

Politechnika Wrocławska

DISCIPLINE OF SCIENCE:

physical sciences

DOCTORAL DISSERTATION

Machine learning for exploring solid-state quantum computing

Mateusz Krawczyk

Supervisor:

dr hab. inż. Jarosław Pawłowski

Keywords: solid-state physics, quantum information, quantum computing, machine learning, neural networks

WROCŁAW 2026

Table of Contents

Acknowledgements i

Abstract ii

List of Abbreviations iii

1 Introduction 1

2 Quantum Information 3
2.1 Quantum states . . . 3
2.2 Quantum correlations . . . 6
2.3 Measurements and quantum state tomography . . . 11

3 Topological Quantum Computing 21
3.1 Topological superconductivity and Majorana zero modes . . . 22
3.2 Quantum dot chains as tunable realizations of Kitaev-like physics . . . 29
3.3 Synthetic data generation for topological systems . . . 33

4 Machine Learning 37
4.1 Basic ideas . . . 37
4.2 Advanced deep learning architectures . . . 42
4.3 Applications . . . 49

5 Synopsis of the published papers 52
5.1 Data-driven criteria for quantum correlations . . . 53
5.2 Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning . . . 55
5.3 Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests . . . 57
5.4 AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes . . . 59

6 Synopsis of the unpublished manuscripts 62
6.1 Learning quantum tomography from incomplete measurements . . . 62
6.2 Learning Hamiltonians for solid-state quantum simulators . . . 65

7 Conclusions 68

Bibliography 71

Publications 88
Data-driven criteria for quantum correlation . . . 89
Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning . . . 101
Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests . . . 114
AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes 127

Unpublished Results **140**
Learning quantum tomography from incomplete measurements 141
Learning Hamiltonians for solid-state quantum simulators 154

Acknowledgements

I would like to express my sincere gratitude to my supervisor, dr hab. inż. Jarosław Pawłowski, for his guidance, continuous support, and many inspiring scientific discussions throughout my doctoral studies. His insight and encouragement were essential in shaping both the physical problems addressed in this dissertation and the machine-learning methods developed to study them.

I would also like to express my special gratitude to my previous supervisor, Maciej M. Maśka, for introducing me to scientific research, for his guidance during the early stages of my doctoral work, and for the many valuable discussions and encouragement that significantly influenced my scientific development.

I am grateful to all co-authors of the publications and manuscripts included in this dissertation: Jarosław Pawłowski, Maciej M. Maśka, Katarzyna Roszak, and Pavel Baláž. I thank them for their collaboration, discussions, critical comments, and joint work on the problems of quantum correlations, quantum tomography, Majorana modes, and Hamiltonian learning. I would also like to thank M. Margańska-Łyżniak for helpful discussions and comments related to the work on AI-enhanced tuning toward Majorana modes.

The research presented in this dissertation was supported by the National Science Centre, Poland, under Grant No. 2021/43/D/ST3/01989. Parts of the included research were also supported within the QuantERA II Programme, which received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017733, with funding from the organization MEYS, and by the research programme of the Strategy AV21 AI: Artificial Intelligence for Science and Society.

I gratefully acknowledge the Polish high-performance computing infrastructure PLGrid, including ACK Cyfronet AGH, for providing computational resources and support through computational grants No. PLG/2024/017284 and No. PLG/2025/018433.

Finally, I would like to thank my family and friends for their patience, encouragement, and support during the preparation of this dissertation.

Abstract

The development of scalable quantum computing requires efficient methods for characterizing, modeling, and controlling complex quantum systems from experimentally accessible data. In solid-state systems, such data are often only indirectly related to the quantities of interest: tomographic measurements must be inverted to reconstruct quantum states, while transport spectra must be interpreted in terms of Hamiltonian parameters or topological regimes. This dissertation investigates how machine learning can bridge this gap, with a particular focus on quantum state analysis, measurement-based inference, and physics-informed control of solid-state quantum systems.

The first part of the dissertation concerns the identification of quantum correlations. An unsupervised neural network approach is developed to learn the structure of separable three-qubit states and detect correlated states as anomalies. This direction is extended through supervised and semisupervised Siamese neural networks, where entanglement-preserving symmetries, such as local unitary transformations and qubit permutations, are incorporated into the learning process to improve entanglement detection and generalization beyond the training data.

The second part focuses on analyzing experimentally motivated measurement representations. Neural networks and random forests are trained to infer entanglement directly from tomography data, revealing that different models may rely on qualitatively different measurement features. The dissertation further develops learning-based quantum tomography from informationally incomplete measurements, using neural networks to refine pseudoinverse reconstruction and recurrent architectures to jointly reconstruct quantum states and optimize the order of measurements.

The final part addresses solid-state quantum simulators and topological platforms. A physics-informed neural network is utilized for autotuning quantum dot Hamiltonians toward regimes supporting Majorana zero modes, using conductance maps as input and a differentiable modes-scoring quasi-metric as the training objective. In addition, a physics decoder autoencoder framework is developed for Hamiltonian learning, where a neural encoder predicts effective Hamiltonian parameters from transport measurements while a scattering-matrix-based decoder enforces the underlying physical model.

Across these directions, the dissertation shows that combining data-driven learning with physical constraints improves accuracy, interpretability, and generalization in quantum-data analysis and device control. The presented results demonstrate that physics-informed machine learning can serve as a unifying framework for extracting quantum correlations, reconstructing states from incomplete data, learning effective Hamiltonians, and guiding quantum systems toward desired operational regimes.

List of Abbreviations

Abbreviation	Meaning
ABS	Andreev bound state
AI	artificial intelligence
AZ	Altland–Zirnbauer
BdG	Bogoliubov–de Gennes
CNN	convolutional neural network
DL	deep learning
ELBO	evidence lower bound
EoF	entanglement of formation
KC	Kitaev chain
LU	local unitary
LSTM	long short-term memory
MDP	Markov decision process
ML	machine learning
MLIP	machine-learned interatomic potential
MLE	maximum likelihood estimation
MLP	multilayer perceptron
MZM	Majorana zero mode
NISQ	noisy intermediate-scale quantum
NLP	natural language processing
NN	neural network
NP	nondeterministic polynomial time
NPT	negative partial transpose
PCA	principal component analysis
PDE	partial differential equation
PHS	particle-hole symmetry
PINN	physics-informed neural network
PINNAT	PINN autotuning
POVM	positive-operator valued measure
PPT	positive partial transpose
PPTES	positive partial transpose entangled state
PVM	projection-valued measurement
QD	quantum dot
QST	quantum state tomography
RBM	restricted Boltzmann machine
ReLU	rectified linear unit
RF	random forest
SGD	stochastic gradient descent
SVM	support vector machine
TFT	temporal fusion transformer
TQC	topological quantum computing
TRS	time-reversal symmetry

Abbreviation	**Meaning**
VAE	variational autoencoder
ViT	vision transformer
ZBP	zero-bias peak

1 Introduction

The drive to achieve scalable quantum computing has been one of the main research areas in recent years. The development of semiconductor quantum dots (QDs) and hybrid superconductor-semiconductor structures, together with the intensive study of currently available noisy intermediate-scale quantum (NISQ) processors, are just a few manifestations of this interest[1,2,3,4]. Future progress in this field depends on the analysis of a large number of experimental data, which can be in the form of transport spectra, charge stability diagrams, or tomographic measurement outcomes[5,6,7,8,9,10]. Unfortunately, in many solid-state and quantum-information experiments, the measured quantities are related only indirectly to the desired physical properties: for example, transport spectra must be interpreted in terms of Hamiltonian parameters or topological regimes, while measured outcome frequencies in quantum-state tomography must be used to reconstruct a density matrix. This raises significant challenges for traditional analysis and appropriate device tuning[3,11].

At the same time, two complementary paradigms have emerged in quantum computing. On the one hand, quantum information theory provides a framework for understanding and quantifying nonclassical resources such as entanglement and quantum correlations, which are essential for quantum advantage. In particular, for pure state quantum computations, unbounded growth of multipartite entanglement with the system size is necessary to obtain an exponential speedup over classical computation[12,13]. On the other hand, topological quantum computing (TQC) aims to achieve intrinsic robustness by encoding quantum information in protected many-body states, such as those formed by Majorana zero modes (MZMs). While these approaches are often treated separately, they are deeply connected: both rely on identifying, characterizing, and controlling complex quantum states and their correlations, typically from experimentally limited data such as incomplete tomographic measurements, noisy measurement statistics, or transport observables that provide only indirect signatures of the underlying quantum state or Hamiltonian.

Simultaneously, efficient usage of modern machine learning (ML) techniques has been observed in many applications, from tabular data predictions to computer vision, natural language processing, and content generation[14,15,16]. Therefore, applying ML methods to the analysis of quantum systems is a natural direction, since such methods are well-suited for processing complex, high-dimensional, and noisy data. In this context, ML provides a unifying framework that can bridge quantum information and topologically protected systems by enabling data-driven extraction of physical properties, detection of quantum correlations, reconstruction of quantum states from incomplete measurements, Hamiltonian learning, and optimization of system parameters.

The central goal of this dissertation is to explore how ML methods can be leveraged to extract physically relevant information from quantum data and to optimize the closed-loop feedback process between measurement and system calibration. These two aspects, i.e., data-driven inference and adaptive control, provide a common methodological framework for the problems studied throughout the dissertation. In the quantum-information setting, ML is used to identify correlations and reconstruct quantum states from incomplete or indirect measurement data[17,18]. In solid-state systems, transport characteristics are used to infer Hamiltonian parameters or determine control updates that steer quantum devices toward desired

operating regimes[19,20]. In this way, the dissertation develops a measurement-driven perspective in which ML connects experimentally accessible observables with the underlying physical description of the system.

A major motivation for the research presented in this dissertation is the realization of large-scale, fault-tolerant quantum computing, for which topologically protected quantum systems offer a promising route by suppressing certain classes of local errors[21]. Engineered platforms such as semiconductor nanowires and QD chains offer promising routes toward this goal, but their practical implementation requires tuning of many coupled parameters and careful discrimination between topological and trivial low-energy states[4]. At the same time, quantum-information tasks such as detecting and quantifying correlations become increasingly difficult when only incomplete measurements are available. These complementary challenges motivate the specific research directions explored in this dissertation: ML-based criteria for quantum correlations (Section 5.1), Siamese neural network methods for entanglement detection (Section 5.2), learning entanglement from tomographic data (Section 5.3), neural network-assisted reconstruction from incomplete measurements (Section 6.1), AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes (Section 5.4), and Hamiltonian learning for solid-state quantum simulators (Section 6.2).

The dissertation starts with three sections that summarize the theoretical and methodological background needed to understand the research papers. Section 2 covers key concepts from quantum information theory and quantum correlations, followed by an overview of quantum state tomography (QST) and measurement formalism. Subsequently, Section 3 describes topological systems, including MZMs, and their realization in hybrid solid-state platforms, with particular emphasis on semiconductor nanowires and coupled QD chains that can realize effective one-dimensional topological superconducting models[22,4]. Finally, Section 4 reviews core ML concepts, architectures, and their recent applications in physics, highlighting their growing role as a unifying tool for quantum data analysis and control.

In Section 5, a synopsis of the published research papers is provided, emphasizing their contribution to the overarching goal of combining ML with quantum information and TQC. The individual contribution of the author of this dissertation to each published work is summarized at the beginning of that section. Section 6 completes the picture of the research results by including the yet unpublished papers which constitute an integral part of the conducted research, together with a corresponding statement of the author's individual contribution. The dissertation ends with Section 7, which summarizes the results and highlights the conclusions of the PhD dissertation.

2 Quantum Information

Quantum information theory provides the basic language for describing how information is represented, processed, and measured in quantum systems. Unlike classical information, which can be described by probability distributions over mutually exclusive states, quantum information is encoded in state vectors or density matrices and can involve coherence, superposition, and correlations that have no direct classical analogue[23]. These features are responsible for the potential advantage of quantum technologies, but at the same time make the characterization of quantum systems substantially more demanding.

A central challenge in quantum information processing is that the relevant mathematical object, the quantum state, is not directly accessible in an experiment. Instead, one observes only measurement outcomes whose probabilities depend on the state through the Born rule. As a result, reconstructing a state, detecting entanglement, or quantifying more general quantum correlations typically requires solving an inverse problem based on incomplete and noisy data[6,7,8]. This difficulty becomes especially important in the context of the ML approaches studied in this dissertation, where the model input data are often density matrices, tomographic measurement outcomes, or reduced descriptions of quantum states.

This section introduces the quantum information concepts used throughout the dissertation. First, the formal description of pure and mixed quantum states is presented, together with the notation for composite systems and reduced density matrices. Next, the section discusses quantum correlations, including separability, entanglement, entanglement measures, and discord-type correlations. Then, the measurement formalism and QST are reviewed, with particular emphasis on informational completeness, incomplete measurements, and data-driven reconstruction methods. Finally, the section briefly discusses the generation of synthetic quantum state and measurement datasets, which provide controlled input data for the ML models used later in the dissertation.

2.1 Quantum states

A finite d-dimensional quantum system is described by a Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$, where $d = \dim(\mathcal{H})$ is the number of basis states available to the system. In particular, for a register of N qubits, one has $d = 2^N$. *Pure states* are unit vectors $|\psi\rangle \in \mathcal{H}$ with inner product $\langle\psi \,|\, \psi\rangle = 1$, while general (possibly mixed) states are described by a *density matrix* (operator) ρ satisfying

$$\rho = \rho^\dagger, \qquad \rho \succeq 0, \qquad \operatorname{Tr}\rho = 1, \tag{1}$$

meaning that ρ is a Hermitian, positive semidefinite matrix with unit trace. In spectral decomposition

$$\rho = \sum_{j=1}^{d} \lambda_j \,|\phi_j\rangle\,\langle\phi_j|\,, \qquad \lambda_j \geq 0, \qquad \sum_j \lambda_j = 1, \tag{2}$$

where $\{|\phi_j\rangle\}$ is an orthonormal eigenbasis. The eigenvalues $\{\lambda_j\}$ can be interpreted as a classical probability distribution over orthogonal pure states.

A state is pure if $\rho^2 = \rho$. The same distinction can also be expressed using the purity,

$$\gamma(\rho) \equiv \mathrm{Tr}(\rho^2) \in \left[\frac{1}{d}, 1\right], \qquad \gamma(\rho) = 1 \Leftrightarrow \rho \text{ pure.} \tag{3}$$

The $\mathrm{rank}(\rho)$ quantifies the minimal number of pure states that are used for density matrix decomposition, in particular, for pure states $\mathrm{rank}(\rho) = 1$ while for the maximally mixed state $\mathrm{rank}(\rho) = d$.

Composite systems and block structure of density matrices

Composite systems are described by tensor products, $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. In particular, a system composed of two subsystems A and B is called *bipartite*. If we set orthonormal bases as $\{|i\rangle_A\}_{i=1}^{d_A}$ for subsystem A and $\{|\mu\rangle_B\}_{\mu=1}^{d_B}$ for subsystem B, then any density matrix in $\mathcal{H}_{AB}$ can be expressed as

$$\rho_{AB} = \sum_{i,j=1}^{d_A} \sum_{\mu,\nu=1}^{d_B} \rho_{i\mu,\,j\nu}\ (|i\rangle_{AA}\langle j|) \otimes (|\mu\rangle_{BB}\langle\nu|)\,. \tag{4}$$

It is often useful to regard ρ_{AB} as a $d_A \times d_A$ *block matrix* whose blocks are $d_B \times d_B$ operators on subsystem B:

$$\rho_{AB} = \begin{pmatrix} (\rho_{11})_B & (\rho_{12})_B & \cdots & (\rho_{1d_A})_B \\ (\rho_{21})_B & (\rho_{22})_B & \cdots & (\rho_{2d_A})_B \\ \vdots & \vdots & \ddots & \vdots \\ (\rho_{d_A1})_B & (\rho_{d_A2})_B & \cdots & (\rho_{d_Ad_A})_B \end{pmatrix}, \qquad (\rho_{ij})_B \equiv \sum_{\mu,\nu=1}^{d_B} \rho_{i\mu,\,j\nu}\ |\mu\rangle_{BB}\langle\nu|\,. \tag{5}$$

In this representation, the *partial trace* over B reduces to taking traces of the individual blocks, which we write as:

$$(\rho_A)_{ij} = \sum_{\mu=1}^{d_B} \rho_{i\mu,j\mu} = \mathrm{Tr}\left[(\rho_{ij})_B\right], \qquad \text{and therefore} \qquad \rho_A = \mathrm{Tr}_B(\rho_{AB}). \tag{6}$$

Similarly, tracing out A corresponds to summing diagonal blocks as operators on B:

$$\rho_B = \mathrm{Tr}_A(\rho_{AB}) = \sum_{i=1}^{d_A} (\rho_{ii})_B. \tag{7}$$

This block viewpoint is especially valuable when considering correlation structure, as the off-diagonal blocks of ρ_{AB} encode coherences between A-basis sectors while the diagonal blocks of ρ_{AB} encode states of B conditioned on the subsystem A being projected on one of the basis states.

Entropy

The *von Neumann entropy* $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$, is the quantum analog of the classical Shannon entropy and quantifies the degree of mixedness or uncertainty associated with a quantum state. For bipartite systems, the total amount of correlations, both classical and quantum, is captured by the *quantum mutual information*,

$$I(A{:}B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}), \tag{8}$$

which measures the deviation of the joint state from the product state of its marginals. Mutual information vanishes if and only if $\rho_{AB} = \rho_A \otimes \rho_B$, and increases with the presence of correlations, reaching its maximum for a maximally entangled pure state[23].

Similarity measures between quantum states

Quantifying the similarity between two quantum states is a central task in quantum information theory, with applications in QST[7,8], benchmarking of quantum devices[24,25], error analysis[23], and quantum control[26]. The common measure of closeness is *quantum fidelity*[27,23],

$$F(\rho_1, \rho_2) = \left[\mathrm{Tr}\left(\sqrt{\sqrt{\rho_1}\, \rho_2 \, \sqrt{\rho_1}} \right) \right]^2 . \tag{9}$$

The fidelity satisfies $0 \leq F(\rho_1, \rho_2) \leq 1$, where $F(\rho_1, \rho_2) = 1$ if and only if $\rho_1 = \rho_2$, and $F(\rho_1, \rho_2) = 0$ for orthogonal states. In case of pure states $|\psi_1\rangle$ and $|\psi_2\rangle$, the fidelity can be simplified to

$$F(|\psi_1\rangle, |\psi_2\rangle) = |\langle \psi_1 | \psi_2 \rangle|^2, \tag{10}$$

corresponding to the squared inner product of the state vectors.

While fidelity provides a measure of similarity, it does not define a metric on the space of quantum states. For that reason, closely related *Bures distance*[28,27] is defined as

$$D_{\mathrm{Bures}}(\rho_1, \rho_2) = \sqrt{2\left(1 - \sqrt{F(\rho_1, \rho_2)}\right)}. \tag{11}$$

The Bures distance is a metric on the space of quantum states, i.e., it is non-negative, symmetric, vanishes if and only if $\rho_1 = \rho_2$, and satisfies the triangle inequality. It also satisfies a contractive property under quantum operations[23].

Useful special cases and parametrizations

For a single quantum bit (qubit, $d = 2$), every density matrix has the Bloch form[23]

$$\rho = \frac{1}{2}\left(\mathbb{1} + \mathbf{r} \cdot \boldsymbol{\sigma}\right), \qquad \mathbf{r} \in \mathbb{R}^3, \qquad \|\mathbf{r}\| \leq 1, \tag{12}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is a vector of Pauli matrices and $\boldsymbol{r} = (r_x, r_y, r_z)$ is the real-valued *Bloch vector*. In such case purity becomes $\mathrm{Tr}(\rho^2) = (1 + \|\mathbf{r}\|^2)/2$, hence $\|\mathbf{r}\| = 1$ for pure states, giving parametrization for so-called *Bloch sphere*. For bipartite systems, the joint density matrix can be expanded in a tensor-product operator basis (such as Pauli matrices for qubits or generalized Gell-Mann matrices for higher dimensions), yielding a parametrization in terms of local Bloch vectors for each subsystem and a correlation tensor describing inter-system correlations[23,29]. It should be noted, however, that for larger systems, such as two qubits, the manifold of pure states no longer has a sphere geometry[30].

A particularly important restricted parametrization for two-qubit systems is given by the class of so-called *X states*[31], whose density matrices possess nonzero

elements only along the main diagonal and anti-diagonal in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, namely

$$\rho_X = \begin{pmatrix} \rho_{11} & 0 & 0 & \rho_{14} \\ 0 & \rho_{22} & \rho_{23} & 0 \\ 0 & \rho_{23}^* & \rho_{33} & 0 \\ \rho_{14}^* & 0 & 0 & \rho_{44} \end{pmatrix}, \qquad \sum_{i=1}^{4} \rho_{ii} = 1, \tag{13}$$

with positivity imposing additional constraints on the matrix elements. X states constitute a seven-parameter family and include many physically relevant states such as Bell states, Werner states, and thermal states of spin-chain Hamiltonians[32]. Their reduced dimensionality allows for closed-form expressions for entanglement measures, quantum discord, and other correlation quantifiers[33,34].

The X states can be naturally generalized to N-qubit systems by restricting the density matrix to have nonzero entries only along the main diagonal and anti-diagonal in the computational basis. An N-qubit X state thus takes the form

$$\rho_X^{(N)} = \sum_{i=1}^{2^N} a_i |i\rangle\langle i| + \sum_{i=1}^{2^{N-1}} \left(z_i |i\rangle\langle \bar{i}| + z_i^* |\bar{i}\rangle\langle i| \right), \tag{14}$$

where $|i\rangle$ denotes computational basis states and $|\bar{i}\rangle$ their bitwise complements. The coefficients a_i are real and satisfy normalization and positivity conditions, while z_i are complex coherence terms. This generalized X structure preserves a sparse parametrization while capturing a wide class of multipartite entangled states and commonly arises in open-system dynamics with certain symmetries[35,36].

2.2 Quantum correlations

Quantum correlations are a fundamental resource for emerging quantum technologies, with applications ranging from quantum cryptography and secure communication[37,38] to quantum computation and information processing[23,39]. Their utility arises from correlations that have no classical counterpart and can enable phenomena such as nonlocal state manipulation and enhanced information transfer[40,41,42,43]. In pure bipartite states, nonclassical correlations are associated with entanglement, whereas mixed states can exhibit different forms of correlations, including entanglement and quantum discord[44,45,46]. Entanglement is absent for separable states, which can be expressed as convex mixtures of product states[47,48,17]. However, separable states may still possess nonzero quantum discord. A state has zero discord with respect to measurements on subsystem B when there exists a local orthonormal measurement basis on B for which the corresponding measurement leaves the global state unchanged[49,50]. Such states may still contain classical correlations, meaning that measurements on one subsystem can reveal classically correlated information about the other. The characteristics of these two types of quantum correlations are the main topic of the current section.

Separability and entanglement

A bipartite state is *separable* if it can be written as a convex combination of product states,

$$\rho_{AB}^{\text{sep}} = \sum_k p_k \, \rho_A^{(k)} \otimes \rho_B^{(k)}, \qquad p_k \geq 0, \; \sum_k p_k = 1. \tag{15}$$

If no such decomposition exists, the state is *entangled*. While the definition of a separable set is explicit, constructing the decision boundary between separable and entangled states is an NP-hard problem[17,51].

The important property of entanglement is that it does not change under *local unitary (LU) operations*, i.e., operations that can be implemented by local quantum gates on A and B. Hence, any entanglement measure must be invariant under LU transformations.

Pure state entanglement: Schmidt decomposition and entropy

For a pure bipartite state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, the *Schmidt decomposition* states that there exist orthonormal bases $\{|i\rangle_A\}$ and $\{|i\rangle_B\}$ such that

$$|\psi\rangle_{AB} = \sum_{i=1}^{r} \sqrt{\lambda_i}\, |i\rangle_A \otimes |i\rangle_B\,, \qquad \lambda_i \geq 0, \quad \sum_i \lambda_i = 1, \tag{16}$$

where r is the *Schmidt rank*, i.e., the minimal number of nonzero eigenvalues $\{\lambda_i\}$ sufficient to represent $|\psi\rangle_{AB}$. From the Schmidt decomposition, one can derive that $|\psi\rangle_{AB}$ is entangled if and only if its Schmidt rank is strictly greater than 1. In other words, a pure state is entangled if and only if its reduced states are mixed. Based on that one can utilize von Neumann entropy as entanglement measure, called also *entanglement entropy*

$$E(|\psi\rangle) \equiv S(\rho_A) = S(\rho_B) = -\sum_i \lambda_i \log \lambda_i. \tag{17}$$

Alternatively, one can use *Rényi entropies*

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho^\alpha), \qquad \alpha > 0,\ \alpha \neq 1, \tag{18}$$

which are useful in one-shot and finite-copy settings as well as characterizing entanglement spectra[52].

Mixed-state entanglement measures

When considering mixed states, as per the definition, von Neumann entropy is always nonzero, and one cannot use it explicitly as a valid entanglement measure. Instead, the generalization of this metric, *entanglement of formation* (EoF)[53] is defined as the minimum average entanglement entropy over all pure state decompositions of $\rho = \sum_k p_k\, |\psi_k\rangle\, \langle\psi_k|$:

$$\mathrm{EoF}(\rho) = \min_{\{p_k, |\psi_k\rangle\}} \sum_k p_k\, E(|\psi_k\rangle). \tag{19}$$

For two qubits, EoF is an explicit function of the *concurrence*[54]

$$\mathrm{EoF}(\rho) = h\left(\frac{1+\sqrt{1-C(\rho)^2}}{2}\right), \qquad h(x) = -x\log_2 x - (1-x)\log_2(1-x), \tag{20}$$

where

$$C(\rho) = \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4), \tag{21}$$

giving an analytically computable benchmark for learning and tomography tasks. Here, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ are the eigenvalues of the matrix

$$R = \sqrt{\sqrt{\rho}(\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)\sqrt{\rho}}, \tag{22}$$

in decreasing order. Typically, σ_y denotes the respective Pauli matrix and ρ^* is the complex conjugate of the two-qubit density matrix ρ.

For systems larger than 2 qubits (strictly for $\mathcal{H}^2 \otimes \mathcal{H}^3$ systems), no efficiently computable closed-form measure completely characterizes arbitrary mixed-state entanglement[30,55]. Existing criteria are either conclusive but not practically usable (as they require optimization in high-dimensional spaces), or easy to compute but not decisive[30]. Therefore, different entanglement measures and criteria are used depending on the problem, including EoF, negativity, and entanglement witnesses discussed below. Each of them has its own advantages and limitations, and not all such quantities are measures in the strict axiomatic sense.

PPT criterion, negativity, and logarithmic negativity

One of such inconclusive entanglement criteria is *positivity under partial transpose* (PPT)[56,57], where *partial transpose* in the subsystem B is defined as

$$\langle i_A j_B| \rho^{T_B} |k_A l_B\rangle = \langle i_A l_B| \rho |k_A j_B\rangle , \tag{23}$$

while positivity means that all eigenvalues are non-negative. The criterion states that all separable states satisfy $\rho^{T_B} \succeq 0$, thus it forms a necessary condition for separability. The PPT criterion is also a sufficient condition up to $\mathcal{H}^2 \otimes \mathcal{H}^3$ systems, but not in higher dimensional spaces where PPT entangled (bound-entangled) states exist[17].

PPT itself is a separability criterion rather than a quantitative measure of entanglement. A computable entanglement measure derived from this criterion is the *negativity*[58],

$$\mathcal{N}(\rho) = \frac{\|\rho^{T_B}\|_1 - 1}{2}, \tag{24}$$

where $\|X\|_1 = \mathrm{Tr}\sqrt{X^\dagger X}$, and the related *logarithmic negativity*[59],

$$E_{\mathcal{N}}(\rho) = \log_2 \|\rho^{T_B}\|_1 = \log_2(2\mathcal{N}(\rho) + 1) . \tag{25}$$

These measures are particularly useful in practice as they rely on the spectrum of ρ^{T_B} and therefore scale more effectively than optimization-defined measures such as EoF.

Entanglement witnesses

Another approach to entanglement detection is provided by so-called *entanglement witnesses*. An entanglement witness is defined as an observable W such that $\mathrm{Tr}(W\rho) \geq 0$ for all separable states but $\mathrm{Tr}(W\rho) < 0$ for at least one entangled state. For every entangled state, there exists at least one witness that detects it. However, no single linear entanglement witness can detect all entangled states, and different witnesses generally detect different subsets of the entangled-state space[17].

The connection with partial transposition follows from the fact that any state with a negative eigenvalue of ρ^{T_B} admits an entanglement witness that detects it. Consequently, every negative partial transpose (NPT) state is entangled, whereas PPT entangled states (PPTES) occurring in higher-dimensional systems cannot be identified by the PPT criterion alone[17].

Beyond entanglement: discord and related correlation notions

Mixed states can display nonclassical correlations even when separable. Such correlations are captured by the notion of *quantum discord*, originally introduced in Refs.[49]. Quantum discord is defined as the difference between two quantum generalizations of mutual information, namely

$$\mathcal{D}_B(\rho_{AB}) = I(A:B) - \mathcal{J}_B(\rho_{AB}), \tag{26}$$

where $I(A:B)$ is the quantum mutual information introduced in Eq. 8, and $\mathcal{J}_B(\rho_{AB})$ denotes the classical correlations obtained by performing a complete set of measurements on subsystem B,

$$\mathcal{J}_B(\rho_{AB}) = \max_{\{\Pi_k^B\}} \left[S(\rho_A) - \sum_k p_k S\left((\mathbb{1}_A \otimes \Pi_k^B)\rho_{AB}(\mathbb{1}_A \otimes \Pi_k^B)/p_k \right) \right], \tag{27}$$

with Π_k^B being a projective measurement (PVM) operator on subsystem B, $p_k = \mathrm{Tr}[(\mathbb{1}_A \otimes \Pi_k^B)\rho_{AB}]$. Details of PVM are further discussed in section 2.3.

Operationally, discord quantifies the amount of information about one subsystem that can be obtained through local measurements on the other subsystem, beyond what can be accounted for by classical correlations. A composite system AB has zero discord (with respect to measurements on subsystem B) if and only if its state is classical on subsystem B, i.e., diagonal in some orthonormal basis of B, while the conditional states of subsystem A may remain arbitrary. Such states, often referred to as classical-quantum states, can be written as

$$\rho_{AB} = \sum_i p_i \, \rho_A^{(i)} \otimes |\phi_i\rangle_B\langle\phi_i| \,, \tag{28}$$

where $\{|\phi_i\rangle_B\}$ forms an orthonormal basis on subsystem B and $\{\rho_A^{(i)}\}$ are arbitrary states on subsystem A.

Quantifying the amount of discord is an NP-complete problem[60], and is in general more demanding than computing entanglement measures even for small systems[61,62]. At the same time, deciding whether a state has zero discord is a comparatively simple task, which can be achieved by writing ρ_{AB} in the block form with respect to the measured subsystem B and checking whether all blocks are normal and mutually commuting,

$$[\rho_{\mu\nu}, \rho_{\mu\nu}^\dagger] = 0, \qquad [\rho_{\mu\nu}, \rho_{\alpha\beta}] = 0 \quad \text{for all } \mu, \nu, \alpha, \beta. \tag{29}$$

If these conditions are satisfied, the blocks can be diagonalized in a common basis, and the state has zero discord with respect to measurements on subsystem B[61,63].

The relation between separability, entanglement, and discord-type correlations is summarized schematically in Fig. 1. The set of separable states consists of two

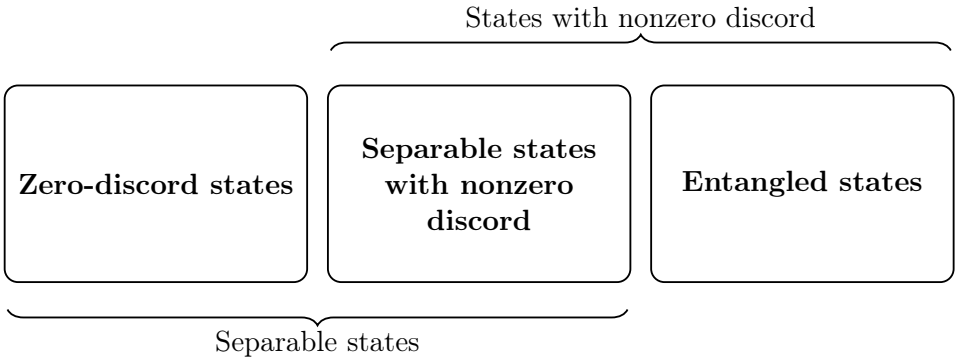


Figure 1: Schematic relation between separability, entanglement, and discord-type quantum correlations for bipartite states. Zero-discord states form a subset of separable states, while separable states with nonzero discord contain quantum correlations without entanglement. Entangled states always have nonzero discord.

qualitatively different classes. The first one contains zero-discord states, whose correlations can be interpreted as fully classical with respect to a suitable local measurement basis. The second one contains separable states with nonzero discord, which are not entangled but still possess genuinely quantum correlations revealed by the disturbance or information loss caused by local measurements. Entangled states form a separate class outside the separable set and always have nonzero discord. Therefore, discord is a broader notion of quantum correlation than entanglement: every entangled state is discordant, but not every discordant state is entangled.

Machine learning approaches to detecting entanglement and quantum correlations

As stated before, the detection of the entanglement is an NP-hard problem in general, and even when computationally efficient criteria exist, i.e., criteria that can be evaluated with polynomial computational cost for a given density matrix (e.g., PPT in low dimensions), experimental constraints may allow to extract only partial information about ρ. These challenges are the main reason for the investigation of different, numerical approaches for entanglement identification [30,64,65].

The most recent methods try to utilize ML approaches for this goal. One line of work suggests training supervised models to decide whether a given state is separable or entangled from a chosen feature representation, e.g., the *separability-entanglement classifier* introduced by Lu *et al.*, which learns a decision rule for two-qubit and two-qutrit states [66].

Instead of simply classifying a full density matrix as separable or entangled, one can use ML to construct or adapt entanglement witnesses to the given measurement constraints or specific state families. For instance, Roik *et al.* [67] compare artificial neural network (NN)-based witnesses with human-designed witnesses for two-qubit states, suggesting that NNs can outperform analytical witnesses under limited information access. Further research suggests that advanced neural architectures are also capable of more accurate identification of the entanglement straight from the density matrices [68,69,70]. For example, one can parameterize a separable state using the NN parameters and then train it to predict the separable state that is closest to a given target entangled state. This turns entanglement detection into a varia-

tional approximation problem and can allow for estimating bounds on entanglement measures[71].

However, for many-body systems, due to the large number of density matrix elements, the full state description is inaccessible, motivating methods that try to detect entanglement directly from undersampled local measurement data. Koutný *et al.*[72] demonstrate on synthetic data that NNs can be trained to quantify the entanglement of an incomplete set of local measurements. A complementary experimental demonstration shows that NN classifiers can be trained directly from laboratory data to identify entanglement without the need to perform full tomography, illustrating an end-to-end approach from measurement to classification[73]. Such approaches are conceptually aligned with the broader trend toward observable-focused characterization (rather than tomography-based) but replace analytical operators with learned estimators.

Beyond bipartite entanglement, ML has also been applied to recognize the multipartite entanglement structure. For example, supervised and semi-supervised learning (including *support vector machine* (SVM)-based approaches) can be utilized to detect genuine multipartite entanglement of three-qubit states[74]. More recent work explores deep NN architectures for distinguishing different classes of entanglement and identifying entanglement in two- and three-qubit systems from state representations[70].

Overall, ML-based methods can be viewed as a data-driven supplement to traditional criteria. They aim to learn effective decision boundaries from labeled or weakly labeled examples, and their practical utility depends on their generalization capability across different state families, robustness to noise, and interpretability in terms of physically meaningful witnesses or features. In the research presented in this dissertation, these ideas are developed further in three complementary directions: by designing NNs to distinguish discord- and entanglement-type quantum correlations, as discussed in Section 5.1; by using Siamese NNs trained in a semi-supervised manner, where entanglement-preserving symmetry transformations are explicitly incorporated to enforce physically motivated invariances, as summarized in Section 5.2; and by training two families of ML models to investigate the importance of tomographic measurements for entanglement detection, as described in Section 5.3.

2.3 Measurements and quantum state tomography

Although the density matrix is a standard way of describing a quantum system, its elements are not directly accessible to the experimentalist. The only observable outcomes of measurements are probabilities associated with projections onto specific states. To reconstruct a density matrix from measurement outcomes, a standard approach is to perform QST. It is specifically useful for extracting complete information about quantum correlations, e.g., in complex photonic systems[75,76].

Projective measurements and Lüders update

A *projection-valued measurement* (PVM) is defined by a set of orthogonal projection operators $\{\Pi_m\}$, where the index m labels the measurement outcomes. Each

projector Π_m satisfies

$$\Pi_m \Pi_{m'} = \delta_{mm'} \Pi_m, \qquad \sum_m \Pi_m = \mathbb{1}, \tag{30}$$

with standard orthogonality criterion denoted by Kronecker delta $\delta_{mm'}$.

Given a quantum state described by the density matrix ρ, the probability $p(m)$ follows the Born rule

$$p(m) = \mathrm{Tr}(\rho \Pi_m)\,. \tag{31}$$

Subsequently, the post-measurement state is given by the Lüders update rule[77]

$$\rho \longmapsto \rho_{|m} = \frac{\Pi_m \rho \Pi_m}{\mathrm{Tr}(\rho \Pi_m)}. \tag{32}$$

Generalized measurements and positive-operator valued measures (POVMs)

More general quantum measurements are described by a collection of measurement (Kraus) operators $\{M_m\}$, each associated with an outcome m. These operators satisfy the completeness relation

$$\sum_m M_m^\dagger M_m = \mathbb{1}, \tag{33}$$

where $M_m^\dagger$ denotes the Hermitian adjoint of M_m. The probability of obtaining outcome m when measuring the state ρ is

$$p(m) = \mathrm{Tr}\big(M_m \rho M_m^\dagger\big) = \mathrm{Tr}(\rho E_m)\,, \tag{34}$$

where the operators

$$E_m \equiv M_m^\dagger M_m, \qquad E_m \succeq 0, \tag{35}$$

define the elements of a *positive-operator valued measure* (POVM). Unlike projectors in a PVM, in general POVM elements are not orthogonal nor idempotent.

The corresponding post-measurement state conditioned on outcome m is

$$\rho \longmapsto \rho_{|m} = \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}\Big(M_m \rho M_m^\dagger\Big)}. \tag{36}$$

This generalized formalism includes PVMs as a special case when $M_m = \Pi_m$.

Naimark dilation and informational completeness

Any generalized measurement described by POVM operators $\{E_m\}$ can be realized as a projective measurement with orthogonal operators $\{\widetilde{\Pi}_m\}$ acting on a larger Hilbert space that includes an auxiliary system. This result, known as *Naimark dilation*, shows that POVMs arise from standard PVMs on an extended space followed by restriction to the system of interest[78].

A measurement scheme is said to be *informationally complete* if the set of outcome probabilities $p(m) = \mathrm{Tr}(\rho E_m)$ uniquely determines the density operator ρ. Informationally complete POVMs provide a natural starting point for tomography, whereas informational incompleteness motivates the use of priors, regularization, and learning-based inference.

Quantum state tomography

QST is the task of reconstructing a quantum state ρ from a finite set of measurement data[7]. It is important to distinguish between the *measurement outcomes* m_ν entering the tomographic reconstruction and the *single measurement events* recorded experimentally. In practice, experiments yield discrete detection events (e.g., photon counts or coincidence counts on detectors), denoted here by n_ν, which are obtained from repeated measurements on identically prepared states. The quantities m_ν used in tomography correspond to normalized frequencies that estimate the Born probabilities,

$$m_\nu \equiv \frac{n_\nu}{N_\nu} \approx p(\nu), \tag{37}$$

where N_ν is the number of measurement repetitions performed for setting ν, and n_ν is the number of events associated with the corresponding outcome. Thus, m_ν are statistical estimators of the Born probabilities, while n_ν represent the underlying raw detection counts[7]. For independent binary measurements, n_ν follows a binomial distribution and the variance of the corresponding frequency estimator is

$$\mathrm{Var}(m_\nu) = \frac{p(\nu)[1-p(\nu)]}{N_\nu}, \tag{38}$$

so that its statistical uncertainty decreases as $1/\sqrt{N_\nu}$ with the number of repetitions.

In the idealized formulation, one considers expectation values of known PVM measurement operators $\{\Pi_\nu\}$ as measurement outcomes m_ν

$$m_\nu = p(\nu) = \mathrm{Tr}(\rho\,\Pi_\nu), \tag{39}$$

and tries to invert the linear map $\rho \mapsto \{m_\nu\}$. Specifically, by choosing an operator basis $\{\Gamma_\mu\}$ for $\mu = \{0, 1, ..., d^2-1\}$, where $d = \dim(\mathcal{H})$, for operators on $\mathcal{H}$ (e.g., Pauli matrices), the density matrix can be expressed as

$$\rho = \sum_\mu r_\mu \Gamma_\mu, \quad \text{where } m_\nu = \sum_\mu B_{\nu\mu} r_\mu, \text{ and } B_{\nu\mu} = \mathrm{Tr}(\Gamma_\mu \Pi_\nu). \tag{40}$$

Here, the operators $\{\Gamma_\mu\}$ are chosen to form a linearly independent basis of the d^2-dimensional operator space. Informational completeness requires the measurement operators $\{\Pi_\nu\}$ to span this operator space, which is equivalent to the measurement matrix B having full column rank,

$$\mathrm{rank}(B) = d^2. \tag{41}$$

In general, an informationally complete measurement set may be overcomplete, in which case B is rectangular and a suitable left inverse, such as the Moore–Penrose pseudoinverse[79], can be used for reconstruction. For a minimal informationally complete set containing exactly d^2 linearly independent measurement operators, B is square and invertible. In this case, tomography reduces to the linear inverse problem

$$\rho = \sum_{\mu,\nu} \Gamma_\mu B^{-1}_{\mu\nu} m_\nu, \tag{42}$$

where $B^{-1}_{\mu\nu}$ denotes the inverse of $B_{\nu\mu}$.

In informationally complete tomography for N qubits ($d = 2^N$), the density matrix contains $d^2 - 1 = 4^N - 1$ independent real parameters after imposing trace normalization. Moreover, since the experimentally accessible quantities m_ν are inferred from a finite number of detection events, they are subject to statistical fluctuations. As a result, naive linear inversion generally produces unphysical estimates (e.g., density matrices with negative eigenvalues). Consequently, modern QST requires a large number of measurement repetitions (samples) and incorporates physical constraints $\rho \succeq 0$ and $\mathrm{Tr}\rho = 1$, thus treating tomography as a statistical inference problem[6,8,7].

Maximum likelihood estimation (MLE)

Maximum likelihood estimation (MLE) is a standard approach in QST for reconstructing a valid density matrix, i.e., preserving hermiticity, positive semidefiniteness, and trace normalization. To enforce these constraints by construction, the d-dimensional density matrix is parametrized as $\rho(\mathbf{t})$, where $\mathbf{t} = (t_1, t_2, \ldots, t_{d^2})$ denotes a set of real parameters used to parametrize a valid quantum state[7]. A common choice is to write

$$\rho(\mathbf{t}) = \frac{T^\dagger(\mathbf{t})T(\mathbf{t})}{\mathrm{Tr}[T^\dagger(\mathbf{t})T(\mathbf{t})]}, \tag{43}$$

where $T(\mathbf{t})$ is a lower-triangular complex matrix whose real diagonal elements and real and imaginary parts of the off-diagonal elements are encoded in the vector $\mathbf{t}$. This form guarantees $\rho(\mathbf{t}) \succeq 0$ and $\mathrm{Tr}\rho(\mathbf{t}) = 1$.

Given measurement operators $\{\Pi_\nu\}$ and the corresponding counts n_ν obtained from N_ν independent repetitions, the model predicts

$$p_\nu(\mathbf{t}) = \mathrm{Tr}[\rho(\mathbf{t})\Pi_\nu]\,. \tag{44}$$

For binary measurements, the appropriate likelihood is the product of binomial probabilities. Omitting combinatorial factors independent of $\mathbf{t}$, it can be written as

$$\mathcal{L}(\mathbf{t}) \propto \prod_\nu p_\nu(\mathbf{t})^{n_\nu} \left[1 - p_\nu(\mathbf{t})\right]^{N_\nu - n_\nu}. \tag{45}$$

Equivalently, one may maximize the log-likelihood function

$$\log \mathcal{L}(\mathbf{t}) = \sum_\nu \left[n_\nu \log p_\nu(\mathbf{t}) + (N_\nu - n_\nu)\log(1 - p_\nu(\mathbf{t}))\right] + \text{const.} \tag{46}$$

Subsequently, using numerical optimization techniques, one finds the optimal parameter set $\mathbf{t}^{(\mathrm{opt})}$ that maximizes the likelihood function. The reconstructed density matrix is then given by $\rho\left(\mathbf{t}^{(\mathrm{opt})}\right)$, obtained by substituting $\mathbf{t}^{(\mathrm{opt})}$ into Eq. 43.

In the asymptotic limit of a large number of measurement samples, MLE estimators are statistically consistent and can be interpreted as minimizing the classical Kullback-Leibler divergence[80] between measured frequencies and model predictions. By explicitly parameterizing only physical states, MLE avoids unphysical reconstructions that may arise from direct linear inversion methods. The cost is a nonlinear optimization problem that may converge to boundary solutions for informationally incomplete or extremely noisy data[6,8,81].

Informationally incomplete tomography

In many experimental scenarios, the available measurement set $\{\Pi_\nu\}$ is not informationally complete, i.e., the number or structure of measurements is insufficient to uniquely determine the density matrix ρ. In such cases, the linear map $\rho \mapsto \{m_\nu\}$ is not invertible, and the reconstruction problem becomes underdetermined, admitting a family of states consistent with the observed data.

A natural extension of linear inversion in this setting is based on the Moore-Penrose pseudoinverse[79]. Writing the measurement model in the operator basis as in Eq. (40), one formally obtains

$$r_\mu = \sum_\nu B^+_{\mu\nu} m_\nu, \tag{47}$$

where B^+ denotes the pseudoinverse of the measurement matrix B. In the case considered here, the selected measurements are linearly independent, so the rows of B are linearly independent, i.e, B has full row rank. Therefore, the Moore-Penrose pseudoinverse can be computed analytically as

$$B^+ = B^T (B B^T)^{-1}, \tag{48}$$

or, more generally, as $B^+ = B^\dagger (B B^\dagger)^{-1}$ for a complex-valued measurement matrix. This construction yields the least-squares solution, i.e., the estimate that minimizes $\sum_\nu |m_\nu - \sum_\mu B_{\nu\mu} r_\mu|^2$.

While computationally efficient, pseudoinverse-based reconstruction suffers from several limitations. In particular, due to statistical noise and incomplete measurement information, the resulting density matrix may violate physical constraints such as positivity ($\rho \not\succeq 0$). Moreover, in the informationally incomplete regime, the pseudoinverse selects a single solution from a degenerate set without incorporating any physical prior, which may lead to reconstructions that are highly sensitive to noise.

These issues motivate the use of constrained and statistical approaches, such as MLE or regularized optimization methods, which incorporate prior information (e.g., positivity or low-rank structure) to select physically meaningful solutions from the set of states compatible with the measured data. In this sense, informationally incomplete tomography naturally leads to a formulation of quantum state reconstruction as an ill-posed inverse problem requiring additional regularization or model assumptions[7,8].

Compressed sensing and low-rank structure

When the quantum state ρ is approximately low rank (e.g., close to a pure state), *compressed sensing* techniques enable accurate reconstruction from substantially fewer measurements than the d^2 parameters required for complete tomography. The central idea is to exploit the low-rank structure by solving a convex optimization problem of the form

$$\min_\rho \ \mathrm{Tr}(\rho) \quad \text{subject to} \quad \sum_\nu |\mathrm{Tr}(\rho\,\Pi_\nu) - m_\nu|^2 \le \epsilon, \qquad \rho \succeq 0, \tag{49}$$

where $\{\Pi_\nu\}$ denote the measurement operators, m_ν are the corresponding experimentally observed measurement outcomes, and ϵ is the assumed threshold that limits the effect of statistical noise and experimental imperfections.

Theoretical and numerical results show that, for suitable measurement settings, e.g., random Pauli measurements, stable recovery is possible with a number of measurements that scale polynomially with the system dimension, rather than with the full Hilbert-space dimension[82]. This reduction in measurement requirements has also been demonstrated experimentally in compressed-sensing tomography of multi-qubit systems[83]. Compressed sensing thus illustrates how restricting state families can dramatically reduce the resources required for scalable QST.

Classical shadows and observable-focused tomography

Instead of focusing on full state reconstruction, the *classical shadow* formalism aims to efficiently predict desired observables. In contrast to standard tomography, one does not first estimate a complete set of probabilities sufficient to reconstruct the full density matrix. Instead, random measurements are performed on individual copies of the quantum state, and each single-shot outcome is classically inverted to produce a *shadow* estimator $\hat{\rho}_{\mathrm{sh}}$. Repeating this procedure many times gives a collection of such estimators, called classical shadows.

The expectation values of observables O are then estimated as averages of $\mathrm{Tr}(O\hat{\rho}_{\mathrm{sh}})$ over the collected shadows. The number of required measurement shots depends on the chosen class of observables, through the corresponding shadow norm, rather than on the full number of density-matrix parameters d^2[84]. This is the main advantage of the method: when one is interested only in selected observables, such as correlation functions, entanglement witnesses, or few-body operators, their expectation values can often be estimated with a number of measurements that scales polynomially with the system size. Classical shadows therefore provide an observable-focused alternative to full QST, especially useful in large systems where reconstructing the entire density matrix is impractical.

Neural network tomography and data-driven inference

Recently, ML-based methods become increasingly popular as an efficient replacement for classical QST. NN-based tomography reconstructs ρ implicitly via a parametrized model trained on measurement data. One important direction uses generative models, such as restricted Boltzmann machines (RBM), to reproduce observed measurement statistics and reconstruct many-body quantum states[18]. Related generative-model approaches have also been developed for quantum state reconstruction and benchmarking beyond the original RBM formulation[85]. Other methods use NNs to recover global quantum states from local or constrained measurement data[86,87], while convolutional NNs can approximate measurement-outcome distributions in a compact variational representation[88]. NN-aided tomography can also demonstrate significantly faster convergence[89,90] than other QST techniques. Moreover, attention-based NNs can be used to effectively denoise measurement data during QST[91].

The use of NNs in tomography is motivated by two related but distinct advantages. First, for noisy but informationally complete data, a NN can learn the geometry of physical density matrices and therefore suppress unphysical reconstructions caused by statistical fluctuations or measurement noise. Second, in informationally incomplete tomography, a NN can learn an effective reconstruction map for a restricted family of states, which can be useful for faster or adaptive reconstruction

when only a subset of measurements is available[87,88].

From a conceptual perspective, neural tomography rephrases QST as a representation-learning problem, i.e., instead of explicitly inverting the measurement procedure, the model learns an effective inverse map constrained by observed data, hence automatically extracting the prior necessary to efficiently scale the reconstruction procedure in higher dimensions. This viewpoint connects tomography to broader themes in inverse problems and motivates hybrid strategies that combine physical constraints with flexible function approximators[18,85,87,88]. In this dissertation, this idea is further developed for informationally incomplete QST by using NNs to reconstruct quantum states from a limited number of measurements, optimize the measurement order, and exploit prior knowledge about restricted state families, in particular X states, with the aim of achieving accurate reconstruction using fewer measurements, as summarized in Section 6.1.

2.4 Synthetic data generation

A central component of the methods developed in this dissertation is the construction of synthetic datasets of quantum states with well-defined correlation properties. While the formal description of quantum states and measurements has been introduced in Sections 2.1-2.3, the practical generation of representative and balanced datasets is a highly nontrivial task, even for systems of a few qubits. In fact, the design of efficient and physically meaningful data-generation procedures constitutes a significant part of the present work.

Generation of pure and mixed random states

The primary approach to constructing datasets relies on random sampling of quantum states using a combination of complementary techniques. Pure states can be generated from the unitarily invariant measure induced by the normalized Haar measure on the unitary group $U(d)$. The Haar measure $d\mu_{\mathrm{H}}(U)$ is the unique normalized measure on $U(d)$ that is invariant under left and right multiplication, i.e.,

$$\int_{U(d)} f(U)\, d\mu_{\mathrm{H}}(U) = \int_{U(d)} f(VU)\, d\mu_{\mathrm{H}}(U) = \int_{U(d)} f(UV)\, d\mu_{\mathrm{H}}(U), \qquad \forall\, V \in U(d), \tag{50}$$

for any integrable function f. This invariance ensures that no basis is preferred and induces a uniform distribution on the space of pure states. In practice, random pure states $|\psi\rangle$ are obtained by applying a Haar-random unitary U to a fixed reference state $|0\rangle$, i.e., $|\psi\rangle = U|0\rangle$, which yields uniform sampling over the complex projective space of pure states[30]. Alternatively, pure states can be generated by constructing parametrized quantum circuits with randomly chosen gates and parameters. The latter approach allows for the controlled introduction of correlations and provides a flexible way of exploring specific regions of the Hilbert space. In the study of correlation identification (Section 5.1-5.3) and quantum state reconstruction (Section 6.1), both approaches were used to generate well-diversified datasets: Haar-random states were used to sample generic pure states, while random quantum circuits containing local single-qubit gates and randomly placed controlled two-qubit gates were used to generate both separable and entangled pure states.

Mixed states can be obtained using several strategies. A direct method consists of forming convex combinations of randomly generated pure states with randomly sampled probability distributions. Alternatively, mixed states can be constructed by embedding the system into a larger Hilbert space and applying a partial trace over auxiliary degrees of freedom, as described in Section 2.1. These procedures yield ensembles with diverse purity and correlation structures, which are essential for training robust ML models. In addition, targeted sampling can be used to enhance the presence of specific classes of states, such as maximally entangled states or states with particular symmetry properties. These strategies were used throughout the entanglement identification research (Section 5.2) to generate mixed separable, mixed entangled, and biseparable states with diverse purity and correlation structure. The resulting datasets were balanced between relevant classes and included separate training, validation, and test subsets, including additional PPTES used to test generalization beyond the states with conclusive PPT labels.

Measurement sampling

In scenarios where only experimentally accessible data are available, the generated density matrices are mapped to measurement outcomes using the tomography framework introduced in Section 2.3. Specifically, expectation values of projective measurements are computed according to the Born rule, yielding vectors that represent realistic experimental observations. In the tomography studies discussed in this dissertation, the respective projectors used for calculating the expectations are taken from a fixed catalog

$$C_N = \left\{ \Pi_\nu = \bigotimes_{k=1}^{N} |\psi_{\nu_k}\rangle \langle \psi_{\nu_k}| \ : \ \nu = (\nu_1, \ldots, \nu_N), \ \nu_k \in \{0,1,2,3\} \right\}. \tag{51}$$

Here, N is the number of qubits, k labels the qubit on which the single-qubit projector acts, and $\nu = (\nu_1, \ldots, \nu_N)$ specifies a choice of single-qubit projectors for all qubits. As a result, the complete catalog C_N contains 4^N product projectors. The single-qubit states are chosen from the standard photonic-tomography states[7]

$$|\psi_{\mu_k}\rangle \in \left\{ |0\rangle, |1\rangle, \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \frac{|0\rangle - i|1\rangle}{\sqrt{2}} \right\}. \tag{52}$$

In complete tomography, all corresponding expectation values can be used, whereas in the informationally incomplete setting only $M < 4^N$ values are provided to the reconstruction algorithm.

This transformation is introduced because several ML models considered in this dissertation are trained not only on density matrices, but also on experimentally motivated measurement representations. In particular, mapping a state ρ to the vector of probabilities $m_\nu = \mathrm{Tr}(\rho\Pi_\nu)$ makes it possible to study how much information about entanglement or the full density matrix can be recovered from a limited set of tomographic measurements, as discussed in Sections 5.3 and 6.1.

Labeling of correlations and entanglement

For supervised and semisupervised learning tasks (see Section 4 for more details of training the ML models), the generated states are labeled according to their corre-

lation properties. Entanglement is typically quantified using computable measures such as negativity or concurrence (Section 2.2), depending on the system size and structure. In practice, care must be taken to avoid ambiguous cases, such as bound entangled states, which cannot be reliably detected by standard criteria and are therefore excluded from certain datasets. Specifically, this problem was addressed during the generation of datasets for entanglement identification, where mixed states with ambiguous PPT-based classification were excluded from the verified training set or treated with a separate labeling strategy.

In addition to entanglement, finer distinctions between correlation types are introduced by identifying discordant and nondiscordant states using operational criteria based on local measurements. This enables the construction of datasets that capture a hierarchy of correlations, ranging from classical to genuinely quantum.

Dataset sizes

The datasets used in the quantum-information part of the dissertation differ depending on the task. In the correlation-identification studies, the models operate mainly on density matrices with known correlation properties. In the tomography-related studies, the density matrices are additionally transformed into vectors of tomographic expectation values. The main dataset sizes are summarized in Table 2.

Table 2: Summary of synthetic datasets used in the quantum-information part of the dissertation. The table reports the main dataset sizes used in the corresponding studies; detailed generation procedures are discussed in the respective synopsis sections and papers.

Study	Main dataset sizes
Section 5.1 *Data-driven criteria for quantum correlations*	**Training:** 530 000 pure and mixed separable states, including nondiscordant states. **Validation:** 50 000 states. **Test:** 30 000 pure states and 65 000 mixed states.
Section 5.2 *Identification of quantum entanglement...*	**Training:** approximately 560 000 states in each main training dataset. **Validation:** 56 000 states and 10 000 PPTES states. **Test:** 30 000 pure states, 40 000 mixed states, and three PPTES test sets of 10 000 states each.
Section 5.3 *Learning entanglement from tomography data....*	**Training:** 460 000 two-qubit density matrices transformed into tomographic measurement vectors. **Validation:** not separately specified. **Test:** 46 000 states.
Section 6.1 *Learning quantum tomography from incomplete measurements*	**Training:** 460 000 density matrices for each of the 2-, 3-, and 4-qubit experiments. **Validation:** not separately specified. **Test:** 46 000 density matrices for each of the 2-, 3-, and 4-qubit experiments.

Overall, the combination of multiple generation techniques, tomographic repre-

sentations, and labeling procedures is necessary to obtain datasets that are both physically meaningful and suitable for ML applications. The complexity of this process highlights that even for few-qubit systems, generating high-quality data is a challenging task that requires careful methodological design rather than straightforward random sampling.

3 Topological Quantum Computing

A central challenge in contemporary quantum computing is the extreme sensitivity of quantum states to environmental noise and decoherence, which leads to significant error rates in current hardware platforms such as superconducting circuits or trapped ions.

In state-of-the-art NISQ devices, the relevant physical error rates are strongly platform- and operation-dependent, but they are typically discussed in the range of approximately 10^{-3}–10^{-2} per operation, i.e., close to the threshold values relevant for the surface code, a leading quantum error-correction architecture based on encoding logical qubits in a two-dimensional array of nearest-neighbor-coupled physical qubits and detecting errors through repeated local stabilizer measurements[3,92,93]. Although such error correction enables fault-tolerant quantum computation in principle, it also introduces a substantial hardware overhead. For example, surface-code estimates indicate that even for physical error rates around 10^{-3}, reaching algorithmically useful logical error rates may require on the order of 10^3–10^4 physical qubits per logical qubit, depending on the target error rate and architecture[93,94].

TQC offers a fundamentally different approach to this problem by encoding quantum information in nonlocal, topologically protected degrees of freedom. In such systems, quantum states are intrinsically robust against local perturbations, leading to a natural suppression of decoherence and operational errors[95]. This built-in protection has the potential to significantly reduce the overhead required for fault-tolerant computation.

Recent experimental and industrial efforts, including the announcement of the *Majorana 1* processor by Microsoft and the related report of interferometric single-shot parity measurements in InAs–Al hybrid devices, have renewed interest in this paradigm[96]. At the same time, the interpretation of such experiments remains under active discussion, since trivial low-energy states and disorder effects can mimic signatures expected from topological superconductivity[97]. Therefore, these developments should be viewed as an important motivation for further studies of Majorana-based platforms, rather than as a completed demonstration of scalable topological quantum computation. In particular, topological qubits are expected to exhibit substantially lower error rates than conventional qubit implementations if the required nonlocal encoding and topological protection are experimentally realized.

This section introduces the theoretical background needed to understand the topological computing aspect of the dissertation. First, the basic concepts of topological superconductivity and MZMs are discussed, including the Bogoliubov-de Gennes (BdG) formalism, particle-hole symmetry, Majorana operators, and topological invariants. Next, QD chains are presented as highly tunable solid-state systems capable of emulating Kitaev-like physics and providing experimentally controllable platforms for studying Majorana modes. Finally, the section describes how synthetic data for topological systems can be generated from model Hamiltonians and used in ML approaches to Hamiltonian learning, phase identification, and device tuning.

3.1 Topological superconductivity and Majorana zero modes

MZMs are zero-energy quasiparticle excitations that can emerge at boundaries, interfaces, or topological defects of superconductors with nontrivial band topology[21]. MZMs appear in spatially separated pairs whose energy remains pinned to zero in the ideal topological limit, with only exponentially small splitting expected in finite systems. They are protected against local perturbations that do not close the superconducting gap or alter the system's topological class. This property underlies the central interest in Majorana modes for quantum computing: they provide a physical platform for topological qubits whose coherence is protected by global properties of the many-body wavefunction rather than fine-tuned microscopic control. Moreover, the exchange (braiding) of MZMs is predicted to implement non-Abelian unitary transformations, enabling quantum gates that depend only on the topology of the braiding paths and are therefore robust to control imperfections[98]. These features position topological superconductivity and MZMs as a promising route toward fault-tolerant solid-state quantum computation[99,100,101]

At the same time, identifying, characterizing, and controlling Majorana modes in realistic materials and device architectures remains a major challenge, due to the complexity of experimental signatures and the presence of competing trivial states[102,4,101,103]. In particular, zero-bias conductance peaks and near-zero-energy bound states can also arise from disorder, smooth confinement, or trivial Andreev bound states (ABSs), which may mimic some expected signatures of MZMs[104,105]. This difficulty was highlighted by the high-profile retraction of the 2018 *Nature* report on quantized Majorana conductance in semiconductor–superconductor nanowires, after reanalysis showed that the original data analysis did not provide sufficiently robust evidence for the claimed topological interpretation[106,107]. This context illustrates why Majorana identification requires stringent diagnostics beyond a single local transport signature. It also motivates the use of data-driven and ML approaches, which offer new tools for exploring high-dimensional parameter spaces, classifying topological phases, and distinguishing genuine Majorana physics from non-topological alternatives.

BdG Hamiltonians and symmetry constraints

A convenient theoretical framework for describing superconducting systems and MZMs is provided by the BdG formalism[98]. In this representation, one introduces fermionic *creation* $c_i^\dagger$ and *annihilation* c_i operators, with $i = 1, \ldots, M$ enumerating the single-particle degrees of freedom of the model, e.g., lattice sites, spin states, orbitals, or their combinations. These operators are then collected into the Nambu spinor $\Psi = (c_1, \ldots, c_M, c_1^\dagger, \ldots, c_M^\dagger)^T$, which explicitly encodes particle-hole redundancy. In this representation, any quadratic fermionic Hamiltonian can be written as

$$H = \frac{1}{2}\Psi^\dagger \mathcal{H}_{\mathrm{BdG}} \Psi + \text{const.}, \tag{53}$$

where $\mathcal{H}_{\mathrm{BdG}}$ is a $2M \times 2M$ Hermitian matrix.

The doubled Nambu description implies intrinsic *particle-hole symmetry* (PHS), reflecting the fact that physical quasiparticle excitations are superpositions of electrons and holes. Assuming Pauli matrices $\{\tau_x, \tau_y, \tau_z\}$ act in particle-hole space, the

PHS operator may be chosen as

$$\mathcal{C} = \tau_x \mathcal{K}, \tag{54}$$

with $\mathcal{K}$ denoting complex conjugation. This symmetry enforces the constraint

$$\mathcal{C}\,\mathcal{H}_{\mathrm{BdG}}\,\mathcal{C}^{-1} = -\mathcal{H}_{\mathrm{BdG}}, \tag{55}$$

which, in translationally invariant systems, typically described in momentum k space, takes the form

$$\mathcal{H}_{\mathrm{BdG}}(k) = -\tau_x \mathcal{H}^T_{\mathrm{BdG}}(-k)\,\tau_x. \tag{56}$$

As a direct consequence, the eigenspectrum is symmetric about zero energy, i.e., both E and $-E$ are valid eigenvalues.

Depending on the physical system, additional symmetries may further constrain $\mathcal{H}_{\mathrm{BdG}}$. If *time-reversal symmetry* (TRS) is present, there exists an antiunitary operator $\mathcal{T}$ that satisfies

$$\mathcal{T}\,\mathcal{H}_{\mathrm{BdG}}(k)\,\mathcal{T}^{-1} = \mathcal{H}_{\mathrm{BdG}}(-k), \qquad \mathcal{T}^2 = \pm 1. \tag{57}$$

When both PHS and TRS are present, their product yields a *chiral symmetry*. Together, these symmetry constraints place the BdG Hamiltonian into an Altland–Zirnbauer (AZ) symmetry class, which determines the topological invariants and the robustness of zero-energy boundary modes. The minimal models relevant for Majorana physics typically break TRS through Zeeman splitting and belong to symmetry class *D*, characterized by a $\mathbb{Z}_2$ invariant in one dimension[102,103]. In the presence of an additional chiral symmetry, an effective BDI class (in AZ typology) description may emerge, supporting an integer winding number and multiple zero modes[99,108], which are discussed later in this section.

Majorana quasiparticles

The operator structure of MZMs can be obtained directly from the BdG formalism introduced before. Considering a normalized zero-energy eigenvector Φ_0 of the BdG Hamiltonian,

$$\mathcal{H}_{\mathrm{BdG}}\Phi_0 = 0, \tag{58}$$

written in the Nambu basis as $\Phi_0 = (u_1, \ldots, u_M, v_1, \ldots, v_M)^T$, one can define the quasiparticle operator using fermionic operators $\{c_i, c_i^\dagger\}$,

$$\Gamma_0 = \sum_{i=1}^{M} \left(u_i c_i + v_i c_i^\dagger \right), \tag{59}$$

where i labels the single-particle degrees of freedom included in the BdG basis. PHS can be represented by the antiunitary operator $\mathcal{C} = \tau_x \mathcal{K}$, where τ_x acts in particle-hole space and $\mathcal{K}$ denotes complex conjugation. Since

$$\mathcal{C}\mathcal{H}_{\mathrm{BdG}}\mathcal{C}^{-1} = -\mathcal{H}_{\mathrm{BdG}}, \tag{60}$$

if Φ_E is an eigenstate with energy E, then $\mathcal{C}\Phi_E$ is an eigenstate with energy $-E$. For a zero-energy state one can therefore choose a Majorana basis satisfying $\mathcal{C}\Phi_0 = \Phi_0$, which gives $v_i = u_i^*$. Substituting this relation into Γ_0 yields $\Gamma_0 = \Gamma_0^\dagger$. In

this case, the quasiparticle operator corresponds to a Majorana operator γ. Thus, within the BdG framework, a MZM arises as a zero-energy solution whose associated quasiparticle operator is invariant under Hermitian conjugation [21,99,102].

The Majorana representation can also be introduced algebraically for any ordinary complex fermionic mode f, which can be decomposed into two Hermitian Majorana operators,

$$\gamma_1 = f + f^\dagger, \qquad \gamma_2 = -i(f - f^\dagger), \tag{61}$$

which obey $\gamma_i = \gamma_i^\dagger$ and $\{\gamma_i, \gamma_j\} = 2\delta_{ij}$. Conversely, γ_1 and γ_2 can be recombined into a standard complex fermion operator,

$$f = \frac{\gamma_1 + i\gamma_2}{2}, \qquad f^\dagger = \frac{\gamma_1 - i\gamma_2}{2}, \tag{62}$$

with occupation number

$$n_f = f^\dagger f = \frac{1 + i\gamma_1\gamma_2}{2}. \tag{63}$$

Equivalently, the fermion parity of this mode is

$$P_f = (-1)^{n_f} = 1 - 2n_f = -i\gamma_1\gamma_2. \tag{64}$$

In a topological superconductor, the operators γ_1 and γ_2 forming such a fermionic mode may be spatially separated and localized near opposite boundaries or defects. The resulting fermionic mode f is then nonlocal, and its occupation determines the joint fermion parity of the pair. Since fermion parity is conserved in an isolated superconducting system, local parity-preserving perturbations cannot change this occupation unless they effectively couple both Majoranas. As a result, local perturbations cannot efficiently change the Majorana state or cause decoherence as long as the modes remain well separated and the superconducting gap stays open. In this way, the BdG zero-mode structure provides the origin of the nonlocal encoding and robustness characteristic of Majorana-based qubits [21,99,102,4].

Another interesting property of MZMs is observed when exchanging, or braiding, two well-separated Majoranas γ_a and γ_b. Specifically, the braiding unitary operator can be written as

$$U_{ab} = \exp\left(\frac{\pi}{4}\gamma_a\gamma_b\right), \tag{65}$$

which implements

$$U_{ab}\gamma_a U_{ab}^\dagger = -\gamma_b, \qquad U_{ab}\gamma_b U_{ab}^\dagger = \gamma_a. \tag{66}$$

Because different braids generally do not commute, MZMs realize non-Abelian statistics. This property is the foundation of topological quantum computation, where gates are implemented by braiding operators, which are insensitive to small deformations of the trajectories [21,99,95].

Importantly, topological protection allows for mitigating the effect of *local perturbations* that do not close the gap and do not break fermion parity. In realistic devices, the main limitations arise from finite-size hybridization of Majoranas, disorder-induced trivial near-zero ABSs, and parity-breaking processes such as quasiparticle poisoning, i.e., unwanted changes of fermion parity caused by nonequilibrium quasiparticles entering or leaving the superconducting system [109]. Consequently, experiments aim not only at the emergence of stable near-zero modes, but also at demonstrating their nonlocal character and non-Abelian braiding properties [4,103].

Topological invariants in one-dimensional superconductors

The existence of MZMs at the boundaries of one-dimensional superconductors is predicted by nontrivial bulk topological invariants protected by certain symmetry classes. These invariants characterize the global properties of the BdG Hamiltonian and cannot change without closing the bulk energy gap.

Class D: $\mathbb{Z}_2$ invariant. For one-dimensional superconductors in symmetry class D (i.e. with PHS and broken TRS), the $\mathbb{Z}_2$ invariant can be obtained directly from the BdG Hamiltonian using the PHS convention introduced before. Defining

$$B(k) = \mathcal{H}_{\mathrm{BdG}}(k)\tau_x, \tag{67}$$

Eq. 56 implies

$$B^T(k) = -B(-k). \tag{68}$$

Therefore, at the particle-hole invariant momenta satisfying $k = -k$ modulo a reciprocal lattice vector, i.e., $k = 0, \pi$ for a one-dimensional lattice with unit lattice spacing, $B(k)$ is antisymmetric and its Pfaffian is well defined. In the fixed Nambu-basis convention used here, the class-D invariant can be written as

$$\nu_{\mathrm{D}} = \mathrm{sgn}[\mathrm{Pf}\, B(0)\, \mathrm{Pf}\, B(\pi)]\,. \tag{69}$$

With this convention, $\nu_{\mathrm{D}} = -1$ corresponds to the topologically nontrivial phase, whereas $\nu_{\mathrm{D}} = +1$ corresponds to the trivial phase[22,98]. A change of the invariant requires closing the bulk excitation gap.

Class BDI: winding number. In the presence of an additional chiral symmetry, corresponding to the one-dimensional BDI symmetry class, the BdG Hamiltonian can be represented in a block off-diagonal form,

$$\mathcal{H}_{\mathrm{BdG}}(k) = \begin{pmatrix} 0 & q(k) \\ q^\dagger(k) & 0 \end{pmatrix}, \tag{70}$$

where, after spectral flattening, $q(k)$ can be taken to be unitary for gapped systems. In a one-dimensional superconductor, the corresponding integer invariant, called the winding number, is defined as[98]

$$\nu_{\mathrm{BDI}} = \frac{i}{2\pi} \int_{-\pi}^{\pi} dk\, \mathrm{Tr}\left[q^{-1}(k) \frac{d}{dk} q(k) \right]. \tag{71}$$

The integer ν_{BDI} counts the number of protected zero-energy edge states in the ideal chiral-symmetric limit. If perturbations break the chiral symmetry, the system is no longer described by the BDI class, but by class D. In such a case, only the parity of the winding number remains topologically protected and corresponds to the $\mathbb{Z}_2$ invariant. Therefore, in realistic finite devices, one often focuses on the even-odd parity of Majorana modes, while the winding number remains useful for understanding idealized symmetry-protected multiple zero modes.

Bulk-boundary correspondence. The central principle connecting these invariants to observable physics is the bulk-boundary correspondence. A nontrivial topological invariant value implies the existence of zero-energy states localized at the edges or defects of a system. This can be understood by considering a topological material ending at a boundary or meeting a trivial medium, such as vacuum. Across such an interface, the topological invariant must change from a nontrivial value in the bulk to a trivial value outside the system. Since the invariant cannot change continuously while the bulk remains gapped, protected boundary states appear at the interface. In one-dimensional topological superconductors, this correspondence underlies the appearance of MZMs at chain edges or domain walls. Importantly, the protection holds as long as the bulk gap remains open and the relevant symmetries are not violated by strong perturbations.

Andreev bound states versus Majorana zero modes

In hybrid superconductor-semiconductor systems, zero modes are not unique to topological phases. A broad class of localized ABSs can arise from multiple Andreev reflections at metal-superconductor interfaces, where electrons are reflected as holes and holes as electrons, leading to constructive interference and the formation of discrete subgap states. ABS can also arise from confinement effects or spatially varying potentials. ABSs are trivial solutions of the BdG equations with energies inside the superconducting gap and are typically localized at one of the edges or interfaces. Because these states remain localized in one region, they lack the nonlocal characteristic of topological Majoranas, which are spatially separated at distant boundaries. A key experimental problem is that certain ABS can be tuned close to zero energy over extended parameter ranges, producing measured zero-bias peaks (ZBPs) that closely resemble those expected from MZMs[108,103].

To distinguish ABS from MZMs, several criteria have been proposed:

1. **Nonlocal correlations.** Correlated responses measured at opposite ends of a chain aim to demonstrate that low-energy modes are spatially separated (MZMs) rather than localized in one region (ABS). Such measurements are experimentally demanding but are among the most promising routes to rule out trivial explanations[103].

2. **Length dependence.** The overlap amplitude of the MZMs decays exponentially with the length of the superconducting nanowire segment, reflecting the exponential localization of the Majorana wavefunctions[110,111]. By contrast, trivial ABS originate from locally overlapping components and do not exhibit such behavior. The major limitation of this method is that it requires the fabrication of multiple devices or dynamically tunable chain lengths.

3. **Parity structure.** In Coulomb-blockaded superconducting islands, where charge fluctuations are suppressed by a finite charging energy, MZMs produce characteristic even-odd effects depending on nonlocal fermion parity, such as parity-dependent ground-state crossings or changes in Coulomb blockade peak spacings[112]. ABS generally correspond to local quasiparticles with no protected nonlocal parity structure. However, similar signatures can also arise from trivial states in nonuniform systems, including partially overlapping ABSs and disorder-induced low-energy modes[112,113,103]. These effects can

produce near-zero-energy excitations and parity signatures that mimic those expected from topological Majorana modes, complicating their unambiguous identification.

4. **Response to perturbations.** MZMs are robust against moderate local perturbations that preserve the bulk gap and fermion parity, whereas ABS energies are typically sensitive to local gate variations.

In practice, realistic disorder, finite-size effects, and smooth potential profiles further blur these distinctions. As emphasized in recent works, no single observable provides a definitive diagnosis, and correct identification of topological Majorana modes requires consistency across multiple probes combined with detailed theoretical modeling[4,103,108]. This motivates the use of problem-specific diagnostic quantities in this dissertation. In particular, in Section 5.4, we introduce a differentiable quasimetric $\mathcal{M}$ that combines zero-energy spectral weight, edge localization, and electron-hole symmetry into a single physics-informed objective. The metric is designed not only to favor Majorana-like edge modes, but also to suppress trivial zero modes such as ABSs, which can otherwise mimic selected Majorana signatures.

Kitaev chain as the canonical 1D model

The simplest theoretical one-dimensional model that realizes MZMs is the *Kitaev chain* (KC)[22]. It describes a spinless chain with nearest-neighbor hopping and superconducting pairing,

$$H_{\mathrm{KC}} = -\mu \sum_n c_n^\dagger c_n - \sum_n \left(t\, c_n^\dagger c_{n+1} + \Delta\, c_n c_{n+1} + \mathrm{h.c.} \right), \tag{72}$$

where μ is the chemical potential, t is the hopping amplitude, and Δ is the superconducting pairing amplitude. Importantly, the pairing term couples neighboring sites and therefore represents an effective spinless p-wave pairing. This feature is essential because the superconductors most commonly used in experiments are conventional s-wave superconductors, in which electrons form Cooper pairs with opposite spins on the same site or in the same local orbital state. Such pairing is much easier to realize experimentally, but it does not by itself produce the spinless nearest-neighbor pairing required in the Kitaev chain model. Therefore, in realistic semiconductor-superconductor devices, the effective p-wave description has to be engineered from induced s-wave pairing using additional ingredients such as spin-orbit coupling and Zeeman splitting.

