Skolkovo Institute of Science and Technology

MODELLING OF EXCITON-POLARITON CONDENSATES FOR UNCONVENTIONAL COMPUTING

Doctoral Thesis
by

NIKITA STROEV

DOCTORAL PROGRAM IN PHYSICS

Supervisor
Professor, Natalia Berloff

Moscow – 2021

MODELLING OF EXCITON-POLARITON CONDENSATES FOR UNCONVENTIONAL COMPUTING

by

NIKITA STROEV

Tuesday 30th November, 2021 16:45

Submitted to the Skoltech center for Photonics and Quantum Materials
on October 2021, in partial fulfillment of the requirements for the
DOCTORAL PROGRAM IN PHYSICS

Abstract

In recent years, a significant demand for computational resources is observed, which resulted in much effort devoted to the theoretical simplification of complex problems and the development of various technological platforms to solve particular classes of hard problems. Exciton-polaritons appear to be a very promising physical system, serving as the perfect foundation for such technological advancement. The main research effort was focused on describing the correspondence between computational problems of high complexity and the physical system states. It was shown that using exciton-polaritons, one can realize k-local Hamiltonians with $k > 2$ with the nontrivial phase configurations. In addition to that, the novel contribution consists in the introduction of the complex coupling switching method, providing a way to significantly increase the success probability of solving optimisation problems using the exciton-polariton platform and applicable to gain-dissipative simulators in general. One can use this method as a useful heuristic on conventional computer architecture with the algorithmic perspective. Moreover, the existing correspondence between different computational tasks and presented the methodology of encoding/decoding of an arbitrary computational task into the optical/photonic hardware was considered. The most generic and complex machine-learning approach was considered with the potential architecture mapping. It was shown that using the nonlinear spin clusters, one can approximate the predetermined architecture with minor accumulated error, pushing the limits of the available computation. This new alternative method allows one to directly implement neural network algorithms on many condensed matter systems, with various advantages, such as reducing the overhead on the additional variables required to implement a more conventional approach for neural network realisation. Since the exciton-polariton has promising and alluring properties with the perspective technologies, the research on the potential applications besides the existing ones was carried out, emphasizing periodic structures and its analytical description. Emphasizing the analytic forms, the introduced approach allows one to identify how the velocity profile of the condensate changes with parameters, such as the trapping and dissipation potentials, which allows avoiding extensive calculations. The behaviour and the phase diagram were built, which opens the way to the controllable laser or polariton flows for ultra-fast information processing and analog simulators. Summarizing, one can say with complete confidence that exciton-polaritons are a promising platform that is not exploited to its full potential.

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Nikita Stroev)

Supervisor (Prof. Natalia Berloff, PhD)

Publications

Main author

1. Nikita Stroev and Natalia G Berloff. Managing the flow of liquid light. *Physical Review B*, 102(20):201114, 2020

2. Nikita Stroev and Natalia G Berloff. Discrete polynomial optimization with coherent networks of condensates and complex coupling switching. *Physical Review Letters*, 126(5):050504, 2021

3. Nikita Stroev and Natalia G Berloff. Xy neural network. *arXiv preprint arXiv:2103.17244*, 2021

4. Nikita Stroev and Natalia G Berloff. Modern optical computing. in preparation, 2021

Acknowledgments

First of all, I would like to express my gratitude to my supervisor Prof. Natalia Berloff for her guidance and inspiration during PhD study, discussion of the results, help with the publication process and maintaining exciting group seminars. These meetings inspired many educational and productive discussions, as well as future joint projects.

Secondly, I would like to thank Skoltech PhD students and postdocs from CPQM and CDISE centres with whom I participated in various courses and projects, which were the sources of tremendous experience. Moreover, I would like to thank RQC researchers for valuable discussions and their perspectives on the area of research.

I am also grateful for the constructive feedback provided on the annual reviews in the CPQM centre and express appreciation to Skoltech for providing research opportunities.

Personally, the author would like to thank Prof. Alexei Buchachenko for his great help in solving organizational issues and support.

Additionally, I am thankful to Jury Members: Dr. Jonathan Keeling, Prof. Sergei Turitsyn, Prof. Nikolay Gippius, Prof. Pavlos Lagoudakis for the valuable feedback that allowed me to get an independent perspective on my work and consequently improve its quality and additionally to Prof. Ildar Gabitov, who agreed to be the Chairman of the Defense process.

More importantly, I would love to thank my mother and relatives for their tremendous support and encouragement over the years of my educational process.

Contents

List of Figures

List of Tables

List of acronyms

BCS Bardeen–Cooper–Schrieffer.

BEC Bose–Einstein condensate.

BKT Berezinskii–Kosterlitz–Thouless.

BN Bayesian network.

BP belief propagation.

CC complex-coupling.

cGLE complex Ginzburg-Landau equation.

CIM coherent Ising machine.

CS computer science.

DBR distributed Bragg reflector.

DL deep learning.

EP exciton-polariton.

FPGA field-programmable gate array.

GD gradient descent.

GP Gross-Pitaevskii.

HOBO high-order binary optimization.

ILP integer linear programming.

KL Kullback-Leibler.

kNN k-nearest neighbors algorithm.

ML machine learning.

MP message passing.

MRF Markov random field.

NBU networks of bistable units.

NFLT no free lunch theorem.

NN neural network.

NP nondeterministic polynomial time complexity class.

OPO optical parametric oscillator.

P PTIME complexity class.

PCA principal component analysis.

PGM probabilistic graphical model.

PP polynomial programming.

PUBO polynomial unconstrained binary optimization.

QP quadratic programming.

QUBO quadratic unconstrained binary optimization.

QW quantum well.

RC reservoir computing.

SAT Boolean satisfiability problem.

SLM spatial light modulator.

SVD singular value decomposition.

SVM support vector machine.

TGD tensor gain-dissipative.

TSP travelling salesman problem.

VLSI very large-scale integration.

WPE Wishart planted ensemble.

Chapter 1

Introduction

My interest in this PhD thesis topic was ignited by recent development in the domain of quantum information science or, more specifically, its subdomain quantum computing, which has tremendously changed during the period of the study. The motivation was to get involved in physical research (due to my background) with computer science (CS) elements; however, I had no desire to rush into the mainstream quantum computing research. Many variations of quantum hardware present a unique and specialized problem solvers. If one wants to obtain the universal quantum computer, it needs to break through a set of fundamental and engineering challenges. Additionally, the breakthroughs were coming from the major corporations, which had a tremendous amount of resources to invest. The best choice for me was to contact Professor Natalia Berloff, who was working at Skoltech CPQM and who with the coworkers recently published a paper establishing the connection between the condensed matter system of Bose-Einstein condensate Bose–Einstein condensate (BEC) of exciton-polariton (EP) and XY Hamiltonian, which was the significant step for computation on such systems. She let me join her group and start PhD research. The developing alternative technologies, like computation based on the BEC physical platform, gave a good perspective on this specific domain and the mainstream field of quantum computing in general as a part of the obligatory review work. My motivation was heated up by the practical significance of the work and lots of potential available projects.

Since the main research direction was already established, my first task was to investigate the novel regime in the system, achieved by the significant change of external pumping and consider the corresponding mathematical optimization problem, which was solved by the system and was the result of several wavefunctions intersection (see 3.2). Such an exotic complicated state was possible due to the matter-wave nature of the EP BEC. Besides the alluring physical parameters, it has many exciting features, which we can take advantage of, for example, its intrinsic wave nature.

Besides, many physical systems, like quantum gases and ultracold atoms in optical lattices, are investigated with similar applied purposes. There are many similarities between the alternative physical systems, which spawns similar perspectives on how to utilize them in the most efficient manner, which makes this field very rich in ideas and models.

I also was involved into the more practical task, like the description of the periodic structures and the velocity profile of the BEC condensate inside such structures (see 3.7), which brings more insights into the knowledge of the physical phenomena and a better understanding of the difference between the theoretical propositions and its experimental realization.

Despite the academic workload, I managed to take the additional course on Foundations of Data Science, which resulted in independent research's exciting byproduct. It is considered in the 3.4, where I describe the correspondence between the classical XY Hamiltonian and the typical machine learning (ML) architecture, which was found using the approximation techniques and gave several exciting outcomes. Pattern recognition is a subdomain of ML and appears to be a general universal task of particular practical interest. It can be performed using the Ising problem formulation, and it's similar to the particular Hopfield networks; however, I tried to avoid trivial combinatorial results of variational nature and focused on the more difficult tasks of correspondence between the specific set of nonlinear functions and mathematical operations, sufficient for the ML.

The development of the ML ideas, together with the investigation of the system's physical properties, led me to quite an exotic topic, which holds the name of reservoir computing (RC). Despite its bizarreness and lack of good unification theory, it seems that it is a promising way to squeeze the system to maximize effectiveness in terms of computation. Nevertheless, it lacks a universal theory based on the recurrent NN description, which makes it harder to understand or map to more traditional tasks.

The general problem I am trying to solve was evident initially but condensed to the final statement through the

research process and became more explicit. There is an increasing demand for computational tools, theoretical tools (algorithms) and physical hardware, which allows one to speed up several tasks of particular interest, which grows across different domains, both academic and industrial. In this thesis, I aim to exploit the particular EP BEC system to solve several problems in the most efficient way. Conventional architectures for modern digital computers are suffering from modern algorithms' scalability problem, which increases the demand for computational resources.

To summarize, every system can be exploited in terms of computation. The best way to solve the problem of mapping a particular task into the hardware system is to use strict formalization of this task and the precise mathematical tools to analyze it and the system's possible states and behaviour. In general, it is possible to connect many CS task with every existing platform. However, the question remains to what degree this connection will be efficient. To exploit the system, one needs to achieve the tradeoff between the complicated computation model and the experimental limitations.

1.1 Thesis structure

The thesis consists of an introduction, three main chapters, represented in Fig. 1-1 a conclusion, additional resources and a bibliography.

Chapter 1 - Introduction. This first chapter is an introduction section, which includes the small context of the personal research, the thesis organization, a brief description, and the literature review's references, the background of the general problem aimed to be solved. The thesis objectives are defined and the consequent research methodology, ending with the scientific novelty results.

Chapter 2 - Physical preface. This section considers the brief history of the middle of the previous century's condensed-matter, emphasizing the superfluidity and BEC theory. The description of the transition to the particular system is followed, where the BEC realization is possible - mainly the EP quasiparticles. I cover the basic description of such a system, present its short experimental history and state of the art research with the latest various applications. This section's primary focus is devoted to the physical properties of the system and physical models since the following chapters will primarily operate with the mathematical models. In the end, the alternative platforms are presented, which compete with the BEC of EPs for the special-purpose hardware systems.

Chapter 3 - System design. This chapter is devoted to the methods of the EP system adjustment for the realization of different tasks. The basic block describes the mechanism behind XY Hamiltonian's realization of the system obeying the Gross-Pitaevskii (GP) description. The next part is the generalization of the previous method, resulting in the tensor sum's minimization, achieved within a similar system but with different parameters. In the following part, I will demonstrate how to perform the transfer of ML tasks into the XY Hamiltonian (or similar system, where it is possible to reproduce such a model) in an analogue approximation way. The general description of the computational framework of reservoir computing and the particular systems is followed, where this way of computation has been realized. The last part of this section is devoted to engineering periodic structure's applied task for reproducing different dynamical states and specific velocities parameters.

Chapter 4 - Algorithmic network. This chapter is aiming at utilizing the established connections (encoding-decoding procedure) between different CS tasks. Using such a graph, one can map the assignment of particular interest in the presented special purpose-hardware even if it was not designed for that specifically. It is followed by a discussion of the applicability of such a diagram for different platforms. The next part discusses the fundamental building block in CS, which is the SAT and its potential, i.e. how one can reduce many practical assignments into this basic form. The final block is devoted to calculating the general error accumulated in the transferred deep learning (DL) architecture.

Chapter 5 - Conclusion. The final chapter summarizes the obtained results and consists of discussing the practical and theoretical significance of the work.

Chapter 6 - Additional resources. This chapter contains the supplementary information for the thesis, including the description of several hardware platforms suitable for the same computational tasks as EP system and some perspectives on constructing hard optimization instances.

1.2 Literature review

Since this work is multidisciplinary by definition, it is essential to give the research scope from different perspectives with the additional literature reviews. This fact results in several parts located across the thesis, each covering its question if one wants to familiarize oneself with the presented supporting literature. First, the next section will present the general

background of the problem, which I am trying to solve utilizing the particular physical system, besides other different approaches. The history of the BEC and its realization in the EP systems is given in the next chapter. The review of the state of the art experiments is given afterwards. The corresponding experiments on different platforms, utilized in the same applied manner, is going afterwards. The brief algorithmic perspectives are given in the same background section, while the general scope and the connections between algorithms are presented in the Section 4.

1.3 Background of the problem

The general definition of a computation can be treated as a purely physical phenomenon, which occurs inside a particular physical system. Such a system is usually referred to as a calculating machine or currently known as the computer. Some computation can be performed on different physical systems that include digital computers, mechanical computers, and even exotic ones like quantum computers, molecular computers, microfluidics-based computers. The dominant paradigm for the computational system is modern stored-program computers. They are instances of a specific form of an abstract machine, a random-access stored-program machine model. This abstract machine is close to the Turing machine's formal concept, giving birth to the computational complexity theory and the famous P vs NP problem.

The intrinsic complexity of many objects and processes around us, which can be hardly grasped with a single analytical formula, makes the computation a routine and necessary procedure in describing the external world. Thus, computational tools are in great demand in scientific research, like chemistry and other domains, like modern industrial design. However, all of the corresponding problems are suffering from the scalability of the modern algorithms, which increase the power and time consumption per sequence of computing operations. All of these challenges must best be solved, which resulted in the emergence of multiple approaches and research directions.

This thesis aims to utilize a specific condensed-matter system to solve several generic algorithmic problems and try to do it most efficiently. However, if one needs to solve computational problems efficiently, it is possible to use alternative approaches to exploit the physical hardware. Thus, I distinguish three main perspectives that deal with the modern challenges of computation. In general, I greatly rely only on the last one (based on the physical hardware itself). However, others give general perspectives on the research background, which greatly affected this work and interests.

