyk, P. S. Zuchowski, and M. Tomza, *Magnetically tunable Feshbach resonances in ultracold gases of europium atoms and mixtures of europium and alkali-metal atoms*, Phys. Rev. A **98**, 032704 (2018).

[343] M. D. Frye, S. L. Cornish, and J. M. Hutson, *Prospects of Forming High-Spin Polar Molecules from Ultracold Atoms*, Phys. Rev. X **10**, 041005 (2020).

[344] A. Dawid and M. Tomza, DOI:10.5281/zenodo.3985911 (2020), GitHub repository: Two-Molecules-1D-Trap (Version arXiv1.0).

[345] A. Doçaj, M. L. Wall, R. Mukherjee, and K. R. A. Hazzard, *Ultracold Nonreactive Molecules in an Optical Lattice: Connecting Chemistry to Many-Body Physics*, Phys. Rev. Lett. **116**, 135301 (2016).

[346] M. L. Wall, R. Mukherjee, S. S. Alam, N. P. Mehta, and K. R. A. Hazzard, *Lattice-model parameters for ultracold nonreactive molecules: Chaotic scattering and its limitations*, Phys. Rev. A **95**, 043636 (2017).

[347] M. L. Wall, N. P. Mehta, R. Mukherjee, S. S. Alam, and K. R. A. Hazzard, *Microscopic derivation of multichannel Hubbard models for ultracold nonreactive molecules in an optical lattice*, Phys. Rev. A **95**, 043635 (2017).

[348] G. E. Astrakharchik, J. Boronat, J. Casulleras, and S. Giorgini, *Beyond the Tonks-Girardeau Gas: Strongly Correlated Regime in Quasi-One-Dimensional Bose Gases*, Phys. Rev. Lett. **95**, 190407 (2005).

[349] E. Tempfli, S. Zöllner, and P. Schmelcher, *Excitations of attractive 1D bosons: Binding versus fermionization*, New J. Phys. **10**, 103021 (2008).

[350] A. Einstein and W. De Haas, *Experimental proof of the existence of Ampere's molecular currents*, in *Proc. KNAW*, Vol. 181 (1915) p. 696.

[351] F. Haake, *Quantum signatures of chaos*, Springer Series in Synergetics, Vol. 54 (Springer, 2010).

[352] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, 2006).

[353] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (The MIT Press, 2016).

[354] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into deep learning*, arXiv:2106.11342 (2021).

[355] H. Moravec, *Mind Children: The Future of Robot and Human Intelligence* (Harvard University Press, Cambridge, Massachusetts, 1988) p. 15.

[356] J. Dai and R. V. Krems, *Interpolation and Extrapolation of Global Potential Energy Surfaces for Polyatomic Systems by Gaussian Processes with Composite Kernels*, J. Chem. Theory Comput. **16**, 1386 (2020).

[357] H. Sugisawa, T. Ida, and R. V. Krems, *Gaussian process model of 51-dimensional potential energy surface for protonated imidazole dimer*, J. Chem. Phys. **153**, 114101 (2020).

[358] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, Ann. Eug. **7**, 179 (1936).

[359] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, Proc. IEEE **86**, 2278 (1998).

[360] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, *Mastering the game of Go with deep neural networks and tree search*, Nature **529**, 484 (2016).

[361] D. Silver and et. al., *Mastering the game of Go without human knowledge*, Nature **550**, 354 (2017).

[362] R. Porotti, A. Essing, B. Huard, and F. Marquard, *Deep reinforcement learning for quantum state preparation with weak nonlinear measurements*, arXiv:2107.08816 (2021).

[363] S. Borah, B. Sarma, M. Kewming, G. J. Milburn, and J. Twamley, *Measurement-Based Feedback Quantum Control with Deep Reinforcement Learning for a Double-Well Nonlinear Potential*, Phys. Rev. Lett. **127**, 190403 (2021).

[364] V. Nguyen, S. B. Orbell, D. T. Lennon, H. Moon, F. Vigneau, L. C. Camenzind, L. Yu, D. M. Zumbühl, G. A. D. Briggs, M. A. Osborne, D. Sejdinovic, and N. Ares, *Deep reinforcement learning for efficient measurement of quantum devices*, npj Quantum Inf. **7**, 100 (2021).