Using the Majorana representation introduced above, one can define two Majorana operators on each site,

$$\gamma_{A,n} = c_n + c_n^\dagger, \qquad \gamma_{B,n} = -i(c_n - c_n^\dagger), \tag{73}$$

which satisfy $\gamma_{\alpha,n} = \gamma_{\alpha,n}^\dagger$. In this basis, the Kitaev chain Hamiltonian can be written, up to an irrelevant constant, as

$$H_{\mathrm{KC}} = -\frac{i\mu}{2} \sum_n \gamma_{A,n}\gamma_{B,n} - \frac{i}{2} \sum_n \left[(t+\Delta)\gamma_{B,n}\gamma_{A,n+1} + (\Delta - t)\gamma_{A,n}\gamma_{B,n+1} \right]. \tag{74}$$

This form gives a simple physical picture of the topological phase. In the trivial limit, the dominant coupling pairs the two Majoranas belonging to the same lattice site. By contrast, at the representative topological point $\mu = 0$ and $t = \Delta$, Eq. (74) reduces to

$$H_{\mathrm{KC}} = -it \sum_n \gamma_{B,n}\gamma_{A,n+1}, \tag{75}$$

so Majorana operators on neighboring sites are paired instead. In a finite open chain, this leaves two unpaired Majorana operators, $\gamma_{A,1}$ and $\gamma_{B,N}$, localized at opposite ends of the system. These unpaired edge operators form the MZMs.

More generally, for $\Delta \neq 0$, the KC is in the topological phase when $|\mu| < 2|t|$, while the phase is trivial for $|\mu| > 2|t|$. This condition is equivalent to obtaining the nontrivial value of the one-dimensional class-D $\mathbb{Z}_2$ invariant, Eq. (69), i.e., $\nu_{\mathrm{D}} = -1$ [22,102]. The KC therefore provides the canonical minimal model explaining why one-dimensional topological superconductors support MZMs at their boundaries.

Experimental pursuit of MZMs

While the theoretical model predicts the existence of MZMs, its experimental realization remains an unsolved challenge. A central experimental method for verifying the emergence of MZMs is tunneling spectroscopy. In the simplest picture, a localized MZM yields a conductance ZBP. Early reports of ZBPs suggested the presence of Majorana fermions in hybrid nanowires [114] or in proximity-induced superconducting Coulomb islands, i.e., so-called Majorana islands [111]. At the same time, it became clear that ZBPs are not unique signatures: previously mentioned ABS, disorder, and confinement effects can produce robust near-zero modes without topological protection, leading to a nontrivial discrimination problem [108,103]. A particularly strong example was the 2018 report of quantized zero-bias conductance in epitaxial hybrid nanowires by the Kouwenhoven group [106], which was later formally retracted [107]. This episode reinforced the need for cross-validated diagnostics and demonstrations that go beyond local tunneling spectra.

Current efforts, therefore, emphasize complementary signatures of topology and nonlocality, including correlated measurements at opposite ends of a wire, fusion-rule experiments that probe the fermion parity structure of paired Majoranas, and branched nanowire or Josephson-junction-network protocols designed to realize controlled braiding operations [115,116]. From the perspective of TQC, specifically important is the preservation of fermion parity over timescales long compared to gate operations, since quasiparticle poisoning, i.e., unwanted tunneling of nonequilibrium quasiparticles into or out of the superconducting system, can change this parity and corrupt the encoded information [109,117]. Recent Microsoft experiments also focus on this parity-based description, reporting single-shot parity readout in hybrid InAs-Al devices and minimal Kitaev chain structures as a step toward measurement-based control of Majorana qubits [96,118]. At the same time, the topological interpretation of transport-based diagnostics remains under active discussion [97]. Although some progress has been made in extending parity lifetimes in hybrid superconducting devices, maintaining coherence in scalable architectures remains an active area of research [4].

One major platform for realizing MZMs consists of spin-orbit-coupled semiconductors proximitized by an s-wave superconductor and subjected to Zeeman splitting [119,120]. MZMs are also actively pursued in other platforms, including chains

of magnetic atoms[121,122], engineered Josephson-junction networks[123], and vortex-core states in superconductors proximitized to topological surface bands[124], with increasing exploration of candidate MZMs in iron-based superconductors[125]. Further works indicate that, through appropriate tuning, it is possible to realize a KC using a chain of coupled QDs with either s-wave superconductors[126,127,128,129] or p-wave superconductors[130,131]. The diversification of platforms is driven by the same overarching goals: larger topological gaps, cleaner spectra, better control of disorder and quasiparticle poisoning, and ultimately device geometries that enable demonstrable non-Abelian operations[21,4,103].

3.2 Quantum dot chains as tunable realizations of Kitaev-like physics

QDs are nanoscale electronic systems in which charge carriers are confined in all spatial directions, resulting in discrete energy levels and strong electron-electron interactions. Their energy spectrum, occupation, and tunnel couplings can be precisely controlled using electrostatic gates, which makes them a versatile platform for a wide range of applications, including photovoltaics (QD solar cells), optoelectronics (QD light-emitting diodes), quantum cryptography (quantum key distribution), or quantum computing[132,133]. Specifically, when coupled to superconducting reservoirs, QDs exhibit proximity-induced pairing and can host Majorana modes in appropriately engineered structures[126,134,108,135]. The high degree of electrostatic tunability in QD chains and the possibility of controlling individual sites distinguish QD arrays from continuous nanowire systems and enable programmable implementations of TQC.

The minimal microscopic description of a QD coupled to metallic or superconducting leads is provided by the Anderson impurity model[136]. In its most common form, the Hamiltonian is written as

$$H = \underbrace{\sum_{k,\sigma} \epsilon_k \, c_{k\sigma}^\dagger c_{k\sigma}}_{H_{\text{leads}}} + \underbrace{\sum_{\sigma} \epsilon_d \, d_\sigma^\dagger d_\sigma + U \, d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow}_{H_{\text{dot}}} + \underbrace{\sum_{k,\sigma} \left(V_k \, c_{k\sigma}^\dagger d_\sigma + V_k^* \, d_\sigma^\dagger c_{k\sigma} \right)}_{H_{\text{T}}} . \quad (76)$$

Here, the operators $d_\sigma^\dagger$ (d_σ) create (annihilate) an electron with spin $\sigma \in \{\uparrow, \downarrow\}$ in the localized orbital of the QD, while $c_{k\sigma}^\dagger$ ($c_{k\sigma}$) create (annihilate) an electron in an extended single-particle state of the leads characterized by quantum number k, which labels the single-particle eigenstates of the leads and spin σ. The first term H_{leads} describes the electron conduction in the leads, with dispersion relation ϵ_k. The second term H_{dot} denotes the dot (site)-level physics with parameter ϵ_d representing the energy of the dot level, which can be tuned experimentally using gate voltages, and U is the on-site Coulomb interaction energy associated with double occupation of the dot. The last term H_{T} represents tunneling between the dot and the leads, where V_k is the hybridization matrix element governing the amplitude for electron transfer between the localized dot state and the lead state k.

This Hamiltonian captures the essential physics of a localized electronic level coupled to external reservoirs, including Coulomb blockade and Kondo correlations[136,137]. When the leads are superconducting, the hybridization induces pairing correlations on the dot, giving rise to subgap quasiparticle states and enabling

the realization of Majorana modes in coupled QD systems under appropriate conditions[126,134,108].

From dot arrays to effective Kitaev-like Hamiltonians

When each dot in the QD chain is proximitized by a superconductor and subjected to spin-orbit coupling and an external magnetic field, the dot array can act as a tunable quantum simulator of the KC[135,127]. The respective Hamiltonian for a chain of spinful single-level QDs can be expressed using dot fermion creation ($d^\dagger$) and annihilation (d) operators as

$$H = \sum_{n,s,s'} \Big[(-\mu_n \sigma_0 + V_Z \sigma_z)_{ss'} \, d^\dagger_{n,s} d_{n,s'} + \tfrac{1}{2} \Big(\big(\Delta_n e^{i\phi_n} i\sigma_y\big)_{ss'} \, d^\dagger_{n,s} d^\dagger_{n,s'} + \text{h.c.} \Big) + \big(t_n (e^{i\boldsymbol{\lambda}_n \cdot \boldsymbol{\sigma}})_{ss'} d^\dagger_{n,s} d_{n+1,s'} + \text{h.c.} \big) \Big], \tag{77}$$

with on-site potentials μ_n, inter-dot hopping amplitudes t_n modulated by the Rashba spin-orbit vector $\boldsymbol{\lambda}_n$, Zeeman energy V_Z, and proximity-induced superconducting (s-wave) pairing $\Delta_n e^{i\phi_n}$, where n describes the dot number, s denotes the spin degree of freedom and σ_i refers to the respective Pauli matrix.

The connection to the KC emerges in the low-energy regime where the Zeeman field and spin-orbit coupling effectively select a single helical, spin-nondegenerate band. Projecting the induced s-wave pairing onto this helical subspace generates an effective nearest-neighbour p-wave pairing term, while the ordinary inter-dot tunnelling gives the hopping term of the Kitaev model[127,135]. In this way, the gate-defined parameters of the QD chain provide direct control over the effective Kitaev chain parameters: the chemical potential, hopping amplitude and induced pairing.

A major advantage of such an approach is the ability to control the relevant system parameters using gate voltages and superconducting phases. Local tuning of parameters toward "sweet spots", where the effective hopping and pairing amplitudes are balanced and the on-site energies are compensated, can stabilize a robust topological regime with Majorana quasiparticles localized near the edges of the chain[127,108]. In principle, disorder effects, common to continuous nanowires, can be mitigated by individually compensating on-site energies and tunneling couplings[138]. As a result, using the QD chain allows for systematic exploration of parameter space and the realization of minimal topological models with a small number of sites.

Experimental realization of controllable QD chains

Although QD chains provide a highly tunable platform for simulating Kitaev-like topological superconducting Hamiltonians, their experimental implementation is complicated by several practical limitations. A primary challenge arises from the large number of electrostatic control parameters. Each QD typically requires multiple gate voltages to tune its chemical potential, tunneling couplings between neighboring dots, and superconducting phases that describe the superconductor proximity effect. Even short chains are characterized by a high-dimensional parameter space with mutual dependencies between control parameters, making manual tuning and stabilization of desired regimes experimentally demanding[139,140].

Another important limitation is the presence of cross-capacitance between neighboring gates, which leads to nonlocal electrostatic effects. Changes in a given gate

voltage can modify not only the target dot but also adjacent sites and tunnel barriers, requiring careful compensation and iterative calibration procedures to maintain independent control over the system[141,140]. In addition, achieving an effective lattice Hamiltonian requires precise alignment of the dot energy levels relative to the induced superconducting gap, as small deviations can lead to the emergence of the previously mentioned ABS[103,108].

Further complications arise from spatial inhomogeneity in the superconducting proximity effect, due to differences in tunneling coupling to the superconductor and local electrostatic variations, thereby reducing the uniformity required for the realization of robust topological phases[142,143]. In addition, quasiparticle poisoning and thermal effects impose fundamental limits on the stability of parity-dependent states in superconducting QD systems. Specifically, such effects can induce stochastic changes in fermion parity and reduce the coherence time of desired modes[117,111]. These processes are particularly relevant in multi-dot structures, where parity conservation across the entire system is required to maintain the integrity of nonlocal Majorana modes.

Nevertheless, recent experiments have demonstrated minimal realizations of Kitaev-like chains using coupled QDs. This direction is often referred to as the *poor man's Majorana* approach, originally proposed as a way to engineer Majorana physics in a small number of coupled QDs rather than in an extended topological superconducting material[126]. A two-site implementation showed the emergence of near-zero energy modes consistent with the simplest topological model[128]. Extending the chain to three sites leads to enhanced stability and increased localization of zero-energy modes, in qualitative agreement with expectations of longer KCs[129].

These works demonstrate the potential of QD arrays as a viable platform for engineered topological superconductivity, while also highlighting the challenges that make manual tuning increasingly impractical as system size grows, hence highlighting the need for automated approaches to device calibration, including ML-assisted optimization of multi-parameter quantum devices.

Measurements in QD chains

The common measurements performed on the QD system, used to provide an indirect probe of the low-energy spectrum and, in particular, zero modes, are electronic transport measurements via the conductance maps. In hybrid dot-superconductor devices, charge transfer involves both conventional electron tunneling and Andreev reflection processes, which are naturally described within the scattering formalism. Specifically, conductance maps can be obtained from the scattering matrix $S(E)$, which relates the incoming and outgoing quasiparticle amplitudes at energy E[144]. In the wide-band limit, the scattering matrix can be computed directly from the BdG Hamiltonian H using the Weidenmüller formula[145,126,129], which expresses the scattering matrix as

$$S(E) = \mathbf{1} - iW^\dagger \left(E - H + \frac{i}{2} W W^\dagger \right)^{-1} W, \tag{78}$$

where W is the coupling matrix that describes the tunneling between the system and the leads, and its structure encodes both the dot-lead connections and the coupling strengths. This is closely related to the famous Meir-Wingreen formula[146] for

quantum transport, which is formulated in terms of retarded and advanced Green's functions. The Hamiltonian H and coupling matrix W are defined in the full Nambu basis, including site, spin, and particle-hole degrees of freedom. The scattering matrix naturally decomposes into electron (e) and hole (h) sectors,

$$S(E) = \begin{pmatrix} s^{ee}(E) & s^{eh}(E) \\ s^{he}(E) & s^{hh}(E) \end{pmatrix}, \tag{79}$$

where the subblocks $s^{ee(hh)}$ and $s^{eh(he)}$ describe normal and Andreev reflection processes, respectively. The differential conductance between leads i and j is obtained from the corresponding reflection amplitudes as

$$G_{ij}(E) = \frac{e^2}{h}\left[2\delta_{ij} - \mathrm{Tr}\left(|s_{ij}^{ee}(E)|^2\right) + \mathrm{Tr}\left(|s_{ij}^{he}(E)|^2\right)\right], \tag{80}$$

where the trace is taken over the spin and orbital indices, and δ_{ij} ensures current conservation.

In practice, the full conductance tensor G_{ij} can be evaluated for arbitrary multi-dot geometries and lead configurations by constructing the BdG Hamiltonian and coupling matrix and applying Eq. (78). Conductance maps are then obtained by evaluating $G_{ij}(E)$ as a function of energy and external control parameters such as gate voltages or magnetic field. Because the scattering matrix depends nonlinearly on the underlying Hamiltonian, the experimentally measured conductance provides only indirect information about microscopic parameters. This inverse relationship complicates both the interpretation of experimental data and the systematic tuning of device parameters, particularly in systems with many coupled QDs and superconducting correlations.

Zero-bias conductance peaks (ZBPs)

To assess the emergence of MZMs, one can commonly use the near-ZBPs of conductance as a direct indicator for nanowire experiments. Its strengths are experimental simplicity and sensitivity to low-energy states. However, ZBPs are not unique to topological MZMs and can arise from the previously mentioned ABS.

Hence, to distinguish between MZMs and ABS, further measurements based on parity structure, length dependence, or nonlocal correlations must be performed. Unfortunately, as mentioned before, no single metric is conclusive on its own. Robust identification of MZMs relies on consistency between multiple indicators combined with theoretical modeling that accounts for disorder and finite-size effects. That is why recent works focus on applying modern ML methods to simplify the detection and creation of MZMs.

Automation, machine learning, and data-driven control of QD devices

Early demonstrations showed that automated algorithms can reliably tune a two-dot system to target regimes, such as single-electron occupation and desired tunnel couplings[147]. Subsequent work introduced ML methods for state recognition from charge stability diagrams and transport measurements, enabling adaptive tuning protocols for high-dimensional gate-voltage landscapes[19]. Fully automated ML-based tuning was then shown to reach target QD regimes faster than human experts[148].

More recent works have moved toward more robust and scalable tuning workflows. Compact NNs have been proposed for charge-transition detection and possible on-chip autotuning[149]. Data-quality-control modules were introduced to make autonomous tuning more robust against noisy or unreliable measurements[150]. Uncertainty-aware NN strategies were later used to guide charge-state autotuning with reduced risk of failed exploration[151]. Cross-architecture approaches further demonstrated that a single ML-based tuning strategy can be transferred between different silicon and SiGe QD devices[152]. Finally, online experimental charge autotuning showed that NNs can be integrated directly into closed-loop calibration procedures[153].

Beyond reaching predefined operating points, ML has increasingly been used for inverse estimation of device properties. In QD-based KCs, adversarial Hamiltonian learning has been used to infer effective Hamiltonian parameters from conductance data[20]. In Majorana nanowires, NNs have also been trained to reconstruct disorder landscapes from conductance maps[154]. Such approaches are particularly relevant for QD chains, where extracting Hamiltonian parameters from measurements is a natural first step toward compensating fabrication disorder and stabilizing desired operating regimes.

In the context of topological superconductivity, ML techniques have also been applied to the detection, classification, and optimization of Majorana-like signatures. For example, ML classifiers have been used to distinguish Majorana-related ZBP signals from trivial signals in simulated tunneling data[155]. Vision-transformer-based approaches have further been used to infer topological indicators from conductance measurements in disordered Majorana nanowires[156]. Related work showed that such models can also guide disorder mitigation and optimize topological indicators[157]. In this dissertation, this direction is extended to QD-based Kitaev chain simulators by introducing a physics-informed NN autotuning framework that learns from conductance maps and proposes Hamiltonian-parameter corrections toward the emergence of Majorana modes, as discussed in Section 5.4.

Together, these developments establish ML as a central methodology for navigating the complexity of QD chains and for realizing automatic platforms aimed at the discovery, tuning, and stabilization of engineered topological superconducting regimes.

3.3 Synthetic data generation for topological systems

In contrast to the quantum state datasets discussed in Section 2.3, the data generation process for topological systems is based on sampling Hamiltonian parameters and computing experimentally motivated observables derived from them. The starting point is a QD chain Hamiltonian $H(\mathbf{P})$, where $\mathbf{P}$ denotes the tunable parameters of the device, such as local chemical potentials, interdot hoppings, spin-orbit couplings, Zeeman field, and proximity-induced superconducting pairing. The synthetic data are then generated by sampling $\mathbf{P}$ from selected parameter ranges and calculating the corresponding transport response.

Hamiltonian parameter sampling

The generated Hamiltonians are not sampled as arbitrary random matrices, but are restricted to physically motivated families describing Rashba QD chains coupled to

a nearby s-wave superconductor. This restriction is important because the learning models are intended to operate on data compatible with experimentally realizable gate-defined devices. In the AI-enhanced tuning study (Section 5.4), random parameter sets are used to generate detuned Hamiltonians, which are subsequently corrected by the NN toward Majorana-supporting regimes. Depending on the assumed experimental scenario, the model is trained either to correct electrically controlled local parameters $\{\mu_n, t_n, \lambda_n\}$ or to tune local potentials together with the global Zeeman field, $\{\mu_n, V_Z\}$.

In the Hamiltonian-learning study (Section 6.2), the parameter vector is instead treated as the latent variable to be inferred from transport maps. For the three-dot Rashba chain simulator, the learned vector is

$$\mathbf{P} = \{\mu_L, \mu_C, \mu_R, t_1, t_2, \lambda_1, \lambda_2\}, \tag{81}$$

while other quantities, such as the induced pairing and Zeeman field, are fixed to define the Hamiltonian family. Sampling is performed around a reference configuration $\mathbf{P}_0$ corresponding to a Majorana-supporting regime, but the training regions can be widened or narrowed to test the generalization capability of the model.

Conductance map generation

For each sampled Hamiltonian, the experimentally accessible representation is obtained from transport simulations. The scattering matrix is calculated using the Weidenmüller formula introduced in Eq. (78), and the differential conductance is obtained from Eq. (80). The resulting conductance maps play the role of high-dimensional observables analogous to measurement vectors in quantum tomography, but here they encode transport signatures of the underlying Hamiltonian rather than probabilities of projective measurements.

In the considered QD chain models, the conductance tensor is composed of several two-dimensional maps. The four conductance components G_{LL}, G_{LR}, G_{RL}, and G_{RR} are computed as functions of the Fermi energy and selected control-parameter detunings, such as local chemical potentials or the Zeeman field. In this way, each Hamiltonian parameter vector $\mathbf{P}$ is mapped to a structured multi-channel image $G(\mathbf{P})$. This representation is especially useful for NN models, since features such as ZBPs, avoided crossings, and gap openings appear as spatial patterns in the conductance maps.

Learning targets and physics-informed quantities

The topological studies concerned in this dissertation use these synthetic conductance maps for different learning objectives. In the Hamiltonian-learning task (Section 6.2), the goal is to reconstruct the hidden Hamiltonian parameters from the observable conductance maps. This is done using an unsupervised autoencoder-like architecture in which the encoder predicts $\mathbf{P}$, while the physics decoder implements the forward map

$$\mathbf{P} \mapsto H(\mathbf{P}) \mapsto S[H(\mathbf{P})] \mapsto G(\mathbf{P}). \tag{82}$$

The model is therefore trained by comparing the input conductance maps with the maps reconstructed from the predicted parameters, rather than by using externally supplied parameter labels.

In the AI-enhanced tuning task (Section 5.4), the synthetic data are generated with a different purpose. Here, the model receives conductance maps and proposes corrections $\delta\mathbf{P}$ to the Hamiltonian parameters. The learning objective is based on a differentiable quasimetric $\mathcal{M}$, introduced to quantify how close a given Hamiltonian is to the Majorana regime. For a Hamiltonian $H(\mathbf{P})$ with eigenpairs $\{\psi_i, E_i\}$ sorted according to increasing $|E_i|$, this quantity is defined as

$$\mathcal{M}(H(\mathbf{P})) = \frac{p_0}{2} \max\left[0,\, 2m_0 - \sum_{i>1} m_i\right], \tag{83}$$

where

$$m_i = |\langle\psi_i|M\rangle|\, e^{-|E_i|/\epsilon}, \qquad p_i = 2\max\left[0,\, 4|u_i|^2|v_i|^2 - \frac{1}{2}\right]. \tag{84}$$

Here, m_i measures the overlap of the eigenstate ψ_i with the left, or equivalently right, Majorana edge mode $M = \gamma_L$, weighted by its distance from zero energy, with $\epsilon = 0.1\,\text{meV}$ setting the energy scale. The factor p_i quantifies electron-hole symmetry of the eigenstate, where u_i and v_i denote its electron and hole components. For an ideal Majorana pair one obtains $m_0 = m_1 = 1$ and $p_0 = p_1 = 1$, which gives $\mathcal{M} = 1$.

This formulation favors Hamiltonians with zero-energy modes localized at the edges and with balanced electron-hole character. At the same time, the term $-\sum_{i>1} m_i$ penalizes additional low-energy states, which helps suppress trivial zero modes such as ABSs. As a result, the generated data are not only examples of possible device responses, but also inputs for a physics-informed optimization procedure that guides the model toward Majorana-supporting configurations.

Noise augmentation

To improve robustness to measurement imperfections, the synthetic conductance maps used in the Hamiltonian-learning study (Section 6.2) were additionally augmented with noise. The augmentation included stochastic distortions of the control-parameter axis, intended to mimic effects such as charge noise and slow gate drift, as well as additive uncorrelated and correlated noise in the conductance signal. The corresponding noise parameters were varied during training, exposing the model to a range of measurement-like distortions rather than to ideal simulated maps only. A detailed description of the noise model and its implementation is provided in the Appendix B of the *Learning Hamiltonians for solid-state quantum simulators* manuscript.

Dataset sizes

The main synthetic datasets used in the topological part of the dissertation are summarized in Table 3.

Overall, synthetic data generation for topological systems is based on constructing physically meaningful Hamiltonian ensembles and converting them into experimentally motivated transport signatures. This makes the datasets directly connected to the measurement protocols available in QD devices. At the same time, it allows the learning algorithms to address two complementary tasks: reconstructing hidden Hamiltonian parameters from conductance maps and actively tuning the system toward regimes supporting MZMs.

Table 3: Summary of synthetic datasets used in the topological part of the dissertation. The table reports the main data-generation procedures used in the final two studies.

Study	Generated data
Section 5.4 *AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes*	**Input:** conductance map tensors $G(\mathbf{P})$ generated from a multiple-dot Rashba chain Hamiltonian (Eq. 77). **Three-dot training:** 10 000 independent Hamiltonian samples per training epoch, with parameters sampled from predefined physical ranges. **Larger chains:** for the seven-dot case, 10 000 filtered samples were generated around configurations satisfying a threshold value of $\mathcal{M}$ before applying random detuning.
Section 6.2 *Learning Hamiltonians for solid-state quantum simulators*	**Input:** conductance map tensors $G(\mathbf{P})$ generated from a three-dot Rashba chain Hamiltonian (Eq. 77). **Training sets:** 10 000–50 000 sampled parameter vectors, depending on the size of the parameter region. **Noise augmentation:** additional noisy conductance maps were generated to test and improve robustness against realistic measurement distortions.

4 Machine Learning

This section introduces the ML concepts and models that form the methodological backbone of this dissertation. The central idea underlying all presented works is a data-driven pipeline, in which large-scale synthetic datasets are generated using physically motivated models, and subsequently used to train ML algorithms for tasks such as classification, regression, or inference. This approach enables systematic exploration of complex physical systems, including both quantum state data and Hamiltonian-based datasets, where ground-truth properties (such as existence of quantum correlations or specific topological properties) are known by construction.

A key contribution of this dissertation is the development and application of such a pipeline as a general framework for studying quantum and solid-state systems. By decoupling data generation from experimental acquisition, one can efficiently produce diverse and well-controlled datasets, which are then used to train models that can be directly applied or further fine-tuned on experimental data. This strategy provides a scalable and flexible alternative to purely experiment-driven analysis and allows ML models to capture physically relevant features already at the training stage.

It should be emphasized that the models, architectures, and learning formalisms introduced in this section are not confined to a single study, but are consistently employed and adapted across all research papers included in this dissertation. Consequently, this section serves as a common methodological reference point, and the reader is encouraged to follow it carefully, as many later results rely on the concepts, notation, and design choices established here.

4.1 Basic ideas

From a theoretical point of view, ML is a branch of statistical learning that aims to construct implementable models that could infer functional relationships or patterns directly from the data. Rather than being explicitly programmed with decision rules, ML systems improve their performance through subsequent observations of data samples. Formally, ML can be viewed as the problem of approximating an unknown mapping or distribution using a parametrized model whose parameters are optimized according to a predefined objective related to the given data[158,159].

Basic learning formalism

An ML model is typically represented as a parametrized function

$$f_\theta : \mathcal{X} \to \mathcal{Y}, \tag{85}$$

where $\theta \in \mathbb{R}^p$ denotes a set of trainable parameters (weights), $\mathcal{X}$ is the input space, and $\mathcal{Y}$ is the output space. Depending on the task, $\mathcal{Y}$ may represent discrete labels (classification), continuous values (regression), or whole probability distributions (generation). Elements of the input space $x \in \mathcal{X}$ are typically represented as feature vectors $x = (x^{(1)}, x^{(2)}, \ldots, x^{(d)})$, where individual components $x^{(j)}$ are referred to as *features*.

Given a data set $\{(x_i, y_i)\}_{i=1}^N$, $x_i \in \mathcal{X}, y_i \in \mathcal{Y}$, sampled from an unknown distribution, the training of an ML model consists of minimizing a target-specific loss

function $\mathcal{L}$ that quantifies the discrepancy between model predictions $f_\theta(x_i)$ and targets y_i [158]

$$\min_\theta \ \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}\big(f_\theta(x_i), y_i\big). \tag{86}$$

When f_θ is differentiable with respect to θ, optimization is commonly performed using gradient-based methods. The fundamental update rule of the *(stochastic) gradient descent* (SGD) method [160] in the following steps t takes the form

$$\theta_{t+1} = \theta_t - \eta\nabla_\theta\mathcal{L}(\theta_t), \tag{87}$$

where $\eta > 0$ is the learning rate. In practice, gradients are estimated on mini-batches of data to improve computational efficiency and introduce beneficial stochasticity, i.e., reducing the probability of converging to a single solution. Numerous adaptive variants of SGD have been proposed to further improve its efficiency. For example, the Adam optimizer incorporates running estimates of the first and second moments of the gradient to accelerate convergence in high-dimensional parameter spaces [161].

In the limit $\eta \to 0$, iterative updates can be approximated by a gradient flow equation,

$$\frac{d\theta}{dt} = -\nabla_\theta\mathcal{L}(\theta), \tag{88}$$

which provides a useful analytical perspective on convergence properties and on the geometry of the loss landscape [162,163]. This viewpoint has been instrumental in theoretical analyses of overparameterized models, implicit bias, and the role of optimization dynamics in generalization [164,165]. Recent work has further separated the effects of SGD bias from architectural overparameterization [166], connected early-stopped gradient descent with implicit regularization [167], and analyzed how explicit regularization shapes the implicit bias of gradient flow [168].

Learning paradigms

ML tasks are commonly categorized according to the availability of labeled data:

1. **Supervised learning.** In supervised learning, each input sample x is paired with a labeled target output y. The objective is to learn a mapping that generalizes to unseen data. This means that the model should not merely minimize the loss on the training set, but should capture regularities that remain valid for new samples drawn from the same or a similar distribution. This generalization ability is one of the main features distinguishing ML from ordinary numerical optimization of a fixed objective on a fixed data set. Typical tasks include classification and regression. The performance of the learned model is evaluated on data unseen by the model during the training phase, i.e., validation and test data sets, to estimate the generalization error.

2. **Unsupervised learning.** Unsupervised learning operates on unlabeled data and seeks to uncover latent structure. Common objectives include dimensionality reduction, clustering, and probability distribution estimation [159,15]. Methods such as principal component analysis (PCA), autoencoders, and probabilistic generative models aim to capture characteristic features or low-dimensional representations underlying high-dimensional observations.

3. **Semi-supervised learning.** Semi-supervised learning combines a limited set of labeled examples with a larger pool of unlabeled data. Using structural assumptions about data distribution, such approaches can improve predictive performance when obtaining labeled data is expensive or impractical.

4. **Reinforcement learning.** In reinforcement learning, the model does not depend on a predefined dataset. Instead, an agent interacts sequentially with an environment under given constraints. At each time step, the agent observes an environment state, selects an action, and receives a scalar reward signal as feedback. The objective is to learn a policy that maximizes the expected cumulative reward over time. Reinforcement learning is commonly formulated in the framework of Markov decision processes (MDPs) and is widely applied in control, robotics, game playing, and sequential decision-making problems[169].

Bias-variance tradeoff and generalization

A central concept in ML is generalization, i.e., the ability of a model trained on a finite data set to perform well on previously unseen samples. In such terms, the expected prediction error at a point x admits the decomposition into bias and variance errors,

$$\mathbb{E}\big[(\hat{f}(x) - f(x))^2\big] = \mathrm{Bias}^2(x) + \mathrm{Var}(x), \tag{89}$$

where $\hat{f}$ is the learned estimator and f is the true function[158]. High-bias models, with overly restrictive hypothesis classes, tend to underfit, whereas high-variance models, excessively sensitive to data changes, risk overfitting. This standard picture is summarized schematically in Fig. 2.

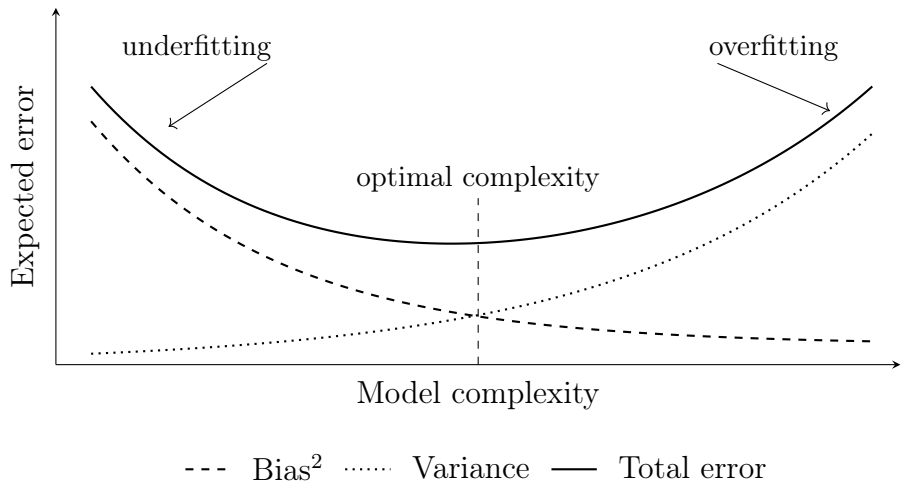


Figure 2: Schematic illustration of the classical bias-variance tradeoff. Increasing model complexity usually decreases bias but increases variance, leading to an intermediate regime with the lowest expected error. Low-complexity models tend to underfit, whereas highly complex models may overfit.

While reducing bias typically comes with increasing the number of parameters in the model, to mitigate the high variance problem, one has to introduce specific regularization techniques to control effective model complexity and improve generalization. Explicit strategies include weight decay, i.e., L_2 regularization known

as ridge regression in linear models, dropout, early stopping, and augmenting the training data[15]. In addition, gradient-based optimization itself can induce implicit regularization, favoring solutions with particular geometric or norm properties, especially in overparameterized regimes.

Modern deep NNs often contain more trainable parameters than training samples, a setting referred to as overparameterization, yet can generalize effectively. This phenomenon has motivated renewed theoretical interest in the interplay between model capacity, optimization dynamics, and data structure. Although comprehensive theory remains an active area of research[170], it is widely recognized that generalization depends not only on parameter count, but also on *architectural inductive biases*, i.e., assumptions and constraints encoded in the network architecture that make certain representations or solutions easier to learn than others, as well as on the data distribution and the specifics of the optimization procedure[171]. In particular, deep learning (DL) does not always follow the classical U-shaped bias–variance curve: in highly overparameterized regimes, the test error can decrease again beyond the interpolation threshold, a phenomenon known as *double descent*[172,173]. Consequently, understanding and controlling the bias-variance tradeoff, as well as incorporating domain-specific prior structure where appropriate, is central to the robust application of ML methods in data-driven scientific research.

Feature importance and Shapley values

Understanding the contribution of individual input features to model predictions is an important aspect of interpretability in ML. A principled and widely used approach to quantify feature importance is based on *Shapley values*[174], originally introduced in cooperative game theory. In this framework, each feature is treated as a "player" contributing to the final prediction of a model.

Formally, the Shapley value ϕ_i associated with feature i for model f is defined as its average marginal contribution over all possible subsets of features,

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} w(S, N) \left[f(S \cup \{i\}) - f(S) \right], \tag{90}$$

where N denotes the full set of features, S is a subset not containing i, $|S|$ denotes the cardinality of this subset, i.e., the number of features it contains, and $w(S, N) = |S|!(|N| - |S| - 1)!/|N|!$ is a combinatorial weighting factor ensuring fair attribution. This definition guarantees that the total model prediction is decomposed into contributions from individual features in a consistent and additive manner.

An important advantage of Shapley values is their model-agnostic nature, allowing them to be applied to a wide class of ML models, including NNs and ensemble methods. Moreover, they provide *local explanations*, i.e., feature contributions can be evaluated for individual data samples and subsequently averaged to obtain global importance measures across a dataset[175].

Beyond individual contributions, the framework can be extended to capture interactions between features via *Shapley interaction values*, which quantify the joint effect of feature pairs beyond their independent contributions. This is particularly relevant in physical systems, where observables often exhibit strong correlations and nontrivial interdependence.

In the context of this dissertation, Shapley-based methods provide a systematic way to analyze which input features (e.g., measurement outcomes or observables) are most relevant for model predictions, offering insights into both the learned representations and the underlying physical structure of the data.

Tree-based models

Tree-based models are supervised learning methods that approximate the target function by recursively partitioning the input space. A decision tree performs this partitioning through a sequence of hierarchical splits, where each internal node tests one feature and each leaf assigns a prediction. For regression, the leaf prediction is typically the average target value of the training samples reaching that leaf, whereas for classification it is usually the majority class. As a result, a decision tree produces a piecewise-constant approximation of the target function. Deep trees are flexible and can capture nonlinear feature interactions, but they often have high variance and are therefore prone to overfitting.

Random forests (RFs) are ensemble learning methods based on collections of decision trees, introduced as a robust extension of *bagging*, i.e., bootstrap aggregating, techniques[176]. In a RF, multiple trees $\{T_b(x)\}_{b=1}^{B}$ are trained on bootstrap samples of the data, meaning samples drawn with replacement from the original training set, and their predictions are aggregated to form the final model,

$$f(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{91}$$

for regression, or by majority voting in classification tasks.

The purpose of this aggregation can be understood using the bias-variance tradeoff discussed above. Averaging many high-variance estimators reduces the variance of the ensemble without substantially increasing the bias. In the idealized case of B independent estimators with variance σ_T^2, the variance of their average is reduced to σ_T^2/B. In practice, individual trees are not independent, so RFs introduce two sources of randomness: sampling of training examples through bagging and random selection of feature subsets at each split. This decorrelates individual trees and significantly reduces variance compared to a single deep tree, while maintaining low bias.

From a functional perspective, RFs approximate nonlinear decision boundaries through an ensemble of locally adaptive partitions of the feature space. Their empirical success stems from combining variance-reducing aggregation with flexible, nonparametric base learners, making them particularly effective when feature interactions are complex but data preprocessing is limited.

Deep learning

Many modern applications of ML methods use NNs consisting of multiple layers with nonlinear transformations[15]. DL is typically based on a feedforward NN with depth L and can be written as a composition

$$f_\theta(x) = f^{(L)} \circ f^{(L-1)} \circ \cdots \circ f^{(1)}(x), \tag{92}$$

where each layer $f^{(\ell)}$ consists of multiple units, also called *neurons* or *perceptrons*. A single perceptron computes an affine transformation of its inputs followed by a nonlinear activation function,

$$h(x) = \sigma(w^T x + b), \tag{93}$$

where w and b are trainable parameters and σ is the activation function. In deep NNs, the layers are commonly split into an input layer, which receives the input sample, the output layer that produces the result, and the intermediate hidden layers that capture increasingly complex data representations. In this sense, deeper layers can combine simpler features extracted in earlier layers into more abstract task-relevant representations.

From an analytic perspective, NNs can be understood as *universal function approximators*[177]. In particular, the universal approximation theorem states that a feedforward network with a single hidden layer of sufficient width and a nonconstant nonlinear activation function can approximate any continuous function on a compact subset of $\mathbb{R}^d$ with arbitrary accuracy. Such a network can be viewed as a two-layer perceptron model of the form

$$F(x) = \sum_{j=1}^{m} a_j \, \sigma(w_j^T x + b_j), \tag{94}$$

where increasing the number of hidden units m increases the expressive capacity of the model. This result explains why nonlinear NNs are, in principle, very flexible function approximators. The practical motivation for using deeper architectures is not only universality itself, but also the fact that depth often provides more efficient representations of complex functions than a very wide shallow network.

Their empirical success is often attributed to their capacity for representation learning, i.e., the automatic construction of task-relevant features from raw data, reducing the need for manual feature engineering.

4.2 Advanced deep learning architectures

The multilayer perceptron (MLP) introduced above provides a general-purpose NN model, but it does not explicitly encode assumptions about the structure of the input data. Many modern DL applications therefore rely on specialized architectures that impose architectural constraints or structured connections adapted to the data type, such as images, sequences, graphs, or high-dimensional measurement maps.

Convolutional neural networks (CNNs)

Convolutional neural networks (CNNs) are a class of deep architectures designed to exploit local correlations and translational equivariance in structured data[178]. Instead of connecting every input unit to every output unit, CNNs apply spatially localized filters, or *kernels*, that are shared across the input domain. This weight sharing substantially reduces the number of trainable parameters and encodes translational symmetry directly in the network architecture.

Consider an input feature tensor $h^{(l-1)} \in \mathbb{R}^{C_{\text{in}} \times H \times W}$ at layer $l-1$, where C_{in} denotes the number of input channels, and H and W are the spatial height and

width, respectively. A convolutional layer l with C_{out} output channels is defined by a set of learnable kernels $W^{(l)} \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}} \times K_h \times K_w}$ and biases $b^{(l)} \in \mathbb{R}^{C_{\text{out}}}$, where K_h and K_w denote the height and width of the kernel. In image-like applications, commonly used kernel sizes include small local windows such as 3×3 or 5×5, while larger effective receptive fields are obtained by stacking multiple convolutional layers. For stride s_h and s_w in the vertical and horizontal directions, respectively, the output at spatial position (i, j) in output channel k is given by

$$h^{(l)}_{k,i,j} = \sigma\left(\sum_{c=1}^{C_{\text{in}}} \sum_{u=0}^{K_h-1} \sum_{v=0}^{K_w-1} W^{(l)}_{k,c,u,v}\, h^{(l-1)}_{c,\, is_h+u,\, js_w+v} + b^{(l)}_k\right), \tag{95}$$

where σ denotes the nonlinear activation function. The indices (u, v) run over the spatial dimensions of the kernel, while (i, j) label the spatial positions of the output feature map. The stride parameters (s_h, s_w) determine the step size with which the kernel is shifted across the input. In the absence of additional input padding, i.e., without additional spatial extension, the dimensions of the output are reduced according to the kernel size and the stride.

This operation can be interpreted as computing a set of local weighted sums that detect specific patterns within receptive fields of size $K_h \times K_w$. By stacking multiple convolutional layers, CNNs progressively increase their effective receptive field and learn hierarchical feature representations. Due to their ability to capture spatially localized and translation-equivariant structures, CNNs are particularly well suited for grid-structured data such as images, time-frequency representations, and spatially resolved measurements. In scientific applications, including condensed matter and quantum systems, NN and CNN-based models have been successfully employed to identify phase transitions and classify many-body states from raw numerical or experimental data[179,180]. CNNs have also been used to detect phase transitions in spin models and quantum many-fermion systems, including settings where conventional correlation-function analysis is difficult[181,182].

Siamese networks and geometric learning

Siamese architectures consist of two (or more) identical networks with shared parameters, producing output embeddings $g_\theta(x)$[183]. Training is guided by similarity-based loss functions such as *contrastive loss*[184],

$$\mathcal{L} = \|g_\theta(x) - g_\theta(x')\|, \tag{96}$$

which compares the representations of related inputs x and x'. Typically, x' is obtained via a transformation that preserves certain properties of interest, and the model is trained to produce similar embeddings for such pairs. This makes Siamese networks particularly useful in settings where relative comparisons are easier to define than absolute labels, for example, when the output should be invariant under a given transformation[185].

From a broader perspective, Siamese architectures can be viewed as a practical realization of *geometric learning* principles, where the goal is to learn representations that respect the intrinsic structure of the data[186]. In particular, they encourage the embedding space to reflect notions of similarity, distance, or symmetry present in the underlying problem. This is closely related to learning invariant or equivariant

representations, where transformations in the input space correspond to predictable changes (or invariances) in the latent space.

Such approaches are especially relevant in physical systems, where symmetries and conservation laws naturally define equivalence classes of states or configurations[187]. By incorporating these geometric relations into the training procedure, Siamese networks enable the construction of embeddings that are both physically meaningful and robust to transformations. This makes Siamese-like approaches well suited for tasks such as phase identification[188,189] or comparison of quantum states[190].

Autoencoders

Autoencoders are unsupervised NN models designed to learn efficient representations of data by compressing inputs into a lower-dimensional *latent space* and subsequently reconstructing them. A schematic representation of an autoencoder is shown in Fig. 3.

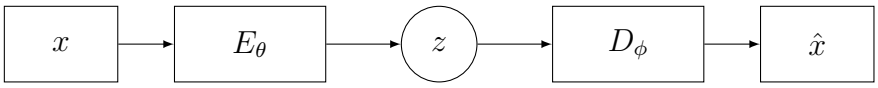


Figure 3: Schematic illustration of the autoencoder model

It consists of two components: an *encoder* E_θ that maps the input x to a latent representation z, and a *decoder* D_ϕ that reconstructs the input from this representation,

$$z = E_\theta(x), \qquad \hat{x} = D_\phi(z) \approx x. \tag{97}$$

The model is trained by minimizing a reconstruction loss, typically of the form

$$\mathcal{L}_{\text{rec}} = \|x - \hat{x}\|^2, \tag{98}$$

which encourages the latent variable z to capture the most relevant features of the data distribution. Additionally, reconstruction error can be used as an anomaly score, allowing the detection of rare events that deviate from the training data distribution[191]. In other words, an *anomaly detector* is an ML model that is trained to detect objects deviating significantly from those observed in the training set.

Typically, the dimensionality or structure of z is restricted, so the autoencoder is forced to learn a compressed representation that retains essential information while avoiding noise or redundancy. As a result, the latent space learned by the encoder often exhibits meaningful structure, making autoencoders useful for tasks such as dimensionality reduction, feature extraction, and denoising[192]. In particular, when trained on structured physical data, the latent variables can implicitly encode underlying parameters, effective physical concepts, or phases of the system[193,194,195].

From a broader perspective, autoencoders can also be interpreted as learning an approximate low-dimensional manifold on which the data lie. Variants such as *variational autoencoders* (VAEs) introduce probabilistic structure in the latent space, while *physics-informed autoencoders*, described further in this dissertation, incorporate known physical models directly into the decoder, allowing for interpretable and constrained representations of complex systems.

Variational autoencoders

VAEs are probabilistic generative models that combine NNs with variational inference. In contrast to deterministic autoencoders, VAEs assume that data x are generated from latent variables z through a likelihood model $p_\phi(x|z)$, while a prior distribution $p(z)$, typically the standard normal distribution $\mathcal{N}(0, I)$, is imposed on the latent space. Here, ϕ denotes the parameters of the decoder, or generative model, whereas θ denotes the parameters of the encoder, or inference model, which produces the approximate posterior distribution $q_\theta(z|x)$. VAEs are typically trained by maximizing the evidence lower bound (ELBO),

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q_\theta(z|x)}[\log p_\phi(x|z)] - D_{\text{KL}}(q_\theta(z|x)\|p(z)) \,, \tag{99}$$

or, equivalently, by minimizing its negative. The first term formulates the reconstruction objective $\mathbb{E}_{q_\theta(z|x)}[\log p_\phi(x|z)]$, i.e., the expected log-likelihood of the data reconstructed by the decoder from latent variables sampled from $q_\theta(z|x)$. The second term regularizes the latent representation by penalizing deviations of the approximate posterior from the prior. Here $D_{\text{KL}}(q\|p)$ denotes the Kullback-Leibler divergence[80], defined for continuous distributions as

$$D_{\text{KL}}(q\|p) = \int q(z) \log \frac{q(z)}{p(z)} \, dz, \tag{100}$$

typically used to evaluate the discrepancy between two probability distributions.

By balancing reconstruction accuracy and latent regularization, VAEs learn structured, continuous latent spaces that enable convenient data sampling and interpolation. Hence, VAEs can be naturally used as generative models and have been widely applied to model complex probability distributions.

Recurrent neural networks (RNNs)

Recurrent neural networks (RNNs) are architectures designed to model sequential or time-dependent data by incorporating internal state variables that evolve over input steps[196]. Unlike feedforward networks, RNNs process an input sequence $\{x_t\}_{t=1}^T$ iteratively, updating a hidden state h_t at step t according to

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b) \,, \tag{101}$$

where W_h and W_x are trainable weight matrices, b is a trainable bias parameter, and σ is a nonlinear activation function. The hidden state serves as a memory that encodes information from previous inputs, enabling the network to capture temporal correlations.

However, standard RNNs suffer from a vanishing gradient problem during training, which makes learning long sequence dependencies increasingly difficult[197]. *Long short-term memory* (LSTM) networks address this limitation by introducing a structured memory cell c_t and gated mechanisms that regulate information flow[198]. A

standard LSTM update can be written as

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)\,, \tag{102}$$
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)\,, \tag{103}$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)\,, \tag{104}$$
$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)\,, \tag{105}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{106}$$
$$h_t = o_t \odot \tanh(c_t), \tag{107}$$

where i_t, f_t, and o_t are the input, forget, and output gates, respectively, $\tilde{c}_t$ is the candidate memory update, and $\odot$ denotes element-wise multiplication. In LSTM, σ typically denotes the sigmoid activation function,

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{108}$$

which maps gate values to the interval $(0, 1)$, while tanh is the hyperbolic tangent function,

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}, \tag{109}$$

which maps its input to the interval $(-1, 1)$. The forget gate controls which information from the previous memory cell is retained, the input gate controls which new information is stored, and the output gate controls which part of the memory is exposed as the hidden state. This structure separates memory storage from the instantaneous hidden representation and improves the ability to learn long-term dependencies.

RNNs and their gated variants have been widely applied to sequential data, including time series, natural language, and dynamical systems. In scientific contexts, they provide a flexible framework for modeling temporal evolution, correlated measurements, and autoregressive generative processes, with examples including LSTM-based forecasting of high-dimensional chaotic systems[199] and RNN-based variational wave functions for quantum many-body states[200].

Transformers and attention-based models

Transformers are modern architectures based entirely on *attention* mechanisms, which replace recurrence or convolution in favor of global, content-dependent interactions between input elements[16]. Given a sequence of input tokens $\{x_i\}_{i=1}^{N}$, each token is first embedded into a d-dimensional representation and linearly projected into three distinct vectors: queries (Q), keys (K), and values (V). For a matrix of input embeddings $X \in \mathbb{R}^{N\times d}$, these projections are defined by standard matrix multiplication as

$$Q = XW_Q, \qquad K = XW_K, \qquad V = XW_V,$$

where $W_Q \in \mathbb{R}^{d\times d_k}$, $W_K \in \mathbb{R}^{d\times d_k}$, and $W_V \in \mathbb{R}^{d\times d_v}$ are learnable weight matrices. Consequently, $Q, K \in \mathbb{R}^{N\times d_k}$ and $V \in \mathbb{R}^{N\times d_v}$.

The self-attention mechanism computes pairwise similarities between queries and keys and uses them to form weighted combinations of the values:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V. \tag{110}$$

Here, the matrix product QK^T contains dot-product similarities between tokens, while the factor $1/\sqrt{d_k}$ prevents these products from becoming too large for high-dimensional keys and queries. The softmax function is applied row-wise and normalizes the attention weights,

$$\text{softmax}(a_j) = \frac{e^{a_j}}{\sum_l e^{a_l}}. \tag{111}$$

Conceptually, queries determine the information sought, keys represent how tokens can be matched, and values provide the aggregated content.

In practice, transformer blocks interleave multi-head attention with position-wise fully connected layers. In multi-head attention, several independent attention operations, called heads, are performed in parallel with different projection matrices. The resulting outputs are concatenated and linearly combined, enabling the model to capture diverse relational patterns within the same layer.

Vision Transformers (ViTs) extend this framework to image data by partitioning an image into non-overlapping patches, flattening them into tokens, and processing them using the standard transformer architecture[201]. Despite the absence of convolutional inductive biases such as locality and translation equivariance, ViTs have demonstrated competitive performance on large-scale visual tasks, highlighting the flexibility of attention-based models in capturing long-range spatial dependencies.

Physics-informed neural networks (PINNs)

A particularly important class of NN models in physics and materials science are *physics-informed neural networks* (PINNs)[202,203]. These architectures incorporate physical laws directly into the learning process by augmenting the loss function with constraints derived from governing equations,

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda \mathcal{L}_{\text{physics}}, \tag{112}$$

where $\mathcal{L}_{\text{data}}$ represents the standard loss function for a given supervised or unsupervised problem, while $\mathcal{L}_{\text{physics}}$ encodes prior knowledge in the form of system differential equations, conservation laws, or symmetry constraints. This formulation enables data-efficient learning and improves generalization by restricting the solution space to physically plausible configurations.

Beyond this standard formulation, a closely related and particularly powerful paradigm is based on embedding physical models directly into the NN architecture, often referred to as *physics-informed decoders*, *physics decoders*, or *physics-constrained autoencoders*. Related ideas have appeared in quantum chemistry and materials science, where NNs are coupled to physically motivated Hamiltonian models rather than used as purely black-box predictors[204,205,206]. In this dissertation, this idea is further developed in the form of physics decoder architectures, where the encoder latent representation corresponds to physically meaningful parameters, while the decoder implements a known forward model mapping of these parameters to observable quantities. Formally, given an input x, e.g., experimental or simulated measurement data, the encoder predicts a parameter vector

$$\mathbf{P} = E_\theta(x), \tag{113}$$

and the decoder reconstructs the observable via a deterministic physical mapping,

$$\hat{x} = D_{\text{phys}}(\mathbf{P}), \tag{114}$$

where D_{phys} is not a generic NN, but a differentiable implementation of the underlying physical model. In practice, this means that the predicted parameters $\mathbf{P}$ are inserted directly into the equations defining the model, and the corresponding physical operators or numerical relations are evaluated to obtain the predicted observable $\hat{x}$.

Such architectures effectively combine the advantages of PINNs and autoencoders. Instead of penalizing violations of physical laws only through an auxiliary loss term, the model enforces consistency by construction, since all predictions must lie within the range of the implemented physical forward model. This significantly improves interpretability, as the latent space directly corresponds to physical parameters, and reduces the risk of unphysical solutions. Moreover, the training objective typically reduces to a reconstruction loss,

$$\mathcal{L}_{\text{rec}} = \left\| x - D_{\text{phys}}(E_\theta(x)) \right\|^2, \tag{115}$$

while the governing physical equations are implicitly satisfied through the decoder.

In the context of solid-state and quantum systems, this framework is particularly well-suited for inverse problems, such as Hamiltonian learning or device characterization, where one aims to infer underlying parameters from experimentally accessible observables. In the Hamiltonian-learning architecture developed in this dissertation (Section 6.2), the encoder predicts the parameters of a QD chain Hamiltonian, while the physics decoder implements the corresponding transport calculation using the scattering-matrix formalism. The model is therefore trained to reconstruct conductance maps from the inferred Hamiltonian parameters, rather than to learn a direct black-box mapping between conductance maps and labels. This makes the predicted latent variables interpretable and physically constrained.

A closely related idea is also relevant for quantum state-based problems studied in this dissertation. By constructing a physics-informed decoder that explicitly generates physically admissible quantum states, and in particular separable mixed states, one can enforce the separability constraint directly at the level of the model architecture (Section 5.1). In this setting, the latent representation parameterizes only separable states, ensuring that all reconstructed density matrices are physically valid and belong to the separable subset of the Hilbert space. This formulation naturally lends itself to anomaly-detection-based approaches for identifying quantum correlations: states that cannot be well reconstructed by such a separability-constrained model can be interpreted as exhibiting nonclassical correlations.

Overall, the physics-informed viewpoint provides a common conceptual basis for several parts of this dissertation. It appears both through physically motivated loss terms, such as the quasi metric quantifying the presence of Majorana modes used for tuning of QD Hamiltonians (Section 5.4) and the symmetry-enforcing terms used in entanglement identification (Section 5.2), and through architecture-level constraints, such as the physics decoder approach discussed above. In this way, the dissertation combines loss-constrained and architecture-constrained strategies for integrating domain knowledge with data-driven learning.

4.3 Applications

ML has become broadly useful across multiple scientific disciplines as it allows for extracting informative representations from complex data, building predictive models when analytical descriptions are incomplete, and accelerating inference by learning approximations for expensive computations. These principles have led to widely adopted ML pipelines for transforming image-like data, text, time series, and structured scientific datasets, often complementing established statistical modeling.

Computer vision

Computer vision addresses the processing of images, videos, and other grid-structured measurements, including classification, detection, and segmentation tasks. A prominent early demonstration of DL at scale is the use of a deep convolutional network trained to achieve substantially improved performance on a large-scale image classification problem[207]. Beyond classification, many scientific tasks require dense, pixel-level predictions. Ronneberger *et al.* introduced U-Net specifically for biomedical image segmentation, proving that an encoder-decoder CNN with skip connections can yield accurate segmentation from relatively few annotated images (supported by strong data augmentation)[208]. Similar models are now routinely adapted to microscopy[209] and X-ray tomography[210]. Another important NN applications are human pose detection[211] and high-fidelity graphics tasks such as image upscaling or super-resolution[212,213].

Natural language processing (NLP)

Modern ML methods, in particular transformer-based approaches, have provided multiple tools that can be effectively used in natural language processing. Although typical applications of transformer models extend their machine translation capabilities[214], the most recent usage of the so-called large language models proves their efficiency in various language tasks, such as text information extraction (e.g., entity recognition[215], relation extraction[216], sentiment analysis[217]), building searchable representations[218], and enabling language interfaces with scientific knowledge[219].

Forecasting and temporal modeling

Forecasting problems arise naturally in the analysis of complex phenomena in dynamical systems. One tool, which aims at predicting future data, given historical information, is DeepAR[220], a probabilistic forecasting method based on an autoregressive recurrent network trained on many related time series. Another approach is to use the Temporal Fusion Transformer (TFT) model, utilizing an attention-based architecture for multi-horizon forecasting that combines strong predictive performance with interpretability-oriented components[221]. For high-dimensional geophysical weather dynamics, Lam *et al.* proposed to use an ML forecasting approach trained on reanalysis data from past atmospheric conditions[222]. These works illustrate how sequence ML models can scale from standard time series to complex temporal fields and weather conditions.

Molecular and material sciences

In molecular modeling, ML is used to learn structure-property relations and to accelerate the exploration of chemical or material space. SchNet is one example of a neural architecture designed for atomistic systems, which demonstrates an accurate prediction of molecular/material properties[223]. In structural biology, the AlphaFold model shows that NN-based modeling can achieve highly accurate protein structure prediction, explicitly addressing protein folding prediction[224]. Finally, generative NNs demonstrated their ability to design complex material architectures and effectively sample new structures[225,226].

Physics applications

Physics has widely adopted ML both for data-driven classification tasks and as a tool for constructing compact representations of high-dimensional systems. As representative applications in quantum correlation identification (Section 2.2), quantum tomography (Section 2.3), as well as Hamiltonian learning and detection of topological properties (Section 3.2) have already been discussed, the following overview is limited to a selection of additional examples highlighting the broader impact of ML in physics.

NN approaches have been successfully applied to quantum many-body systems, where they can serve as variational ansatzes for many-body wave functions, commonly referred to as neural quantum states[227]. This idea has also been extended to quantum chemistry, where fermionic NN wave functions can be used to approximate solutions of the many-electron Schrödinger equation[228]. In the context of quantum control, Bukov *et al.* demonstrated that reinforcement learning can discover high-fidelity state-preparation protocols in interacting many-body systems[229]. More broadly, ML techniques are increasingly used across diverse areas of physics. In gravitational wave research, they are employed for signal detection, noise reduction, and parameter estimation directly from interferometric data[230]. In materials science, machine-learned interatomic potentials (MLIPs) provide accurate and computationally efficient mappings from atomic configurations to potential energies, bridging the gap between first-principles calculations and empirical models[231,232]. This direction has been further strengthened by geometric and equivariant DL methods, which explicitly encode symmetries such as rotations, translations, or permutations into the model architecture[186,233].

Another rapidly developing direction is operator learning, which can be viewed as a data-driven extension of physics-informed modeling for systems governed by differential equations. Instead of learning a single solution, neural operators learn mappings between function spaces, for example from initial conditions or material parameters to full solution fields[234,235]. This makes them particularly relevant for surrogate modeling of nonlinear PDEs and complex multiphysics systems. In fluid dynamics and astrophysics, physics-informed NNs and related operator-learning approaches have been used to model nonlinear partial differential equations, including magnetohydrodynamic turbulence and subgrid-scale processes, demonstrating strong predictive capabilities in complex multiphysics settings[236]. ML is also widely applied in high-energy physics, where it supports event classification, particle reconstruction, and acceleration of numerical simulations[237].

Across these domains, ML serves as a versatile computational framework for ex-

tracting structure from data, compressing complex representations, and optimizing analysis or control protocols in regimes where traditional methods face significant limitations.

Overall, the impact of ML in scientific research is derived from combining expressive models with problem-specific inductive biases and rigorous validation, allowing predictive performance and scientifically meaningful inference.

5 Synopsis of the published papers

This section summarizes the published papers forming the peer-reviewed part of the dissertation. These works establish the main methodological direction of the dissertation: using machine learning to extract physically meaningful information from quantum data. They cover data-driven detection of quantum correlations, symmetry-aware entanglement identification, learning from tomographic measurement data, and physics-informed autotuning of QD Hamiltonians toward Majorana modes. Together, these papers constitute a core part of the dissertation, contributing to its central goal of developing machine-learning methods for quantum state analysis, measurement-based inference, and physics-informed control of solid-state quantum systems.

Since all papers included in this section are multi-author works, it is important to distinguish the individual contribution of the author of this dissertation. Table 4 summarizes the author's contribution to the conception and methodology, numerical implementation and data generation, analysis and visualization, and preparation of each publication.

Table 4: Individual contribution of the author of this dissertation to the published papers.

Paper	**Author's contribution**
Section 5.1 *Data-driven criteria for quantum correlations*	Contributed to the conception and development of the ML methodology; designed and implemented the separability-preserving and permutation-equivariant NN architecture and its training pipeline; developed and generated synthetic quantum state datasets; performed numerical experiments, benchmarking, and analysis of the results; prepared visualizations and contributed substantially to the preparation and revision of the manuscript.
Section 5.2 *Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning*	Contributed substantially to the conception and development of the supervised and semisupervised entanglement-detection framework; implemented and tested CNN, Siamese, and ensemble models together with the relevant symmetry transformations; developed and conducted synthetic states generation and numerical benchmarking, including tests on PPT entangled states; analyzed the numerical results, prepared visualizations, and contributed to preparation and revision of the manuscript.
Section 5.3 *Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests*	Contributed to the design of the measurement-based ML study; implemented and trained NN models for entanglement inference from tomography data; generated tomographic synthetic data and contributed to the comparative analysis of NN and RF models, including feature-importance and noise studies; participated in the interpretation and visualization of the results and in preparation and revision of the manuscript.

Paper	Author's contribution
Section 5.4 *AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes*	Designed the architecture of the physics-informed auto-tuning framework; implemented the NN, training pipeline and closed-loop tuning procedure; developed and used the numerical simulation pipeline for generating conductance maps and evaluating the Majorana quasi-metric; performed numerical studies of parameter tuning, iterative feedback, and model scaling; prepared the numerical analysis and visualizations and contributed substantially to preparation and revision of the manuscript.

5.1 Data-driven criteria for quantum correlations

The first paper of this dissertation, *Data-driven criteria for quantum correlations*, addresses one of the central challenges in quantum information science: the identification and characterization of quantum correlations in experimentally relevant quantum states. It is a fundamental task in quantum information science, as these correlations constitute essential resources for quantum communication, computation, and cryptography[23,17,45]. In the context of solid-state quantum computing platforms, such as semiconductor QDs or hybrid nanowires, access to the full quantum state is typically indirect and limited. Consequently, efficient and scalable methods for detecting nonclassical correlations from incomplete or noisy data are of key importance for device characterization and control. This difficulty is further compounded by the lack of scalable, conclusive analytical methods for entanglement identification in an arbitrary quantum state[30].

The paper proposes a novel, fully data-driven approach to detecting quantum correlations using ML. Specifically, an unsupervised NN model is constructed to learn the structure of separable quantum states. The model is trained exclusively on separable three-qubit states and subsequently used as an anomaly detector: states that cannot be well reconstructed by the network are classified as quantum correlated. This formulation bypasses the need for explicit entanglement measures, which are known to be computationally intractable for mixed states, and instead leverages the representational power of NNs to learn physically meaningful structures directly from data. Importantly, this approach shifts the paradigm from evaluating predefined analytical criteria toward learning implicit representations of quantum correlations, which may capture more general or experimentally relevant features than standard measures.

A key technical contribution of the work is the design of a NN architecture that incorporates physical constraints. The model is constructed to preserve separability and to be equivariant under permutations of qubits, ensuring consistency with the symmetries of the underlying quantum system. In this sense, the approach fits into the broader trend of geometric DL, where known symmetries and structural constraints are incorporated into the model architecture[186]. The network operates similarly to an autoencoder, where the encoder maps a three-qubit density matrix into single-qubit components, and the decoder reconstructs a separable approximation via tensor products. The reconstruction error, defined via an L_1 loss, serves as a quantitative measure of correlation strength. In this construction, appropri-

ately defined convolutional layers effectively generalize the partial trace operation, while additional fully connected layers introduce nonlinear transformations that enable the model to capture more complex structures beyond simple factorization. As demonstrated in the paper, removing these nonlinear components causes the model to collapse to a trivial baseline equivalent to the partial trace, highlighting the importance of architectural expressivity for detecting nontrivial correlations.

A crucial component of the approach is the generation of large synthetic datasets of quantum states with known properties. In order to obtain a diverse and representative dataset, several state-generation methods are employed, including sampling random states from the Haar measure and constructing parametrized families of separable states. The training dataset contains over half a million separable states, while independent test sets include both separable and entangled states, allowing the performance of the method to be evaluated systematically. Such synthetic data generation is a common strategy in ML-based studies of quantum systems, where analytical labels are available for simulated states but not necessarily for experimental ones. Additionally, careful dataset design ensures coverage of different subclasses of states, such as nondiscordant, discordant, and entangled states, which is essential for understanding how the learned representation generalizes beyond the training distribution.

The results demonstrate that the model achieves very high accuracy ($> 99\%$) in distinguishing separable and entangled pure states, while maintaining strong performance for mixed states. However, one of the most significant and unexpected findings is that the network is considerably more effective at detecting quantum discord than entanglement. For mixed states, the balanced accuracy reaches approximately 93% for discord detection, compared to about 82% for entanglement detection (see performance analysis in Fig. 2 of the paper). Moreover, the reconstruction loss landscape closely reproduces the geometry of nondiscordant states in Hilbert space, indicating that the learned representation aligns more naturally with discord than with entanglement.

This observation has important conceptual implications. While entanglement is traditionally regarded as the primary resource for quantum technologies, quantum discord is more directly related to measurable quantities and may be more accessible in experimental settings. The fact that an unsupervised ML model, trained without explicit labels for discord, preferentially captures this form of correlation suggests that data-driven approaches may uncover physically relevant structures that are not immediately apparent from standard theoretical definitions. In particular, it indicates that the notion of similarity learned by the NN is more closely aligned with operational or measurement-based definitions of quantum correlations, rather than strictly formal separability criteria.

From the perspective of the full PhD dissertation, this work provides a foundational methodology for using ML to extract physically meaningful information from quantum data. The proposed framework demonstrates that NNs can learn physically meaningful representations of separability without requiring explicit supervision. This is particularly relevant for realistic experimental situations, where full QST may be incomplete or noisy, and where analytical correlation criteria may not be directly applicable. Moreover, the anomaly-detection paradigm used here provides a general strategy for identifying nontrivial quantum features as deviations from learned reference distributions, which can be extended to other problems such

as phase detection, Hamiltonian learning, or device calibration.

In addition, the architectural design principles introduced in this work – such as incorporating physical constraints (separability preservation, symmetry equivariance) into an autoencoder-like network – provide a general framework that is further developed in subsequent papers. These ideas are crucial for ensuring that ML models remain interpretable and physically consistent when applied to quantum systems. Embedding prior physical knowledge into model design not only improves performance but also reduces the risk of learning spurious correlations, which is particularly important in high-dimensional quantum systems with limited data availability.

Finally, the results highlight the potential of ML methods to complement traditional analytical tools in quantum information. By revealing that data-driven models more naturally capture certain types of quantum correlations (notably discord), this work motivates a broader shift toward learning-based approaches in quantum device characterization.

5.2 Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning

The second paper of this dissertation, *Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning*, builds directly upon the data-driven framework introduced in the previous work and addresses one of its key limitations: the difficulty of reliably identifying entanglement in mixed quantum states using analytical criteria. While measures such as negativity and the PPT criterion provide operational tools for entanglement detection, they are either computationally demanding or fundamentally inconclusive for higher-dimensional systems. In particular, the existence of PPTES, which satisfy the PPT criterion despite being entangled, highlights the intrinsic limitations of such approaches[56,57]. As emphasized in this work, no general, practically applicable criterion exists for entanglement identification beyond low-dimensional systems, making the problem particularly challenging in realistic experimental scenarios.

To overcome these limitations, the paper proposes a supervised and semisupervised ML framework based on deep CNNs, extending the unsupervised paradigm developed in the previous paper. In contrast to anomaly detection, where separability is learned implicitly, here the task is formulated as a classification problem, where the model predicts entanglement for each possible bipartition of a multiqubit system. This formulation provides a more detailed characterization of quantum correlations, enabling the identification of entanglement structure rather than a single scalar indicator. However, a central difficulty arises from the lack of reliable labels for mixed states, especially in the PPT regime. To address this issue, a *verified* training strategy is introduced, where only states with unambiguous labels, e.g., non-positive partial transpose states or explicitly constructed separable states, are used for supervised training, while ambiguous PPT states are excluded from the labeled training set. This is not treated as a final solution of the PPTES problem, but rather as a way to avoid training on unreliable labels. The missing information is then partially compensated by the semisupervised symmetry-based learning scheme discussed below.

A major methodological advancement of this work is the use of symmetry-aware Siamese NN architectures for entanglement identification. Specifically, the application of siamese networks is tailored to the quantum-information problem of learning an entanglement detector consistent with physical invariances. Building on the idea that entanglement is invariant under local unitary transformations and permutations of subsystems, the model explicitly incorporates these physical symmetries into the learning process. This is achieved by constructing a triple Siamese network, where multiple copies of the same CNN, sharing parameters, process different symmetry-transformed versions of the input state, including locally rotated and permuted density matrices. The training objective is augmented with additional regularization terms that enforce consistency between the outputs for symmetry-related inputs, effectively guiding the model to learn representations that respect fundamental physical invariances. Such an approach represents a natural continuation of the physically constrained architectures introduced in the previous paper, now extended to supervised learning settings.

The use of semisupervised learning further enhances the robustness of the model. Initially, the network is trained in a supervised manner using verified labels, after which the influence of symmetry-enforcing terms is gradually increased during training. This strategy allows the model to first learn coarse classification boundaries and then refine them by incorporating physically motivated constraints. Importantly, this approach mitigates the risk of overfitting to imperfect labels and improves generalization to previously unseen classes of states, particularly PPTES, which are absent from the training dataset.

Another key innovation is the introduction of an ensemble learning framework, which directly leverages the unsupervised model developed in the first paper. Specifically, the reconstruction loss of the autoencoder is used as a learned separability measure to partition the state space into distinct domains. Separate Siamese models are then trained on each domain, effectively specializing in different regions of the Hilbert space. This strategy is conceptually analogous to constructing multiple entanglement witnesses tailored to different classes of states[238]. As shown in the results, such an ensemble significantly improves detection performance for challenging PPTES classes, particularly those that are close to separable states in terms of standard metrics.

The results demonstrate that the proposed approach achieves high accuracy for both pure and mixed states, exceeding 97% for pure states and reaching above 90% for mixed states when using Siamese architectures. More importantly, the model exhibits strong generalization capabilities to PPTES, despite not being trained on them. For example, high detection accuracy is achieved for well-known classes of PPTES, such as Horodecki states[239] and unextendible product basis states[240], while more challenging families (like Acin et al. states[241]) require the ensemble approach for improved performance. These findings indicate that the model captures features of entanglement that go beyond those encoded in standard analytical criteria, effectively learning a more general notion of quantum correlations.

The study also compares models that operate directly on density matrices with models trained on measurement data. The results indicate that using the density matrix representation leads to higher accuracy, suggesting that the structural information contained in the matrix representation is beneficial for NN learning. Additionally, preliminary experiments demonstrate that the proposed approach can

be extended to four-qubit states, demonstrating applicability beyond the three-qubit setting.

From the perspective of the overall PhD dissertation, this work represents a crucial step toward bridging the gap between theoretical entanglement measures and practical data-driven detection methods. While the first paper established that unsupervised learning can capture general features of quantum correlations, this study demonstrates how supervised and semisupervised techniques can refine these representations to achieve higher accuracy and better interpretability. In particular, the integration of physically motivated symmetries into the learning process provides a general design principle for ML models in quantum physics, improving robustness and encouraging consistency with underlying physical laws.

In the broader context of quantum technologies, the results of this paper demonstrate that ML can serve not only as a practical tool for entanglement detection but also as a means of extending beyond traditional analytical frameworks. By successfully identifying entanglement in regimes where standard criteria fail, the proposed methods open new possibilities for automated characterization and control of quantum devices. This is especially relevant for scalable quantum computing architectures, where efficient and reliable identification of quantum resources is essential for system optimization and error mitigation.

5.3 Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests

The third paper of this dissertation, *Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests*, naturally extends the previous results by shifting the focus from density-matrix-based learning to direct measurement-based entanglement detection. Although the second paper (Section 5.2) indicated that density matrices provide a more convenient and structured input representation for learning entanglement, the present work considers the experimentally closer setting in which the model operates directly on tomographic measurement outcomes rather than on an explicitly reconstructed density matrix. This setting is particularly relevant experimentally, as reconstructing the full density matrix requires an exponentially growing number of measurements and precise control of measurement bases, which becomes impractical even for relatively small systems[7].

Relying on the measurement outcomes as model input introduces a fundamental complication: information about the quantum state is no longer localized but rather distributed across multiple interdependent measurements. As shown in the analysis of the tomography scheme (Fig. 2 of the paper), different groups of measurements encode different aspects of the density matrix, i.e., diagonal elements (occupations), single-qubit off-diagonal elements, referred to here as local coherences, and two-qubit off-diagonal elements, referred to as nonlocal coherences, which are directly relevant for entanglement[54]. This fragmentation of information fundamentally alters the learning landscape compared to the previous papers.

To explore this regime, two conceptually distinct ML approaches are compared, i.e., deep NNs and ensemble-based RF. Although both models are trained to predict entanglement (quantified through the concurrence measure) from tomography data,

their internal representations and decision mechanisms differ significantly. The research focuses on 2-qubit systems, where NNs achieve a higher predictive accuracy overall, exceeding 92% classification accuracy compared to approximately 88% for RF models. However, the central contribution of this work lies not in performance comparison alone, but in uncovering the qualitatively different ways in which these models extract information from measurements.

A key methodological innovation of this study is the systematic analysis of *measurement importance*, performed using multiple complementary techniques, including feature importance metrics, perturbation analysis, and Shapley values[174] for interpretability. These analyses reveal a striking and initially counterintuitive result: NNs and RFs rely on fundamentally different subsets of measurements to infer entanglement. NNs follow physically intuitive behavior, assigning the highest importance to measurements containing information about nonlocal coherences, followed by local coherences, and only marginal importance to diagonal elements. This hierarchy aligns with the structure of entanglement itself and reproduces the behavior observed when concurrence is computed analytically from reconstructed density matrices.

In contrast, RF models exhibit an inverted hierarchy of importance. They assign the highest relevance to measurements corresponding to diagonal elements (occupations), while measurements encoding nonlocal coherences are deemed least important. At first glance, this appears to contradict the fundamental role of coherence in entanglement. However, the paper demonstrates that this behavior arises from the interdependence of measurement data: occupation measurements simultaneously include the information required for reconstructing all other elements of the density matrix. As a result, RF models prioritize the features that are most informative for global inference, rather than those that directly encode entanglement. This is consistent with a standard distinction between tree-based models and NNs: RFs often rely strongly on the quality and preprocessing of the supplied feature representation, whereas deep NNs can learn task-adapted internal features directly from structured data[176,14]. This distinction highlights a key conceptual difference between models that learn physical structure (NNs) and those that exploit statistical dependencies (RFs).

Importantly, this discrepancy disappears when both models are trained directly on density matrix elements instead of measurement outcomes. In that case, both approaches recover the physically expected importance ordering, confirming that the observed differences are not intrinsic to the models themselves, but rather emerge from the structure of tomography data. This observation provides a deeper understanding of how information encoding affects ML representations in quantum systems and clarifies why the use of full density matrices, as emphasized in the previous paper, leads to more physically interpretable models.

Another contribution of this work is the analysis of robustness to noise in measurement data. By introducing controlled perturbations to individual measurements, it is shown that NNs are more sensitive to noise in measurements carrying direct physical information about entanglement, while RFs are more robust due to their ensemble nature. This robustness arises from averaging over many decision trees trained on different subsets of features, effectively mitigating the impact of noise in any single measurement. Consequently, while NNs provide higher accuracy in ideal conditions, RFs may outperform them in realistic experimental scenarios with noisy

data.

From the perspective of the dissertation, this paper moves the progression from architecture-constrained unsupervised learning and symmetry-aware supervised and semisupervised models developed in the previous works to a fully measurement-driven framework that closely reflects experimental constraints. It demonstrates that the success of ML in quantum physics depends not only on model architecture, but also critically on how physical information is encoded in the input data. In particular, the results highlight that incorporating physical structure, such as symmetries or explicit representations of quantum states, can guide models toward more interpretable and physically meaningful solutions, whereas purely data-driven approaches may exploit indirect correlations that are less transparent.