The first way to deal with complex problems (particularly NP-hard problems) is to use so-called approximation algorithms. They are efficient algorithms that find approximate solutions to problems (often of optimization type) with additional provable guarantees on a specific solution's particular metric to be the optimal one. The origins of this approach can be traced with the use of [5, 6, 7, 8]. Approximation algorithms are usually studied in theoretical computer science due to the famous P vs NP conjecture, which states that a broad class of optimization problems is impossible to solve exactly in polynomial time. Besides, approximation algorithms try to understand to which extinct it is possible to approximate optimal solutions. The old works about the approximate algorithms on the particular famous tasks, like travelling salesperson problem, bin packing and MAX CUT [9, 10, 11] are still relevant up to date.

The approximation techniques take advantage of the mathematical structure of the considered object. It is of particular use when the problem can be represented in a simple, clear, mathematical object. Many problems pose such a structure which sometimes is not seen at first sight. Thus the approximation play increasingly important roles in computational mathematics and numerical analysis [12, 13, 14]. One of the common mathematical objects in such tasks is matrices and their multidimensional generalizations - tensors, which appear across different domains. Several methods allow one to effectively represent a tensor using fewer parameters because problems of large sizes cannot be handled by direct numerical methods due to the so-called curse of dimensionality. Among them are Tucker decomposition [15], singular value decomposition (SVD) generalization [16], which is a subtype of the previous method, Tensor-Train decomposition [17], Tensor-Ring decomposition [18], for a review see [19, 20, 21].

The general approach appeared to be very fruitful. For example, the implementation of Tensor-Train decomposition has many successful cases [22, 23]. For example, Tensor-train's application to compress NN architecture leads to the dramatic reduction of its parameters [24]. It is essential to emphasize the chosen difference between the presented approximation methods and the next part, which exploits the statistical physics for investigating the algorithmic properties. These two approaches can be considered very similar since both of them are getting rid of small parameters in a particular problem but differently. However, each case's approximation comes from a different source since the first approach is essentially mathematical, while the second relies on physical insights.

One way to deal with complex algorithms is to use methods from statistical physics. In general, the statistical approach was long ago used in the CS domain. However, statistical physics has its unique tools that are different from standard statistics yet valuable for algorithmic analysis. The significant interest in adapting techniques from statistical physics, particularly mean-field theory, can be traced until the late 1990s. The initial idea was to create deterministic heuristic algorithms to approximately solve several optimization problems [25, 26, 27, 28, 29]. This set of algorithms

are usually referred to as deterministic annealing and has intrinsic connections with the simulated annealing [30, 31]. The general approach matches the particular optimization problem with the cost function and defines the corresponding Gibbs distribution.

Simulated annealing samples, the Gibbs probability distribution with the corresponding temperature-analogue parameter is reduced to zero. In contrast, the deterministic annealing attempts follow the approximate value of the mean of the distribution. The additional usefulness lies in the simplicity of translating them into the special-purpose hardware like very large-scale integration (VLSI) chips [32] or field-programmable gate array (FPGA). At the beginning stage, the mean-field approach's effectiveness was shown heuristically, while the theoretical analysis came later.

The paper [33] gives the historical remarks on the earliest period of such connections between mean-field theory methods and algorithmic applications. Additionally, it considers explicitly several particular algorithms [34, 35, 36, 37, 38, 39, 40] with the particular emphasis on the mathematical aspects, like convergence and bounds on the convergence times and relations to other optimization algorithms.

The ideas behind this interesting connection between the algorithms and physics did not vanish and reappeared with the problems coming with the enormous amounts of data in the modern setting. One of the main algorithmic types is inference, aiming to extract as much information as possible from the noisy data. It possesses a lot of challenges, resulting in the birth of the state of the art algorithms [41]. The inference theory's central questions are at what conditions do the measurements provide enough quality information, sufficient for a fair inference, and the most efficient algorithms for this task? The community's accumulated physical insight allows one to understand and locate these challenges by using analogies with the phase transitions in the same sense used in statistical physics. This relation is discussed in various sources [42, 43, 44, 45] and source with an emphasis on spin glass theory [46].

Understanding such problems came from the merging of two large domains that appeared to be very fruitful. Among the major works and significant results, one can stress the study of the stochastic block model and the problem of inferring functional groups or communities from the topology of the network with the emphasis on the properties of the detectability/undetectability phase transition or the easy/hard phase transition [47]. The study of the two prototypical random k-SAT and q-coloring of random regular graphs problems (random constraint satisfaction problems) resulted in the formal definition of each phase transition in terms of different notions of correlation between distinct variables [48] and highlighting domain of the effective parameters and the study of the cavity method, which resulted in the final formula for the mutual information in statistical inference problems, coming from the graphs and the corresponding phase transitions in random graph models [49].

The current interdisciplinary field of study is gaining momentum after these successful results. The latest ML results across different domains [50, 51, 52, 53] can be related to the same field and possess even more challenging and exciting questions, where the known methods and tools can be used with the same efficiency. For example, the idea of the general landscape (or multidimensional surface of the particular problem) of the cost function is still alive and shows itself in the ML problems [54]. There are many more questions within the domain of DL [55, 56], and some problems are hard to formalize, giving the space for new future directions.

The central part of the thesis will be devoted to the last approach, i.e. exploiting the particular physical system with its advantages. Various technological platforms act as quantum or classical analogue simulators to solve certain classes of hard classical optimization problems [57, 58, 59, 60, 61, 62, 63].

It is expected that these kinds of platforms would help to efficiently solve many tasks of significant computational complexity, ranging from modeling microscopic effects and processes like the behavior of electrons in complex materials [64, 65] and finding the ground state of spin glasses [66], to the applied combinatorial optimization problems such as the travelling salesman problem (TSP) [67].

In some cases, it is useful to take advantage of the characteristic speed of a platform [68, 69] or use the benefits of essentially different type of computation, like the intrinsic parallelism coming with the qubits [70, 71], for which the so-called "quantum supremacy" on a particular task was claimed to be achieved [72]. Even after the field of quantum computing is slowly coming to saturation, there are even more bizarre ideas, like topological quantum computing [73].

There are particular kinds of systems, which dynamic behaviour can be exploited. The idea is to use the system as a type of recurrent NN called the reservoir computing [74, 75, 76, 77, 78]. Together with the broader scope of platforms, see further in Section 5.

Several methods to deal with the high complexity algorithmic problems and their demanding nature in terms of power and speed were presented to summarise this section. Each of the approaches possesses good advantages, based either on the intrinsic mathematical properties of the problem or the characteristic values of the considered physical platforms exploited in the computational domains. In that sense, high-performance computing is developing from both theoretical and practical sides, satisfying rapidly growing demands. However, to reach maximal efficiency, one needs to utilize the effectiveness from all perspectives. That means that the best way is to create a hierarchical approach to the

problem - use all three methods in a combination. For example, the first stage is to use the statistical results and analyze the task with the given data approach; the second is to use the approximation techniques after the task is analyzed and then try to map the resulting problem into a particular hardware system. The general perspective is that one needs to utilize all the available tools in the most efficient manner.

1.4 Thesis objectives

The goals/questions of the PhD thesis are the following:

• Investigate the influence of the various EP system manipulation techniques (using numerical/analytical methods) on the realization of a particular kind of computation and to what degree it affects the final computational task;

• How can one design the physical system to solve more tasks of different nature and complexity? To what extent can we exploit the given system and the most complex task we can solve?

• What kind of applications besides the existing ones we can engineer?

• What kind of periodic structures can it be realized in the EP settings, and how to manipulate their parameters? What kind of behaviour can we expect?

• Explore in the general sense the existing connections/correspondence between different computational tasks and give a brief picture for practical tasks with the available knowledge about the current EP setting.

• Present the methodology of encoding/decoding an arbitrary computational task on the EP system and give several methodological examples.

1.5 Research methodology

This section covers the main methodological tools used in the current research throughout this thesis.

The standard methods of applied mathematics were extensively used to derive analytical equations, see for example, the rate equations in Section 3 for tensor sum minimization tasks. Approximation techniques for solving differential equations or finding approximate solutions were used. For the particular one-dimensional nonlinear Schrodinger equation, the property of the nonlinear basis of Jacobi elliptic functions was exploited.

The majority of the code for modelling purposes was written in the C++/Python programming languages. Several cases involved the nonlinear system of differential equations numerical integrations. In contrast, the modelling of the GP equations involved the finite difference schemes approximation techniques, using either Euler numerical procedure of first order for time and the second order for the spatial Runge-Kutta with the fourth-order for time and second order for the coordinate derivative.

In many parts, the differential equations similar to the gradient descent or annealing techniques were used for testing them on the particular set of generated instances. gradient descent (GD), Hopfield neural network (NN) [25], Gain-Dissipative algorithm [79] and complex-coupling (CC) post-processing [2] methods were used.

The Pytorch library [80] was extensively used in working with simple NN architectures. Moreover, the analogue set of computational blocks was written to transfer the given NN architecture into the XY model and outputs the corresponding parameters.

The computational assignments analysis was performed using the encoding/decoding techniques, with the particular emphasis on the Ising/quadratic unconstrained binary optimization (QUBO) formulations.

1.6 Scientific novelty

All results obtained in this thesis are new. The following main results were obtained:

1. A new way to adjust and hence to complicate the system of EP condensates was shown, which results in the increasing complexity of the computational task. The connection between the physical parameters and the complexity of the assignment was explicitly demonstrated.

2. A new algorithmic heuristics, which we refer to as complex-coupling (CC) switching was obtained. The initial idea of the procedure is to be applied to the experimental setup to enhance its performance. The heuristic itself was shown to outperform the conventional Hopfiled NNS on instances of the random tensor sum minimization.

3. Using the approximation techniques, the connection between the mathematical operation and the solution to the XY Hamiltonians equilibration cluster was established. It allows one to transfer any arbitrary ML architecture to any condensed matter system to realize XY Hamiltonian and potentially serve as the pre-trained ML hardware. A new methodology for such transfer was reported.

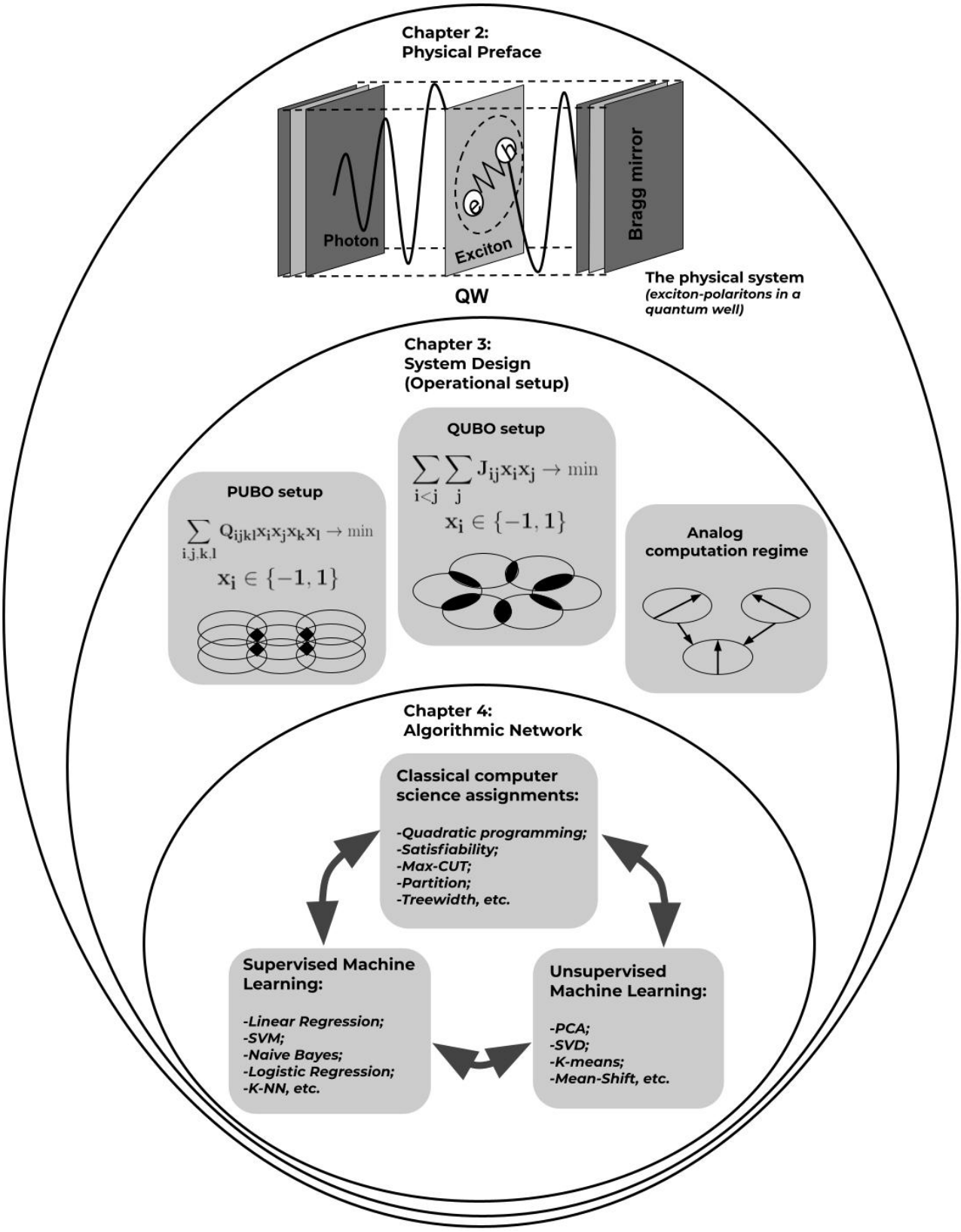


Figure 1-1: The scheme represents the general structure of this thesis. Chapter 2 discusses the physical properties of the EP system. Chapter 3 investigates the possible operational way of manipulating the considered system. Finally, Chapter 4 is focused on several suitable operational schemes and their connections with the applied assignments coming from the CS domain.