[365] T. Fösel, M. Yuezhen Niu, F. Marquardt, and L. Li, *Quantum circuit optimization with deep reinforcement learning*, arXiv:2103.07585 (2021).

[366] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, *Reinforcement Learning with Neural Networks for Quantum Feedback*, Phys. Rev. X **8**, 031084 (2018).

[367] C. Wang and H. Zhai, *Machine learning of frustrated classical spin models. I. Principal component analysis*, Phys. Rev. B **96**, 144432 (2017).

[368] W. Hu, R. R. Singh, and R. T. Scalettar, *Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination*, Phys. Rev. E **95**, 062122 (2017).

[369] Y. Ming, C. T. Lin, S. D. Bartlett, and W. W. Zhang, *Quantum topology identification with deep neural networks and quantum walks*, Npj Comput. Mater. **5**, 88 (2019).

[370] P. Rosson, M. Kiffner, J. Mur-Petit, and D. Jaksch, *Characterizing the phase diagram of finite-size dipolar Bose-Hubbard systems*, Phys. Rev. A **101**, 013616 (2020).

[371] A. Lidiak and Z. Gong, *Unsupervised Machine Learning of Quantum Phase Transitions Using Diffusion Maps*, Phys. Rev. Lett. **125**, 225701 (2020).

[372] P. Broecker, F. F. Assaad, and S. Trebst, *Quantum phase recognition via unsupervised machine learning*, arXiv:1707.00663 (2017).

[373] R. Bagheri, *Weight Initialization in Deep Neural Networks*, https://towardsdatascience.com/weight-initialization-in-deep-neural-networks-268a306540c0 (2020), accessed: 2022-02-16.

[374] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, *Optuna: A next-generation hyperparameter optimization framework*, in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* (2019).

[375] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, *Gradient Descent Converges to Minimizers*, arXiv:1602.04915 (2016).

[376] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, *On the importance of initialization and momentum in deep learning*, in *ICML 2013 - 30th Int. Conf. Mach. Learn.*, Vol. 28 (2013) pp. 1139–1147.

[377] Y. Liu, Y. Gao, and W. Yin, *An improved analysis of stochastic gradient descent with momentum*, arXiv:2007.07989 (2020).

[378] D. P. Kingma and J. L. Ba, *Adam: A method for stochastic optimization*, in *ICLR 2015 - 3rd Int. Conf. Learn. Represent.* (2015).

[379] Z. Zhang, *Improved Adam Optimizer for Deep Neural Networks*, in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)* (IEEE, 2018).

[380] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, *Algorithm 778: L-BFGS-B*, ACM Trans. Math. Softw. **23**, 550 (1997).

[381] L. M. Rios and N. V. Sahinidis, *Derivative-free optimization: A review of algorithms and comparison of software implementations*, J. Glob. Optim. **56**, 1247 (2012).

[382] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning* (The MIT Press, Cambridge, 2009).

[383] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, and J. Snoek, *Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift*, arXiv:1906.02530 (2019).

[384] B. Nushi, E. Kamar, and E. Horvitz, *Towards accountable AI: Hybrid human-machine analyses for characterizing system failure*, arXiv:1809.07424 (2018).

[385] M. A. Nielsen, *Neural networks and deep learning* (Determination Press, 2015).

[386] S. Dreiseitl and L. Ohno-Machado, *Logistic regression and artificial neural network classification models: a methodology review*, J. Biomed. Inf. **35**, 352 (2002).

[387] K. O'Shea and R. Nash, *An Introduction to Convolutional Neural Networks*, arXiv:1511.08458 (2015).

[388] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, *Going deeper with convolutions*, in *CVPR 2015 - IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2015) pp. 1–9.

[389] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, arXiv:1409.1556 (2014).

[390] H. Bourlard and Y. Kamp, *Auto-association by multilayer perceptrons and singular value decomposition*, Biol. Cybern. **59**, 291 (1988).

[391] G. E. Hinton and R. S. Zemel, *Autoencoders, minimum description length and Helmholtz free energy*, in *NIPS 1993 - 6th Adv. Neural. Inf. Process. Syst.* (1993) pp. 3–10.