Furthermore, this study reinforces a central theme of the dissertation: the interplay between physical principles and ML methodologies. While the first two papers already incorporated physical constraints and symmetries, here the comparison between NNs and RF models reveals how different learning paradigms implicitly encode (or fail to encode) physical constraints when operating on experimentally accessible data. This insight is crucial for the design of future quantum ML models, particularly in scenarios where only partial or noisy measurements are available.

5.4 AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes

The last published paper, *AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes*, extends the central theme of learning from measurement data into a qualitatively new regime: instead of reconstructing quantum states or inferring their properties, the focus shifts to actively controlling and engineering quantum systems toward desired physical phases. In particular, the work addresses the problem of autotuning Hamiltonian parameters in QD simulators to realize MZMs, a key ingredient for TQC. This represents a natural progression from the previous paper, where learning-based methods were used to detect quantum entanglement from different sets of measurements. Here, the same paradigm is elevated from quantum system analysis to *closed-loop, physics-informed control* of quantum devices.

Building on the insight that measurement data can encode relevant information about underlying quantum systems, this work leverages conductance maps as the primary input to a learning framework. Unlike the tomography setting, where measurement outcomes are used to reconstruct a static object (the density matrix), conductance maps provide indirect but highly structured information about the system Hamiltonian. Moreover, transport measurements are a standard tool for probing and identifying nanostructures, making conductance maps a natural experimental representation for QD-based devices. As illustrated schematically in Fig. 1 of the paper, the proposed PINN autotuning (PINNAT) framework establishes a feedback loop in which measured conductance maps are processed by a NN to predict parameter updates that drive the system closer to a topological regime supporting MZMs.

At the heart of the approach lies a physics-informed unsupervised learning strategy. The model is trained not to reproduce target labels, but to optimize a differentiable quasi-metric $\mathcal{M}$ that quantifies the presence of Majorana modes. This metric combines several physical indicators, including edge-state localization, zero-energy

spectral weight, and electron-hole symmetry, while explicitly penalizing trivial zero-energy states such as ABS. As described in the Section 3.3 and paper's Methods section, the metric is constructed from the eigenstructure of the BdG Hamiltonian and serves as a guiding objective that embeds domain knowledge directly into the learning process. This design ensures that the network does not merely fit data correlations, but instead learns to navigate the physically meaningful landscape of Hamiltonian parameters.

The underlying physical model is a chain of QDs described by a Rashba-Zeeman superconducting Hamiltonian (see Eq. 77 for the exact formalism), whose parameters, such as local chemical potentials, hopping amplitudes, spin-orbit coupling, and Zeeman energy, must be carefully tuned to reach the topological phase. Sufficient deviations from the Kitaev chain sweet spot can reduce or close the topological gap and degrade the Majorana modes, making manual tuning highly challenging in realistic, noisy devices. In contrast, the PINNAT framework learns to associate patterns in conductance maps with regions of parameter space that support MZMs, effectively constructing an implicit inverse map from experimental observables to optimal control directions.

A key methodological innovation of this work is the use of a vision-transformer (ViT) architecture to process conductance maps as structured visual data. Each input consists of multiple conductance components measured as functions of system parameters and the Fermi energy, forming a high-dimensional tensor representation of transport behavior. The network outputs corrective updates to selected Hamiltonian parameters, which are then applied to the system. As shown in the results (Fig. 2 and Fig. 3 of the paper), even a single update step significantly enlarges the region of parameter space where the Majorana metric $\mathcal{M}$ is positive, indicating successful movement toward the topological phase. Moreover, iterative application of the model, where updated measurements are fed back into the network, leads to progressive refinement and substantial improvement, as demonstrated in Fig. 4 of the paper.

An important observation is that the model learns nontrivial correlations between different Hamiltonian parameters, allowing it to compensate for limited controllability in realistic setups. For example, even when only a subset of parameters (such as local potentials and global Zeeman field) is tunable, the network can effectively steer the system toward the desired regime by exploiting underlying physical dependencies. As discussed in the analysis of parameter maps in the paper's Results section, tuning the Zeeman field proves particularly effective, reflecting its dominant role in opening the topological gap. This highlights a practical advantage of ML-based tuning: the model can automatically identify useful patterns in conductance maps and translate them into parameter corrections expressed directly in the physically meaningful parameter space of the Hamiltonian. Notably, since the predicted control variables correspond to Hamiltonian parameters, the resulting tuning procedure remains interpretable in terms of the underlying physical model rather than operating purely as a black-box optimization.

Crucially, the proposed framework learns to distinguish configurations supporting Majorana modes from tested trivial zero-energy states without requiring direct supervision based on explicitly labeled ABS examples. As illustrated in Fig. 5 of the paper, this discrimination emerges from the physics-informed training objective: the predefined Majorana quasi-metric $\mathcal{M}$ favors configurations satisfying the chosen

Majorana criteria while penalizing configurations associated with trivial zero-energy states. This robustness is essential for practical applications, where experimental signatures can be ambiguous and prone to false positives.

From a broader perspective, this paper represents a conceptual culmination of the dissertation trajectory. Earlier works demonstrated how ML can extract quantum correlations, incorporate symmetries, and reconstruct states from measurement data, even in incomplete regimes. The present study advances this paradigm further by showing that learning-based models can close the loop between measurement and control, enabling autonomous navigation of complex quantum parameter spaces. In this sense, the role of ML evolves from a passive analysis tool into an active component of experimental design and operation.

At the same time, the work reveals new challenges associated with scalability and realism. As discussed in Appendix D of the paper, the computational cost increases substantially with the number of QDs. Although the dimension of the effective BdG Hamiltonian grows only linearly with system size and the corresponding diagonalization has polynomial complexity, the dimensionality of the Hamiltonian parameter space also increases, making configurations with large values of $\mathcal{M}$ progressively more difficult to sample. Combined with repeated conductance calculations and differentiable diagonalizations required during training, this resulted in an approximately exponential increase in the observed data-generation and training times over the system sizes considered in the paper. Furthermore, the reliance on simulated conductance maps highlights the need for future extensions incorporating experimental noise and hardware imperfections. Nevertheless, the demonstrated ability to generalize beyond the training regime and to operate iteratively suggests that such approaches can form the basis for practical, adaptive control protocols in quantum devices.

In summary, this paper establishes a powerful framework for physics-informed, data-driven control of quantum systems, bridging the gap between measurement-based inference and Hamiltonian engineering. By integrating transport measurements, neural architectures, and topological criteria into a unified loop, it opens a pathway toward automated realization of Majorana modes and, more broadly, toward intelligent control of complex quantum matter.

6 Synopsis of the unpublished manuscripts

This section summarizes two yet unpublished manuscripts that constitute an integral part of the research carried out within this dissertation. Although these works have not yet completed the publication process, they are not separate from the main dissertation topic. Instead, they extend the same central ideas developed in the published manusripts: learning from incomplete quantum measurements, incorporating physical constraints into NN models, and using measurement data to infer the structure of quantum systems. The first manuscript develops learning-based reconstruction of density matrices from informationally incomplete tomography data, while the second introduces a physics decoder framework for Hamiltonian learning in solid-state quantum simulators. Including these results is therefore necessary to present the full methodological progression of the dissertation.

The two manuscripts summarized in this section are likewise collaborative works. Table 5 specifies the individual contribution of the author of this dissertation to their development and preparation.

Table 5: Individual contribution of the author of this dissertation to the unpublished manuscripts.

Manuscript	**Author's contribution**
Section 6.1 *Learning quantum tomography from incomplete measurements*	Contributed substantially to the conception and development of the learning-based approach to informationally incomplete tomography; designed and implemented recurrent LSTM selector–reconstructor models and implemented the corrector networks; generated and processed the synthetic quantum state and measurement datasets; performed numerical benchmarking for two-, three-, and four-qubit systems; contributed to analyses of measurement ordering, restricted state families, and positive semidefiniteness; prepared visualizations and contributed substantially to preparation and revision of the manuscript.
Section 6.2 *Learning Hamiltonians for solid-state quantum simulators*	Contributed to the conception and development of the physics decoder approach to Hamiltonian learning; contributed to the numerical transport simulations used for data generation and reconstruction; contributed to the implementation of the neural network encoder and physics decoder; developed and implemented a noise model, performed numerical studies of measurement noise impact, model mismatch, and increasing QD chain size; contributed to preparation of visualizations and revision of the manuscript.

6.1 Learning quantum tomography from incomplete measurements

The first unpublished manusript, *Learning quantum tomography from incomplete measurements*, constitutes a natural continuation of the previous study by moving

from entanglement detection based on measurement data to the more demanding task of full density matrix reconstruction under measurement-resource constraints. While the third published paper (Section 5.3) demonstrated that meaningful quantum properties can be inferred directly from measurement outcomes without explicit state reconstruction, the present work revisits QST itself, focusing on the fundamentally challenging regime of informationally incomplete measurements. In this setting, the number of available measurements is insufficient to uniquely determine the quantum state, making standard reconstruction techniques intrinsically ill-posed.

Building upon the limitations identified in the previous paper, where the structure of measurement data significantly influenced model behavior, this work explicitly addresses the question of how to optimally reconstruct quantum states when only partial information is available. In conventional tomography, the complete product-projector catalog produces 4^N expectation values for an N-qubit system, while a generic normalized density matrix contains $4^N - 1$ independent real parameters. Therefore, when the number of measurements M is smaller than the complete tomographic set, the reconstruction problem becomes underdetermined, and standard linear inversion methods fail due to the non-invertibility of the measurement matrix. The commonly used workaround, i.e., pseudoinverse (Moore-Penrose inverse)[79] based reconstruction, provides only approximate solutions and may violate physical constraints such as trace normalization or positivity constraints.

To overcome these limitations, two complementary NN-based approaches are introduced to reconstruct quantum states directly from incomplete measurement data. The first approach, referred to as the *corrector network*, can be interpreted as a data-driven refinement of the pseudoinverse reconstruction. Instead of relying solely on the Moore-Penrose inverse, the network learns correction terms that depend on the available measurement operators and, in some variants, also on the corresponding measurement outcomes. This effectively transforms the reconstruction problem into a supervised learning task, where the network approximates an improved mapping from incomplete measurement statistics to density-matrix estimates. In the basic versions, the correction coefficients depend only on the chosen measurement operators, preserving the pseudoinverse-based linear reconstruction as the analytical starting point while learning how this reconstruction should be corrected. When the correction coefficients also depend on measurement outcomes, or when quadratic variants are used, the overall estimator becomes more expressive, yet remains anchored in the pseudoinverse reconstruction scheme.

The second framework employs recurrent NNs (i.e. LSTM architectures) to perform sequential reconstruction. In contrast to static reconstruction schemes, this model processes measurement data step by step, updating its estimate of the quantum state as new measurements are incorporated. Crucially, this architecture is coupled with a measurement selection mechanism, enabling the network not only to reconstruct the state but also to adaptively choose the most informative next measurement. As illustrated in the scheme presented in Fig. 1 of the manuscript, the model alternates between selecting measurement operators and refining the state estimate, effectively learning a measurement protocol optimized for incomplete tomography. The details of both the corrector networks and the selector-reconstructor LSTM models are discussed in Appendix A of the manuscript.

The numerical results demonstrate that both approaches achieve lower reconstruction errors than traditional methods such as pseudoinverse reconstruction and

MLE[7]. Importantly, this comparison is not fully equivalent in terms of prior information available to the estimators. The neural models are trained on an ensemble of admissible quantum states and therefore learn statistical regularities of this distribution, which act as an implicit prior during reconstruction. In contrast, the pseudoinverse contains no comparable learned prior, while MLE incorporates physical constraints and a statistical measurement model but is not trained on the state distribution used by the NNs. The observed advantage of the neural estimators should therefore be understood partly as a consequence of exploiting information learned from the training distribution. For two-qubit systems, the neural models consistently achieve a lower reconstruction error, quantified using a raw-output Bures-type score, across the analyzed range of measurement numbers. As shown in Fig. 3 of the manuscript, the improvement is particularly pronounced in the strongly undercomplete regime, where standard methods struggle the most. Furthermore, the recurrent models that jointly optimize measurement selection and reconstruction often exhibit the strongest performance in the undercomplete and intermediate-measurement regimes, indicating that adaptive measurement strategies play a crucial role in efficient tomography.

An important insight emerging from this work is that the effectiveness of reconstruction depends not only on the number of measurements but also on their ordering and informational content. In particular, the LSTM-based selector learns nontrivial measurement sequences that avoid redundancy and prioritize measurements contributing complementary information. This behavior becomes evident when the state space is restricted, for example, to the family of X states. In such cases, the model identifies structured measurement sequences that first determine diagonal elements and subsequently refine coherences, as discussed in the analysis of measurement ordering in Appendix D of the manuscript. This behavior is also reflected in Fig. 4 of the manuscript, where training on the restricted X-state family improves reconstruction compared with training on general mixed states. This observation provides an operational realization of the no-free-lunch principle for quantum tomography. In informationally incomplete tomography, no reconstruction method can uniquely determine an arbitrary unknown state from insufficient measurement information without introducing additional assumptions or prior information. In the neural approaches considered here, such prior information is supplied implicitly by the training distribution: the models learn regularities of the admissible state ensemble and exploit them when selecting measurements and reconstructing states.

From a broader perspective, this manuscript deepens the central theme developed throughout the dissertation: the interplay between physical structure, measurement representation, and ML. While the third published paper revealed that different models extract different types of information from measurement data when performing entanglement detection, the present work shows how learning-based methods can improve undercomplete reconstruction by exploiting statistical regularities of the training ensemble and the geometry of physical density matrices, especially positive semidefiniteness. As shown in Section VI and Fig. 7 of the manuscript, positive semidefiniteness is not imposed during evaluation, yet the neural models tend to reconstruct matrices whose lowest eigenvalue is nonnegative on average. This does not mean that missing information is uniquely recovered. Instead, the neural models use the prior implicitly learned from the training distribution to select physically plausible reconstructions consistent with the available measurements.

Moreover, the results highlight a crucial trade-off between model expressivity and robustness. The corrector networks retain interpretability by extending known analytical methods, whereas the recurrent models achieve higher performance at the cost of increased complexity and potential memory limitations for larger systems. As shown in Fig. 5 of the manuscript, both approaches were extended from the two-qubit case to three- and four-qubit systems in the presented experiments. These results demonstrate applicability beyond the smallest systems, but do not establish asymptotic scalability, particularly given the exponential growth of the generic density-matrix parameter space. The larger-system results also reveal that LSTM-based methods may retain a small but finite error close to the complete-measurement regime, which is attributed to partial forgetting of the earliest measurement expectations in long sequences.

It is also important to note that the numerical benchmarks in the present manuscript are performed in the infinite-shot regime, where measurement expectation values are evaluated exactly from the target state. Thus, the study isolates the effect of informational incompleteness and the reconstruction map itself. Validation under realistic experimental conditions remains an important next step, including finite-shot statistics, calibration errors, detector and measurement noise, and possible noise-aware training or data augmentation.

In the context of the dissertation, this manuscript complements the earlier quantum-correlation studies by extending the use of physically motivated structure to informationally incomplete tomography. The preceding works incorporated such structure through separability preservation, permutation equivariance, and symmetry-aware learning, whereas here it enters directly through the reconstruction and measurement selection procedures. In particular, the corrector network is constructed as a refinement of the analytical pseudoinverse reconstruction and can incorporate constraints derived from the measurement subspace, while in the sequential approach the Born rule relation is included explicitly in the training loop, allowing the reconstruction loss to propagate back to the measurement selector. The resulting models can therefore learn not only how to reconstruct states from incomplete data, but also how to select informative measurements, enabling more efficient and adaptive quantum characterization protocols.

6.2 Learning Hamiltonians for solid-state quantum simulators

The second unpublished manuscript, *Learning Hamiltonians for solid-state quantum simulators*, serves as a methodological and conceptual complement to the preceding work on AI-assisted autotuning toward MZMs (Section 5.4). While the previous study focused on steering a quantum system toward a desired topological phase through closed-loop updates, the present contribution addresses a more fundamental inverse problem: how to infer effective Hamiltonian parameters directly from experimental observables. In this sense, it completes the transition from measurement-driven learning, through control, to effective model identification, a cornerstone for autonomous quantum device operation.

At the core of the work, similarly to the previous works, lies a physics-informed, unsupervised learning framework designed to extract effective Hamiltonian parameters from transport measurements in solid-state systems. Specifically, the approach

targets QD chains that emulate Kitaev-like models capable of hosting MZMs. As in the previous paper, the primary data modality consists of high-dimensional conductance maps, encoding rich information about the system's spectral and transport properties. However, rather than using these maps to guide parameter updates, they are now treated as inputs to an inverse model that reconstructs the Hamiltonian parameters themselves.

The proposed architecture follows a physics decoder autoencoder paradigm inspired by PINNs. As illustrated schematically in Fig. 1 of the manuscript, the encoder network (implemented as a ViT) maps conductance maps to a latent representation corresponding to Hamiltonian parameters, while the decoder is not a generic NN but an explicit implementation of the underlying physical model. In particular, the decoder uses the S-matrix formalism and transport equations to reconstruct conductance maps from predicted parameters, ensuring that all outputs remain physically consistent. This architecture embeds the governing equations directly into the learning loop, effectively constraining the hypothesis space to physically realizable solutions.

The manuscript considers the same class of Hamiltonians as the AI-enhanced tuning work (Section 5.4), namely a Rashba-Zeeman superconducting QD chain, described in detail in Eq. 77. Hamiltonian parameters define a high-dimensional control space in which sweet spots give rise to Majorana modes. As shown in the eigenstate diagrams in Fig. 2 of the manuscript, these regimes are characterized by zero-energy states with strong edge localization and electron-hole symmetry, key signatures of topological behavior. The learning task is therefore not only to reconstruct parameters but also to do so accurately in regions of physical relevance, including the vicinity of such topological regimes.

A central result of the manuscript is that the model learns an effective inverse mapping from conductance maps to Hamiltonian parameters without direct supervision on the parameter values. Instead, the encoder is trained through the physics decoder by requiring that the conductance maps reconstructed from the predicted parameters reproduce the input maps. As demonstrated in Fig. 3 of the manuscript, this reconstruction remains accurate not only within the training region but also over a finite region outside it, despite the relatively sparse sampling of the seven-dimensional parameter space. Since the mapping from conductance observables to Hamiltonian parameters is not guaranteed to be globally unique, agreement between input and reconstructed conductance maps should not be interpreted as proof of unique parameter identification. Rather, within the physically relevant parameter ranges imposed in the study, the model learns parameter sets that reproduce the observed transport response and generalize beyond the training domain.

This generalization is further analyzed by varying the size of the training domain. As shown in Figs. 4 and 5 of the manuscript, there exists a trade-off between coverage and precision: larger training regions improve global generalization but slightly reduce local accuracy, reflecting the finite capacity of the NN. Notably, the vicinity of the Majorana sweet spot appears particularly favorable for learning, as conductance features in this regime exhibit similar patterns, making them easier to encode and generalize. This observation provides an important physical insight: the structure of experimentally accessible observables can directly influence the learnability of Hamiltonian parameters.

Another key aspect of the work is the robustness to experimentally motivated

noise and model imperfections. By augmenting synthetic conductance maps with noise models that capture both instrumental noise and charge fluctuations, as described by Eq. (5) of the manuscript and detailed in Appendix B of the manuscript, it is demonstrated that the framework can be adapted to handle imperfect data. As shown in Fig. 6 of the manuscript, while noise initially degrades performance, retraining the model with noise-augmented data restores accurate parameter prediction. This supports the potential applicability of the approach to experimental settings, where noise is unavoidable.

The robustness analysis is further extended to Hamiltonian-level imperfections. In these tests, the input conductance maps are generated from a perturbed Hamiltonian that includes effects not represented in the decoder, such as variations in fixed parameters, Rashba-field orientation disorder, or cross-capacitance between gates. The physics decoder, however, retains the ideal single-level Hamiltonian family. As shown in Fig. 7 of the manuscript, the reconstruction error increases smoothly rather than abruptly with disorder strength, indicating that the inferred effective parameters remain meaningful under moderate model mismatch.

An even stronger test is performed by generating input data from an extended Hamiltonian containing a second orbital level on each QD, while keeping the decoder restricted to the original single-level Hamiltonian. In this case, the goal is not to recover the full microscopic Hamiltonian, which is outside the decoder's representation, but to determine whether the data can still be described by an effective Hamiltonian within the assumed model family. As shown in Figs. 8 and 9 of the manuscript, the model can still reconstruct a close conductance pattern and infer meaningful effective parameters, while the remaining reconstruction error serves as an indicator that the assumed Hamiltonian family is incomplete.

Taken together, these results outline a unified paradigm in which ML acts as an intermediary between experimental data and theoretical models, enabling both inference and control of quantum systems. In this paradigm, measurement data are no longer merely analyzed post hoc, but are continuously integrated into learning loops that reconstruct, interpret, and optimize the underlying physical systems. The manusript also analyzes architecture complexity and scalability, showing that the ViT encoder performs better than the tested CNN-based alternatives and that the approach can be extended from three to five and seven QDs with a gradual increase in reconstruction error. This establishes a foundation for autonomous quantum simulators, capable not only of exploring complex phases such as those supporting Majorana modes, but also of identifying and adapting their own effective models in real time.

In summary, this final manusript elevates the role of ML from a tool for state reconstruction and parameter tuning to a framework for discovering and validating effective Hamiltonians. By embedding physical laws directly into neural architectures and demonstrating robust performance under experimentally motivated noise and model imperfections, it provides a crucial step toward autonomous, self-calibrating quantum devices, thereby closing the conceptual arc of the dissertation.

7 Conclusions

This dissertation explored the use of ML methods for extracting physically relevant information from quantum systems and for optimizing their control, with a particular focus on solid-state platforms relevant for scalable quantum computing. The central goal was to bridge the gap between experimentally accessible data and the underlying physical models, enabling both efficient characterization of quantum states and automated tuning of complex quantum devices. Across the published papers and unpublished manuscripts included in the dissertation, a consistent methodological framework has been developed, combining data-driven learning with physically informed constraints applied to ML models to ensure both accuracy and interpretability.

The first problem addressed in the dissertation was the identification of quantum correlations without relying on a predefined analytical correlation measure. To this end, an unsupervised neural network architecture was developed to learn the structure of separable three-qubit states, with separability preservation and qubit permutation equivariance incorporated directly into the model. The resulting anomaly-based criterion successfully distinguished correlated from uncorrelated states and, for mixed states, was found to reflect quantum discord more closely than entanglement. These results were obtained using synthetically generated quantum states. Extending the approach to larger systems and to experimentally reconstructed states therefore remains an important direction for further validation.

The second study addressed the difficulty of supervised entanglement detection when complete and universally reliable labels are not available, particularly for PPT entangled states. The original contribution was a symmetry-aware Siamese architecture that incorporates entanglement-preserving transformations, including local unitary operations and qubit permutations, together with supervised and semisupervised learning strategies. The models demonstrated improved generalization to states outside the directly labeled training domain, including several classes of PPT entangled states. Nevertheless, the method does not constitute a universal entanglement criterion, and its performance remains dependent on the state families and labels represented during training. Further studies on larger systems and broader state distributions are therefore required.

The third published work considered entanglement inference directly from tomographic measurement data, avoiding explicit reconstruction of the density matrix. Neural networks and RFs were compared not only in terms of classification accuracy, but also through the measurement features on which their predictions relied. The results showed that NNs primarily emphasized measurements related to nonlocal coherences, whereas RFs gave greater importance to occupation-type measurements that also encode information relevant to other matrix elements. This demonstrated that the apparent physical interpretation of feature importance depends both on the learning algorithm and on the chosen representation of the measurement data. The main limitation of this analysis is that it relies exclusively on simulated tomography data. Therefore, validating the conclusions on experimental hardware data, including realistic measurement imperfections, represents an important next step.

The fourth published work moved from inference to device control and addressed the problem of tuning QD Hamiltonians toward regimes supporting MZMs. A physics-informed closed-loop framework was developed in which conductance maps

are processed by a neural network trained using a differentiable Majorana quasimetric combining zero-energy spectral weight, edge localization, and electron-hole symmetry while suppressing trivial zero-energy states such as ABS. Numerical experiments demonstrated that the trained model can propose Hamiltonian-parameter updates that increase the occurrence of Majorana-like regimes in simulated QD chains. The result establishes the feasibility of integrating transport observables, neural models, and a physics-based objective into a single feedback loop, but the demonstrated tuning remains simulation-based. Direct integration with an experimental device, together with tests under calibration errors, parameter drift, measurement noise, and model mismatch, is therefore an essential next step.

The first unpublished manuscript addressed quantum state reconstruction from informationally incomplete measurements. Since an arbitrary unknown state cannot be uniquely reconstructed from insufficient information without additional assumptions, the central contribution was to investigate how such prior information can be learned from a training ensemble. Two complementary approaches were developed: corrector networks that refine pseudoinverse-based reconstruction and recurrent LSTM models that sequentially reconstruct the state while also selecting informative measurements. For two-qubit systems, both approaches achieved lower reconstruction errors than pseudoinverse and MLE-based baselines in the undercomplete regime, and the methods were further extended to three- and four-qubit systems. These results demonstrate that statistical regularities learned from the training ensemble can be exploited to improve undercomplete tomography and measurement selection, without implying unique recovery of information absent from the measurements. The presented benchmarks assume exact measurement expectation values in the infinite-shot regime, so validation under finite-shot statistics, calibration errors, and realistic measurement noise remains an important next step.

The second unpublished manuscript addressed the complementary problem of inferring an effective Hamiltonian from transport measurements. A physics-decoder autoencoder was developed in which a vision-transformer encoder predicts Hamiltonian parameters, while an explicit S-matrix-based decoder reconstructs the corresponding conductance maps. This architecture embeds the assumed transport model directly into the learning procedure and allows the inferred latent variables to retain a physical interpretation. Numerical experiments on simulated QD chains showed that the method can infer effective Hamiltonian parameters outside parts of the training region, remain robust to noisy conductance maps and moderate Hamiltonian-level disorder, and return meaningful effective parameters even when data are generated by an extended Hamiltonian outside the decoder model family. These results should nevertheless be interpreted as effective-model identification rather than exact recovery of an unknown microscopic Hamiltonian. Consequently, experimental validation and systematic characterization of model mismatch remain necessary before applying the method to real devices.

A common practical component of these studies was the construction of the numerical pipelines required to generate training and validation data. In the quantum-information problems, standard random-state sampling was combined with structured state families and controlled parametrizations appropriate to the corresponding task. For the solid-state studies, numerical models of QD systems and their transport response were developed to generate conductance maps and Hamiltonian-dependent observables. These simulation pipelines made it possible to benchmark

the proposed ML approaches under controlled conditions and to study situations for which sufficiently large labeled experimental datasets are not presently available.

Another important aspect of the presented research is the emphasis on reproducibility and openness. The developed models, datasets, and simulation frameworks have been organized into publicly accessible repositories, enabling further validation, reuse, and extension by the research community. This provides a practical basis for reproducing the numerical results and for adapting the presented methods to related quantum-information and solid-state problems.

Taken together, the dissertation demonstrates several complementary ways in which physical knowledge can be incorporated directly into the ML workflow: through symmetry-preserving architectures, analytically motivated reconstruction schemes embedded into the learning process, learned measurement protocols, physics-informed objective functions, and explicit physical decoders. The demonstrated benefits include improved inference of quantum correlations, reconstruction from incomplete measurements, effective Hamiltonian identification, and closed-loop parameter optimization. At the same time, most of these results were obtained using synthetic data generated from numerical models, even when the input representations were chosen to correspond closely to experimentally accessible quantities such as tomography outcomes or conductance maps. The results should therefore be viewed as demonstrations of methods and principles rather than as completed experimental validation.

The most important next step is consequently to test the proposed approaches directly on experimental data and, where applicable, integrate them into experimental feedback loops. Such validation should account for finite sampling statistics, detector and calibration errors, parameter drift, device-specific noise, and discrepancies between the physical system and the model assumed during training. Further work should also examine how the methods behave as the Hilbert-space or Hamiltonian parameter dimension increases, without assuming that the successful extensions demonstrated here for small systems imply favorable asymptotic scaling. Addressing these issues would determine which of the methods developed in this dissertation can provide practical advantages in the characterization and control of solid-state quantum devices.

Bibliography

[1] Loss, D.; DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **1998**, *57*, 120–126.

[2] Vandersypen, L. M. K.; Bluhm, H.; Clarke, J. S.; Dzurak, A. S.; Ishihara, R.; Morello, A.; Reilly, D. J.; Schreiber, L. R.; Veldhorst, M. Interfacing spin qubits in quantum dots and donors-hot, dense, and coherent. *npj Quantum Information* **2017**, *3*, 34.

[3] Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79.

[4] Lutchyn, R. M.; Bakkers, E. P. A. M.; Kouwenhoven, L. P.; Krogstrup, P.; Marcus, C. M.; Oreg, Y. Majorana zero modes in superconductor–semiconductor heterostructures. *Nature Reviews Materials* **2018**, *3*, 52–68.

[5] van der Wiel, W. G.; De Franceschi, S.; Elzerman, J. M.; Fujisawa, T.; Tarucha, S.; Kouwenhoven, L. P. Electron transport through double quantum dots. *Rev. Mod. Phys.* **2002**, *75*, 1–22.

[6] Hradil, Z. Quantum-state estimation. *Phys. Rev. A* **1997**, *55*, R1561–R1564.

[7] James, D. F. V.; Kwiat, P. G.; Munro, W. J.; White, A. G. Measurement of qubits. *Phys. Rev. A* **2001**, *64*, 052312.

[8] Paris, M., Reháček, J., Eds. *Quantum State Estimation*, 1st ed.; Lecture Notes in Physics; Springer-Verlag Berlin Heidelberg: Berlin, Heidelberg, 2004; pp XIII, 520.

[9] Foulk, N. L.; Das Sarma, S. Theory of charge stability diagrams in coupled quantum dot qubits. *Phys. Rev. B* **2024**, *110*, 205428.

[10] Cha, M.-H.; Hwang, J. Quantum transport in a chain of quantum dots with inhomogeneous size distribution and manifestation of 1D Anderson localization. *Scientific Reports* **2020**, *10*, 16701.

[11] Zwolak, J. P.; Taylor, J. M. Colloquium: Advances in automation of quantum dot devices control. *Rev. Mod. Phys.* **2023**, *95*, 011006.

[12] Jozsa, R.; Linden, N. On the role of entanglement in quantum-computational speed-up. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2003**, *459*, 2011–2032.

[13] Vidal, G. Efficient Classical Simulation of Slightly Entangled Quantum Computations. *Phys. Rev. Lett.* **2003**, *91*, 147902.

[14] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.

[15] Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016; `http://www.deeplearningbook.org`.

[16] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L. u.; Polosukhin, I. Attention is All you Need. Advances in Neural Information Processing Systems. 2017.

[17] Horodecki, R.; Horodecki, P.; Horodecki, M.; Horodecki, K. Quantum entanglement. *Rev. Mod. Phys.* **2009**, *81*, 865–942.

[18] Torlai, G.; Mazzola, G.; Carrasquilla, J.; Troyer, M.; Melko, R.; Carleo, G. Neural-network quantum state tomography. *Nature Physics* **2018**, *14*, 447–450.

[19] Kalantre, S. S.; Zwolak, J. P.; Ragole, S.; Wu, X.; Zimmerman, N. M.; Stewart, M. D.; Taylor, J. M. Machine learning techniques for state recognition and auto-tuning in quantum dots. *npj Quantum Information* **2019**, *5*, 6.

[20] Koch, R.; van Driel, D.; Bordin, A.; Lado, J. L.; Greplova, E. Adversarial Hamiltonian learning of quantum dots in a minimal Kitaev chain. *Phys. Rev. Appl.* **2023**, *20*, 044081.

[21] Nayak, C.; Simon, S. H.; Stern, A.; Freedman, M.; Das Sarma, S. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **2008**, *80*, 1083–1159.

[22] Kitaev, A. Y. Unpaired Majorana fermions in quantum wires. *Physics-Uspekhi* **2001**, *44*, 131.

[23] Nielsen, M. A.; Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*; Cambridge University Press, 2010.

[24] Magesan, E.; Gambetta, J. M.; Emerson, J. Scalable and Robust Randomized Benchmarking of Quantum Processes. *Phys. Rev. Lett.* **2011**, *106*, 180504.

[25] Flammia, S. T.; Liu, Y.-K. Direct Fidelity Estimation from Few Pauli Measurements. *Phys. Rev. Lett.* **2011**, *106*, 230501.

[26] Dong, D.; Petersen, I. Quantum control theory and applications: a survey. *IET Control Theory & Applications* **2010**, *4*, 2651–2671.

[27] Jozsa, R. Fidelity for Mixed Quantum States. *Journal of Modern Optics* **1994**, *41*, 2315–2323.

[28] Bures, D. An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite W*-algebras. *Transactions of the American Mathematical Society* **1969**, *135*, 199–212.

[29] Bertlmann, R. A.; Krammer, P. Bloch vectors for qudits. *Journal of Physics A: Mathematical and Theoretical* **2008**, *41*, 235303.

[30] Bengtsson, I.; Życzkowski, K. *Geometry of Quantum States: An Introduction to Quantum Entanglement*; Cambridge University Press, 2017; Chapter Quantum entanglement.

[31] Quesada, N.; Al-Qasimi, A.; James, D. F. Quantum properties and dynamics of X states. *Journal of Modern Optics* **2012**, *59*, 1322–1329.

[32] Rau, A. R. P. Algebraic characterization of X-states in quantum information. *Journal of Physics A: Mathematical and Theoretical* **2009**, *42*, 412002.

[33] Yu, T.; Eberly, J. H. Evolution from entanglement to decoherence of bipartite mixed X states. *Quantum Info. Comput.* **2007**, *7*, 459–468.

[34] Ali, M.; Rau, A. R. P.; Alber, G. Quantum discord for two-qubit X states. *Phys. Rev. A* **2010**, *81*, 042105.

[35] Hashemi Rafsanjani, S. M.; Huber, M.; Broadbent, C. J.; Eberly, J. H. Genuinely multipartite concurrence of N-qubit X matrices. *Phys. Rev. A* **2012**, *86*, 062303.

[36] Vinjanampathy, S.; Rau, A. R. P. Generalized X states of N qubits and their symmetries. *Phys. Rev. A* **2010**, *82*, 032336.

[37] Gisin, N.; Ribordy, G.; Tittel, W.; Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **2002**, *74*, 145–195.

[38] Pirandola, S. *et al.* Advances in quantum cryptography. *Adv. Opt. Photon.* **2020**, *12*, 1012–1236.

[39] Demirel, B.; Weng, W.; Thalacker, C.; Hoban, M.; Barz, S. Correlations for computation and computation for correlations. *npj Quantum Information* **2021**, *7*, 29.

[40] Bennett, C. H.; Brassard, G.; Crépeau, C.; Jozsa, R.; Peres, A.; Wootters, W. K. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **1993**, *70*, 1895–1899.

[41] Bouwmeester, D.; Pan, J.-W.; Mattle, K.; Eibl, M.; Weinfurter, H.; Zeilinger, A. Experimental quantum teleportation. *Nature* **1997**, *390*, 575–579.

[42] Boschi, D.; Branca, S.; De Martini, F.; Hardy, L.; Popescu, S. Experimental Realization of Teleporting an Unknown Pure Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels. *Phys. Rev. Lett.* **1998**, *80*, 1121–1125.

[43] Roszak, K.; Korbicz, J. K. Purifying teleportation. *Quantum* **2023**, *7*, 923.

[44] Horodecki, M.; Horodecki, P.; Horodecki, R.; Oppenheim, J.; Sen(De), A.; Sen, U.; Synak-Radtke, B. Local versus nonlocal information in quantum-information theory: Formalism and phenomena. *Phys. Rev. A* **2005**, *71*, 062307.

[45] Modi, K.; Brodutch, A.; Cable, H.; Paterek, T.; Vedral, V. The classical-quantum boundary for correlations: Discord and related measures. *Rev. Mod. Phys.* **2012**, *84*, 1655–1707.

[46] Roszak, K.; Cywiński, Ł. The relation between the quantum discord and quantum teleportation: The physical interpretation of the transition point between different quantum discord decay regimes. *EPL (Europhysics Letters)* **2015**, *112*, 10002.

[47] Werner, R. F. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A* **1989**, *40*, 4277–4281.

[48] Plenio, M. B.; Virmani, S. An Introduction to Entanglement Measures. *Quantum Info. Comput.* **2007**, *7*, 1–51.

[49] Ollivier, H.; Zurek, W. H. Quantum Discord: A Measure of the Quantumness of Correlations. *Phys. Rev. Lett.* **2001**, *88*, 017901.

[50] Henderson, L.; Vedral, V. Classical, quantum and total correlations. *Journal of Physics A: Mathematical and General* **2001**, *34*, 6899.

[51] Gurvits, L. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences* **2004**, *69*, 448–484, Special Issue on STOC 2003.

[52] Wang, Y.-X.; Mu, L.-Z.; Vedral, V.; Fan, H. Entanglement Rényi α entropy. *Phys. Rev. A* **2016**, *93*, 022324.

[53] Wootters, W. K. In *Quantum Communication, Computing, and Measurement 2*; Kumar, P., D'Ariano, G. M., Hirota, O., Eds.; Springer US: Boston, MA, 2002; pp 69–73.

[54] Wootters, W. K. Entanglement of Formation of an Arbitrary State of Two Qubits. *Phys. Rev. Lett.* **1998**, *80*, 2245–2248.

[55] Plenio, M. B.; Virmani, S. S. In *Quantum Information and Coherence*; Andersson, E., Öhberg, P., Eds.; Springer International Publishing: Cham, 2014; pp 173–209.

[56] Peres, A. Separability Criterion for Density Matrices. *Phys. Rev. Lett.* **1996**, *77*, 1413–1415.

[57] Horodecki, M.; Horodecki, P.; Horodecki, R. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A* **1996**, *223*, 1–8.

[58] Vidal, G.; Werner, R. F. Computable measure of entanglement. *Phys. Rev. A* **2002**, *65*, 032314.

[59] Plenio, M. B. Logarithmic Negativity: A Full Entanglement Monotone That is not Convex. *Phys. Rev. Lett.* **2005**, *95*, 090503.

[60] Huang, Y. Computing quantum discord is NP-complete. *New Journal of Physics* **2014**, *16*, 033027.

[61] Dakić, B.; Vedral, V.; Brukner, Č. Necessary and sufficient condition for nonzero quantum discord. *Phys. Rev. Lett.* **2010**, *105*, 190502.

[62] Tufarelli, T.; MacLean, T.; Girolami, D.; Vasile, R.; Adesso, G. The geometric approach to quantum correlations: Computability versus reliability. *Journal of Physics A: Mathematical and Theoretical* **2013**, *46*, 275308.

[63] Huang, J.-H.; Wang, L.; Zhu, S.-Y. A new criterion for zero quantum discord. *New Journal of Physics* **2011**, *13*, 063045.

[64] Doherty, A. C.; Parrilo, P. A.; Spedalieri, F. M. Distinguishing Separable and Entangled States. *Phys. Rev. Lett.* **2002**, *88*, 187904.

[65] Hulpke, F.; Bruß, D. A two-way algorithm for the entanglement problem. *Journal of Physics A: Mathematical and General* **2005**, *38*, 5573–5579.

[66] Lu, S.; Huang, S.; Li, K.; Li, J.; Chen, J.; Lu, D.; Ji, Z.; Shen, Y.; Zhou, D.; Zeng, B. Separability-entanglement classifier via machine learning. *Phys. Rev. A* **2018**, *98*, 012315.

[67] Roik, J.; Bartkiewicz, K.; Černoch, A.; Lemr, K. Accuracy of Entanglement Detection via Artificial Neural Networks and Human-Designed Entanglement Witnesses. *Phys. Rev. Appl.* **2021**, *15*, 054006.

[68] Chen, Y.; Pan, Y.; Zhang, G.; Cheng, S. Detecting quantum entanglement with unsupervised learning. *Quantum Science and Technology* **2021**, *7*, 015005.

[69] Asif, N.; Khalid, U.; Khan, A.; Duong, T. Q.; Shin, H. Entanglement detection with artificial neural networks. *Scientific Reports* **2023**, *13*, 1562.

[70] Ureña, J.; Sojo, A.; Bermejo-Vega, J.; Manzano, D. Entanglement detection with classical deep neural networks. *Scientific Reports* **2024**, *14*, 18109.

[71] Girardin, A.; Brunner, N.; Kriváchy, T. Building separable approximations for quantum states via neural networks. *Phys. Rev. Res.* **2022**, *4*, 023238.

[72] Koutný, D.; Ginés, L.; Moczała-Dusanowska, M.; Höfling, S.; Schneider, C.; Predojević, A.; Ježek, M. Deep learning of quantum entanglement from incomplete measurements. *Science Advances* **2023**, *9*, eadd7131.

[73] Gao, J.; Qiao, L.-F.; Jiao, Z.-Q.; Ma, Y.-C.; Hu, C.-Q.; Ren, R.-J.; Yang, A.-L.; Tang, H.; Yung, M.-H.; Jin, X.-M. Experimental Machine Learning of Quantum States. *Phys. Rev. Lett.* **2018**, *120*, 240501.

[74] Luo, Y.-J.; Liu, J.-M.; Zhang, C. Detecting genuine multipartite entanglement via machine learning. *Phys. Rev. A* **2023**, *108*, 052424.

[75] Huet, H.; Ramesh, P. R.; Wein, S. C.; Coste, N.; Hilaire, P.; Somaschi, N.; Morassi, M.; Lemaître, A.; Sagnes, I.; Doty, M. F.; Krebs, O.; Lanco, L.; Fioretto, D. A.; Senellart, P. Deterministic and reconfigurable graph state generation with a single solid-state quantum emitter. *Nature Communications* **2025**, *16*, 4337.

[76] Li, B.; Economou, S. E.; Barnes, E. Photonic resource state generation from a minimal number of quantum emitters. *npj Quantum Information* **2022**, *8*, 11.

[77] Lüders, G. Über die Zustandsänderung durch den Meßprozeß. *Annalen der Physik* **1950**, *443*, 322–328.

[78] Busch, P.; Lahti, P. J.; Mittelstaedt, P. *The Quantum Theory of Measurement*, 2nd ed.; Lecture Notes in Physics Monographs; Springer: Berlin, Heidelberg, 2008; p 181.

[79] Penrose, R. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* **1955**, *51*, 406–413.

[80] Kullback, S.; Leibler, R. A. On Information and Sufficiency. *The Annals of Mathematical Statistics* **1951**, *22*, 79–86.

[81] Blume-Kohout, R. Optimal, reliable estimation of quantum states. *New Journal of Physics* **2010**, *12*, 043034.

[82] Gross, D.; Liu, Y.-K.; Flammia, S. T.; Becker, S.; Eisert, J. Quantum State Tomography via Compressed Sensing. *Phys. Rev. Lett.* **2010**, *105*, 150401.

[83] Riofrío, C. A.; Gross, D.; Flammia, S. T.; Monz, T.; Nigg, D.; Blatt, R.; Eisert, J. Experimental quantum compressed sensing for a seven-qubit system. *Nature Communications* **2017**, *8*, 15305.

[84] Huang, H.-Y.; Kueng, R.; Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nature Physics* **2020**, *16*, 1050–1057.

[85] Carrasquilla, J.; Torlai, G.; Melko, R. G.; Aolita, L. Reconstructing quantum states with generative models. *Nature Machine Intelligence* **2019**, *1*, 155–161.

[86] Xin, T.; Lu, S.; Cao, N.; Anikeeva, G.; Lu, D.; Li, J.; Long, G.; Zeng, B. Local-measurement-based quantum state tomography via neural networks. *npj Quantum Information* **2019**, *5*, 109.

[87] Quek, Y.; Fort, S.; Ng, H. K. Adaptive quantum state tomography with neural networks. *npj Quantum Information* **2021**, *7*, 105.

[88] Schmale, T.; Reh, M.; Gärttner, M. Efficient quantum state tomography with convolutional neural networks. *npj Quantum Information* **2022**, *8*, 115.

[89] Koutný, D.; Motka, L.; Hradil, Z. c. v.; Řeháček, J.; Sánchez-Soto, L. L. Neural-network quantum state tomography. *Phys. Rev. A* **2022**, *106*, 012409.

[90] Ahmed, S.; Sánchez Muñoz, C.; Nori, F.; Kockum, A. F. Quantum State Tomography with Conditional Generative Adversarial Networks. *Phys. Rev. Lett.* **2021**, *127*, 140502.

[91] Palmieri, A. M.; Müller-Rigat, G.; Srivastava, A. K.; Lewenstein, M.; Rajchel-Mieldzioć, G.; Płodzień, M. Enhancing quantum state tomography via resource-efficient attention-based neural networks. *Phys. Rev. Res.* **2024**, *6*, 033248.

[92] Barends, R. *et al.* Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **2014**, *508*, 500–503.

[93] Fowler, A. G.; Mariantoni, M.; Martinis, J. M.; Cleland, A. N. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **2012**, *86*, 032324.

[94] Litinski, D. A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery. *Quantum* **2019**, *3*, 128.

[95] Lahtinen, V. T.; Pachos, J. K. A short introduction to topological quantum computation. *SciPost Phys.* **2017**, *3*, 021.

[96] Aghaee, M. *et al.* Interferometric single-shot parity measurement in InAs–Al hybrid devices. *Nature* **2025**, *638*, 651–655.

[97] Legg, H. F. On the robustness of topological gap detection via transport. *Nature* **2026**, *654*, E22–E26.

[98] Stanescu, T. D. *Introduction to Topological Quantum Matter & Quantum Computation*, 2nd ed.; CRC Press, 2024.

[99] Alicea, J. New directions in the pursuit of Majorana fermions in solid state systems. *Reports on Progress in Physics* **2012**, *75*, 076501.

[100] Sarma, S. D.; Freedman, M.; Nayak, C. Majorana zero modes and topological quantum computation. *npj Quantum Information* **2015**, *1*, 15001.

[101] Flensberg, K.; von Oppen, F.; Stern, A. Engineered platforms for topological superconductivity and Majorana zero modes. *Nature Reviews Materials* **2021**, *6*, 944–958.

[102] Beenakker, C. Search for Majorana Fermions in Superconductors. *Annual Review of Condensed Matter Physics* **2013**, *4*, 113–136.

[103] Prada, E.; San-Jose, P.; de Moor, M. W. A.; Geresdi, A.; Lee, E. J. H.; Klinovaja, J.; Loss, D.; Nygård, J.; Aguado, R.; Kouwenhoven, L. P. From Andreev to Majorana bound states in hybrid superconductor–semiconductor nanowires. *Nature Reviews Physics* **2020**, *2*, 575–594.

[104] Pan, H.; Das Sarma, S. Physical mechanisms for zero-bias conductance peaks in Majorana nanowires. *Phys. Rev. Res.* **2020**, *2*, 013377.

[105] Das Sarma, S.; Pan, H. Disorder-induced zero-bias peaks in Majorana nanowires. *Phys. Rev. B* **2021**, *103*, 195158.

[106] Zhang, H. *et al.* Quantized Majorana conductance. *Nature* **2018**, *556*, 74–79, Retracted article.

[107] Zhang, H. *et al.* Retraction Note: Quantized Majorana conductance. *Nature* **2021**, *591*, E30–E30.

[108] Aguado, R. Majorana quasiparticles in condensed matter. *La Rivista del Nuovo Cimento* **2017**, *40*, 523–593.

[109] Rainis, D.; Loss, D. Majorana qubit decoherence by quasiparticle poisoning. *Phys. Rev. B* **2012**, *85*, 174533.

[110] Ricco, L. S.; Marques, Y.; Sanches, J. E.; Shelykh, I. A.; Seridonio, A. C. Interaction induced hybridization of Majorana zero modes in a coupled quantum-dot–superconducting-nanowire hybrid system. *Phys. Rev. B* **2020**, *102*, 165104.

[111] Albrecht, S. M.; Higginbotham, A. P.; Madsen, M.; Kuemmeth, F.; Jespersen, T. S.; Nygård, J.; Krogstrup, P.; Marcus, C. M. Exponential protection of zero modes in Majorana islands. *Nature* **2016**, *531*, 206–209.

[112] Liu, J.; Song, J.; Sun, Q.-F.; Xie, X. C. Even-odd interference effect in a topological superconducting wire. *Phys. Rev. B* **2017**, *96*, 195307.

[113] Moore, C.; Stanescu, T. D.; Tewari, S. Two-terminal charge tunneling: Disentangling Majorana zero modes from partially separated Andreev bound states in semiconductor-superconductor heterostructures. *Phys. Rev. B* **2018**, *97*, 165302.

[114] Mourik, V.; Zuo, K.; Frolov, S. M.; Plissard, S. R.; Bakkers, E. P. A. M.; Kouwenhoven, L. P. Signatures of Majorana Fermions in Hybrid Superconductor-Semiconductor Nanowire Devices. *Science* **2012**, *336*, 1003–1007.

[115] Fu, J.-B., *et al.* Experimental review on Majorana zero-modes in hybrid nanowires. *Sci. China Phys. Mech. Astron.* **2021**, *64*, 287431.

[116] Yazdani, A. Hunting for Majoranas. *Science* **2023**, *379*, 1135–1136.

[117] Higginbotham, A. P.; Albrecht, S. M.; Kiršanskas, G.; Chang, W.; Kuemmeth, F.; Krogstrup, P.; Jespersen, T. S.; Nygård, J.; Flensberg, K.; Marcus, C. M. Parity lifetime of bound states in a proximitized semiconductor nanowire. *Nature Physics* **2015**, *11*, 1017–1021.

[118] van Loo, N.; Zatelli, F.; Steffensen, G. O.; Roovers, B.; Wang, G.; Van Caekenberghe, T.; Bordin, A.; van Driel, D.; Zhang, Y.; Huisman, W. D.; Badawy, G.; Bakkers, E. P. A. M.; Mazur, G. P.; Aguado, R.; Kouwenhoven, L. P. Single-shot parity readout of a minimal Kitaev chain. *Nature* **2026**, *650*, 334–339.

[119] Oreg, Y.; Refael, G.; von Oppen, F. Helical Liquids and Majorana Bound States in Quantum Wires. *Phys. Rev. Lett.* **2010**, *105*, 177002.

[120] Lutchyn, R. M.; Sau, J. D.; Das Sarma, S. Majorana Fermions and a Topological Phase Transition in Semiconductor-Superconductor Heterostructures. *Phys. Rev. Lett.* **2010**, *105*, 077001.

[121] Nadj-Perge, S.; Drozdov, I. K.; Bernevig, B. A.; Yazdani, A. Proposal for realizing Majorana fermions in chains of magnetic atoms on a superconductor. *Phys. Rev. B* **2013**, *88*, 020407.

[122] Nadj-Perge, S.; Drozdov, I. K.; Li, J.; Chen, H.; Jeon, S.; Seo, J.; MacDonald, A. H.; Bernevig, B. A.; Yazdani, A. Observation of Majorana fermions in ferromagnetic atomic chains on a superconductor. *Science* **2014**, *346*, 602–607.

[123] Hegde, S. S.; Yue, G.; Wang, Y.; Huemiller, E.; Van Harlingen, D.; Vishveshwara, S. A topological Josephson junction platform for creating, manipulating, and braiding Majorana bound states. *Annals of Physics* **2020**, *423*, 168326.

[124] Fu, L.; Kane, C. L. Superconducting Proximity Effect and Majorana Fermions at the Surface of a Topological Insulator. *Phys. Rev. Lett.* **2008**, *100*, 096407.

[125] Sang, L., *et al.* Majorana zero modes in iron-based superconductors. *Matter* **2022**, *5*, 3387–3417.

[126] Leijnse, M.; Flensberg, K. Introduction to topological superconductivity and Majorana fermions. *Semiconductor Science and Technology* **2012**, *27*, 124003.

[127] Sau, J. D.; Sarma, S. D. Realizing a robust practical Majorana chain in a quantum-dot-superconductor linear array. *Nature Communications* **2012**, *3*, 964.

[128] Dvir, T. *et al.* Realization of a minimal Kitaev chain in coupled quantum dots. *Nature* **2023**, *614*, 445–450.

[129] Bordin, A. *et al.* Enhanced Majorana stability in a three-site Kitaev chain. *Nature Nanotechnology* **2025**, *20*, 726–731.

[130] Mohseni, M.; Allami, H.; Miravet, D.; Gayowsky, D. J.; Korkusinski, M.; Hawrylak, P. Majorana Excitons in a Kitaev Chain of Semiconductor Quantum Dots in a Nanowire. *Nanomaterials* **2023**, *13*.

[131] Ezawa, M. Even-odd effect on robustness of Majorana edge states in short Kitaev chains. *Phys. Rev. B* **2024**, *109*, L161404.

[132] Agarwal, K.; Rai, H.; Mondal, S. Quantum dots: an overview of synthesis, properties, and applications. *Materials Research Express* **2023**, *10*, 062001.

[133] de Arquer, F. P. G.; Talapin, D. V.; Klimov, V. I.; Arakawa, Y.; Bayer, M.; Sargent, E. H. Semiconductor quantum dots: Technological progress and future challenges. *Science* **2021**, *373*, eaaz8541.

[134] Souto, R. S.; Baran, V. V.; Nitsch, M.; Maffi, L.; Paaske, J.; Leijnse, M.; Burrello, M. Majorana modes in quantum dots coupled via a floating superconducting island. *Phys. Rev. B* **2025**, *111*, 174501.

[135] Fulga, I. C.; Haim, A.; Akhmerov, A. R.; Oreg, Y. Adaptive tuning of Majorana fermions in a quantum dot chain. *New J. Phys.* **2013**, *15*, 045020.

[136] Anderson, P. W. Localized Magnetic States in Metals. *Phys. Rev.* **1961**, *124*, 41–53.

[137] Hewson, A. C. *The Kondo Problem to Heavy Fermions*; Cambridge Studies in Magnetism; Cambridge University Press, 1993.

[138] Thamm, M.; Rosenow, B. Conductance based machine learning of optimal gate voltages for disordered Majorana wires. *Phys. Rev. B* **2024**, *109*, 045132.

[139] Hanson, R.; Kouwenhoven, L. P.; Petta, J. R.; Tarucha, S.; Vandersypen, L. M. K. Spins in few-electron quantum dots. *Rev. Mod. Phys.* **2007**, *79*, 1217–1265.

[140] van Diepen, C. J.; Eendebak, P. T.; Buijtendorp, B. T.; Mukhopadhyay, U.; Fujita, T.; Reichl, C.; Wegscheider, W.; Vandersypen, L. M. K. Automated tuning of inter-dot tunnel coupling in double quantum dots. *Applied Physics Letters* **2018**, *113*, 033101.

[141] Volk, C.; Zwerver, A. M. J.; Mukhopadhyay, U.; Eendebak, P. T.; van Diepen, C. J.; Dehollain, J. P.; Hensgens, T.; Fujita, T.; Reichl, C.; Wegscheider, W.; Vandersypen, L. M. K. Loading a quantum-dot based "Qubyte" register. *npj Quantum Information* **2019**, *5*, 29.

[142] Deacon, R. S.; Tanaka, Y.; Oiwa, A.; Sakano, R.; Yoshida, K.; Shibata, K.; Hirakawa, K.; Tarucha, S. Tunneling Spectroscopy of Andreev Energy Levels in a Quantum Dot Coupled to a Superconductor. *Phys. Rev. Lett.* **2010**, *104*, 076805.

[143] Souto, R. S.; Martín-Rodero, A.; Yeyati, A. L. Andreev Bound States Formation and Quasiparticle Trapping in Quench Dynamics Revealed by Time-Dependent Counting Statistics. *Phys. Rev. Lett.* **2016**, *117*, 267701.

[144] Donarini, A.; Grifoni, M. *Quantum Transport in Interacting Nanojunctions: A Density Matrix Approach*; Springer Cham, 2024.

[145] Christiansen, T. J.; Zworski, M. A mathematical formulation of the Mahaux–Weidenmüller formula for the scattering matrix. *Journal of Physics A: Mathematical and Theoretical* **2009**, *42*, 415202.

[146] Meir, Y.; Wingreen, N. S. Landauer formula for the current through an interacting electron region. *Phys. Rev. Lett.* **1992**, *68*, 2512–2515.

[147] Baart, T. A.; Eendebak, P. T.; Reichl, C.; Wegscheider, W.; Vandersypen, L. M. K. Computer-automated tuning of semiconductor double quantum dots into the single-electron regime. *Applied Physics Letters* **2016**, *108*, 213104.

[148] Moon, H.; Lennon, D. T.; Kirkpatrick, J.; van Esbroeck, N. M.; Camenzind, L. C.; Yu, L.; Vigneau, F.; Zumbühl, D. M.; Briggs, G. A. D.; Osborne, M. A.; Sejdinovic, D.; Laird, E. A.; Ares, N. Machine learning enables completely automatic tuning of a quantum device faster than human experts. *Nature Communications* **2020**, *11*, 4161.

[149] Czischek, S.; Yon, V.; Genest, M.-A.; Roux, M.-A.; Rochette, S.; Camirand Lemyre, J.; Moras, M.; Pioro-Ladrière, M.; Drouin, D.; Beilliard, Y.; Melko, R. G. Miniaturizing neural networks for charge state autotuning in quantum dots. *Machine Learning: Science and Technology* **2021**, *3*, 015001.

[150] Ziegler, J.; McJunkin, T.; Joseph, E.; Kalantre, S. S.; Harpt, B.; Savage, D.; Lagally, M.; Eriksson, M.; Taylor, J. M.; Zwolak, J. P. Toward Robust Autotuning of Noisy Quantum dot Devices. *Phys. Rev. Appl.* **2022**, *17*, 024069.

[151] Yon, V.; Galaup, B.; Rohrbacher, C.; Rivard, J.; Godfrin, C.; Li, R.; Kubicek, S.; De Greve, K.; Gaudreau, L.; Dupont-Ferrier, E.; Beilliard, Y.; Melko, R. G.; Drouin, D. Robust quantum dots charge autotuning using neural network uncertainty. *Machine Learning: Science and Technology* **2024**, *5*, 045034.

[152] Severin, B. *et al.* Cross-architecture tuning of silicon and SiGe-based quantum devices using machine learning. *Scientific Reports* **2024**, *14*, 17281.

[153] Yon, V.; Galaup, B.; Rohrbacher, C.; Rivard, J.; Morel, A.; Leclerc, D.; Godfrin, C.; Li, R.; Kubicek, S.; Greve, K. D.; et al., Experimental Online Quantum Dots Charge Autotuning Using Neural Networks. *Nano Letters* **2025**,

[154] Taylor, J. R.; Sau, J. D.; Das Sarma, S. Machine Learning the Disorder Landscape of Majorana Nanowires. *Phys. Rev. Lett.* **2024**, *132*, 206602.

[155] Cheng, M.; Okabe, R.; Chotrattanapituk, A.; Li, M. Machine learning detection of Majorana zero modes from zero-bias peak measurements. *Matter* **2024**, *7*, 2507–2520.

[156] Taylor, J. R.; Das Sarma, S. Vision transformer based deep learning of topological indicators in Majorana nanowires. *Phys. Rev. B* **2025**, *111*, 104208.

[157] Taylor, J. R.; Das Sarma, S. Mitigating disorder and optimizing topological indicators with vision-transformer-based neural networks in Majorana nanowires. *Phys. Rev. B* **2025**, *112*, L041110.

[158] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, 2009.

[159] Bishop, C. M. *Pattern Recognition and Machine Learning*; Information Science and Statistics; Springer: New York, NY, 2006.

[160] ichi Amari, S. Backpropagation and stochastic gradient descent method. *Neurocomputing* **1993**, *5*, 185–196.

[161] Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. **2017**,

[162] Su, W.; Boyd, S.; Candès, E. J. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *J. Mach. Learn. Res.* **2016**, *17*, 5312–5354.

[163] Pascanu, R.; Dauphin, Y. N.; Ganguli, S.; Bengio, Y. On the saddle point problem for non-convex optimization. 2014; `https://arxiv.org/abs/1405.4604`.

[164] Chizat, L.; Bach, F. On the global convergence of gradient descent for over-parameterized models using optimal transport. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA, 2018; pp 3040–3050.

[165] Barrett, D. G. T.; Dherin, B. Implicit Gradient Regularization. 2022; `https://arxiv.org/abs/2009.11162`, Published as a conference paper at ICLR 2021.

[166] Peleg, A.; Hein, M. Bias of Stochastic Gradient Descent or the Architecture: Disentangling the Effects of Overparameterization of Neural Networks. Proceedings of the 41st International Conference on Machine Learning. 2024; pp 40154–40184.

[167] Wu, J.; Bartlett, P.; Telgarsky, M.; Yu, B. Benefits of Early Stopping in Gradient Descent for Overparameterized Logistic Regression. Proceedings of the 42nd International Conference on Machine Learning. 2025; pp 67081–67110.

[168] Jacobs, T.; Zhou, C.; Burkholz, R. Mirror, Mirror of the Flow: How Does Regularization Shape Implicit Bias? Proceedings of the 42nd International Conference on Machine Learning. 2025; pp 26673–26699.

[169] Sutton, R. S.; Barto, A. G. *Reinforcement learning : an introduction*, 2nd ed.; Adaptive computation and machine learning; The MIT Press: Cambridge, Massachusetts, 2018.

[170] Grohs, P., Kutyniok, G., Eds. *Mathematical Aspects of Deep Learning*; Cambridge University Press, 2022.

[171] Kawaguchi, K.; Bengio, Y.; Kaelbling, L. P. In *Mathematical Aspects of Deep Learning*; Grohs, P., Kutyniok, G., Eds.; Cambridge University Press, 2022; pp 112–148.

[172] Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences* **2019**, *116*, 15849–15854.

[173] Nakkiran, P.; Kaplun, G.; Bansal, Y.; Yang, T.; Barak, B.; Sutskever, I. Deep Double Descent: Where Bigger Models and More Data Hurt. International Conference on Learning Representations. 2020.

[174] Shapley, L. S. In *Contributions to the Theory of Games*; Kuhn, H. W., Tucker, A. W., Eds.; Princeton University Press, 1953; Vol. 2; pp 307–317.

[175] Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA, 2017; pp 4768–4777.

[176] Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.

[177] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* **1989**, *2*, 303–314.

[178] Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; Chen, T. Recent advances in convolutional neural networks. *Pattern Recognition* **2018**, *77*, 354–377.

[179] Carrasquilla, J.; Melko, R. G. Machine learning phases of matter. *Nature Physics* **2017**, *13*, 431–434.

[180] Ch'ng, K.; Carrasquilla, J.; Melko, R. G.; Khatami, E. Machine Learning Phases of Strongly Correlated Fermions. *Phys. Rev. X* **2017**, *7*, 031038.

[181] Tanaka, A.; Tomiya, A. Detection of Phase Transition via Convolutional Neural Networks. *Journal of the Physical Society of Japan* **2017**, *86*, 063001.

[182] Broecker, P.; Carrasquilla, J.; Melko, R. G.; Trebst, S. Machine learning quantum phases of matter beyond the fermion sign problem. *Scientific Reports* **2017**, *7*, 8823.

[183] Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; Shah, R. Signature verification using a "Siamese" time delay neural network. Proceedings of the 7th International Conference on Neural Information Processing Systems. San Francisco, CA, USA, 1993; pp 737–744.

[184] Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). 2006; pp 1735–1742.

[185] Chicco, D. *Artificial Neural Networks*; Springer, 2021; pp 73–94.

[186] Bronstein, M. M.; Bruna, J.; Cohen, T.; Veličković, P. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. 2021; `https://arxiv.org/abs/2104.13478`.

[187] Wetzel, S. J.; Melko, R. G.; Scott, J.; Panju, M.; Ganesh, V. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *Phys. Rev. Res.* **2020**, *2*, 033499.

[188] Patel, Z.; Merali, E.; Wetzel, S. J. Unsupervised learning of Rydberg atom array phase diagram with Siamese neural networks. *New Journal of Physics* **2022**, *24*, 113021.

[189] Shen, J.; Chen, S.; Liu, F.; Li, W.; Liu, Y. Learning phase transitions by siamese neural network. *Physica A: Statistical Mechanics and its Applications* **2025**, *676*, 130857.

[190] Xiao, T.; Huang, J.; Li, H.; Fan, J.; Zeng, G. Intelligent certification for quantum simulators via machine learning. *npj Quantum Information* **2022**, *8*, 138.

[191] An, J.; Cho, S. Variational autoencoder based anomaly detection using reconstruction probability. *Special lecture on IE* **2015**, *2*, 1–18.

[192] Li, P.; Pei, Y.; Li, J. A comprehensive survey on design and application of autoencoder in deep learning. *Applied Soft Computing* **2023**, *138*, 110176.

[193] Wetzel, S. J. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Phys. Rev. E* **2017**, *96*, 022140.

[194] Iten, R.; Metger, T.; Wilming, H.; del Rio, L.; Renner, R. Discovering Physical Concepts with Neural Networks. *Phys. Rev. Lett.* **2020**, *124*, 010508.

[195] Ng, K.-K.; Yang, M.-F. Unsupervised learning of phase transitions via modified anomaly detection with autoencoders. *Phys. Rev. B* **2023**, *108*, 214428.

[196] Medsker, L. R., Jain, L. C., Eds. *Recurrent Neural Networks: Design and Applications*; CRC Press, 2001.

[197] Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* **1994**, *5*, 157–166.

[198] Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Computation* **1997**, *9*, 1735–1780.

[199] Vlachas, P. R.; Byeon, W.; Wan, Z. Y.; Sapsis, T. P.; Koumoutsakos, P. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *474*, 20170844.

[200] Hibat-Allah, M.; Ganahl, M.; Hayward, L. E.; Melko, R. G.; Carrasquilla, J. Recurrent neural network wave functions. *Phys. Rev. Res.* **2020**, *2*, 023358.

[201] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S., *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations. 2021.

[202] Raissi, M.; Perdikaris, P.; Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* **2019**, *378*, 686–707.

[203] Cuomo, S.; Di Cola, V. S.; Giampaolo, F.; Rozza, G.; Raissi, M.; Piccialli, F. Scientific Machine Learning Through Physics-Informed Neural Networks: Where we are and What's Next. *Journal of Scientific Computing* **2022**, *92*, 88.

[204] Zubatiuk, T.; Nebgen, B.; Lubbers, N.; Smith, J. S.; Zubatyuk, R.; Zhou, G.; Koh, C.; Barros, K.; Isayev, O.; Tretiak, S. Machine learned Hückel theory: Interfacing physics and deep neural networks. *The Journal of Chemical Physics* **2021**, *154*, 244108.

[205] Gu, Q.; Zhouyin, Z.; Pandey, S. K.; Zhang, P.; Zhang, L.; E, W. Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy. *Nature Communications* **2024**, *15*, 6772.

[206] Schwade, M.; Zhang, S.; Vonhoff, F.; Delgado, F. P.; Egger, D. A. Physics-informed Hamiltonian learning for large-scale optoelectronic property prediction. *Nature Communications* **2026**, *17*, 2652.

[207] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90.

[208] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham, 2015; pp 234–241.

[209] Xing, F.; Xie, Y.; Su, H.; Liu, F.; Yang, L. Deep Learning in Microscopy Image Analysis: A Survey. *IEEE Transactions on Neural Networks and Learning Systems* **2018**, *29*, 4550–4568.

[210] Huang, Y.; Wang, S.; Guan, Y.; Maier, A. Limited angle tomography for transmission X-ray microscopy using deep learning. *Journal of Synchrotron Radiation* **2020**, *27*, 477–485.

[211] Toshev, A.; Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014; pp 1653–1660.

[212] Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C. C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. Computer Vision – ECCV 2018 Workshops. Cham, 2019; pp 63–79.

[213] Nehme, E.; Weiss, L. E.; Michaeli, T.; Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **2018**, *5*, 458–464.

[214] Zhao, Y.; Zhang, J.; Zong, C. Transformer: A General Framework from Machine Translation to Others. *Machine Intelligence Research* **2023**, *20*, 514–538.

[215] Zhou, W.; Zhang, S.; Gu, Y.; Chen, M.; Poon, H. UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition. The Twelfth International Conference on Learning Representations. 2024.

[216] Campillos-Llanos, L.; Valverde-Mateos, A.; Capllonch-Carrión, A.; González-Quevedo, D.; López-Urbán, M. R.; Hernando-Tundidor, M. S.; Zakhir-Puig, S.; Heras, J. Transformer-based relation extraction and concept normalization using an annotated clinical trials corpus. *Scientific Data* **2026**, *13*, 280.

[217] Nguyen, D. Q.; Vu, T.; Tuan Nguyen, A. BERTweet: A pre-trained language model for English Tweets. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Online, 2020; pp 9–14.

[218] Karpukhin, V.; Oguz, B.; Min, S.; Lewis, P.; Wu, L.; Edunov, S.; Chen, D.; Yih, W.-t. Dense Passage Retrieval for Open-Domain Question Answering. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020; pp 6769–6781.

[219] Taylor, R.; Kardas, M.; Cucurull, G.; Scialom, T.; Hartshorn, A.; Saravia, E.; Poulton, A.; Kerkez, V.; Stojnic, R. Galactica: A Large Language Model for Science. 2022; `https://arxiv.org/abs/2211.09085`.

[220] Salinas, D.; Flunkert, V.; Gasthaus, J.; Januschowski, T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting* **2020**, *36*, 1181–1191.

[221] Lim, B.; Arık, S. Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* **2021**, *37*, 1748–1764.

[222] Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421.

[223] Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **2018**, *148*, 241722.

[224] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

[225] Mao, Y.; He, Q.; Zhao, X. Designing complex architectured materials with generative adversarial networks. *Science Advances* **2020**, *6*, eaaz4169.

[226] Dan, Y.; Zhao, Y.; Li, X.; Li, S.; Hu, M.; Hu, J. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials. *npj Computational Materials* **2020**, *6*, 84.

[227] Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602–606.

[228] Pfau, D.; Spencer, J. S.; Matthews, A. G. D. G.; Foulkes, W. M. C. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Res.* **2020**, *2*, 033429.

[229] Bukov, M.; Day, A. G. R.; Sels, D.; Weinberg, P.; Polkovnikov, A.; Mehta, P. Reinforcement Learning in Different Phases of Quantum Control. *Physical Review X* **2018**, *8*, 031086.

[230] Cuoco, E.; Cavaglià, M.; Heng, I. S.; Keitel, D.; Messenger, C. Applications of machine learning in gravitational-wave research with current interferometric detectors. *Living Reviews in Relativity* **2025**, *28*, 2.

[231] Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

[232] Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Science Advances* **2017**, *3*, e1701816.

[233] Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications* **2022**, *13*, 2453.

[234] Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G. E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence* **2021**, *3*, 218–229.

[235] Kovachki, N.; Li, Z.; Liu, B.; Azizzadenesheli, K.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Neural operator: learning maps between function spaces with applications to PDEs. *J. Mach. Learn. Res.* **2023**, *24*, 1–97.

[236] Karpov, P. I.; Huang, C.; Sitdikov, I.; Fryer, C. L.; Woosley, S.; Pilania, G. Physics-informed Machine Learning for Modeling Turbulence in Supernovae. *The Astrophysical Journal* **2022**, *940*, 26.

[237] Radovic, A.; Williams, M.; Rousseau, D.; Kagan, M.; Bonacorsi, D.; Himmel, A.; Aurisano, A.; Terao, K.; Wongjirad, T. Machine learning at the energy and intensity frontiers of particle physics. *Nature* **2018**, *560*, 41–48.

[238] Lewenstein, M.; Kraus, B.; Cirac, J. I.; Horodecki, P. Optimization of entanglement witnesses. *Phys. Rev. A* **2000**, *62*, 052310.

[239] Horodecki, P. Separability criterion and inseparable mixed states with positive partial transposition. *Physics Letters A* **1997**, *232*, 333–339.

[240] Bennett, C. H.; DiVincenzo, D. P.; Mor, T.; Shor, P. W.; Smolin, J. A.; Terhal, B. M. Unextendible Product Bases and Bound Entanglement. *Phys. Rev. Lett.* **1999**, *82*, 5385–5388.

[241] Acín, A.; Bruß, D.; Lewenstein, M.; Sanpera, A. Classification of Mixed Three-Qubit States. *Phys. Rev. Lett.* **2001**, *87*, 040401.

Publications

PHYSICAL REVIEW A **109**, 022405 (2024)

Data-driven criteria for quantum correlations

Mateusz Krawczyk,[1] Jarosław Pawłowski,[1] Maciej M. Maśka,[1] and Katarzyna Roszak[2]
[1]*Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*
[2]*Institute of Physics (FZU), Czech Academy of Sciences, Na Slovance 2, 182 00 Prague, Czech Republic*

(Received 4 August 2023; accepted 9 January 2024; published 5 February 2024)

We build a machine learning model to detect correlations in a three-qubit system using a neural network trained in an unsupervised manner on randomly generated states. The network is forced to recognize separable states, and correlated states are detected as anomalies. Quite surprisingly, we find that the proposed detector performs much better at distinguishing a weaker form of quantum correlations, namely, the quantum discord, than entanglement. In fact, it has a tendency to grossly overestimate the set of entangled states even at the optimal threshold for entanglement detection, while it underestimates the set of discordant states to a much lesser extent. In order to illustrate the nature of states classified as quantum correlated, we construct a diagram containing various types of states—entangled, as well as separable, both discordant and nondiscordant. We find that the near-zero value of the recognition loss reproduces the shape of the nondiscordant separable states with high accuracy, especially considering the nontrivial shape of this set on the diagram. The network architecture is designed carefully: it preserves separability, and its output is equivariant with respect to qubit permutations. We show that the choice of architecture is important to get the highest detection accuracy, much better than for a baseline model that just utilizes a partial trace operation.

DOI: 10.1103/PhysRevA.109.022405

I. INTRODUCTION

Quantum correlations play a pivotal role in near-future quantum technologies, from quantum cryptography [1,2], to networks [3,4], to computation [5,6]. The source of their usefulness is that they allow for the state of one subsystem to be influenced by operations and measurements on the other, which yields effects that cannot be reproduced by classical physics. A good example here is teleportation [7–10], which allows for the transfer of an unknown quantum state between distant locations by means of local operations and measurements, and the transfer of classical information, due to a maximally entangled state being shared by two parties.

In quantum information theory, there is a long-standing question about quantum correlations in mixed states [11–13]. For pure states, any correlations between subsystems are quantum in their nature (entanglement), and the set of separable (not entangled) states is equivalent to the set of product states. In the case of mixed states, the fact that a density matrix cannot be written in product form does not guarantee the existence of quantum correlations, since this formalism allows for the description of classical (purely statistical) correlations, i.e., $\hat{\rho}_{AB} = 1/2(|00\rangle\langle 00| + |11\rangle\langle 11|)$, whereas a product state contains no correlations whatsoever.

One definition of the set of states that contain no quantum correlations, namely the set of mixed separable states, is given by [14–16]

$$\hat{\rho}_{AB}^{\mathrm{sep}} = \sum_i p_i |\psi_i\rangle_{AA}\langle\psi_i| \otimes |\phi_i\rangle_{BB}\langle\phi_i|, \tag{1}$$

where p_i are probabilities and the indices A and B label the subsystems. It is relevant to note here that there is no limitation on states $|\psi_i\rangle_A$ or $|\phi_i\rangle_B$, which do not have to form a basis. By this definition, a bipartite state is separable if and only if it can be created from a product state using only local operations and classical communication. The quantification of entanglement (for mixed states, entangled states are defined as not separable) is classed as an NP-complete problem [17], and it does not simplify even when a state only needs to be qualified as separable or entangled.

A state with no quantum correlations can also be defined as a state for which no information about one subsystem can be learned by measurements performed on the other subsystem. Such states are said to contain zero *discord* [18–20], and are defined by Eq. (1) with the added constraints that the states $|\psi_i\rangle_A$ must be part of an orthonormal basis, and the same for states $|\phi_i\rangle_B$. Quantifying the amount of discord in a given state is also NP-complete [21], but it is harder than quantifying entanglement for small systems [22–24]. On the other hand, simply qualifying a state as discordant or not is a straightforward task [22,25]. The set of nondiscordant states is a subset of separable states. Currently, entanglement is the accepted, standard metric for quantum correlations, yet it should be noted that discord has one important advantage over entanglement, namely that it has a much stronger relationship to measurable quantities [26], while the detection of entanglement generally requires quantum tomography [27,28].

In the following, we would like to go beyond the typical scheme of describing the correlations in quantum states using analytical metrics and propose the study of quantum correlations using machine learning (ML) models forcing them to learn similarities between different classes of quantum states. We use neural networks (NNs) which have proven their usefulness in many applications, such as computer vision or signal processing [29,30], but recently also in quantum physics, e.g., in many-body theory [31–34], in entanglement

 ©2024 American Physical Society

modeling [35–40], or in state reconstruction [41]. However, in our approach, we carefully design the model architecture to be separability preserving and qubit permutations equivariant, which is part of the general trend of building NN models that respect symmetries of a given problem, for example, being rotation and translation group equivariant [42–45].