4. The analytical description of periodic structures in the EP settings was obtained. A new phase diagram for the condensate behaviour was built, and the range of possible parameters, including the velocity profile, was derived.

5. A network of the variety of computational tasks was summarized. Using this correspondence, one can encode/decode the assignment of particular interest in the task for which the particular hardware was designed.

This thesis's results are primarily theoretical but of great importance for practical applications, with the significant connections being explicitly demonstrated and discussed. They may be helpful to specialists working in condensed matter theory and experiments, emphasizing the EP system or other photonic hardware. The ML section is quite generic and appears useful for a wide range of systems, not to mention the established approximation itself. Many mathematical details and a specific derived heuristic algorithm will help computer scientists or researchers work in unconventional computing or alternative technologies for computing purposes.

Chapter 2

Physical preface

This section of the thesis is called the physical preface, introducing the reader into the history of the physical ideas of the 20th century developed mainly in the condensed matter domain with the emphasis on superfluidity and BEC state of matter. Simultaneously, it appears to be the sole chapter where the connection between the manifestation of physical phenomena and models can be easily traced and remains obvious through text. After the historical remarks, the chapter focuses on the BEC systems' physical properties and its Gross-Pitaevskii (GP) description. Finally, the chapter's ending presents the state-of-the-art experiments on the BEC of EPs and a short review of the existing platforms competing in unconventional applications.

The presented history and the beginning of this chapter are built around physical models, which came through developing the condensed matter field and attempts to understand the related concepts of nature. The official definition of a model is a simplified object that preserves only the essential properties of a real existing object or system and is intended for their study. We emphasise the importance of models in the research settings since the following chapters will be mainly devoted to studying them. The main scientific activity by itself is the understanding of the world through models. Besides, it aims to make a particular part or feature of the world more accessible to understand, define, quantify, visualise or simulate by referencing the existing and usually commonly accepted knowledge.

The engineering and design tasks appear when we want to take advantage of the particular systems using the obtained knowledge about them, given many options for combining known elements into something useful. Since the CS field is dealing with computational systems, it has close interconnections with the engineering devices based on tiny electron-based transistors. Moreover, a lot of theoretical concepts that represent the capabilities of these devices or are closely connected with them, such as Turing machine or complexity classification, are currently undergoing lots of extensions due to attempts to incorporate many bizarre aspects of nature into the computational systems (for example the simple quantum computing elements - qubits).

The thesis is called «Modelling of exciton-polariton condensates for unconventional computing», because its primary goal is to exploit BEC's model in the GP framework concerning various computing purposes. Although the topic's connection with CS is very unusual and can be seen through the lens of physical science, the central part of this work is thus can be treated as applied science.

The general idea of unconventional computing is to use physical systems to simulate the process, which resembles algorithmic problem solving, finally giving the solution to the practical tasks. To investigate this correspondence between the problem of interest and the physical system, we do not usually need the experimental setup (which will be helpful for validation purposes). That is why it is enough to work with the theoretical models, but the connection with the physical world become very elusive. That is another reason to call the chapter a physical introduction, stressing this connection with the physical experimental setup.

For those who want to skip the historical details, I briefly describe BEC's essential properties. Initially, a Bose–Einstein condensate (BEC) is a group of atoms cooled close to absolute zero. While losing the portions of energy, they begin to clump together and enter the same energy states. Since the corresponding statistics, they become identical, and the whole group starts behaving as a single entity in the same quantum state. This behaviour is observed when the thermal de Broglie wavelength becomes comparable or more significant than the average inter-particle spacing.

This final condensed perspective was formed during the long historical process that is rich for other brilliant ideas. It started with the London ideas in 1938 about the macroscopic wavefunction, which is the Bose and Einstein works' development. Then, understanding a more detailed picture of superfluid 4He got another perspective with minor overlaps, like the Feynman idea of the many-body wavefunction and phenomenological two-fluid description by Landau. Moreover,

many of the critical concepts of modern physics like elementary excitations, collective modes, broken symmetry, order parameter were first introduced in dealing with superfluid 4He. I give little emphasis on the supporting formulas, using them mainly in the EP description or stressing the core concepts since representing all of the critical developments in condensed matter physics with the proper rigour is an impossible task.

2.1 History of BEC

To give a brief history of the scientific development concerned the ideas about macroscopic wave function, BEC, and superfluidity as the preparatory material for description of BEC in the EP system, the help of «truncated version» from the source [81] was used. To get the additional perspectives on the historical aspects from other sources one can use the scientific biography of Fritz London by Gavroglu [82]; the autobiography of Andronikashvilii [83] with an insider's perspectives on the low-temperature group under Landau and Kapitza in Moscow; the Ph.D. thesis of Jurkowitz [84], which examines the development of modern theories of superfluidity and superconductivity developed in the 1930s; the collection of papers by P.W. Anderson [85]. As we can see, much of the profound physicists were involved in the condensed matter theory to answer the most intriguing questions of the 20th century.

2.1.1 BEC prediction and the study of superfluidity

Satyendra Nath Bose, an Indian physicist and mathematician, first sent Einstein a paper, where he investigated the properties of quantum statistics of photons, where he derived Planck's quantum radiation law without any reference to classical physics. Einstein greatly admired this work and helped Bose translate the document into German to submit it to the publication in Zeitschrift für Physik. The «Planck's Law and Hypothesis of Light Quanta» was published in German in 1924 under Bose's name.

Einstein then developed Bose's ideas. Final efforts resulted in the concept of a Bose gas, which is governed by Bose-Einstein statistics. In general, it describes the statistical distribution of indistinguishable particles with integer spin, or in modern terms the bosons. Both photons and Helium-4 atoms appear to be bosons and can occupy the same quantum state. Einstein proposed that cooling bosonic atoms to very low temperatures would result in their condensation to the available lowest state, leading to the completely new state of matter.

The first bricks in the development of foundations of quantum mechanics were laid after this noticeable Einstein's paper's appearance. Additionally, this paper referenced "invisible" for that time works of Louis de Broglie [86, 87], using to some extent the ideas of matter waves. Einstein used the same idea as a justification for calculating the photons statistics, which have the properties of the bosons. There is an opinion that the origin of Schrodinger's wave equation appeared out of Einstein's BEC paper and not reverse [81].

London brothers Fritz and Heinz introduced a novel concept of superconductivity based on the already mentioned core idea of a macroscopic wave-function in 1935-1937 years. The illustration of this idea together with the criterion for Bose-Einstein condensation are presented in Fig. 2-1. The same ideas strongly influenced the development of the modern microscopic BCS theory [82]. At the same time, the experimental work on the superfluidity in liquid Helium was in progress. The main experimental effect was discovered by Pyotr Kapitsa [90], and John F. Allen and Don Misener [91] in 1937 independently and later published in January 1938 in *Nature*. London heard about it as well as about the most significant and important physical results, such as the phase transition at $T_c = 2.17K$, a peak in the specific heat and other aspects of strange behavior (like the sudden absence of boiling below the critical temperature T_c [92], zero viscosity appearing in the microchannels, approaching the infinity thermal conductivity, etc). London immediately tried to put the ideas about the bosons and the macroscopic wave-function together, which resulted in the proposal of the BEC pattern involved in this strange Helium phase transition in late January. Moreover, the basic Einstein's ideas with the corresponding formulas for the critical temperature and the specific heat agreed with these experiments. The final idea was submitted as a one-page work in March and later published in *Nature* in April [93] by London.

Another noticeable Hungarian-born American physicist, Lazlo Tisza, initiated the two-fluid theory after London shared his ideas about BEC's involvement into the Helium phase transition (with the later consequent help of Edward Teller and Lev Landau, with whom he worked in Kharkov during 1935-1937). The concept of the collective degree of freedom, which forces the matter to move coherently with the consequent loss of a friction mechanism, appeared. In general, this mechanism should give rise to superfluid behaviour. In such a way, the general phenomenological theory began to develop with later improvements. Tisza submitted a brief note after London in May 1938 [94].

To demonstrate the idea of the two-fluid model it is useful to make a few steps further and show the factorization of the GP equation, which will be described further with all possible modifications and similar mathematical tricks. The common field-theoretic description of Bose condensation is to separate the classical field of Bose condensate degree of

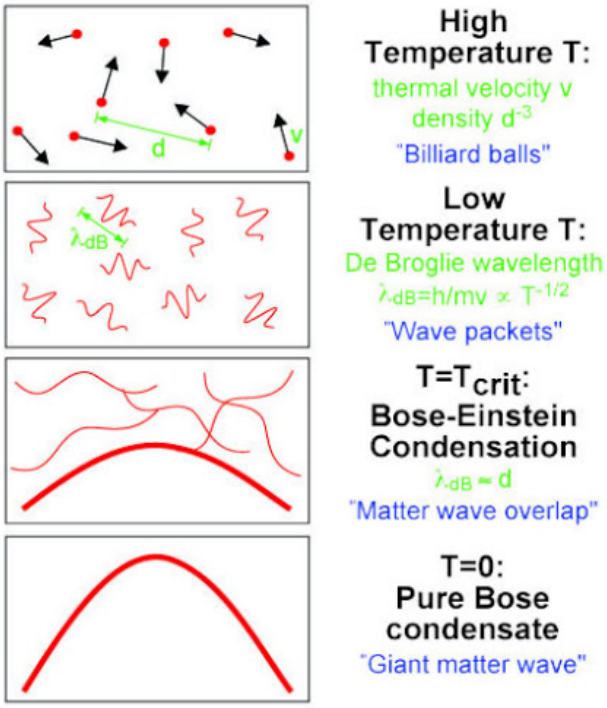


Figure 2-1: The picture representing the criterion for Bose-Einstein condensation [88]. At high temperatures, a weakly interacting gas can be treated as a system of mechanical particles. In a simplified quantum description, the wave effects come into play, atoms can be regarded as wavepackets with an extension δx, approximately given by Heisenberg's uncertainty relation $\delta x = \hbar/\delta p$, where δp denotes the width of the thermal momentum distribution. δx is approximately equal to the thermal de Broglie wavelength λ_{dB}. Cooling down the gas is increasing the de Broglie wavelength. At the BEC transition temperature, λ_{dB} is comparable to interatomic distance, and the Bose condensates are characterized by a macroscopic population of the ground state of the system. Finally, the temperature approaches absolute zero, the thermal cloud disappears, leaving a pure Bose condensate. The figure is taken from [88].

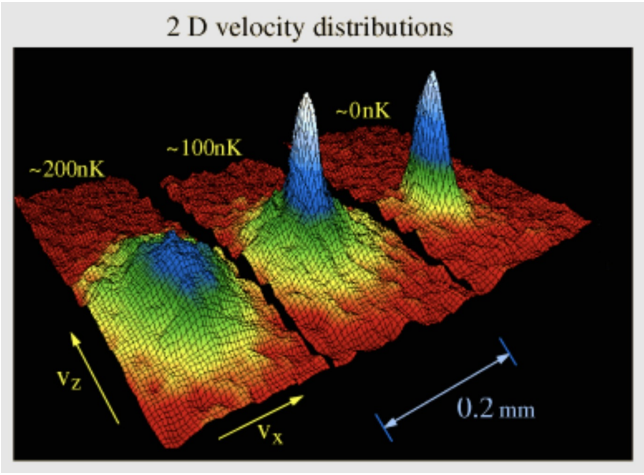


Figure 2-2: The figure shows the result of the momentum distribution measurement for the Rubidium atoms. In the left plot, no Bose-Einstein Condensation took place. The Bose-Einstein statistics gives the energy distribution of the atoms. In the middle plot, the conditions for condensation were hardly achieved. There are statistically distributed atoms with the ground state's additional overpopulation, expressed by the sharp blue peak. In the picture to the right, condensation conditions were fulfilled so that hardly any statistically distributed atoms can be seen. One can observe a high concentration of atoms in the ground state. This image is specifically credited to Mike Matthews of the JILA research team, University of Colorado, Boulder. The picture is taken from [89].

freedom from the non-condensate component of the quantum field, which satisfies Bose commutation relations. At zero temperature we have a dilute Bose gas with all the atoms described by the complex common condensate wavefunction $\Phi(\mathbf{r},t)=\sqrt{n_c(\mathbf{r},t)}e^{i\theta(\mathbf{r},t)}$, with the superfluid velocity $m\mathbf{v}_s(\mathbf{r})=\hbar\boldsymbol{\nabla}\theta(\mathbf{r},t)$. The wavefunction satisfies the GP equation of motion:

$$i\hbar\frac{\partial\Phi(\mathbf{r},t)}{\partial t}=\left[-\frac{\hbar^2\nabla^2}{2m}+V_{ex}(\mathbf{r})+gn_c(\mathbf{r},t)\right]\Phi(\mathbf{r},t), \tag{2.1}$$

with $V_{ex}(\mathbf{r})$ is the trap potential and $gn_c(\mathbf{r},t)$ is the Hartree field term produced by the condensate atoms. It is possible to rewrite the GP equation in terms of the density $n_c(\mathbf{r},t)$ and velocity $\mathbf{v}_s(\mathbf{r},t)$ of the condensate (this separation as well as the Madelung transformation will be extensively used further):

$$\begin{aligned}\frac{\partial n_c(\mathbf{r},t)}{\partial t}&=-\boldsymbol{\nabla}\cdot n_c(\mathbf{r},t)\mathbf{v}_s(\mathbf{r},t)\\ m\left(\frac{\partial\mathbf{v}_s}{\partial t}+\frac{1}{2}\boldsymbol{\nabla}\mathbf{v}_s^2\right)&=-\boldsymbol{\nabla}\mu_c(\mathbf{r},t),\end{aligned} \tag{2.2}$$

where the condensate chemical potential is $\mu_c(\mathbf{r},t)\equiv-\frac{\hbar^2\nabla^2\sqrt{n_c}}{2m\sqrt{n_c}}+V_{ex}(\mathbf{r})+gn_c(\mathbf{r},t)$. This separation into the two "hydrodynamic" looking equations is possible due to the complex form of the wavefunction parameter, describing a large number of atoms in the same single-particle quantum state. For the additional discussion see [95].