[392] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, *Extracting and composing robust features with denoising autoencoders*, in *ICML 2008 - 25th Int. Conf. Mach. Learn.* (2008) p. 1096–1103.

[393] J. Xie, L. Xu, and E. Chen, *Image denoising and inpainting with deep neural networks*, in *NIPS 2012 - Adv. Neural Inf. Process. Syst.*, Vol. 25 (2012) pp. 341–349.

[394] F. Baldassarre, D. G. Morín, and L. Rodés-Guirao, *Deep Koalarization: Image Colorization using CNNs and Inception-ResNet-v2*, arXiv:1712.03400 (2017).

[395] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, arXiv:1312.6114 (2014).

[396] D. J. Rezende, S. Mohamed, and D. Wierstra, *Stochastic backpropagation and approximate inference in deep generative models*, in *ICML 2014 - 31st Int. Conf. Mach. Learn.* (2014).

[397] C. Doersch, *Tutorial on Variational Autoencoders* (2016).

[398] N. S. Keskar, J. Nocedal, P. T. P. Tang, D. Mudigere, and M. Smelyanskiy, *On large-batch training for deep learning: Generalization gap and sharp minima*, in *ICLR 2017 - Int. Conf. Learn. Represent.* (2017).

[399] L. Wu, Z. Zu, and W. E, *Towards understanding generalization of deep learning: Perspective of loss landscapes*, arXiv:1706.10239 (2017).

[400] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, *Averaging weights leads to wider optima and better generalization*, in *UAI 2018 - 34th Conf. Uncertain. Artif. Intell.*, Vol. 2 (2018) pp. 876–885.

[401] H. He, G. Huang, and Y. Yuan, *Asymmetric valleys: Beyond sharp and flat local minima*, in *NeurIPS 2019 - Adv. Neural Inf. Process. Syst.*, Vol. 32 (2019).

[402] L. Dinh, R. Pascanu, S. Bengio, and Y. Bengio, *Sharp minima can generalize for deep nets*, in *ICML 2017 - 34th Int. Conf. Mach. Learn.*, Vol. 3 (International Machine Learning Society (IMLS), 2017) pp. 1705–1714.

[403] Y. N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, *Identifying and attacking the saddle point problem in high-dimensional non-convex optimization*, in *NIPS 2014 - Adv. Neural. Inf. Process. Syst.*, Vol. 27 (2014) pp. 2933–2941.

[404] L. Sagun, L. Bottou, and Y. LeCun, *Eigenvalues of the Hessian in deep learning: Singularity and beyond*, arXiv:1611.07476 (2016).

[405] G. Alain, N. Le Roux, and P. A. Manzagol, *Negative eigenvalues of the Hessian in deep neural networks*, in *ICLR 2018 - Int. Conf. Learn. Represent.* (2018).

[406] A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, and Y. LeCun, *The loss surfaces of multilayer networks*, in *AISTATS 2015 - 18th Int. Conf. Artif. Intell. Stat.*, Vol. 38 (2015) pp. 192–204.

[407] P. Schulam and S. Saria, *Can you trust this prediction? Auditing pointwise reliability after learning*, in *AISTATS 2019 - 22nd Int. Conf. Artif. Intell. Stat.*, Vol. 89 (2020) pp. 1022–1031.

[408] L. Sagun, U. Evci, V. U. Güney, Y. Dauphin, and L. Bottou, *Empirical analysis of the Hessian of over-parametrized neural networks*, in *ICLR 2018 - 6th Int. Conf. Learn. Represent.* (2018).

[409] B. Ghorbani, S. Krishnan, and Y. Xiao, *An investigation into neural net optimization via Hessian eigenvalue density*, in *ICML 2019 - 36th Int. Conf. Mach. Learn.*, Vol. 97 (2019) pp. 2232–2241.

[410] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, *Visualizing the loss landscape of neural nets*, in *NIPS 2018 - Adv. Neural. Inf. Process. Syst.* (2018) pp. 6389–6399.

[411] P. W. Koh and P. Liang, *Understanding black-box predictions via influence functions*, in *ICML 2017 - 34th Int. Conf. Mach. Learn.*, Vol. 70 (2017) pp. 1885–1894.