We present results where the ML model is first trained to distinguish between separable and entangled states (on both mixed and pure states) and then tested on similar sets. These sets (both in the training and testing phases), apart from entangled states, contain different variants of separable states, with special emphasis lain on discordant separable states. While the ML model trained only on pure states can only distinguish between product and nonproduct mixed states, hence signifying correlations, but not being able to discriminate between classical, quantum-classical (discord), and quantum-quantum (entanglement) correlations, quite surprisingly, we find that the performance of the separator trained also on mixed states is much better at signifying quantum discord than entanglement. This corroborates the intuition that a type of quantum correlations which are defined on the properties of measurement outcomes, as opposed to correlations defined on specifics of state preparation, should be more evident in a given state, even though the ML model was not trained directly on nondiscordant states (but on separable ones).

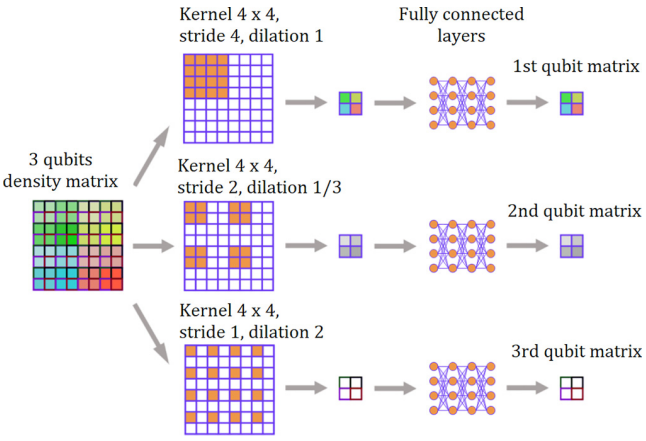


FIG. 1. The three-qubit neural network *separator* model. The input density matrix is convolved separately with three convolution layers, each with 4×4 kernels that work in parallel. The *stride* and *dilation* parameters defining these convolution layers are arranged in such a way that they operate on each qubit subspace separately (details can be found in the Appendix A). After the convolution layers, shape-preserving fully connected layers are applied independently for each qubit matrix.

II. NEURAL NETWORK MODEL

Since the goal of using the neural network is not only to classify the discordance or entanglement of states, but also to produce a similarity measure, we propose to train the model in the unsupervised anomaly detector scheme [46]. Anomaly detection is a standard problem in ML where a model is trained to reconstruct normal data and then fails when trying to reconstruct abnormal data (anomaly). The reconstruction error, i.e., the model "loss function," measures the similarity of a given unknown sample to the training data, and if it is greater than some assumed threshold, then the sample is treated as an anomaly. In our case, anomalies are states that are quantum correlated. The model is trained to recognize separable states of three qubits. We study three-qubit states to have a nontrivial yet manageable Hilbert space, keeping in mind that the proposed model construction can be easily extended to larger qubit registers.

To build the model, we use convolutional NNs (with a discrete convolution operation [47–49]) as they are commonly applied to build robust representations of correlations on data on regular grids such as matrices. First, we need to organize the convolution layers, so that the resulting NN can separate the input density matrix into single-qubit matrices. The constructed NN should preserve pure separable (product) states by definition and also not change the separability of mixed states. A closer look at the model scheme presented in Fig. 1 shows that the convolution layers are organized in a way that is a generalization of the partial trace operation, in the sense that the kernel simplified to the identity matrix would calculate the partial trace over the remaining qubits. Such an architecture implies qubit permutation equivariance, meaning that permutation of qubits in the input matrix gives the same permutation of the reconstructed state and preserves separability of any reconstructed state.

Additionally, to extend the model complexity we include four fully connected (FC) layers with nonlinear activation functions (multilayer perceptron) after the convolutional ones. This extension has a physical justification that will become clear later. Importantly, to preserve the separability property, these layers are applied independently to each block that represents the qubit matrix. Moreover, to allow the neural network to generalize and reconstruct also separable mixed states, more sets of convolutions, i.e., *channels*, have been used, giving a separable density matrix that is a sum over N_K channels. Details can be found in the Appendix A. The reconstructed density matrix is given by

$$\hat{\rho}_{ABC} = \frac{1}{N}\sum_{i=1}^{N_K} \hat{\rho}_{Ai} \otimes \hat{\rho}_{Bi} \otimes \hat{\rho}_{Ci}, \tag{2}$$

where we calculate the sum of the Kronecker product of output single-qubit matrices ($\hat{\rho}_{Ai}$, $\hat{\rho}_{Bi}$, $\hat{\rho}_{Ci}$) over different N_k channels, numbered by i. Note that the reconstructed matrix is separable, but, in general, it does not fulfill the additional criteria which would make it a nondiscordant state from definition. Since a perfect decoupling into single-qubit matrices is possible only for separable states, the minimal distance between the original and reconstructed density matrices, obtained by the training of the neural network, can be seen as a measure of the separability of states.

Both the original and reconstructed density matrices have to be included in the NN loss function, which describes how different the reconstructed state is from the original state. In quantum information theory, one of the natural choices for a difference between density matrices is the Bures distance [50], but numerical experiments showed much better efficiency for

training our NN model with the well-established L^1 norm:

$$\mathcal{L}(\rho_{ABC}) = \frac{1}{d^2}\sum_{i,j}\left|\rho^{ij}_{ABC} - \hat{\rho}^{ij}_{ABC}\right|. \tag{3}$$

Here ρ^{ij}_{ABC} and $\hat{\rho}^{ij}_{ABC}$ denote the elements of the original and reconstructed density matrices labeled by the indices i and j, and d^2 is the number of matrix elements. Using this loss function, we train the NN model so that the total reconstruction loss, summed over all training examples, is minimized. Note that to train the network, which we will call a *separator*, we do not need the labeled data. Instead, we compare inputs with outputs, which means that our training scheme is fully unsupervised. These types of ML models are also called *autoencoders* (we comment on this in more detail in the Appendix A).

Importantly, by choice of the model architecture, the separator network is suited to reconstruct separable states; hence, for these states we expect the value of the loss function to be close to zero. In contrast, for entangled states, the model is not able to reconstruct the input, and the difference between the two matrices will result in a significantly higher value of the loss $\mathcal{L}$. In order to distinguish between separable and entangled states, we set the anomaly threshold τ in such a way that if $\mathcal{L}(\rho_{ABC})$ is greater than τ for a given state ρ_{ABC}, then this state is classified as entangled. Otherwise, it is marked as separable.

III. DATA GENERATION

The most critical aspect of training an ML model is having a well-balanced and diversified data set. The modern approach in ML is to synthesize data with already known properties (e.g., forming separate classes in case of classifier training). To achieve this goal, we select different methods for generating random density matrices. The main idea is to generate random states using the quantum circuits approach [38], but in order to have a well-diversified data set, we also use a technique based on sampling from the uniform Haar measure [51] and an additional parametrized method designed by us (generation details can be found in the Appendix E).

We construct a training set composed of 530 000 pure and mixed separable states including a class of nondiscordant states and a validation set of 50 000 is constructed in the same way. Furthermore, we generate two test sets involving both separable and entangled states. The first of those (S_{pure}) is constructed from 15 000 random pure separable states and 15 000 random pure entangled states. This data set is generated evenly using both the quantum circuit approach and sampling from the Haar measure. The second test set (S_{mixed}) includes 65 000 mixed states, constructed from pure states either by mixing different states with random probabilities or by reducing a larger multiqubit space to the desired few-qubit size via the partial trace.

To classify states as separable during the generation process, we use the *negativity* [52,53], an entanglement measure based on the PPT (positive partial transpose) criterion of entanglement [54,55], which is one of the few measures that can be found directly from the density matrix for systems larger than two qubits. It is defined as the absolute sum of the negative eigenvalues of a bipartite density matrix after partial transposition is performed on either of the subsystems. The generation methods for mixed entangled states are specifically adjusted not to produce bound entangled states [56], which cannot be detected by negativity. To qualify discord, we use the criterion of Ref. [25], which is very straightforward to check, especially if one of the subsystems is small (relevant details are given in the Appendix F). Incidentally, it is common that a state is discordant only with respect to one part of a bipartition, so a state has to contain no discord with respect to both subsystems to be classified as a nondiscordant state [20]. For both entanglement and discord, the lack of correlations is checked three times: between qubit i and the rest of the three-qubit system, with $i = A, B, C$.

TABLE I. Average separator loss for different subsets in pure S_{pure} and mixed S_{pure} test sets quantifying similarity between different types of states and separable ones.

	$\langle\mathcal{L}\rangle$ for different subsets			
Test set	Separable	Nondiscordant	Discordant	Entangled
S_{pure}	5.9×10^{-4}	5.9×10^{-4}	0.042	0.042
S_{mixed}	5.2×10^{-3}	1.0×10^{-3}	0.017	0.021

IV. MODEL TRAINING AND RESULTS

In ML, a single pass through the entire training set, during which we update model parameters (typically using the gradient of a loss function in the parameter space), is called an epoch. The separator model was trained for 20 epochs, in a way that the model parameters for the epoch with the lowest averaged reconstruction loss $\langle\mathcal{L}\rangle$ on the validation set were saved and used for evaluation (testing) purposes.

For the test set that contains only pure states, S_{pure}, the average loss $\langle\mathcal{L}\rangle$ for separable states is equal to 5.9×10^{-4}, while in the case of entangled states, it is much larger, giving $\langle\mathcal{L}\rangle = 0.042$. Obviously, for the discord the numbers are the same—see Table I. Choosing the threshold value τ to obtain the highest possible accuracy [balanced accuracy (BA), defined in the Appendix B] of the model predictions yields $\tau = 1.25\times10^{-3}$ with the corresponding classification accuracy of 99.63% for the whole S_{pure} set.

For the test set with mixed states, S_{mixed}, in the case of nondiscordant states, the average loss $\langle\mathcal{L}\rangle$ is only about 1.5 times higher than for uncorrelated pure states, meaning that the original and reconstructed states are the same to an equivalent extent, while for mixed separable states it is 40 times higher, signaling that, on average, mixed separable states differ significantly more from the reconstruction. The same behavior, with mixed discordant states more similar to pure separable than mixed separable, can also be observed in supervised ML models—further details can be found in the Appendix C. Moreover, discordant and entangled states display roughly half $\langle\mathcal{L}\rangle$ for mixed states compared to pure correlated states, which is in agreement with the fact that mixed states are on average less correlated than pure states. Overall, the difference between zero and nonzero discord

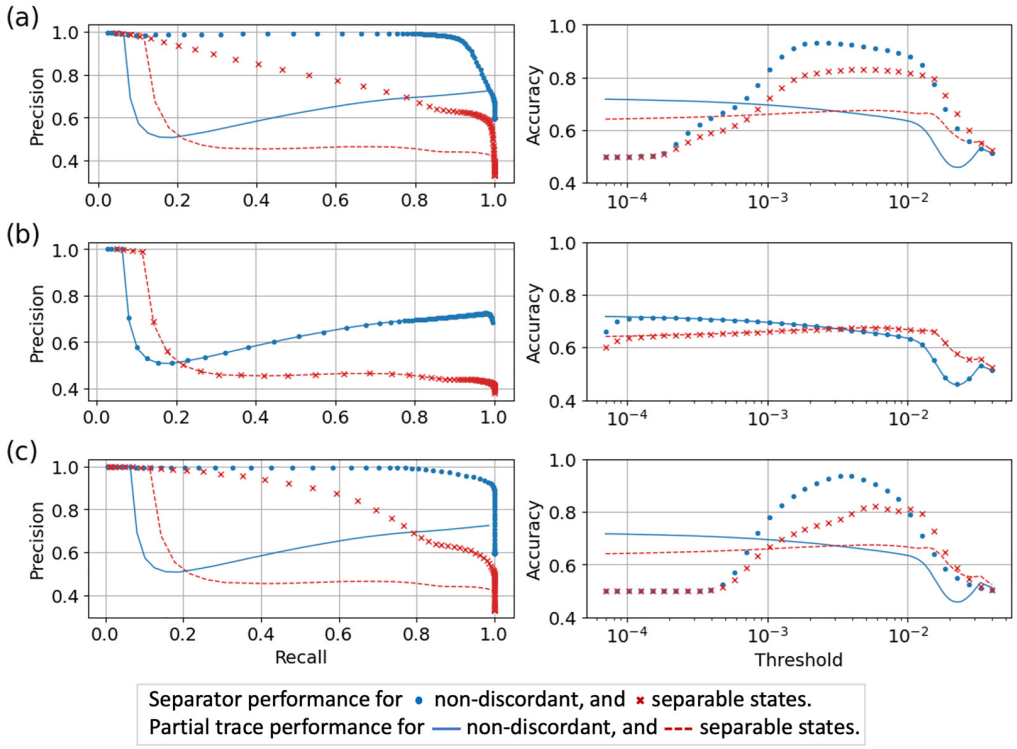


FIG. 2. (a) Separator performance for detecting discordant (blue dots) and entangled (red crosses) states, tested on S_{mixed}. Left: Precision vs recall curves. Right: Balanced accuracy, depending on the threshold τ value. For comparison, partial trace-based baseline model performance is also presented by blue solid and red dashed curves, respectively. (b) Same as (a) but for the separator model with removed FC layers—cf. Fig. 1. (c) Same as in (a) but for the separator trained on the nonproduct states only.

states (measured by reconstruction loss $\mathcal{L}$) is over an order of magnitude larger than between separable and entangled states.

Let us now test the separator performance using the mixed test set, S_{mixed}, and compare its practical usefulness for either discord or entanglement detection. We performed two independent tests. In the first, we took discord labels and tested the ability to classify discord, and in the second, we took separability labels for the same set S_{mixed} and tested the ability to classify entanglement. Since for mixed states a discordant state can also be separable, one cannot simultaneously balance nondiscordant vs discordant and separable vs entangled states in a single mixed test set; therefore, we use BA (see the Appendix B) as the model performance measure.

Figure 2 contains results of separator performance for detecting discordant (green curves) and entangled (red) states on the S_{mixed} set. Figure 2(a), right, shows that the model has much higher BA when detecting discord than entanglement in a wide range of thresholds, reaching 93% for $\tau = 0.0023$ in case of discord detection, while for entanglement a maximum of 82% is reached at $\tau = 0.0051$. Moreover, when training the model on the subspace of the original training domain where the product states have been excluded, as presented in Fig. 2(c), we clearly get better performance for discord in terms of the larger area under the precision-recall curve (left). Performance comparisons for training separately on some other classes of states can be found in the Appendix A.

The convolution part in the separator performs an operation that is a generalization of the partial trace. When removing the FC layers that extend the convolution layers, we observe that the model simply collapses to the partial trace baseline (example kernels for both cases can be found in the Appendices) and is analogous to what is obtained when training only on pure separable states. Figure 2(b) contains results for a separator with the FC layers removed: comparing accuracy with a partial trace baseline (blue and orange curves) we get practically the same accuracies.

We present the confusion matrices for the separator serving as a separability vs. entanglement classifier in Table II and a nondiscordance vs. discordance classifier in Table III. Overall, the separator tends to grossly overestimate the set of entangled states as seen in Table II, where for both thresholds presented ($\tau = 0.0023$ corresponding to the best discord classification and $\tau = 0.0051$ corresponding to the best entanglement classification) the number of separable states wrongly labeled as entangled is of the same order of magnitude as the number of correctly detected entangled states. The number of erroneously classified separable states is comparatively small. In

TABLE II. Confusion matrix for separable vs. entangled states classification at $\tau = 0.0023$ and 0.0051.

Labels \ Predictions		Separable	Entangled
Separable	$\tau = 0.0023$	28340	15257
Entangled	$\tau = 0.0023$	305	21098
Separable	$\tau = 0.0051$	30696	12901
Entangled	$\tau = 0.0051$	942	20461

TABLE III. Confusion matrix for nondiscordant vs. discordant states classification at $\tau = 0.0023$ and 0.0051.

Labels \ Predictions		Nondiscordant	Discordant
Nondiscordant	$\tau = 0.0023$	25114	1192
Discordant	$\tau = 0.0023$	3531	35163
Nondiscordant	$\tau = 0.0051$	25822	484
Discordant	$\tau = 0.0051$	5816	32878

contrast, the set of discordant states tends to be underestimated, with a discordant state being more likely to be labeled as not correlated than the other way around. This occurs to a much lesser extent than the overestimation of entanglement as seen in Table III. The results here are more balanced, with erroneous classifications averaging at a number an order of magnitude lower than the correct predictions.

V. CORRELATION MAP

In order to acquire a deeper understanding of how well the sets of nondiscordant and separable states are reconstructed by the ML classifier at different thresholds, we analyze a family of three-qubit states which are easy to parametrize, and study a two-dimensional (2D) diagram forming a map with regions differing in the type of correlations present (or absent) as seen in the inset of Fig. 3.

The family of states is constructed as follows. We take a separable mixed state of the form $\rho = p\rho_1 + (1-p)\rho_2$, where probability p is one of the parameters, and the density matrices ρ_1 and ρ_2 correspond to three-qubit pure states, $|\Psi_i\rangle = \bigotimes_k |\psi_i\rangle$, with $k = A, B, C$ and

$$|\psi_1\rangle = a|0\rangle + \sqrt{1-a^2}|1\rangle, \tag{4a}$$

$$|\psi_2\rangle = e^{-i\frac{\phi}{2}}\sqrt{1-a^2}|0\rangle - ae^{i\frac{\phi}{2}}|1\rangle, \tag{4b}$$

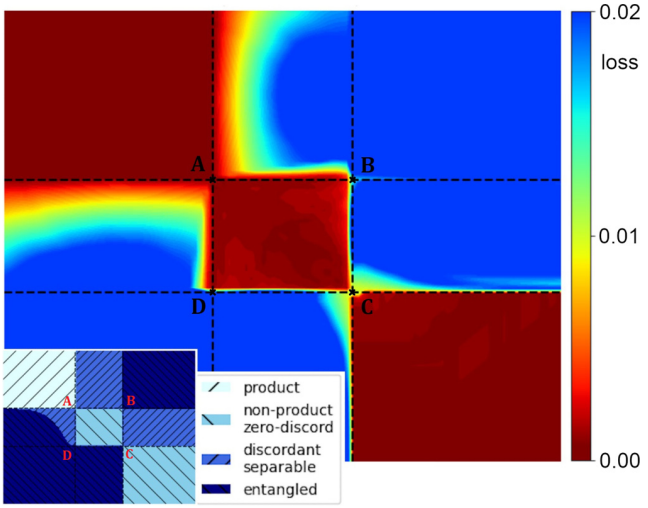


FIG. 3. Reconstruction loss $\mathcal{L}$ for the already trained separator model tested on the family of three-qubit states with their separability and discordance conditions known and parametrized on a 2D map (see inset for the map division into the classes and the Appendix D for detailed parametrization of the classes).

respectively. Coefficient a and phase factor ϕ constitute two other parameters that characterize the family of states. It now contains only separable states. For $p = 0, 1$ the state is pure (and consequently of product form), while it is mixed nondiscordant for $\phi = 0$ or $a = 0, 1$. To introduce entanglement into the system we find the spectral decomposition of $\rho = \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|$ and increase the largest eigenvalue, λ_0:

$$\rho \to \frac{\sum_{i\neq 0} \lambda_i |\varphi_i\rangle\langle\varphi_i| + (\lambda_0 + c)|\varphi_0\rangle\langle\varphi_0|}{\sum_i \lambda_i + c}, \tag{5}$$

making c the final parameter which characterizes this family of states.

This allows us to plot the loss function (3) for selected cross sections of the Hilbert space and visualize how surfaces of equal $\mathcal{L}$ compare to the sets of nondiscordant and separable states. In Fig. 3 only the darkest regions in the inset are entangled; all other regions are separable and are further divided into product states, nondiscordant mixed states, and discordant separable states. The respective regions are obtained by careful choice of parameters p, a, ϕ, and c along the horizontal and vertical axes.

The main plot in Fig. 3 shows the loss $\mathcal{L}$ on this parameter space. The specifics of how the different parameters change in each region are described in detail in the Appendix D. It is noticeable that, while the partial trace-based model is capable only of distinguishing between product and nonproduct states (we show the plot and discuss this in detail in the Appendix A), the neural network distinguishes all nondiscordant states. The results presented in Fig. 3 are in agreement with the general results of Fig. 2, and Tables II and III, demonstrating why the discord detection results in false negatives (overestimating the set of nondiscordant states), while entanglement detection results in so many false positives.

VI. CONCLUSIONS

In conclusion, we designed and trained a neural network to distinguish between separable and entangled states on a set of mixed and pure states. For optimum threshold values, the network performs with 82% balanced accuracy for mixed states, as compared to over 99% classification accuracy of the same network for pure state entanglement. To understand this discrepancy outside of the trivial acknowledgment that determining mixed state entanglement is a much more complicated task than pure state entanglement, we have compared the results obtained on the mixed test set with other classes of uncorrelated states. We have found that the set of states labeled as uncorrelated by the separator overlaps best with the set of nondiscordant states. This is confirmed by the 93% balanced accuracy of detecting such states, outperforming the detection of both separable and product states. This result is surprising, because it shows that, for two different ML models (supervised and unsupervised), random mixed nondiscordant states are much more similar to states from a random set of pure separable states (as the lowest loss is observed for such states) than random mixed separable states. We supplement plots of ML metrics and confusion matrices with a map of numerically detected quantum correlations for cross sections of the three-qubit Hilbert space, which allows us to understand the nature of the sets of states which are qualified

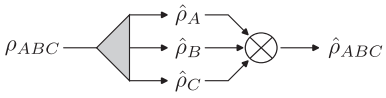


FIG. 4. The separator autoencoder composed of the trainable encoder neural network (gray triangle) and the analytical decoder that simply calculates the Kronecker product.

as uncorrelated by ML. The surface where the loss is near zero (no quantum correlations) reproduces the set of nondiscordant states extremely well, while the discordant separable states are largely classified as quantum correlated, illustrating why the separator grossly overestimates entanglement while only marginally underestimating discordance. The network architecture is designed carefully to preserve separability (i.e., it does not introduce any entanglement when reconstructing the output) and its output is equivariant with respect to qubit permutations. We observe that the simplified version of the separator model, which realizes just generalization of a partial trace operation, simply collapses to partial trace during the training. However, when extending it to a neural model capable to learn other representations by adding fully connected layers working separately within each qubit subspace, it learns more complex patterns and improves its performance over a partial trace baseline.

ACKNOWLEDGMENTS

M.K. and J.P. acknowledge support from the National Science Centre (Poland) under Grant No. 2021/43/D/ST3/01989. M.M.M. acknowledges that this research is part of Project No. 2021/43/P/ST3/03293 cofunded by the National Science Centre and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant No. 945339.

APPENDIX A: DETAILS ON THE SEPARATOR NEURAL NETWORK

In the following section, we analyze the structure of the NN model trained to reconstruct separable states and detect nonseparable states as an anomaly. Anomaly detection is a standard problem in ML where a model is trained to reconstruct normal data and then fails when trying to reconstruct abnormal data (anomaly). The reconstruction error, i.e., the model loss function, measures the similarity of a given unknown sample to the training data, and if it is greater than some assumed threshold, then the sample is treated as an anomaly. It is easy to construct a many-qubit density matrix as a product of single-qubit density matrices. However, the opposite operation is much more complicated. Then, this task can be seen as an *inverse problem*. Here, we propose to solve it with the help of a modified autoencoder-type NN, schematically depicted in Fig. 4. The typical autoencoder is composed of two connected networks: the *encoder* and the *decoder*. Usually, the output of the encoder is narrower than its input, so this part compresses the data. Next, the decoder decompresses it, and both parts of the autoencoder are trained to have the output as close as possible to the input. Our idea is to use the encoder to transform three-qubit density matrices into three single-qubit matrices in such a way that their product, calculated by the decoder, is as close to the original density matrix as possible. The autoencoder will be implemented as a NN, with only the encoder part trainable. The decoder simply calculates the Kronecker product of the single-qubit density matrices, as described in Eq. (2).

The backbone of the encoder is composed of convolution layers, which map the input density matrix into single-qubit subspaces. They are followed by FC layers that extend the model and introduce nonlinearity; however, they do not introduce entanglement, since they are applied independently for each of the reconstructed qubits.

Let us focus on a $d = 3$ qubit system ρ_{ABC}, bearing in mind that the proposed reconstruction can be easily extended to larger qubit registers. We require a neural network model $\mathcal{N}$, which we call a separator, to have the following properties:

$$\mathcal{N}(\rho_A \otimes \rho_B \otimes \rho_C) = c\,\rho_A \otimes \rho_B \otimes \rho_C, \tag{A1}$$

$$\mathcal{N}\left(\sum_i p_i\,\rho_{Ai} \otimes \rho_{Bi} \otimes \rho_{Ci}\right) = \sum_i \hat{p}_i\,\hat{\rho}_{Ai} \otimes \hat{\rho}_{Bi} \otimes \hat{\rho}_{Ci}, \tag{A2}$$

with some multiplicative constant c. To do so one can adjust the convolution layers $\mathcal{C}^{s,l}$, with kernel $\mathcal{K}$, step size (stride) s, and dilation l:

$$\mathcal{C}_{s,l}(\rho)^{i,j} = (\mathcal{K} *_l \rho)^{si,sj} = \sum_{k,q} \mathcal{K}^{k,q}\,\rho^{si+l(k),sj+l(q)}, \tag{A3}$$

where $i, j = 0, 1$, and $k, q = 0, \ldots, 2^{d-1} - 1$, so that the kernels overlap with the respective blocks of the density matrix and therefore convolving them against the density matrix results with output matrices representing systems with a reduced number of qubits. Dilation can be just a simple scaling of the sum index: $l(k) = l\,k$, with rate l meaning that the kernel only touches the signal at every lth entry, or more complex indexing.

Specifically, choosing $\mathcal{C}_A = \mathcal{C}_{4,1}$ with kernel $\mathcal{K}$ of size 4×4 allows the neural network to learn the structure of repeating blocks $\rho_B \otimes \rho_C$ and therefore to reduce the input density matrix to

$$\mathcal{C}_{4,1}(\rho_A \otimes \rho_B \otimes \rho_C) = c\,\rho_A,$$

with c being some multiplicative constant, as in Eq. (A1). To extract the last one, we need $\mathcal{C}_C = \mathcal{C}_{1,2}$. To extract the second qubit, more complex dilation is needed in $\mathcal{C}_B$. All kernels are visualized in Fig. 1. The input density matrix, divided into two channels (respectively for the real and imaginary parts), is convolved separately with three convolution layers, each with 4×4 kernels. The first one is set up using a stride equal to 4, as a result producing a matrix corresponding to the first qubit subspace. In the second convolution, the kernel is divided into four blocks of size 2×2, separated with dilations resulting in the second qubit subspace. The third convolution consists of the kernel dilated with extra space between every weight, therefore extracting the information about the last qubit. Finally, after applying all the convolutions, one can combine them and construct the output separable density matrix $\mathcal{C}_A(\rho_{ABC}) \otimes \mathcal{C}_B(\rho_{ABC}) \otimes \mathcal{C}_C(\rho_{ABC}) = \hat{\rho}_A \otimes \hat{\rho}_B \otimes \hat{\rho}_C$.

It is obvious that in this way the defined separator model $\mathcal{N}$ fulfils Eq. (A1). However, to enable the neural network to

generalize and reconstruct separable mixed states, i.e., obey Eq. (A2), a larger number of (triples of) kernels is chosen (we use $N_K = 24$ kernels in each type of convolution layer). To simplify the implementation and keep neural network values real, we doubled the number of kernels, that is, $2N_K$, so half of the channels processed the real and the other half processed the imaginary part of the density matrix. At this point, we also add four FC layers for each qubit (A, B, C) separately to extend the complexity of the model. Subsequently, we calculate the sum of the Kronecker product of output single-qubit matrices $(\hat{\rho}_{Ai}, \hat{\rho}_{Bi}, \hat{\rho}_{Ci})$ over different kernels numbered by i and arrive at the reconstructed density matrix as in Eq. (2).

Both the original and reconstructed density matrices are included in the loss function of the neural network defined in Eq. (3). Using this loss function, we train the separator model so that the total reconstruction loss, summed over all training examples, is minimized. Note that to train the network, we do not need the data to be labeled. Instead, we simply compare inputs with outputs, and therefore, our model scheme is fully unsupervised. The trainable weights of the kernels are adjusted using the procedure called gradient descent, where the gradients of the loss are calculated with respect to the weights. Subsequently, all parameters of the model, e.g., each kernel weight K_{ij}, are updated as follows:

$$K^{ij}_{\text{new}} = K^{ij}_{\text{old}} - \alpha \frac{\partial \mathcal{L}}{\partial K^{ij}_{\text{old}}}, \tag{A4}$$

where α is the learning rate. We proceed similarly for weights in FC layers. This procedure is repeated until the reconstruction loss [Eq. (3)] reaches (we hope) the global minimum, which means that the model is already trained. Importantly, by the choice of the model architecture, the separator network is suited to reconstruct separable states, and for these states, we expect the value of the loss function to be close to zero. In contrast, in the case of an entangled state, the model will not be able to reconstruct the input, and the output density matrix should differ from the input one, resulting in a significantly higher value of the loss. Therefore, to distinguish between separable and entangled states, we set the anomaly threshold τ in such a way that if $\mathcal{L}(\rho_{ABC})$ is larger than τ for a given state ρ_{ABC}, then this state is classified as entangled and, otherwise, it is marked as separable. Interestingly, in addition to classification purposes, the loss value $\mathcal{L}(\rho_{ABC})$ can also be interpreted as a similarity measure between the states used in the training dataset and given tested state ρ_{ABC}. In Fig. 5 we present the extended results for the separator model with and without four FC layers trained on different subsets of the original training set. As before, the neural network is evaluated on the S_{mixed} set. We start to train the model in the simplest possible scenario where the dataset consists of 190 000 pure separable states. The model trained on such a dataset collapses to the baseline trace model. It is expected behavior since in this case the baseline model is a sufficient criterion for distinguishing between entangled and separable states. The same situation occurs when we extend our training set to the wider set of product states by adding 120 000 mixed product states. Interestingly, in this case, the neural network reflects partial trace operation even more precisely, which is revealed by a wider range of thresholds producing

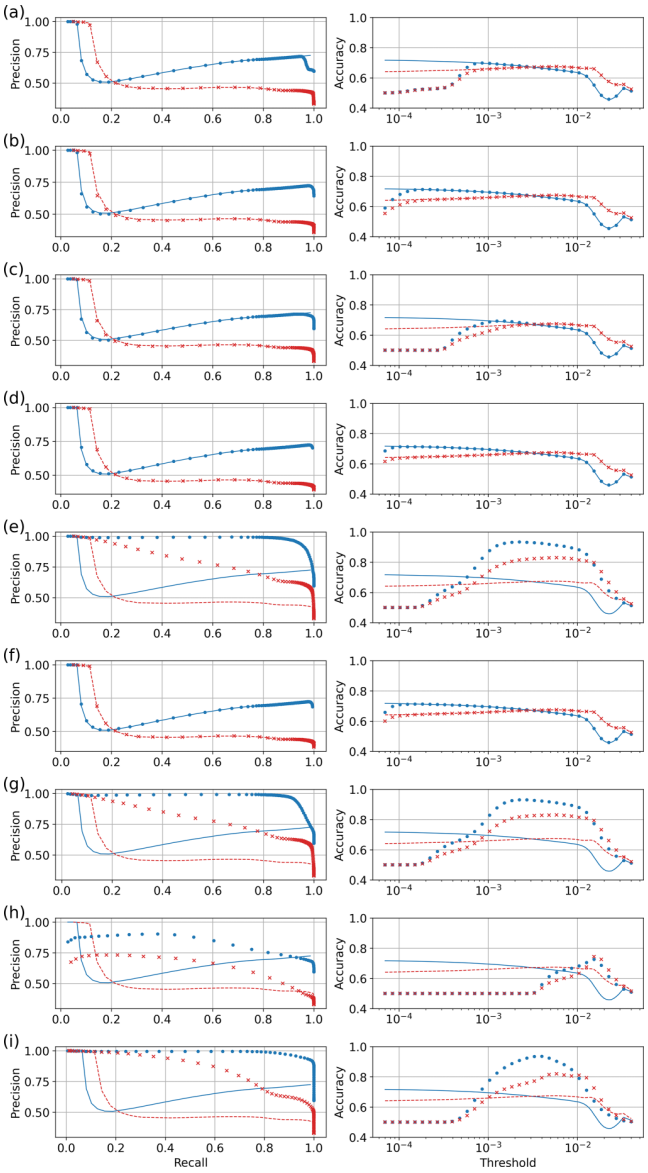


FIG. 5. Same as in Fig. 2, but for the separator trained only on some part of the original training set: (a) pure separable states, (b), (c) product states, (d), (e) nondiscordant states, (f), (g) separable states, and (h), (i) nonproduct separable states. Results are shown for the (standard) separator with four fully connected layers after convolutional ones (c), (e), (g), (i) and without it, containing only convolutional layers (a), (b), (d), (f), (h).

overlapping accuracy. Including FC layers, and thus introducing the nonlinearity, slightly limits this phenomenon; however, since the baseline model is still a precise criterion of separability in the case of product states, hence the neural network is not penalized for not learning more complex operations. The situation changes when we append 100 000 nonproduct nondiscordant states. Now, the baseline is no longer reliable and the network needs to learn extra features in order to correctly separate and reconstruct such states. Notably, the separator model is capable of achieving this task only when extra FC layers are present. Otherwise, it collapses to the baseline. This situation reoccurs for the full training

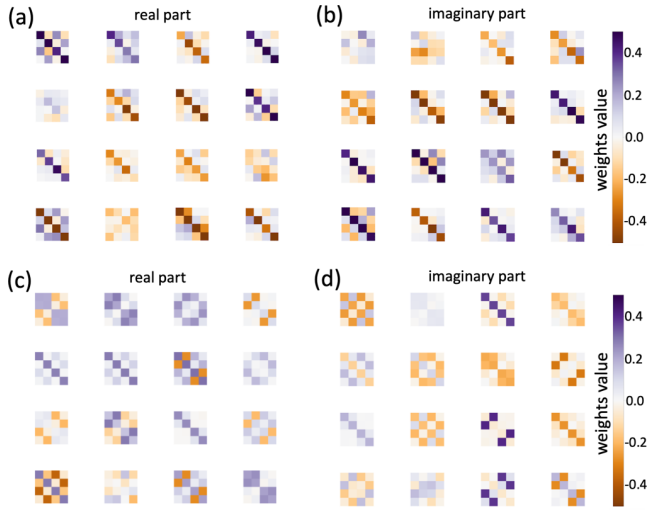


FIG. 6. Example weights of 4×4 convolutional kernels for the already trained three-qubit separator model. The kernels reconstruct (a) real and (b) imaginary parts of the first qubit (weights for two other qubits are similar) for the plain separator model, with just convolutional layers. Similarly, (c) and (d) show real and imaginary parts of kernels for the separator with additional fully connected layers.

set consisting of 530 000 separable states. For the sake of completeness, we additionally test our model after training it on the reduced domain of nonproduct separable states. The data set in this case is constructed separately from 360 000 states, which are generated with the same methodology as for the original training set. The main difference of the model trained on such a subdomain is that even without any FC layers it does not collapse to a partial trace operation, which may suggest that using only nonproduct separable states for training can also be beneficial for the model with FC layers. In Fig. 6 we present exemplary kernels learned for the separator model with (right) and without FC layers (left) trained on the full data set (as in the main text). The model with only convolutional layers collapses to a partial trace baseline which manifests itself in the fact that its kernels resemble identity matrices (the partial trace is equivalent to the convolutional kernel being equal to the identity matrix).

In Fig. 7 we also present a reconstruction loss map analogous to the map in Fig. 3 in the main text, but obtained using a partial trace as a baseline replacement for the separator model. On this map, only product states are clearly detected as not containing correlations, as is to be expected; however, the region in the lower right square of the map, which represents some space of mixed zero-discord states, is signified as containing fewer correlations. Note that the other region of similar type (in the center of the plot) is not correspondingly distinguished.

APPENDIX B: CLASSIFICATION METRICS

In machine learning to evaluate classification or detection results, we typically use various types of metrics that measure model performance.

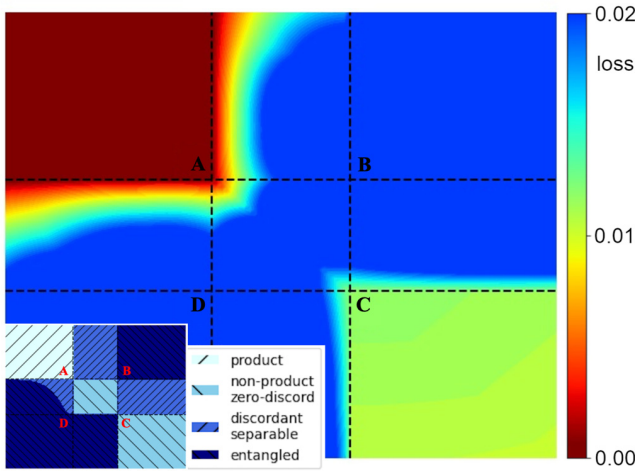


FIG. 7. Reconstruction loss $\mathcal{L}$ for the partial trace-based model tested on a family of three-qubit states with their separability and discordance conditions parametrized on a 2D map (see inset for the map division into the classes and Appendix D for detailed parametrization of the classes).

The standard metrics [57] used in binary classification, also to tune the model hyperparameters, e.g., classification threshold τ, are precision $\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ and recall $\mathrm{RC} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$. Here TP, TN, FP, and FN denote respectively the number of true and false positives and negatives; by "negatives" we mean nondiscordant (or separable), while by "positives" we mean discordant (or entangled) states.

The procedure called "precision-recall tradeoff" involves choosing a threshold that maximizes the area under the precision-recall curve. However, due to the fact that our dataset is imbalanced (we cannot have balanced separable or entangled and nondiscordant or discordant classes at the same time), we use a metric typically defined for imbalanced data sets. Balanced accuracy is defined as the average of recall obtained on each class [58]: $\mathrm{BA} = \frac{1}{2}(\frac{\mathrm{TP}}{\mathrm{TP+FN}} + \frac{\mathrm{TN}}{\mathrm{TN+FP}})$.

APPENDIX C: SIMILARITY BETWEEN STATES

In order to verify whether the statement from the main text, that the mixed nondiscordant states are much more similar to states from a set of pure separable states (as the lowest loss is observed for such states in Table I) than mixed separable states, also holds for other machine learning models, we tested two different models, as introduced in Ref. [38]. Both of them, CNN and Siamese CNN, were trained in a supervised manner to classify if a given three-qubit state is entangled [38]. In Table IV there are presented average output probabilities for different classes of states in the test set for both models. We can clearly observe that the outputs show the same similarities between states as for the unsupervised model, used in the main text.

APPENDIX D: PARAMETER VARIATIONS ON A THREE-QUBIT MAP

The family of states in the map in Fig. 3 is constructed as follows. We take a separable mixed state of the form

TABLE IV. Average output probability p for CNN (top) and Siamese CNN (bottom) models trained to predict if a given state is entangled.

	$\langle p\rangle$ for CNN model			
Test set	Separable	Nondiscordant	Discordant	Entangled
S_{pure}	0.104	0.104	0.968	0.968
S_{mixed}	0.221	0.115	0.686	0.887
	$\langle p\rangle$ for Siamese CNN model			
Test set	Separable	Nondiscordant	Discordant	Entangled
S_{pure}	0.027	0.027	0.969	0.969
S_{mixed}	0.105	0.038	0.570	0.824

$\rho = p\rho_1 + (1-p)\rho_2$, where probability p is one of the parameters, and the density matrices ρ_1 and ρ_2 correspond to three-qubit pure states, $|\Psi_i\rangle = \bigotimes_k |\psi_i\rangle$, with $k = A, B, C$ and

$$|\psi_1\rangle = a|0\rangle + \sqrt{1-a^2}|1\rangle, \tag{D1a}$$

$$|\psi_2\rangle = e^{-i\frac{\phi}{2}}\sqrt{1-a^2}|0\rangle - ae^{i\frac{\phi}{2}}|1\rangle, \tag{D1b}$$

respectively. Coefficient a and phase factor ϕ constitute two other parameters that characterize the family of states. It now contains only separable states. For $p = 0, 1$ the state is pure (and consequently of product form), while it is mixed nondiscordant for $\phi = 0$ or $a = 0, 1$. To introduce entanglement into the system we find the spectral decomposition of $\rho = \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|$ and increase the largest eigenvalue, λ_0:

$$\rho \to \frac{\sum_{i\neq 0}\lambda_i|\varphi_i\rangle\langle\varphi_i| + (\lambda_0 + c)|\varphi_0\rangle\langle\varphi_0|}{\sum_i \lambda_i + c}, \tag{D2}$$

making c the final parameter which characterizes this family of states.

Now we detail how the specific parameters are changed in the different regions of the map to generate states corresponding to different classes in terms of quantum and classical correlations. Specifically, we start from point A, where the parameter $a = 1/\sqrt{2}$ takes the maximal value, while all others are set to minimal values: $p = 0$, $\phi = 0$, and $c = 0$. Moving along the axis AB we increase the parameter p so that it takes the maximal value $1/2$ in point B. Similarly, coefficient a is decreased along the axis BC, so that in point C it is equal to zero. Moving outside the square region $ABCD$ we fix parameters a and p to the boundary values and start increasing the parameters ϕ and c. The phase ϕ grows linearly along both horizontal and vertical axes in such a way that inside the square $ABCD$ it is equal to zero and it is equal to $\pi/2$ at the edges of the map. The parameter c is increased along the counterdiagonal in the top right square, taking the minimal value zero in point B and maximal value 1 in the top right corner of the map. The parametrization is performed symmetrically on both sides of the diagonal of the map. Therefore, the axis AB in the upper right triangle corresponds to the axis AD in the lower left triangle. Similarly, BC is reflected to DC. Keeping that in mind, we note that parameter c is additionally increased by Δc in the lower left triangle. This extra term grows linearly along the counterdiagonal from $\Delta c = 0$ in the center of square $ABCD$ to $\Delta c = 1$ in the bottom left corner of the map.

APPENDIX E: PARAMETRIZED METHOD OF GENERATING THREE-QUBIT DENSITY MATRICES

The most critical aspect of training an ML model is having a well-balanced and diversified data set. The modern approach in ML is to synthesize data with already known properties (e.g., forming separate classes in case of classifier training). To achieve this goal, we select different methods for generating random density matrices. The main idea is to generate random states using the quantum circuits approach [38], but in order to have a well-diversified data set, we also use a technique based on sampling from the uniform Haar measure [51].

We construct a training set composed of 530 000 pure and mixed separable states including a class of nondiscordant states and a validation set of 50 000 is constructed in the same way. Furthermore, we generate two test sets involving both separable and entangled states. The first of those (S_{pure}) is constructed from 15 000 random pure separable states and 15 000 random pure entangled states. This data set is generated evenly using both the quantum circuit approach and sampling from the Haar measure. The second test set (S_{mixed}) includes 65 000 mixed states, constructed from pure states either by mixing different states with random probabilities or by reducing a larger multiqubit space to the desired few-qubit size via the partial trace.

Now we describe the original parametrized method used to generate random mixed states when building the training set. To generate such states, we also used the quantum circuits approach and a technique based on sampling from the uniform Haar measure—both are described in Ref. [38].

Let us start by defining a three-qubit pure separable state:

$$\begin{aligned}|\psi_{\text{sep}}\rangle &= |\psi_1\rangle \otimes |\psi_2\rangle \otimes |\psi_3\rangle \\ &= (a_1|0\rangle + b_1|1\rangle) \otimes (a_2|0\rangle + b_2|1\rangle) \otimes (a_3|0\rangle + b_3|1\rangle) \\ &= a_1a_2a_3|000\rangle + a_1a_2b_3|001\rangle + a_1a_3b_2|010\rangle \\ &\quad + a_1b_2b_3|011\rangle + a_2a_3b_1|100\rangle + a_2b_1b_3|101\rangle \\ &\quad + a_3b_1b_2|110\rangle + b_1b_2b_3|111\rangle,\end{aligned} \tag{E1}$$

where the complex coefficients of each state fulfill

$$|a_i|^2 + |b_i|^2 = 1. \tag{E2}$$

We then introduce entanglement into the system for each pair of qubits via phases ϕ_{ij}:

$$\begin{aligned}|\psi_{\text{ent}}\rangle &= a_1a_2a_3|000\rangle + a_1a_2b_3|001\rangle + a_1a_3b_2|010\rangle \\ &\quad + a_1b_2b_3e^{i\phi_{23}}|011\rangle + a_2a_3b_1|100\rangle \\ &\quad + a_2b_1b_3e^{i\phi_{13}}|101\rangle + a_3b_1b_2e^{i\phi_{12}}|110\rangle \\ &\quad + b_1b_2b_3e^{i(\phi_{12}+\phi_{13}+\phi_{23})}|111\rangle.\end{aligned} \tag{E3}$$

Finally, the decoherence coefficients c_i are included for each qubit separately in order to obtain mixed states, mimicking the operation of single-qubit dephasing channels:

$$\rho_i = |\psi_i\rangle\langle\psi_i| = \begin{pmatrix} |a_i|^2 & a_ib_i^* \\ a_i^*b_i & |b_i|^2 \end{pmatrix} \tag{E4}$$

$$\xrightarrow{\text{decoherence}} \begin{pmatrix} |a_i|^2 & c_ia_ib_i^* \\ c_i^*a_i^*b_i & |b_i|^2 \end{pmatrix} = \tilde{\rho}_i. \tag{E5}$$

Combining all of the above, one arrives at the following:

$$\rho = |\psi_{\text{ent}}\rangle\langle\psi_{\text{ent}}| \xrightarrow{\text{decoherence}} \tilde{\rho}, \tag{E6}$$

where the density matrix $\tilde{\rho}$ is parametrized with independent coefficients a_i and c_i as well as phases ϕ_{ij}. In order to obtain a diversified set of states, we propose to choose these parameters uniformly at random over the interval $[0, 1]$ for a_i and c_i, and over the interval $[0, 2\pi]$ for phases ϕ_i. Moreover, we randomize the states by evolving them with local single-qubit gates:

$$U(\theta, \phi, \lambda) = \begin{pmatrix} \cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2} \end{pmatrix}, \tag{E7}$$

where θ, ϕ, and λ are some arbitrary Euler angles. Hence, sampling the angles randomly over the interval $[0, 2\pi]$, independently for separate qubits, one can write the final randomized density matrix as

$$\tilde{\rho}_{\text{rand}} = (U_1 \otimes U_2 \otimes U_3)\tilde{\rho}(U_1 \otimes U_2 \otimes U_3)^\dagger. \tag{E8}$$

APPENDIX F: CLASSIFYING STATES AS NONDISCORDANT

To classify a state as discordant or not, we use an if-and-only-if criterion for nondiscordant states from Ref. [25]. The criterion states that a bipartite state with subsystems of arbitrary dimensions N and M has zero quantum discord with respect to the system of dimension M, if and only if all blocks of the $(NM)\times(NM)$ density matrix, which are obtained by partitioning the density matrix into N^2 square matrices of dimension M, are normal matrices and commute with each other.

The required partition is performed starting from the density matrix written in the form

$$\hat{\sigma} = \sum_{kq}\sum_{nm} P_{kq}^{nm}|k\rangle\langle q| \otimes |n\rangle\langle m|, \tag{F1}$$

where the indices and corresponding states k and q correspond to the system of dimension N, while the indices and states n and m correspond to the subsystem of dimension M. The partition into blocks requires leaving only the elements which correspond to matrix elements $|k\rangle\langle q|$ in the subspace of the system of dimension N:

$$\hat{\sigma}_{kq} = \langle k|\hat{\sigma}|q\rangle. \tag{F2}$$

Each choice of k and q yields a different matrix.

A normal matrix is one that commutes with its Hermitian conjugate (the matrix obtained through the partition does not need to be a density matrix), thus fulfilling

$$[\hat{\sigma}_{kq}, \hat{\sigma}_{kq}^\dagger] = 0, \tag{F3}$$

whereas commutation obviously requires that for all k and q and for all k' and q' one has

$$[\hat{\sigma}_{kq}, \hat{\sigma}_{k'q'}] = 0. \tag{F4}$$

Both criteria are fulfilled, if and only if the state has zero discord with respect to the subsystem of dimension M.

It is relevant to keep in mind that a zero discord state must not have any correlations of this type with respect to the subsystems of both size M and size N, so in the case of our three-qubit system and each of the three-qubit–two-qubit partitions we will require two checks for zero discord, one with respect to the qubit and the other with respect to the two-qubit subsystem. In total, this means that checking for the presence of discord in a given three-qubit state is performed six times due to the different partitions and the asymmetry of the discord.

APPENDIX G: NUMERICAL METHODS

Data generation is performed with the usage of the QISKIT [59] library. All data sets can be generated in ≈8 h. The neural network model (encoder and decoder part) is implemented using the PYTORCH [60] library with an automatic differentiation engine (autograd) used to track gradients of the model parameters. The training of the single model on the GPU Nvidia RTX 3070 takes ≈1 day.

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145 (2002).

[2] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, Adv. Opt. Photon. **12**, 1012 (2020).

[3] H. J. Kimble, The quantum internet, Nature (London) **453**, 1023 (2008).

[4] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, Cambridge, England, 2000).

[6] B. Demirel, W. Weng, C. Thalacker, M. Hoban, and S. Barz, Correlations for computation and computation for correlations, npj Quantum Inf. **7**, 29 (2021).

[7] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **70**, 1895 (1993).

[8] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Experimental quantum teleportation, Nature (London) **390**, 575 (1997).

[9] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **80**, 1121 (1998).

[10] K. Roszak and J. K. Korbicz, Purifying teleportation, Quantum **7**, 923 (2023).
[11] M. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen, U. Sen, and B. Synak-Radtke, Local versus nonlocal information in quantum-information theory: Formalism and phenomena, Phys. Rev. A **71**, 062307 (2005).
[12] K. Modi, A. Brodutch, H. Cable, T. Paterek, and V. Vedral, The classical-quantum boundary for correlations: Discord and related measures, Rev. Mod. Phys. **84**, 1655 (2012).
[13] K. Roszak and Ł. Cywiński, The relation between the quantum discord and quantum teleportation: The physical interpretation of the transition point between different quantum discord decay regimes, Europhys. Lett. **112**, 10002 (2015).
[14] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, Phys. Rev. A **40**, 4277 (1989).
[15] M. B. Plenio and S. Virmani, An introduction to entanglement measures, Quantum Inf. Comput. **7**, 1 (2007).
[16] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).
[17] L. Gurvits, Classical complexity and quantum entanglement, J. Comput. Syst. Sci. **69**, 448 (2004).
[18] H. Ollivier and W. H. Zurek, Quantum discord: A measure of the quantumness of correlations, Phys. Rev. Lett. **88**, 017901 (2001).
[19] L. Henderson and V. Vedral, Classical, quantum and total correlations, J. Phys. A **34**, 6899 (2001).
[20] K. Modi, A pedagogical overview of quantum discord, Open Syst. Inf. Dyn. **21**, 1440006 (2014).
[21] Y. Huang, Computing quantum discord is NP-complete, New J. Phys. **16**, 033027 (2014).
[22] B. Dakić, V. Vedral, and Č. Brukner, Necessary and sufficient condition for nonzero quantum discord, Phys. Rev. Lett. **105**, 190502 (2010).
[23] A. Miranowicz, P. Horodecki, R. W. Chhajlany, J. Tuziemski, and J. Sperling, Analytical progress on symmetric geometric discord: Measurement-based upper bounds, Phys. Rev. A **86**, 042123 (2012).
[24] T. Tufarelli, T. MacLean, D. Girolami, R. Vasile, and G. Adesso, The geometric approach to quantum correlations: Computability versus reliability, J. Phys. A **46**, 275308 (2013).
[25] J.-H. Huang, L. Wang, and S.-Y. Zhu, A new criterion for zero quantum discord, New J. Phys. **13**, 063045 (2011).
[26] K. Roszak and L. Cywiński, Equivalence of qubit-environment entanglement and discord generation via pure dephasing interactions and the resulting consequences, Phys. Rev. A **97**, 012306 (2018).
[27] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).
[28] M. Christandl and R. Renner, Reliable quantum state tomography, Phys. Rev. Lett. **109**, 120403 (2012).
[29] J. Chai, H. Zeng, A. Li, and E. W. Ngai, Deep learning in computer vision: A critical review of emerging techniques and application scenarios, Mach. Learn. Applic. **6**, 100134 (2021).
[30] M. Alam, M. Samad, L. Vidyaratne, A. Glandon, and K. Iftekharuddin, Survey on deep neural networks in speech and vision systems, Neurocomputing **417**, 302 (2020).
[31] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).
[32] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 662 (2017).
[33] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Bridging many-body quantum physics and deep learning via tensor networks, in *Mathematical Aspects of Deep Learning*, edited by P. Grohs and G. Kutyniok (Cambridge University Press, New York, 2022), pp. 439–474.
[34] X. Cheng, S. Zhang, P. C. H. Nguyen, S. Azarfar, G.-W. Chern, and S. S. Baek, Convolutional neural networks for large-scale dynamical modeling of itinerant magnets, arXiv:2306.11833.
[35] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum entanglement in deep learning architectures, Phys. Rev. Lett. **122**, 065301 (2019).
[36] J. Roik, K. Bartkiewicz, A. Černoch, and K. Lemr, Accuracy of entanglement detection via artificial neural networks and human-designed entanglement witnesses, Phys. Rev. Appl. **15**, 054006 (2021).
[37] Y. Chen, Y. Pan, G. Zhang, and S. Cheng, Detecting quantum entanglement with unsupervised learning, Quantum Sci. Technol. **7**, 015005 (2022).
[38] J. Pawłowski and M. Krawczyk, Identification of quantum entanglement with siamese convolutional neural networks and semi-supervised learning, arXiv:2210.07410.
[39] N. Asif, U. Khalid, A. Khan, T. Q. Duong, and H. Shin, Entanglement detection with artificial neural networks, Sci. Rep. **13**, 1562 (2023).
[40] J. Ureña, A. Sojo, J. Bermejo, and D. Manzano, Entanglement detection with classical deep neural networks, arXiv:2304.05946.
[41] A. M. Palmieri, G. Müller-Rigat, A. K. Srivastava, M. Lewenstein, G. Rajchel-Mieldzioć, and M. Płodzień, Enhancing quantum state tomography via resource-efficient attention-based neural networks, arXiv:2309.10616.
[42] T. Cohen and M. Welling, Group equivariant convolutional networks, in *Proceedings of the 33rd International Conference on Machine Learning*, Proc. Mach. Learn. Res. Vol. 48 (PMLR, 2016), pp. 2990–2999.
[43] E. J. Bekkers, M. W. Lafarge, M. Veta, K. A. J. Eppenhof, J. P. W. Pluim, and R. Duits, Roto-translation covariant convolutional networks for medical image analysis, in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, edited by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Springer, New York, 2018), pp. 440–448.
[44] T. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling, Gauge equivariant convolutional networks and the icosahedral CNN, in *Proceedings of the 36th International Conference on Machine Learning* (PMLR, 2019), pp. 1321–1330.
[45] Y. Nagai and A. Tomiya, Self-learning Monte Carlo with equivariant transformer, arXiv:2306.11527.
[46] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, Long short term memory networks for anomaly detection in time series, in *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (ESANN, 2015).
[47] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, in *European Conference on Computer Vision* (Springer, New York, 2014), pp. 818–833.

[48] F. Yu and V. Koltun, Multi-scale context aggregation by dilated convolutions, in *Proceedings of the 4th International Conference on Learning Representations* (ICLR, 2016).

[49] V. Dumoulin and F. Visin, A guide to convolution arithmetic for deep learning, arXiv:1603.07285.

[50] A. Laha, A. Aggarwal, and S. Kumar, Random density matrices: Analytical results for mean root fidelity and the mean-square Bures distance, Phys. Rev. A **104**, 022438 (2021).

[51] F. Mezzadri, How to generate random matrices from the classical compact groups, Notices AMS **54**, 592 (2007).

[52] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[53] M. B. Plenio, Logarithmic negativity: A full entanglement monotone that is not convex, Phys. Rev. Lett. **95**, 090503 (2005).

[54] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).

[55] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, Phys. Lett. A **223**, 1 (1996).

[56] P. Horodecki, M. Horodecki, and R. Horodecki, Bound entanglement can be activated, Phys. Rev. Lett. **82**, 1056 (1999).

[57] D. Powers, Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation, J. Mach. Learn. Tech. **2**, 37 (2011).

[58] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, The balanced accuracy and its posterior distribution, in *Proceedings of the 2010 20th International Conference on Pattern Recognition* (IEEE, New York, 2010), pp. 3121–3124.

[59] H. Abraham *et al.*, Qiskit: An open-source framework for quantum computing (2019).

[60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems*, Vol. 32 (2019).

PHYSICAL REVIEW APPLIED **22,** 014068 (2024)

Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning

Jarosław Pawłowski* and Mateusz Krawczyk

Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, Wrocław 50-370, Poland

(Received 14 November 2023; revised 23 February 2024; accepted 8 July 2024; published 26 July 2024)

Quantum entanglement is a fundamental property commonly used in various quantum information protocols and algorithms. Nonetheless, the problem of identifying entanglement has still not reached a general solution for systems larger than 2×3. In this study, we use deep convolutional NNs, a type of supervised machine learning, to identify quantum entanglement for any bipartition in a three-qubit system. We demonstrate that training the model on synthetically generated datasets of random density matrices excluding challenging positive-under-partial-transposition entangled states (PPTES), which cannot be identified (and correctly labeled) in general, leads to good model accuracy even for PPTES states, that were outside the training data. Our aim is to enhance the model's generalization on PPTES. By applying entanglement-preserving symmetry operations through a triple Siamese network trained in a semisupervised manner, we improve the model's accuracy and ability to recognize PPTES. Moreover, by constructing an ensemble of Siamese models, even better generalization is observed, in analogy with the idea of finding separate types of entanglement witnesses for different classes of states.

DOI: 10.1103/PhysRevApplied.22.014068

I. INTRODUCTION

Modern deep-learning (DL) architectures, that use multilayer neural networks (NNs), have enabled unprecedented achievements in various domains like computer vision or natural language processing. Convolutional neural networks (CNNs) with many hidden layers and complex network structures are extremely powerful in feature learning, enabling automatic extraction of the most optimal features. Deep CNNs easily outperform classical algorithms in image classification [1], object detection [2,3], or face-recognition tasks [4]. In physics, one natural application of DL involves the study of quantum many-body systems [5–8], where the extreme complexity of many-body states often makes theoretical analysis intractable. Nonetheless, employment of machine learning (ML) in physics for entangled-state representations is typically focused on more traditional architectures such as Boltzmann machines [7,9] or fully connected (FC) NNs [10]. However, there are recent signals that deep convolutional and recurrent networks can better represent highly entangled quantum systems [11,12]. Therefore, it is worth trying more modern architectures, which are closer to the state-of-the-art approaches in DL, to study entanglement.

Entanglement is a fundamental feature of quantum physics, where the correlation between subsystems (i.e., particles) cannot be described within a *local* classical model. In quantum information theory, entanglement is regarded as a useful resource in achieving various tasks, such as quantum computation, cryptography, and teleportation. Simultaneously, the problem of identifying entanglement has not reached a general solution for systems larger than $2 \otimes 3$ [13,14]. There exist analytic criteria for separability of a given mixed state, which are either sufficient and necessary, but not practically usable (require minimization); or easy to use, but not conclusive [13]. One of the criterion of separability, which can be considered is the PPT (positive under partial transpose) criterion [15]. It is based on the idea, that having a density matrix partially transposed for a given bipartition, one can determine separability by simply looking at the eigenvalues of this matrix. Namely, if all eigenvalues are positive, then the state is separable in the given bipartition. However, in the case of systems larger than two qubits it is only a necessary condition of separability—there are no separable states, which have negative eigenvalues, but it is not a sufficient condition—there exist mixed states, which are entangled although they are PPT. Such states are called PPT entangled states (PPTES). Similarly, all other operationally feasible analytic criteria provide partial solutions only.

An interesting idea in this area is to implement NNs on a quantum computer [16,17] and thus create a bridge between DL and quantum computing [18], leading to

*Contact author: jaroslaw.pawlowski@pwr.edu.pl

 © 2024 American Physical Society

exponential speedup in training deep NNs [17]. Recently discovered quantum convolutional NNs [19] allow their efficient training and implementation on realistic quantum devices. Such networks can be used for efficient entanglement detection in multiqubit systems [20], as well as for solving different classes of quantum many-body problems such as quantum phase recognition [19]. ML can also be applied to a key problem in quantum algorithm design, the automatic optimization of large unitary matrices into sequences of two-qubit gates [21]. Entanglement can also assist ML image classifiers by compressing images and thus reducing computational resources needed [22].

Proposed studies will attempt to identify entanglement using modern DL architectures and confront it with traditional analytic entanglement metrics. The first applications of advanced DL architectures to such problems are already appearing, both with the utilization of unsupervised [23] and supervised manner [24,25]. In this work, we test a combined approach, which allows us to leverage the strengths and limit the weaknesses of both techniques. Specifically, a lack of appropriate conclusive criterion, which could be used as a label, is a huge problem for supervised learning, and susceptibility to mode collapsing of unsupervised models may lead to incomplete coverage of the desired space, and therefore misclassification of omitted states. Simultaneously, we aim to solve the more general problem of precise entanglement identification by predicting entanglement for each bipartition in a system. In this sense, our model will predict a vector of probabilities that identifies entanglement distribution in the system, rather than just a single entanglement metric. Moreover, Chen *et al.* [23] validate their method using PPT criterion only. Our method tries to deal with more general quantum systems in classifying PPTES, i.e., in situations where the PPT criterion fails, and one cannot obtain reliable labels for supervised learning, testing more advanced NN architectures than simple multilayer perception (MLP) networks [10]. We also obtain better results than using standard supervised trained NNs [24,25] and classical, non-neural ML algorithms [26,27]. Therefore, our main goal is to build a semisupervised machine-learning system, composed mainly of CNNs, that trained on verified states can only generalize at best on the PPTES family—as depicted in Fig. 1.

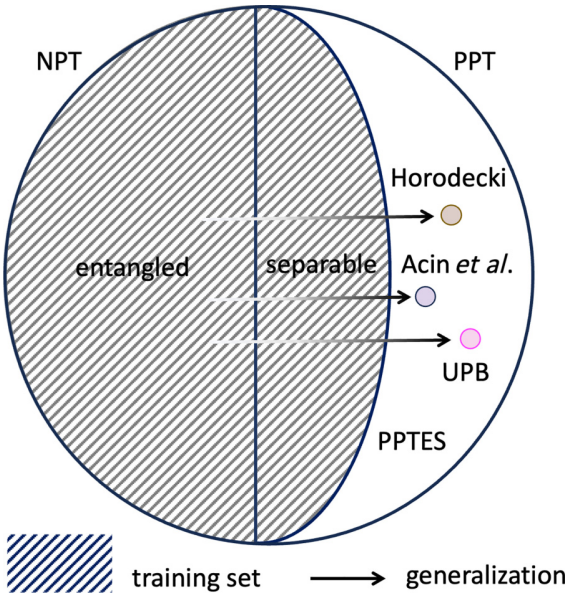


FIG. 1. Idea of the supervised training with the training set composed of states we know whether they are entangled or not (*verified* strategy). The goal is to build a model capable of generalizing to PPTES that are hard to recognize by analytical methods.

II. QUANTUM ENTANGLEMENT

Before discussing the introduced methods for the identification of entanglement in a given n-qubit quantum system, let us introduce basic concepts and definitions used when discussing this phenomenon. The foundations of quantum mechanics state that in order to describe a system that forms a n-qubit register, one must define a 2^n-dimensional state space (known as the Hilbert space). The most convenient (and general) way to describe these multiqubit systems is to construct a density matrix:

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{1}$$

where p_i is a probability that the system is in a given state $|\psi_i\rangle$. In this general approach, a quantum system is represented by a mixture of states, or in other words a mixed state. However, if only one state is used to determine the whole system, then it is said, that it is a pure state. In such a case, the density matrix is reduced to the outer product of a state vector $|\psi\rangle$ with itself $\rho = |\psi\rangle \langle\psi|$.

Having a complex system, i.e., a system consisting of at least two qubits, one can define a separable state as the one, that can be separated to a tensor product of subsystem states. This definition is valid in the case of pure states, however, is has to be extended to capture mixed states. In the latter case, the biseparable mixed state can be written as

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i, \tag{2}$$

where $\sum_i p_i = 1$, while $\{\rho_A^i\}$ and $\{\rho_B^i\}$ are the density matrices representing states of the corresponding subsystems A and B. If the mixed state cannot be represented as such a convex sum ($p_i \geq 0$) of product states, then the state is entangled.

In general, determining the separability of a given state using the above-mentioned definition is a nontrivial problem [13]. Therefore, many different criteria have been constructed to decide whether a state is entangled. One noteworthy example is the negativity [28] metric, which is defined as follows:

$$\mathrm{Neg}(\rho_{AB}) = \sum_{\lambda_i<0} |\lambda_i|, \tag{3}$$

where λ_i are the negative eigenvalues of the partially transposed density matrix ρ^{T_B} associated with the given bipartition $A|B$ of the whole system. The partial transpose operation for its density matrix $\rho_{AB} = \sum_{i,j,r,s} a_{i,r} a^*_{j,s} |i\rangle_A \langle j|_A \otimes |r\rangle_B \langle s|_B$, with respect to the subsystem B is defined as

$$\begin{aligned}\rho^{T_B} &= \sum_{i,j,r,s} a_{i,r} a^*_{j,s} |i\rangle_A \langle j|_A \otimes (|r\rangle_B \langle s|_B)^T \\ &= \sum_{i,j,r,s} a_{i,r} a^*_{j,s} |i\rangle_A \langle j|_A \otimes |s\rangle_B \langle r|_B \,. \end{aligned} \tag{4}$$

The negativity metric Neg(.) is based on the PPT separability criterion [29,30]. It states, that non-negative eigenvalues of the partially transposed density matrix are the necessary condition of separability. However, it is a sufficient condition only for $2 \otimes 2$ or $2 \otimes 3$ systems. In the case of larger systems, one can be certain that if the negativity is larger than zero (NPT states), then the state is entangled. On the other hand, if the negativity is equal to zero (PPT states) it is not always true that the state is separable, i.e., we may have PPTES.

There exist some other methods of calculating entanglement, for instance, entanglement of formation or entanglement witnesses, however they are far more complex (time consuming) because they require optimization in high-dimensional spaces [31–33], and therefore impractical to use in case of large datasets. All in all, one arrives with analytical methods that are either hard to calculate or sufficient only for pure states or low-dimensional mixed states. On the other hand, there exist numerical algorithms, which are able to qualify the entanglement after a finite time with a desired precision [34–36], however, our attempts to use them as *numerical* labels were rather unsuccessful due to a problem with convergence of these algorithms. That is why artificial NNs, which are also approximate numerical methods, but statistical in nature, are supposed to cope with this task.

III. NEURAL NETWORK ARCHITECTURES

Detecting quantum entanglement in many-qubit systems is inseparably connected with analyzing the quantum state of the system, which is represented by a density matrix. The most natural concept of a NN designed to process such an input (in the form of a finite two-dimensional array) is a convolutional NN (CNN). This type of ML model has already proved its efficiency in dealing with quantum entangled systems [11,37–40]. In this paper, we propose to use CNN as a feature-extracting backbone followed by FC classification layers in three different configurations presented in Fig. 2.

A. Deep convolutional networks

The first network architecture composed using a single CNN is shown in Fig. 2(a). The input layer has size $K \times K \times 2$, where $K = 2^N$ is the Hilbert-space dimension and the density matrix is split into the two channels by taking the real and the imaginary part separately. Subsequently, there are three convolutional layers with ReLU activation functions, each with a kernel of size 2×2, and the number of channels in the ith layer $c_i = \lfloor r_i c_{i-1} \rfloor$. Here, the $r_i = \sqrt{r_{i-1}}$ parameter is the layers increasement ratio initialized with $r_1 = 16$, and $\lfloor \cdot \rfloor$ denotes *floor* operation. Convolutional layers are followed by five fully connected layers with 128 units each and ReLU activation functions. Finally, the m-unit output vector with sigmoid activation functions is used, where m corresponds to the number of possible bipartitions in the N qubits' system. The architecture of the CNN model for the three-qubit system is presented in Fig. 2(a), however, this construction can be easily extended to a larger N-qubit register.

The network is trained in a supervised manner. To evaluate the loss function we use binary cross entropy averaged over all possible bipartitions:

$$\begin{aligned}\mathcal{L}_{\mathrm{CNN}} = L_{\mathrm{BCE}}(\mathbf{p}(\rho), \mathbf{q}) = \frac{-1}{mn} \sum_{j}^{m} \sum_{i}^{n} q_{ij} \log(p_j(\rho_i)) \\ + (1 - q_{ij}) \log(1 - p_j(\rho_i)),\end{aligned} \tag{5}$$

where m is the number of bipartitions, n is the batch size, q_{ij} is the label for the j th bipartition of the given state ρ_i (with 0 meaning separable, and 1—entangled bipartition), and $p_j(\rho_i)$ is the predicted probability that ρ_i has parts for j th bipartition entangled.

B. Siamese networks

Entanglement should be invariant to local operations taken on separate qubits. In other words, applying single-qubit rotations or phase modification does not influence the entanglement between different qubits. Moreover, permutations of the qubits in the system also should not have any impact on the entanglement identification. Having those two facts in mind, we propose to extend the CNN into a triple Siamese NN, to force the model to be immune to such symmetry operations, and thus improve its robustness. Specifically, in this approach instead of having only one density matrix on the input, we augment it

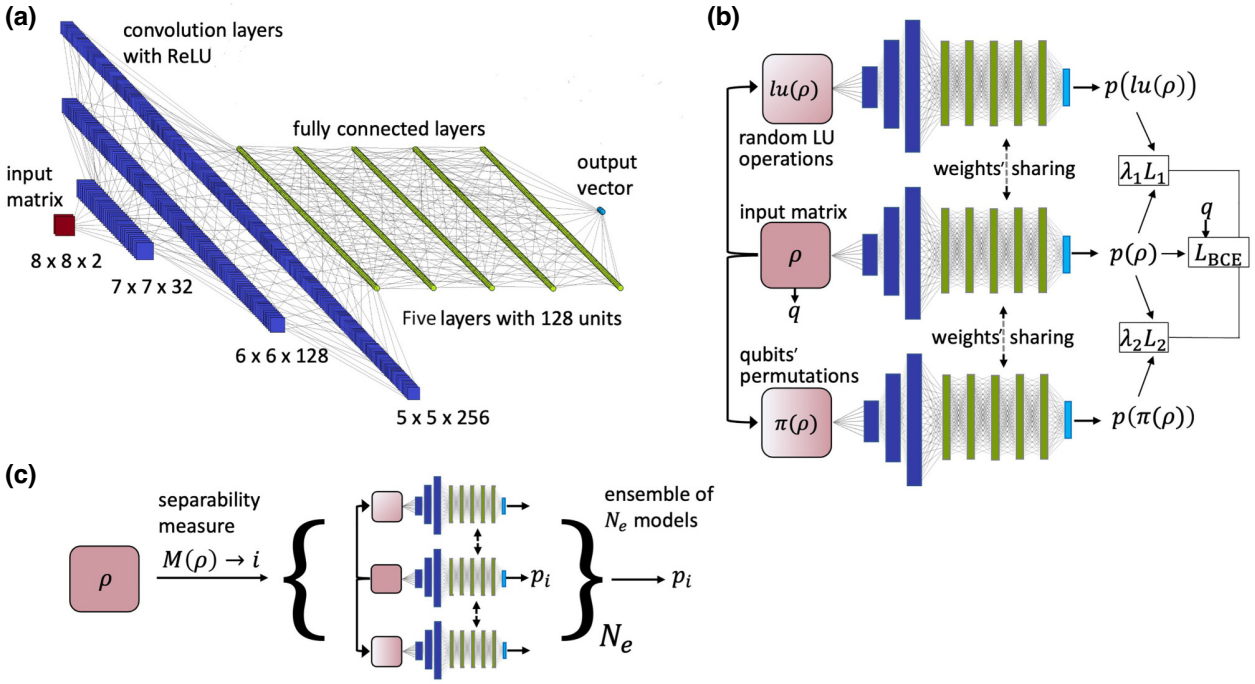


FIG. 2. Model architectures: (a) NN for three-qubit input density matrix (red): CNN backbone with three convolutional layers, each of 2×2 kernels (dark blue blocks); five fully connected layers (green blocks); and finally, three-element output vector identifying entanglement in three possible bipartitions (light blue). (b) Architecture of the triple Siamese network, where each subnetwork, being the CNN from the model (a), is fed by a different input (original or transformed), but shares the same network parameters (weights). (c) Ensemble of Siamese networks with separate N_e Siamese models trained on different domains determined using separability measure M (being a pretrained autoencoder model).

into three matrices, which are processed parallelly. The characteristic is that each subnetwork is an exact copy of the original CNN but with the model weights *shared* between them—see Fig. 2(b). The first input is the original matrix, the second one represents the original state with extra random local unitary operations (LU) applied to each qubit, and the third one corresponds to the density matrix, which is permuted with respect to randomly selected qubits. Hence, now the loss for a given batch is defined as follows:

$$\begin{aligned}\mathcal{L}_{\text{siam}} = \lambda_0 L_{\text{BCE}}\left(\mathbf{p}(\rho), \mathbf{q}\right) \\ + \lambda_1 \frac{1}{mn} \sum_{j}^{m} \sum_{i}^{n} \left| p_j(\rho_i) - p_j(\text{lu}(\rho_i)) \right| \\ + \lambda_2 \frac{1}{mn} \sum_{k=\pi(j)}^{m} \sum_{i}^{n} \left| p_j(\rho_i) - p_k(\pi(\rho_i) \right|, \end{aligned} \tag{6}$$

where by $\text{lu}(\rho_i)$ we denote some random LU operation on state ρ_i, $\pi(\rho_i)$ is the permutation of state ρ_i chosen independently for a given batch, and λ_1, λ_2 are hyperparameters, which allow adjustment of the influence of such extensions of the loss function. At the beginning $\lambda_0 = 1$, $\lambda_1 = 0$, and $\lambda_2 = 0$, while during the training after 10 epochs, we increased the λ_2 and λ_3 regularizers to 0.5. This shifts the impact from the label-based part of the loss (controlled by λ_0) to the symmetrizing (i.e., enforcing symmetries) part (λ_1 and λ_2), which is a typical approach for *semisupervised* training schemes. One remark is, that permutation π can rearrange the output vector in the case of certain bipartitions. For example, three-qubit permutation $\pi = (1, 3, 2)$ does not change the predicted probability of entanglement for the bipartition $(1|23)$. Contrarily, the same permutation switches the probabilities for the other two bipartitions $(2|13$ and $3|12)$. Therefore, corresponding indexes must be selected carefully, which we denoted by writing $k = \pi(j)$.

C. Ensemble of networks

During the experiments, we observed that there is a significant difficulty in forcing NN models to recognize mixed separable and entangled states at the same time. Thus, to cope with this issue we decide to train several Siamese models, forming *ensemble* of size N_e, that work separately on their domains—see the scheme in Fig. 2(c). Different classes of states falling into separate domains can be distinguished on the basis of some external measure M that quantifies their separability. To obtain such a measure, we use a pretrained autoencoder NN model, introduced in our

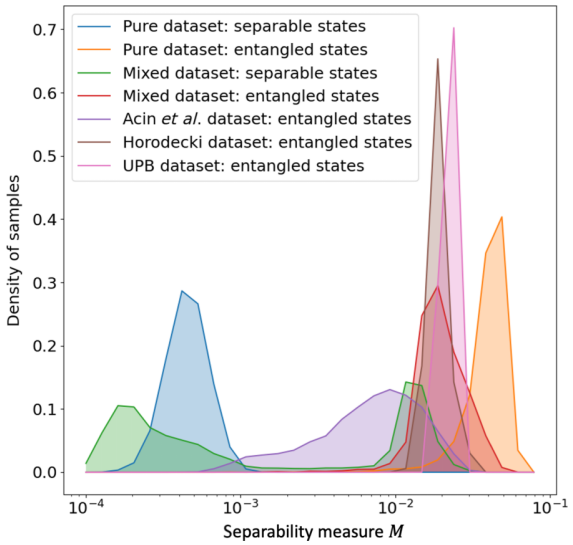


FIG. 3. Idea of the ensemble method: various states can be distinguished on the basis of some external measure (here defined by a pretrained autoencoder loss M) and then learned separately by *ensemble* of models that work independently in their M domains.

recent work [41]. The autoencoder was trained to reconstruct separable mixed states, hence the reconstruction loss can measure whether a given state is similar to separable states, and consequently serve as machine-learned separability measure M. Having values of M for different states, we form $N_e = 10$ domains with logarithmic intervals $[0, 0.0001)$, $[0.0001, 0.0002)$, $[0.0002, 0.0005)$, $[0.0005, 0.001)$, $[0.001, 0.002)$, $[0.002, 0.005)$, $[0.005, 0.01)$, $[0.01, 0.02)$, $[0.02, 0.05)$ and $[0.05, \infty)$.