The first papers of London and Tisza created much interest in ideas about BEC's relevance to liquid Helium. Both scientists developed these concepts in more detail further in [96] and in [97]. However, London works lacked quantitative description. The necessary theoretical tools required for filling these gaps in the mathematical construction of the theory describing the liquid Helium appeared only in the late 1950's [81] when the many-body theory started to develop. Still, the basic London ideas and concepts can be found in modern microscopic theories of superfluidity and superconductivity [98].

The original idea of a "two-fluid hydrodynamics" based on the notion of a superfluid and normal fluid was the consequence of merging the giant matter-wave concepts with the classical hydrodynamics. Tisza managed to explain all the experiments exhibiting superfluidity, using the two-fluid model moving in opposite directions. It is worth noting that the Hungarian-American physicist predicted the particular kind of hydrodynamic oscillation, which is famously known as a second sound.

Unfortunately, London and Tisza's ideas were overshadowed by the emphasis on operator formalism in quantum mechanics and its place in the general picture of science with the appearing questions of interpretation of quantum mechanics. Nevertheless, these concepts have received renewed interest close to the 1950's with the consequent work of Bogoliubov in 1947 [99] and Oliver Penrose in 1951 [100].

The significant role in the history of the superfluidity development theory was played by Lev Landau, the Soviet physicist who made many other fundamental contributions to the different branches of theoretical physics, such as density matrix method [101] in quantum mechanics, the quantum mechanical theory of diamagnetism, the theory of second-order phase transitions, the Ginzburg–Landau theory of superconductivity, the theory of Fermi liquid along with many others. Still, the 1962 Nobel Prize in Physics was awarded to him for the mathematical theory of superfluidity. The phenomenological paper [102] by Landau on superfluid 4He is considered to be view-changing on perspectives concerning the condensed matter systems, despite a lot of original ideas of "new hydrodynamics" of two fluids with normal and superfluid components from the work of Tisza [97].

Moreover, Landau came up with many other brilliant ideas. He introduced the novel but powerful idea that the liquid (or another many-body system) could be described in terms of a "gas of weakly interacting quasiparticles" (or excitations), which by themselves have a relatively simple energy spectrum for two kinds of quasiparticles: phonons and rotons. The quasiparticles concept describes the emergent phenomena for the complicated system, usually very useful in the microscopic condensed matter settings, where such collective effects essentially behave like weakly interacting particles in the free space. Modern physics deals with the wide variety of such quasiparticles [103]. Currently, the modern condensed matter theoreticians are trying to overcome the limitation of the proposed description, which does not diminish the usefulness of the quasiparticle approach. Despite the historical remarks, the concept is twice helpful in the description of EP systems, for which there will be devoted quite a lot of attention within the scope of this thesis.

Among other ideas of Landau was the introduction of collective modes, distinctively from the above-mentioned quasiparticles. By the 1950's the London-Tisza model was not so popular because of the competing Landau-Khalatnikov theory. Its close connection with the experiment (direct measurement of the phonon-roton spectrum using inelastic neutron scattering [104, 105]) confirmed the correctness of the Landau scenario. Nowadays, the Landau-Khalatnikov theory of superfluid 4He is considered the standard theory, primarily used to describe the superfluid 4He properties. However, the idea of the bosonic nature of 4He atoms was not mentioned in Landau's landmark paper, overlooking not

only the BEC origins but, more importantly, the ideas of the matter-wave system.

The Bose condensate ideas' Landau resistance can be explained from his personal perspectives on the quantum mechanics [81]. He sincerely believed in the Copenhagen view of quantum mechanics, as can be concluded from the beginning of [106]. In his opinion, to obtain an accurate description, one needs to take a classical description and then quantize it by initializing the operators corresponding to the physical observables. Similarly, the correct way of understanding a "quantum" liquid was to quantize the standard hydrodynamical theory of a "classical" liquid. However, this approach never succeeded. This point of view changed after the two excellent papers by Beliaev in 1957 [107]. The final statement of Beliaev is that «the difference between liquid He and a non-ideal Bose gas is only a quantitative one, and that no qualitatively new phenomena arise in the transition from gas to liquid". Despite the Landau initial resistance of a Bose condensate being relevant to superfluid 4He, he quite frequently formulated and used the concept of an order parameter, which can be thought of as a very much similar idea with a matter-wave.

The big breakthrough came with the paper of Bogolyubov in 1947 [99], which took a lot of time time to be generally understood and accepted by the whole community of condensed matter physicists. One key point of this work was the demonstration that "BEC was not much altered by interactions in a weakly interacting dilute Bose gas, something which was not obvious at the time" [81]. The interactions were shown to change the long wave-length behaviour of Bose-condensed gas completely. Consequently, Bogolyubov filled the gaps in Landau's phenomenological theory with his rigorous mathematical tools, which also resulted in the predicted phonon spectrum at low momentum.

The connection between London and Tisza's original works and Bogolyubov's paper can be traced through his work. Merging the ideas soon resulted in developing a "complete" theory by the early 1960s, based on Bose's critical role in broken symmetry. However, these critical results did not attract significant attention until 1957. This happened due to the quite impressive use of mathematical tools, such as the second quantization technique (unfamiliar at that time) and several non-conserving approximation.

Landau recognized the correctness of Bogolyubov's results and acknowledged them [108]. Soon the BCS theory of superconductivity appeared in 1957 [109], which quickly triggered the attention to the previous works of Bogolyubov and his treatment of a dilute Bose gas since both theories involve similar ideas.

The ideas about the giant matter-wave and corresponding experimental BEC realization were hidden behind the mainstream works on superfluidity. The initial perception of the liquid 4He was heavily influenced by the hydrodynamic description, taking into account the simplicity of the two-fluid model. The corresponding focus was on the collision-dominated hydrodynamics domain, while the majority of the original experiments in the 1930s were on low frequency, hydrodynamic phenomena.

In 1946 Peshkov succeeded in observing the second sound as a temperature oscillation and showed that the temperature dependence of the second sound velocity agreed with the prediction of two-fluid hydrodynamics [110]. In contrast to the superfluidity, to get the direct experimental evidence of BEC was very hard in 4He. However, the BEC state was immediately observed much later in atomic gases in the first successful experiment [89], while the two-fluid hydrodynamic region was difficult to access because of the low density.

The Bose complex order parameter in modern theory is defined as $\Phi(\mathbf{r},t) = \sqrt{n_c(\mathbf{r},t)}e^{i\theta(\mathbf{r},t)}$. The superfluid velocity which is given by the expression $m\mathbf{v_s}(\mathbf{r},t) = \hbar\nabla\theta(\mathbf{r},t)$ can be easily measured by a variety of experiments. Comparing it with the n_c, which is "hard to catch" in the experiments with superfluid 4He. Additionally, the superfluid density ρ_s is easier to measure, which depends on n_c but in a complicated manner. These facts were additional reasons for the BEC's delayed realisation and the shifted attention of the community. The Bose complex parameter is very similar to the Madelung ansatz, used in quantum hydrodynamics, allowing one to equivalently reformulate the Schrödinger equation in terms of hydrodynamical variables, similar to the Navier–Stokes equations of fluid dynamics and will be used extensively in the derivations included in this thesis.

Penrose and Onsager [111] extended the concept of $\Phi(\mathbf{r},t)$ to a uniform Bose liquid and discussed the corresponding long-range correlations in 1956. They estimated the n_c density at $T = 0$ using Feynman approach which will be mentioned below and found $n_c = 0.08n$ in liquid 4He. Further experiments in 1980s gave more convincing values. Additionally, after a good analysis the estimates on $n_c = 0.1n$ at $T = 0$ can be made with the important limit of $n_c \to 0$ as $T \to T_c$.

Another prominent American theoretical physicist Richard Feynman made his contributions to superfluidity and BEC quantum mechanics. Starting with the various papers, he succeeded to obtain the “microscopic” picture of the roton part of the quasiparticle spectrum. The initial scope of his works aimed to answer the critical question on the connection between many-body wavefunctions $\Psi_N(r_1, r_2, ...r_N)$ for the ground state the lowest excited states of liquid 4He and Bose statistics [112, 113]. These works were the beginning of a direction to create various approximations for the many-particle quantum states wavefunctions. The approach was successful in calculating ground-state properties. Besides, part of Feynman's works was devoted to vortices and their role in superfluid 4He. He stressed the idea that so-called “rotons” were intrinsically connected to the phenomena of superfluidity - and these were related to vortices.

A significant number of insights on the BEC behaviour was obtained during the 1957 - 1965 period. In these years, the interacting Bose-condensed gas problem was a hot topic and attacked by many theorists. The main key aspects can be found in [98, 107, 114, 115, 116, 99].

Another critical development, which can not be separated from these contributions is the famous GP equation for $\Phi(\mathbf{r}, t)$ [117] which describes the ground state of a quantum system of identical bosons and is relevant for the description of the trapped gases. It was obtained using the Hartree–Fock approximation and the pseudopotential interaction model from 1959 to 1961. Nevertheless, Pitaevskii's real contribution lies in introducing the whole idea of a macroscopic wave-function that could depend on both position and time. In some sense, it serves as a development of the giant matter wave idea. Pitaevskii's work was also inspired by the use of a similar "wave-function" approach from Ginzburg and Landau pioneering work on spatially inhomogeneous superconductors [118].

The Bose statistics and corresponding effects played a crucial role in physics during the 20th century with the essential manifestation in the phenomenon of superfluidity of liquid 4He. It also inspired alternative activities, like the first experiments on liquid 3He (a Fermi liquid) in 1949, which were not so fruitful since it showed no evidence for superfluidity down to 1K.

The modern understanding of superfluidity in Bose fluids is intrinsically connected with the understanding of superconductivity [85], where the general macroscopic wavefunction is the Bose order parameter in superfluids [107], while is almost equivalent to the Cooper pair amplitude in superconductors [119]. Developing these ideas lead to the realization that the Bardeen–Cooper–Schrieffer (BCS) theory [109] can be treated as a "BEC of Cooper pairs" [119]. This similarity became clear in 1980s when Leggett, Nozieres and others (see the article by Randeria in [120]) stressed out that the BCS theory in the strong coupling limit intrinsically reduces to a Bogolyubov theory of a dilute Bose gas composed of non-overlapping Cooper pairs.

Some of the research directions of modern physics are aiming towards studying this BCS phenomenon with the emphasis on BCS to BEC transition [121]. Before that, there were lots of systems investigated for the BEC properties. The review [120] lists the major systems that can exhibit the BEC properties. Among them are laser-cooled alkali atoms, spin-polarized hydrogen atoms and optically-excited excitons in semiconductors which will be in the centre of attention, devoted in this thesis. The following subsection will give a brief overview of the experimental achievements, which went hand in hand with the described history of theoretical developments before jumping into the particular EP settings.

2.1.2 Experimental verification of BEC

Repeating the previous experimental key results - 1938 was the year when Pyotr Kapitsa and independently John Allen and Don Misener discovered superfluidity in liquid Helium at a temperature below the λ-point $T = 2.17K$ [90, 91]. The obtained substance had many bizarre properties: zero viscosity and the existence of quantised vortices. However, the idea of experimental realization of BEC, fundamental for the superfluid helium-4, was delayed due to several potential reasons listed before.

The first form of a "pure" gaseous BEC was produced by Eric Cornell, Carl Wieman, and their team at the University of Colorado at Boulder NIST–JILA lab (National Institute of Standards and Technology - Joint Institute for Laboratory Astrophysics US) with a gas of rubidium atoms cooled down to 170 nanokelvins [89] on 5 June 1995. Fig. 2-2 and Fig. 2-3 shows the momentum distribution measurements and the experimental setup respectively. Shortly after this event, Wolfgang Ketterle produced the same state of matter at MIT but with a gas of sodium atoms [122]. These achievements were rewarded with the 2001 Nobel Prize in Physics for Cornell, Wieman, and Ketterle. Another result around the same time was achieved by a Randall Hulet group at Rice University with lithium atoms with a slight delay of one month after the JILA work [123].

To obtain the Bose-Einstein condensation, one needs to operate with the techniques to cool gases to sub-microkelvin temperatures and confine them at high density while keeping them away from the vacuum chamber's hot walls [120]. The sequential cooling technique, which first slows the atomic beam, then optically traps them, and laser cools with the final magnetic trapping and evaporative cooling. The overall cooling process goes from 600 K by nine orders of magnitude to 1mK.

To identify the BEC formation within the experimental setup, one can use four noticeable features, among which are the sudden density increase, the sudden appearance of a bimodal cloud consisting of a diffuse, normal component and a dense core, anisotropic velocity distribution of the condensate and agreement between the predicted and measured transition temperatures. The condensate image is usually recorded by illuminating the cloud with resonant laser light and imaging the shadow of the cloud onto a charged-coupled device camera [89].

Two critical experimental techniques were practically perfected during these experimental achievements. In the first work, a dilute vapour of nearly two thousand rubidium-87 atoms was cooled below 170 nK using a combination of laser cooling and magnetic evaporative cooling. Laser cooling is a technique that uses the photon exerted pressure to slow

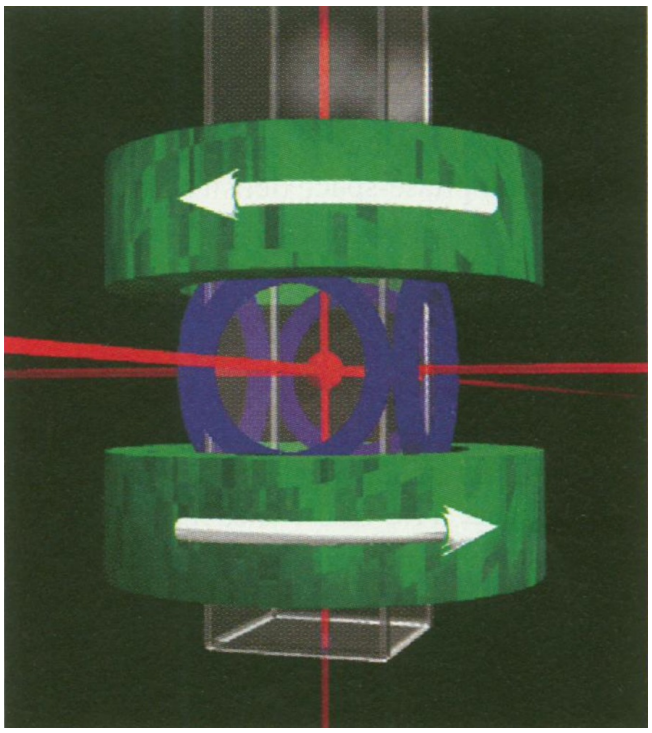


Figure 2-3: The picture representing apparatus for the first experimental observation of the BEC [89]. Six laser beams intersect in a glass cell, creating a magneto-optical trap. The cell is 2.5 cm square by 12 cm long, and the beams are 1.5 cm in diameter. The coils generating the fixed quadrupole and rotating transverse components of the top trap magnetic fields are shown in green and blue, respectively. The glass cell hangs down from a steel chamber (not shown) containing a vacuum pump and rubidium source.

down an atom moving against a laser beam since it encounters a higher frequency than an atom moving with the beam's direction. The inventors of laser cooling, Steven Chu, Claude Cohen-Tannoudji, and William D. Phillips, got the 1997 Nobel Prize in Physics.