[412] D. Madras, J. Atwood, and A. D'Amour, *Detecting extrapolation with local ensembles*, in *ICLR 2020 - Int. Conf. Learn. Represent.* (2020).

[413] A. Dawid, P. Huembeli, M. Tomza, M. Lewenstein, and A. Dauphin, *Hessian-based toolbox for reliable and interpretable machine learning in physics*, Mach. Learn.: Sci. Technol. **3**, 015002 (2021).

[414] A. Dawid, P. Huembeli, M. Tomza, M. Lewenstein, and A. Dauphin, *Phase detection with neural networks: Interpreting the black box*, New J. Phys. **22**, 115001 (2020).

[415] R. D. Cook, *Detection of Influential Observation in Linear Regression*, Technometrics **19**, 15 (1977).

[416] R. D. Cook and S. Weisberg, *Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression*, Technometrics **22**, 495 (1980).

[417] R. D. Cook and S. Weisberg, *Residuals and Influence in Regression* (Chapman and Hall, New York and London, 1982).

[418] P. W. Koh, K.-S. Ang, H. H. K. Teo, and P. Liang, *On the accuracy of influence functions for measuring group effects*, in *NeurIPS 2019 - Adv. Neural. Inf. Process. Syst.*, Vol. 32 (2019).

[419] E. Barshan, M.-E. Brunet, and G. Karolina Dziugaite, *RelatIF: Identifying Explanatory Training Examples via Relative Influence*, in *AISTATS 2020 - 23rd Int. Conf. Artif. Intell. Stat.*, Vol. 108 (2020) pp. 1899–1909.

[420] D. Hendrycks, M. Mazeika, and T. Dietterich, *Deep anomaly detection with outlier exposure*, arXiv:1812.04606 (2018).

[421] P. Huembeli, *Machine learning for quantum physics and quantum physics for machine learning*, Ph.D. thesis, Universitat Politècnica De Catalunya (2021).

[422] A. Graves, *Practical variational inference for neural networks*, in *NIPS 2011 - 25th Neural. Inf. Process. Syst.*, Vol. 24 (2011).

[423] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian approximation: Representing model uncertainty in deep learning*, in *ICML 2016 - 33rd Int. Conf. Mach. Learn.*, Vol. 48 (2016) pp. 1050–1059.

[424] B. Efron and R. Tibshirani, *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy*, Stat. Sci. **1**, 54 (1986).

[425] D. Teney, K. Kafle, R. Shrestha, E. Abbasnejad, C. Kanan, and A. van den Hengel, *On the value of out-of-distribution testing: An example of Goodhart's law*, in *NeurIPS 2020 - 34th Conf. Neural. Inf. Process. Syst.* (2020).

[426] N. Agarwal, B. Bullins, and E. Hazan, *Second-order stochastic optimization for machine learning in linear time*, J. Mach. Learn. Res. **18**, 1 (2017).

[427] J. Martens and R. Grosse, *Optimizing neural networks with Kronecker-factored approximate curvature*, in *ICML 2015 - 302nd Int. Conf. Mach. Learn.*, Vol. 3 (2015) pp. 2398–2407.

[428] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Natl. Bur. Stand. **45**, 255 (1950).

[429] Z. Yao, A. Gholami, K. Keutzer, and M. W. Mahoney, *PyHessian: Neural networks through the lens of the Hessian*, in *2020 IEEE Int. Conf. Big Data* (2020) pp. 581–590.

[430] O. Dutta, M. Gajda, P. Hauke, M. Lewenstein, D.-S. Lühmann, B. A. Malomed, T. Sowiński, and J. Zakrzewski, *Non-standard Hubbard models in optical lattices: a review*, Rep. Prog. Phys. **78**, 66001 (2015).

[431] E. Hallberg, E. Gagliano, and C. Balseiro, *Finite-size study of a spin-1/2 Heisenberg chain with competing interactions: Phase diagram and critical behavior*, Phys. Rev. B **41**, 9474 (1990).

[432] T. Mishra, J. F. Carrasquilla, and M. Rigol, *Phase diagram of the half-filled one-dimensional t-V-V' model*, Phys. Rev. B **84**, 115135 (2011).