In Fig. 3 distributions of samples from various subsets are presented, appearing in the training and test sets, according to the value of separability measure. It is seen that the chosen measure separates well the separable and entangled states, apart from some specific group of mixed separable states, which are confused by the network with entangled states. In fact, if this subset had not existed, one could have simply utilized the unsupervised autoencoder as the perfect tool to distinguish between separable and entangled states, as proposed in Chen *et al.* [23] However, with the current approach of training supervised models specifically on the subsets of the data set, we can force the networks to learn the characteristic patterns of the given subgroups, and therefore allow them to detect states omitted by the autoencoder. One can notice, that it is in full agreement with the analytical approach of finding separate types of entanglement witnesses for different classes of states.

What may also be beneficial, the various PPTES classes are separated somewhat with respect to the M measure (cf. Fig. 3). Interestingly, the distribution of the Acin *et al.* states seems to be the closest to the separable states, hence correct classification of these states may be the most problematic.

Nevertheless, due to its specification, the ensemble model should be able to recognize entanglement patterns with higher accuracy than standard models and generalize well on PPTES states, which were not present in the training datasets as shown on the scheme in Fig. 1.

IV. TRAINING DATA GENERATION

The most crucial thing to train a ML model is to have a well-balanced and diversified dataset. Thus, in order to achieve this goal, we propose various methods of density matrices' generation, that will be used to build synthetic datasets for training our NN models.

Two main techniques were used to generate pure random states. The first idea is to use a quantum circuit model that contains local single-qubit gates U that randomize separate qubits, and the random number of two-qubit controlled gates CU, that can entangle qubits. The CU gates are put into random pairs of qubits. Combining these two types of quantum gates with their appropriate parameters (angles), one can obtain a suitable generator of pure states, both entangled and separable ones. The detailed parametrization of the gates and circuit organization can be found in the Appendix. An example of a three-qubit circuit generated by this method is presented in Fig. 4. The second idea of generating random pure states is based on sampling from the uniform Haar measure [42] described in the Appendix.

Having these two methods, one can suspect that using both of them produces a sufficiently diversified dataset of the pure-state vectors and consequently density matrices. However, apart from them, we also included three specific types of quantum states, which may hardly occur when using both procedures. Namely, using the circuit approach we intentionally generate pure separable states (here CU gates are absent in the circuit), the GHZ state, and the W state [43], each of them being locally randomized with the U gates. It is obvious, that having only pure states included in the training dataset may result in missing the possibility of detecting mixed states. That is why, we should also think of generating such states straight from the definition of a mixed state:

$$\rho = \sum_{i}^{d} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{7}$$

where for training purposes we select d at random over the interval $[2, 2^N]$, and N is a number of qubits. In this naïve method one simply mixes many pure states $|\psi_i\rangle$ with random probabilities p_i. However, to have as diversified dataset as possible, we also use the second method

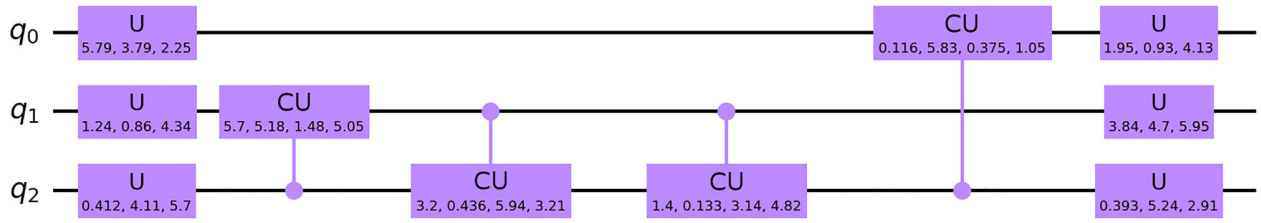


FIG. 4. Example of a circuit generating a three-qubit random state. It consists of U and CU gates (violet blocks) with random parameters written in each block (and defined in the Appendix).

of generating mixed states:

$$\rho_{AB} = \mathrm{tr}_C(|\psi_{ABC}\rangle \langle\psi_{ABC}|), \tag{8}$$

with a partial trace operation $\mathrm{tr}_C(\cdot)$. Namely, we generate a larger pure system $|\psi_{ABC}\rangle$ and trace out some part of it (subspace C), until the desired size of the system AB is achieved. This way, if the global space ABC is entangled, one can be sure that the reduced density matrix ρ_{AB} is mixed.

To train a NN in a supervised learning scheme one needs to have labels q—see Fig. 2(b), which classify each bipartition of the system in a given state as entangled or separable. Then, each state ρ_i is labeled by a binary vector q_{ij}, where subsequent elements j encode the presence of entanglement for the given bipartition. In the case of pure states, we calculate the negativity measure [Eq. (3)] for each bipartition and mark it accordingly to its value: if $\mathrm{Neg} = 0$, it is marked as separable, while if $\mathrm{Neg} > 0$ it is marked as entangled. In the case of mixed states, where it is not guaranteed that the negativity measure distinguishes separable and entangled states, we either generate states that are separable (by definition) or entangled (by taking only NPT, i.e., where the negativity is conclusive), or in case of PPT we propose two different strategies for labeling unknown states. This latter group is key because it contains PPTES, and therefore allows us to train a model to capture them.

The most naïve strategy is based on negativity metric and simply labels mixed states accordingly to the Neg value—we will call this strategy as *negativity labeled*. However, the main problem with this approach is that it may falsely label mixed PPT states, i.e., where the criterion is inconclusive. That is why we introduce a second strategy called *verified*, where the PPT states are simply excluded from the data. The main disadvantage of such a solution is that it automatically excludes all of mixed separable states. To minimize this drawback, we supplement the dataset by generating them straight from the definition. Further descriptions of the labeling strategies can be found in the Appendix.

Having defined various methods for the generation of pure and mixed states together with two different strategies for labeling mixed states we have generated training datasets. They are summarized in Table I, along with the number of generated states for each group. In total, we have synthesized two training sets, each with the same structure of the pure (120 000 states) and mixed separable (160 000 states), but with different strategies for labeling mixed entangled (160 000 states) and mixed biseparable (120 000 states).

TABLE I. Training datasets composition. Both sets have the same structure of pure and mixed separable states but have different strategies for labeling mixed entangled states.

State type	Generating method	Labeling method	Amount
Pure separable	Circuit with random one-qubit gates only (separable circuit)	Negativity	40 000
Pure biseparable	Kronecker product of one-qubit state and two-qubit state, sampling from Haar measure	Negativity	20 000
Pure entangled	Random circuits, including W and GHZ, and sampling from the Haar measure	Negativity	60 000
Mixed separable	Separable circuit, Kronecker product (A4) of random one-qubit states, and Kronecker product (A4) of traced one-qubit states	By definition	20 000+ 120 000+ 20 000
Mixed biseparable	Kronecker product (A5) of one-qubit state, and two-qubit state	Two strategies: *negativity*, *verified*	120 000
Mixed entangled	from the definition (7), and using partial trace (8)	Two strategies: *negativity*, *verified*	100 000+ 60 000

TABLE II. Averaged bipartite entanglement detection accuracy for the three-qubit system using different models trained on *verified* dataset. The results are presented for test sets composed of pure and mixed states, together with the various classes of PPTES. The last row shows results for the MLP model that operates directly on measurements set while not on the (reconstructed) density matrix.

Models	Pure states	Mixed states	Horodecki PPTES	UPB PPTES	Acin PPTES
CNN	97.40	86.39	83.01	88.85	57.45
Siamese CNN	98.31	90.79	86.99	94.89	50.81
Ensemble	98.18	87.88	88.63	89.58	87.36
MLP based on measurements	79.31	79.31	89.67	99.81	36.30

In order to correctly evaluate the model, apart from the training sets, the validation and test sets were also created. The first validation set is generated with the same structure as the *verified* training dataset but with 10 times smaller number of examples, i.e., 56 000 states generated. Additionally, the second validation data set consisting of 10 000 $2 \otimes 2^{N-1}$ dimensional PPTES states [44] is generated as well. Next, we defined two test datasets for pure (30 000 states) and mixed states separately. The latter contains 20 000 separable states, and 20 000 entangled containing NPT states only (*verified* strategy), generated both from the definition [Eq. (7)] and using partial trace [Eq. (8)]. Therefore, the test dataset contains only the correctly labeled states. Further details on states' generation can be found in the Appendix.

Since the main reason for creating the artificial classifier is to have the ability to detect not only NPT entangled states but also PPTES, we generate three separate test datasets, each containing 10 000 well-defined PPTES of different types. Precisely, the first PPTES test set includes $2 \otimes 4$ PPTES introduced by Horodeccy [15] and then generalized to N qubits, $2 \otimes 2^{N-1}$-dimensional states [44]. The second test set is generated with the use of unextendible product basis (UPB) [45], with a similar approach being described by Ma *et al.* [10] The third one is composed of fully PPT entangled states proposed by Acin *et al.* [46] All of these states are further randomized by applying local operations to the separate qubits.

V. RESULTS

After we have defined the deep NN models and generated the labeled training and test sets, we can move on to training and then testing our models. To verify the models' fidelity we used the standard accuracy metric, which is defined as follows:

$$\mathrm{Acc} = \frac{1}{mn} \sum_{i=1}^{m} TP_i + TN_i, \tag{9}$$

where m is the number of possible bipartitions, TP_i and TN_i represent the number of true positive (entangled classified as entangled) and true negative (separable classified as separable) examples for a given bipartition i, and n is the total number of samples in the test datasets.

We also optimized the architecture hyperparameters by training the models on the *verified* set and evaluating on the validation sets—discussion can be found in the Appendix. The results show that the most efficient architecture is the network (CNN or Siamese CNN) with three convolutional layers and a kernel of shape 2×2.

Let us analyze the results for the smallest configuration where nontrivial entanglement relations may occur, i.e., three-qubit system. In Tables II and III results are presented for CNN, Siamese, and Ensemble models, trained on both training sets: *verified* and *negativity labeled*. The accuracy was calculated for all generated test sets: with pure states, mixed states, and with various PPTES classes.

When analyzing Table II, the first thing, that one can notice, is that all models work pretty well in the case of pure states, reaching accuracies over 97%. However, as one might expect, the detection of mixed states is slightly weaker. When it comes to comparing CNN, Siamese CNN, and Ensemble networks one can see the advantage of the Siamese CNN model in better recognizing mixed states than CNN: 91% vs. 86%. In the case of PPTES, we also observe the advantage of using the Siamese model, which gives better for Horodecki (87% vs. 83%) and UPB states (95% vs 89%). However, in the case of Acin states the fidelity for the Siamese network is slightly lower (51% vs. 57%)—we suspect that this is due to the fact that Acin states are "closer" to separable states than two other PPTES (cf. Figure 3), causing the Siamese detector (which is more sensitive to PPTES) to generate more false-positive predictions. It is worthwhile to recall here, that all models were trained on the datasets, which do not include PPTES. This means, that the models can generalize well to Horodecki and UPB states, and slightly worse to Acin states.

To handle unsatisfactory results for the Acin states, we introduced the ensemble of networks and thus were able to train several models to capture different PPTES classes separately. This approach improved accuracy for Acin states up to 87% with slightly lower (than for Siamese) accuracy for UPB: 90%, and slightly better for Horodecki: 89%. In Fig. 5 we plot accuracies for the subsequent models' domains in the ensemble showing that Acin PPTES are quite well detected over almost all covered domains. This proves the usefulness of the ensemble approach where each

TABLE III. Test results as in Table II, but now the models trained on the *negativity labeled* dataset.

Models	Pure states	Mixed states	Horodecki PPTES	UPB PPTES	Acin PPTES
CNN	97.37	88.89	82.80	95.78	5.71
Siamese CNN	98.43	91.49	83.04	87.39	1.04
Ensemble	98.13	89.71	84.51	90.61	4.01

model specializes in detecting entangled states included only within its own domain.

As the standard representation for the quantum system is a density matrix, hence naturally we proposed to use it as the model input. However, in principle, one could also train the models to operate directly on measurements set. Such an approach may seem beneficial as it reduces the necessity to tomographically reconstruct the density matrix from measurements [47]. On the other hand, it simultaneously limits the models' ability to capture the structural representation of the data, which is contained only in the arrangement of matrix coefficients. To compare those two approaches we trained a simple MLP, with measurements on the input, getting slightly worse results, as presented in Table II(last row), besides UPB states, for which the model works quite well. Therefore, we conclude that using a density matrix is more efficient strategy, especially that one can always either perform the standard tomography [47] before proceeding with NNs, or embed equivalent linear transformation into the NN model, or even train a separate NN to reconstruct ρ from the complete measurement set.

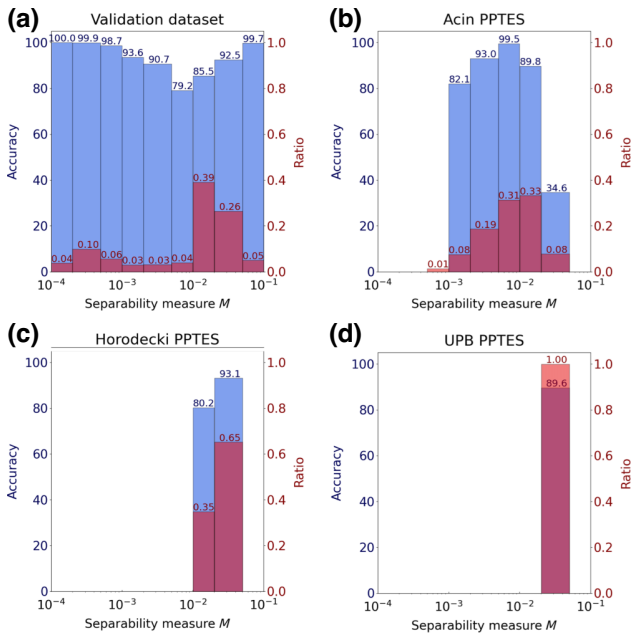


FIG. 5. Accuracy (blue bars) of the Ensemble models plotted separately for the subsequent models' domains, determined using the separability measure M. Ratio (red bars) characterizes the population of the subsequent domains by the states included in various datasets: (a) validataion, (b) Acin PPTES, (c) Horodecki PPTES, (d) UPB PPTES. Purple color means that blue and orange bars overlap.

Contrasting the above results with those for *negativity labeled* training set, presented in Table III, we notice slightly higher accuracy for the general set of mixed states, but worse for the PPTES classes, especially in the Acin states. However, we still observe some ability to detect Horodecki and UPB states, suggesting that NNs learned to recognize some features of entanglement more general than the negativity criterion itself.

VI. DISCUSSION

The main problem with building a general classifier of entanglement in quantum states is a lack of reliable methods for labeling mixed states. If we label some states with mistakes, the ML methods may learn from the mistakes as well. Thus, while it is good to see the high accuracy of the models for pure or mixed NPT states, PPTES are our main concern.

However, we were able to show that if we use a properly diversified training dataset, we can force the model to generalize correctly to capture PPTES even if they were not present in the training set. This effect can even be strengthened if we use the architecture of Siamese CNN trained in a semisupervised scheme with regularization terms that force the model to respect the system symmetries, that do not change the entanglement configuration (LU or respective permutations).

Improved generalization efficiency through the use of Siamese networks shows the possible direction for the development of ML models for describing physical systems with their inherent symmetries.

Unfortunately, Siamese CNN performs worse with a certain group of states that are inherently closer to separable states, i.e., the Acin states. To overcome this, we introduced the Ensemble model, which is combined with a separability measure (unsupervised autoencoder) that enables the training of multiple models tuned to different groups of PPTES, and is capable of learning the characteristic patterns of the given subgroups. The autoencoder itself, which is trained in an unsupervised way (on separable states), does not fully distinguish between mixed separable and entangled states, but extending it with an ensemble of supervised models leads to synergistic improvement. There are no PPTES in the training set (training is PPTES agnostic), thus we can suspect that the

TABLE IV. Scaling properties of the Ensemble model: comparison of detection accuracies between three- and four-qubit systems.

System size	Train dataset	Pure states	Mixed states
Three qubits	92.29	98.18	87.88
Four qubits	96.28	97.65	91.58

Ensemble model should give equally good results for any other type of PPTES.

During the experiments, we also tested other NN architectures. However, both generic multilayer FC networks, as well as more advanced tensor networks [48], used as a backbone model instead of CNN, gave a worse performance. This shows that convolutional networks can be well suited to efficiently represent entanglement in multiqubit systems.

To verify whether the proposed approach can be scaled to larger multiqubit systems we calculate results presented in Table IV, showing that the Ensemble model works correctly also for detecting bipartite entanglement in a four-qubit system. Accuracy for the detection of pure states is only slightly worse, while for mixed states is better, which should be taken with caution. Particularly, if we look at the accuracy for the training set we notice that it is also higher in the four-qubit case, which misleadingly implies that identifying the entanglement is easier in a larger space. In fact, it simply means that the generated dataset does not cover the complexity of this space (especially in the case of mixed states), and future works should focus on improving the generation methods of mixed states in higher-dimensional space. Nevertheless, the model architecture seems to scale correctly.

Entanglement is a key feature of quantum systems with huge application potential in quantum computation. Controllable entangled states are resources for quantum communication, quantum teleportation, or quantum key distribution protocols (quantum cryptography). Not all quantum states are equally valuable as a resource [49]. To quantify this value, different entanglement measures can be used. In real situations, entanglement measures are difficult to compute for an arbitrary mixed state (in a situation when a state is a part of a larger system) as the dimension of the entangled system grows. From this perspective, searching for more reliable and scalable methods of entanglement identification in quantum resources is desirable, in particular usage of neural-network-based representations seems to be an interesting method for developing alternative entanglement detectors.

VII. METHODS AND CODE AVAILABILITY

To implement the quantum circuit model, during the training dataset generation, the python Qiskit [50] library is used. Also, when sampling from the uniform Haar measure during generation we use the built-in Qiskit method. All the NN models were implemented, trained, and evaluated using the PyTorch [51] library that supports autograd mechanism. The typical training time depends on the model size and ranges from 1 h for the pure CNN model to 3–4 h for the Siamese CNN and 20 h for the Ensemble model. To train the NN models we used Nvidia RTX 4090 GPU. The whole dataset generation lasts about 1 day for three-qubit density matrices, using a typical few-core CPU. The code with the neural models's definition and detailed training schemes, as well as the data generation procedures, are available at Ref. [53].

ACKNOWLEDGMENTS

Authors acknowledge support from National Science Centre, Poland, under Grant No. 2021/43/D/ST3/01989.

APPENDIX

1. Circuit for random states' generation

Here we present two methods for generating random pure states.

The first idea is to use a quantum circuit model. It is a common technique in quantum computing when one needs to process states of many qubits. In this approach, firstly we initialize the set (register) of N qubits in a given state (usually $|0\rangle^{\otimes N}$), and then apply so-called quantum gates. One can distinguish two main types of such operations: the local—single-qubit gates, and the two-qubit controlled gates. Although both of them can be understood as rotations on the qubit sphere by the given angles, in the latter such operation is performed only on the single qubit while the rest of the qubits can be used to determine (control) whether this operation should or should not be performed. A controlled gate can be used to entangle two qubits. In contrast, single qubit gates can only modify the local states of qubits and not influence the global entanglement properties. Therefore, by combining these two types of quantum gates with appropriate parameters (angles), one can obtain a suitable generator of pure states, both entangled and separable ones. The universal single-qubit gate is defined as follows:

$$U(\theta,\phi,\lambda)=\begin{pmatrix}\cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2}\\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2}\end{pmatrix}, \tag{A1}$$

where θ, ϕ, and λ are three Euler angles. Subsequently, the controlled universal gate is described as

$$CU(\theta,\phi,\lambda,\gamma)=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & e^{i\gamma}\cos\frac{\theta}{2} & -e^{i(\gamma+\lambda)}\sin\frac{\theta}{2}\\ 0 & 0 & e^{i(\gamma+\phi)}\sin\frac{\theta}{2} & e^{i(\gamma+\phi+\lambda)}\cos\frac{\theta}{2}\end{pmatrix}, \tag{A2}$$

with additional global phase $e^{i\gamma}$. To obtain the most possible diversified dataset, the parameters of these two gates are chosen at random, where the range for each parameter θ, ϕ, λ, and γ is equal to $[0, 2\pi)$.

Let us now describe the circuit organization. At first, the system is initialized in the random separable state by applying the U gates to each of the qubits in the register. Next, the random number of CU gates is picked from the range $[1, 2\binom{N}{2})$, where N is the number of qubits. The CU gates are put to random pairs of qubits. And finally, the state is locally randomized, again, by applying the U gates.

The second idea of generating random pure states is based on sampling from the uniform Haar measure. At first, we generate a vector $\mathbf{x}$ of size $K = 2^N$ (N is the number of qubits) with its elements being uniformly sampled over an interval $(0, 1]$. Secondly, we replace each coordinate x_i with $y_i = -\log(x_i)$, therefore obtaining vector $\mathbf{y}$ with logarithmic distribution over an interval $[0, \infty)$. Next, we generate a K-dimensional phase vector γ, where each coordinate is chosen at random from the range $[0, 2\pi)$. Finally, we construct a state vector as

$$|\psi_{\text{Haar}}\rangle = \sqrt{\frac{y_1}{\sum_i y_i}} e^{i\gamma_1} |00\cdots0\rangle + \cdots + \sqrt{\frac{y_K}{\sum_i y_i}} e^{i\gamma_K} |11\cdots1\rangle . \tag{A3}$$

2. Strategies for state labeling

We propose two different strategies for labeling PPT mixed states, i.e., in situations where the negativity criterion is inconclusive.

The most simple solution is to trust the negativity metric and always label mixed state accordingly to the Neg value—we will call this strategy as *negativity labeled*. However, such an approach may result in a network trained to reflect just the PPT criterion instead of actual entanglement. That is why we can think of the second method, which simply excludes mixed PPT states (with Neg $= 0$). We will call this strategy as *verified*, because all the labels here are again correct. The main disadvantage of such a solution is that it automatically excludes all of the mixed separable states from the training set. To minimize this drawback, we generate some extra mixed separable states straight from the definition:

$$\rho_{abc} = \sum_i p_i \rho_a^i \otimes \rho_b^i \otimes \rho_c^i, \tag{A4}$$

where one can generate single-qubit states ρ_a^i, ρ_b^i, ρ_c^i, and combine them with Kronecker product into the three-qubit mixed separable state ρ_{abc}. This approach can be easily generated for N-qubit states just by extending the number of single-qubit states. Similarly, we can generate biseparable states as

$$\rho_{abc} = \sum_i p_i \rho_a^i \otimes \rho_{bc}^i, \tag{A5}$$

where single-qubit state ρ_a^i is combined with two-qubit state ρ_{bc}^i. Subsequently, in order to obtain states biseparable in different bipartition qubits are randomly permuted.

3. Generated datasets' structure

Having two strategies for labeling mixed states (labeling pure states is obvious), we consider generating separate training datasets, each corresponding to the different strategy: *negativity labeled* or *verified*. All of them retain the same structure of pure states: 40 000 pure separable states, and 60 000 pure entangled states—generated using random circuits (using U [Eq. (A1)] and CU [Eq. (A2)] gates) including W and GHZ, and sampling from Haar measure method [Eq. (A3)]. Additionally, 20 000 pure biseparable states are generated by sampling one- and two-qubit states from Haar measure and then combining them with the Kronecker product.

In the case of mixed states, we generate 20 000 mixed separable states using separable circuits and formula [Eq. (A4)]. Further 60 000 separable states are generated as a Kronecker product of mixed single-qubit states obtained with the usage of quantum circuits. A similar approach is used in order to generate another 60 000 states, with the remark that single-qubit states are sampled from Haar measure, and 20 000 states, where single-qubit states are obtained by tracing from the larger system. Next, we generate 120 000 biseparable mixed states, by creating single and two-qubit states and then, combining them into three-qubit states by applying the Kronecker product. The first half of these states is obtained by the usage of quantum circuits and another half is sampled from Haar measure. Then, we generate 100 000 mixed entangled states by mixing pure entangled states explicitly from the definition [Eq. (7)], and keeping only those with Neg > 0 (NPT). Finally, we add 60 000 mixed states generated with the use of the partial trace from a global pure entangled state [Eq. (8)]. For the last group in which we have no guarantee that states are either separable or entangled, we apply the two labeling strategies. Additionally, the first training dataset, with the *verified* labeling strategy, is extended with 20 000 of mixed states with maximal ranks calculated as a partial trace of the larger fully entangled global system generated with the quantum circuit. Those extra states are handily labeled as entangled.

To verify that generated sets homogeneously cover the areas of three-qubit space for separable and entangled states, we present Fig. 6 showing distributions of above-mentioned groups of states as a function of two measures: negativity and separability M. It can be observed

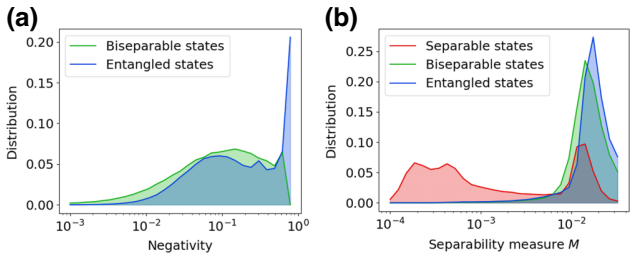


FIG. 6. Distribution of the training dataset plotted against (a) negativity metric and (b) separability measure M. Note that for separable states Neg $= 0$ by definition, thus they are not visible on the (a) plot.

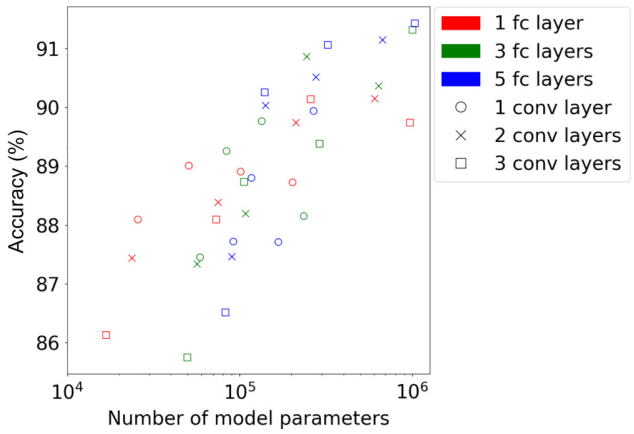


FIG. 7. CNN model architecture scaling. Number of parameters is increased either by changing the number of fully connected layers (denoted by different colors), convolutional layers (different marker types) or directly by increasing the number of convolutional channels. The accuracy is measured on the first validation dataset.

that all these three sets fill the corresponding areas fairly homogeneously.

In order to correctly evaluate the model, apart from the training sets, the validation and test sets were also created. The first validation set is generated with the same distribution as the verified training set but with a reduced number of examples, i.e., 10% meaning 56 000 states generated. The second one consists of 10 000 $2 \otimes 2^{N-1}$-dimensional PPTES states [44]. Next, we define two test datasets for pure and mixed states separately. The former consists of 15 000 separable states, (separable circuits and Kronecker product of single-qubit states), and 15 000 entangled states (random circuits and sampling from Haar measure). The latter contains 20 000 separable states (mixture of separable circuits and Kronecker product), and 20 000 entangled containing NPT states only, generated both from the definition [Eq. (7)] and using partial trace [Eq. (8)].

4. Network optimization

Having defined the CNN and the extended Siamese network model, we optimized their architecture hyperparameters (e.g., number of layers or kernel size) by training them on the weakly labeled training set and evaluating on the validation sets. In particular, each model was trained for 20 epochs. Then, we measured the averaged bipartition accuracy on both validation sets and chose the most accurate model.

Having performed the above procedure for different models, we can compare their architectures and consequently find the most optimal one. This is demonstrated in Fig. 7, where the accuracy for the CNN network models is plotted against the number of parameters. One can notice that using deeper (consisting of more convolutional and FC layers) models tends to have better accuracy. However, as the complexity of the networks increases, by adding more convolutional and FC layers, we observe a saturation in prediction accuracy for the models containing approximately 10^6 parameters. This is why we decided to limit our architectures, presented in Fig. 2, to three convolutional and five FC layers only.

It is worth mentioning that the parameter count significantly exceeds the number of linearly independent real parameters needed to characterize a three-qubit system, i.e., $8^2 - 1 = 63$. This is due to a fact that modern NNs are usually *overparameterized*, that is, unlike traditional statistical and ML models where taking too complex models results in overfitting (the so-called bias-variance trade-off), in DL models the usage of more parameters even improves their properties [52]. It is then observed as a "double descent" curve that generalizes the standard "U-shaped" bias-variance curve by showing that increasing model capacity beyond some critical point results in improved performance. Increasing the size of the model can act as a kind of regularization. This suggests that parameter counting cannot indicate the true complexity of deep NNs.

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. **25**, 84 (2012).

[2] S. Ren, K. He, R. Girshick, and J. Sun, Faster r-CNN: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. **28**, 1137 (2015).

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016).

[4] F. Schroff, D. Kalenichenko, and J. Philbin, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Boston, MA, USA, 2015), p. 815.

[5] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[6] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 1 (2017).
[7] G. Carleo, Y. Nomura, and M. Imada, Constructing exact representations of quantum many-body systems with deep neural networks, Nat. Commun. **9**, 1 (2018).
[8] S. Tibaldi, G. Magnifico, D. Vodola, and E. Ercolessi, Unsupervised and supervised learning of interacting topological phases from single-particle correlation functions, SciPost Phys. **14**, 005 (2023).
[9] D.-L. Deng, X. Li, and S. Das Sarma, Quantum entanglement in neural network states, Phys. Rev. X **7**, 021021 (2017).
[10] Y.-C. Ma and M.-H. Yung, Transforming Bell's inequalities into state classifiers with machine learning, npj Quantum Inf. **4**, 1 (2018).
[11] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum entanglement in deep learning architectures, Phys. Rev. Lett. **122**, 065301 (2019).
[12] J. Carrasquilla and G. Torlai, How to use neural networks to investigate quantum many-body physics, PRX Quantum **2**, 040201 (2021).
[13] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2017), Chapter Quantum Entanglement.
[14] M. B. Plenio and S. S. Virmani, in *Quantum Information and Coherence*, edited by E. Andersson and P. Öhberg (Springer International Publishing, Cham, 2014), p. 173.
[15] P. Horodecki, Separability criterion and inseparable mixed states with positive partial transposition, Phys. Lett. A **232**, 333 (1997).
[16] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 808 (2020).
[17] Z. Yang and X. Zhang, Entanglement-based quantum deep learning, New J. Phys. **22**, 033041 (2020).
[18] L. Schatzki, A. Arrasmith, P. J. Coles, and M. Cerezo, Entangled datasets for quantum machine learning, ArXiv: 2109.03400.
[19] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).
[20] P.-H. Qiu, X.-G. Chen, and Y.-W. Shi, Detecting entanglement with deep quantum neural networks, IEEE Access **7**, 94310 (2019).
[21] L. Banchi, N. Pancotti, and S. Bose, Quantum gate learning in qubit networks: Toffoli gate without time-dependent control, npj Quantum Inf. **2**, 16019 (2016).
[22] Y. Liu, W.-J. Li, X. Zhang, M. Lewenstein, G. Su, and S.-J. Ran, Entanglement-based feature extraction by tensor network machine learning, Front. Appl. Math. Stat. **7**, 716044 (2021).
[23] Y. Chen, Y. Pan, G. Zhang, and S. Cheng, Detecting quantum entanglement with unsupervised learning, Quantum Sci. Technol. **7**, 015005 (2021).
[24] N. Asif, U. Khalid, A. Khan, T. Q. Duong, and H. Shin, Entanglement detection with artificial neural networks, Sci. Rep. **13**, 1562 (2023).
[25] J. Ureña, A. Sojo, J. Bermejo, and D. Manzano, Entanglement detection with classical deep neural networks, ArXiv:2304.05946.
[26] C. B. Goes, A. Canabarro, E. I. Duzzioni, and T. O. Maciel, Automated machine learning can classify bound entangled states with tomograms, Quantum Inf. Process. **20**, 1 (2021).
[27] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, Separability-entanglement classifier via machine learning, Phys. Rev. A **98**, 012315 (2018).
[28] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Volume of the set of separable states, Phys. Rev. A **58**, 883 (1998).
[29] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).
[30] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, Phys. Lett. A **223**, 1 (1996).
[31] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett. **80**, 2245 (1998).
[32] A. R. R. Carvalho, F. Mintert, and A. Buchleitner, Decoherence and multipartite entanglement, Phys. Rev. Lett. **93**, 230501 (2004).
[33] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki, Optimization of entanglement witnesses, Phys. Rev. A **62**, 052310 (2000).
[34] I. Bengtsson and K. Zyczkowski, Geometry of Quantum States: An Introduction to Quantum Entanglement (Cambridge University Press, 2006).
[35] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Distinguishing separable and entangled states, Phys. Rev. Lett. **88**, 187904 (2002).
[36] F. Hulpke and D. Bruß, A two-way algorithm for the entanglement problem, J. Phys. A: Math. Gen. **38**, 5573 (2005).
[37] X. Liang, S.-J. Dong, and L. He, Hybrid convolutional neural network and projected entangled pair states wave functions for quantum many-particle states, Phys. Rev. B **103**, 035138 (2021).
[38] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated J_1-J_2 model studied with neural network quantum states, Phys. Rev. B **100**, 125124 (2019).
[39] 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).
[40] C. Roth and A. H. MacDonald, Group convolutional neural networks improve quantum state accuracy, ArXiv:2104.05085.
[41] M. Krawczyk, J. Pawłowski, M. M. Maśka, and K. Roszak, Data-driven criteria for quantum correlations, Phys. Rev. A **109**, 022405 (2024).
[42] F. Mezzadri, How to generate random matrices from the classical compact groups, Not. Am. Math. Soc. **54**, 592 (2006).
[43] D. Cruz, R. Fournier, F. Gremion, A. Jeannerot, K. Komagata, T. Tosic, J. Thiesbrummel, C. L. Chan, N. Macris, M.-A. Dupertuis, and C. Javerzac-Galy, Efficient quantum algorithms for GHZ and W states, and implementation on

the IBM quantum computer, Adv. Quantum Technol. **2**, 1900015 (2019).

[44] J. Tura, R. Augusiak, P. Hyllus, M. Kuś, J. Samsonowicz, and M. Lewenstein, Four-qubit entangled symmetric states with positive partial transpositions, Phys. Rev. A **85**, 060302 (2012).

[45] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal, Unextendible product bases and bound entanglement, Phys. Rev. Lett. **82**, 5385 (1999).

[46] A. Acín, D. Bruß, M. Lewenstein, and A. Sanpera, Classification of mixed three-qubit states, Phys. Rev. Lett. **87**, 040401 (2001).

[47] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[48] C. Roberts, A. Milsted, M. Ganahl, A. Zalcman, B. Fontaine, Y. Zou, J. Hidary, G. Vidal, and S. Leichenauer, Tensor network: A library for physics and machine learning, ArXiv:1905.01330.

[49] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).

[50] H. Abraham, *et al.*, Qiskit: An open-source framework for quantum computing (2019).

[51] A. Paszke, *et al.*, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., Red Hook, New York, 2019), Vol. 32, p. 8024.

[52] M. Belkin, D. Hsu, S. Ma, and S. Mandal, Reconciling modern machine-learning practice and the classical bias–variance trade-off, Proc. Natl. Acad. Sci. **116**, 15849 (2019).

[53] github.com/Maticraft/quantum_correlations

PHYSICAL REVIEW A **112**, 022431 (2025)

Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests

Pavel Baláž,[1] Mateusz Krawczyk,[2] Jarosław Pawłowski,[2] and Katarzyna Roszak[1]

[1]*FZU—Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 00 Prague, Czech Republic*

[2]*Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*

(Received 10 May 2025; accepted 30 July 2025; published 25 August 2025)

We study the effectiveness of two distinct machine learning techniques, neural networks and random forests, in the quantification of entanglement from two-qubit tomography data. Although we predictably find that neural networks yield better accuracy, we also find that the way the two methods reach their prediction is starkly different. This is seen by the measurements which are the most important for the classification. Neural networks follow the intuitive prediction that measurements containing information about nonlocal coherences are most important for entanglement, but random forests signify the dominance of information contained in occupation measurements. This is because occupation measurements are necessary for the extraction of data about all other density matrix elements from the remaining measurements. The same discrepancy does not occur when the models are used to learn entanglement directly from the elements of the density matrix, so it is the result of the scattering of information and interdependence of measurement data. As a result, the models behave differently when noise is introduced to various measurements, which can be harnessed to obtain more reliable information about entanglement from noisy tomography data.

DOI: 10.1103/zgfy-yyw8

I. INTRODUCTION

Pure state entanglement is in principle straightforward to determine, since there is a direct correspondence between the amount of entanglement between subsystems and the purity of the density matrix of either subsystem. This is obvious, since a measure of entanglement, analogous to entanglement entropy but based on linear entropy instead of von Neumann entropy, is a good and reliable entanglement measure. Experimental determination of purity cannot be performed directly and quantum state tomography is required, but it is enough to measure the state of the smaller of the entangled subsystems to describe pure state entanglement.

For mixed states, even theoretical determination of entanglement is challenging and its experimental study requires quantum state tomography (QST) of the whole density matrix [1] followed by the calculation of one of the more numerically accessible entanglement measures. Since quantum state tomography requires $\mathcal{N}^2 - 1$ measurements, where $\mathcal{N}$ is the dimension of the whole system, the number of required measurements grows rapidly with system size, i.e., $\mathcal{N} = 2^{\#\,\text{qubits}}$, and becomes unmanageable for pretty small systems. For larger systems quantum state reconstruction can be effectively (with learned positivity constraint) performed using deep neural networks (NNs) [2–9], which can converge much faster [7] than standard QST, even for noisy measurement data [9].

The number of measurements required for full state tomography of the smallest possible system that can be entangled, namely, a two-qubit system, is already non-negligible. Although it is standard in optics, in many systems it is not possible to perform the required $4^2 - 1 = 15$ two-qubit measurements, especially since very good control of the measurement basis is required to obtain all of the information about the two-qubit density matrix. Thus a situation where not all of the measurements can be performed and estimation of entanglement must be performed from incomplete or noisy data is very natural.

Machine learning (ML) has proven useful in recognizing entanglement with (classical) deep NNs [10–15] or non-neural models [16], e.g., using ensemble learning [17–19]. Ensemble learning is a powerful paradigm in ML in which multiple models are combined to enhance overall predictive performance. This technique capitalizes on the diversity of base learners, leading to a more generalized and reliable model. In scientific applications, ensemble learning has demonstrated significant benefits, from reducing overfitting to enhancing the stability of predictions, making it an indispensable tool in the modern ML toolkit [20]. The random forest (RF) model [21] is one of the best known examples of ensemble methods that construct multiple decision trees (DTs) [22,23] using the principle of bagging [24] and selecting random features.

In the past decade, the revolution in AI has been driven by the rapid development of deep NNs [25,26], which have demonstrated remarkable success in tasks such as image processing and natural language analysis [27]. Unlike DTs or dimensionality reduction techniques, NNs take a different approach to machine learning, with a structure inspired by biology but also physical models [28]. While they allow for the training of highly complex models far beyond traditional ML methods [29], their interpretability presents a challenge. Nevertheless, efforts are underway to develop techniques that make NNs more explainable [30], especially in science where understanding how a model works is as important as the quality of its predictions.

In this work we study the possibility of obtaining reliable information about two-qubit entanglement from quantum

 ©2025 American Physical Society

tomography measurements on the basis of incomplete or perturbed experimental data. To this end, we employ two distinct ML techniques, the NN model and the RF model. We predictably find that the NN model is better at evaluating entanglement from the full set of measurement data. We further evaluate the importance of the different measurments both directly and by introducing noise for a chosen measurement. Here we consistently find that the NN model behaves intuitively, meaning that measurements containing information about nonlocal coherences are the most relevant, followed by those with information about local coherences, and occupations are least relevant. This behavior is in agreement with feature importance displayed when concurrence is calculated from tomography data, without ML (as quantified by the Shapley values).

The RF model displays completely opposite behavior. It signifies that measurements that quantify occupations are the most important, while those that depend on the nonlocal coherences are least important. This discrepancy occurs because the data about the system state are provided as a set of measurements, in which the information is dispersed. A similar calculation performed using density matrix elements directly showed no discrepancy and the feature importance followed the intuitive order of importance for entanglement: nonlocal coherences, local coherences, and occupations.

This means that under circumstances where some data may be unreliable, predictions made by substantially different ML techniques are likely to vary and the better choice of method depends on the type of problem under study. Neural network models act in a more direct way and for the specific problem of entanglement quantification perform better for a full set of tomography data. The inner working of RF models is more intricate and they are much more dependent on the interdependence of the measurement data, thus shifting importance onto occupations. This is an advantage, since occupations are typically easier to measure than coherences.

The paper is organized as follows. In Secs. II and III we provide the definition and discussion of the concurrence and introduce basic notions about two-qubit quantum state tomography, respectively. In Sec. IV we discuss the datasets that are used in the study. Section V introduces the two types of machine learning models. Section VI contains a description of the results and their discussion. Section VII summarizes the paper.

II. CONCURRENCE

The classification of two-qubit entanglement is reasonably straightforward even for mixed states, since there exists a formula for the (numerical) calculation of entanglement of formation (EOF) [31] directly from the density matrix (no minimization is required and bound states [32,33] do not exist). In the following we use the concurrence as the measure of entanglement. The concurrence is unambigously related to EOF [31], while it is a bit easier to use.

The concurrence is a unique measure that exists only for two-qubit systems (contrarily to EOF, which can be defined for a system of any size). It is defined as

$$C(\rho) = \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4), \tag{1}$$

where $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \lambda_4$ are the eigenvalues of the matrix

$$R = \sqrt{\sqrt{\rho}(\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)\sqrt{\rho}}, \tag{2}$$

in decreasing order. Here σ_y denotes the appropriate Pauli matrix and ρ^* is the complex conjugate of the two-qubit density matrix. Concurrence is an entanglement monotone, which ranges between 0 and 1. A concurrence value $C(\rho) = 0$ signifies that the state ρ is separable (there is no entanglement between the qubits), while $C(\rho) = 1$ indicates a maximally entangled state.

Let us now look at a generic two-qubit density matrix written in the standard separable basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$,

$$\begin{pmatrix} \rho_{00,00} & \rho_{00,01} & \rho_{00,10} & \rho_{00,11} \\ \rho^*_{00,01} & \rho_{01,01} & \rho_{01,10} & \rho_{01,11} \\ \rho^*_{00,10} & \rho^*_{01,10} & \rho_{10,10} & \rho_{10,11} \\ \rho^*_{00,11} & \rho^*_{01,11} & \rho^*_{10,11} & \rho_{11,11} \end{pmatrix}, \tag{3}$$

with $\rho_{11,11} = 1 - \rho_{00,00} - \rho_{01,01} - \rho_{10,10}$. We color coded the elements of the density matrix using colored boxes, marking the diagonal elements (occupations) red, the nonlocal coherences blue, and leaving the other off-diagonal elements black, for clarity in the subsequent discussion. The importance of the different elements in this basis for entanglement is not uniform. For instance, a state that has nonzero elements only on the diagonal is a statistical mixture and can contain only classical correlations. Nonzero nonlocal coherences are, on the other hand, a necessary condition for two-qubit entanglement, since they contain information about the phase relations between the qubits. This is immediately evident when studying X states [34,35], but is true for any two-qubit state. The interplay of the nonlocal coherences with each other and with other off-diagonal elements plays a significant role for both the existence of entanglement and its quantity. The magnitude of the different off-diagonal elements with respect to diagonal elements contains information about the purity of the state and as such is also critical for entanglement.

III. TWO-QUBIT QUANTUM STATE TOMOGRAPHY

In the following we use the projective-measurement tomography scenario as in the original proposal of Ref. [1]. To determine the state of a single qubit via quantum state tomography, only three distinct measurements are required, which are typically chosen to correspond to the three Pauli operators. For two qubits, the minimum required number of measurements is 15, but in the following we always include the superfluous 16th measurement for the sake of symmetry. In general, the outcome of a projective measurement can be written as [1] $m = \mathcal{N}\,\mathrm{Tr}(\hat{\rho}\hat{\mu})$, where $\hat{\rho}$ is the density matrix, $\hat{\mu}$ is the projection operator, and $\mathcal{N}$ is a constant of proportionality which can be determined from the data. For two qubits, the outcomes of joint measurements required for tomography can be expressed as

$$m_{ij} = \mathcal{N}\,\mathrm{Tr}[\hat{\rho}(\hat{\mu}_i \otimes \hat{\mu}_j)], \tag{4}$$

where $\hat{\mu}_i$ and $\hat{\mu}_j$ $(i, j = 0, 1, 2, 3)$ are projectors on a given single-qubit state. Note that this formulation limits the measurements to measurements in separable bases.

The choice of measurements required to obtain full information about a quantum state is not unique. In the following, we use the set of measurements proposed in Ref. [1], where the single-qubit measurements $\hat{\mu}_i = |i\rangle\langle i|$ correspond to projections on states

$$|0\rangle, |1\rangle, |2\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |3\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle - i|1\rangle). \tag{5}$$

Below we provide expectation values for each measurement, expressed with the help of the elements of the density matrix (3) and retaining the color coding in order to facilitate the discussion on how the information about the density matrix is distributed within the measurements. Throughout the paper we group the measurements into four blocks as follows:

$$\begin{aligned} \boldsymbol{M}_{\mathrm{A}} &= \{m_{00}, m_{01}, m_{10}, m_{11}\}, \\ \boldsymbol{M}_{\mathrm{B}} &= \{m_{02}, m_{03}, m_{12}, m_{13}\}, \\ \boldsymbol{M}_{\mathrm{C}} &= \{m_{20}, m_{21}, m_{30}, m_{31}\}, \\ \boldsymbol{M}_{\mathrm{D}} &= \{m_{22}, m_{23}, m_{32}, m_{33}\}. \end{aligned} \tag{6}$$

The blocks contain information about different types of density matrix elements, as is evident below.

Block $\boldsymbol{M}_{\mathrm{A}}$ contains all measurement outcomes that can be obtained in the $|ij\rangle$ $(i, j = 0, 1)$ two-qubit basis and thus the measurements only yield information about the diagonal elements of the matrix (3),

$$m_{0,0} = \boxed{\rho_{00,00}}, \tag{7a}$$

$$m_{0,1} = \boxed{\rho_{01,01}}, \tag{7b}$$

$$m_{1,0} = \boxed{\rho_{10,10}}, \tag{7c}$$

$$m_{1,1} = \boxed{\rho_{11,11}}. \tag{7d}$$

Blocks $\boldsymbol{M}_{\mathrm{B}}$ and $\boldsymbol{M}_{\mathrm{C}}$ contain measurement outcomes where one of the qubits is measured in the $\{|0\rangle, |1\rangle\}$, basis while the other is measured with respect to one of the other two states used in the tomography [see Eqs. (5)]. Thus the dependence of outcomes on the elements of the density matrix from block $\boldsymbol{M}_{\mathrm{B}}$ can be obtained by exchanging qubit indices in the outcomes from block $\boldsymbol{M}_{\mathrm{C}}$ and vice versa. Note that this symmetry does not translate into any of the measurements being superfluous, since a density matrix does not have to contain any symmetry with respect to the exchange of qubits. The explicit formulas for blocks $\boldsymbol{M}_{\mathrm{B}}$ and $\boldsymbol{M}_{\mathrm{C}}$ are given by

$$m_{0,2} = \tfrac{1}{2}\big[\boxed{\rho_{00,00}} + \boxed{\rho_{01,01}} + 2\,\mathrm{Re}(\rho_{00,01})\big], \tag{8a}$$

$$m_{0,3} = \tfrac{1}{2}\big[\boxed{\rho_{00,00}} + \boxed{\rho_{01,01}} + 2\,\mathrm{Im}(\rho_{00,01})\big], \tag{8b}$$

$$m_{1,2} = \tfrac{1}{2}\big[\boxed{\rho_{10,10}} + \boxed{\rho_{11,11}} + 2\,\mathrm{Re}(\rho_{10,11})\big], \tag{8c}$$

$$m_{1,3} = \tfrac{1}{2}\big[\boxed{\rho_{10,10}} + \boxed{\rho_{11,11}} + 2\,\mathrm{Im}(\rho_{10,11})\big] \tag{8d}$$

and

$$m_{2,0} = \tfrac{1}{2}\big[\boxed{\rho_{00,00}} + \boxed{\rho_{10,10}} + 2\,\mathrm{Re}(\rho_{00,10})\big], \tag{9a}$$

$$m_{2,1} = \tfrac{1}{2}\big[\boxed{\rho_{01,01}} + \boxed{\rho_{11,11}} + 2\,\mathrm{Re}(\rho_{01,11})\big], \tag{9b}$$

$$m_{3,0} = \tfrac{1}{2}\big[\boxed{\rho_{00,00}} + \boxed{\rho_{10,10}} + 2\,\mathrm{Im}(\rho_{00,10})\big], \tag{9c}$$

$$m_{3,1} = \tfrac{1}{2}\big[\boxed{\rho_{01,01}} + \boxed{\rho_{11,11}} + 2\,\mathrm{Im}(\rho_{01,11})\big], \tag{9d}$$

respectively. These measurements contain information about the local coherences, but not the nonlocal coherences. This means that the information contained in the two blocks is insufficient to distinguish between an entangled state and a separable state.

The last block $\boldsymbol{M}_{\mathrm{D}}$ is the block that contains information about the critical nonlocal coherences,

$$\begin{aligned} m_{2,2} = \tfrac{1}{4}\big[2\,\mathrm{Re}\big(&\rho_{00,01} + \rho_{00,10} + \boxed{\rho_{00,11}} + \boxed{\rho_{01,10}} \\ &+ \rho_{01,11} + \rho_{10,11}\big) + 1\big], \end{aligned} \tag{10a}$$

$$\begin{aligned} m_{2,3} = \tfrac{1}{4}\big[2\,\mathrm{Im}\big(&\rho_{00,01} + \boxed{\rho_{00,11}} - \boxed{\rho_{01,10}} + \rho_{10,11}\big) \\ &+ 2\mathrm{Re}(\rho_{00,10} + \rho_{01,11}) + 1\big], \end{aligned} \tag{10b}$$

$$\begin{aligned} m_{3,2} = \tfrac{1}{4}\big[2\,\mathrm{Im}\big(&\rho_{00,10} + \boxed{\rho_{00,11}} + \boxed{\rho_{01,10}} + \rho_{01,11}\big) \\ &+ 2\mathrm{Re}(\rho_{00,01} + \rho_{10,11}) + 1\big], \end{aligned} \tag{10c}$$

$$\begin{aligned} m_{3,3} = \tfrac{1}{4}\big[&2\,\mathrm{Im}(\rho_{00,01} + \rho_{00,10} + \rho_{01,11} + \rho_{10,11}) \\ &+ 2\mathrm{Re}\big(\boxed{\rho_{01,10}} - \boxed{\rho_{00,11}}\big) + 1\big]. \end{aligned} \tag{10d}$$

Note that the measurement outcomes also depend on other off-diagonal elements; thus it is impossible to determine the magnitude of the nonlocal coherences without the other measurements.

IV. DATASET

We generate a set of two-qubit density matrices by using different random-sampling methods. The main idea is to generate random states using the quantum circuit approach [11], but in order to have a well-diversified ensemble, we also use a technique based on sampling from the uniform Haar measure [36]. Additionally, to increase the number of maximally entangled states, we specifically include 20 000 of them. In total, we generate 460 000 density matrices for the training dataset and 46 000 matrices for the test set. All density matrices are labeled with the concurrence as defined by Eq. (1). We then transform the density matrices into feature vectors that contain the outcomes of the two-qubit measurements [Eq. (4)]. Feature vectors are composed of measurements together with respective labels C representing the concurrence values from the training dataset $\boldsymbol{S}_{\mathrm{train}}$. During the generation process, we specifically restrict our sampling methods to generate a dataset balanced in the number of separable and entangled states. Due to this, the number of samples generated by various methods is adjusted to obtain a similar number of states with $C(\rho) < \tau$ and $C(\rho) >= \tau$, with $\tau = 10^{-6}$, which is the chosen threshold value between separable and entangled states. More details regarding generation methods with a similar approach can be found in Ref. [11]. In the end, our training dataset $\boldsymbol{S}_{\mathrm{train}}$ consists of about 53% entangled states and 47% separable states.

We show the distribution (histogram) of the concurrence values C across the training set $\boldsymbol{S}_{\mathrm{train}}$ in Fig. 1. The bars show the proportion of samples with C within a given range. The main figure shows the whole range of C values in a logarithmic scale. We observe two regions corresponding to separable and entangled states. In the inset of Fig. 1 we see the distribution of concurrence of the entangled states only (with

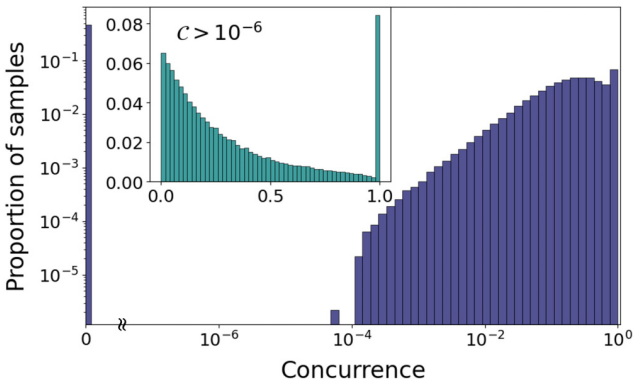


FIG. 1. Proportion of samples in the training dataset with the concurrence within a given range. The main figure shows data for all samples. In the inset we show just entangled samples, i.e., with the concurrence $C > 10^{-6}$.

$C > 10^{-6}$). The number of samples decreases with increasing concurrence; however, there is a peak for samples with $C(\rho) \simeq 1$, i.e., maximally entangled, reflecting the presence of artificially included randomized Bell states. The test set $\boldsymbol{S}_{\text{test}}$ follows the same distribution.

V. MODELS

We examine two predominant types of machine learning models utilized in data classification and regression, i.e., an ensemble method (RF) and a deep NN [multilayer perceptron (MLP)], and elucidate their respective advantages and disadvantages in the context of quantum state recognition.

A. Ensemble methods

Ensemble models are primarily divided into two categories: bagging [24] and boosting [37,38]. One of the most popular bagging ensembles is RFs [21]. Random forests utilize an ensemble of DTs (as base learners), which is a supervised learning technique that generates predictions by recursively partitioning data into subsets according to feature values [23]. In this study, we utilize the RF models for the classification of measurement vectors into two categories: separable (negative) or entangled (positive). Further details on the training of RFs and DTs, as well as their hyperparameters, can be found in Appendix B.

B. Deep neural networks

The second type of ML model employed is the NN MLP model. Multilayer perceptrons are supervised models that have been successfully applied to both classification and regression tasks. In this work we employ an MLP to accurately predict the value of concurrence C using measurement values as inputs, as presented in Fig. 2(a). Our MLP architecture comprises just two fully connected hidden layers, each containing 128 units (neurons) and utilizing a rectified linear unit activation function [39]. Finally, it has a single output unit with a linear activation function that predicts the C value. We used a linear output activation, although a sigmoid function may seem more natural. However, combining the mean-square error (MSE) as the loss function with a sigmoid activation often leads to optimization issues, as the resulting cost surface becomes nonconvex. To ensure that the predictions for C remain within its natural range, we apply additional clipping. We limit the model to only two hidden layers, which proves sufficient for making two-qubit predictions. However, we also train a much larger model consisting of ten hidden layers, but it achieves comparable results.

The MLP is trained using the training dataset $\boldsymbol{S}_{\text{train}}$ (of size N) using a standard MSE loss function,

$$\mathcal{L}(\boldsymbol{S}_{\text{train}}) \equiv \text{MSE}(\boldsymbol{C}, \hat{\boldsymbol{C}}) = E_j[(C_j - \hat{C}_j)^2], \tag{11}$$

where C_j is the ground truth value for concurrence of a given state j, while $\hat{C}_j$ is the the concurrence estimated by the NN model. Moreover, $E_j[X_j] = (1/N)\sum_{j=1}^{N} X_j$ denotes a standard average over an index j.

C. Measurements perturbation in QST

To improve the understanding of the measurements' impact on concurrence prediction, we also analyze subsequent measurements' impact on QST by introducing perturbations to the standard tomography. Figure 2(b) shows how measurement perturbations propagate through quantum tomography on a

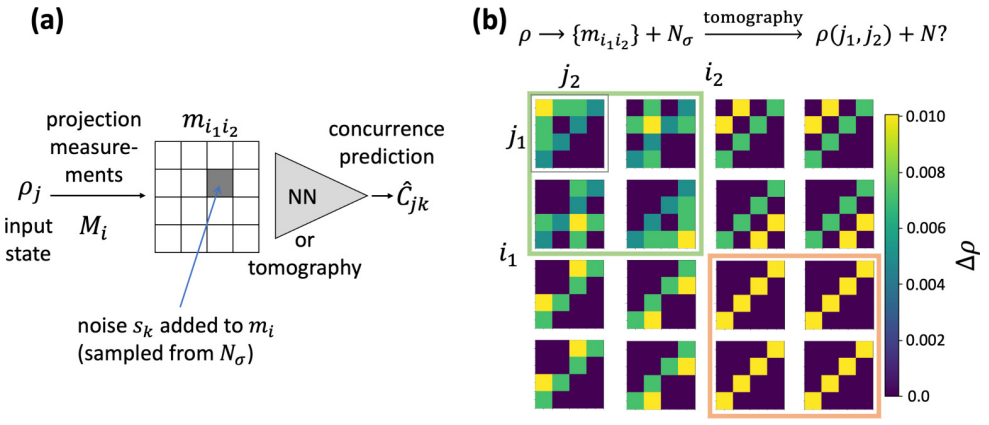


FIG. 2. Measurement perturbations. (a) Methodology. (b) Perturbation $\Delta\rho$ propagation within the quantum tomography. Two distinct blocks of measurements are marked by green ($\boldsymbol{M}_{\text{A}}$, upper left) and orange ($\boldsymbol{M}_{\text{D}}$, lower right) boxes.

TABLE I. Evaluation scores in entanglement classification for the RF classifiers and NN regressor on the $\boldsymbol{S}_{\text{test}}$ dataset.

Model	Accuracy	Precision	Recall
RF1	0.87930	0.87706	0.90026
RF2	0.86854	0.88570	0.86562
RF3	0.85837	0.86027	0.87739
MLP	0.92212	0.92794	0.92608

reconstructed state ρ. There we present average errors in the reconstructed state $\rho(j_1, j_2)$, with density matrix elements numbered by j_1 and j_2, upon the addition of a Gaussian noise N_σ, with $\sigma = 0.01$, to subsequent measurements m_{i_1,i_2}. In this way we obtain information on how given measurements impact ρ element reconstruction. We observe that various measurements play different roles during the reconstruction. For example, the $\boldsymbol{M}_{\text{D}}$ block (orange box) explains only the coherences of the density matrix ρ. On the other hand, block $\boldsymbol{M}_{\text{A}}$ (green box) has an impact on all the elements of ρ, which is the consequence of the fact that the diagonal elements which are measured by this block have to be used to infer the off-diagonal elements from measurement blocks $\boldsymbol{M}_{\text{B}}$ and $\boldsymbol{M}_{\text{C}}$. These in turn are critical in determining the interqubit coherences. Block $\boldsymbol{M}_{\text{D}}$ measurements do not impact any other elements of the density matrix besides the nonlocal coherences.

VI. RESULTS

In this section we conduct an evaluation of our models utilizing the test dataset $\boldsymbol{S}_{\text{test}}$, comprising 46 000 samples that were not incorporated during the training phase. Subsequently, we apply interpretability methods to elucidate the most salient aspects of the models' predictions.

A. Predictive power of the ML models

We start with ensemble models (RFs listed in Table II in Appendix B). For these models, we compute the accuracy, which is defined as

$$\text{ACC}_{\text{RF}}(\boldsymbol{S}_{\text{test}}) = E_j[1 - |P_j - \hat{P}_j|], \tag{12}$$

where P_j is the ground truth value for the predicted sample class (binary: separable or entangled), while $\hat{P}_j$ is the model's prediction. In the scenario where $\hat{P}_j = 0$ holds, the model infers that the observations pertain to two separable qubits, thereby classifying them as the negative class. Conversely, $\hat{P}_j = 1$ indicates a prediction of an entangled qubit pair, thus assigning it to the positive class.

Let us analyze first the RF1 model. The accuracy of this model evaluated on $\boldsymbol{S}_{\text{test}}$ is of order of 0.88 (see Table I). This value describes the overall performance of the model regardless of the concurrence value of the input samples. To gain a deeper understanding of its error characteristics, we present Fig. 3, which illustrates the relationship between the accuracy of the RF1 model [Eq. (12)] and the concurrence of the input sample. The concurrence value depicted on the x axis of Fig. 3 represents the ground truth value, calculated in accordance with Eq. (1).

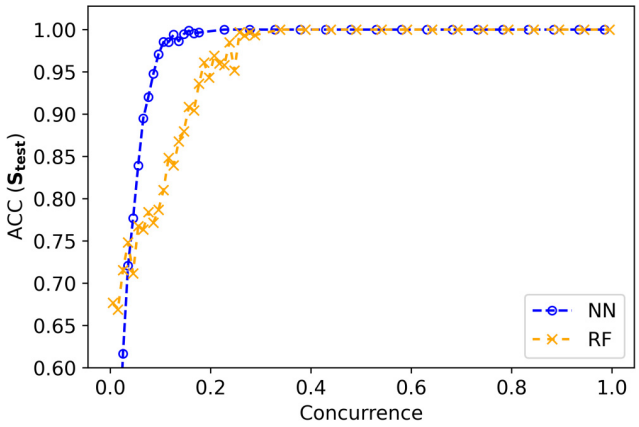


FIG. 3. RF1 and NN model accuracy for the input states (from the test set $\boldsymbol{S}_{\text{test}}$) with different concurrence values.

In Fig. 3 it is evident that the RF1 model accuracy is at its lowest when the C value approaches 0, indicating minimal entanglement among the qubits. Consequently, the model is more likely to misclassify such inputs as separable cases. As the C value rises, the model's accuracy correspondingly improves. It achieves an accuracy of 100% when the concurrence of the input samples reaches $C \gtrsim 0.3$. We obtain similar findings for the RF2 and RF3 models. This observation is in accord with the t-distributed stochastic neighbor embedding analysis of the data, illustrated in Fig. 8 in Appendix A 2, which shows that entangled samples with elevated concurrence values are more prone to being well separated and therefore easily distinguishable from separable qubits. On the other hand, if we take into account just separable (negative) samples from the $\boldsymbol{S}_{\text{test}}$, i.e., those with $C(\rho) < \tau$, the accuracy of their correct classification is of order of 0.86. The misclassified negative samples might be confused with entangled samples of low concurrence value.

For the NN model, the prediction error metrics are defined as

$$\text{RMSE}(\boldsymbol{S}_{\text{test}}) = \text{MSE}(\boldsymbol{S}_{\text{test}})^{1/2}. \tag{13}$$

In order to compare the NN model with the RF models, we define also accuracy for the NN regressor as

$$\text{ACC}_{\text{NN}}(\boldsymbol{S}_{\text{test}}) = E_j[1 - |\Theta(\tau - C_j) - \Theta(\tau_{\text{NN}} - \hat{C}_j)|], \tag{14}$$

where we explicitly count the number of correct classifications (instances where NN predictions match the data classes). The Heaviside function $\Theta(\cdot)$ compares the output (or true) C value with the assumed threshold. The NN classification threshold $\tau = 10^{-6}$ splits the predictions into separable (with $C_i \leqslant \tau$) and entangled ($C_i > \tau$) classes. The threshold $\tau_{\text{NN}} = 0.03$ is chosen to maximize the NN model accuracy by maximizing the area under the so-called precision-recall curve [40]. The resulting NN predictor performance, presented in Fig. 3, is significantly better than RF over different C values.

The overall comparison of the models is shown in Table I. For the sake of completeness, we also define the models' precision, which measures the proportion of correctly identified entangled instances (true positives) out of all instances that

were predicted to be entangled

$$\mathrm{PREC}(\boldsymbol{S}_{\mathrm{test}}) = E_j[P_j\hat{P}_j]/E_j[\hat{P}_j], \tag{15}$$

as well as express the proportion of actual entangled instances that were correctly identified by the classifier

$$\mathrm{REC}(\boldsymbol{S}_{\mathrm{test}}) = E_j[P_j\hat{P}_j]/E_j[P_j]. \tag{16}$$

Notably, the NN model demonstrates superior performance compared to all of the RF models across the three metrics evaluated. This result is expected, as deep NNs have proven to be more flexible and effective in entanglement recognition [10–15] than classical non-neural ML algorithms [16,18]. In the following sections we explore the underlying reasons. Among the RF classifiers, RF1, which comprises DTs with minimal regularization, appears as the most accurate one. Additional regularization of the DT sizes (present in RF2 and RF3; cf. Table II in Appendix B) leads to a decrease in the overall accuracy of the ensemble.

B. Measurement importance

We now direct our attention to the interpretability of the applied models. Here we assess the significance of the measurements that constitute our dataset. We employ a variety of methods to assess feature importance, enabling a comparative analysis of data handling between ensemble and NN models. Initially, we examine the mean decrease in impurity (MDI) measure, defined in Appendix B, that is uniquely applicable to models derived from DTs. Consequently, it is not suitable for comparing ensemble models with NNs. Hence, we proceed to assess the significance of measurements through their perturbations, facilitating a direct comparison between the two model types employed. Ultimately, we implement a more advanced approach predicated on Shapley values (SVs).

1. Mean decrease in impurity

One of the key advantages of RF models is their ability to assess feature importance, providing insights into the contributions of different variables in predicting the target outcome. In an RF, feature importance is typically determined by measuring the impact of each feature on the model's predictive accuracy, e.g., using MDI [21] (see Appendix for details). Figure 4(a) shows MDI values obtained for the RF1 model, which has been trained utilizing the dataset $\boldsymbol{S}_{\mathrm{train}}$. Within the heat map, each location, specified by indices i and j ($i, j = 0, 1, 2, 3$), corresponds to the m_{ij} measurement, defined in Eq. (4). The data demonstrate that individual features exhibit varying levels of importance (measurements that are most crucial for sample classification possess the highest MDI values). Quite surprisingly, these measurements are situated in block $\boldsymbol{M}_{\mathrm{A}}$, which only contains information about the diagonal elements of the density matrix in a separable basis. Conversely, the lowest MDI values are found in block $\boldsymbol{M}_{\mathrm{D}}$, which contains information about the off-diagonal elements which are critical for entanglement (nonlocal coherences). Blocks $\boldsymbol{M}_{\mathrm{B}}$ and $\boldsymbol{M}_{\mathrm{C}}$ encompass measurements of moderate significance. It is relevant to note that the differences between the feature significance within the different blocks are moderate and all features are essential to achieve accurate classification.

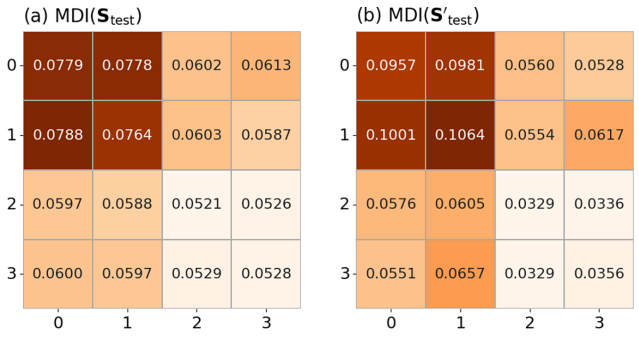


FIG. 4. Mean decrease in impurity of the RF1 model trained and tested using (a) the samples with arbitrary values of concurrence and (b) samples with concurrence $C < \tau$ (separable) and $C > 0.99$ (strongly entangled).

The seemingly disproportionate importance of blocks that on their own do not contain information about entanglement (blocks $\boldsymbol{M}_{\mathrm{A}}$, $\boldsymbol{M}_{\mathrm{B}}$, and $\boldsymbol{M}_{\mathrm{C}}$) is due to two factors. First, it is impossible to determine the nonlocal coherences from the measurements of block $\boldsymbol{M}_{\mathrm{D}}$ without the knowledge of the other off-diagonal elements. This in turn can be found from the measurements of blocks $\boldsymbol{M}_{\mathrm{B}}$ and $\boldsymbol{M}_{\mathrm{C}}$ only when the occupations are known, and occupations are directly measured via block $\boldsymbol{M}_{\mathrm{A}}$. Second, nonzero nonlocal coherences are only necessary for entanglement, while the actual entanglement strongly depends on the interplay between all elements of the density matrix.

To determine if the importance of the different blocks is not a numerical peculiarity, we provide a second set of data where the RF1 model is trained on a subset of the training dataset $\boldsymbol{S}'_{\mathrm{train}} \subset \boldsymbol{S}_{\mathrm{train}}$, which includes only states that are either separable ($C < \tau$) or strongly entangled ($C > 0.99$). The model is tested on an analogous subset of the test dataset $\boldsymbol{S}'_{\mathrm{test}} \subset \boldsymbol{S}_{\mathrm{test}}$. The RF1 model can distinguish between separable and strongly entangled qubit pairs with 100% accuracy. Figure 4(b) shows the MDI values for this situation. We note that the relative importance of the different blocks remains unchanged, but the features became significantly more prominent, so while the importance of blocks $\boldsymbol{M}_{\mathrm{B}}$ and $\boldsymbol{M}_{\mathrm{C}}$ remains similar, the importance of measurements in the $\boldsymbol{M}_{\mathrm{A}}$ block increase, while measurements in $\boldsymbol{M}_{\mathrm{D}}$ have their significance further reduced. This is highly surprising, but it does demonstrate that the hierarchy of feature importance is retained even if the analysis is performed on datasets which are chosen to amplify the differences between separable and entangled states and is thus not a numerical peculiarity.

2. Measurement importance via perturbations

Since RF and NN models are fundamentally different approaches, our main goal is to determine how the NN predictor works compared to ensemble RF models. To this end, we evaluate the importance of features of both models in the same way. The employed methodology for testing the importance of features (measurements) is presented in Fig. 2(a). For subsequent states from the test set $\boldsymbol{S}_{\mathrm{test}}$, we perturb a given measurement by adding uniform noise N_σ within the interval $[-\sigma, \sigma]$ and then study how the prediction of the concurrence

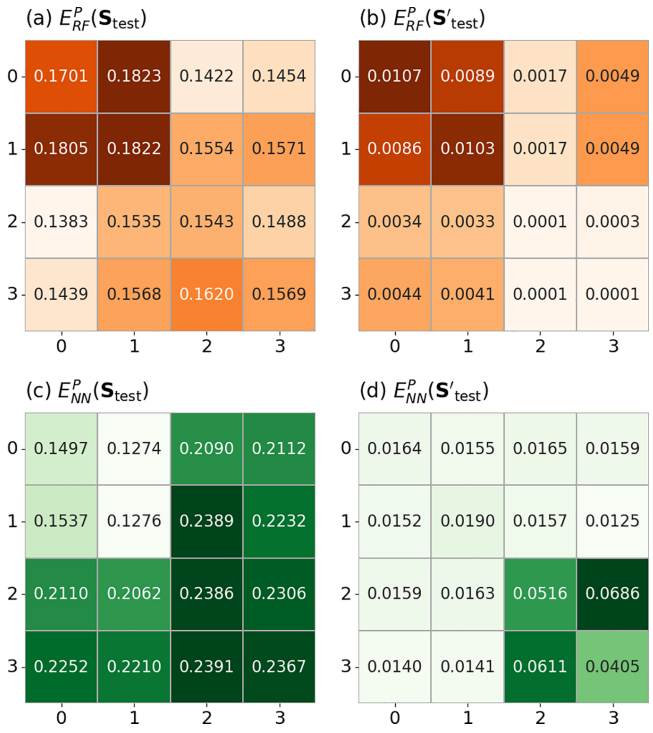


FIG. 5. Perturbation measure E^p for random noise applied to a single measurement outcome with $\sigma = 0.05$ for (a) the RF1 model trained using $\boldsymbol{S}_{\text{train}}$, (b) the RF1 model trained using $\boldsymbol{S}'_{\text{train}}$, i.e., samples with concurrence $\mathcal{C} < 10^{-6}$ and $\mathcal{C} > 0.99$, and the NN model trained using (c) $\boldsymbol{S}_{\text{train}}$ and (d) $\boldsymbol{S}'_{\text{train}}$.

C is perturbed. This way we describe the measurement importance in concurrence prediction. Figure 5 summarizes the prediction fidelity when a single measurement characterized by indices $i, j = 0, 1, 2, 3$ is disturbed. The RF classifier used is given by

$$E^{\text{P}}_{\text{RF}}(\sigma, i) = |\text{ACC}_{\text{RF}}(\boldsymbol{S}_{\text{test}}) - \text{ACC}_{\text{RF}}[\tilde{\boldsymbol{S}}_{\text{test}}(\sigma, i)]|, \qquad (17)$$

where $\tilde{\boldsymbol{S}}_{\text{test}}$ is the test dataset modified by the perturbation in such a way that it only affects the measurement $m_{i,j}$. It is perturbed by a random value from the interval $[-\sigma, \sigma]$, as described above. Analogously, the NN classifier is defined as

$$E^{\text{P}}_{\text{NN}}(\sigma, i) = E_{jk}[|\hat{C}_{ijk} - C_j|], \qquad (18)$$

where C_j is the ground truth concurrence for the $\rho_j \in \boldsymbol{S}_{\text{test}}$ state. These measures estimate how important a given feature (measurement) is for the prediction of the exact value of the state concurrence.

The data show that the RF and NN models learn to predict their targets in very different way. The RF classifier shadows the counterintuitive behavior of the feature importance classified by MDI, so it is most vulnerable to perturbations of measurements in the $\boldsymbol{M}_{\text{A}}$ block, as seen in Fig. 5(a). There is a lesser difference between the effect of perturbations between the remaining three blocks. Results obtained on the datasets limited to separable and highly entangled states [Fig. 5(b)] recapture all of the MDI features, thus proving that feature importance classified by MDI and the effect of perturbations produce the same result.

Contrarily, the NN results for both the full dataset [Fig. 5(c)] and the restricted one [Fig. 5(d)] show fully intuitive behavior. The block containing information about the nonlocal coherences ($\boldsymbol{M}_{\text{D}}$) is the most important, closely followed by blocks with information about other off-diagonal elements ($\boldsymbol{M}_{\text{B}}$ and $\boldsymbol{M}_{\text{C}}$), while the block that only provides statistical information ($\boldsymbol{M}_{\text{A}}$) is comparatively irrelevant.

These results highlight that the two types of models work very differently in entanglement quantification for tomography. The NN model is more disturbed by the perturbations of the measurements which directly contain information about the more relevant elements of the density matrix. The RF model obtains its results in a more convoluted way and the features that in themselves contain no information about entanglement are most relevant through their interplay with the information contained in other measurements.

Incidentally, such a discrepancy in feature importance does not occur when the two techniques are used to quantify the concurrence directly from the elements of the density matrix. In such a case, both RF and NN models indicate that the nonlocal coherences are of highest importance, followed by the other off-diagonal elements and then by occupations. This means that the difference between the models stems from the way that information about the density matrix is inferred from the measurement outcomes.

In addition, in Appendix E we compare the average performance of RF and NN models under a single measurement perturbation. It is shown that the NN model dominates at lower noise levels. However, as the noise amplitude increases, the accuracy of the NN model decreases faster than that of the RF. As a result, the RF model outperforms the NN at higher noise level (see Fig. 10). Random forest models combine the predictions of many decision trees trained on bootstrap-resampled subsets of data. Because each tree sees a different slice of the input features, their averaged vote smooths out individual errors, making the overall model noticeably sturdier against noisy features and labels.

3. Shapley additive explanations

In the preceding section we showed that RFs and NNs make their predictions in a way which is sensitive to different measurements at the input. Thus, we compute the so-called Shapley values [41] to explain how the models learn and predict. Shapley values, originally developed in cooperative game theory, provide a fair way to distribute the total contribution of all players (or features) in a system by considering all possible coalitions. In the context of ML, SVs offer a way to quantify the contribution of each feature to the model's prediction [42]. Importantly, SVs form a model-agnostic method, meaning they can be applied to any ML model without requiring specific modifications. Detailed definitions of SVs and Shapley interaction values (SIVs) are described in Appendix C. To compute SVs and SIVs, we employ the Shapley additive explanations (shap) library [42], which provides a unified framework for interpreting ML models based on SVs.