Another technique - evaporative cooling occurs in a magnetic trap, where the atoms with the highest energy escape while the remaining ones collide with each other, apportioning out the remaining energy. During this time, the atoms slow down and are closely packed at the bottom of the trap so that their quantum nature becomes more pronounced. The particles' wave nature starts to manifest itself, and the particle-wave packets become less distinct and start to overlap. The chain reaction of atoms going down the trap into the lowest energy state results in merging into the single ground-state wavefunction which is a BEC itself.

After this tremendous success, the foundation for the experimental field of ultracold atoms was layed. Hundreds of research groups are now producing BECs of dilute atomic vapours in their laboratories routinely, leading to an increase in experimental and additional theoretical activity. The list of isotopes for which BECs had been obtained is the following: ^{7}Li, ^{23}Na, ^{39}K, ^{41}K, ^{85}Rb, ^{87}Rb, ^{133}Cs, ^{52}Cr, ^{40}Ca, ^{84}Sr, ^{86}Sr, ^{88}Sr, ^{174}Yb, ^{164}Dy, ^{168}Er [124]. They are mostly alkali metals, alkaline earth metals, and lanthanide atoms since their nuclear properties are a perfect fit for working in traps. Some alkali gases' bosonic nature arises from an interplay of electronic and nuclear spins, despite their nuclei having half-integer total spin. At low temperatures, the half-integer total spin of the electronic structure together with the half-integer total spin of the nucleus are coupled by the weak interaction. Nevertheless, such systems' chemical properties at room temperature are determined mostly by the electronic properties, which is essentially fermionic, because the energy of typical thermal excitations is much higher than the electronic-nuclear spin coupling values.

After the significant 1995 year, different atomic species have been condensed. BEC state was realized using molecules, quasiparticles in solids (magnons, excitons, polaritons, since they have integer spin, representing their bosonic nature) and photons [125, 126]. Additionally, BEC is an excellent platform to study various physical phenomena. Among them are interference effects, superfluidity, and quantized vortices, bright matter-wave solitons, light pulses slowed with the electromagnetically induced transparency [127]. Besides, the BEC platform allows one to reproduce exotic phenomena, like the correspondence between analogue gravity black holes and vortices, realization of "optical lattices", usefull to explore the transition between a superfluid and a Mott insulator [128], the solids that flow through themselves [129], magnetic liquid droplets [130], neutral particles that act like charged [131] or explosion analgues in BEC[132].

2.2 BEC of exciton-polaritons

2.2.1 What is an exciton-polariton?

Microcavity polariton is a quasiparticle resulting from the strong coupling between a Wannier-Mott exciton (a bound pair of electron and hole) and light components (strongly confined photon field in a semiconductor microcavity). Having the quasiparticle concept for the model isolated system in mind, the vacuum analogue for a semiconductor is the state with a filled valence band and an empty conduction band. At the same time, excitons are treated as units of elementary excitations or quasiparticles. They consist of an excited from the valence band electron with charge $-e$ and a hole, a vacancy having the opposite charge $+e$. The electron-hole system is bound by a Coulomb interaction, which energy is approximately of 10-100 meV and the Bohr radius is 10-100 Å [133]. Using conventional for the condensed matter notation, one can write the exciton creation operator in the following way:

$$\hat{e}^{\dagger}_{\boldsymbol{K},n} = \sum_{\boldsymbol{k},\boldsymbol{k}'} \delta_{\boldsymbol{K},\boldsymbol{k}+\boldsymbol{k}'} \boldsymbol{\phi}_n \left(\frac{m_h \boldsymbol{k} - m_e \boldsymbol{k}'}{m_e + m_h} \right) \hat{a}^{\dagger}_{\boldsymbol{k}} \hat{b}^{\dagger}_{\boldsymbol{k}'} \tag{2.3}$$

where $\hat{a}^{\dagger}_{\boldsymbol{k}}$ and $\hat{b}^{\dagger}_{\boldsymbol{k}'}$ are the creation operators of a conduction-band electron with momentum $\boldsymbol{k}$ and of the valence-band hole with $\boldsymbol{k}'$ respectively. m_e and m_h are their corresponding masses, $\boldsymbol{K}$ is the center-of-mass momentum of the exciton and the $\boldsymbol{\phi}_n(\boldsymbol{k})$ is the wave-function of the mutual dynamics of an electron and a hole similar to the system of an electron and a proton in a hydrogen atom with n being the quantum number of the relative motion. One can easily derive commutation relations from the presented form:

$$\begin{aligned} &[\hat{e}_{\boldsymbol{K}',n'}, \hat{e}_{\boldsymbol{K},n}] = 0 \\ &\left[\hat{e}^{\dagger}_{\boldsymbol{K}',n'}, \hat{e}^{\dagger}_{\boldsymbol{K},n}\right] = 0 \\ &\left[\hat{e}_{\boldsymbol{K}',n'}, \hat{e}^{\dagger}_{\boldsymbol{K},n}\right] = \delta_{\boldsymbol{K}\boldsymbol{K}'}\delta_{nn'} - O\left(n_{\text{exc}} a_B^3\right), \end{aligned} \tag{2.4}$$

which allows us to treat excitons as bosons with the assumption that exciton interparticle distance is much larger than its Bohr radius $n_{\text{exc}} \ll a_B^{-3}$. So far, excitons are electron-hole pairs which are bounded together by the Coulomb interaction. Their binding energy depends on their mass and the Bohr radius, similar to the hydrogen atom case. Following [134], the brief introduction to the EP physics is presented.

By introducing the microcavity, one can achieve the photon confinement in the z direction. A semiconductor quantum well (QW) is required to enhance polariton properties and achieve stronger EP coupling. It is a thin layer of semiconductors with a small thickness compared to the exciton Bohr radius. QW is surrounded by barrier layers with a much larger bandgap. QW excitons behave like two-dimensional quasiparticles in the presented setting if only one quantized level of their motion is concerned. The quantum confinement modifies the valence band structure and, hence, changes the optical transition strengths and selection rules. Moreover, the momentum conservation happens only in the QW plane but not along the confinement axis. Thus, the exciton couple to light with the in-plane wave number $k_{||} = k_{xy}$ and any transverse $k_{\perp} = k_z$. The confinement of a QW exciton leads to its smaller Bohr radius and larger binding energy than its bulk analogue.

A distributed Bragg reflector (DBR) is a building block of a semiconductor microcavity. Its role is to serve as a high-reflectance mirror when the wavelength of the incident light is in the stopband range, and the destructive interface effects prevent the transmission of light. The DBR structure consists of layers of different high and low refraction indices with the optical path equals to a quarter of a target wavelength. It is possible to get high values of reflectivities using this kind of structure and, consequently, dramatically increasing the photon lifetimes. More important is that one can form a resonance at λ_c with two such high-reflectance DBRs that surround a layer with an optical thickness proportional to the half of the λ_c and some integer number. The well-known formula gives the transmission at this wavelength:

$$T = \frac{(1-R_1)(1-R_2)}{\left[1-\sqrt{R_1 R_2}\right]^2 + 4\sqrt{R_1 R_2}\sin^2(\phi/2)}, \tag{2.5}$$

with ϕ is the photon phase shift after reflection and R_1, R_2 are the reflectances of the DBRs (depending on the refraction indices). Another important characteristic of the cavity is the quality factor Q:

$$Q = \frac{\lambda_c}{\Delta\lambda_c} \simeq \frac{\pi (R_1 R_2)^{1/4}}{1-(R_1 R_2)^{1/2}}, \tag{2.6}$$

where $\Delta\lambda_c$ is the width of the resonance, which for an ideal case would go to zero. In the special case of cavity equals half the wave-length, Q will represent the average number of photon reflections between the cavity borders.

The behaviour of photon field inside the cavity depends on the considered axis. The confinement takes place only in the xy plane, while the incident light from the angle θ relative to the z axis would have a resonance at $\lambda_c/\cos\theta$. One can define the energy dispersion for the photon field (also called cavity dispersion):

$$E_{ph} = \frac{\hbar c}{n_c}\sqrt{k_z^2 + k_{xy}^2}, \tag{2.7}$$

with $k_z = k_\perp = n_c\,(2\pi/\lambda_c)$, with n_c being the refraction index of the cavity. Each resonance mode with in-plane wave number k_{xy} has the unique correspondence with the incidence angle θ, so if $k_{xy} \ll k_z$:

$$k_{xy} = n_c\frac{2\pi}{\lambda_c}\tan\left[\sin^{-1}\left(\frac{\sin\theta}{n_c}\right)\right] \approx \frac{2\pi}{\lambda_c}\theta, \tag{2.8}$$

This allows one to rewrite the dispersion relation in the following form:

$$E_{ph} \approx \frac{\hbar c}{n_c}k_z\left(1 + \frac{k_{xy}^2}{2k_z^2}\right) = E_{\text{cav}}\,(k_{xy} = 0) + \frac{\hbar^2 k_{xy}^2}{2m_{\text{cav}}}, \tag{2.9}$$

which consequently makes it possible to evaluate the cavity-photon effective mass, which is typically of $10^{-5}m_e$. It is clear that the low mass of the particles participating in the condensation transition leads to the higher critical condensation temperatures, which is close to the room temperature.

The storage of the combined excitations in the system can be achieved when the energy exchange rate between the cavity field and excitons exceeds the decay and decoherence rates of the photons and the excitons. Then, the creation of new eigenstates that are coherent superpositions of the two initial states happens. One can make a working assumption about the fully two-dimensional behaviour of the system (all corresponding wavevectors will be in the xy plane so that $k = k_{xy} = k_{||}$ and the same works for the dispersion relations $E(k)$). Thus, the linear Hamiltonian of such a system can be written in the simple form of its constituents and the interaction term:

$$\hat{H}_{pol} = \hat{H}_{ph} + \hat{H}_{exc} + \hat{H}_{int} = \sum_k E_{ph}(k)\hat{a}_k^\dagger\hat{a}_k + \sum_k E_{exc}(k)\hat{b}_k^\dagger\hat{b}_k + \sum_k g_0\left(\hat{a}_k^\dagger\hat{b}_k + \hat{a}_k\hat{b}_k^\dagger\right), \tag{2.10}$$

where $\hat{a}_k$ is the same photon creation operator, $\hat{b}_k$ is the exciton creation operators, $g_0 = \hbar\Omega$ is the exciton-photon dipole interaction strength. The diagonalization of the Hamiltonian (2.10) will give the eigenenergies of the new eigenstates. To find them, one has to calculate the eigenvalues of the following matrix:

$$\tilde{H} = \begin{bmatrix} E_{ph}(k) & g_0 \\ g_0 & E_{exc}(k) \end{bmatrix} \tag{2.11}$$

This can be achieved by the linear transformations $\hat{P} = X_k\hat{b}_k + C_k\hat{a}_k$ and $\hat{Q} = -C_k\hat{b}_k + X_k\hat{a}_k$, which lead to the simpler form of the Hamiltonian (2.10):

$$\hat{H}_{pol} = \sum_k E_{\text{L}}(k)\hat{P}_k^\dagger\hat{P}_k + \sum_k E_{\text{U}}\,(k)\,\hat{Q}_k^\dagger\hat{Q}_k, \tag{2.12}$$

where two branches of lower and higher eigenenergies corresponds to the new quasiparticles of the system, which are referred as the lower polaritons (L) and upper polaritons (U).

The eigenenergies of the polariton Hamiltonian are derived from the diagonalization procedure of the matrix $\tilde{H}$:

$$E_{\text{L,U}}(k) = \frac{1}{2}\left[E_{exc}(k) + E_{ph}(k) \pm \sqrt{\left(E_{exc}(k) - E_{ph}(k)\right)^2 + 4g_0^2}\right] \tag{2.13}$$

The phenomenon of splitting of the eigenvalues of a Hermitian matrix that cannot become equal in value is also known as avoided crossing, also referred to as the von Neumann-Wigner theorem. The energy separation between the two new states is connected with the coupling energy between the photons and excitons. Introducing the energy difference $\Delta E(k) = E_{exc}(k) - E_{ph}(k)$, the resonance case $\Delta E(k) = 0$ has the lowest separation value of $2g_0$ between L and U energies, which is often called the normal-mode splitting.

Since the polariton is a linear superposition of an exciton and a photon with the same $k = k_{xy}$, it inherits the bosonic nature of the excitons and photons. Each fractions in the branches are given by the Hopfield coefficients [135]:

$$\begin{aligned} |X_k|^2 &= \tfrac{1}{2}\left(1 + \tfrac{\Delta E(k)}{\sqrt{\Delta E(k)^2 + 4g_0^2}}\right) \\ |C_k|^2 &= \tfrac{1}{2}\left(1 - \tfrac{\Delta E(k)}{\sqrt{\Delta E(k)^2 + 4g_0^2}}\right), \end{aligned} \tag{2.14}$$

which in case of $\Delta E = 0$, $|X|^2 = |C|^2 = \frac{1}{2}$ one has half photon and half exciton state.