[433] P. Weinberg and M. Bukov, *QuSpin: A Python package for dynamics and exact diagonalisation of quantum many body systems part I: Spin chains*, SciPost Phys. **2**, 003 (2017).

[434] A. Dawid, P. Huembeli, M. Tomza, M. Lewenstein, and A. Dauphin, DOI:10.5281/zenodo.3759432 (2020), GitHub repository: Interpretable-Phase-Classification (Version arXiv1.1).

[435] A. Dawid, P. Huembeli, M. Tomza, M. Lewenstein, and A. Dauphin, DOI:10.5281/zenodo.5148870 (2021), GitHub repository: Hessian-Based-Toolbox (Version arXiv1.0).

[436] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, U. Dogan, M. Kloft, F. Orabona, T. Tommasi, and A. Ganin, *Domain-adversarial training of neural networks*, J. Mach. Learn. Res. **17**, 1 (2016).

[437] P. Huembeli and A. Dauphin, *Characterizing the loss landscape of variational quantum circuits*, Quantum Sci. Technol. **6**, 025011 (2021).

[438] A. Mari, T. R. Bromley, and N. Killoran, *Estimating the gradient and higher-order derivatives on quantum hardware*, Phys. Rev. A **103**, 012405 (2021).

[439] N. Fläschner, *Ultracold fermions in tunable hexagonal lattices: From high-precision spectroscopy to the measurement of Berry curvature*, Ph.D. thesis, Universität Hamburg (2016).

[440] N. Fläschner, B. S. Rem, M. Tarnowski, D. Vogel, D.-S. Lühmann, K. Sengstock, and C. Weitenberg, *Experimental reconstruction of the Berry curvature in a Floquet Bloch band*, Science **352**, 1091 (2016).

[441] M. Tarnowski, F. N. Ünal, N. Fläschner, B. S. Rem, A. Eckardt, K. Sengstock, and C. Weitenberg, *Measuring topology from dynamics by obtaining the Chern number from a linking number*, Nat. Commun. **10**, 1728 (2019).

[442] L. Asteria, D. T. Tran, T. Ozawa, M. Tarnowski, B. S. Rem, N. Fläschner, K. Sengstock, N. Goldman, and C. Weitenberg, *Measuring quantized circular dichroism in ultracold topological matter*, Nat. Phys. **15**, 449 (2019).

[443] F. D. M. Haldane, *Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the "parity anomaly"*, Phys. Rev. Lett. **61**, 2015 (1988).

[444] A. Akhmerov, J. Sau, B. van Heck, S. Rubbert, R. Skolasiński, B. Nijholt, I. Muhammad, and T. Örn Rosdahl, *Online course on topology in condensed matter* (2015).

[445] Y. Aharonov and J. Anandan, *Phase change during a cyclic quantum evolution*, Phys. Rev. Lett. **58**, 1593 (1987).

[446] H. M. Price and N. R. Cooper, *Mapping the Berry curvature from semiclassical dynamics in optical lattices*, Phys. Rev. A **85**, 033620 (2012).

[447] A. Dauphin and N. Goldman, *Extracting the Chern number from the dynamics of a Fermi gas: Implementing a quantum Hall bar for cold atoms*, Phys. Rev. Lett. **111**, 135302 (2013).

[448] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger, *Experimental realisation of the topological Haldane model with ultracold fermions*, Nature **515**, 237 (2014).

[449] M. Aidelsburger, M. Lohse, C. Schweizer, M. Atala, J. T. Barreiro, S. Nascimbène, N. R. Cooper, I. Bloch, and N. Goldman, *Measuring the Chern number of Hofstadter bands with ultracold bosonic atoms*, Nat. Phys. **11**, 162 (2015).

[450] L. Duca, T. Li, M. Reitter, I. Bloch, M. Schleier-Smith, and U. Schneider, *An Aharonov-Bohm interferometer for determining Bloch band topology*, Science **347**, 288 (2015).

[451] D. T. Tran, A. Dauphin, A. G. Grushin, P. Zoller, and N. Goldman, *Probing topology by "heating": Quantized circular dichroism in ultracold atoms*, Sci. Adv. **3** (2017), 10.1126/sciadv.1701207.