Let us first analyze the SVs of the RF model. Since SVs provide an explanation for a single sample, they allow us to analyze how the RF model classifies samples with concurrence in a given range of values. Figures 6(a) and 6(b) show

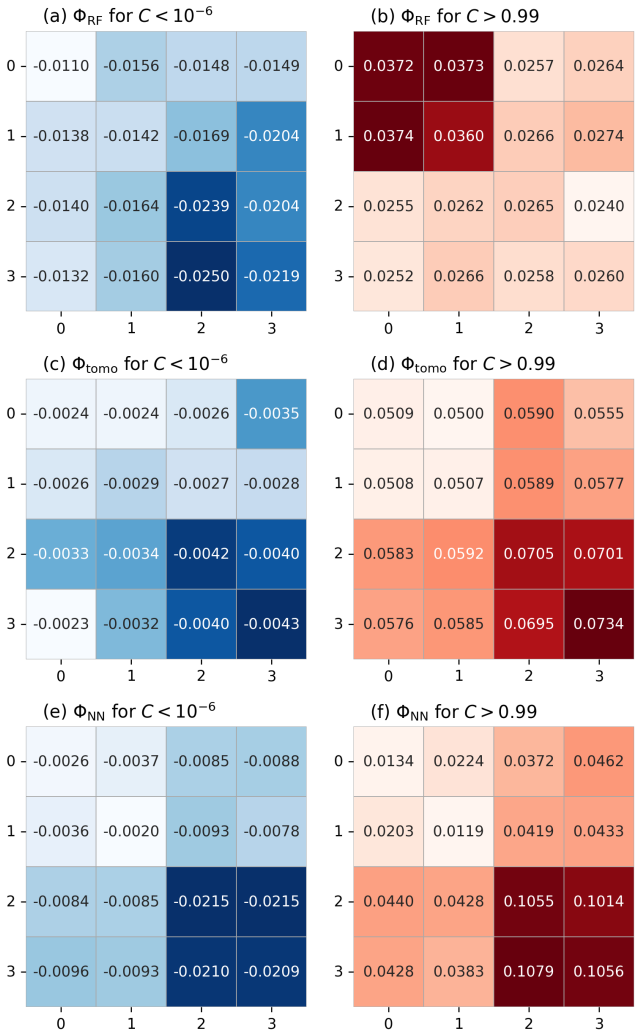


FIG. 6. Shapley values Φ_i calculated using samples with concurrence (a), (c), and (e) $C < \tau$ (separable) and (b), (d), and (f) $C > 0.99$ (strongly entangled) for (a) and (b) the RF1 model, (c) and (d) states reconstructed via quantum tomography, and (e) and (f) the NN predictor.

SVs averaged over a subset of samples from the testing dataset with concurrence below the threshold τ [Fig. 6(a)], i.e., for separable qubits, and with concurrence above 0.99 [Fig. 6(b)]. In our sign notation, the positive mean Shapley value means that the corresponding feature contributes to increasing the likelihood of class 1 (entangled). Oppositely, when a Shapley value is negative it drives the prediction towards class 0 (separable). Thus, it is obvious that the model treats the two groups in a very different manner. While the most important features contributing to correct classification of separable samples are those in segment $\boldsymbol{M}_{\mathrm{D}}$, the strongly entangled samples are classified based on the measurement values from set $\boldsymbol{M}_{\mathrm{A}}$.

In addition, the contribution of feature interactions, i.e., cases where features jointly influence the prediction in a way that goes beyond their individual effects, is studied in Appendix D using the SIVs. In the case of separable states, it is shown that diagonal SIVs push the model's prediction towards the opposite class (entagled). The contribution leading to the correct prediction originates from the feature interactions. Among them, the interactions of the measurements (m_{00},m_{11}) and (m_{01},m_{10}) are dominant. Conversely, if we take into account just the strongly entangled states, the model prediction is based mainly on the single measurement values.

Before we start the analysis of SVs for the NN model, let us look at Figs. 6(c) and 6(d), which show SVs for the concurrence C calculated from QST, i.e., where the concurrence is calculated directly for a state reconstructed from measurements via quantum tomography analytically. Here all the measurements are of similar importance, with the $\boldsymbol{M}_{\mathrm{A}}$ block slightly less relevant than the others, which is intuitive from the point of view of quantum information theory.

Shapley values of the NN model presented in Figs. 6(e) and 6(f) follow the same behavior as the QST-reconstructed results [cf. Figs. 6(c) and 6(d)]. The $\boldsymbol{M}_{\mathrm{D}}$ block has the most impact on C prediction. Interestingly, $\boldsymbol{M}_{\mathrm{D}}$ increases C predictions for separable states [see Fig. 6(e)] but decreases for entangled ones [Fig. 6(f)]. The $\boldsymbol{M}_{\mathrm{B}}$ and $\boldsymbol{M}_{\mathrm{C}}$ blocks are slightly weaker but still more important than block $\boldsymbol{M}_{\mathrm{A}}$, which only contains information about the occupation of separable states. Note that a similar structure of measurement importance for NN was also observed with the perturbation measure E^P, as presented in Fig. 5(c). We arrive at the intriguing result that, depending on the type of ML model (RF vs NN), different measurement groups are relevant for the C prediction, which strongly suggests that the two types of machine learning models operate very differently when trying to evaluate the amount of entanglement present in a given state.

VII. CONCLUSION

We have studied the performance of NN and RF models in the quantification of entanglement for two qubits from quantum tomography data. We found that NN models, which are known to be more flexible and efficient in entanglement prediction from quantum states, perform better also when tomography measurement outcomes are given on input instead of density matrix elements.

More surprisingly, we found that the importance of different measures differs greatly between the two classes of ML algorithms. Neural network feature importance follows the intuitive behavior in which the most relevant measurements are those that carry information about the qubit coherences which appear, e.g., in Bell or X states, followed by measurements that capture other coherences, and finally the ones responsible for diagonal elements of the density matrix (written in a separable basis). The same feature importance is exhibited when the concurrence is calculated directly from quantum state reconstruction, i.e., without ML, which we assessed using Shapley values. The RF feature importance is fully contradictory: The most important measurements are the ones which directly yield the diagonal elements and the ones that carry information about interqubit coherence were found to be least important. This points to a vast difference in the way that the two ML techniques acquire information from the tomography measurements. When the same analysis was performed for learning of the concurrence directly from elements of the density matrix, such a discrepancy was not observed.

Neural networks have the ability to build efficient representations (intrinsic dimensionality reduction) using large data vectors, while classical ML models need fewer better quality and more universal features and cannot cope with more diffused information. Therefore, the RF model uses measurements that are more relevant for the extraction of information (also from other measurements), whereas the NN model behaves in accordance with the QST, being able to effectively predict concurrence from the full measurement set.

On the other hand, the RF models can outperform the networks in the presence of noise perturbing a single measurement, as demonstrated in Appendix E. The higher robustness of the RF model to the noise perturbation stems from the cooperation of a big number of classifiers relying on different features of the input vector.

In general, any model performs better when all measurements are reliable, but the different feature importance displayed by the models which translated into corresponding behavior in the presence of noise on the corresponding measurement suggests that the choice of ML strategy is of vital importance when some measurements are likely to be less reliable. Machine learning models can be used to test the universality and interdependence of measurements as an information resource for quantum state reconstruction.

ACKNOWLEDGMENTS

K.R. received funding within the QuantERA II Programme, which has received funding from the EU H2020 research and innovation program under Grant Agreement No. 101017733 and from the funding organization MEYS. M.K. and J.P. acknowledge support from National Science Centre, Poland, under Grant No. 2021/43/D/ST3/01989. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support through computational Grant No. PLG/2024/017284. This work was supported by the research program of the Strategy AV21 AI: Artificial Intelligence for Science and Society.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: DATASET INSPECTION

In order to have a more comprehensive picture of the studied datasets, we analyzed the data using standard unsupervised ML techniques for feature vector dimensionality reduction: principal component analysis (PCA) and the more recent t-distributed stochastic neighbor embedding (t-SNE).

1. Principal component analysis

First, we examined the training dataset using the PCA method [43], an unsupervised dimensionality reduction technique used to reduce the dimensionality of complex datasets while retaining their most important characteristics. Principal component analysis finds a linear transformation of the original features into a set of new variables, called principal components. The principal components are ordered by their explained variance ratio, which is defined as the amount of variance they capture in the data. The explained variance ratio helps us determine how many principal components to retain for analysis, as components with higher explained variance ratios contain more information about the dataset's structure.

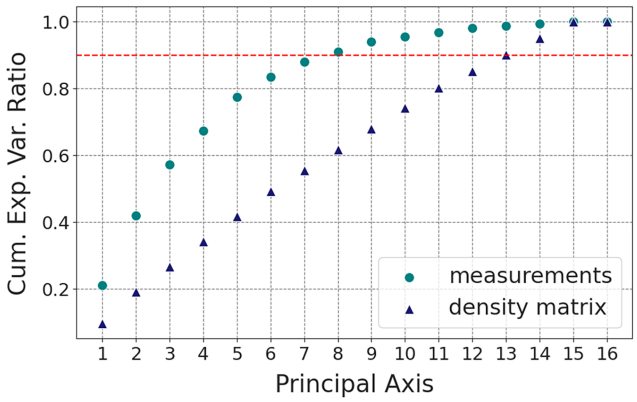


FIG. 7. Cumulative explained variance ratio of PCA performed on dataset vectors of measurements (circles) and on components of density matrices (triangles). The red dashed lines show the boundary of 90% of preserved information.

Our goal is to estimate how many components of the feature vectors we need to preserve a majority of the information in the dataset. To this end we performed PCA on the training dataset and evaluated the explained variance ratios of each component r_i, where $i = 1, 2, \ldots, 16$. For comparison, we performed PCA on both training datasets consisting of measurement outcomes as well as on the components of the density matrices. Figure 7 compares the cumulative explained variance ratio R_j of the PCA performed on both datasets, defined as

$$R_j = \sum_{i=1}^{j} r_i, \tag{A1}$$

where $j = 1, 2, \ldots, 16$. By definition the total sum of explained variance ratios equals 1; thus $R_{16} = 1$. Obviously, both dependences strongly differ. In the case of the dataset consisting of the density matrix elements, we need 13 principal components to reach the boundary of 0.9, which corresponds to 90% of the preserved information. On the other hand, when we deal with measurements outcomes, we need just the first eight PCA components to overcome this threshold. This observation speaks in favor of the possibility of quantum tomography based on a limited number of measurements. In addition, we notice that in the case of both datasets we reach a cumulative explained variance ratio equal to 1 for 15 principal components. This stems from the fact that the density matrix components are not independent, namely, the trace equals 1, which allows one to calculate one diagonal component using the others.

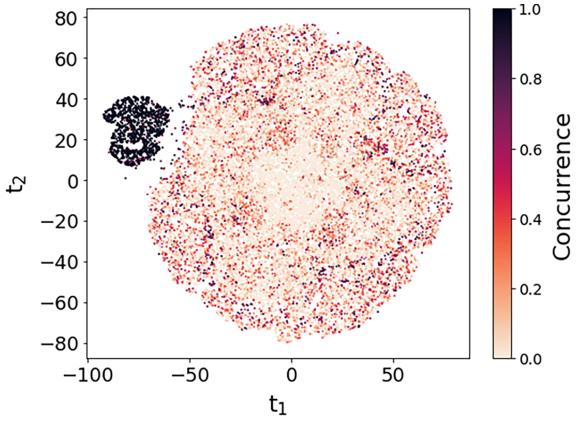


FIG. 8. Measurement vectors projected into two-dimensional space using the t-SNE method. Note that both dimensions $t_{1,2}$ defining embedding space do not have a strict interpretation.

2. t-distributed stochastic neighbor embedding

Second, we employed the t-SNE method [44], which is a dimensionality reduction technique particularly well suited for visualizing high-dimensional data by mapping them into a lower-dimensional space. Importantly, it captures the local structure of the data while preserving global relationships in the dataset. Figure 8 shows the projection of measurement vectors in a two-dimensional embedding space. Analyzing the t-SNE embedded space, one can see that the entangled states seem to cluster nicely, but it is worth noting that they also occur largely between separable states (as black dots in the bright cluster). Thus, the detection of entanglement based on this representation is worse than the ones obtained using RF or NN predictors.

APPENDIX B: RANDOM FOREST DETAILS

Each DT in a RF constitutes a tree graph structure in which each node signifies a decision rule, while each leaf node corresponds to a specific class. The purity of the kth node is specified by the Gini impurity, G_k, defined as [21]

$$G_k = 1 - \sum_{i=1}^{N} p_{i,k}^2, \tag{B1}$$

where N is the number of classes and $p_{i,k}$ is the proportion of samples of class i in the node k. The mean decrease in impurity [21,45] can be calculated from the training samples and serves as a metric within DT-based models to ascertain feature importance. It quantifies the extent to which a feature contributes to the reduction of Gini impurity in the dataset when it is utilized for data splitting at a node within the tree.

For this purpose, the `RandomForestClassifier` from the scikit-learn library [46] was employed. We applied a grid search algorithm alongside cross-validation to fine-tune the model's hyperparameters. The most effective RF hyperparameters are presented in Table II, where `n_estimators` denotes the number of DTs within the ensemble. During the search for optimal splits, only a random subset of features is considered, with its size determined by `max_features`. The Boolean parameter `bootstrap` indicates whether bootstrap samples are utilized in tree construction. Additionally, `max_depth` represents the maximum allowable tree depth and `min_samples_leaf` defines the minimum number of samples required for a leaf node. Should the aforementioned parameters be configured to `None`, this constraint is not factored into the regularization process.

TABLE II. Details of the hyperparameters of the most successful random forest models. The explanations of these hyperparameters can be found within the main text.

Model	RF1	RF2	RF3
`n_estimators`	1000	1000	1000
`max_features`	4	16	4
`max_depth`	`None`	20	`None`
`min_samples_leaf`	`None`	`None`	100
`bootstrap`	`True`	`True`	`True`

APPENDIX C: DEFINITION OF SHAPLEY VALUES AND SHAPLEY INTERACTION VALUES

The Shapley value for a feature i in a model f is defined as

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} w_1(S, N) \Delta f(S, i), \tag{C1}$$

where $w_1(S, N) = |S|!(|N| - |S| - 1)!/|N|!$ is a weighting factor, with N the set of all features and S representing a subset of features excluding i. Here

$$\Delta f(S, i) = f(S \cup \{i\}) - f(S) \tag{C2}$$

represents the marginal contribution of feature i. This definition ensures that contributions are fairly allocated based on marginal contributions across all possible subsets.

For binary classification problems, SVs can be computed separately for each class. The SVs for class 0 and class 1 are equal in magnitude but differ in sign, $\phi_i(1) = -\phi_i(0)$. This sign of the SV $\phi_i(c)$ reflects the contribution of a feature towards increasing or decreasing the probability of a given class c. A positive SV for class 1 indicates that the feature increases the probability of predicting class 1, while a negative value suggests it supports class 0 instead.

Importantly, SVs are computed for individual samples, providing a local explanation of how a particular instance is classified. To obtain a global understanding of feature importance across the dataset, we compute the global SVs over all instances in a selected set of samples $\boldsymbol{X} \subset \boldsymbol{S}_{\text{test}}$,

$$\Phi_i = E_k\left[\phi_i^{(k)}\right], \tag{C3}$$

where $\phi_i^{(k)}$ represents the SV of feature i for sample $X_k \in \boldsymbol{X}$.

While SVs explain individual feature contributions, they do not capture interactions between features. The SIV extends the concept by quantifying the pairwise interactions between features in a model. It measures how much the joint presence of two features contributes to the prediction beyond their individual effects.

Similarly, for single-value SVs, a SIV for two features i and j in a model f is defined as

$$\phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} w_2(S,N) \Delta f(S,i,j), \tag{C4}$$

where the weighting factor for subset size is $w_2(S,N) = |S|!(|N| - |S| - 2)!/2(|N|!)$ and

$$\Delta f(S,i,j) = f(S \cup \{i,j\}) - f(S \cup \{i\}) - f(S \cup \{j\}) + f(S) \tag{C5}$$

quantifies the additional contribution of both features appearing together. This formulation ensures that the interaction term captures the additional contribution of both features appearing together, beyond what would be expected from their individual SVs. Shapley interaction values are particularly useful for understanding dependences between features and identifying synergistic or redundant effects in complex models.

The global SIVs are defined as

$$\Phi_{i,j} = E_k\left[\phi_{i,j}^{(k)}\right], \tag{C6}$$

where $\phi_{i,j}^{(k)}$ is the SIV of features i and j for sample $X_k \in \boldsymbol{X}$.

Shapley interaction values are closely related to standard SVs. The total contribution of a feature i can be decomposed into its individual SV and its interactions with other features

$$\phi_i = \phi_i^{\text{ind}} + \sum_{j \neq i} \phi_{i,j}, \tag{C7}$$

where ϕ_i^{ind} represents the independent contribution of feature i and the second term accounts for the total interaction effects with other features. This decomposition highlights that the SV of a feature includes both its standalone effect and all pairwise interactions with other features. If there are no significant interactions, then $\phi_{i,j} \approx 0$, reducing the model to an additive form where each feature contributes independently.

In this study, we have calculated SVs and SIVs for the RF classifier using shap's `TreeExplainer` class. The `TreeExplainer` is specifically optimized for tree-based models, leveraging their structure to efficiently compute SVs and interactions [42].

APPENDIX D: SHAPLEY INTERACTION VALUES FOR RFS

To elucidate the different approaches of the RF classifier to separable and entangled samples, we extend our analyses to SIVs. The global SIVs for separable and strongly entangled samples are shown in Figs. 9(a) and 9(b), respectively. Since the computational costs of SIVs for a large model with 16 features are enormous, we did the calculations using a smaller RF model consisted of 100 decision trees regularized by `min_samples_leaf` = 20. After training the model using the $\boldsymbol{S}_{\text{train}}$ dataset, it shows Shapley values similar to those calculated for the RF1 model. Thus, we believe that the SIVs computed for the simplified model might shed some light on the performance of the large RF models studied in this paper.

If a SIV in Fig. 9 is positive, it drives the prediction of the model towards class 1 (entangled), while the negative

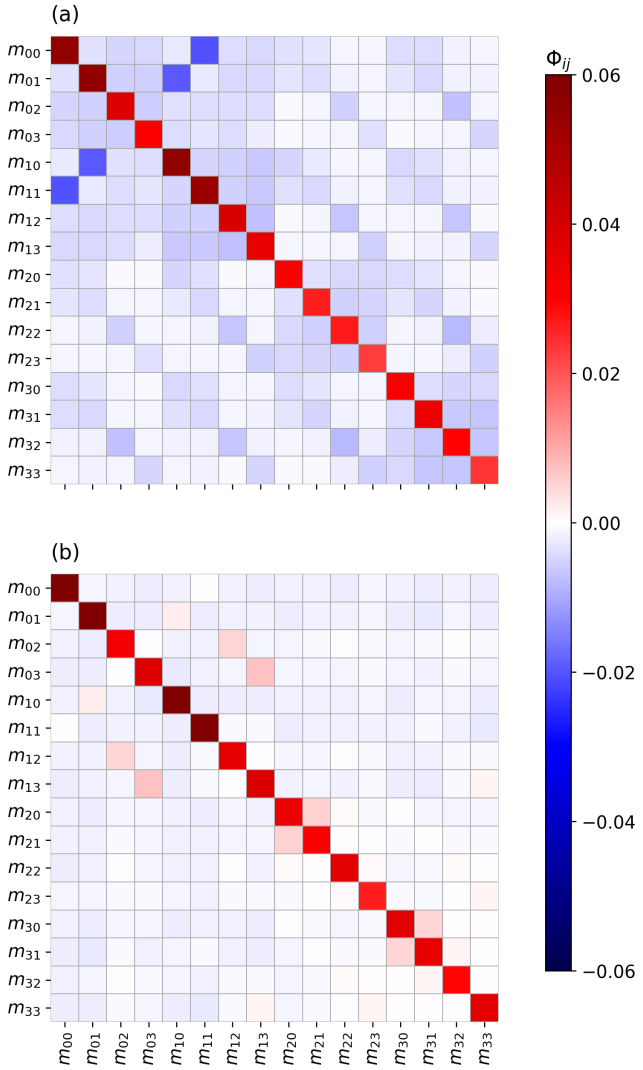


FIG. 9. Global SIVs calculated for a simplified random forest model using a subset of $\boldsymbol{S}_{\text{test}}$ with (a) $C < 10^{-6}$ (separable) and (b) $C > 0.99$ (strongly entangled).

SIVs increase the probability of predicting class 0 (separable). Thus, a first look at the case with strongly entangled samples reveals that the classifier classifies these samples based on the single features. Among them the four measurements in the block $\boldsymbol{M}_{\text{A}}$ appear to be most important, which is in accord with our previous results for the RF classifier. In addition, one might observe a minor contribution of interactions between features. All the other SIVs are close to zero. On the other hand, in the case of separable samples, the diagonal elements of the Shapley interaction matrix drive the predictions towards the opposite class (entangled). In this case, however, the strong contribution of feature interactions prevails and draw the prediction towards the correct result. Among the interactions, the most important appears to be the interaction between measurements (m_{00},m_{11}) and (m_{01},m_{10}), which are also located in block $\boldsymbol{M}_{\text{A}}$.

The observations from Figs. 6(a) and 6(b) show that in the case of the samples with $C > 0.99$ the most important features are in the $\boldsymbol{M}_{\text{A}}$ block, while for classification of the separable

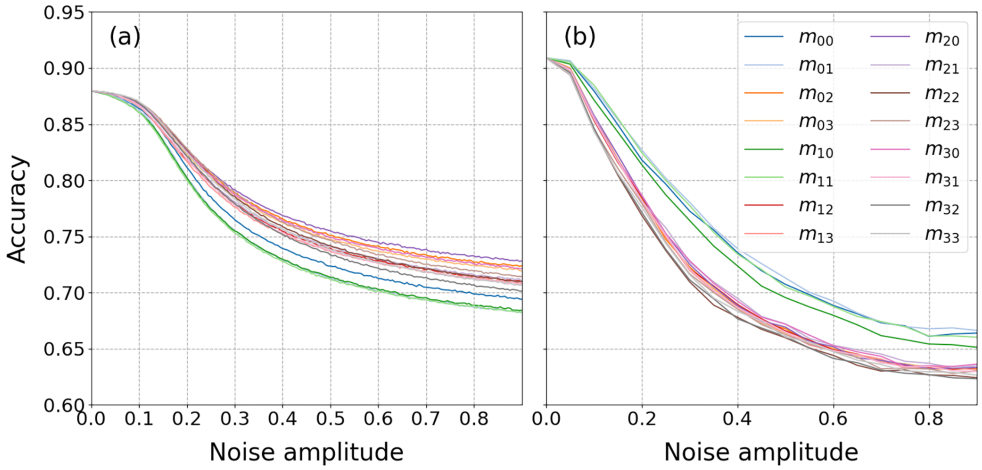


FIG. 10. Perturbation measure for (a) the random forest classifier and (b) the neural network regressor as a function of the noise amplitude σ.

samples the most important measurements are located in the $\boldsymbol{M}_{\mathrm{D}}$ block. Figure 9 shows that the structure of the diagonal SIV does not significantly change for the separable and strongly correlated samples. A crucial effect is the negative contributions of the feature interactions, which compensate the strong positive effect of diagonal elements. Moreover, this result suggests that the RF models tend to identify the separable samples not based on the single measurements but on the mutual correlations of the features.

APPENDIX E: PERTURBATION MEASURE

In addition, we compared the performance of the random forest and neural network for random noise applied to a single measurement outcome, as described in Sec. VI B 2. Figure 10 shows the accuracy calculated for the RF1 and NN models. Although in the case of zero noise the accuracy of the NN model exceeds that of the RF model, in the presence of a certain noise level the RF model might on average perform better.

The reason for this behavior is the large number of classifiers in the RF ensemble, which makes the method more robust against noise and errors. Because each tree in the RF ensemble classifies a sample in a different way, a random change of one feature is unlikely to affect the prediction of every tree in the ensemble. Therefore, even stronger noise in the data does not significantly affect the overall performance of the model. On the other hand, the NN is a highly nonlinear model, which often relies not just on the single parameters, but also on their correlations. Thus, a systematic perturbation of one of the correlated features might significantly reduce the model accuracy.

[1] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).
[2] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).
[3] T. Xin, S. Lu, N. Cao, G. Anikeeva, D. Lu, J. Li, G. Long, and B. Zeng, Local-measurement-based quantum state tomography via neural networks, npj Quantum Inf. **5**, 109 (2019).
[4] A. Melkani, C. Gneiting, and F. Nori, Eigenstate extraction with neural-network tomography, Phys. Rev. A **102**, 022412 (2020).
[5] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, Quantum state tomography with conditional generative adversarial networks, Phys. Rev. Lett. **127**, 140502 (2021).
[6] Y. Quek, S. Fort, and H. K. Ng, Adaptive quantum state tomography with neural networks, npj Quantum Inf. **7**, 105 (2021).
[7] D. Koutný, L. Motka, Z. Hradil, J. Řeháček, and L. L. Sánchez-Soto, Neural-network quantum state tomography, Phys. Rev. A **106**, 012409 (2022).
[8] T. Schmale, M. Reh, and M. Gärttner, Efficient quantum state tomography with convolutional neural networks, npj Quantum Inf. **8**, 115 (2022).
[9] H. Ma, D. Dong, I. R. Petersen, C.-J. Huang, and G.-Y. Xiang, Neural networks for quantum state tomography with constrained measurements, Quantum Inf. Process. **23**, 317 (2024).
[10] M. Krawczyk, J. Pawłowski, M. M. Maśka, and K. Roszak, Data-driven criteria for quantum correlations, Phys. Rev. A **109**, 022405 (2024).
[11] J. Pawłowski and M. Krawczyk, Identification of quantum entanglement with siamese convolutional neural networks and semisupervised learning, Phys. Rev. Appl. **22**, 014068 (2024).
[12] N. Taghadomi, A. Mani, A. Fahim, and A. Bakouei, Effective detection of quantum discord by using convolutional neural networks, Quantum Mach. Intell. **7**, 40 (2025).
[13] Y. Chen, Y. Pan, G. Zhang, and S. Cheng, Detecting quantum entanglement with unsupervised learning, Quantum Sci. Technol. **7**, 015005 (2022).
[14] N. Asif, U. Khalid, A. Khan, T. Q. Duong, and H. Shin, Entanglement detection with artificial neural networks, Sci. Rep. **13**, 1562 (2023).
[15] J. Ureña, A. Sojo, J. Bermejo-Vega, and D. Manzano, Entanglement detection with classical deep neural networks, Sci. Rep. **14**, 18109 (2024).

[16] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, Separability-entanglement classifier via machine learning, Phys. Rev. A **98**, 012315 (2018).

[17] B. C. Hiesmayr, Free versus bound entanglement, a NP-hard problem tackled by machine learning, Sci. Rep. **11**, 19739 (2021).

[18] C. B. D. Goes, A. Canabarro, E. I. Duzzioni, and T. O. Maciel, Automated machine learning can classify bound entangled states with tomograms, Quantum Inf. Process. **20**, 99 (2021).

[19] B. Wang, Learning to detect entanglement, arXiv:1709.03617.

[20] M. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, Ensemble deep learning: A review, Eng. Appl. Art. Intell. **115**, 105151 (2022).

[21] L. Breiman, Random forests, Mach. Learn. **45**, 5 (2001).

[22] J. Fürnkranz, Decision tree, in *Encyclopedia of Machine Learning*, edited by C. Sammut and G. I. Webb (Springer, Boston, 2010), pp. 263–267.

[23] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees* (Routledge, Abingdon, 2017).

[24] L. Breiman, Bagging predictors, Mach. Learn. **24**, 123 (1996).

[25] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature (London) **521**, 436 (2015).

[26] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, edited by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Springer, Cham, 2015), pp. 234–241.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, 2017*, edited by U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, and R. Fergus (Curran, Red Hook, 2017), pp. 5998–6008.

[28] E. Gibney and D. Castelvecchi, Physics nobel scooped by machine-learning pioneers, Nature (London) **634**, 523 (2024).

[29] M. Belkin, D. Hsu, S. Ma, and S. Mandal, Reconciling modern machine-learning practice and the classical bias–variance trade-off, Proc. Natl. Acad. Sci. USA **116**, 15849 (2019).

[30] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. D. Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, R. Jiang, H. Khosravi, F. Lecue, G. Malgieri, A. Páez, W. Samek, J. Schneider, T. Speith, and S. Stumpf, Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions, Inf. Fusion **106**, 102301 (2024).

[31] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett. **80**, 2245 (1998).

[32] P. Horodecki, Separability criterion and inseparable mixed states with positive partial transposition, Phys. Lett. A **232**, 333 (1997).

[33] M. Horodecki, P. Horodecki, and R. Horodecki, Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? Phys. Rev. Lett. **80**, 5239 (1998).

[34] T. Yu and J. H. Eberly, Evolution from entanglement to decoherence of bipartite mixed "X" states, Quantum Inf. Comput. **7**, 459 (2007).

[35] P. E. Mendonça, M. A. Marchiolli, and D. Galetti, Entanglement universality of two-qubit X-states, Ann. Phys. (N.Y.) **351**, 79 (2014).

[36] F. Mezzadri, How to generate random matrices from the classical compact groups, Not. Am. Math. Soc. **54**, 592 (2006).

[37] R. E. Schapire, The strength of weak learnability, Mach. Learn. **5**, 197 (1990).

[38] L. Breiman, Arcing classifier (with discussion and a rejoinder by the author), Ann. Stat. **26**, 801 (1998).

[39] V. Nair and G. E. Hinton, Rectified linear units improve restricted Boltzmann machines, in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, 2010, edited by J. Fürnkranz and T. Joachims (Omnipress, Madison, 2010), pp. 807–814.

[40] V. Raghavan, P. Bollmann, and G. S. Jung, A critical investigation of recall and precision as measures of retrieval system performance, ACM Trans. Inf. Syst. **7**, 205 (1989).

[41] L. S. Shapley, A value for n-person games, in *Contributions to the Theory of Games*, edited by H. W. Kuhn and A. W. Tucker (Princeton University Press, Princeton, 1953), Vol. 2, pp. 307–317.

[42] S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Ref. [27]), pp. 4768–4777.

[43] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics (Springer, New York, 2002).

[44] G. E. Hinton and S. T. Roweis, Stochastic neighbor embedding, in *Advances in Neural Information Processing Systems 15*, edited by S. Becker, S. Thrun, and K. Obermayer (MIT Press, Cambridge, 2003), pp. 857–864.

[45] E. Scornet, Trees, forests, and impurity-based variable importance in regression, Ann. Inst. H. Poincaré Probab. Statist. **59**, 21 (2023).

[46] 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).

PHYSICAL REVIEW APPLIED **25,** 064032 (2026)

AI-enhanced tuning of quantum dot Hamiltonians toward Majorana modes

Mateusz Krawczyk and Jarosław Pawłowski*

Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

(Received 19 January 2026; revised 20 March 2026; accepted 11 May 2026; published 8 June 2026)

We propose a neural-network-based model capable of learning the broad landscape of working regimes in quantum dot simulators and using this knowledge to autotune these devices—based on transport measurements—toward obtaining Majorana modes in the structure. The model is trained in an unsupervised manner on synthetic data in the form of conductance maps, using a physics-informed loss that incorporates key properties of Majorana zero modes. We show that, with appropriate training, a deep vision-transformer network can efficiently memorize the relations between Hamiltonian parameters and structures on conductance maps and use them to propose parameter updates for a quantum dot chain that drives the system toward a topological phase. Starting from a broad range of initial detunings in parameter space, a single update step is sufficient to generate nontrivial zero modes. Moreover, by enabling an iterative tuning procedure—where the system acquires updated conductance maps at each step—we demonstrate that the method can address a much larger region of the parameter space.

DOI: 10.1103/xkbl-ctwn

I. INTRODUCTION

Majorana zero modes (MZMs) are emergent quasiparticles predicted to obey non-Abelian statistics, enabling topologically protected quantum information processing [1–4]. Their realization in hybrid superconductor-semiconductor systems requires precise control over local parameters such as chemical potentials, interdot couplings, and spin-orbit interactions [5,6]. Small deviations from the so-called *sweet-spot* conditions of the Kitaev chain (KC) [7] can destroy the topological gap and delocalize MZMs [8]. Moreover, it is not trivial to distinguish between MZMs and different nontopological zero-bias peaks [9–12]. Consequently, achieving robust Majorana states experimentally remains a major challenge, particularly in the presence of fabrication disorder and parameter noise.

To address this challenge, Fulga *et al.* [13] proposed an adaptive tuning protocol for a superconductor-proximitized chain of quantum dots (QDs) that simulates a KC generalized on a nonuniform parameter distribution across dots, trying to demonstrate how gate-controlled voltages and superconducting phases can be adjusted to reach a topologically nontrivial regime. In this paper we revisit this problem by trying to automate it using deep neural estimators. More recently, further works have indicated that through appropriate tuning, it is possible to realize a KC with emerging MZMs using QDs coupled through elastic cotunneling and crossed Andreev reflection [5,14–17]. All these proposals utilize coupling with more readily available *s*-wave superconductors; however, there are also proposals employing *p*-wave superconductors [18,19].

At the same time, a different strategy for machine learning (ML)-assisted autotuning of QD-based quantum simulators [20–22] using transport measurements is gaining significant interest [23–27]. The use of transport measurements in the form of *conductance maps* [22,28] to derive insights about the system Hamiltonian [29–35] appears to be a promising path toward the automatic tuning of its parameters. In particular, inverting the measured conductance matrices to determine electrostatic potential disorder, using evolutionary optimization [33], supervised neural networks (NNs) [34–36], or both [37], is a natural first step that can allow precise tuning of Hamiltonian parameters toward MZMs. Although these methods already allow for successful mitigation of disorder in nanowires [33,37], they rely on indirect, nondifferentiable cost functions and heuristic evolutionary searches, rather than learning the broader behavior of the underlying physical system.

In contrast, here we introduce an unsupervised *physics-informed* [38,39] NN-based autotuning (PINNAT) model with Majorana physics embedded in the loss function, and show that with proper training the model can memorize various structures on conductance maps, which enables it to build an efficient quantum simulator autotuning system. Our AI-enhanced adaptive tuning framework directly extends QD simulators by leveraging *vision transformer* (ViT) [40] NN architecture. The scheme of the proposed approach is presented in Fig. 1, which shows the

*Contact author: jaroslaw.pawlowski@pwr.edu.pl

 © 2026 American Physical Society

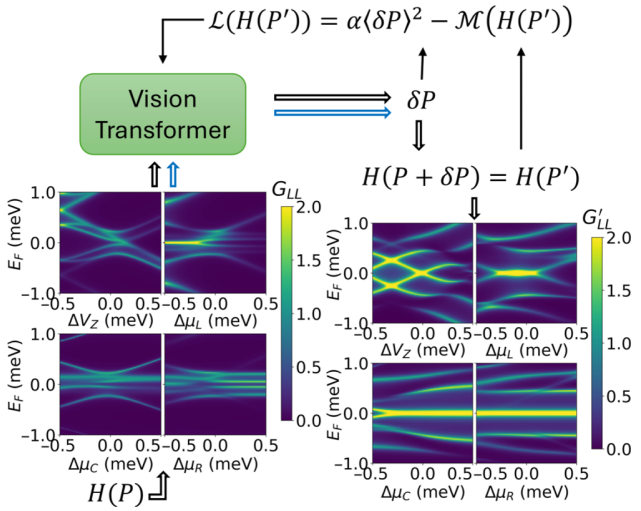


FIG. 1. Scheme of the quantum dot–based Kitaev-chain-simulator autotuning system (physics-informed neural-network-based autotuning). Black (blue) arrows show the system training (inference) path. The vision-transformer model is fed with conductance maps to predict parameter corrections that should decrease the $\mathcal{M}$-based loss function, simultaneously increasing the probability of the emergence of Majorana zero modes.

information flow during the model training (black arrows) and evaluation (*inference*, blue arrows).

II. MODEL

Following a lattice model from Ref. [13], the effective superconductor-proximitized Rashba-Zeeman Hamiltonian for a chain of spinful single-level QDs can be expressed as follows:

$$\begin{aligned} H = \sum_{n,s,s'} \Big\{ & (-\mu_n\sigma_0 + V_Z\sigma_z)_{ss'}\, c^\dagger_{n,s}c_{n,s'} \\ & + \frac{1}{2}\Big[\left(\Delta_n e^{i\phi_n} i\sigma_y\right)_{ss'} c^\dagger_{n,s}c^\dagger_{n,s'} + \text{H.c.}\Big] \\ & + \left[t_n (e^{i\boldsymbol{\lambda}_n\cdot\boldsymbol{\sigma}})_{ss'} c^\dagger_{n,s}c_{n+1,s'} + \text{H.c.}\right]\Big\}, \end{aligned} \tag{1}$$

with the on-site potentials μ_n, interdot hopping amplitudes t_n modulated by the (Rashba) spin-orbit vector $\boldsymbol{\lambda_n} = \lambda_n(\sin\rho_n\cos\xi_n, \sin\rho_n\sin\xi_n, \cos\rho_n)$, Zeeman energy V_Z, proximity-induced superconducting (s-wave) pairing $\Delta_n e^{i\phi_n}$, with n describing the dot number, and s the spin degree of freedom. The values of those parameters are then carefully adjusted to have MZMs emerge in the system. One set of tuned (yet uniform) parameters (which we call *reference parameters* P_0) for the $N = 3$ QD chain includes: $\mu = 0.6$ meV, $t = 0.25$ meV, $\lambda = 0.27\pi$, $\rho = \xi = \pi/2$, $V_Z = 0.5$ meV, and $\Delta = 0.25$ meV. The Rashba length $\lambda = 0.27\pi$ was tuned so that at $\mu = 0.6$ meV, two energy levels touch at zero energy [c.f., Fig. 5(a)]. Also note that some parameters, i.e., $\{\mu_n, t_n, \lambda_n\}$ (seven in total) are easier to tune—electrically (via local gating)—than the others: the global Zeeman field V_Z and the proximity-induced superconducting gap $\{\Delta_n\}$.

To train the ViT model in a physics-informed manner, we introduce the differentiable quasimetric $\mathcal{M}$ to quantify how close a given Hamiltonian is to the MZM regime. The proposed phenomenology combines several physical indicators: edge-state localization, zero-energy spectral weight, and parity (electron-hole) symmetry. For a detailed $\mathcal{M}$ definition, see the Methods section. We also note that we tested the popular Majorana *polarization* [41–43] measure, directly related to the parity operator, but it yielded unsatisfactory training results. This is because it fails to discriminate topologically trivial zero modes that localize in the central QD or antilocalize in the left and right QDs (see the Supplementary Information of Ref. [17] for examples of such states). We define our measure in a way that also discriminates these states.

We started with random set of parameters P and collected conductance maps $G(H(P))$—details on G calculations can be found in the Methods section. The PINNAT model fed by G maps predicts corrective updates δP to a subset of Hamiltonian parameters. We assume that, depending on the experimental setup, different subsets of parameters may be available for tuning. Two versions of the PINNAT model to mimic different experimental arrangements were trained. The first one was trained to predict corrections to electrically controlled local parameters $\{\mu_n, t_n, \lambda_n\}$, each of which can be adjusted independently, while leaving $\{\Delta_n, V_Z\}$ as background variables. The second one was allowed to adjust the $\{\mu_n, V_Z\}$ pair, where μ_n can be corrected locally, while V_Z was adjusted globally. We also have to keep in mind that the PINNAT model (as well as the experimentalist) *does not know* which parameters were detuned.

Maximizing $\mathcal{M}$ drives the NN to adjust P such that edge states become increasingly localized and the strong midgap energy signal emerges, which is visualized in Fig. 1 by output conductance maps [this example is further analyzed in Fig. 5(d)]. Specifically, the PINNAT model was trained to minimize the loss,

$$\mathcal{L}(H(P')) = \alpha\,\langle\delta P\rangle^2 - \mathcal{M}(H(P')), \tag{2}$$

for the Hamiltonian with tuned parameters $H(P') = H(P+\delta P)$. Additionally, to force the PINNAT model to predict the smallest possible correction δP to the parameters P, an extra regularization term $\alpha\langle\delta P\rangle^2$ was included with the factor $\alpha = 0.1$. Training details, including the parameter sampling used to generate the synthetic training set of conductance maps, are described in the Methods section.

III. RESULTS

To present the PINNAT performance in parameter tuning, we plot the values of $\mathcal{M}$ when selected three-QD Hamiltonian parameters are disturbed. Specifically, we shift (detune) the values of the given parameters (the rest are kept at default values P_0)—to obtain P—and verify what corrections δP are predicted by the NN models. Then the $\mathcal{M}(H(P+\delta P))$ value is plotted. In Figs. 2(b) and 2(c) we show how $\mathcal{M}$ changes with uniform (i.e., the same for each QD) parameter shifts for models tuning $\{\mu_n, t_n, \lambda_n\}$ and $\{\mu_n, V_Z\}$, respectively. The values for $\mathcal{M}(H(P))$, i.e., before correction, are presented as reference values in Fig. 2(a). For example, in the t vs μ plot we shift parameters to $t_1 = t_2 = t$ and $\mu_L = \mu_C = \mu_R = \mu$ obtaining P [Fig. 2(a)], then predict correction δP and present $\mathcal{M}(H(P+\delta P))$ in respective positions in Figs. 2(b) and 2(c). Additionally, in Fig. 2(a), orange dashed curves mark the analytical conditions: (1) $\Delta = \sqrt{V_Z^2 - \mu^2}$ for the existence of two zero modes in decoupled QDs, and (2) $\lambda = \arctan(\mu/\Delta)$ corresponding to the KC sweet spot (see the Appendix A for details).

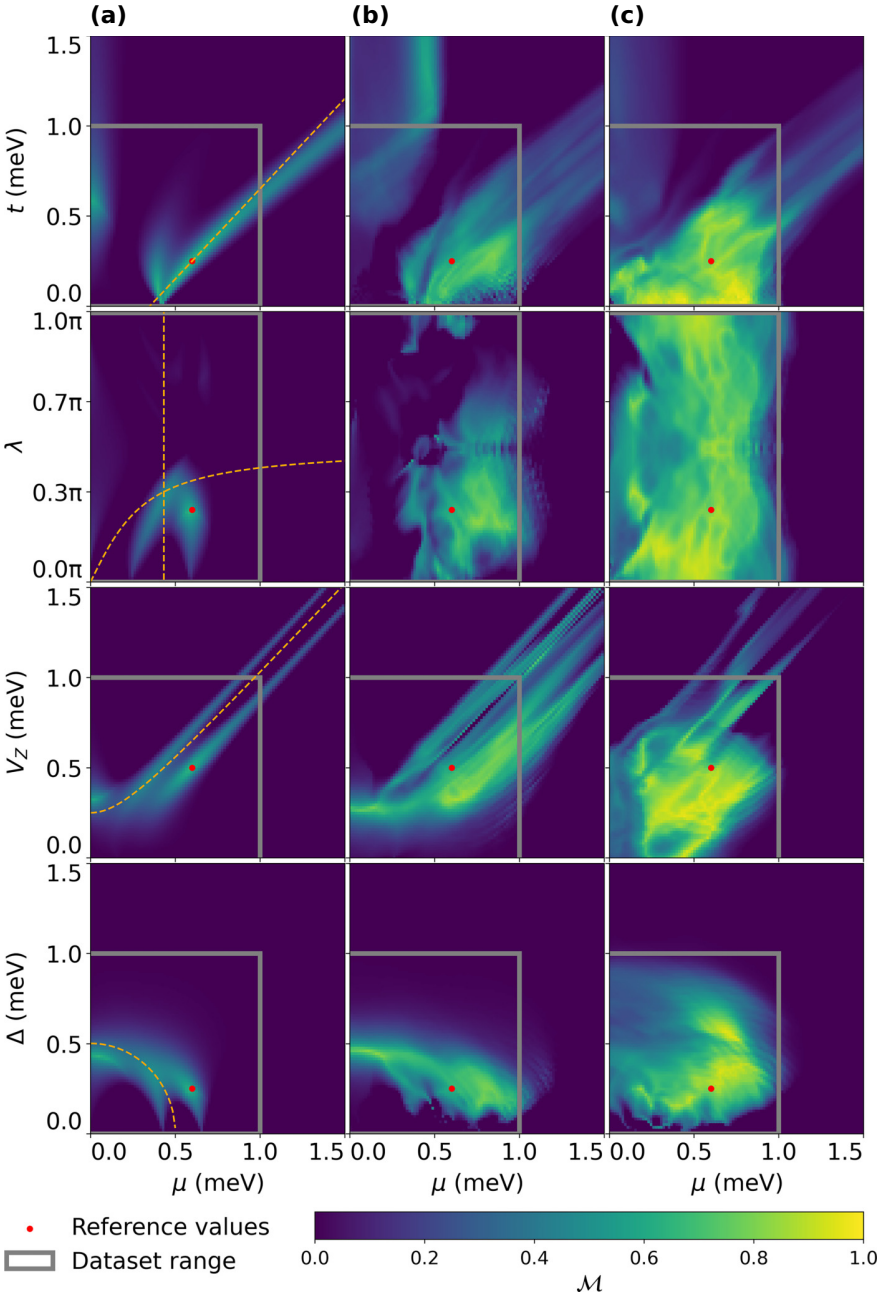


FIG. 2. $\mathcal{M}$ metric map for uniform (across the dots) change of different parameters—resulting in shifted P—for (b) model adjusting $\{\mu_n, t_n, \lambda_n\}$, and (c) adjusting $\{\mu_n, V_Z\}$. The ranges of the parameter sampling that were used in the model training set are marked with a gray rectangle. The reference parameter values P_0 are marked with a red dot. In (a), we present values of $\mathcal{M}(H(P))$ before parameter tuning, while in (b) and (c) we present $\mathcal{M}(H(P+\delta P))$ after tuning.

Results in Figs. 2(b) and 2(c) show that the PINNAT models can effectively learn to identify various $H(P)$ regimes from G maps and use them to significantly increase—in comparison with Fig. 2(a)—the regions with $\mathcal{M} > 0$ by proposing corrections δP to the subset of the parameters. Unsurprisingly, the models are most effective in the parameter regime covered in the training data, although some ability to *generalize* on unseen parameter ranges can be noticed, especially for t and V_Z shifts in Fig. 2(b). On the other hand, the PINNAT model that updates $\{\mu_n, t_n, \lambda_n\}$ exhibits difficulties in some regions in predicting appropriate corrections for the values of t and λ, even though they are allowed to be modified.

Interestingly, the model that adjusts $\{\mu_n, V_Z\}$, presented in Fig. 2(c), is slightly more effective in proposing corrections, both in terms of the parameter range that can be effectively corrected and the probability of the emergence of MZMs, as indicated by higher values of $\mathcal{M}$. This behavior is specifically worth highlighting as the model does not modify the t_n, λ_n, and Δ parameters, although it can effectively amend them by updating μ_n and V_Z. These results show that V_Z plays a more important role in controlling the system than the hoppings t_n and spin-orbit interaction (SOI) λ_n. This is consistent with observations known from Rashba wires [1,44]. Having a (perpendicular) magnetic field removes the so-called fermion doubling problem by opening a gap (at $k = 0$). The size of this gap increases with the Zeeman field, which means that for larger Zeeman splitting there is a wider range of μ_n for which the nontrivial phase is present. This is what we observe. However, it should be noted that when the hopping parameters t, λ are close to zero in the first and second rows of Fig. 2(c), the model likewise selects symmetric zero modes localized at the edge QDs with a high $\mathcal{M}$; nevertheless, in this case (in the limit $t \to 0$) the system is clearly in a trivial phase [44]. For $\lambda = 0$ there is no possibility of realizing p-wave pairing in the KC, and the entire model breaks down.

One way to mitigate this behavior is to include regularization terms in the training phase that push these parameters (t and λ) away from zero. However, to simplify the training procedure, we decided not to include that in this work.

PINNAT behavior is further investigated in Fig. 3, which shows how $\mathcal{M}$ changes with shifts of (a pair of selected) local parameters. In the case of a shift in the μ_n parameter, both models produce comparable results; although tuning $\{\mu_n, t_n, \lambda_n\}$ allows one to cover a wider range of parameters, adjusting $\{\mu_n, V_Z\}$ allows one to reach higher values of $\mathcal{M}$. Simultaneously, sweeping t_n and λ_n produces larger discrepancies in favor of the PINNAT model that corrects the parameters $\{\mu_n, V_Z\}$. Nevertheless, both models demonstrate the ability to effectively correct the dot-specific parameters and increase the probability of observing desired MZM modes.

Since the models tend to work in a rather limited area of parameters covered in the training data, we propose an iterative modification to the proposed procedure. After applying the initial corrections proposed by PINNAT, one can measure the conductance maps for the corrected system and pass them to the NN in the subsequent step of autotuning. As a result, in Fig. 4, we can notice that within ten steps of corrections, the $\mathcal{M}$ metric can be significantly increased even in the regions with an initially zero value of $\mathcal{M}$.

Finally, we present a specific example of the autotuning procedure for the λ_n parameter. In Fig. 5, we subsequently plot the eigenvalue spectrum of the three-QD-chain Hamiltonian H as a function of the global offset μ added to local potentials $\mu_n + \mu$. In Fig. 5(a) we have exactly the reference parameter case, $H(P_0)$, where we can clearly observe MZMs separated (by a topological gap) from the rest of the spectrum with maintained electron-hole symmetry and (partially) localized on the edges. Conductance maps for the reference parameter P_0 case, $G(H(P_0))$, are presented in Fig. 6.

Adding noise to λ_n (by setting $\lambda_1 = 0.62\pi$ and $\lambda_2 = 0.41\pi$)—presented in Fig. 5(b)—results in reduced $\mathcal{M}$ values and vanishing MZMs. Then the noisy parameter can be corrected using either of the two presented NN models. Tuning $\{\mu_n, t_n, \lambda_n\}$—presented in Fig. 5(c)—makes MZMs reappear with the desired properties, i.e., they are localized on the edges, electron-hole symmetry is preserved, and the edge states are slightly gapped from the rest of the spectrum. On the other hand, tuning $\{\mu_n, V_Z\}$—Fig. 5(d)—not only leads to the restoration of MZMs at the exact point in the tuned parameter space with the larger topological gap, but also increases the probability of observing MZMs in a wider range of μ values. An additional example of analogous autotuning, but for the t parameter, is provided in Appendix B.

Importantly, Fig. 5 shows that the $\mathcal{M}$ metrics achieve high values only for the topological states. Specifically, trivial zero modes like the Andreev bound state (ABS) [45,46] indicated by the orange arrows in Figs. 5(a) and 5(b) do not increase the $\mathcal{M}$ value. Since the PINNAT model is trained to increase $\mathcal{M}$, only nontrivial zero-energy states emerge after parameter tuning, highlighting the robustness of the proposed approach in distinguishing between trivial and topological states.

IV. DISCUSSION

These findings align with and extend current efforts in ML-assisted control of Majorana platforms [33,36,37]. Most notably, the recent work of Taylor and Das Sarma [37] introduced a ViT-based framework for mitigating strong disorder in Majorana nanowires. Their study demonstrated that disorder can be compensated—

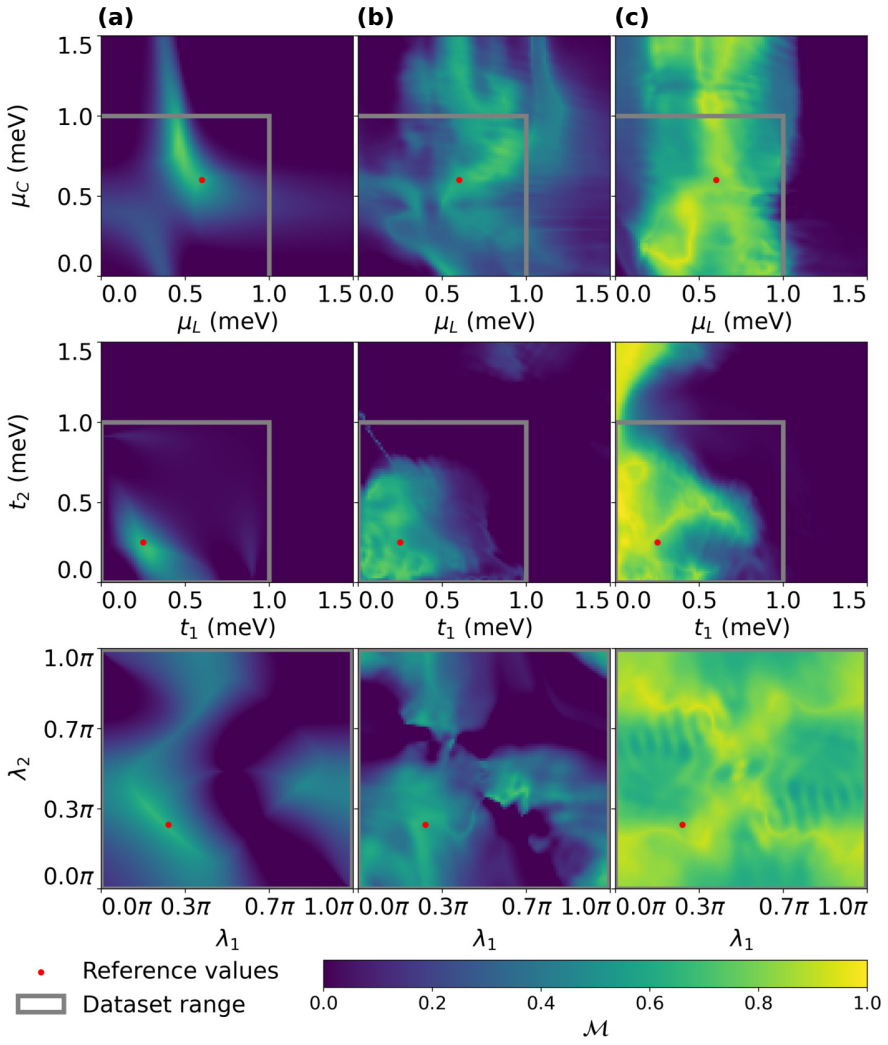


FIG. 3. Same as in Fig. 2: $\mathcal{M}$ metric maps (a) before, and (b),(c) after parameter tuning, but now with *local* shifts of a selected pair of parameters.

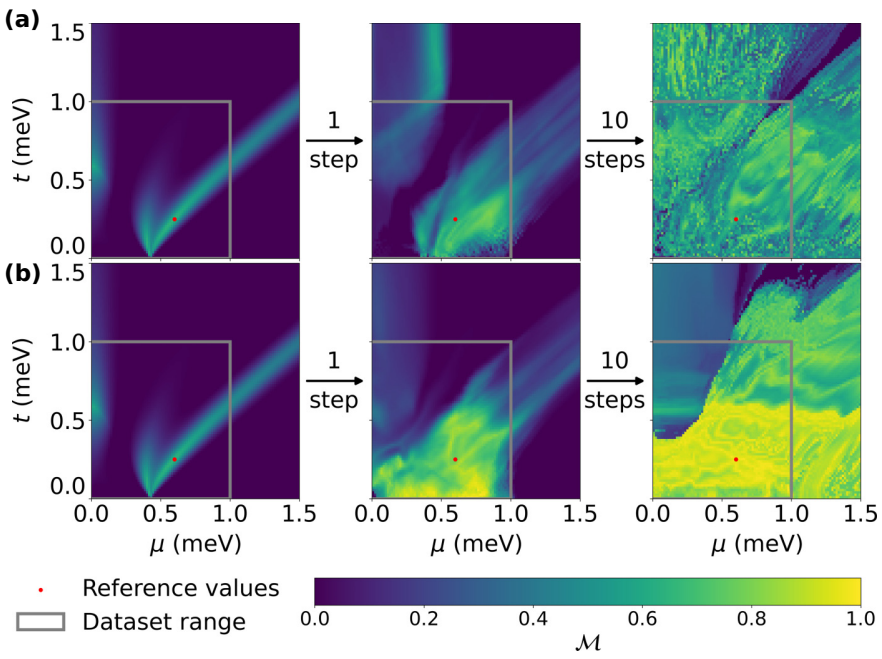


FIG. 4. Iterative autotuning procedure for the model adjusting (a) $\{\mu_n, t_n, \lambda_n\}$ and (b) $\{\mu_n, V_Z\}$. In the first (left) plot, we present the $\mathcal{M}$ map before tuning. Then, in the center, parameters are tuned with a single step of neural network (NN) corrections. Finally, in the last (right) plot, there is a map plotted for parameters tuned within ten subsequent steps of NN corrections.

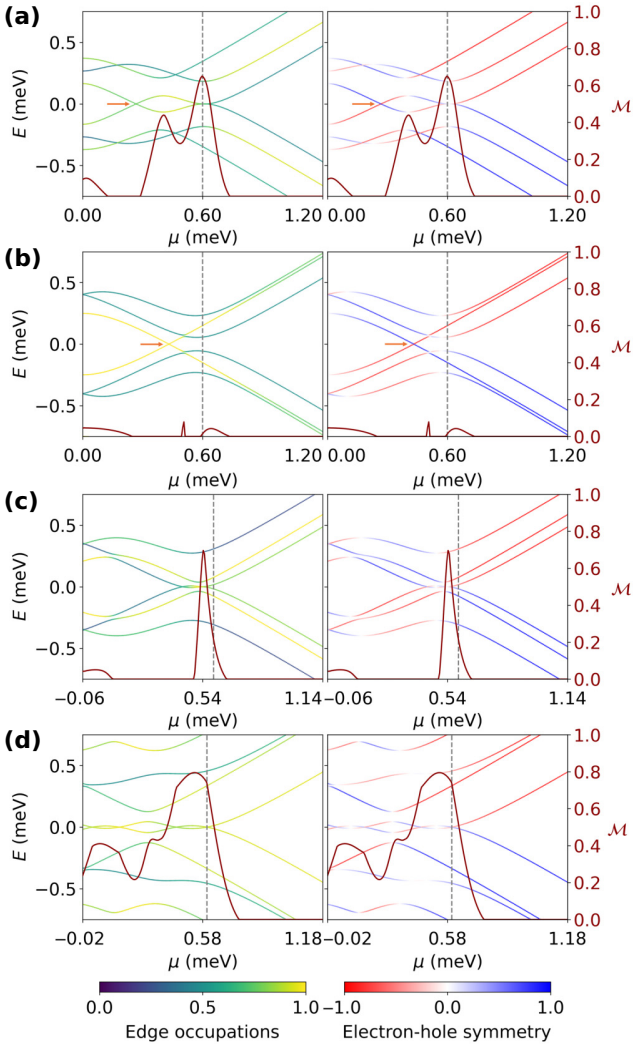


FIG. 5. Three-quantum-dot-chain Hamiltonian H as a function of the offset μ added to local potentials: $\mu_n \to \mu_n + \mu - 0.6$ meV (dashed vertical line marks the reference $\mu = 0.6$ meV). Additionally, the $\mathcal{M}$ metric is plotted, and eigenvalues are colored with edge occupations $O^{(i)}_{\text{edge}}$ (left column) or electron-hole symmetry $S^{(i)}_{eh}$ (right column), as defined in the Methods section. Plots are presented for (a) reference parameters; (b) modified λ_1 and λ_2; (c) parameters from (b) but with neural-network (NN)-tuned $\{\mu_n, t_n, \lambda_n\}$; (d) parameters from (b) with NN-tuned $\{\mu_n, V_Z\}$. Orange arrows point to the trivial zero modes.

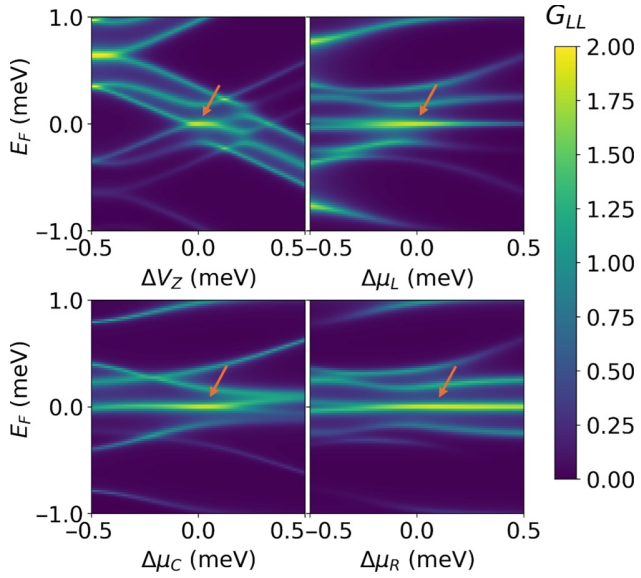


FIG. 6. Conductance maps for the reference parameters P_0. Zero-bias peaks are indicated by orange arrows.

sometimes restoring topology in systems that originally show no nontrivial regions—using conductance-based NN inference combined with an evolutionary optimizer (CMA-ES). Our approach differs in several important respects. First, instead of relying on a second-stage optimizer such as CMA-ES, our network learns Hamiltonian behaviors manifested by transport measurements to *directly* predict parameter corrections, making the tuning process faster and conceptually closer to a experimental tuning protocol. Second, whereas Ref. [37] focuses primarily on restoring scattering-invariant and local density-of-states (LDOS)-based indicators in bulk nanowires, our work addresses *QD chains*, a complementary Majorana platform with discrete QDs and gate-defined tunnel couplings between them. Third, our method integrates physics-informed NNs via the differentiable quasimetric $\mathcal{M}$, capturing edge localization, electron-hole symmetry, and spectral proximity to zero energy; this contrasts with supervised learning to predict topology-based indicators.

It is instructive to compare the parameter regimes identified by the PINNAT model with theoretical sweet-spot relations discussed in Appendix A. The parameter sets obtained in our NN-based autotuning follow the same qualitative trends, namely Zeeman energy exceeding the interdot couplings and remaining comparable to $\sqrt{\Delta^2 + \mu_n^2}$, and spin-orbit coupling following the expected sweet-spot relation with on-site potential. These relations are satisfied at the order-of-magnitude level and for averaged parameters, rather than exactly at each site, reflecting the fact that PINNAT optimizes the metric $\mathcal{M}$ instead of enforcing analytical constraints explicitly. A detailed comparison is provided in Appendix C.

We also emphasize that our results highlight the possibility of *generalization beyond the training regime*. For several parameters, notably t and V_Z, the networks propose meaningful corrections even when initial values fall outside the training set distribution. Specifically, after increasing the number of consecutive steps of corrections, the network is capable of effective autotuning in a broad area of parameters. This indicates that the unsupervised

PINNAT architecture is capable of learning physically sensible parameter correlations rather than just approximating between samples (Hamiltonians) from the training set. Such generalization is essential if these methods are to be deployed on real devices, where fabrication imperfections and drift may move operating points far from any simulated training set.

Importantly, the proposed PINNAT scheme is fully physics-informed: the MZM metric $\mathcal{M}$, together with the Hamiltonian family definition and eigensolver implementation, is explicitly encoded in the structure of the NN loss function. This approach aligns with the recent strong trend toward designing NNs that respect the underlying physics of the problem [47], for example, in studies of quantum entanglement [48,49], quantum control [50], or in the optimization of quantum tomography protocols [51]. Moreover, we argue that similar approaches based on conductance map inputs could be developed to optimize a broad class of physics-informed metrics. In particular, since direct estimation of Hamiltonian parameters from conductance maps appears feasible [52], a wide range of parameter-tuning problems could potentially be addressed using PINNAT-like models. Tuning the QD system into the single-electron (Coulomb blockade) regime could constitute a standard application in this context.

While our current proof-of-concept study is restricted mainly to a three-QD system, we show the ability to scale the proposed method to a seven-QD chain in Appendix D. With some optimization adjustments in the training procedure, we claim that it is possible to have the PINNAT model trained to propose significant corrections in larger systems, and that it can maintain its ability to autotune the Hamiltonian parameters beyond the training regime.

Although our demonstrations focus on QD chains, the PINNAT framework is not restricted to this platform. Since proximitized semiconductor nanowires are routinely modeled using discretized lattice Hamiltonians of the same type [6,7], the proposed approach could be extended to nanowire systems—albeit with a limited number of lattice sites, as discussed in Appendix D—by replacing the underlying Hamiltonian and adapting transport formalism.

The major intrinsic limitation of this work is that our method relies on simulated conductance maps, whereas experimental data may include additional noise sources, nonidealities, and systematic offsets not captured in the numerical model. Future work should therefore incorporate synthetic noise or hybrid train-on-simulation–fine-tune-on-experiment protocols to bridge this gap.

V. METHODS

The proposed method uses conductance maps as high-dimensional visual inputs for the ViT model. The conductance G is calculated using the S-matrix formalism in the wide-band limit [28] via the Weidenmüller formula [17,53] for the $N = 3$ QD chain,

$$S(E_F) = \begin{pmatrix} s^{ee} & s^{eh} \\ s^{he} & s^{hh} \end{pmatrix} = \mathbf{1} - iW^\dagger \left(E_F - H + \frac{i}{2} WW^\dagger \right)^{-1} W, \quad (3)$$

with the tunneling matrix defined as $W = \mathrm{diag}(1,0,1) \otimes \mathrm{diag}(\sqrt{\Gamma}, \sqrt{\Gamma}, -\sqrt{\Gamma}, -\sqrt{\Gamma})$ and the dot-lead coupling $\Gamma = 0.1$ meV. If we reshape the S-matrix accordingly, $S = S_{n,p,s,n',p',s'}$ ($n = L, C, R$ indexing dots; $p = 1, 2$ particles, and s spin subspaces) then the respective reflection matrices $s^{ee}_{s,s'}(i,j) = S_{i,1,s,j,1,s'}$ and $s^{he}_{s,s'}(i,j) = S_{i,2,s,j,1,s'}$ give differential conductance as, $ss^\dagger \equiv |s|^2$,

$$G_{ij}(E_F) \equiv \frac{\mathrm{d}I_i}{\mathrm{d}V_j} = 2\delta_{ij} - \mathrm{tr}\left(|s^{ee}(i,j)|^2\right) + \mathrm{tr}\left(|s^{he}(i,j)|^2\right) \quad (4)$$

in units of e^2/h (where e is electron charge and h is Planck constant), where $i,j = L, R$ denote left or right lead, and E_F is the Fermi energy in the leads. The input maps include four conductance G_{ij} components: G_{LL}, G_{LR}, G_{RL}, and G_{RR}, where for instance $G_{LL} = dI_L/dV_L$, with I_L denoting current through the left lead, and V_L the bias voltage of the left lead. Similarly, other components can be defined by using different combinations of left and right leads. Noteably, each conductance map is defined by a two-dimensional plot of G_{ij} as a function of some parameter and E_F. We utilize four maps per component—three for $\mu_{i=L,C,R}$ variation and one for V_Z variation— giving a total of 16 maps that serve as the input. Figure 6 shows the conductance maps (G_{LL} components) for the reference parameters P_0, highlighting the emergent zero-biased peaks.

The $\mathcal{M}$ metric, utilized to train the PINNAT model, is the following:

$$\begin{aligned} \mathcal{M}(H(P)) &= \frac{p_0}{2} \max\left[0, 2m_0 - \sum_{i>1} m_i\right], \\ m_i &= |\langle \psi_i | M \rangle| e^{-|E_i|/\epsilon}, \\ p_i &= 2 \max\left[0, 4|u_i|^2 |v_i|^2 - \frac{1}{2}\right], \end{aligned} \quad (5)$$

with $\{\psi_i, E_i\}$ being the collection of $H(P)$ eigenpairs. The index i enumerates the subsequent eigenvalues: both m_i and p_i are sorted in ascending order by $|E_i|$. The eigenstate projection m_i on the left (or equivalently right) Majorana mode $M = \gamma_L$ is weighted by its distance from zero

energy $|E_i|$ with some threshold $\epsilon = 0.1$ meV. The second factor, p_i, quantifies electron-hole symmetry: u_i, v_i are the electron-hole components of eigenvectors ψ_i. By the H symmetry, $m_0 = m_1$ and $p_0 = p_1$. For the ideal MZM pair we should have $m_0 = m_1 = 1$ and $p_0 = p_1 = 1$, giving $\mathcal{M} = 1$. The form of the metric, Eq. (5), is designed to discriminate trivial zero modes such as ABSs, whose discrimination from MZMs is described, for example, in Refs. [17,45,46]. In particular, the $-\sum_{i>1} m_i$ term penalizes the presence of trivial zero modes, even when the topological ones are there as well.

Although the metric defined in Eq. (5) contains piecewise operations such as $\max(0, \cdot)$ and sorting by $|E_i|$, gradients required for training are computed using automatic differentiation provided by the AUTOGRAD framework. Such piecewise functions are handled using standard subgradient rules (analogous to rectified linear unit (ReLU) activations), so that they propagate whenever the argument of the $\max(0, \cdot)$ operation is positive. The eigenvalues and eigenvectors are obtained from a differentiable eigensolver applied to the Hamiltonian. Degeneracies and level crossings occur only on sets of measure zero in parameter space and did not cause practical issues during training. However, eigenvector-based gradients may become numerically ill-conditioned near degeneracies or for larger matrices, which can introduce noise in the optimization signal. This potential limitation is discussed further in Appendix D.

For the visualization of the spectra in Figs. 5 and 7, two auxiliary quantities are computed from the normalized Bogoliubov–de Gennes (BdG) eigenvectors ψ_i, utilizing their electron and hole components u_i, v_i. The site occupation of dot n for eigenstate i is defined as

$$O_n^{(i)} = |u_{i,n}|^2 + |v_{i,n}|^2. \tag{6}$$

The quantity referred to as edge occupation is the total weight on the outer dots,

$$O_{\text{edge}}^{(i)} = O_L^{(i)} + O_R^{(i)}. \tag{7}$$

In addition, electron-hole symmetry is defined as the difference between the total electron and hole weights

$$S_{eh}^{(i)} = \sum_n \left(|u_{i,n}|^2 - |v_{i,n}|^2 \right). \tag{8}$$

In our work we adapt typical ViT architecture [40] adjusted to process a 16-channel 50×50 input tensor (representing all generated conductance map components) and returning the vector of the Hamiltonian parameter corrections δP. The architecture hyperparameters such as number of attention heads (four), number of attention blocks (six), hidden size (256), and patch size (one) are tuned to decrease the value of $\mathcal{L}$ after 100 epochs below a given threshold $\tau = -0.7$.

During a single epoch of training we iterate through 10 000 independent H samples generated using Eq. (1), varying parameters $\{\mu_n, t_n, V_Z, \Delta_i\}$ in the range $[0, 1]$ meV and λ_n in the range $[0, \pi]$. The training proceeds until the convergence of $\mathcal{L}$, Eq. (2), is observed—usually between 150 and 200 epochs.

VI. CONCLUSIONS

In this work, we demonstrated that a ViT-based physics-informed neural network model using the $\mathcal{M}$ metric trained on conductance maps can effectively autotune a QD-chain Hamiltonian toward the emergence of MZMs. Our results show that the proposed framework succeeds in correcting both global deviations of system parameters—those that shift all QDs uniformly—and local parameter noise that affects individual QDs independently. In both regimes, the PINNAT model consistently increases the Majorana metric $\mathcal{M}$, thereby restoring the formation of zero-energy edge states where possible.

The proposed framework merges quantum transport simulation with ML-based parameter feedback for topological phase tuning. By integrating experimental observables (conductance maps) as the model input and theoretical descriptors ($\mathcal{M}$ measure) as physics-informed targets for model training, our approach paves the way for autonomous tuning into robust topological regimes in noisy mesoscopic systems.

ACKNOWLEDGMENTS

We would like to thank M. Margańska-Łyżniak for interesting discussions and helpful comments on the manuscript. We gratefully acknowledge the Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational Grant No. PLG/2025/018433.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: ANALYTICAL CONSIDERATIONS FOR THE THREE-QD MODEL

Let us start with the Rashba-Zeeman-BCS Hamiltonian, Eq. (1), written in so-called Nambu spinor representation $\hat{\Psi}_n = (c_{n\uparrow}, c_{n\downarrow}, c_{n\uparrow}^\dagger, c_{n\downarrow}^\dagger)^T$: $H = \frac{1}{2}\sum_{nm} \hat{\Psi}_n^\dagger H_{nm}^{\text{BdG}} \hat{\Psi}_m$, where H^{BdG} is a first-quantized Hamiltonian, also called the BdG Hamiltonian. Its explicit matrix form reads

$$H^{\mathrm{BdG}} = \sum_{n=1}^{N} \begin{pmatrix} -\mu_n + V_Z & 0 & 0 & \Delta_n e^{i\phi_n} \\ 0 & -\mu_n - V_Z & -\Delta_n e^{i\phi_n} & 0 \\ 0 & -\Delta_n e^{-i\phi_n} & \mu_n - V_Z & 0 \\ \Delta_n e^{-i\phi_n} & 0 & 0 & \mu_n + V_Z \end{pmatrix}_{n,n} + \sum_{n=1}^{N-1} t_n \begin{pmatrix} e^{i\boldsymbol{\lambda}_n \cdot \boldsymbol{\sigma}} & 0 \\ 0 & -e^{-i\boldsymbol{\lambda}_n \cdot \boldsymbol{\sigma}^T} \end{pmatrix}_{n,n+1} + \mathrm{H.c.}, \tag{A1}$$

with the on-site (dot) block $(\quad)_{n,n}$ and the off-site hopping (interdot) block $(\quad)_{n,n+1}$ both represented by 4×4 matrices. For a system consisting of three quantum dots ($N = 3$), the full BdG Hamiltonian H^{BdG} is a 12×12 matrix.

Our goal is to simulate the KC model, i.e., to tune the H^{BdG} (parameters) to mimic the KC Hamiltonian form. The first step is that each QD should have a pair of zero energy levels $E_n = 0$. Diagonalizing a (separated) single dot block $(\quad)_{n,n}$ yields the condition for this to happen: $V_Z = \sqrt{\Delta_n^2 + \mu_n^2}$ or equivalently $\Delta_n = \sqrt{V_Z^2 - \mu_n^2}$. When this condition is satisfied, each dot has a pair of fermionic excitations $a_n, a_n^\dagger$, where

$$a_n = \frac{1}{\sqrt{2V_Z}} \left(e^{-\frac{i\phi_n}{2}} \sqrt{V_Z + \mu_n} c_{n\uparrow}^\dagger - e^{\frac{i\phi_n}{2}} \sqrt{V_Z - \mu_n} c_{n\downarrow} \right). \tag{A2}$$

The energy of a_n ($a_n^\dagger$) is zero. Two other modes are separated by the energy $\pm 2V_Z$. If we now assume that the hopping is much smaller than the energy of the excited state, $t \ll 2V_Z$, we may project the Hamiltonian (A1) onto the Hilbert space spanned by a set $\{a_1, a_1^\dagger, a_2, a_2^\dagger, \ldots, a_N, a_N^\dagger\}$. To simplify the system definition, we take $\boldsymbol{\lambda}_n \cdot \boldsymbol{\sigma} = \lambda_n \sigma_y$, which follows from choosing the Rashba vector direction $\rho = \xi = \pi/2$. The resulting projected Hamiltonian is

$$\tilde{H}^{\mathrm{BdG}} = \sum_{n=1}^{N-1} \frac{t_n}{V_Z} \begin{pmatrix} T_n & -D_n \\ D_n & -T_n^* \end{pmatrix}_{n,n+1} + \mathrm{H.c.}, \tag{A3}$$

where $T_n = -[\mu_n \cos(\phi_{n+1} - \phi_n)/2 - iV_Z \sin(\phi_{n+1} - \phi_n)/2] \cos\lambda_n$, and $D_n = -\Delta_n \cos(\phi_{n+1} - \phi_n)/2 \sin\lambda_n$. The goal is to simulate the KC model, which, for example, for $N = 3$ takes the form [in the same Nambu representation $(a_1, a_1^\dagger, a_2, a_2^\dagger, a_3, a_3^\dagger)$ as $\tilde{H}^{\mathrm{BdG}}$]

$$H^{\mathrm{KC}} = \begin{pmatrix} \mu_1 & 0 & t_1' & -\Delta_1' & 0 & 0 \\ 0 & -\mu_1 & \Delta_1' & -t_1' & 0 & 0 \\ t_1' & \Delta_1' & \mu_2 & 0 & t_2' & -\Delta_2' e^{i\Phi} \\ -\Delta_1' & -t_1' & 0 & -\mu_2 & \Delta_2' e^{i\Phi} & -t_2' \\ 0 & 0 & t_2' & \Delta_2' e^{-i\Phi} & \mu_3 & 0 \\ 0 & 0 & -\Delta_2' e^{-i\Phi} & -t_2' & 0 & -\mu_3 \end{pmatrix}. \tag{A4}$$

The KC Hamitonian, Eq. (A4), at the sweet spot $\mu_n = 0$, $t_n' = \Delta_n'$ possesses two MZMs. For the Rashba-Zeeman s-wave Hamiltonian $\tilde{H}^{\mathrm{BdG}}$, Eq. (A3), to simulate H^{KC}, Eq. (A4), at the sweet spot, the following conditions must be satisfied (assuming $\Phi = 0$):

$$\begin{aligned} \phi_{n+1} &= \phi_n, \\ \mu_n \cos\lambda_n &= \Delta_n \sin\lambda_n. \end{aligned} \tag{A5}$$

The latter yields $\lambda_n = \arctan(\mu_n/\Delta_n)$, which, together with $\mu_n = \sqrt{V_Z^2 - \Delta_n^2}$, gives the approximate position of the parameters' sweet spot for the QD Hamiltonian H. Both conditions are also discussed in the adaptive tuning proposal of Fulga *et al.* [13], as well as in the recent experimental realization of short KCs of Bordin *et al.* [54].