2.2.2 Gross-Pitaevskii description

As previously mentioned, an ideal gas of noninteracting bosonic particles can undergo the BEC transition, which Einstein predicted in 1925. However, the essential properties of BEC rely on the particle-particle interaction and resulting peculiar excitation spectra. Bogoliubov was the first to develop a quantum field-theoretical formulation for a weakly interacting Bose condensed system. It is helpful to show on the simple model how one can obtain the mean-field description from the Bogoliubov theory with the following quantities, such as interaction effects, the excitation spectra and the consequent superfluidity.

The Hamiltonian of the EP condensate in terms of the LP complex-field operator ψ reads:

$$\hat{H} = \int \left(\frac{\hbar^2}{2m} \nabla \hat{\psi}^+ \nabla \hat{\psi} \right) dr + \frac{1}{2} \int \hat{\psi}^+ \hat{\psi}'^+ V\left(r' - r\right) \hat{\psi} \hat{\psi}' dr' dr, \tag{2.15}$$

with $V(r)$ being the two-body interaction potential. The Heisenberg equation of motion is then defined as:

$$i\hbar \frac{\partial}{\partial t} \hat{\psi}(r,t) = [\hat{\psi}(r,t), \hat{H}] = \left[-\frac{\hbar^2 \nabla^2}{2m} + \int \hat{\psi}^+ \left(r',t\right) \times V\left(r' - r\right) \hat{\psi}\left(r',t\right) dr' \right] \hat{\psi}(r,t) \tag{2.16}$$

If one writes the field operator in the momentum basis form $\hat{\psi}(r) = \frac{1}{\sqrt{V}} \sum_p \hat{a}_p e^{ip \cdot r/\hbar}$, where V is a quantization volume, $\hat{a}_p$ is the annihilation operator for the single-particle state with a momentum p, the Hamiltonian can be rewritten as:

$$\hat{H} = \sum_p \frac{p^2}{2m} \hat{a}_p^+ \hat{a}_p + \frac{1}{2V} \sum_{p_1} \sum_{p_2} \sum_q V_q \hat{a}_{p_1+q}^+ \hat{a}_{p_2-q}^+ \hat{a}_{p_1} \hat{a}_{p_2}, \tag{2.17}$$

where $V_q = \int V(r) \exp[-iq \cdot r/\hbar] dr$ is the Fourier transform of the interparticle potential.

Initially, we are interested in the macroscopic description of the condensate, we need to take into account the small momenta and take $q = 0$ value for the Fourier component, thus $V_0 = g = \int V(r) dr$ and the interaction term in the Hamiltonian is written as

$$\frac{g}{2V} \sum_{p_1} \sum_{p_2} \sum_q \hat{a}_{p_1+q}^+ \hat{a}_{p_2-q}^+ \hat{a}_{p_1} \hat{a}_{p_2} \tag{2.18}$$

The corresponding c-number (c being the number replacing the operator $\hat{a}_0$) field amplitude, which appears to be an order parameter, $\psi_0(r,t)$ does change slowly over distances in comparison with the range of the interparticle force, thus we can change the potential form to the delta function and obtain the GP equation:

$$i\hbar \frac{\partial}{\partial t} \psi_0(r,t) = \left(-\frac{\hbar^2 \nabla^2}{2m} + g \left|\psi_0(r,t)\right|^2 \right) \psi_0(r,t) \tag{2.19}$$

Initially, the GP picture was obtained for bosonic fluid-like 4He and described its ground state using the Hartree–Fock approximation and the pseudopotential interaction model. The nonlinear term in the equation has its origin in the interaction between the particles. In the general case, the GP equation does not have an analytic solution. If one wants to model the behaviour and especially the dynamic properties, it is necessary to use different numerical methods.

The GP model's weaknesses come from the physical approximation valid for certain classes of BECs, like the contact two-body type interaction, which does not cover the general case. Moreover, the GP description of the EP system has more limitations on the GP model's applicability.

2.2.3 Quantum description and related effects

The key model of this thesis is the rate equations (which will be derived at the beginning of the next chapter), which represent the reduced description of the separate EP condensates. Such a system consists of the macroscopic wavefunction description of classical units (or their degrees of freedom), which interacts via harmonic potential. At the same time, the noise which plays a crucial part in the condensation process is treated as a classical random process.

However, one should understand that to reach the point of rate equations; one needs to perform many simplifications with the initial physical description. The central derivations of this thesis are based on the GP equation, which is "blind" to the microscopic quantum effects, essential for EP systems. To ignore these properties, we will give a partial description of this system from the quantum mechanical perspective, conventional for the quantum optics domain. This subsection contains standard models for such a system with the corresponding effects omitted in the classical description.

Superradiance is a phenomenon describing a group of several emitters, which interact with a common light field collectively and coherently due to the small separation of these sources compared to the wavelength of the light. The

related term superradiance was introduced in 1954 by Robert H. Dicke. This effect was considered in [136] and following works of Hepp and Lieb in 1973 on the steady-state superradiance [137] with the example of the so-called Dicke model, which became a fundamental model of quantum optics and the basic model to describe the interaction between light and matter. The Dicke Hamiltonian reads:

$$H = \omega_c a^\dagger a + \omega_z \sum_{j=1}^{N} \sigma_j^z + \frac{2\lambda}{\sqrt{N}} \left(a + a^\dagger\right) \sum_j \sigma_j^x, \tag{2.20}$$

where the light component is described as a single quantum mode with the creation and annihilation operators $a^\dagger$ and a, while the matter is described as a set of two-level systems with spin operators σ_i^α. The photon frequency is denoted by the ω_c parameter, ω_z represents the energy gap for the atom system and the photon-atom subsystems coupling is paramatrized by λ. The coefficient $\frac{1}{\sqrt{N}}$ is required for appropriate scaling of the number of photons n with the N atoms in the superradiant phase. The Hamiltonian (2.20) can be rewritten in more compact form through the total spin operators $S^\alpha = \sum_j \sigma_j^\alpha$, which greatly simplifies corresponding analysis. This model undergoes transition from normal to superradiant phase in the thermodynamic limit, which shares the mean-field Ising universality class.

Various physical realizations of the Dicke model have been considered, accounting for different behaviour or setups. For example, the generalized Dicke Hamiltonian has the form:

$$H = \omega_c a^\dagger a + \omega_z \sum_{j=1}^{N} \sigma_j^z + \frac{\lambda}{\sqrt{N}} \sum_{j=1}^{N} \left(a\sigma_j^+ + a^\dagger \sigma_j^-\right) + \frac{\lambda'}{\sqrt{N}} \sum_{j=1}^{N} \left(a\sigma_j^- + a^\dagger \sigma_j^+\right), \tag{2.21}$$

where λ and λ' determines the weights of interpolation between Dicke and Tavis–Cummings models. Such generalisation spawns different phase diagram with unequal weights, which makes the coexistence of the normal and superradiant stable states possible. Other extensions include Raman driving scheme with the additional term $Ua^\dagger a\sigma^z$ in the model Hamiltonian or an additional coherent light source coupled directly to the cavity mode with $H_F = F\left(a + a^\dagger\right)$ part.

Another variation of the Hamiltonian (2.20) can be extended for the individual subsystems with different characteristic constants:

$$H = \omega_c a^\dagger a + \sum_{j=1}^{N} \omega_z^j \sigma_j^z + \frac{2}{\sqrt{N}} \left(a + a^\dagger\right) \sum_j \lambda^j \sigma_j^x, \tag{2.22}$$

which is addressing the role of disorder and thus usually called disordered Dicke model, which is relevant to the study of microcavity polaritons. The discussion of other models and corresponding variations, non-equilibrium aspects of these models and some analytical and numerical approximations, used to study them can be found in [138].

Polariton condensation can be described by a model of localized excitons coupled to a continuum of radiation modes in a two-dimensional microcavity in case of neglecting Coulomb interactions between different units, based on the previous Hamiltonian. The thermodynamic properties of such model were investigated in great detail in [139, 140, 141, 142].

After defining the Hamiltonian, one is usually interested in the dynamics of the corresponding system. In general, one can not simply treat real physical systems as isolated due to the interaction with the environment, and the EPs are not an exception. The canonical formulation of quantum mechanics uses unitary dynamics to describe a system's time evolution, which entails the absence of decay and phase decoherence processes. Specific mathematical techniques have been introduced to describe the interaction of a quantum system with its environment. For example, the Caldeira-Leggett model treats the environment Hamiltonian as an infinite set of free modes with a continuous spectrum [143], which allows describing dissipative dynamics with the environment memory effects.

However, a typical quantum optics setup do not require to account for such complicated effects, and a reasonable simplification is usually enough. For this purpose, a Markovian open quantum systems framework has been developed [144, 145]. Such effective dynamics for the reduced density matrix of the system give rise to the Lindblad-form master equation, which reads:

$$\dot{\rho} = \mathcal{L}\rho = -i[H, \rho] + \mathcal{D}\rho, \tag{2.23}$$

where the first term represents the Von Neumann equation, which can be formally solved for the density matrix ρ by double multiplication of the system's density matrix by the evolution operators $U(t, t') = e^{-iH\left(t-t'\right)}$ and $U(t', t)$. $\mathcal{L}$ is called the Liouvillian super-operator. The second term $\mathcal{D}$ is referred as Lindblad super-operator and dramatically changes the dynamics of the isolated system. It is given by the following expression:

$$\mathcal{D}\rho = \sum_i \gamma_i (2L_i \rho L_i^\dagger - \left\{L_i^\dagger L_i, \rho\right\}), \tag{2.24}$$

where L_i are jump operators that are usually determined by describing the coupling between system and bath. Index i for rates γ_i and jump operators correspond to different sources of dissipation. In the case of the initial Dicke model, the main

sources are collective effects (such as cavity decay and collective atomic decay), and single-atom decay and dephasing [138]. The Lindblad master equation allows one to trace such effects as equilibration with the pump and decay processes, thermalization of the system and different aspects of interaction with the environment. Such an approach is closely related to Keldysh field theory. For a broader description (a non-Markovian case), one usually develops approximated schemes that account for the omitted effects.

Besides, EP systems have many more exciting effects, which include the charge-imbalanced electron-hole system in an optical microcavity [146], where the exciton condensation is expected to have unpaired carriers, occupying a crescent-shaped part of the momentum space due to the strong coupling to photons; polariton condensation and lasing over a wide range of cavity frequencies with organic molecules, which have multiple vibrational modes and were studied in the framework of a non-equilibrium Dicke-Holstein model [147], including both strong coupling to vibrational modes and strong matter-light coupling in the thermodynamic limit; consequent study of the organic polariton lasing using the microscopic picture beyond-mean-field approach and accounting for all photon modes in a system [148]. Furthermore, one can consider more complex models with a non-equilibrium scenario for the steady-state properties of the photons, taking into account the pump and decay processes [149] and consequent work on the thermalization process of photon gas [150].

In such a way, EPs represent an interesting platform for exploring its computational analogue properties, in addition to the physical properties. However, due to the enormous complexity of the system, it is either investigated using reduced macroscopic spatially dependent models (using different assumptions, such as the BEC condensation) similar to GP description or using localized quantum mechanical models with the particular Hamiltonian (such as presented above). In the last case, the behaviour of the system is richer, while in the first framework, many effects are ignored. Among them are quantum fluctuations, complex dynamics of the excitation levels, non-equilibrium processes and many others. The choice of the model usually depends on the quantities of interest.

One should pay attention to other microscopic processes in the EP system since such consideration gives more degrees of freedom to inspect compared to the simple mean-field theory [151, 152, 153]. These effects make a rich landscape for the potential future research directions and should be considered for the proper design of optoelectronic devices. An additional direction of work is to trace the influence of these effects on the algorithmic side of the operational regime.

2.2.4 Physical realization of exciton-polariton condensate

The early theoretical propositions of polariton BEC are attributed to the Atac Imamoglu and coauthors [154]. In that work they analyzed elementary properties of exciton and polariton lasers and the process of generating coherent optical and matter waves. They stressed the significant difference between these types of lasing and tried to realize it in the subsequent work The authors claimed the observation of the BEC state in [155], however it was soon disproved [156]. Preliminary work was later peformed with the collaboration with the research group of Jacqueline Bloch. In 2002 the evidence for nonequilibrium condensation was reported [157]. The paper claimed a phase transition from a classical thermal mixed state to a quantum-mechanical pure state of EPs in GaAs multiple QW microcavity. It was supported by the decrease of the second-order coherence function, a nonlinear threshold behavior in the pump-intensity dependence of the emission, a polariton-like dispersion relation above threshold, and a decrease of the relaxation time of the lower polariton state. However, the widely accepted claim of nonequilibrium Bose–Einstein condensation of polaritons was achieved by the group of Benoit Deveaud in 2006 [158].

The work presented the compelling evidence for BEC of polaritons, among which are the observation of massive occupation of the ground state by a polariton gas at thermal equilibrium at 19 K, an increase of temporal coherence, long-range spatial coherence and linear polarization. Since the BEC of dilute atom gas of rubidium atoms happens at the temperatures below 200 nanokelvin and the EP quasiparticles are 109 times lighter, the theoretical estimate for the critical temperature is of order of cryogenic temperatures.

The sample studied in the work consists of a CdTe/CdMgTe microcavity grown by molecular beam epitaxy and contained 16 quantum wells displaying a vacuum field Rabi splitting of 26 meV. The continuous-wave Ti:Sapphire laser was used to excite the microcavity with an acousto-optic modulator. Aside from the [158] the set of additional paramaters, like pulse duration, laser beam shape, excitation energy, DBR parameters can be found in the PhD thesis of Jacek Kasprzak [159]. Fig. 2-4 and Fig. 2-5 show the schematic QW structure together with the different dispersion branches with the corresponding technical parameters.

After this significant demonstration, the research on EP BEC did not decline, but began to gain momentum. In 2007, the group of David Snoke demonstrated nonequilibrium BEC of polaritons in a trap [160]. The group of Jaqueline Bloch observed polariton condensation in 2009 [161] after which many other experimental groups shifted their attention to different aspects of EP BEC. Alberto Amo reported the the polariton superfluidity later [162]. Moreover, this effect was demonstrated more recently at room temperature [163] which is the first manifestation of room temperature superfluidity.