[452] E. Alba, X. Fernandez-Gonzalvo, J. Mur-Petit, J. K. Pachos, and J. J. Garcia-Ripoll, *Seeing Topological Order in Time-of-Flight Measurements*, Phys. Rev. Lett. **107**, 235301 (2011).

[453] P. Hauke, M. Lewenstein, and A. Eckardt, *Tomography of band insulators from quench dynamics*, Phys. Rev. Lett. **113**, 045303 (2014).

[454] N. Fläschner, D. Vogel, M. Tarnowski, B. S. Rem, D. S. Lühmann, M. Heyl, J. C. Budich, L. Mathey, K. Sengstock, and C. Weitenberg, *Observation of dynamical vortices after quenches in a system with topology*, Nat. Phys. **14**, 265 (2018).

[455] T. Oka and H. Aoki, *Photovoltaic Hall effect in graphene*, Phys. Rev. B **79**, 081406 (2009).

[456] M. C. Rechtsman, J. M. Zeuner, Y. Plotnik, Y. Lumer, D. Podolsky, F. Dreisow, S. Nolte, M. Segev, and A. Szameit, *Photonic Floquet topological insulators*, Nature **496**, 196 (2013).

[457] A. Dauphin, D.-T. Tran, M. Lewenstein, and N. Goldman, *Loading ultracold gases in topological Floquet bands: the fate of current and center-of-mass responses*, 2D Mater. **4**, 024010 (2017).

[458] T. Kitagawa, E. Berg, M. Rudner, and E. Demler, *Topological characterization of periodically driven quantum systems*, Phys. Rev. B **82**, 235114 (2010).

[459] M. S. Rudner, N. H. Lindner, E. Berg, and M. Levin, *Anomalous edge states and the bulk-edge correspondence for periodically driven two-dimensional systems*, Phys. Rev. X **3**, 031005 (2013).

[460] A. Kumar, M. Rodriguez-Vega, T. Pereg-Barnea, and B. Seradjeh, *Linear response theory and optical conductivity of Floquet topological insulators*, Phys. Rev. B **101**, 174314 (2020).

[461] N. Käming, A. Dawid, K. Kottmann, M. Lewenstein, K. Sengstock, A. Dauphin, and C. Weitenberg, DOI:10.5281/zenodo.4700379 (2021), GitHub repository: unsupervised-ML-of-topological-experimental-data (Version 2.0.

[462] S. van der Walt, S. C. Colbert, and G. Varoquaux, *The NumPy array: A structure for efficient numerical computation*, Comput. Sci. Eng. **13**, 22 (2011).

[463] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, *Automatic differentiation in PyTorch*, in *NIPS 2017 - 31st Conf. Neural Inf. Process. Syst.* (2017).

[464] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: large-scale machine learning on heterogeneous systems* (2015), software available from tensorflow.org.

[465] G. E. Hinton and R. R. Salakhutdinov, *Reducing the dimensionality of data with neural networks*, Science **313**, 504 (2006).

[466] A. Fitzgibbon, M. Pilu, and R. B. Fisher, *Direct least square fitting of ellipses*, IEEE Trans. Pattern Anal. Mach. Intell. **21**, 476 (1999).

[467] C. F. Ockeloen, A. F. Tauschinsky, R. J. C. Spreeuw, and S. Whitlock, *Detection of small atom numbers through image processing*, Phys. Rev. A **82**, 061606 (2010).

[468] R. Fung, A. M. Hanna, O. Vendrell, S. Ramakrishna, T. Seideman, R. Santra, and A. Ourmazd, *Dynamics from noisy data with extreme timing uncertainty*, Nature **532**, 471 (2016).

[469] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Trans. Image Process. **13**, 600 (2004).

[470] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12**, 2825 (2011).

[471] F. Crameri, G. Shephard, and P. Heron, *The misuse of colour in science communication*, Nat. Commun. **11**, 5444 (2020).

[472] G. N. Watson, *A note on the polynomials of Hermite and Laguerre*, J. London Math. Soc. **s1-13**, 29 (1938).

[473] Wolfram Research, Inc., *The mathematical functions site* (2020), function ID: 05.01.21.0009.01.

[474] B. A. Pearlmutter, *Fast exact multiplication by the Hessian*, Neural Comput. **6**, 147 (1994).