APPENDIX B: ANALYZING THE TUNING PROCEDURE FOR HOPPING PARAMETER *t*

Here we present an additional example of an autotuning procedure for the parameter t. In Fig. 7(a), we plot the H eigenspectrum for detuned parameters $t_0 = t_1 = 0.06$ meV, fixed to reduce the $\mathcal{M}$ value at reference $\mu = 0.6$ meV. It can be seen that such a setting of t_0 and t_1 does not make zero modes vanish completely, but instead moves them to a lower μ regime. Simultaneously, the zero-energy

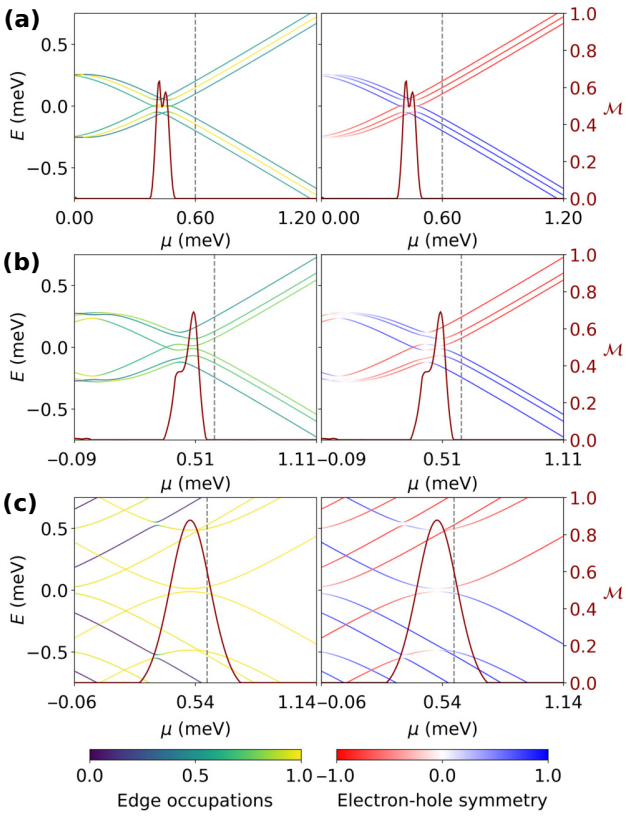


FIG. 7. Three-QD-chain Hamiltonian H as a function of the offset μ added to local potentials: $\mu_n \to \mu_n + \mu - 0.6$ meV (vertical dashed line marks the reference $\mu = 0.6$ meV). Additionally (twin y axis), the $\mathcal{M}$ metric is plotted and eigenvalues are colored (left column) with edge occupation $O^{(i)}_{\text{edge}}$ or (right column) electron-hole symmetry $S^{(i)}_{eh}$, as defined in the Methods section. Plots are presented for (a) reference parameters modified by setting $t_0 = 0.06$ meV and $t_1 = 0.06$ meV; (b) parameters from (a) with physics-informed neural-network-based autotuned $\{\mu_n, t_n, \lambda_n\}$; (c) parameters from (a) with physics-informed neural-network-based autotuned $\{\mu_n, V_Z\}$.

states are far less gapped from the rest of the spectrum than for the reference $t = 0.25$ meV setup.

In the next step we try to tune the parameters $\{\mu_n, t_n, \lambda_n\}$ using the NN model, as presented in Fig. 7(b). We observe that PINNAT adjusts μ_n, reducing the average $\langle\mu_n\rangle = 0.51$, and thereby shifting the energy offset and restoring the previously observed zero modes. Additionally, further setting $t_0 = 0.11$ meV, $t_1 = 0.1$ meV, $\lambda_0 = 0.38\pi$, and $\lambda_1 = 0.26\pi$ parameters allows the NN to increase the $\mathcal{M}$ value and, as a result, increase the topological gap—the separation between zero modes and the bulk spectrum.

Finally, we can adjust the detuning in t_0 and t_1 using the second version of the PINNAT model that sets $\{\mu_n, V_Z\}$ parameters—Fig. 7(c). As before, μ_n values are shifted so that their average is decreased: $\langle\mu_n\rangle = 0.54$ meV and zero modes are seen to emerge. Moreover, the NN model proposes to set $V_Z = 0.24$ meV, yielding a higher value of $\mathcal{M}$ and an even larger topological gap than that in Fig. 7(b), where the model tunes t_n and λ_n.

Nevertheless, similar to the example presented in the main text, we can observe that both models are capable of correcting given parameters to produce H with emergent zero-energy states, which are likely to maintain MZM properties.

APPENDIX C: COMPARISON OF TUNED PARAMETERS WITH MAJORANA SWEET-SPOT CONDITIONS

To facilitate comparison with previously proposed parameter regimes for Majorana zero modes in few-site Kitaev chains, Table I lists representative sets of PINNAT parameters together with the reference values used in the simulations. In addition to the raw parameters, dimensionless ratios are reported that quantify proximity to the sweet-spot relations discussed in Appendix A.

The PINNAT configurations consistently fall within the expected Majorana-supporting regime. In particular, the Zeeman energy exceeds the interdot couplings in all cases, as indicated by $V_Z/\max_n t_n > 1$, often by a significant margin. This places the system in a regime where Zeeman splitting dominates over interdot hybridization, which is favorable for the emergence of topological phases.

Furthermore, the ratio $V_Z/\sqrt{\Delta^2 + \langle\mu_n\rangle^2}$ remains of order unity, indicating that the condition $V_Z = \sqrt{\Delta^2 + \mu_n^2}$ for having a single fermionic pair at zero energy per dot is approximately satisfied. Finally, the relation between the

TABLE I. Comparison of reference and PINNAT parameter sets for the three-dot chain. Energies are given in meV. In the presented sets Δ_n does not change with n: $\Delta_n = \Delta$.

Configuration	$\langle\mu_n\rangle$	V_Z	Δ	t_1	t_2	λ_1	λ_2	$\frac{V_Z}{\max_n t_n}$	$\frac{V_Z}{\sqrt{\Delta^2 + \langle\mu_n\rangle^2}}$	$\frac{\langle\lambda_n\rangle}{\arctan\left(\frac{\langle\mu_n\rangle}{\Delta}\right)}$
Reference parameters	0.60	0.50	0.25	0.25	0.25	0.27π	0.27π	2.0	0.77	0.72
PINNAT $\{\mu_n, t_n, \lambda_n\}$ [Fig. 5(c)]	0.54	0.50	0.25	0.20	0.25	0.67π	0.41π	2.0	0.83	1.49
PINNAT $\{\mu_n, V_Z\}$ [Fig. 5(d)]	0.58	0.31	0.25	0.25	0.25	0.62π	0.41π	1.24	0.49	1.39
PINNAT $\{\mu_n, t_n, \lambda_n\}$ [Fig. 7(b)]	0.51	0.50	0.25	0.11	0.10	0.38π	0.26π	4.55	0.88	0.90
PINNAT $\{\mu_n, V_Z\}$ [Fig. 7(c)]	0.54	0.24	0.25	0.06	0.06	0.27π	0.27π	4.0	0.4	0.75

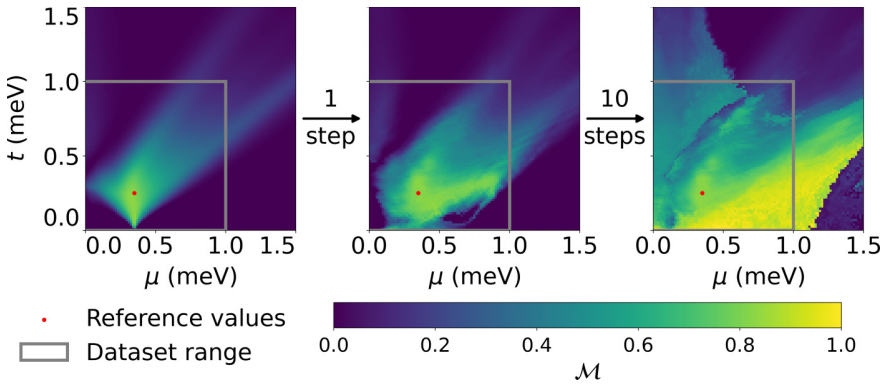


FIG. 8. Iterative autotuning procedure for the model adjusting $\{\mu_n, V_Z\}$ for a system composed of seven quantum dots. In the left plot, we present the $\mathcal{M}$ map before tuning. Then, in the center, parameters are tuned with a single step of neural network (NN) corrections. Finally, in the right plot, there is a map plotted for parameters tuned within ten subsequent steps of NN corrections.

Rashba spin-orbit length and the on-site potential, captured by $\langle\lambda_n\rangle/\arctan\frac{\langle\mu_n\rangle}{\Delta}$, also remains of order unity across the considered configurations, indicating consistency with the analytical condition derived in Appendix A.

Overall, while the analytical sweet-spot conditions are not exactly fulfilled, the PINNAT parameters systematically lie in their approximate regime of validity, demonstrating that the autotuning procedure identifies physically consistent Majorana-supporting configurations.

APPENDIX D: SCALING THE SYSTEM SIZE

While analyzing results for a quantum chain composed of only three QDs proves the conceptual validity of the proposed approach, the natural question arises about whether it is possible to repeat a similar procedure for larger systems. For this reason, we manually tuned the Rashba-Zeeman-BCS Hamiltonian, Eq. (1), to realize MZMs for $N = 7$ quantum dots. Such reference parameters include $\mu = 0.35$ meV, $t = 0.25$ meV, $\lambda = 0.27\pi$, $\rho = \xi = \pi/2$, $V_Z = 0.5$ meV, and $\Delta = 0.35$ meV. In contrast to three-QD systems, to train the PINNAT model for the seven-QD chain, we could not use entirely random parameter sampling due to a much larger parameter space and far smaller number of samples with $\mathcal{M} > 0$. Instead, we generated 10 000 samples, randomly shifting the parameters from the reference values with the restriction that the $\mathcal{M}$ value is larger than $\tau = 0.7$. Subsequently, in each training epoch, the samples were independently randomized to decrease the value of $\mathcal{M}$. For half of the samples, $\mathcal{M}$ was reduced to 0, and for the remaining 5000 samples, $\mathcal{M}$ was decreased to the interval $(0, \tau]$. Next, a set of 16 conductance maps was determined: to reduce the conductance tensor size we used only the G_{LL} and G_{RR} pair as a function of eight parameters: μ_n, for n iterating over each dot, and V_Z (16 maps in total). Finally, the PINNAT model was trained to increase the value of $\mathcal{M}$ by adjusting the $\{\mu_n, V_Z\}$ parameters.

The results of training the model are shown in Fig. 8, where the $\mathcal{M}$ metric is plotted for the previously introduced ten-step autotuning procedure. In contrast to the three-QD system, Fig. 4(b), the PINNAT model trained to produce corrections for the seven-QD chain tends to tune the system to a lesser extent. However, the increase in the $\mathcal{M}$ value is still visible even after a single step of corrections. Moreover, within the ten steps, the PINNAT model demonstrates its ability to correct parameter detuning in the whole training distribution and also shows the potential of generalization beyond the training regime.

While these results prove that it is possible to scale the PINNAT model with an autotuning approach to larger systems, several limitations to the scaling should be highlighted. In particular, the dataset construction requires filtering Hamiltonians that exhibit emerging Majorana modes, while the training procedure involves repeated direct diagonalization within the loss evaluation (via the $\mathcal{M}$ metric) and the computation of conductance maps. As a result, both the data generation time and the training time scale approximately exponentially with the number of quantum dots in the system. This behavior is illustrated in Table II, which summarizes the observed computational costs for systems of different sizes. While small systems can be handled almost instantaneously, the computational cost increases rapidly. For example, generating the dataset

TABLE II. Observed scaling of data generation and model training times with system size.

Number of dots	Data generation time (h)	Training time per epoch (h)
3	online	0.28
5	2	0.42
7	5	1.17
10	17	5.50

for a ten-dot system requires approximately 17 h, and a single training epoch takes around 5.5 h. Therefore, extending the method to larger systems will likely require additional optimizations, such as limiting the number of input conductance maps, reducing the number of tuned parameters, narrowing the parameter search ranges, or constructing a more time-effective target function.

Additionally, although the eigendecomposition used in the loss evaluation is differentiable almost everywhere, the gradients associated with eigenvectors can become numerically ill-conditioned near degeneracies or level crossings. This effect may introduce noise in the backpropagated gradients and can potentially limit the scalability of eigenvalue-based learning procedures. Various approaches have been proposed to mitigate this issue in differentiable eigensolvers [55,56]. However, in our experiments, we observed stable and well-behaved training for systems up to ten quantum dots (corresponding to 40×40 Hamiltonian matrices), with no noticeable degradation of the optimization process.

[1] J. Alicea, New directions in the pursuit of Majorana fermions in solid state systems, Rep. Prog. Phys. **75**, 076501 (2012).

[2] C. W. J. Beenakker, Search for Majorana fermions in superconductors, Annu. Rev. Condens. Matter Phys. **4**, 113 (2013).

[3] J. D. Sau, R. M. Lutchyn, S. Tewari, and S. Das Sarma, Generic new platform for topological quantum computation using semiconductor heterostructures, Phys. Rev. Lett. **104**, 040502 (2010).

[4] Y. Oreg, G. Refael, and F. von Oppen, Helical liquids and Majorana bound states in quantum wires, Phys. Rev. Lett. **105**, 177002 (2010).

[5] M. Leijnse and K. Flensberg, Introduction to topological superconductivity and Majorana fermions, Semicond. Sci. Technol. **27**, 124003 (2012).

[6] R. M. Lutchyn, E. P. A. M. Bakkers, L. P. Kouwenhoven, P. Krogstrup, C. M. Marcus, and Y. Oreg, Majorana zero modes in superconductor–semiconductor heterostructures, Nat. Rev. Mater. **3**, 52 (2018).

[7] N. Leumer, M. Grifoni, B. Muralidharan, and M. Marganska, Linear and nonlinear transport across a finite Kitaev chain: An exact analytical study, Phys. Rev. B **103**, 165432 (2021).

[8] A. Y. Kitaev, Unpaired Majorana fermions in quantum wires, Phys. Usp. **44**, 131 (2001),translation from: Usp. Fiz. Nauk 171 (2001) 131–136, supplement No. 10.

[9] M. Valentini, F. Peñaranda, A. Hofmann, M. Brauns, R. Hauschild, P. Krogstrup, P. San-Jose, E. Prada, R. Aguado, and G. Katsaros, Nontopological zero-bias peaks in full-shell nanowires induced by flux-tunable Andreev states, Science **373**, 82 (2021).

[10] S. Sasaki, S. D. Franceschi, J. M. Elzerman, W. G. van der Wiel, M. Eto, S. Tarucha, and L. P. Kouwenhoven, Kondo effect in an integer-spin quantum dot, Nature **405**, 764 (2000).

[11] D. I. Pikulin, J. P. Dahlhaus, M. Wimmer, H. Schomerus, and C. W. J. Beenakker, A zero-voltage conductance peak from weak antilocalization in a Majorana nanowire, New J. Phys. **14**, 125011 (2012).

[12] G. Kells, D. Meidan, and P. W. Brouwer, Low-energy subgap states in multichannel p-wave superconducting wires, Phys. Rev. B **85**, 060507(R) (2012).

[13] I. C. Fulga, A. Haim, A. R. Akhmerov, and Y. Oreg, Adaptive tuning of Majorana fermions in a quantum dot chain, New J. Phys. **15**, 045020 (2013).

[14] M. Leijnse and K. Flensberg, Parity qubits and poor man's Majorana bound states in double quantum dots, Phys. Rev. B **86**, 134528 (2012).

[15] J. D. Sau and S. D. Sarma, Realizing a robust practical Majorana chain in a quantum-dot-superconductor linear array, Nat. Commun. **3**, 964 (2012).

[16] T. Dvir, G. Wang, N. van Loo, C.-X. Liu, G. P. Mazur, A. Bordin, S. L. D. ten Haaf, J.-Y. Wang, D. van Driel, F. Zatelli, X. Li, F. K. Malinowski, S. Gazibegovic, G. Badawy, E. P. A. M. Bakkers, M. Wimmer, and L. P. Kouwenhoven, Realization of a minimal Kitaev chain in coupled quantum dots, Nature **614**, 445 (2023).

[17] A. Bordin, C.-X. Liu, T. Dvir, F. Zatelli, S. L. D. ten Haaf, D. van Driel, G. Wang, N. van Loo, Y. Zhang, J. C. Wolff, T. V. Caekenberghe, G. Badawy, S. Gazibegovic, E. P. A. M. Bakkers, M. Wimmer, L. P. Kouwenhoven, and G. P. Mazur, Enhanced Majorana stability in a three-site Kitaev chain, Nat. Nanotechnol. **20**, 726 (2025).

[18] M. Mohseni, H. Allami, D. Miravet, D. J. Gayowsky, M. Korkusinski, and P. Hawrylak, Majorana excitons in a Kitaev chain of semiconductor quantum dots in a nanowire, Nanomaterials **13**, 2293 (2023).

[19] M. Ezawa, Even-odd effect on robustness of Majorana edge states in short Kitaev chains, Phys. Rev. B **109**, L161404 (2024).

[20] F. Borsoi, N. W. Hendrickx, V. John, M. Meyer, S. Motz, F. van Riggelen, A. Sammak, S. L. de Snoo, G. Scappucci, and M. Veldhorst, Shared control of a 16 semiconductor quantum dot crossbar array, Nat. Nanotechnol. **19**, 21 (2024).

[21] A. R. Mills, D. M. Zajac, M. J. Gullans, F. J. Schupp, T. M. Hazard, and J. R. Petta, Shuttling a single charge across a one-dimensional array of silicon quantum dots, Nat. Commun. **10**, 1063 (2019).

[22] A. Shandilya, S. Kapila, R. Krishnan, B. Weber, and B. Muralidharan, Unified simulation framework for experimentally-observed spin-valley locking in MoS_2 quantum dots: Implications for qubit applications, ACS Appl. Nano Mater. **8**, 14949 (2025).

[23] J. Darulová, S. J. Pauka, N. Wiebe, K. W. Chan, G. C. Gardener, M. J. Manfra, M. C. Cassidy, and M. Troyer, Autonomous tuning and charge-state detection of gate-defined quantum dots, Phys. Rev. Appl. **13**, 054005 (2020).

[24] L. Lunczer, P. Leubner, M. Endres, V. L. Müller, C. Brüne, H. Buhmann, and L. W. Molenkamp, Approaching quantization in macroscopic quantum spin Hall devices through gate training, Phys. Rev. Lett. **123**, 047701 (2019).

[25] J. P. Zwolak and J. M. Taylor, Colloquium: Advances in automation of quantum dot devices control, Rev. Mod. Phys. **95**, 011006 (2023).

[26] M.-A. Roux *et al.*, Rapid autotuning of a SiGe quantum dot into the single-electron regime with machine learning and RF-reflectometry FPGA-based measurements, IEEE Trans. Quantum Eng. **7**, 3101707 (2026).

[27] M. P. R. Losert, D. Denora, B. van Straaten, M. Chan, S. D. Oosterhout, L. Stehouwer, G. Scappucci, M. Veldhorst, and J. P. Zwolak, Automated electrostatic characterization of quantum dot devices in single- and bilayer heterostructures, arXiv:2601.00067 [cond-mat.mes-hall].

[28] A. Donarini and M. Grifoni, *Quantum Transport in Interacting Nanojunctions* (Springer Nature Switzerland AG, 2024).

[29] G. E. Blonder, M. Tinkham, and T. M. Klapwijk, Transition from metallic to tunneling regimes in superconducting microconstrictions: Excess current, charge imbalance, and supercurrent conversion, Phys. Rev. B **25**, 4515 (1982).

[30] R. Koch, D. van Driel, A. Bordin, J. L. Lado, and E. Greplova, Adversarial Hamiltonian learning of quantum dots in a minimal Kitaev chain, Phys. Rev. Appl. **20**, 044081 (2023).

[31] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, Experimental quantum Hamiltonian learning, Nat. Phys. **13**, 551 (2017).

[32] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nat. Rev. Phys. **5**, 141 (2023).

[33] M. Thamm and B. Rosenow, Conductance based machine learning of optimal gate voltages for disordered Majorana wires, Phys. Rev. B **109**, 045132 (2024).

[34] J. R. Taylor, J. D. Sau, and S. Das Sarma, Machine learning the disorder landscape of Majorana nanowires, Phys. Rev. Lett. **132**, 206602 (2024).

[35] J. R. Taylor and S. Das Sarma, Neural network based deep learning analysis of semiconductor quantum dot qubits for automated control, Phys. Rev. B **111**, 035301 (2025).

[36] D. van Driel *et al.*, Cross-platform autonomous control of minimal Kitaev chains, arXiv:2405.04596 [cond-mat.mes-hall].

[37] J. R. Taylor and S. Das Sarma, Mitigating disorder and optimizing topological indicators with vision-transformer-based neural networks in Majorana nanowires, Phys. Rev. B **112**, L041110 (2025).

[38] M. Raissi, P. Perdikaris, and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. **378**, 686 (2019).

[39] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, Neural operator: Learning maps between function spaces with applications to PDEs, J. Mach. Learn. Res. **24**, 1 (2023).

[40] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16 × 16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929 (2020).

[41] D. Sticlet, C. Bena, and P. Simon, Spin and Majorana polarization in topological superconducting wires, Phys. Rev. Lett. **108**, 096802 (2012).

[42] M. M. Maśka and T. Domański, Polarization of the Majorana quasiparticles in the Rashba chain, Sci. Rep. **7**, 16193 (2017).

[43] P. Szumniak, D. Chevallier, D. Loss, and J. Klinovaja, Spin and charge signatures of topological superconductivity in Rashba nanowires, Phys. Rev. B **96**, 041401(R) (2017).

[44] T. Stanescu, *Introduction to Topological Quantum Matter & Quantum Computation* (CRC Press, Boca Raton, Florida, 2024).

[45] C.-X. Liu, J. D. Sau, T. D. Stanescu, and S. Das Sarma, Andreev bound states versus Majorana bound states in quantum dot-nanowire-superconductor hybrid structures: Trivial versus topological zero-bias conductance peaks, Phys. Rev. B **96**, 075161 (2017).

[46] E. Prada, P. San-Jose, M. W. A. de Moor, A. Geresdi, E. J. H. Lee, J. Klinovaja, D. Loss, J. Nygård, R. Aguado, and L. P. Kouwenhoven, From Andreev to Majorana bound states in hybrid superconductor–semiconductor nanowires, Nat. Rev. Phys. **2**, 575 (2020).

[47] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, arXiv:2104.13478.

[48] J. Pawłowski and M. Krawczyk, Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning, Phys. Rev. Appl. **22**, 014068 (2024).

[49] M. Krawczyk, J. Pawłowski, M. M. Maśka, and K. Roszak, Data-driven criteria for quantum correlations, Phys. Rev. A **109**, 022405 (2024).

[50] A. Norambuena, M. Mattheakis, F. J. González, and R. Coto, Physics-informed neural networks for quantum control, Phys. Rev. Lett. **132**, 010801 (2024).

[51] M. Krawczyk, P. Baláž, K. Roszak, and J. Pawłowski, Learning quantum tomography from incomplete measurements, arXiv:2506.19428 [quant-ph].

[52] J. Pawłowski and M. Krawczyk, Learning Hamiltonians for solid-state quantum simulators, arXiv:2603.02889 [cond-mat.mes-hall].

[53] T. J. Christiansen and M. Zworski, A mathematical formulation of the Mahaux–Weidenmüller formula for the scattering matrix, J. Phys. A: Math. Theor. **42**, 415202 (2009).

[54] A. Bordin, F. J. Bennebroek Evertsz', B. Roovers, J. D. Torres Luna, W. D. Huisman, F. Zatelli, G. P. Mazur, S. L. D. ten Haaf, G. Badawy, E. P. A. M. Bakkers, C.-X. Liu, R. Seoane Souto, N. van Loo, and L. P. Kouwenhoven, Probing Majorana localization of a phase-controlled three-site Kitaev chain with an additional quantum dot, Nat. Commun. **17**, 2313 (2026).

[55] W. Wang, Z. Dang, Y. Hu, P. Fua, and M. Salzmann, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2019).

[56] T. Li, Z. Shi, J. Zhao, and M. Lin, in *Advances in Neural Information Processing Systems*, edited by A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (Curran Associates, Inc., Red Hook, NY, 2024), Vol. 37, pp. 138809–138835.

Unpublished Results

Learning quantum tomography from incomplete measurements

Mateusz Krawczyk,[1] Pavel Baláž,[2] Katarzyna Roszak,[2] and Jarosław Pawłowski[1,*]

[1] *Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*

[2] *FZU - Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 00 Prague, Czech Republic*

We revisit quantum tomography in an informationally incomplete scenario and propose improved state reconstruction methods using deep neural networks. In the first approach, the trained network predicts an optimal linear or quadratic reconstructor with coefficients depending only on the collection of (already taken) measurement operators. This effectively refines the undercomplete tomographic reconstructor based on pseudoinverse operation. The second, based on an LSTM recurrent network performs state reconstruction sequentially. It can also optimize the measurement sequence, which suggests a no-free-lunch theorem for tomography: by narrowing the state space, we gain the possibility of more efficient tomography by learning the optimal sequence of measurements. Numerical experiments for a 2-qubit system show that both methods outperform standard maximum likelihood estimation and also scale to larger 3- and 4-qubit systems. Our results demonstrate that neural networks can effectively learn the underlying geometry of multi-qubit states and use it for their reconstruction.

arXiv:2506.19428v5 [quant-ph] 9 Sep 2026

I. INTRODUCTION

Quantum state tomography (QST) [1–3] aims to determine the full state of a quantum system via a series of quantum measurements. Although the methods for complete state reconstruction are well known [2, 4], QST for incomplete data does not have a simple straightforward solution as the positivity criterion imposed on the state estimators makes analytical studies of such problems extremely difficult [5]. Approximate statistical-based methods for QST with an undercomplete or noisy measurement set are the maximum entropy (MaxEnt) principle [5–7], maximum likelihood estimation (MLE) [5, 7–9], or least-squares inversion [10].

Although QST scales exponentially in the system size, approximate methods that scale much better (polynomially) already appear using either the singular value thresholding method [11], or neural networks (NNs) [12]. In the case of permutationally invariant states (which occur in many practical situations, such as Dicke states or spin-squeezed states), exact tomography methods exist [13] that scale quadratically with system size [14]. On the other hand, classical shadow tomography [15, 16] utilizes the fact that predicting important properties of a many-body state (without necessarily fully characterizing the quantum state) requires only a polynomially-growing number of measurements [17]. Restricted Boltzmann machine (RBM) [18, 19], multilayer perceptron (MLP) [20, 21], convolutional NN [22], recurrent NN [12, 23], or conditional generative adversarial NN (CGAN) [24] fed by subsequent measurement snapshots that form an *informationally complete* system can be used to reconstruct certain classes of states, but most models (beside that of Ref. [23], or Ref. [25], but using integer programming) do not optimize the sequence of measurements. NN-aided tomography can also converge much faster [21, 24] than other QST techniques, while attention-based NNs can be used to effectively denoise measurement data during QST [26]. Deep NNs have also shown their usability in recognizing quantum entanglement [27–31], or discord [32].

In this work, we focus on a different scenario: we examine various approaches to QST of general mixed states using *undercomplete* measurement data, under the assumption that a single measurement expectation value is known exactly, and show that NNs outperform standard estimators such as MLE on a diverse test set (S_{test}) of 2- to 4-qubit random mixed states. We explicitly demonstrate that neural estimators yield optimal reconstruction for the 1-qubit case and identify the best measurement sequence for 2-qubit X states.

In general, the expectation value of a projection-valued measurement (PVM) is given by the Born rule [33] $m = \mathrm{Tr}(\Pi\rho)$, where Π is the projection operator and ρ is the density matrix. Since Π is a projector, this expectation value is also the probability of obtaining the outcome associated with Π. Experimentally, it is estimated from a finite number N_c of copies as $\hat{m} = \frac{n_c}{N_c}$, where n_c is a number of coincidence counts in a measurement setup. For N qubits, the expectation value of joint measurements required for QST can be expressed as

$$m_{i_1,..,i_N} = \mathrm{Tr}((\mu_{i_1} \otimes \ldots \otimes \mu_{i_N})\,\rho) = \mathrm{Tr}(\Pi_{i_1,..,i_N}\rho)\,, \quad (1)$$

where $\mu_{i_k} = |\psi_{i_k}\rangle\langle\psi_{i_k}|$ are single-qubit projectors acting on the k-th qubit state. In the special case of rank-one projectors considered here, it can equivalently be interpreted as the corresponding measurement probability. In the following, unless stated otherwise, we use the standard set of operators employed in photonic tomography by James *et al.* [2] as a fixed catalog of single-qubit rank-

* jaroslaw.pawlowski@pwr.edu.pl

one projectors,

$$|\psi_i\rangle \in \left\{ |0\rangle,\ |1\rangle,\ \frac{|0\rangle + |1\rangle}{\sqrt{2}},\ \frac{|0\rangle - i|1\rangle}{\sqrt{2}} \right\}. \tag{2}$$

For an N-qubit system, this defines the product-projector catalog

$$\mathcal{C}_N = \left\{ \Pi_\nu \equiv \Pi_{i_1,\ldots,i_N} = \bigotimes_{k=1}^{N} |\psi_{i_k}\rangle\langle\psi_{i_k}| : i_k \in \{0,1,2,3\} \right\}, \tag{3}$$

$|\mathcal{C}_N| = 4^N$. The index ν is therefore a shorthand for the multi-index $(i_1, \ldots, i_N)$.

It is important to stress that the set in Eq. (2) is not used as a single four-outcome PVM. In particular, the projectors onto $|0\rangle$ and $|1\rangle$ are complementary outcomes of a computational-basis measurement, and their probabilities are not independent once the trace of ρ is fixed. Nevertheless, we keep both projectors in the catalog for two practical reasons. First, the resulting product catalog is the conventional James *et al.* tomographic frame and leads to a simple linear reconstruction matrix in the fully measured case. Second, in the present work, the central object is not to use a complete set of PVMs but an incomplete list of expectation values $m_\nu = \mathrm{Tr}(\Pi_\nu \rho)$ chosen from a fixed, experimentally transparent catalog. Thus, the apparent redundancy of some probabilities is intentionally retained in the initial setting. When measurements are selected adaptively, the *selector* model can learn to avoid redundant choices.

The projectors in Eq. (3) are Hermitian, idempotent, and positive semidefinite, but they do not form a PVM as they are not trace-orthonormal [34]. This does not prevent them from forming an informationally complete tomographic frame (the completeness of PVM is not the same as the completeness of QST itself). Indeed, the Gramian matrix

$$G_{\nu\nu'} = \mathrm{Tr}(\Pi_\nu \Pi_{\nu'}) \tag{4}$$

is nonsingular, equivalently

$$\det G = \det[\mathrm{Tr}(\mu_i \mu_j)]^{N4^{N-1}} = \left(\frac{1}{4}\right)^{N4^{N-1}} \neq 0, \tag{5}$$

hence operators Π_ν form a linearly independent set of dimension 4^N and give a complete measurement basis [5].

After collecting a set of expectation values m_ν one can tomographically reconstruct the state via the Γ_μ-matrix basis [2] (Einstein summation convention assumed):

$$\rho = \Gamma_\mu \left(\langle\psi_\nu|\Gamma_\mu|\psi_\nu\rangle \right)^{-1}_{\mu\nu} m_\nu, \tag{6}$$

where the tomographic state $|\psi_\nu\rangle = |\psi_{i_1}\rangle \otimes \ldots \otimes |\psi_{i_N}\rangle$ is composed of states defining subsequent projectors, and $\Gamma_\mu = \bigotimes_{k=1}^{N} \sigma_{j_k}$ is the Kronecker product composed out of the Pauli matrices σ_{j_k}, with $\mu \equiv j_1, .., j_N$ and $j_k = 0, 1, 2, 3$ indexing the Pauli matrix of qubit k. Note that σ_0, being the identity matrix, is included to complete the Pauli matrices. Wherever possible, we adopt the Einstein summation convention. For instance, Eqs. 6 and 7 involves two contractions: with respect to the indices ν and μ. The matrix $B_{\nu\mu} = \langle\psi_\nu|\Gamma_\mu|\psi_\nu\rangle$ relates subsequent measurement expectations m_ν with the state representation in the Γ_μ basis. The choice of measurement basis ($\mu_{i_k} = |\psi_{i_k}\rangle\langle\psi_{i_k}|$) required to obtain full information about a quantum state ρ is not unique, but $B_{\nu\mu}$ must be invertible. $B_{\nu\mu} = \mathrm{Tr}((\bigotimes_{k=1}^{N} \mu_{i_k})(\bigotimes_{k=1}^{N} \sigma_{i_k})) = \bigotimes_{k=1}^{N} \mathrm{Tr}(\mu_{i_k}\sigma_{j_k}) = \mathrm{Tr}(\mu_i\sigma_j)^{\otimes N}$, and $\mathrm{Det}(B_{\nu\mu}) = \mathrm{Det}(\mathrm{Tr}(\mu_i\sigma_j))^{N4^{N-1}}$ is nonzero because $\mathrm{Det}(\mathrm{Tr}(\mu_i\sigma_j)) = -2 \neq 0$ in our case, so the condition is met. Its inverse is denoted by $(B_{\nu\mu})^{-1} = B^{-1}_{\mu\nu}$ (note the index swap).

QST is extremely resource-intensive, so it is natural to ask what happens if we do not have a complete set of measurements or if we have measurements with a significant uncertainty. Here we study the undercomplete regime in which fewer than a complete set of tomographic measurement expectation values is available [5]. Operationally, this means that only M values $m_\nu = \mathrm{Tr}(\Pi_\nu \rho)$ are provided to the reconstruction algorithm, with $M < 4^N$, or if trace normalization is accounted for, fewer than the $4^N - 1$ independent parameters required for a generic N-qubit density matrix. In the numerical results presented in this work, these values are evaluated exactly from the sampled target state ρ. Thus, the reported benchmarks correspond to the infinite-shot limit and isolate the error caused by informational incompleteness and by the reconstruction map itself. In an experiment, the same algorithms would take as input empirical estimates $\hat{m}_\nu = n_\nu / N_c$ instead of the exact expectation values m_ν.

When the number of measurement operators M is smaller than the dimension of the complete set spanned by the 4^N Γ_μ matrices in the linear QST, $B_{\nu\mu}$ with dimension $M \times 4^N$ is no longer a square matrix, and becomes non-invertible. The most natural way to keep the formula of Eq. (6) operational is to use the pseudoinverse (Moore–Penrose inverse) [35] operation: $(.)^+$. Since measurement expectation values are linearly independent, so will the rows of matrix $B_{\nu\mu}$ be, hence the formula for pseudoinverse: $B^+ = B^\dagger (B B^\dagger)^{-1}$. Then the reconstruction formula is (Einstein summation assumed):

$$\rho \simeq \Gamma_\mu B^+_{\mu\nu} m_\nu. \tag{7}$$

Using the pseudoinverse as a QST estimator was also discussed in Ref. [21]. In the results section we show that pseudoinverse-based reconstruction does not give the best possible solution, therefore, we aim to construct more accurate reconstruction schemes based on NN regressors.

It is important to emphasize that Eq. (7) defines a raw linear reconstruction map. In the informationally incomplete regime this map is not constrained to return a physical density matrix: the reconstructed operator may fail to have unit trace and may also fail to be positive semidefinite. In the present work we do not post-process

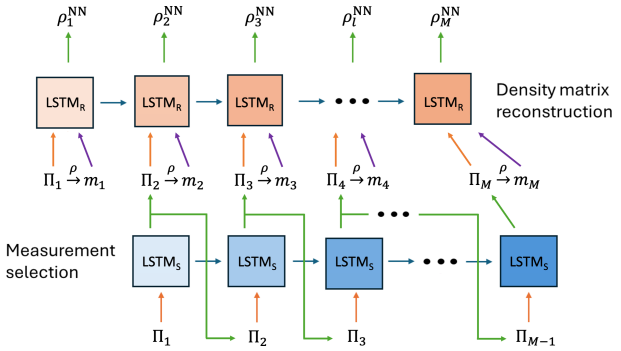


FIG. 1. Scheme for the neural-based measurements selection and reconstruction network. Having a previous selection Π_{l-1}, the goal for the recurrent autoregressive $\mathrm{LSTM_S}$ unit is to propose the next measurement operator Π_l, which is used to calculate the value m_l. Simultaneously, a pair (Π_l, m_l) is used as an input to $\mathrm{LSTM_R}$ cell, which outputs increasingly better reconstructions ρ_l, where l denotes the number of already performed measurements.

such reconstructions by trace normalization or eigenvalue clipping. This choice is deliberate: our goal is to compare the reconstruction maps as estimators of the density operator, without adding an external correction step that could artificially improve the performance of a given method.

II. NEURAL MODELS

In the first attempt, we train a NN just to correct the pseudoinverse $B^{+}_{\mu\nu}$ matrix by adding some correction term $b^{\mathrm{NN}}_{\mu\nu}$ in order to get the smallest possible reconstruction error averaged over the training set:

$$\rho^{\mathrm{NN}}_{\mathrm{corr}}(\mathcal{M}) = \Gamma_\mu\left(\left(B^{+}_{\mu\nu} + b^{\mathrm{NN}}_{\mu\nu}(\mathcal{M})\right) m_\nu + c^{\mathrm{NN}}_{\mu}(\mathcal{M})\right), \quad (8)$$

where an incomplete measurement set $\mathcal{M} = \{(\Pi_\nu)^M_{\nu=1}, (m_\nu)^M_{\nu=1}\}$ is a subset sampled without replacement from the catalog $\mathcal{C}_N$ in Eq. (3); while c^{NN}_μ is added to enable the model to predict normalized ρ^{NN}, irrespectively of the m_ν values. Thus, for each value of M, the measurement subset is random, but contains no repeated projectors. The network input, therefore, contains both the selected projectors and, depending on the model variant, the corresponding measurement expectation values. For each M, we train and test the network for 10 independently sampled subsets and report the average reconstruction performance. This procedure is used to characterize the average behavior of the reconstruction map for undercomplete data, rather than the performance for one specially chosen measurement sequence. We call this model the *corrector* network, and its details are discussed in the Appendix A.

We also introduced a modified corrector version $\bar{\rho}^{\mathrm{NN}}_{\mathrm{corr}}(\Pi)$ that predicts coefficients using only the information about measurement operators (that do not depend on measurement expectation values) – defined in the same way as in Eq. 8 but now $\mathrm{NN}_{\mathrm{corr}}(\Pi) = (b^{\mathrm{NN}}_{\mu\nu}(\Pi), c^{\mathrm{NN}}_{\mu}(\Pi))$, where we abbreviated $\Pi \equiv (\Pi_\nu)^M_{\nu=1}$. One can also imagine improved version of the corrector NN, i.e. one containing a *quadratic* term:

$$\bar{\rho}^{\mathrm{NN}}_{\mathrm{M^2\text{-}corr}}(\Pi) = \bar{\rho}^{\mathrm{NN}}_{\mathrm{corr}}(\Pi) + \Gamma_\mu S^{\mathrm{NN}}_{\mu\nu\nu'}(\Pi)\, m_\nu m_{\nu'}, \quad (9)$$

with the network $S^{\mathrm{NN}}_{\mu\nu\nu'}(\Pi)$ architecture designed in a way that the output tensor $S_{\mu\nu\nu'}$ is symmetric – invariant under any measurements (ν and ν') permutation.

NNs may become useful not only for reconstructing the density matrix from the reduced number of measurement expectations, but also for selecting the most important measurements [36]. The general idea is to use two separate long short-term memory (LSTM) networks presented in Fig. 1, where the first one, $\mathrm{LSTM_S}$, chooses the order of the measurements, and the second, $\mathrm{LSTM_R}$, exploits the information contained in the measurements to reconstruct the density matrix effectively. At step l, the selected by LSTM_S pair (Π_l, m_l) is passed to LSTM_R, which returns an intermediate reconstruction ρ^{NN}_l. The reconstruction error of ρ^{NN}_l is then used in the training loss. In this way, LSTM_S is trained indirectly through the quality of the reconstruction produced by LSTM_R. More details about the *selector* model can be found in the Appendix A.

We compare three sequential scenarios. In the first one, called LSTM with *random basis sequence*, only the reconstruction network LSTM_R is used: the projectors are chosen randomly without replacement from $\mathcal{C}_N$, and LSTM_R reconstructs the state from the resulting sequence. This provides a baseline showing the effect of recurrent reconstruction without learned measurement selection. In the second scenario, called LSTM with *predefined* (James *et al.*) *basis*, LSTM_S chooses projectors from the discrete catalog $\mathcal{C}_N$, again without repetitions. In the third scenario, called LSTM with *adjusted basis*, LSTM_S is not restricted to the discrete catalog and instead outputs a product projector directly. We refer to these projectors as adjusted measurement operators: they are still valid product projectors, but their local measurement directions need not coincide with the catalog in Eq. (3).

The overall numerical procedure is as follows. First, training and test sets of random density matrices, S_{train} and S_{test}, are generated as described in Appendix B. For the corrector networks and a fixed value of M, we sample incomplete measurement sets $\mathcal{M} \subset \mathcal{C}_N$ without replacement, compute the exact expectation values $m_\nu = \mathrm{Tr}(\Pi_\nu \rho)$, and train the network parameters by minimizing the mean L_2 reconstruction error over S_{train}. For the sequential LSTM models, the same reconstruction loss is accumulated over all steps $l = 1, \ldots, M$: LSTM_S selects the measurement projectors, LSTM_R reconstructs the density matrix after each new measurement, and the two modules are optimized jointly. Finally, to describe the fidelity of a state reconstruction, the trained models are evaluated on S_{test} using the so-

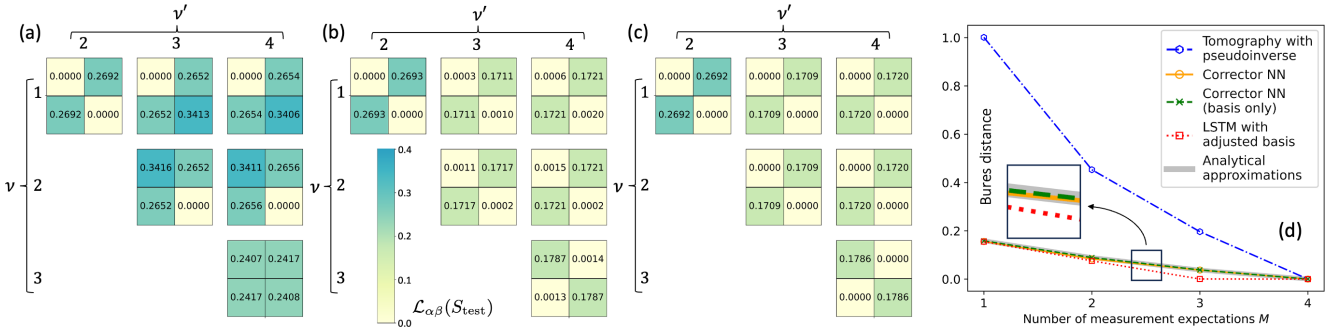


FIG. 2. Undercomplete 1-qubit density matrix tomography with two measurements (ν, ν'). (a-c) Element-wise reconstruction error $\mathcal{L}_{\alpha\beta}(S_{\text{test}})$ using: (a) pseudoinverse reconstruction, (b) NN corrector, (c) best possible analytical formula. (d) Bures distance to target matrix averaged over S_{test} for the different techniques with increasing number of measurements. All panels show errors of the raw reconstruction outputs. No trace normalization or eigenvalue clipping is applied before comparison.

called Bures distance [37, 38]:

$$\mathcal{B}(\rho^{\text{NN}}, \rho) = 2 - 2\sqrt{F(\rho^{\text{NN}}, \rho)}, \tag{10}$$

with the *fidelity* $0 \leq F \leq 1$ defined as

$$F(\rho^{\text{NN}}, \rho) = \text{Tr}\left(\sqrt{\sqrt{\rho^{\text{NN}}}\rho\sqrt{\rho^{\text{NN}}}}\right)^2.$$

Here ρ^{NN} denotes the direct output of the considered reconstruction method. No additional physical projection is applied before evaluating the error. When a method produces a non-normalized or non-positive operator, this unphysicality contributes to the increased reconstruction error. The quantity should therefore be understood as a raw-output Bures-type score rather than as a post-processed physical-state distance. To accumulate statistics for different values of M, the whole procedure is then repeated for $M \in \{1, ..., 4^N\}$

III. 1-QUBIT RESULTS

As the first step of the evaluation, we verify the proposed approach for the trivial 1-qubit system. In Fig. 2 we present the reconstruction error for a 1-qubit density matrix, where we calculate the average error element-wise: $\mathcal{L}_{\alpha\beta}(S_{\text{test}}) = \frac{1}{N_{\text{test}}}\sum_i |\rho_{i,\alpha\beta} - \rho^{\text{NN}}_{i,\alpha\beta}|$, $\rho_i \in S_{\text{test}}$. We plot the results for different pairs of measurement operators (ν, ν'). One can notice that the NN corrector, as presented in Fig. 2(b), has almost the same structure of errors as the analytical formulas – Fig. 2(c), which give the best possible reconstruction for a given pair of measurements. These formulas can be easily found in a 1-qubit system and are discussed in Appendix C. What is characteristic, the optimal selection of $b^{\text{NN}}_{\mu\nu}$ and c^{NN}_{μ} does not depend on the measurement results m_ν but only on the measurement operators Π_ν. Therefore, $\rho^{\text{NN}}_{\text{corr}}(\mathcal{M})$ (as well as $\bar{\rho}^{\text{NN}}_{\text{M}^2\text{-corr}}(\mathcal{M})$) simply collapses to $\bar{\rho}^{\text{NN}}_{\text{corr}}(\Pi)$. Measurement expectations m_ν become relevant for the reconstruction of the 2-qubit and larger systems. It is worth noting that the optimal reconstruction found by NN outperforms an approximate one using the pseudoinverse – Fig. 2(a). The latter often has problems with preserving the trace-norm, as discussed in Appendix C.

Similar error precedence can be observed when examining the (S_{test}-averaged) reconstruction (error) measure $\langle\mathcal{B}\rangle$, for different numbers of measurements of a single qubit. Fig. 2(d) shows that the Bures distance for various QST techniques decreases with the number of measurements and approaches zero when $M = 4$. The reconstruction obtained with the NN corrector (orange curve) is much better than pseudoinverse (blue) and practically as good as (the best) analytical approximation (gray), which is further discussed in Appendix C. The NN corrector that takes only information about the basis, i.e. Π_ν, (green) performs similarly. The LSTM model (red) additionally optimizes the measurement order, avoiding redundancy, and obtains a complete reconstruction just with $M = 3$ measurements, that for 1-qubit setup with optimized choice of measurements (i.e. one that utilizes trace-normalization – see Appendix C) are sufficient. Note that if the measurement set is not optimal (i.e., is redundant), then even three measurements do not provide complete information about the state. Consequently, even in the best analytical reconstruction, the gray curve in Fig. 2(d) exhibits a nonzero error for $M = 3$.

IV. 2-QUBIT RESULTS

Let us now move to a 2-qubit system, whose geometry is far more nontrivial [39–41] and even describing semi-positivity is nontrivial itself. The average Bures distance $\langle\mathcal{B}\rangle$ vs. the number of measurements M for different reconstruction approaches for the set S_{test} of random mixed states is presented in Fig. 3. Although all NN models achieve far better scores in density matrix reconstruction than the pseudoinverse-based QST (blue), they also outperform the MLE method-based reconstructions (computed using `Quantum-Tomography` library [42]) both when calculated with (light blue) and without (dark

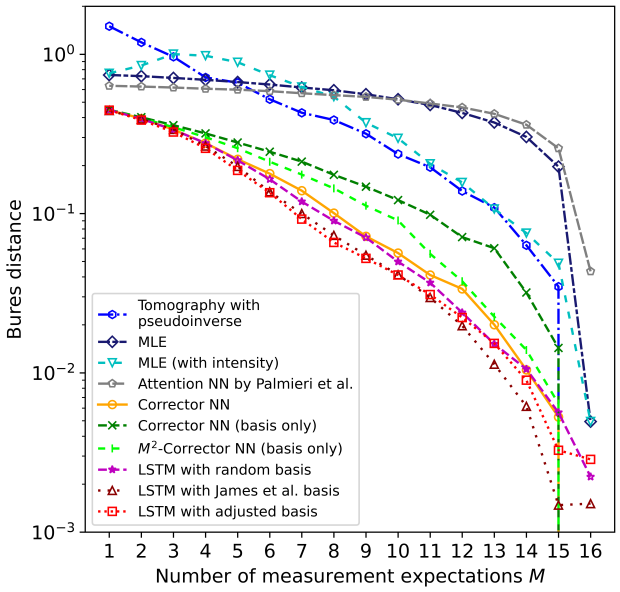


FIG. 3. Undercomplete tomography of 2-qubit random mixed states: S_{test}-averaged Bures distance $\langle \mathcal{B} \rangle$ to target matrix for the different reconstruction techniques with increasing number of measurement outcomes M.

blue) the *intensity* parameter. It informs the MLE which basis measurements were present (with intensity 1) or absent (with zero intensity) during the fitting process. Our protocols also outperform the attention-based NN QST (LI-NN model) introduced by Palmieri *et al.* [26] (grey curve). In the LI-NN model, the attention mechanism assigns larger weights to the most informative measurement expectation values, enabling the NN to more efficiently extract the information relevant for QST. To adapt the method to our setting, we add substantially stronger Gaussian noise (with 100 times larger standard deviation, $\sigma = 0.158$) to those measurement expectation values that are not used in a given test (i.e., those that are *out of the box*). This differs from the original methodology [26] in which variable noise levels were added uniformly to *all* measurements.

When it comes to the NN corrector model, it is clear that including information about the measurement values in $\rho^{\text{NN}}_{\text{corr}}(\mathcal{M})$ (orange), helps the NN to predict slightly more accurate corrections to the pseudoinverse method than $\bar{\rho}^{\text{NN}}_{\text{corr}}(\Pi)$ (dark green). One can also observe that the inclusion of the quadratic term in $\bar{\rho}^{\text{NN}}_{\text{M}^2\text{-corr}}(\Pi)$ (light green), Eq. 9, gives better reconstruction than the linear model (dark green), but still worse than the measurement value-dependent one (orange).

The sequential LSTM results show the benefit of learning the measurement order. The violet curve in Fig. 3 corresponds to the baseline LSTM with a random basis sequence: only LSTM_R is used, while the measurement projectors are drawn randomly without replacement from $\mathcal{C}_N$. The brown curve corresponds to LSTM with predefined basis, where LSTM_S learns to choose projectors from the catalog $\mathcal{C}_N$. In this case, after training, the selector typically assigns almost all probability to one ordered list of catalog projectors. In other words, we can say that the selector "collapses" to a single measurement sequence. However, different training runs may still converge to different, nearly equivalent sequences. The red curve corresponds to the LSTM with adjusted measurement basis, where LSTM_S is allowed to output product projectors not restricted to the discrete catalog. In this case, the learned measurement directions can depend more continuously on the previous measurement history, and the trajectories in projector space need not coincide with a fixed list of catalog elements.

Also in Fig. 3, we can observe that for the complete measurement set ($M = 16$), the corrector schemes exhibit no error at all (the correction is trivial, no correction is needed), while the LSTM models that reconstruct a state from scratch show a small but finite error. From this perspective, the LSTM with predefined basis (brown curve) performs better.

In general, for an arbitrary state, there is no predefined optimal sequence of measurements, which are linearly independent operators. However, if we restrict the state space to a certain subclass, e.g. to so-called X states [43], certain sequences of measurements are preferable to others. Typically, the first three gave full information about the diagonal elements; while the next four gave full information about the coherences. It is further discussed in Appendix D. In Fig. 4, we compare the performance of the LSTM model when trained on a restricted space of X states versus training on general mixed states (as in Fig. 3). We observe that the model can adapt when trained on restricted states. Increasing the system size

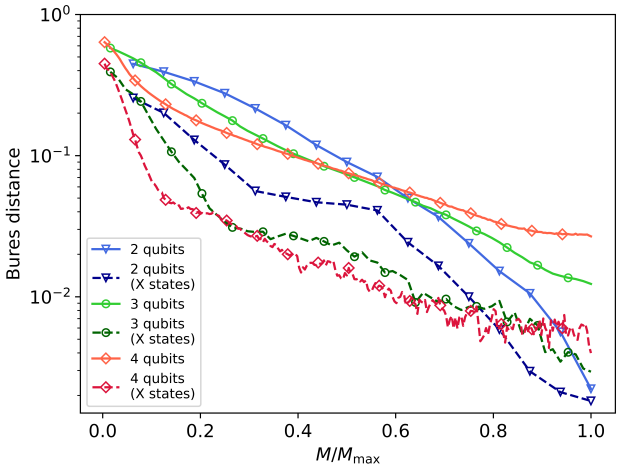


FIG. 4. Undercomplete tomography for restricted states: average Bures reconstruction distance $\langle \mathcal{B} \rangle$ as a function of the number of measurement outcomes M ($M_{\text{max}} = 4^N$), for model trained on general mixed states (solid lines) and on the restricted X-states family (dashed lines) across different system sizes $N = 2, 3, 4$. Results for LSTM with random basis.

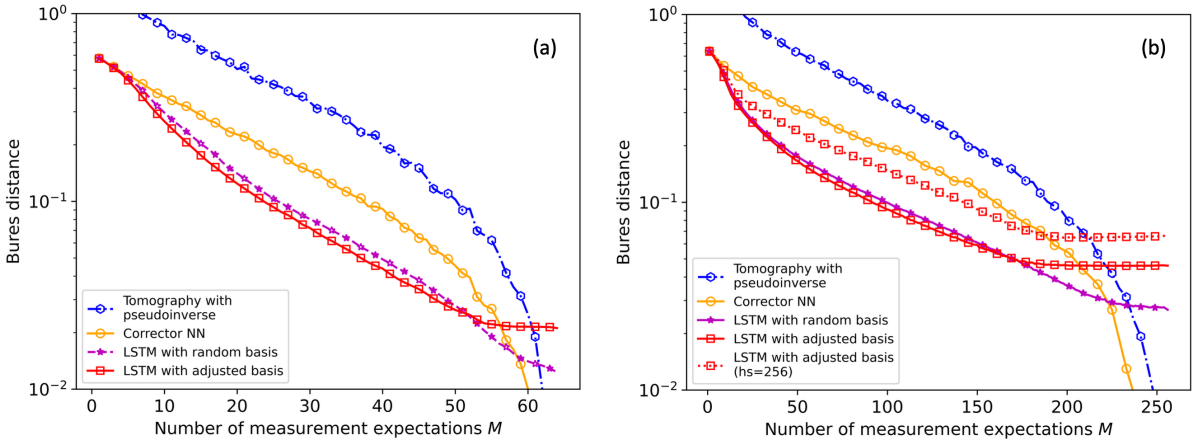


FIG. 5. Undercomplete (a) 3-qubit and (b) 4-qubit tomography of random mixed states: average Bures distance $\langle \mathcal{B} \rangle$ to a target matrix for the different reconstruction techniques with increasing number of measurement expectation values M: (a) $M = 1, \ldots, 4^3$, and (b) $M = 1, \ldots, 4^4$. Additional dotted red curve in (b) presents results for the smaller LSTM with adjusted basis model.

also affects the model's performance. In particular, for X states, where the number of effective parameters scales as 2^{N+1} compared to $4^N - 1$ for general mixed states, it becomes evident that significant information about the quantum state accumulates in the initial measurements and can be efficiently learned by the NN.

V. SCALING UP TO LARGER SYSTEMS

All the presented methods can be scaled to larger systems – up to 3 (4) qubits with $4^3 = 64$ ($4^4 = 256$) measurements, as presented in Fig. 5(a) and (b), respectively. These results illustrate tomographic reconstruction for random mixed states of 3 and 4 qubits. Comparing them with the 2-qubit results shown in Fig. 3, we observe that, as before, our proposed protocols significantly outperform standard methods based on the pseudoinverse; however, it is worth adding that the pseudoinverse reconstruction decreases with the number of measurement expectation values M in the same exponential way as the ML-based methods. We also see that the results scale qualitatively: all curves exhibit a similar character in the plots for 3 and 4 qubits. Two characteristic features can also be observed: LSTM-based methods perform slightly better than the corrector NN; on the other hand, for larger systems, a small but finite error appears in the LSTM methods toward the end of reconstruction, when the measurement basis is nearly complete. This is related to the fact that recurrent networks (even LSTMs) may "forget" the earliest measurement expectations.

Regarding the scaling of model size, which determines their capacity when the system size N changes – at each size, the capacity was chosen to ensure comparable performance between models at different system sizes. The criterion was to achieve a Bures distance about 5 times smaller at the halfway measurement range ($M = 4^N/2$) for the LSTM-based models compared to the pseudoinverse. For LSTM models, we had to increase the number of layers from 1 to 2 for 3 and 4-qubit systems. Additionally, for LSTM with adjusted basis, to keep the satisfactory level of performance for 4 qubits, we further modified the hidden size to 1024 units. In the case of the corrector NN, the increase in the number of parameters follows directly from the input and output scaling, while the architecture itself was kept unchanged.

The total number of learnable parameters in the analyzed models are listed in Table I and illustrated in Figure 6. From Fig. 6, we see that as the system size N increases, the capacity of the NN models must also be increased. The increase of the NN corrector (orange curve) is exponential and is directly related to the change in the dimension 2^N of the state space. For both LSTM-

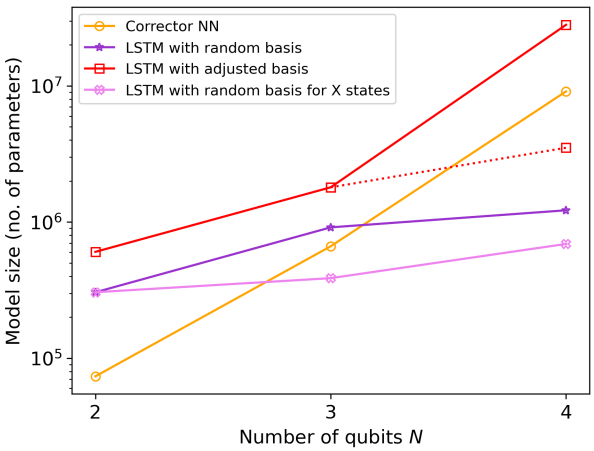


FIG. 6. Scaling of the analyzed neural network models with the system size N. Additional dotted red curve presents scaling for the smaller LSTM with adjusted basis model.

Model	2 qubits	3 qubits	4 qubits
Corrector NN	7.36×10^4	6.64×10^5	9.11×10^6
LSTM with random basis	3.05×10^5	9.14×10^5	1.22×10^6
LSTM with adjusted basis	6.04×10^5	1.80×10^6	2.81×10^7 (3.51×10^6)

TABLE I. Number of neural network parameters for different models and system sizes. The number of parameters for LSTM with adjusted basis for 4 qubits is presented both for the model with 1024 hidden units and 256 units (in brackets).

based models (red and violet curves), their size also grows exponentially, which is natural since we tested them on general mixed states. For example, in the case of LSTM with an adjusted basis for $N = 4$ qubits – shown as the solid red curve in Fig. 5(b) – we intentionally increased the number of hidden units in the LSTM cell to 1024 to achieve a performance comparable to that of $N = 2, 3$. If instead we keep the number of hidden units the same as for $N = 2, 3$ qubits (i.e., hidden size (hs) = 256) – giving the scaling shown in Fig. 6 by the dotted red curve – the performance decreases, as indicated by the dotted red curve in Fig. 5(b). Extrapolating the observed parameter count shown in Fig. 6 suggests that analogous architectures could be constructed for larger systems, with the parameter count remaining below 10^9 up to approximately 6 qubits for the architectural choices considered here.

VI. POSITIVE SEMIDEFINITENESS

As additional tests for the models' performance, we also verified that both the corrector NN and the LSTM models reconstruct a positive semidefinite matrix: regardless of the measurement collection size M, the reconstructed matrix has the lowest eigenvalue that is on average always nonnegative, as presented in Fig. 7. Here the two lowest eigenvalues: 1st and 2nd (marked as an intense and faded color, respectively) are represented as the average eigenvalue and its standard deviation (marked as an error bar). This shows that NNs tend to preserve the geometry of mixed states, which is not true for the pseudoinverse method. MLE often returns rank-deficient density matrices whose smallest eigenvalues are exactly zero because the optimum lies on the boundary of the physical state space [44]. We stress that the positivity analysis shown here is diagnostic only. Positive semidefiniteness is not imposed on the reconstructed matrices during evaluation. Instead, the eigenvalues are computed directly from the raw outputs of the corresponding reconstruction methods.

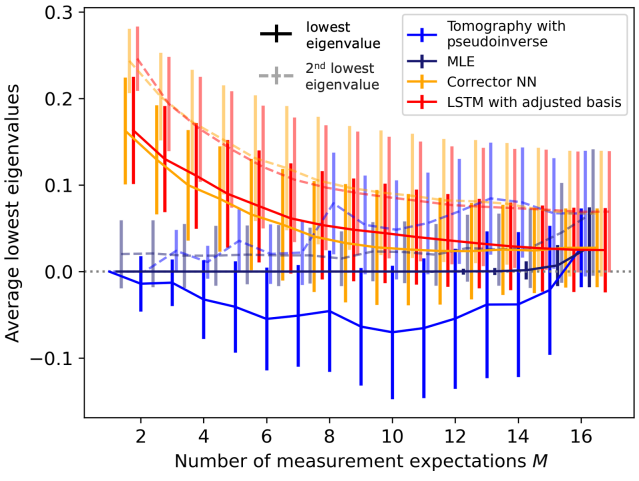


FIG. 7. Positive semidefiniteness of density matrices reconstructed using different techniques. Shown are the average lowest (intense color) and second-lowest (faded color) eigenvalues, with error bars indicating the standard deviation.

VII. DISCUSSION

The results presented here demonstrate that learning-based approaches can significantly improve QST for informationally incomplete measurement data. Across all tested system sizes, NN estimators consistently outperform standard reconstruction techniques, including pseudoinverse-based inversion and MLE, even when trained and evaluated on general mixed states. Results for 3- and 4-qubit random mixed states show that the qualitative behavior observed for two qubits persists for larger systems: the NN-based methods remain substantially more accurate than the pseudoinverse reconstruction in the undercomplete regime, and the LSTM-based schemes typically outperform the corrector network at intermediate values of M. At the same time, the larger-system results reveal a limitation of the recurrent approach. Close to the complete-measurement regime, the LSTM reconstructions may retain a small but finite error, even when almost all measurement expectation values are available. We attribute this effect to the finite memory of recurrent architectures: despite the gating mechanism of LSTM cells, information from the earliest measurements can be partially forgotten after a long sequence of later inputs. This suggests that future implementations may benefit from architectures with more direct access to the full measurement history, such as attention-based or transformer-like models.

Two complementary strategies were explored: (1) learning optimal linear or quadratic reconstruction maps conditioned on the chosen measurement operators; (2) pair of recurrent LSTM networks performing sequential reconstruction while adaptively selecting subsequent measurements. In both cases, the networks can effectively learn the geometry of quantum states, e.g., positive semidefiniteness.

Training on restricted families of states leads to faster convergence and reveals preferred measurement sequences, indicating that the networks can adaptively tailor the reconstruction protocol to the relevant region of

the state space. In this sense, the results provide an operational manifestation of a *no-free-lunch* principle for QST: efficiency gains are achievable when some assumptions of the underlying state space can be learned.

Learning-assisted QST complements existing approaches such as compressed sensing [45], Bayesian inference, and shadow tomography. Whereas these methods rely on analytically specified priors or targeted observable prediction, NNs offer a general and adaptable framework capable of learning reconstruction protocols directly from measurement statistics. The results presented here suggest that such methods can enhance state characterization in both general and structured scenarios, particularly in experimentally relevant regimes where measurement resources are limited.

A limitation of the present numerical analysis is that it is performed in the infinite-shot regime. For a finite number N_c of copies, each input probability m_ν should be replaced by an empirical frequency $\hat{m}_\nu = n_\nu/N_c$, with binomial variance $m_\nu(1-m_\nu)/N_c$. Such sampling noise would propagate through both the pseudoinverse estimator and the neural reconstructions, and its quantitative impact is expected to depend on N_c, the selected projectors, and the conditioning of the reconstruction map. The present work deliberately separates this effect from the error due to the explicit focus on the informational incompleteness scenario. Incorporating realistic finite-shot fluctuations, detector noise, and possible noise-aware training or data augmentation is a natural next step toward experimental deployment of the proposed schemes.

ACKNOWLEDGMENTS

MK and JP acknowledge support from National Science Centre, Poland, under grant no. 2021/43/D/ST3/01989. KR is funded within the QuantERA II Programme that has received funding from the EU H2020 research and innovation programme under Grant Agreement No. 101017733, and with funding organization MEYS. This work was supported by the research programme of the Strategy AV21 AI: Artificial Intelligence for Science and Society. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2024/017284.

DATA AVAILABILITY

The datasets generated and analyzed during the current study are available in the Zenodo repository, `https://doi.org/10.5281/zenodo.17276582`.

Appendix A: Neural models

1. Corrector network details

The *corrector* NN, introduced in the Eq. 8 of the main text, has two outputs: $\mathrm{NN}_{\mathrm{corr}}(\mathcal{M}) = (b^{\mathrm{NN}}_{\mu\nu}(\mathcal{M}), c^{\mathrm{NN}}_{\mu}(\mathcal{M}))$. The second output, c^{NN}_{μ} vector, is conceived to enable the model to predict normalized ρ^{NN}, irrespectively of the expectation values $(m_\nu)_{\nu=1}^{M}$, $M < 4^N$ (which no longer form a complete tomographic set). Moreover, if we take Eq. 8 and calculate element $\langle\psi_{\nu'}|\,.\,|\psi_{\nu'}\rangle$ of both sides, we arrive at $B_{\nu'\mu}v_\mu = 0$, $v_\mu = b^{\mathrm{NN}}_{\mu\nu}m_\nu + c^{\mathrm{NN}}_{\mu}$ for all $\nu' = 1...M$ (see Appendix E for derivation), which means that the two NN outputs are not independent, and vector v_μ must belong to the orthogonal complement of the subspace spanned by rows of $B_{\nu'\mu}$. To fulfill this condition, we added an extra component to the corrector NN loss during its training. Although the orthogonality condition speeds up training, it is not necessary: if we remove it, the corrector NN learns it anyway. It turns out that in the 1-qubit case, i.e., $N = 1$, taking an appropriate (depending on the measurement basis $(\Pi_\nu = \mu_{i_1})_{i_1=1}^{M<4}$) orthogonal vector v_μ (from the orthogonal complement) and choosing its norm so that $\mathrm{Tr}(\rho^{\mathrm{NN}}) = 1$ is equivalent to taking the best possible analytical formula for the reconstruction, as discussed in the Appendix C. However, for two and more qubits the choice is not so obvious anymore. Also, for a single qubit, the optimal $b^{\mathrm{NN}}_{\mu\nu}$ and c^{NN}_{μ} do not depend on the results of measurements m_ν (we can simply remove this information from training, getting the same results), which is not the case for $N > 1$.

Note that the input layer size in all NN correctors $(\rho^{\mathrm{NN}}_{\mathrm{corr}}(\mathcal{M}), \bar{\rho}^{\mathrm{NN}}_{\mathrm{corr}}(\Pi), \bar{\rho}^{\mathrm{NN}}_{\mathrm{M^2\text{-}corr}}(\mathcal{M}))$ depends on M. Due to its universality [46, 47], as our baseline model we choose a 6-layer fully connected (FC) network (multilayer perceptron) with 64 units (neurons) in each hidden layer and ReLU activations. The input/output layers are then adapted to meet the size M. Given the collection $\mathcal{M}$ of measurements of size M and the training set of random density matrices S_{train} the NNs were trained to minimize the supervised training loss function:

$$\mathcal{L}(\mathcal{M}, S_{\mathrm{train}}) = \frac{1}{N_b}\sum_i \|\rho_i - \rho_i^{\mathrm{NN}}(\mathcal{M})\|, \tag{A1}$$

where the difference between the original $\rho_i \in S_{\mathrm{train}}$ and the reconstructed ρ^{NN} is calculated averaging over a batch of size $N_b = 64$, and $\|\cdot\|$ denotes a standard L^2 norm. The batch is a random fraction of S_{train} used in a single step of the training procedure (typically when training NN the stochastic gradient descent method [47] is used). Random-sampling methods used to generate S_{train} and S_{test} sets are described in Appendix B.

2. Measurements selector network details

The selector model consists of two recurrent neural networks. A recurrent network processes a sequence step by step and stores information from previous steps in an internal hidden state. Specifically, we utilize a popular LSTM [48] architecture. Here, LSTM_S is responsible for measurement selection, while LSTM_R is responsible for state reconstruction. The input size of both networks is independent of the final number of measurements M, because the same cells are reused at each measurement step.

Randomly selecting a reduced set of measurements is not the best possible approach. Having partial information about the system (i.e. few already performed measurements) may allow the model to predict the next most promising measurement operator, which gives the most valuable information about the system. Especially, we propose to feed the LSTM_S with initial measurement basis $\Pi_1 = {\mu_0}^{\otimes N}$ and force it to predict the next measurement operator Π_2, which is used to determine value m_2 according to Eq. 1 in the main text. Subsequently, this procedure is repeated until the final choice Π_M is obtained.

Since there is no predetermined correct order of measurements, the typical approach of supervised training cannot be applied in this case, i.e., LSTM_S cannot be trained by simply comparing its output with a prescribed correct sequence. Instead, LSTM_S and LSTM_R are trained together. At each step, LSTM_R produces a reconstruction ρ_l^{NN}, and the loss compares this reconstruction with the true training state. The loss is then summed over all steps, so that measurement choices are rewarded when they lead to better subsequent reconstructions. To enable free information flow during *backpropagation* [47] (NN training), the measurement expectation values algebra is incorporated directly into the NN architecture, i.e., values $m_l = \mathrm{Tr}(\Pi_l \rho)$ are calculated explicitly during the training phase. At the test stage, however, the reconstruction network does not require access to the target state ρ: the internally computed value m_l can simply be replaced by an externally supplied experimental estimate $\hat{m}_l = n_l/N_c$. Finally, both networks are trained with the use of L_2 loss

$$\mathcal{L}(S_{\mathrm{train}}) = \frac{1}{N_b N_l} \sum_{i,l} \|\rho_i - \rho_{i,l}^{\mathrm{NN}}\|, \quad \rho_i \in S_{\mathrm{train}}, \quad \text{(A2)}$$

where N_b denotes the number of samples in the S_{train} and N_l is the number of selection-reconstruction steps.

Two versions of the *selector* network LSTM_S were proposed and tested. The first approach at step l simply selects a previously unused measurement operator Π_l from the catalog $\mathcal{C}_N$, defined in Eq. (3) of the main text. The second strategy is to allow LSTM_S to produce a *custom* measurement operator Π_l. The selected operator is then used to collect measurements and both the operator and the resulting measurement expectation values are passed to the reconstruction network LSTM_R. One should also note that the first version needs discrete selections that typically utilize the *argmax* function, which is not differentiable. Therefore, to make the selection process differentiable and thus enable backpropagation, we applied the softmax function, i.e., the so-called *Gumbel-softmax* activation [49], to select the operator Π_l. However, further experiments have shown that another approach with an additional loss term for the *selector* LSTM_S, as discussed in Appendix F, may be equally effective.

We also note that, since the density matrix is complex, each NN processes the real and imaginary parts separately, treating them as distinct channels.

Appendix B: Training and test datasets

We generate a set of density matrices by using different random-sampling methods. The main idea is to generate random states using the quantum circuits approach, but in order to have a well-diversified ensemble, we also use a technique based on sampling from the uniform Haar measure [50]. Additionally, to increase the number of maximally entangled states, we specifically include 20000 of them. All methods are described in detail in Ref. [27]. In total, we generate 460000 matrices for the training dataset S_{train} and 46000 matrices for the test set S_{test} for 2-qubit experiments. The sizes of the 3- and 4-qubit datasets were the same. The datasets generated and/or analyzed during the current study are available in the Zenodo repository, `https://doi.org/10.5281/zenodo.17276582`.

Appendix C: Single-qubit tomography from incomplete measurement set – best approximation

Here we analyze the simplest single-qubit system ($N = 1$) and show how the standard quantum tomography scheme extended via the pseudoinverse operation into an incomplete measurements scenario, i.e. $M = 2 < 4$, behaves and in what way this behavior can be corrected.

Let us assume a density matrix in the following form:

$$\rho = \begin{pmatrix} a & b \\ b^* & 1-a \end{pmatrix}, \quad \text{(C1)}$$

with real $a \in [0,1]$, complex b with $|b| \in [0, \frac{1}{2}]$ and $a \geq a^2 + |b|^2$ to fulfill positive semidefiniteness condition. Then, the measurement expectation values (in the projector basis $\Pi_\nu = \mu_{i_1} = |\psi_{i_1}\rangle\langle\psi_{i_1}|$, with $|\psi_{i_1}\rangle$ (defined in Eq. 2) are $(m_\nu)_{\nu=1}^4 = (a, 1-a, \frac{1}{2} + \mathrm{Re}(b), \frac{1}{2} + \mathrm{Im}(b))$, thus $B_{\nu\mu}$ matrix and its inverse are:

$$B_{\nu\mu} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \end{pmatrix}, \; B^{-1}_{\mu\nu} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 2 & 0 \\ 1 & 1 & 0 & -2 \\ 1 & -1 & 0 & 0 \end{pmatrix}. \quad \text{(C2)}$$

Finally, calculating $\rho = \Gamma_\mu B^{-1}_{\mu\nu} m_\nu$ restores the original matrix ρ defined in Eq. (C1).

The situation changes if we reduce the number of projectors to two, e.g., $\nu = 1, 2$ by taking μ_1, μ_2 with $m_1 = a$ and $m_2 = 1 - a$ expectations. Now $B_{\nu\mu}$ [formed by two first rows of $B_{\nu\mu}$ from Eq. (C2)] is no longer a square matrix and its inverse does not exist. However, one can calculate its pseudoinverse getting:

$$B^{+}_{\mu\nu} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix}^T, \tag{C3}$$

and then reconstruct the state via $\rho \simeq \Gamma_\mu B^{+}_{\mu\nu} m_\nu$ (note that $\nu = 1, 2$, $\mu = 1, 2, 3, 4$) formula:

$$\rho_{\text{pseudoinv}} = \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix}. \tag{C4}$$

One can notice that the diagonal part of $\rho_{\text{pseudoinv}}$ in Eq. (C4) is reconstructed correctly, while off-diagonal terms are absent. The most important observation is that since m_1 and m_2 measurement expectaion values contain only the information about populations, i.e. a, better reconstruction than Eq. (C4) is not possible and pseudoinverse scheme gives the best possible reconstruction and no corrections are needed.

However, when we select two other measurement expectation values: $\nu = 1, 3$, that is $m_1 = a$ and $m_3 = \frac{1}{2} + \text{Re}(b)$, pseudoinverse will not give the best possible reconstruction. Now the pseudoinverse is

$$B^{+}_{\mu\nu} = \frac{1}{3}\begin{pmatrix} 1 & -1 & 0 & 2 \\ 1 & 2 & 0 & -1 \end{pmatrix}^T, \tag{C5}$$

leading to the reconstruction:

$$\rho_{\text{pseudoinv}} = \frac{1}{3}\begin{pmatrix} 3a & 1-a+2\text{Re}(b) \\ 1-a+2\text{Re}(b) & 1-a+2\text{Re}(b) \end{pmatrix}. \tag{C6}$$

that is not even correctly normalized. Nevertheless, if we add the following correction terms:

$$b_{\mu\nu} = \frac{1}{3}\begin{pmatrix} -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix}^T, \; c_\mu = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & -1 \end{pmatrix}^T, \tag{C7}$$

this via the formula $\rho \simeq \Gamma_\mu((B^{+}_{\mu\nu} + b_{\mu\nu})m_\nu + c_\mu)$ gives the best possible approximation:

$$\rho_{\text{corrected}} = \begin{pmatrix} a & \text{Re}(b) \\ \text{Re}(b) & 1-a \end{pmatrix}. \tag{C8}$$

This shows that there is room for action for the NN *corrector* model compared to the scheme based on pseudoinverse. One can also verify that for terms defined in Eq. (C7) vector $b_{\mu\nu}m_\nu + c_\mu = \frac{1-a-\text{Re}(b)}{3}(1, -1, 0, -1)^T$ is orthogonal to 1st and 3rd row in matrix $B_{\nu\mu}$ from Eq. (C2). The same can be done for the other pairs of measurements obtaining various $\rho_{\text{corrected}}$ for pairs of measurements (ν, ν')

$$\begin{array}{c} \\ \nu \end{array} \begin{array}{c} \nu' \\ \left[\begin{array}{ccc} \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix} & \begin{pmatrix} a & \text{Re}(b) \\ \text{Re}(b) & 1-a \end{pmatrix} & \begin{pmatrix} a & \text{Im}(b) \\ -\text{Im}(b) & 1-a \end{pmatrix} \\ & \begin{pmatrix} a & \text{Re}(b) \\ \text{Re}(b) & 1-a \end{pmatrix} & \begin{pmatrix} a & \text{Im}(b) \\ -\text{Im}(b) & 1-a \end{pmatrix} \\ & & \begin{pmatrix} \frac{1}{2} & b \\ b^* & \frac{1}{2} \end{pmatrix} \end{array}\right] \end{array}, \tag{C9}$$

where rows denotes subsequent $\nu = 1, 2, 3$ measurement expectations, while columns $\nu' = 2, 3, 4$ expectation. The expressions in Eq. (C9) were used to generate the results shown in Fig. 2(c) of the main text.