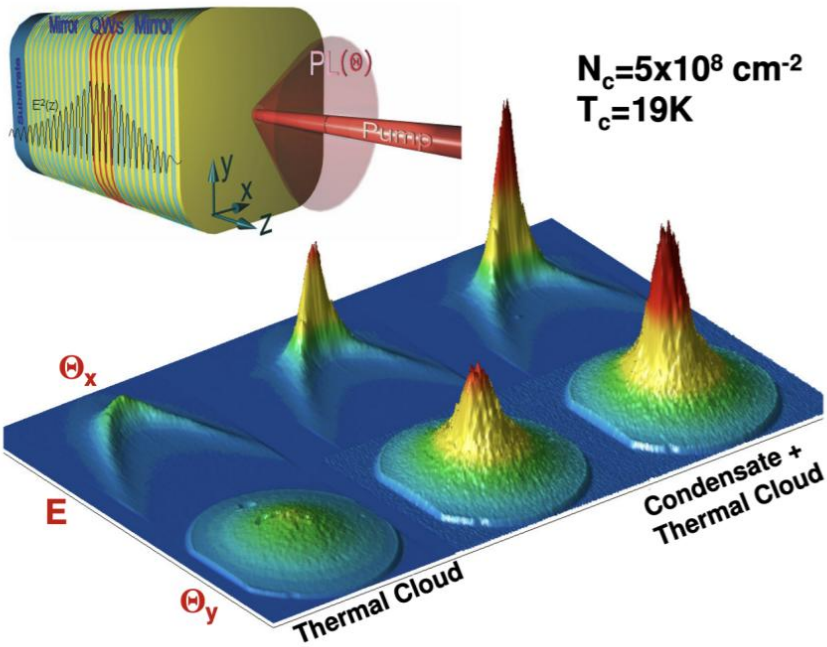


Figure 2-4: QW picture with the corresponding pumping and field distribution. The lower picture represents the momentum distribution, similar to one found in BEC, but in the EP system.and the Schematic picture representing the criterion for Bose-Einstein condensation. Schematic picture is taken from the PhD thesis [159].

2.2.5 Polariton condensate notable properties

This subsection lists several alluring properties, which represent a significant interest from considering both practical and fundamental perspectives.

The previous subsection gave the theoretical description of the EP system and the corresponding characterization in dispersion relations. It was shown that one of the distinctive features of EPs is their *light effective mass*, orders of magnitude lower than the bare electron has. That means the corresponding de Broglie wavelength is larger than the electron and the BEC condensation criterion for the critical temperature and density is easier to achieve. The ordinary $T_c = 10K$ using materials such as GaAs and CdTe, which leads to the fascinating fact that using different materials can achieve the room-temperature manifestation of the BEC effects.

Using GaN, ZnO and other organic semiconductors—polariton condensates at higher temperatures, including *room temperature*, have already been realized [164, 165, 166]. This broadens the range of applicability for the EPs. The potential practical interest allows one to use this platform in an environment where different quantum systems can hardly be exploited.

As the necessary confinement for the BEC of EP production was discussed before, EP condensates essentially form two-dimensional structures. This allows one to investigate the exciting physics aspects connected with the particular two dimensions. One such phenomenon is the continuous-wave Berezinskii–Kosterlitz–Thouless (BKT) transition [167] of the two-dimensional XY model in statistical physics. It is a transition from bound vortex-antivortex pairs to unpaired vortices and anti-vortices at the critical temperature, where the interplay of long-range order and thermally excited vortices is balancing.

The BEC itself, when the particles spontaneously occupy the ground state, appears a very interesting phenomenon with many corresponding interesting properties, among which are the *thermalization* mechanism, the existence of the *macroscopic wavefunction* forming an order parameter with the off-diagonal *long-range order* is a central concept.

The consequent property of the *superfluidity*, similar to the 4He realization or superconductivity, viewed as a condensation of Cooper pairs. This spectacular property manifests itself in the frictionless flow due to the spontaneous coherence or the impossibility of rotating the superfluid.

The spatial coherence and the wave matter nature result in the existence of the nontrivial phases of polariton condensates and the existence of quantized vortices, which are topological defects with zero density at its core and a multiple of phase rotation around it. Single vortices and vortex pairs in EP condensates excited by non-resonant pumping have been observed either pinned at defects or imprinted from the phase fluctuation of lasers. Vortices are detected via

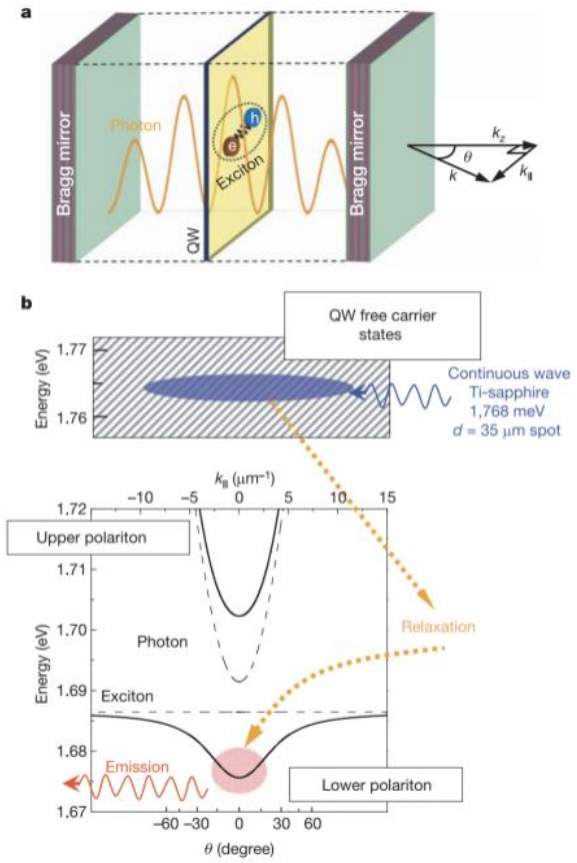


Figure 2-5: The picture showing the microcavity diagram and energy dispersion from the seminal work [158]. (a) A microcavity is a planar Fabry–Perot resonator with two Bragg mirrors at resonance with excitons in QWs. (b) Energy levels as a function of the in-plane wavevector k in a CdTe-based microcavity. Interaction between exciton and photon modes, with parabolic dispersions (dashed curves), gives rise to lower and upper polariton branches (solid curves) with dispersions featuring an anticrossing typical of the strong coupling regime.

the phase output, where it is possible to reconstruct the spatial phase and intensity distributions.

Despite the limitation for studying the standard BEC physics, the dynamical nature of a polariton condensate provides a new experimental tool for the research of the nonequilibrium open system consisting of interacting bosons. In other words, the system is essentially *gain-dissipative*. Through the photon leakage, it is possible to access the condensate's quantum-statistical properties through the experiment. Alternatively, the realization of the optical lattices in cold atomic gases allowed one to manipulate BEC states, with the experiments such as the superfluid–Mott insulator transition [168].

The creation of periodic potentials can lead to applications concerning quantum simulation, where complicated quantum many-body systems are realized, which can hardly be accessible by the reach of computer simulations [169]. The corresponding branch aims to recreate the exotic Hamiltonians for models of metamaterials that have no natural realization. Many engineering methods were created to shape in-plane potentials by modifying either the photonic or excitonic modes. Among them are chemical etching for forming pillars, strips and 2D lattices, laser pump spot formation, piezoelectric acoustic lattices - for the complete list, see the review [170]. The exciting question property is the quantum and classical noise description in the EP system, allowing one to improve the analog simulators.

2.3 Alternative platforms

In the next section, we will show how to utilize the lattices of EP condensates to efficiently simulate the XY Hamiltonian when operating at the condensation threshold. The system time evolution process tends to minimize the amount of matter, which depends on the condensates' relative phases. The underlying mechanism dramatically relies on the macroscopic wave functions' interference, produced by the external sources in the gain-dissipative setting. Using this mechanism to our advantage, it is possible to create special-purpose hardware, which operates like simulated annealing due to the equations' gradient terms. It is possible to extend the system parameters to minimize Ising and q-state Potts models [171].

To exploit the system effectively, one needs to map an optimization problem of interest into the QUBO. There is the possibility to do so and into the connection matrix of the Ising network. The problem of finding the optimal solution of a QUBO problem reduces to finding the ground state of the Ising Hamiltonian, which can be related to finding the 'maximum occupancy' of the collective supermode of the underlying network, as a system-specific gain mechanism is continuously increased to reach the coherence threshold [172, 68].

There are lots of similar platforms that are operating in an analogic way. A small part of them was mentioned before in the background section. Here we give an additional list of them with little discussion, which mostly operates in the same annealer-like manner.

In recent years, much effort has been devoted to developing such platforms that act as quantum or classical analog simulators aimed at solving certain classes of hard classical optimization problems [57, 58, 59, 60, 61, 62, 63]. It is expected that these kinds of platforms would help to efficiently solve many tasks of significant computational complexity, ranging from modeling microscopic effects and processes like the behavior of electrons in complex materials [64, 65] and finding the ground state of spin glasses [66], to the applied combinatorial optimization problems such as the TSP [67].

These activities have a reverse effect, which spawns the algorithmic analogs like the Simulated Bifurcation algorithms inspired by quantum adiabatic optimization using a nonlinear oscillator network [173], destabilization of local minima based on degenerate parametric oscillator networks [174], parallel tempering Monte Carlo [175], and the gain-dissipative algorithm based on the operation of the polariton graph simulator at the condensation threshold [171].

Nonequilibrium condensates, optical parametric oscillators, lasers, memristor crossbars, and other platforms have been considered annealing-inspired accelerators and demonstrated success in finding the ground state of spin Hamiltonians with continuous or discrete variables [58, 62, 68, 176]. In particular, the coherent Ising machine has been shown to significantly outperform classical simulated annealing in terms of accuracy and computation time to solve MAX-CUT problems [58] efficiently. It has shown better scalability than the quantum annealers [177]. Memristor - Hopfield NN with massively parallel operations performed in a dense crossbar array was shown to solve NP-hard MAX-Cut problems predicting over four orders of magnitude advantage over digital and optical hardware [176]. Integrated photonic circuits that use self-phase modulation in two microring resonators were shown to act as an optical coherent Ising machine [178, 179]. Research groups are also utilizing ultracold bosonic and fermionic atoms, molecular gases in optical lattices [180, 181, 182, 183] photons [184], trapped ions [185, 186] and superconducting q-bits [187]. The separate part with the detailed description of operational regimes of lasers, coherent Ising machine and photonic networks can be found in Chapter 6 with additional information.

Another subdomain of such works is closely connected with reservoir computing, which also exploits different physical settings, but does it in a unique way, similar to recurrent NNs. This topic will be discussed separately in 3.5.5.

Chapter 3

System design

This chapter considers several ways of manipulating the system parameters to achieve a certain performance with the correspondence of the existing computational task. One can view it as a system design or alternatively, the description of how to manipulate the system to use it in a certain way.

3.1 Quadratic programming/minimization

This section will cover the principal results concerning the process of simulating classical XY Hamiltonian on a polariton platform, which (results) were considered in great detail in [188, 68] with the particular emphasis on an analytical solution for a single condensate and obtaining the XY model. Other physical systems have previously simulated the XY model, particularly - ultracold atomic optical lattices [189] and coupled photon lasers network [62]. To design an analog Hamiltonian, one needs first to map the target Hamiltonian into the elements of the simulator and the corresponding interactions between the elements, prepare the simulator in a state that is relevant to the problem of interest, and finally, perform the measurements on the simulator with the required precision.

The XY Hamiltonian's choice is motivated by the critical work, tailoring the universal models with the classical spin models [190]. In general, the XY model can be treated as a quadratic constrained optimization model, which is known to be an NP-hard problem for non-convex and dense matrices [191].

Polariton graph configuration has an enormous degree of flexibility in engineering the geometrical configuration of elements and so the corresponding interactions between them, going beyond a standard lattice configuration. As well as the spatial coordinates, more parameters can be changed for manipulating the system setup; see for example [192].

After preparing the simulator in the relevant state, it is possible to realize the initial configuration, searching for the global minimum of the XY Hamiltonian through the dynamic evolution via a bottom-up approach, potentially having an advantage over classical or quantum annealing methods.

The steady-state of such a system is reached through the balancing of pumping and dissipation of photons. Their decay (through the Bragg reflectors) carries all necessary information of the corresponding polariton state wavefunction. It is enough for a polariton condensate characterization in its steady state.