Appendix D: 2-qubit tomography – optimal ordering for X states.

The geometry of 2-qubit states is much more complicated than single-qubit ones [39–41]. To verify effectiveness of NN reconstruction in 2-qubit case we introduce a family of so-called X states that generalize many important classes of mixed quantum states [43], but have much simpler structure than the general matrix:

$$X = \begin{pmatrix} a & 0 & 0 & w \\ 0 & b & z & 0 \\ 0 & z^* & c & 0 \\ w^* & 0 & 0 & d \end{pmatrix}, \tag{D1}$$

with $a + b + c + d = 1$, $a, b, c, d \geq 0$, and $|z| \leq \sqrt{bc}$, $|w| \leq \sqrt{ad}$. Then the measurements in our basis, $\Pi_\nu = \mu_{i_1} \otimes \mu_{i_2}$, with $\mu_{i_k} = |\psi_{i_k}\rangle\langle\psi_{i_k}|$ defined in Eq. 3, are:

$$\begin{aligned} &(m_\nu)_{\nu=1}^{16} = \\ &\big(a, b, \frac{1}{2}(a+b), \frac{1}{2}(a+b), \\ &c, d, \frac{1}{2}(c+d), \frac{1}{2}(c+d), \\ &\frac{1}{2}(a+c), \frac{1}{2}(b+d), \\ &\frac{1}{4} + \frac{1}{2}\left(\text{Re}(w) + \text{Re}(z)\right), \frac{1}{4} + \frac{1}{2}\left((\text{Im}(w) - \text{Im}(z)\right), \\ &\frac{1}{2}(a+c), \frac{1}{2}(b+d), \\ &\frac{1}{4} + \frac{1}{2}\left((\text{Im}(w) + \text{Im}(z)\right), \frac{1}{4} + \frac{1}{2}\left((\text{Re}(z) - \text{Re}(w)\right)\big). \end{aligned} \tag{D2}$$

Thus, some of the measurements are redundant, meaning that some orderings of measurements might be more optimal than others. This leaves room for the NN LSTM model to develop an optimization of the measurement sequence.

By training the LSTM on X states, we obtained sequences like $\nu = (1, 2, 6, 11, 16, 12, 15, ...)$ or $(1, 3, 14, 16, 12, 15, 11, ...)$. Typically, in the acquired sequences, the first three gave full information about the

diagonal elements; while the next four (any permutation of $(11, 12, 15, 16)$ measurements) gave full information about the coherences. This result shows that there is a tomographic equivalent of the *no free lunch* theorem [51, 52]: by restricting the state space, we gain the possibility of more efficient QST by learning the optimal (redundancy-free) sequence of measurements.

Appendix E: Relation between correction terms

We start from Eq. 8 by taking the matrix element $\langle\psi_{\nu'}| \cdot |\psi_{\nu'}\rangle$ of both sides. Assuming perfect reconstruction, $\langle\psi_{\nu'}|\rho_{\rm NN}|\psi_{\nu'}\rangle \simeq m_{\nu'}$, we obtain

$$m_{\nu'} = B_{\nu'\mu}\left[\left(B^{+}_{\mu\nu} + b^{\rm NN}_{\mu\nu}\right) m_\nu + c^{\rm NN}_\mu\right]. \tag{E1}$$

This yields ($B_{\nu'\mu}B^{+}_{\mu\nu} = \delta_{\nu'\nu}$),

$$m_{\nu'} = \delta_{\nu'\nu} m_\nu + B_{\nu'\mu}\left(b^{\rm NN}_{\mu\nu} m_\nu + c^{\rm NN}_\mu\right), \tag{E2}$$

and therefore

$$B_{\nu'\mu} v_\mu = 0, \tag{E3}$$

where

$$v_\mu = b^{\rm NN}_{\mu\nu} m_\nu + c^{\rm NN}_\mu, \tag{E4}$$

for all $\nu' = 1, \ldots, M$. This implies that the two corrector network outputs are not independent: the vector v_μ must belong to the null space (kernel) of $B_{\nu'\mu}$, which is the orthogonal complement of the row space of $B_{\nu'\mu}$.

Appendix F: Loss in selector NN with predefined measurement basis

The issue with the LSTM_S model, which selects measurement operators from the predefined catalog $\mathcal{C}_N$ is that the *argmax* function typically used in classifiers (selectors) is not differentiable.

The main approach that we utilize in this work is to use the softmax activation (e.g., Gumbel-softmax [49]) to construct the operator Π'_l used only as input to LSTM_R during the training stage as:

$$\Pi'_l = \sum_\nu \text{Softmax}(P^l_\nu)\Pi_\nu, \tag{F1}$$

where P^l is the probability vector returned by the LSTM_S at step l, containing the probabilities P^l_ν of choosing the measurement operator Π_ν from the discrete catalog $\mathcal{C}_N$. Additionally, we enforce the uniqueness of the choice, so that Π'_l is different from previously selected operators $\{\Pi'_1, ..., \Pi'_{l-1}\}$. Importantly, while Π'_l is used as LSTM_R input, we keep $\Pi_l = \Pi_{\text{argmax}(P^l)}$ as input of LSTM_S and use it to determine the expectation value m_l. As a result, we allowed smooth gradient backpropagation through Π'_l passed from LSTM_S to LSTM_R.

Simultaneously, a different approach resulted in a similar reconstruction accuracy. Here, during the training stage, in the given step l each possible measurement pair $(\Pi_{\nu_l},\ m_{\nu_l})$ is constructed and independently passed through the LSTM_R, resulting in set of reconstructed matrices $\{\rho^{\rm NN}_{\nu_l}\}_{\nu_l=1,\ldots,4^N}$. Then, $\mathcal{L}_{\nu_l}(S_{\rm train}) = \frac{1}{N_b}\sum_i \|\rho_i - \rho^{\rm NN}_{i,\nu_l}\|$, $\rho_i \in S_{\rm train}$ is calculated for each ν_l to find ν^*_l that minimizes the loss. Such ν^*_l is then set as a target in the cross-entropy loss for the vector P^l returned by the LSTM_S:

$$\mathcal{L}_{\rm CE}(P^l, \nu^*_l) = -\sum_{\nu_l} y_{\nu_l}\log(p_{\nu_l}), \tag{F2}$$

where $y_{\nu_l} = 1$ when ν_l is equal to ν^*_l, and $y_{\nu_l} = 0$ otherwise. Finally, the Π_l is set as $\Pi_{\text{argmax}(P^l)}$, meaning that the measurement operator Π_l is chosen as the Π_ν with the highest probability P^l_ν in a given step l, but with additional restriction to preserve the uniqueness of measurements among all steps. Such Π_l, together with respective m_l, are subsequently used as input for the reconstruction network LSTM_R trained using the $\mathcal{L}_l(S_{\rm train})$.

In contrast, the LSTM_S network that directly produces (a custom) measurement operator that can be passed to the reconstruction network, allows for straightforward simultaneous training of both LSTM_S and LSTM_R models using the single loss $\mathcal{L}(S_{\rm train})$. This approach allows for much faster training.

[1] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[2] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[3] M. Paris and J. Rehacek, eds., *Quantum state estimation*, Lecture Notes in Physics, Vol. 649 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).

[4] J. Altepeter, E. Jeffrey, and P. Kwiat, Photonic state tomography (Academic Press, 2005) pp. 105–159.

[5] Y. S. Teo, J. Řeháček, and Z. Hradil, Informationally incomplete quantum tomography, Quantum Measurements and Quantum Metrology **1**, 57 (2013).

[6] V. Bužek, 6 quantum tomography from incomplete data via maxent principle, in *Quantum State Estimation*, edited by M. Paris and J. Řeháček (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004) pp. 189–234.

[7] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil, Quantum-state reconstruction by maximizing likelihood and entropy, Phys. Rev. Lett. **107**, 020404 (2011).

[8] Z. Hradil, Quantum-state estimation, Phys. Rev. A **55**,

R1561 (1997).
[9] Z. Hradil and J. Řeháček, Efficiency of maximum-likelihood reconstruction of quantum states, Fortschritte der Physik **49**, 1083 (2001).
[10] T. Opatrný, D.-G. Welsch, and W. Vogel, Least-squares inversion for density-matrix reconstruction, Phys. Rev. A **56**, 1788 (1997).
[11] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, Nature Communications **1**, 149 (2010).
[12] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, Nature Machine Intelligence **1**, 155 (2019).
[13] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Permutationally invariant quantum tomography, Phys. Rev. Lett. **105**, 250403 (2010).
[14] C. Schwemmer, G. Tóth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter, Experimental comparison of efficient tomography schemes for a six-qubit state, Phys. Rev. Lett. **113**, 040503 (2014).
[15] A. Acharya, S. Saha, and A. M. Sengupta, Shadow tomography based on informationally complete positive operator-valued measure, Phys. Rev. A **104**, 052418 (2021).
[16] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).
[17] S. Aaronson, Shadow tomography of quantum states, SIAM Journal on Computing **49**, STOC18 (2020), https://doi.org/10.1137/18M120275X.
[18] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nature Physics **14**, 447 (2018).
[19] A. Melkani, C. Gneiting, and F. Nori, Eigenstate extraction with neural-network tomography, Phys. Rev. A **102**, 022412 (2020).
[20] T. Xin, S. Lu, N. Cao, G. Anikeeva, D. Lu, J. Li, G. Long, and B. Zeng, Local-measurement-based quantum state tomography via neural networks, npj Quantum Information **5**, 109 (2019).
[21] D. Koutný, L. Motka, Z. c. v. Hradil, J. Řeháček, and L. L. Sánchez-Soto, Neural-network quantum state tomography, Phys. Rev. A **106**, 012409 (2022).
[22] T. Schmale, M. Reh, and M. Gärttner, Efficient quantum state tomography with convolutional neural networks, npj Quantum Information **8**, 115 (2022).
[23] Y. Quek, S. Fort, and H. K. Ng, Adaptive quantum state tomography with neural networks, npj Quantum Information **7**, 105 (2021).
[24] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, Quantum state tomography with conditional generative adversarial networks, Phys. Rev. Lett. **127**, 140502 (2021).
[25] J. Li, S. Huang, Z. Luo, K. Li, D. Lu, and B. Zeng, Optimal design of measurement settings for quantum-state-tomography experiments, Phys. Rev. A **96**, 032307 (2017).
[26] A. M. Palmieri, G. Müller-Rigat, A. K. Srivastava, M. Lewenstein, G. Rajchel-Mieldzioć, and M. Płodzień, Enhancing quantum state tomography via resource-efficient attention-based neural networks, Phys. Rev. Res. **6**, 033248 (2024).
[27] J. Pawłowski and M. Krawczyk, Identification of quantum entanglement with Siamese convolutional neural networks and semisupervised learning, Phys. Rev. Appl. **22**, 014068 (2024).
[28] N. Taghadomi, A. Mani, A. Fahim, and A. Bakouei, Effective detection of quantum discord by using convolutional neural networks, Quantum Machine Intelligence **7**, 40 (2025).
[29] Y. Chen, Y. Pan, G. Zhang, and S. Cheng, Detecting quantum entanglement with unsupervised learning, Quantum Science and Technology **7**, 015005 (2021).
[30] N. Asif, U. Khalid, A. Khan, T. Q. Duong, and H. Shin, Entanglement detection with artificial neural networks, Scientific Reports **13**, 1562 (2023).
[31] J. Ureña, A. Sojo, J. Bermejo-Vega, and D. Manzano, Entanglement detection with classical deep neural networks, Scientific Reports **14**, 18109 (2024).
[32] M. Krawczyk, J. Pawłowski, M. M. Maśka, and K. Roszak, Data-driven criteria for quantum correlations, Phys. Rev. A **109**, 022405 (2024).
[33] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[34] R. Yuan, A brief introduction to POVM measurement in quantum communications (2022), arXiv:2201.07968 [quant-ph].
[35] R. Penrose, A generalized inverse for matrices, Mathematical Proceedings of the Cambridge Philosophical Society **51**, 406–413 (1955).
[36] P. Baláž, M. Krawczyk, J. Pawłowski, and K. Roszak, Learning entanglement from tomography data: Contradictory measurement importance for neural networks and random forests, Phys. Rev. A **112**, 022431 (2025).
[37] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006).
[38] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).
[39] O. Gamel, Entangled bloch spheres: Bloch matrix and two-qubit state space, Phys. Rev. A **93**, 062320 (2016).
[40] L. Jakóbczyk and M. Siennicki, Geometry of bloch vectors in two-qubit system, Physics Letters A **286**, 383 (2001).
[41] M. Kuś and K. Życzkowski, Geometry of entangled states, Phys. Rev. A **63**, 032307 (2001).
[42] S. Turro, `Quantum-Tomography`: A python library to help perform tomography on a quantum state, `https://quantumtomo.web.illinois.edu/` (2023), version 1.0.7.0.
[43] A. A.-Q. Nicolás Quesada and D. F. James, Quantum properties and dynamics of X states, Journal of Modern Optics **59**, 1322 (2012), https://doi.org/10.1080/09500340.2012.713130.
[44] R. Blume-Kohout, Hedged maximum likelihood quantum state estimation, Phys. Rev. Lett. **105**, 200504 (2010).
[45] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, Phys. Rev. Lett. **105**, 150401 (2010).
[46] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural networks **2**, 359 (1989).
[47] P. Grohs and G. Kutyniok, *Mathematical aspects of deep*

learning (Cambridge University Press, 2023).

[48] Y. Yu, X. Si, C. Hu, and J. Zhang, A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures, Neural Computation **31**, 1235 (2019), https://direct.mit.edu/neco/article-pdf/31/7/1235/1053200/neco_a_01199.pdf.

[49] E. Jang, S. Gu, and B. Poole, Categorical reparameterization with gumbel-softmax, arXiv preprint arXiv:1611.01144 (2016).

[50] F. Mezzadri, How to generate random matrices from the classical compact groups, Notices of the American Mathematical Society **54** (2006).

[51] D. H. Wolpert, The lack of a priori distinctions between learning algorithms, Neural computation **8**, 1341 (1996).

[52] D. Wolpert and W. Macready, No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation **1**, 67 (1997).

Learning Hamiltonians for solid-state quantum simulators

Jarosław Pawłowski* and Mateusz Krawczyk
Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

We introduce a generalizable framework for identifying effective Hamiltonians directly from experimental data in solid-state quantum systems. Our unsupervised autoencoder-based approach incorporates the governing physics (here, the S-matrix formalism) directly into the decoder, enabling physically meaningful Hamiltonian inference without labeled Hamiltonian parameters. Through numerical experiments on a triple quantum dot chain, we demonstrate automated characterization of programmable solid-state simulators from transport measurements. The model accurately infers Hamiltonian parameters and generalizes beyond the training domain, while exhibiting finite capacity as the parameter space becomes increasingly broad or diverse. We further demonstrate robustness to noisy measurements and show that the model can be successfully trained on transport data generated outside the Hamiltonian family assumed by the physics decoder, while still identifying meaningful effective parameters.

arXiv:2603.02889v2 [cond-mat.mes-hall] 4 Sep 2026

INTRODUCTION

Machine learning (ML) techniques have been widely explored in solid-state physics, including phase identification [1–4] and materials discovery through the prediction of chemical and electronic properties [5–8]. A central challenge in this area is the extraction of effective Hamiltonians $\mathcal{H}$ from experimental or simulated observables [9]. Existing approaches broadly fall into two categories: optimization-based inverse methods, such as evolutionary algorithms and heuristic searches [10–13], and supervised learning schemes [14–19], which often struggle with robustness and generalization.

Hamiltonian learning (HL) has shown particular promise in nanoscale solid-state platforms, including the identification of quantum nanomagnets from spectral data [14, 15], classification of skyrmionic magnetic textures [16], and ML-assisted control and readout of molecular spin qubits [20]. Beyond transport measurements, HL has also been applied to density-of-states data for automated band-structure inference [21] and to the analysis of local patterns in moiré materials [22, 23]. Learning tight-binding Hamiltonians is likewise emerging as an important research direction [24, 25]. Finally, learning Hamiltonian dynamics constitutes another active line of research, with applications ranging from classical systems [26, 27] and generic quantum dynamics [28, 29] to superconducting quantum processors [30].

Gate-defined quantum dots (QDs) hosting single electrons or holes are promising platforms for solid-state quantum computing due to their electrical tunability and scalability, and have recently attracted interest as programmable quantum simulators [31–34]. In particular, minimal realizations of the Kitaev chain in QD arrays—based on elastic cotunneling and crossed Andreev reflection and supporting so-called *poor man's* Majorana zero modes (MZMs) [35]—have attracted significant attention [36, 37]. Rashba quantum dot chains have also been proposed as alternative simulator platforms capable of hosting MZMs [38, 39].

At the same time, ML-assisted autotuning of QD-based quantum simulators using transport measurements is gaining increasing interest [10, 34, 40–43]. The use of transport measurements in the form of *conductance maps* [33, 44] to extract information about the system Hamiltonian [9, 13, 17, 18, 45–47] appears to be a promising route toward automated parameter tuning. In particular, inverting measured conductance matrices to determine electrostatic potential disorder—using evolutionary optimization [13], supervised neural networks [46–48], or hybrid approaches [49]—provides a natural starting point for precise inference of $\mathcal{H}$ parameters relevant to MZMs.

Physical constraints can be incorporated into neural network (NN) models in two principal ways: by enforcing symmetry-respecting representations (geometric deep learning [50]) or by constraining solutions to satisfy governing differential equations, as in physics-informed neural networks (PINNs) [51–53]. Both approaches typically rely on supervised training, which can be limiting in realistic HL scenarios where the parameters of effective Hamiltonians are not directly accessible. Here, we adopt a physics-decoder (PD) architecture [54–57], shown in Fig. 1(b), inspired by PINNs but formulated as a fully unsupervised autoencoder-based framework. We employ this paradigm specifically for unsupervised Hamiltonian identification from transport measurements, with the central goal of learning a generalizable inverse map from experimentally accessible observables to effective microscopic Hamiltonian parameters. In this scheme, the encoder infers Hamiltonian parameters, while the decoder explicitly implements the underlying physics for the assumed Hamiltonian family—a strategy that can be readily generalized beyond the present setting. HL with physics-informed training schemes remains relatively underexplored and constitutes an active area of research [24, 25, 58, 59].

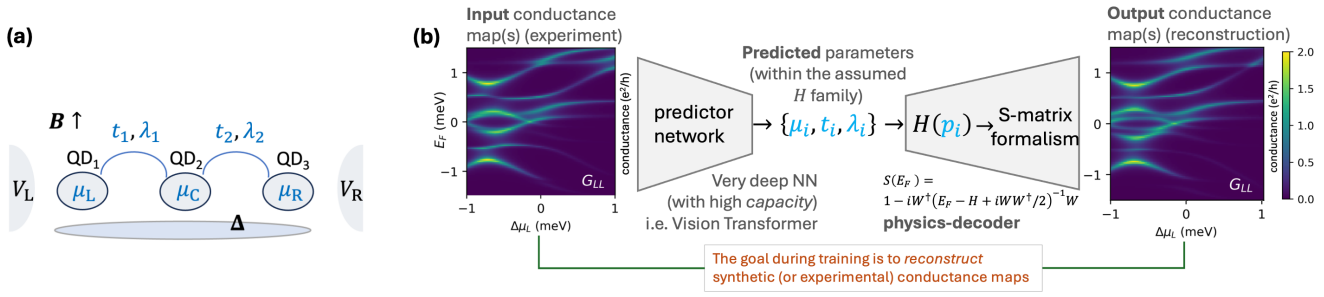


FIG. 1. (a) Chain of the Rashba QDs. (b) Proposed HL architecture composed of encoder—predictor and physics-decoder.

QUANTUM DOTS CHAIN

To demonstrate the working principle of the method, we define a simple simulator consisting of a chain of $N = 3$ QDs, yet sufficiently complex to illustrate the robustness of the approach. We consider a $\mathcal{H}$ family that defines the Kitaev-chain simulator based on a Rashba QD chain coupled with a nearby s-wave superconductor [38], depicted in Fig. 1(a). The Hamiltonian describing chain of spinful single-level QDs with local potential μ_n, Zeeman energy $V_{\rm Z}$, proximity-induced superconducting (s-wave) pairing Δ_n, and Rashba spin-orbit vector $\boldsymbol{\lambda_n}$, which modifies the inter-dot hopping t_n, is defined as:

$$\begin{aligned} H^{\rm QDs} = \sum_{n,s,s'} \Big[& (-\mu_n\sigma_0 + V_{\rm Z}\sigma_z)_{ss'}\, c^\dagger_{n,s} c_{n,s'} \\ & + \tfrac{1}{2}\Big((\Delta_n i\sigma_y)_{ss'}\, c^\dagger_{n,s} c^\dagger_{n,s'} + \text{h.c.}\Big) \\ & + \big(t_n (e^{i\boldsymbol{\lambda}_n\cdot\boldsymbol{\sigma}})_{ss'} c^\dagger_{n,s} c_{n+1,s'} + \text{h.c.}\big)\Big], \end{aligned} \tag{1}$$

where σ_i are Pauli matrices in spin space (indexed by s and s'), and n numbers QDs in the chain. For simplicity we assume uniform pairing $\Delta_n = 0.25$ meV, the Rashba vector in a form $\boldsymbol{\lambda}_n = \lambda_n[0, 1, 0]$, and the global Zeeman field $V_{\rm Z} = 0.5$ meV. The rest control parameters, i.e., $\{\mu_n,\, t_n,\, \lambda_n\} \equiv P$ (7 in total for $N = 3$ QDs chain) can be tuned electrically (via local gating). One set (yet uniform) of parameters includes: $\mu_n = 0.6\,$meV, $t_n = 0.25\,$meV, $\lambda_n = 0.27\,\pi$ we call *reference parameters*, P_0. They define so-called *sweet spot* leading to MZMs emergence, discussed in detail in Appendix A. The configuration of eigenstates for $N = 7$ and $N = 3$ is shown in Fig. 2(a) and Figs. 2(b,c) respectively. The Rashba length $\lambda_n = 0.27\,\pi$ was tuned so that at $\mu = 0.6$ meV two energy levels touch at zero energy (c.f. Fig. 2(b,c)).

TRANSPORT MEASUREMENTS

The analyzed system can be characterized via transport measurements. The conductance G through the QD chain is calculated using the S-matrix formalism in the wide-band limit [44] via the Weidenmüller formula [60, 61] for QDs coupled to normal leads:

$$S[H](E_F) = \mathbf{1} - iW^\dagger(E_F - H + \frac{i}{2}WW^\dagger)^{-1}W, \tag{2}$$

with the tunneling matrix defined as $W = \text{diag}(1, 0, 1) \otimes \text{diag}(\sqrt{\Gamma}, \sqrt{\Gamma}, -\sqrt{\Gamma}, -\sqrt{\Gamma})$ and the dot-lead coupling $\Gamma = 0.1$ meV. If we reshape the S-matrix accordingly: $S = S_{n,p,s,n',p',s'}$ ($n = L, C, R$ indexing dots, $p = 1, 2$ particle, and s—spin subspaces), then the respective reflection matrices: $s^{ee}_{s,s'}(i,j) = S_{i,1,s,j,1,s'}$ and $s^{he}_{s,s'}(i,j) = S_{i,2,s,j,1,s'}$ give differential conductance as,

$$\begin{aligned} G[S(E_F)]_{ij} = & \frac{\mathrm{d}I_i}{\mathrm{d}V_j} = 2\delta_{ij} - \text{tr}\big(s^{ee}(i,j)s^{ee}(i,j)^\dagger\big) \\ & + \text{tr}\big(s^{he}(i,j)s^{he}(i,j)^\dagger\big) \end{aligned} \tag{3}$$

in unit of e^2/h, $i, j = L, R$ denoting left (L) or right (R) lead—see Fig. 1(a), and E_F being the Fermi energy in the leads. Note that the same formula is implemented in the PD shown in Fig. 1(b). The input maps include 4 conductance G_{ij} components: G_{LL}, G_{LR}, G_{RL}, and G_{RR}, where for instance $G_{LL} = \frac{dI_L}{dV_L}$, with I_L denoting current through the left lead, and V_L is the bias voltage of the left lead. Similarly, other components can be defined by using different combinations of the leads. Notably, each conductance map is defined by 100×100 2D plot of G_{ij} as a function of some parameter and E_F. We utilize 4 maps per each component $G(E_f, \Delta l)$: 3 for $l = \mu_{n=L,C,R}$ variation (with respect to actual parameter P) and 1 for $l = V_{\rm Z}$ variation: 16 maps in total serving as the input. Figs. 2(d-f) shows the conductance maps (G_{LL} component) for the reference parameters P_0, highlighting the emerged zero-bias peak (orange arrow).

NEURAL ARCHITECTURE AND TRAINING

The proposed method uses the conductance maps tensor G as a high-dimensional visual input for the predictor NN (encoder), as well as to define the physics-inspired decoder—PD (teacher network) for unsupervised training. The model architecture is presented in Fig. 1(b). It

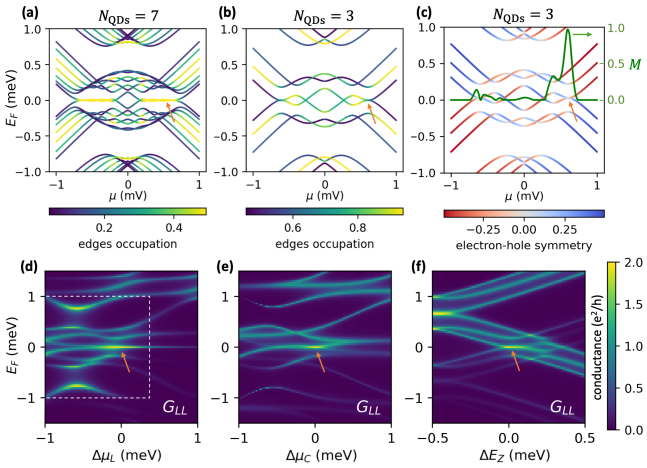


FIG. 2. Rashba chain of QDs proximitized by an s-wave superconductor. Top row: eigenstates as a function of voltage offset with colors encoding: (a,b) edges (L and R dot) occupation, (c) electron-hole symmetry. Reference parameters configuration P_0, giving MZM-like states, is marked by orange arrow. Bottom row: conductance maps as a function of (d) voltage applied to the left dot, (e) voltage on the center dot, and (f) the Zeeman field.

is of autoencoder-form with predictor NN coupled to PD network through parameters P *latent* space, and can be denoted as

$$\hat{G} = \mathrm{PD}(\mathrm{NN}(G) = P), \quad \mathrm{PD}(P) \equiv G[S[\mathcal{H}(P)]]. \qquad (4)$$

The encoder analyzes the G maps and predicts the parameter vector P; PD explicitly implements the transport physics using Eqs. 2 and 3, with the assumed systems family $\mathcal{H} = H^{\mathrm{QDs}}$, defined in Eq. 1. The objective is for the reconstructed map $\hat{G}$ (the PD output) to match the (encoder) input G—by minimizing loss function $\mathcal{L}(G, \hat{G}) = ||G - \hat{G}||^2$. We adapt a pretrained vision transformer (ViT) architecture [62] (patch size 16×16, 12 heads, 12 layers, $\sim$ 86M parameters) with input layers adjusted to process 16-channel ($16 \times 100 \times 100$) input tensor G and returning vector of the Hamiltonian parameters P. Training ($\mathcal{L}$ minimization) is performed for 300 epochs (with a properly tuned learning-rate scheduler) using 10-50k data points ($G(P_{\mathrm{in}})$ maps) for parameter vectors P_{in} sampled from different regions of the parameter space, as described in the Results section.

RESULTS

To demonstrate the validity of the PD concept, we performed HL of $\mathcal{H} = H^{\mathrm{QDs}}$ parameters, i.e. $P = \{\mu_{\mathrm{L}}, \mu_{\mathrm{C}}, \mu_{\mathrm{R}}, t_1, t_2, \lambda_1, \lambda_2\}$ for the triple-QD Rashba chain, as described above. After the training using synthetic $G(P_{\mathrm{in}})$ maps with P_{in} sampled from training sets denoted as white boxes in Figs. 3-6, the predictor NN

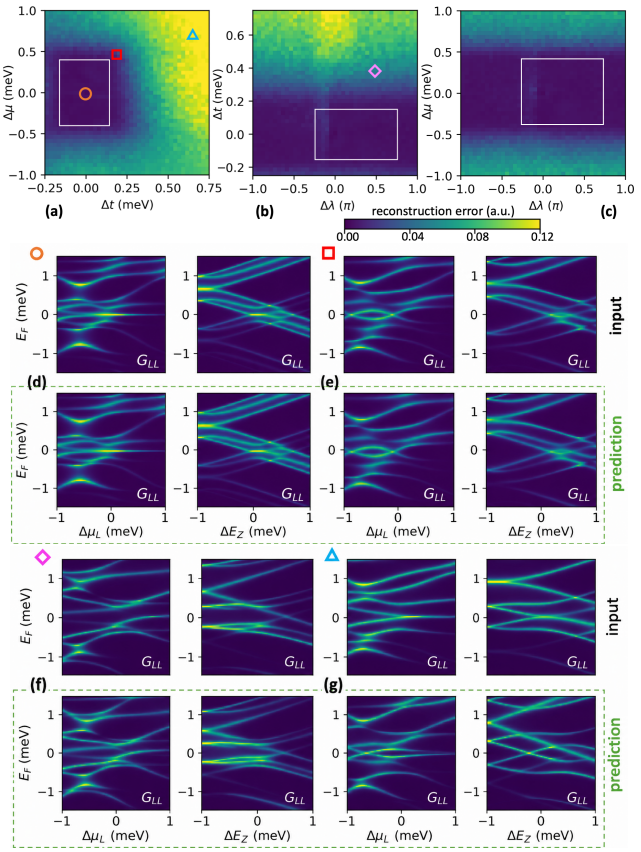


FIG. 3. Results for the HL of the Rashba chain, Eq. 1, of three quantum dots proximitized by an s-wave superconductor. Top row (a-c): error for the conductance maps reconstruction (via the predicted parameters) as a function of parameter variations—the white box denotes the region where training data was sampled. Middle (d,e) and bottom (f,g) rows: example input and predicted conductance maps (shown as middle/bottom pairs) for various parameter configurations indicated by the respective symbols. The default configuration that yields MZMs is marked with an orange circle

can predict the (learned) Hamiltonian parameters P for a given G. Sampling Δl in the range $[l_a, l_b]$ means that parameter $l \in [l_a, l_b] + P_{0,l}$, e.g., all μ_n were sampled from some $[\mu_a, \mu_b] + 0.6$ meV. The robustness of the predictor is shown in Fig. 3. The middle and bottom rows (d-g) show pairs of: input conductance maps (middle), and conductance maps *reconstructed* (calculated) using the predicted parameters (bottom). The better the parameter prediction, the closer the maps in each pair. The pairs presented are for different parameter configurations, indicated by the respective symbols in the top row (a-c). In the top row the reconstruction error $\langle |G - \hat{G}| \rangle$ is presented as a function of various parameters variation $(\Delta l_1, \Delta l_2)$ giving three different cuts in the parameter P space. Reconstruction error is averaged over all channels and pixels on G_{ij} maps, and over variations of different parameters of a given category (e.g., $\Delta\mu$ means all variations of po-

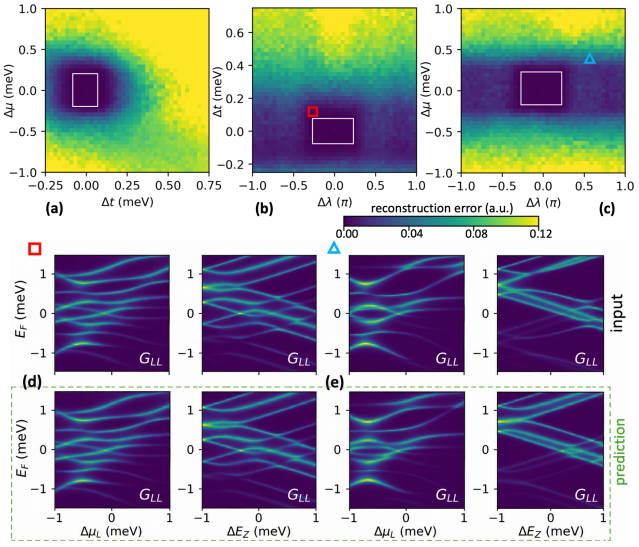


FIG. 4. HL results for a smaller training range, denoted by white boxes on the conductance reconstruction error maps (a-c) within the parameter space. Two specific G reconstruction cases (d,e) are indicated in the parameter space by a red square in (b) and a blue triangle in (c), respectively.

tentials $\Delta\mu_n$).

Inspection of the G_{LL} maps in Fig. 3 for parameters (d) within the training set (denoted by an orange circle on (a), corresponding to P_0), or (e) close to it (red square) shows that the reconstructed maps match the input maps almost perfectly. Moving slightly further away (f, pink diamond), small reconstruction errors begin to appear, although the overall structure of the conductance bands remains qualitatively well reproduced. However, when we move farther from the training set (g, blue triangle), the resulting structure of the conductance peaks in the G maps becomes noticeably different. These results demonstrate that the PD architecture is capable of learning the complex structure of the conductance features that arise even in this simple triple-QD system. More importantly, the model provides accurate predictions for regions of parameter space outside the training domain (outside the white rectangles in Fig. 3(a-c). This indicates that the physics-informed PD architecture possesses a strong ability to generalize. Importantly, the sampling of the 7-dimensional training space is relatively sparse: 10k samples correspond to approximately $(10^4)^{\frac{1}{7}} \simeq 3.7$ points per dimension.

To better investigate this property, we evaluated the NN predictor performance for a smaller training region—indicated in Fig. 4, and for a larger one—shown in Fig. 5. As expected, the smaller training region (reduced by a factor of 2^3) leads to accurate parameter predictions only within a correspondingly smaller volume of the parameter space P. Nevertheless, even in this case there remains a region outside the training set where the model generalizes well. Moreover, the ratio of this generalization volume to the training volume is slightly larger than in the configuration shown in Fig. 3. This can be attributed to the fact that, in the vicinity of the reference parameters P_0 (emergence of MZMs), the structure of peaks on the G map is relatively similar, making it easier for the NN to learn.

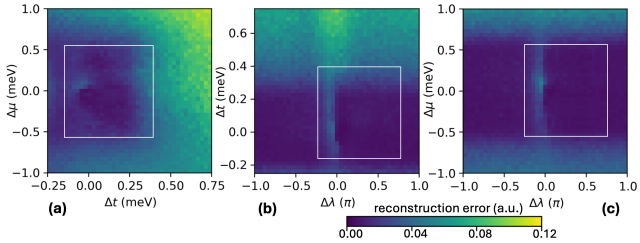


FIG. 5. HL results for a larger training range, indicated by wider white boxes on the conductance reconstruction error maps (a-c). In this case, 50k training samples were used.

In the opposite extreme, presented in Fig. 5—we observe that for a larger training volume (increased by a factor of $1.5 \times 2 \times 1 = 3$) the model correctly predicts the parameters $\mathcal{H}(P)$ over nearly the entire explored parameter space P. Importantly, in this case we increased the number of training samples to 50k (in the previous configurations shown in Figs. 3 and 4, we used 10k samples). However, as a characteristic trade-off, the price of covering such a broad range of G-maps variability is a slightly reduced average accuracy of parameter predictions, even within the training region itself. This clearly indicates that we are approaching the NN capacity limit.

It is also important to note that the inverse mapping from G maps to $\mathcal{H}$ parameters P is not guaranteed to be globally unique. In particular, symmetries and gauge freedoms of the Hamiltonian may lead to different parameterizations producing identical or nearly identical transport observables G. In our case, an example is the transformation: $t \to -t$ and $\lambda \to \lambda + \pi$. In the present approach, such ambiguities are mitigated by restricting the predictor output to physically and experimentally relevant parameter ranges $[l_a, l_b] + P_{0,l}$, thereby selecting a particular branch of the inverse problem. More generally, this restriction could be relaxed by employing a multi-head encoder that predicts several candidate parameter sets, allowing the network to explicitly represent multiple branches of a non-unique inverse map. Such a generalization would require further investigation; here, for simplicity, we restrict the predicted parameters to predefined physically relevant ranges.

RESILIENCE TO NOISE

In the proposed approach, we assume that the NN model is first trained on synthetically generated data and

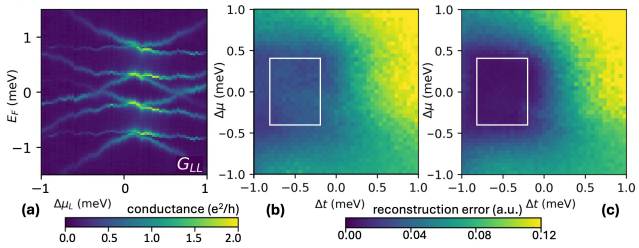


FIG. 6. HL from noisy data: (a) noisy conductance maps lead to (b) slightly degraded predictions. However, retraining the model with noisy samples included in the training data (c) restores the prediction performance.

subsequently transferred to experimental measurement data. However, experimental data may exhibit characteristics different from those of the synthetic dataset; in particular, they may contain noise.

Noise in conductance

To test this and emulate realistic experimental conditions, we augmented the synthetic conductance maps $G_{ij}(E_f, \Delta l)$ by including two different noise contributions, according to

$$\tilde{G}_{ij}(E_f, \Delta l) = G_{ij}^{\mathrm{warp}}(E_f, \Delta l) + \delta G_{ij}^{\mathrm{add}}(E_f, \Delta l), \quad (5)$$

where E_F is the Fermi energy in the leads and l denotes the detuned control parameter, corresponding either to the chemical potential of a selected quantum dot ($l = \mu_n$) or to the Zeeman field ($l = V_Z$). The term G_{ij}^{warp} describes distortions caused by charge noise and gate drift, while $\delta G_{ij}^{\mathrm{add}}$ accounts for additive instrumental noise. Both are discussed in detail in Appendix B.

We next examine the robustness of the proposed method to noise defined in this way—an example of a noisy conductance map $\tilde{G}$ is shown in Fig. 6(a). When testing the model presented in Fig. 3 on such noisy maps, the prediction quality naturally degrades—see Fig. 6(b). However, if the noise present in the experimental setup is well characterized, we can sample $\tilde{G}$ and add to the training data (10k samples were added). In Fig. 6(c) we show the reconstruction error for the NN model trained on the augmented conductance map set. It is visible that the model is now capable of predicting the $\mathcal{H}$ parameters even from the noisy conductance maps.

Note also that the PD architecture can be used for efficient GPU-parallelized synthetic data generation, as was done in this work for conductance map sets.

Hamiltonian-level disorder

Having established the robustness of the model against noise in the conductance maps, a natural next step is to examine its sensitivity to imperfections at the Hamiltonian level. This distinction is important because measurement noise modifies the observed signal, whereas Hamiltonian-level disorder introduces a mismatch between the physical system generating the data and the idealized Hamiltonian family assumed by the physics decoder. We consider two representative sources of such model mismatch. First, we introduce static sample-to-sample variations of parameters that are fixed in the decoder, namely the proximity-induced pairing Δ_n, the local Zeeman field $V_{\mathrm{Z},n}$, and the orientation of the Rashba spin-orbit field $\hat{\boldsymbol{\lambda}}_n$. These perturbations are kept fixed during the acquisition of each complete set of conductance maps and therefore represent device-level Hamiltonian disorder rather than measurement noise. Second, we account for cross-capacitance between electrostatic gates. Instead of assuming that detuning gate k modifies only the corresponding local potential μ_k, we introduce a capacitance matrix $\mathbf{A}$ such that $\delta\boldsymbol{\mu} = \mathbf{A}\mathbf{e}_k\delta V_k$, with random off-diagonal elements describing the response of neighboring QDs.

Importantly, in both tests the perturbed model, with disorder strength s defined in detail in Appendix C, is used only to generate the input conductance maps, while the physics decoder retains the ideal Hamiltonian $\mathcal{H}(P)$ of Eq. 1 and assumes perfectly local gate control. The resulting tests therefore directly probe the robustness of Hamiltonian inference against model mismatch not represented explicitly in the latent parameterization.

In Fig. 7, the conductance reconstruction error $\langle|G - \hat{G}|\rangle$, averaged over the same parameter ranges as in Fig. 3(a-c), is shown as a function of the disorder strength s (see Appendix C). It increases smoothly with s, as expected from the increasing mismatch between the Hamil-

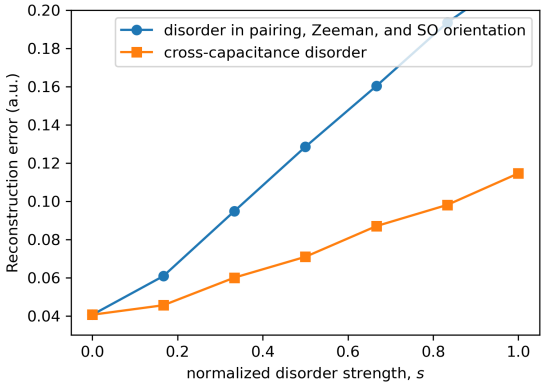


FIG. 7. Prediction robustness against disorder in Hamiltonian parameters outside the predicted parameter space P for two cases: (blue) local disorder in the pairing, Zeeman, and Rashba orientation terms; (orange) correlated detuning of the QDs. The disorder strength s is normalized to the range $[0, 1]$, with its detailed definition provided in Appendix C.

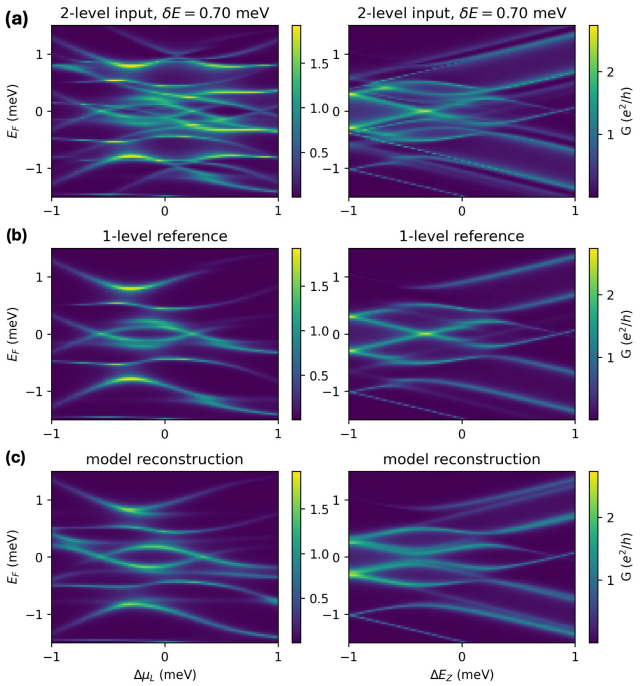


FIG. 8. Example conductance maps illustrating model predictions for an extended Hamiltonian with a second-level splitting of $\delta E = 0.7$ meV: (a) input maps generated with an additional orbital level on each QD, and (c) their reconstruction by the model retaining the single-level physics decoder; (b) reference generated with the single-level Hamiltonian.

tonian generating the data and the clean model assumed by the decoder. Importantly, the degradation remains gradual rather than abrupt, indicating that the learned inference is robust against moderate deviations from the assumed Hamiltonian family. This also suggests that the additional conductance features generated by the hidden degrees of freedom cannot simply be absorbed into some shifts of the seven inferred Hamiltonian parameters P.

Beyond the assumed Hamiltonian family

We finally consider a more stringent test in which the system generating the input conductance maps no longer belongs to the Hamiltonian family $\mathcal{H}$ encoded in PD. In a realistic device, additional degrees of freedom that are absent from the effective model may contribute to transport. In such a situation, the objective is not necessarily to recover the full microscopic Hamiltonian, which by construction cannot be represented by the decoder, but rather to determine whether the measured conductance can still be described by an effective Hamiltonian within the assumed model family. This observation also provides a useful interpretation of the reconstruction error. A failure to accurately reconstruct conductance maps generated by a Hamiltonian outside the assumed family does not necessarily constitute a failure of the inference procedure. Instead, a large reconstruction error can serve as an indicator that the assumed effective Hamiltonian is no longer adequate to describe the measured system.

To investigate this scenario in a controlled setting, we extend the Hamiltonian used to generate the input data by introducing a second orbital level on each QD, while keeping the Hamiltonian implemented in the PD unchanged. The energy of the additional level is shifted by δE relative to the original level. Thus, δE provides a continuous control parameter for the mismatch between the data-generating Hamiltonian and the single-level model assumed by the decoder. For large δE, the additional level is energetically separated from the relevant transport window and the original single-level description is recovered. Upon decreasing δE, the additional levels enter the transport window and generate conductance features that cannot, in general, be represented by the Hamiltonian family used by the decoder.

In Fig. 8(a), we show example conductance maps for a two-level system with $\delta E = 0.7$ meV, while the corresponding reference single-level maps are shown in Fig. 8(b). The additional orbital produces multiple resonant features that are absent from the corresponding single-level conductance maps. As can be seen in Fig. 8(c), when these two-level maps are provided as input to the network, the predictor does not collapse or produce an unphysical reconstruction. Instead, it successfully reconstructs a close conductance pattern accessible within its single-level Hamiltonian family $\mathcal{H}$. This is an encouraging result, as it shows that even when the input data contain features originating from degrees of freedom absent from the assumed model, the network can still identify a meaningful effective Hamiltonian within the model family encoded in the PD.

However, the remaining discrepancy between the input and reconstructed maps is reflected in an increased reconstruction error, presented in Fig. 9, providing a natural indicator that the assumed Hamiltonian family is incomplete. As δE increases in Fig. 9, the additional level becomes effectively decoupled and the reconstruction error approaches the single-level baseline (gray dashed line). At the same time, the parameter prediction error (orange curve), calculated as the normalized RMSE with respect to the reference single-level parameters $P[\mathcal{H}]$, $(\frac{1}{7}\sum_i (P_i[\mathcal{H}] - \hat{P}_i)^2/\Delta P_i^2)^{1/2}$, where ΔP_i denotes the natural range of the i-th parameter, increases by only about a factor of two. This indicates that, despite the mismatch with the assumed Hamiltonian family $\mathcal{H}$, the prediction of meaningful effective parameters does not collapse.

Finally, we also verified that the model can be successfully trained directly on two-level data with $\delta E = 0.7$ meV, while retaining the standard single-level Hamiltonian in the PD. The resulting reconstruction error is marked by the red star in Fig. 9. Interestingly, the model trained locally at a given δE achieves a slightly lower

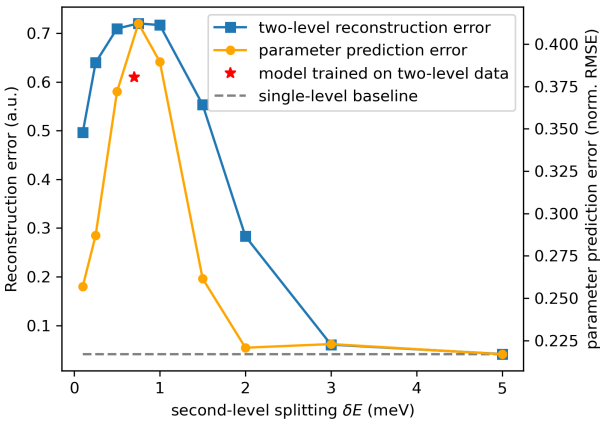


FIG. 9. Prediction beyond the $\mathcal{H}$ family assumed in the PD. Reconstruction (blue) and normalized parameter prediction (orange) error for a model trained exclusively on single-level data and subsequently tested on G maps generated with an additional orbital level on each QD, as functions of the second-level splitting δE. The red star shows the reconstruction error for a model trained and tested on two-level data at $\delta E = 0.7$ meV, while retaining the single-level $\mathcal{H}$ in the PD.

reconstruction error than the model trained exclusively on single-level data and subsequently tested on two-level data (blue curve).

ARCHITECTURE COMPLEXITY AND SCALABILITY

Predictor network complexity

To examine whether the performance of the ViT encoder originates simply from its large number of trainable parameters, we compared several encoder architectures spanning more than two orders of magnitude in model size, while keeping the physics decoder and training procedure unchanged. Fig. 10 shows the reconstruction error as a function of the number of trainable encoder parameters.

The results indicate that performance depends not only on model size but also strongly on the encoder architecture. A simple CNN yields the largest reconstruction error, while increasing the depth and complexity of residual CNN architectures leads to substantial improvement. Importantly, even the small ViT, with approximately an order of magnitude fewer parameters than the ViT used in the main calculations, outperforms the tested CNN-based architectures of comparable or larger size. This suggests different scaling behavior of transformer- and convolution-based architectures for the present conductance-map inference problem. At the same time, the ViT performance appears to approach saturation with increasing model size, at least for the

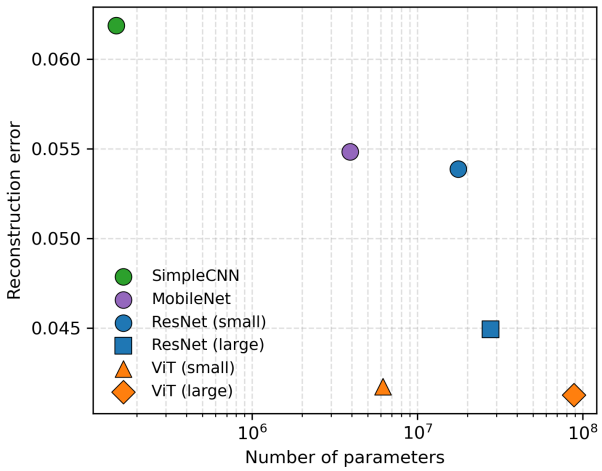


FIG. 10. Comparison of various NN architectures serving as the encoder. Predictor complexity is measured by the number of network parameters.

considered training-set size and corresponding problem complexity.

Model scalability

To assess the scalability of the proposed approach, we repeated the same learning experiment for longer QD chains, extending the system from three to five and seven QDs. The input and output dimensions were scaled accordingly, while the ViT (large) encoder architecture and training procedure were otherwise kept unchanged. Fig. 11 shows the reconstruction error averaged over the same parameter ranges as in Fig. 3(a-c). As expected, the reconstruction error increases with system size, reflecting the growing complexity of the inverse problem. The degradation is, however, gradual rather than abrupt: the five-QD system retains good reconstruction accuracy, while the seven-QD system shows a larger but still

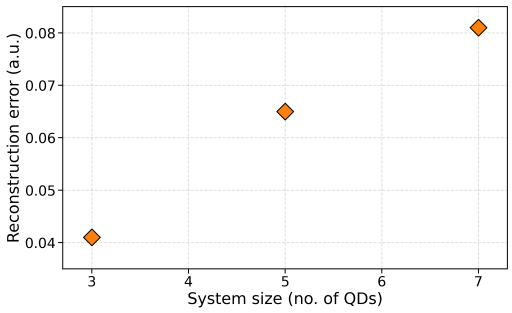


FIG. 11. Scaling with the QD-chain length. Reconstruction error obtained with the ViT (large) encoder for 3,5,7-QD systems, averaged over the same ranges as in Fig. 3(a-c).

controlled increase in error. These results indicate that the method remains applicable to longer chains without architectural modifications, although increasingly large systems are expected to require additional model capacity and/or training data.

SUMMARY

In this work we present an architecture and training algorithms for deep neural networks capable of identifying (predicting or discovering—depending on the application context) the parameters of Hamiltonians in quantum systems, including quantum simulators based on nanostructures. This work focuses on quantum simulators described by lattice Hamiltonians implemented in chains of quantum dots (QDs), together with their basic characterization through transport measurements. Specifically, it addresses the reconstruction of the effective Hamiltonian of a gated QD chain from conductance maps as a function of the applied gate voltages. We propose a natural mechanism for enforcing the underlying physics within the learning procedure. Moreover, the proposed scheme for defining and training a physics-informed neural network model can be applied to other areas of physics—wherever the underlying equations can be embedded in the form of differentiable formulas.

ACKNOWLEDGMENTS

We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2025/018433.

Appendix A: Chain of N Rashba quantum dots

Analytical considerations

We start with the Rashba-Zeeman-BCS Hamiltonian, defined in Eq. 1, written in the Nambu spinor representation $\hat{\Psi}_n = (c_{n\uparrow}, c_{n\downarrow}, c^\dagger_{n\uparrow}, c^\dagger_{n\downarrow})^T$ giving $H^{\mathrm{QDs}} = \frac{1}{2}\sum_{nm} \hat{\Psi}^\dagger_n H^{\mathrm{BdG}}_{nm} \hat{\Psi}_m$, where H^{BdG} is a first-quantized Hamiltonian, also called the Bogoliubov-de Gennes (BdG) Hamiltonian. It's explicit matrix form reads:

$$H^{\mathrm{BdG}} = \sum_{n=1}^{N} \begin{pmatrix} -\mu_n + V_{\mathrm{Z}} & 0 & 0 & \Delta_n \\ 0 & -\mu_n - V_{\mathrm{Z}} & -\Delta_n & 0 \\ 0 & -\Delta_n & \mu_n - V_{\mathrm{Z}} & 0 \\ \Delta_n & 0 & 0 & \mu_n + V_{\mathrm{Z}} \end{pmatrix}_{n,n} + \sum_{n=1}^{N-1} t_n \begin{pmatrix} e^{i\lambda_n \sigma_y} & 0 \\ 0 & -e^{-i\lambda_n \sigma_y^*} \end{pmatrix}_{n,n+1} + \mathrm{h.c.}, \tag{6}$$

with the onsite (dot) block $(\ \ )_{n,n}$ and the offsite hopping (inter-dot) block $(\ \ )_{n,n+1}$—both represented by 4×4 matrices. In Eq. 6 we assumed $\boldsymbol{\lambda}_n \cdot \boldsymbol{\sigma} = \lambda_n \sigma_y$, which follows from choosing the Rashba vector in a form $\boldsymbol{\lambda}_n = \lambda_n [0, 1, 0]$. For a system consisting of N quantum dots, the full BdG Hamiltonian H^{BdG} is a $4N \times 4N$ matrix.

Our goal is to simulate the Kitaev chain (KC) model, i.e., to tune the H^{BdG} (parameters) to mimic KC Hamiltonian form. The first step is that each QD should have a pair of zero energy levels $E_n = 0$. Diagonalizing a (separated) single dot block $(\ \ )_{n,n}$ yields the condition for this to happen: $V_{\mathrm{Z}} = \sqrt{\Delta_n^2 + \mu_n^2}$ or equivalently $\Delta_n = \sqrt{V_{\mathrm{Z}}^2 - \mu_n^2}$. When this condition is satisfied, each dot has a pair of fermionic excitations $a_n, a_n^\dagger$, where

$$a_n = \frac{1}{\sqrt{2V_{\mathrm{Z}}}} \left(\sqrt{V_{\mathrm{Z}} + \mu_n}\, c^\dagger_{n\uparrow} - \sqrt{V_{\mathrm{Z}} - \mu_n}\, c_{n\downarrow} \right). \tag{7}$$

The energy of a_n ($a_n^\dagger$) is zero. Two other modes are separated by the energy $\pm 2V_{\mathrm{Z}}$. Let's now assume that the hopping is much smaller than the energy of the excited state, $t \ll 2V_{\mathrm{Z}}$, we may project the Hamiltonian (6) onto the Hilbert space spanned by a set $\{a_1, a_1^\dagger, a_2, a_2^\dagger, \ldots, a_N, a_N^\dagger\}$. The resulting projected Hamiltonian is:

$$\tilde{H}^{\mathrm{BdG}} = \sum_{n=1}^{N-1} \frac{t_n}{V_{\mathrm{Z}}} \begin{pmatrix} -\mu_n \cos\lambda_n & \Delta_n \sin\lambda_n \\ -\Delta_n \sin\lambda_n & \mu_n \cos\lambda_n \end{pmatrix}_{n,n+1} + \mathrm{h.c.}. \tag{8}$$

The goal is to simulate the KC model, which, e.g. for $N = 2$ takes the form (in the same Nambu representation $(a_1, a_1^\dagger, a_2, a_2^\dagger)$ as $\tilde{H}^{\mathrm{BdG}}$):

$$H^{\mathrm{KC}} = \begin{pmatrix} \mu_1' & 0 & -t_1' & \Delta_1' \\ 0 & -\mu_1' & -\Delta_1' & t_1' \\ -t_1' & -\Delta_1' & \mu_2' & 0 \\ \Delta_1' & t_1' & 0 & -\mu_2' \end{pmatrix}. \tag{9}$$

KC Hamitonian, Eq. 9, at the *sweet spot* $\mu_n' = 0$, $t_n' = \Delta_n'$ possesses two MZMs. For the Rashba–Zeeman s-wave Hamiltonian $\tilde{H}^{\mathrm{BdG}}$, Eq. 8, to simulate H^{KC}, Eq. 9, at the sweet spot, the following condition must be satisfied by the Rashba length

$$\lambda_n = \arctan\left(\frac{\mu_n}{\Delta_n}\right), \tag{10}$$

which together with $\mu_n = \sqrt{V_{\mathrm{Z}}^2 - \Delta_n^2}$, gives the approximate position of the parameters sweet spot for the QDs Hamiltonian H^{QDs}.

Parameters sweet spot

If we put the assumed $V_{\mathrm{Z}} = 0.5$ meV and $\Delta_n = 0.25$ meV, $\sqrt{V_{\mathrm{Z}}^2 - \Delta_n^2}$ gives approximately 0.43 meV, and $\arctan(\frac{\mu_n}{\Delta_n}) \simeq 0.33\,\pi$. However, the numerically

tuned sweet spot values for $N = 3$ are slightly different: $\mu_n = 0.6$ meV and $\lambda_n = 0.27\,\pi$. These parameters are optimized to ensure that the formed MZMs exhibit both large edge localization and preserved particle-hole symmetry, as shown in Figs. 2(b,c) of the main text. The latter values are selected as the reference parameters, P_0 in the main text.

Appendix B: Conductance noise models

The synthetic conductance maps $G_{ij}(E_F, \Delta l)$ were augmented with two dominant noise contributions according to

$$\tilde{G}_{ij}(E_F, \Delta l) = G_{ij}^{\mathrm{warp}}(E_F, \Delta l) + \delta G_{ij}^{\mathrm{add}}(E_F, \Delta l), \quad (11)$$

where E_F is the Fermi energy in the leads and l denotes the detuned control parameter.

The warping contribution was implemented as a stochastic deformation of the detuning axis,

$$G_{ij}^{\mathrm{warp}}(E_F, \Delta l) = G_{ij}\big(E_F, \Delta l + \delta l(E_F)\big), \quad (12)$$

where $\delta l(E_F)$ represents a random shift of the effective detuning value for each scan line at fixed E_F (G map is composed of 100 lines). This shift was composed of a smooth drift component $\delta l_{\mathrm{d}}(E_F)$ and a random telegraph component $\delta l_{\mathrm{t}}(E_F)$:

$$\delta l(E_F) = \delta l_{\mathrm{d}}(E_F) + \delta l_{\mathrm{t}}(E_F). \quad (13)$$

The smooth drift component was generated as a low-frequency random process along the scan direction,

$$\delta l_{\mathrm{d}}(E_F) = \sigma_{\mathrm{d}}\, f(E_F), \quad (14)$$

where $f(E_F)$ is a zero-mean Gaussian random function smoothed along the E_F axis, with a correlation length of 5-30 scan lines, meaning that significant changes of $\delta l_{\mathrm{d}}(E_F)$ occur only after several consecutive lines rather than between neighboring conductance pixels; and σ_{d} controls the drift amplitude (randomly sampled between 0.2% and 2% of the detuning range).

The random telegraph component was defined as a piecewise-constant process,

$$\delta l_{\mathrm{t}}(E_F) = \sum_{n=1}^{N_{\mathrm{j}}} A_n\, \Theta(E_F - E_n), \quad (15)$$

where N_{j} is the number of telegraph jumps (randomly sampled between 0 and 3), E_n are randomly chosen scan-line positions, A_n are jump amplitudes uniformly sampled between 0.5% and 5% of the detuning range, and Θ denotes the Heaviside step function. Each term produces a sudden shift of conductance features for all subsequent scan lines.

The additive noise contribution was defined as

$$\delta G_{ij}^{\mathrm{add}}(E_F, \Delta l) = \eta_{\mathrm{w}}(E_F, \Delta l) + \eta_{\mathrm{c}}(E_F, \Delta l), \quad (16)$$

where $\eta_{\mathrm{w}}(E_F, \Delta l)$ is an uncorrelated Gaussian white noise and $\eta_{\mathrm{c}}(E_F, \Delta l)$ is a Gaussian noise correlated along the scan direction. The standard deviation of η_{w} was randomly sampled between 0.5% and 3% of the maximal conductance value $G_{\max}$. The correlated component η_{c} had correlation length between 2 and 10 scan lines and amplitude between 0.2% and 2% of $G_{\max}$. During training, all noise parameters were randomly sampled within these ranges.

Appendix C: Hamiltonian disorder parametrizations

The Hamiltonian disorder is introduced through static random variations of the local pairing, Zeeman energy, and Rashba spin-orbit orientation: $\Delta_n^{\mathrm{dis}} = \Delta_n\left(1 + s\sigma_\Delta \xi_n^\Delta\right)$, $V_{\mathrm{Z},n}^{\mathrm{dis}} = V_{\mathrm{Z}}\left(1 + s\sigma_{\mathrm{Z}}\xi_n^{\mathrm{Z}}\right)$, and ($\rho_j^{\mathrm{dis}} = \rho_j + s\sigma_{\mathrm{SO}}\xi_j^\rho, \chi_j^{\mathrm{dis}} = \chi_j + s\sigma_{\mathrm{SO}}\xi_j^\chi$), with (ρ, χ) being spherical parameterization of the $\hat{\boldsymbol{\lambda}}$ versor. All ξ are independent random variables drawn from the standard normal distribution, $\mathcal{N}(0, 1)$. We use $\sigma_\Delta = \sigma_{\mathrm{Z}} = 0.1$ and $\sigma_{\mathrm{SO}} = 10°$, while $s \in [0, 1]$ controls the overall disorder strength.

As a second test, we account for cross-capacitance between electrostatic gates. Instead of assuming that detuning gate k modifies only the corresponding local potential μ_k, we introduce a lever-arm matrix $\mathbf{A}$ such that

$$\delta\boldsymbol{\mu} = \mathbf{A}\mathbf{e}_k\, \delta V_k, \qquad \mathbf{A} = \mathbf{I} + \mathbf{C}, \quad (17)$$

where $\mathbf{e}_k$ selects the detuned gate. The diagonal elements satisfy $C_{ii} = 0$, while the off-diagonal lever arms are independently sampled according to uniform distribution,

$$C_{ij} \sim \mathcal{U}(-cs, cs), \qquad i \neq j. \quad (18)$$

We vary the maximum cross-capacitance strength from 0 to 0.2, i.e., $c = 0.2$ and $s \in [0, 1]$.

* jaroslaw.pawlowski@pwr.edu.pl

[1] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, Nature Physics **13**, 435 (2017).

[2] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine learning phases of strongly correlated fermions, Phys. Rev. X **7**, 031038 (2017).

[3] P. Zhang, H. Shen, and H. Zhai, Machine learning topological invariants with neural networks, Phys. Rev. Lett. **120**, 066401 (2018).

[4] Y. Mao, Q. He, and X. Zhao, Designing complex architectured materials with generative adversarial networks, Science Advances **6**, eaaz4169 (2020), https://www.science.org/doi/pdf/10.1126/sciadv.aaz4169.

[5] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, Schnet – a deep learning architecture for molecules and materials, The Journal of Chemical Physics **148**, 241722 (2018), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5019779/16655678/241722_1_online.pdf.
[6] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, Nature **624**, 80 (2023).
[7] A. K. Cheetham and R. Seshadri, Artificial intelligence driving materials discovery? perspective on the article: Scaling deep learning for materials discovery, Chemistry of Materials **36**, 3490 (2024).
[8] J. Qi, T. W. Ko, B. C. Wood, T. A. Pham, and S. P. Ong, Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling, npj Computational Materials **10**, 43 (2024).
[9] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nature Reviews Physics **5**, 141 (2023).
[10] L. Lunczer, P. Leubner, M. Endres, V. L. Müller, C. Brüne, H. Buhmann, and L. W. Molenkamp, Approaching quantization in macroscopic quantum spin hall devices through gate training, Phys. Rev. Lett. **123**, 047701 (2019).
[11] A. J. Lew, K. Jin, and M. J. Buehler, Designing architected materials for mechanical compression via simulation, deep learning, and experimentation, npj Computational Materials **9**, 80 (2023).
[12] K. Inui and Y. Motome, Inverse Hamiltonian design of highly entangled quantum systems, Phys. Rev. Res. **6**, 033080 (2024).
[13] M. Thamm and B. Rosenow, Conductance based machine learning of optimal gate voltages for disordered Majorana wires, Phys. Rev. B **109**, 045132 (2024).
[14] G. Lupi and J. L. Lado, Hamiltonian-learning quantum magnets with nonlocal impurity tomography, Phys. Rev. Appl. **23**, 054077 (2025).
[15] N. Karjalainen, Z. Lippo, G. Chen, R. Koch, A. O. Fumega, and J. L. Lado, Hamiltonian inference from dynamical excitations in confined quantum magnets, Phys. Rev. Appl. **20**, 024054 (2023).
[16] D. Feng, Z. Guan, X. Wu, Y. Wu, and C. Song, Classification of skyrmionic textures and extraction of Hamiltonian parameters via machine learning, Phys. Rev. Appl. **21**, 034009 (2024).
[17] R. Koch, D. van Driel, A. Bordin, J. L. Lado, and E. Greplova, Adversarial Hamiltonian learning of quantum dots in a minimal Kitaev chain, Phys. Rev. Appl. **20**, 044081 (2023).
[18] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, Experimental quantum Hamiltonian learning, Nature Physics **13**, 551 (2017).
[19] J. R. Taylor, J. D. Sau, and S. Das Sarma, Machine learning the disorder landscape of majorana nanowires, Phys. Rev. Lett. **132**, 206602 (2024).
[20] C. Bonizzoni, M. Tincani, F. Santanni, and M. Affronte, Machine-learning-assisted manipulation and readout of molecular spin qubits, Phys. Rev. Appl. **18**, 064074 (2022).
[21] P. Henderson, A. Ghazaryan, A. A. Zibrov, A. F. Young, and M. Serbyn, Deep learning extraction of band structure parameters from density of states: A case study on trilayer graphene, Phys. Rev. B **108**, 125411 (2023).
[22] M. Khosravian, R. Koch, and J. L. Lado, Hamiltonian learning with real-space impurity tomography in topological moiré superconductors, Journal of Physics: Materials **7**, 015012 (2024).
[23] D. Liu, A. B. Watson, M. Hott, S. Carr, and M. Luskin, Learning the local density of states of a bilayer moiré material in one dimension (2025), arXiv:2405.06688 [math-ph].
[24] Q. Gu, Z. Zhouyin, S. K. Pandey, P. Zhang, L. Zhang, and W. E, Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy, Nature Communications **15**, 6772 (2024).
[25] K. Choudhary, Slakonet: A unified slater-koster tight-binding framework using neural network infrastructure for the periodic table, The Journal of Physical Chemistry Letters **16**, 11109 (2025).
[26] M. Mattheakis, D. Sondak, A. S. Dogra, and P. Protopapas, Hamiltonian neural networks for solving equations of motion, Phys. Rev. E **105**, 065305 (2022).
[27] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, Advances in neural information processing systems **32** (2019).
[28] A. Mirani and P. Hayden, Learning interacting fermionic Hamiltonians at the Heisenberg limit, Phys. Rev. A **110**, 062421 (2024).
[29] D. Stilck França, L. A. Markovich, V. V. Dobrovitski, A. H. Werner, and J. Borregaard, Efficient and robust estimation of many-qubit Hamiltonians, Nature Communications **15**, 311 (2024).
[30] D. Hangleiter, I. Roth, J. Fuksa, J. Eisert, and P. Roushan, Robustly learning the Hamiltonian dynamics of a superconducting quantum processor, Nature Communications **15**, 9595 (2024).
[31] F. Borsoi, N. W. Hendrickx, V. John, M. Meyer, S. Motz, F. van Riggelen, A. Sammak, S. L. de Snoo, G. Scappucci, and M. Veldhorst, Shared control of a 16 semiconductor quantum dot crossbar array, Nature Nanotechnology **19**, 21 (2024).
[32] A. R. Mills, D. M. Zajac, M. J. Gullans, F. J. Schupp, T. M. Hazard, and J. R. Petta, Shuttling a single charge across a one-dimensional array of silicon quantum dots, Nature Communications **10**, 1063 (2019).
[33] A. Shandilya, S. Kapila, R. Krishnan, B. Weber, and B. Muralidharan, Unified simulation framework for experimentally-observed spin-valley locking in MoS2 quantum dots: Implications for qubit applications, ACS Applied Nano Materials **8**, 14949 (2025).
[34] V. Hernandes, J. Rogers, R. Koch, T. Spriggs, B. Undseth, A. Chatterjee, L. M. K. Vandersypen, and E. Greplova, Reconstructing quantum dot charge stability diagrams with diffusion models (2026), arXiv:2603.26432 [quant-ph].
[35] M. Leijnse and K. Flensberg, Parity qubits and poor man's majorana bound states in double quantum dots, Phys. Rev. B **86**, 134528 (2012).
[36] T. Dvir, G. Wang, N. van Loo, C.-X. Liu, G. P. Mazur, A. Bordin, S. L. D. ten Haaf, J.-Y. Wang, D. van Driel, F. Zatelli, X. Li, F. K. Malinowski, S. Gazibegovic, G. Badawy, E. P. A. M. Bakkers, M. Wimmer, and L. P. Kouwenhoven, Realization of a minimal kitaev chain in coupled quantum dots, Nature **614**, 445 (2023).

[37] A. Bordin, C.-X. Liu, T. Dvir, F. Zatelli, S. L. D. ten Haaf, D. van Driel, G. Wang, N. van Loo, Y. Zhang, J. C. Wolff, T. Van Caekenberghe, G. Badawy, S. Gazibegovic, E. P. A. M. Bakkers, M. Wimmer, L. P. Kouwenhoven, and G. P. Mazur, Enhanced majorana stability in a three-site kitaev chain, Nature Nanotechnology 10.1038/s41565-025-01894-4 (2025).

[38] I. C. Fulga, A. Haim, A. R. Akhmerov, and Y. Oreg, Adaptive tuning of majorana fermions in a quantum dot chain, New Journal of Physics **15**, 045020 (2013).

[39] M. M. Maśka and T. Domański, Polarization of the Majorana quasiparticles in the Rashba chain, Scientific Reports **7**, 16193 (2017).

[40] J. Darulová, S. Pauka, N. Wiebe, K. Chan, G. Gardener, M. Manfra, M. Cassidy, and M. Troyer, Autonomous tuning and charge-state detection of gate-defined quantum dots, Phys. Rev. Appl. **13**, 054005 (2020).

[41] J. P. Zwolak and J. M. Taylor, Colloquium: Advances in automation of quantum dot devices control, Rev. Mod. Phys. **95**, 011006 (2023).

[42] M.-A. Roux, J. Rivard, V. Yon, A. Morel, D. Leclerc, C. Rohrbacher, E. B. Ndiaye, F. F. Tafuri, B. Bono, S. Kubicek, R. Loo, Y. Shimura, J. Jussot, C. Godfrin, D. Wan, K. D. Greve, M.-A. Tétrault, D. Drouin, C. Lupien, M. Pioro-Ladrière, and E. Dupont-Ferrier, Rapid autotuning of a sige quantum dot into the single-electron regime with machine learning and RF-reflectometry FPGA-based measurements (2025), arXiv:2509.19537 [cond-mat.mes-hall].

[43] M. P. R. Losert, D. Denora, B. van Straaten, M. Chan, S. D. Oosterhout, L. Stehouwer, G. Scappucci, M. Veldhorst, and J. P. Zwolak, Automated electrostatic characterization of quantum dot devices in single- and bilayer heterostructures (2025), arXiv:2601.00067 [cond-mat.mes-hall].

[44] A. Donarini and M. Grifoni, *Quantum Transport in Interacting Nanojunctions* (Springer, 2024).

[45] G. E. Blonder, M. Tinkham, and T. M. Klapwijk, Transition from metallic to tunneling regimes in superconducting microconstrictions: Excess current, charge imbalance, and supercurrent conversion, Phys. Rev. B **25**, 4515 (1982).

[46] J. R. Taylor, J. D. Sau, and S. Das Sarma, Machine learning the disorder landscape of Majorana nanowires, Phys. Rev. Lett. **132**, 206602 (2024).

[47] J. R. Taylor and S. Das Sarma, Neural network based deep learning analysis of semiconductor quantum dot qubits for automated control, Phys. Rev. B **111**, 035301 (2025).

[48] D. van Driel, R. Koch, V. P. M. Sietses, S. L. D. ten Haaf, C.-X. Liu, F. Zatelli, B. Roovers, A. Bordin, N. van Loo, G. Wang, J. C. Wolff, G. P. Mazur, T. Dvir, I. Kulesh, Q. Wang, A. M. Bozkurt, S. Gazibegovic, G. Badawy, E. P. A. M. Bakkers, M. Wimmer, S. Goswami, J. L. Lado, L. P. Kouwenhoven, and E. Greplova, Cross-platform autonomous control of minimal kitaev chains (2024), arXiv:2405.04596 [cond-mat.mes-hall].

[49] J. R. Taylor and S. Das Sarma, Mitigating disorder and optimizing topological indicators with vision-transformer-based neural networks in Majorana nanowires, Phys. Rev. B **112**, L041110 (2025).

[50] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, arXiv preprint arXiv:2104.13478 (2021).

[51] M. Raissi, P. Perdikaris, and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics **378**, 686 (2019).

[52] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, Neural operator: learning maps between function spaces with applications to pdes, J. Mach. Learn. Res. **24** (2023).

[53] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, Physics-informed neural operator for learning partial differential equations, ACM / IMS J. Data Sci. **1**, 10.1145/3648506 (2024).

[54] M. Krawczyk, J. Pawłowski, M. M. Maśka, and K. Roszak, Data-driven criteria for quantum correlations, Phys. Rev. A **109**, 022405 (2024).

[55] M. Kliczkowski, L. Keyes, S. Roy, T. Paiva, M. Randeria, N. Trivedi, and M. M. Maśka, Autoencoder-based analytic continuation method for strongly correlated quantum systems, Phys. Rev. B **110**, 115119 (2024).

[56] P. Leszczyński, K. Kutorasiński, M. Szewczyk, and J. Pawłowski, Machine-learned models for power magnetic material characteristics, IEEE Transactions on Power Electronics **40**, 1554 (2025).

[57] M. Krawczyk and J. Pawłowski, AI-enhanced tuning of quantum dot Hamiltonians toward majorana modes (2026), arXiv:2601.02149 [cond-mat.mes-hall].

[58] M. Elhamod, J. Bu, C. Singh, M. Redell, A. Ghosh, V. Podolskiy, W.-C. Lee, and A. Karpatne, Cophy-pgnn: Learning physics-guided neural networks with competing loss functions for solving eigenvalue problems, ACM Trans. Intell. Syst. Technol. **13**, 10.1145/3530911 (2022).

[59] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation, Nature Computational Science **2**, 367 (2022).

[60] A. Bordin, C.-X. Liu, T. Dvir, F. Zatelli, S. L. D. ten Haaf, D. van Driel, G. Wang, N. van Loo, Y. Zhang, J. C. Wolff, T. V. Caekenberghe, G. Badawy, S. Gazibegovic, E. P. A. M. Bakkers, M. Wimmer, L. P. Kouwenhoven, and G. P. Mazur, Enhanced Majorana stability in a three-site Kitaev chain, Nature Nanotechnology **20**, 726 (2025).

[61] T. Christiansen and M. Zworski, A mathematical formulation of the Mahaux–Weidenmüller formula for the scattering matrix, Journal of Physics A: Mathematical and Theoretical **42**, 415202 (2009).

[62] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in *International Conference on Learning Representations* (2021).