The next step is to explain how a polariton condensate's particular arrangement can lead to the reproduction of the XY model starting from the mean-field model of polariton condensates and derive analytical solutions for a single condensate. As was discussed in the previous chapters, the mean field of polariton condensates can be described by the driven-dissipative GP equation, coupled to the rate equation for the density of the exciton reservoir n_R. After the assumption of the stationary state of the reservoir $\frac{\partial n_R}{\partial t} = 0$, rewriting the equations lead us to the form of the complex Ginzburg-Landau equation (cGLE) with a saturable nonlinearity:

$$i\frac{\partial \Psi}{\partial t} = -(1 - i\eta_d n_R)\Delta\Psi + |\Psi|^2\Psi + g n_R \Psi + i[n_R - \gamma]\Psi, \tag{3.1}$$

$$n_R = p(r)/(1 + b|\Psi|^2), \tag{3.2}$$

Where η_d is s the energy relaxation coefficient specifying the rate at which gain decreases with increasing energy, g is the strength of effective polariton-exciton interactions and γ is the is the rate of the condensate losses in the cavity. By taking the Taylor expansion for small densities in the expression for the reservoir n_R and we arrive at the different form of the cGLE:

$$i\frac{\partial\Psi}{\partial t} = -(1-i\eta p)\Delta\Psi + (1-gbp)|\Psi|^2\Psi + gp\Psi + i[p-\gamma-pb|\Psi|^2]\Psi, \tag{3.3}$$

Considering the approximation of the steady state solutions for a radially symmetric Gaussian pumping profile $p(r) = p_0 \exp\left(-\sigma r^2\right)$, with the p_0 being the maximum pumping intensity and σ is the inverse width of the Gaussian, one can use the Madelung transformation $\Psi = \sqrt{\rho}\exp[\mathrm{i}S]$ to separate the real and imaginary parts of equation 3.3 :

$$\frac{1}{r\rho}\frac{\mathrm{d}(r\rho u)}{\mathrm{d}r} = \frac{p(r)}{1+b\rho}\left(1 + \frac{\eta\left(r(\sqrt{\rho})'\right)'}{r\sqrt{\rho}} - \eta u^2\right) - \gamma \tag{3.4}$$

$$\mu = -\frac{(\sqrt{\rho})''}{\sqrt{\rho}} - \frac{(\sqrt{\rho})'}{r\sqrt{\rho}} + u^2 + \rho + \frac{p(r)}{1+b\rho}\left(g - \frac{\eta}{r\rho}\frac{\mathrm{d}(r\rho u)}{\mathrm{d}r}\right) \tag{3.5}$$

In the asymptotic behaviour where $p(r) = 0$, the velocity $u = |\mathbf{u}| = \frac{dS}{\mathrm{d}r}$ is given by the outflow wavenumber $k_c = const$ with $r\rho_r + \rho = -\gamma\rho r/k_c$, which gives $\rho \sim \exp\left[-r\gamma/k_c\right]/r$ after the integration. The chemical potential at the infinity is given by $\mu = k_c^2 - \gamma^2/4k_c^2$. Utilizing these equations, we can write the density and the velocity in the following forms:

$$\rho(r) = \frac{a_0}{\gamma r \exp\left(\gamma r k_c^{-1}\right) k_c^{-1} + \xi - \gamma r/k_c + a_3 r^3} \sim a_0 k_c \exp\left[-r\gamma/k_c\right]/r\gamma, r \to \infty, \tag{3.6}$$

$$u(r) = k_c \tanh\left(lr/k_c\right) \sim k_c, r \to \infty, \tag{3.7}$$

using their behaviour at the initial spot and infinity. The next step is to parametrize the potential solutions with ξ, a_0, a_3 and l. One can find them using the previous equations and matching with the asymptotic. The expansion to $O(r^2)$ of equation 3.4 determines k_c, while the expansions to $O(r^2)$ of equation 3.5 define the remaining parameters through two nonlinear equations. The details of the calculations together with the final coefficients can be found in the original article [188].

Equations 3.6 and 3.7 with the parameters found for the given system together with the pumping parameters p_0 and σ fully specify the approximate analytical solution of equations 3-1 (see Figs. 3-1).

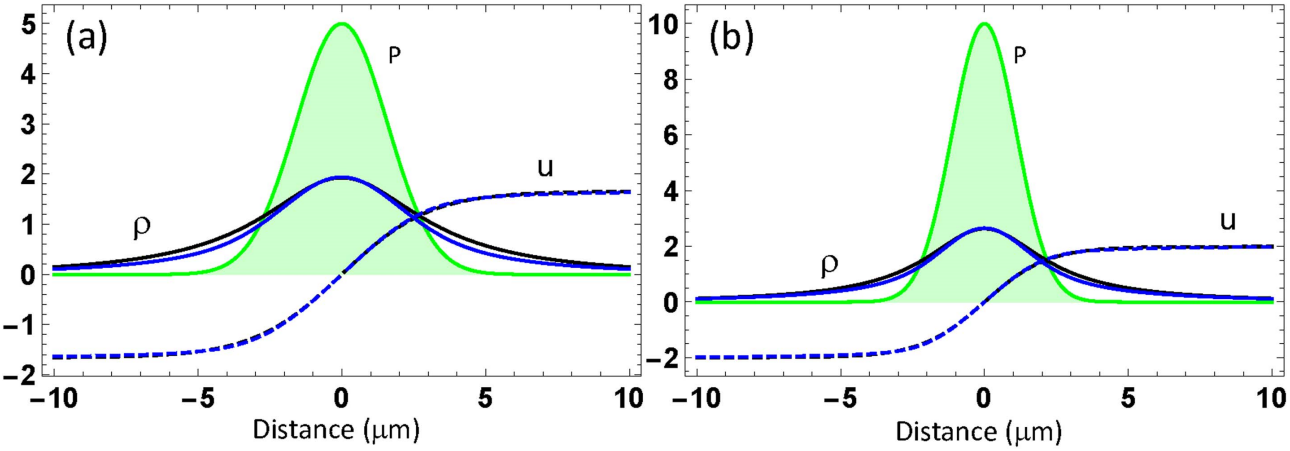


Figure 3-1: Approximate analytical (blue lines) and numerical (black lines) solutions for density (solid lines) and velocity (dashed lines) of equation 3.3 for the pumping profile given by $p(r) = p_0 \exp\left(-\sigma r^2\right)$ (green shaded area). The system parameters are $b = 1.5, \gamma = 0.2, g = 0.5$ and (a) $\sigma = 0.2, p_0 = 5$;(b)$\sigma = 0.4, p_0 = 10$. The picture is taken from the [188].

A spatial light modulator (SLM) can be used to pump condensates at the vertices of a distributed graph via $p(\mathbf{r}) = \sum_{i=1}^{N} p_i \exp\left[-\sigma_i\left|\mathbf{r}-\mathbf{r}_i\right|^2\right]$. Assuming that all vertices are pumped identically and additionally assuming that all condensates are well-separated, we can use the linear approximation for the the resulting condensate wavefunction $\Psi_N(\mathbf{r},t) \approx \sum_{i=1}^{N} \Psi_0\left(|\mathbf{r}-\mathbf{r}_i|\right)\exp\left(\mathrm{i}\theta_i\right)$, where Ψ_0 is the single condensate solution, which is given by:

$$\Psi_0(r) = \sqrt{\rho_0(r)}\exp\left[\mathrm{i}\frac{k_c^2}{l}\log\left(\cosh\left(\frac{l}{k_c}r\right)\right)\right] \sim \sqrt{\rho_0(r)}\exp\left[\mathrm{i}k_c r\right], r \to \infty. \tag{3.8}$$

The total amount of matter $\mathcal{M}$ in the system can be written:

$$\mathcal{M} = \int |\Psi_N|^2\,\mathrm{d}\mathbf{r} = \frac{1}{(2\pi)^2}\int\left|\tilde{\Psi}_N(\mathbf{k})\right|^2\mathrm{d}\mathbf{k} = 2\pi N\int_0^\infty |\Psi_0|^2\, r\mathrm{d}r + \sum_{i<j} J_{ij}\cos\left(\theta_i-\theta_j\right), \tag{3.9}$$

$$J_{ij} = \frac{1}{\pi}\int_0^\infty \left|\tilde{\Psi}_0(k)\right|^2 J_0\left(k\left|\mathbf{r}_i - \mathbf{r}_j\right|\right) k\mathrm{d}k, \tag{3.10}$$

where the modulus expression is the Fourier transform of the original macroscopic wavefunction:

$$\tilde{\Psi}_N(\mathbf{k}) = \int \exp(-\mathbf{ik}\cdot\mathbf{r})\Psi_N(\mathbf{r})\mathrm{d}\mathbf{r} = \tilde{\Psi}_0(k)\sum_{i=1}^{N}\exp\left(\mathbf{ik}\cdot\mathbf{r}_i + \mathrm{i}\theta_i\right), \tag{3.11}$$

where $\tilde{\Psi}_0(k) = 2\pi\int_0^\infty \Psi_0(r)J_0(kr)rdr$ and J_0 is the Bessel function. Since the system tends to maximize the total number of particles given by the Equation 3.9, this is equivalent to minimizing the XY Hamiltonian (the last term in the expression).

Such polariton graphs offer the excellent scalability of optical lattices. The condensates' arrangement allows one to study disordered systems and control the sign and the strength of the coupling for each element independently. There are lots of variations for the spin configurations. The coupling between the adjacent polaritons can be modified by either varying the distance between the pumping spots or changing the individual spots' pumping parameters. Additionally, the cGLE can be reformulated as the Kuramoto model which is a paradigm for describing the spontaneous emergence of collective synchronization [193].

The simpler perspective is to consider ordinary one-dimensional Schrodinger equation with the delta-type potential $\delta(x)$, for which there is a symmetric solution, described by the exponential function. Adding the phase shift due to the time evolution, we can approximate the condensate by the function $\exp(-\mathbf{i}k\cdot|x|)\exp(-\beta\cdot|x|)$. Calculating the density of two condensates, we come to the underlying interference pattern responsible for the correspondence between the EP system reproducing the XY Hamiltonian.

3.2 Tensor sum minimization and complex coupling switching

Gain-dissipative platforms consisting of lasers, optical parametric oscillators and nonequilibrium condensates operating at the condensation/coherence threshold have been recently proposed as efficient analog simulators of 2-local spin Hamiltonians with continuous or discrete degrees of freedom. In this section, it will be shown that nonequilibrium condensates *above the threshold* arranged in an interacting network may realize k-local Hamiltonians with $k > 2$ and lead to nontrivial phase configurations. The derivation of the rate equations in the generalization settings for the higher-order binary optimization problems will be repeated. Additionally, the method of facilitating the search for the global solution by invoking complex couplings in the system will be presented, and the efficiency of the method on the sets of complex problems will be demonstrated.

Discrete Ising 'spins' or continuous XY 'spins' are encoded in individual phase modes of the nonlinear networks in all the presented technological platforms. An optimization problem of interest is mapped into the quadratic unconstrained binary optimization (QUBO) and, therefore, into the Ising network's connectivity matrix. The problem of finding the optimal solution of a QUBO problem reduces to finding the ground state of the Ising Hamiltonian, which can be related to finding the 'maximum occupancy' of the collective supermode of the underlying network, as a system specific gain mechanism is continuously increased to reach the coherence threshold [172, 68].

The focus of all these technological and inspired implementations of the annealer-based optimization has been on QUBO; however, there is a large class of optimization problems — the high-order binary optimization (HOBO) – that are more naturally encoded by the k-local Hamiltonians [194, 195]. HOBO is concerned with optimizing a (high degree) multivariate polynomial function in binary variables. A basic model is to maximize or minimize a k-th degree polynomial function $f(\mathbf{x})$ where $\mathbf{x} = (x_1, ..., x_i, ..., x_n)$, $x_i \in \{-1, +1\}$. HOBO examples are ubiquitous from Hypergraph max-covering problem to Frobenius and "market split" problems [195]. HOBO is a fundamental problem in integer programming and is also known as the Fourier support graph problem. Any HOBO can be mapped into the QUBO [190]; however, the overhead in the number of nodes becomes prohibitive in an actual technological platform, so it is important to consider ways to solve HOBO directly. The purpose of this section is three-fold. First, I show that Ising machines based on nonequilibrium condensates can be used to address 4-local HOBO when operating *above* the threshold. Secondly, inspired by the operation of the nonequilibrium networks, a new optimization algorithm for solving HOBO of arbitrary degree will be proposed. Finally, I show that another physics-inspired method of turning on and off the complex coupling between the nonlinear condensates greatly enhances the global minimum search.

Polynomial optimization with coherent networks. The optimization problem studied in this section is

$$\min_{\mathbf{x}\in\{-1,+1\}^N} -\sum_{\Omega}\mathbf{A}^k_{i_1,i_2,\cdots,i_k}x_{i_1}x_{i_2}\cdots x_{i_k}, \tag{3.12}$$

where $\Omega = \{i_j : 1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq N\}$ and $\mathbf{A}^k$ is the super-symmetric tensor of degree k.

To formulate the gain-dissipative platform that reaches the ground state of HOBO by finding the 'maximum occupancy' collective supermode of the underlying network of nonequilibrium condensates we consider the mean-field equations that govern such a network based on the Ginzburg-Landau equation [196, 197]. This is a universal driven-dissipative equation that describes the behaviour of systems in the vicinity of a symmetry–breaking instability and has been used to describe lasers, thermal convection, nematic liquid crystals, and various non-equilibrium condensates [198, 199]. When derived asymptotically from a generic laser model given by Maxwell-Bloch equations it has a saturable nonlinearity and can be written as

$$\mathrm{i}\frac{\partial\psi}{\partial t} = -\nabla^2\psi + \tilde{U}|\psi|^2\psi + i\left(\frac{P(\mathbf{r},t)}{1+b|\psi|^2} - \gamma_c\right)\psi, \tag{3.13}$$

where $\psi(\mathbf{r},t)$ is the wavefunction of the system, $\tilde{U}$ is the strength of the delta-function interaction potential, γ_c is the rate of linear losses, b parametrizes the effective strength of nonlinear losses, $P(\mathbf{r},t)$ describes the gain mechanism that adds particles to the system. First of all, we assume that $b \ll U$, so that the saturation term can be replaced by its Taylor expansion:

$$\frac{P(\mathbf{r},t)}{1+b|\psi|^2} \approx P(\mathbf{r},t) - P(\mathbf{r},t)b|\psi|^2. \tag{3.14}$$

We define by $p(\mathbf{r}) = P(\mathbf{r},t)$ the injection profile that gives rise to a single condensate centered at the origin that is described by a normalized wavefunction $\phi(\mathbf{r})$. Mathematically, $\phi(\mathbf{r})$ satisfies

$$\mu\phi = -\nabla^2\phi + \tilde{U}|\phi|^2\phi + i\left(p - pb|\phi|^2 - \gamma_c\right)\phi, \tag{3.15}$$

$$\int_\Gamma |\phi(\mathbf{r})|^2\, d\mathbf{r} = 1, \tag{3.16}$$

where μ is the chemical potential (the Lagrange multiplier) and Γ is the entire system space. Based on the wavefunction of the single isolated condensate, we can construct an approximation for N localized condensates noting that the well-separated condensates interact by the outflow of the particles from the positions where they were created [198, 199]. This is in a contrast with the conservative condensates, such as ultracold atomic Bose-Einstein condensates, where spatially separated condensates (with separation much larger than the condensate width) do not interact.

We shall assume that pumping $P(\mathbf{r},t)$ adds particles in N spatial locations centered at $\mathbf{r}_i$, $i = 1, ..., N$, so that $P(\mathbf{r},t) = \sum_i f_i(t)p_i(\mathbf{r})$, where f_i is the time-dependent part of the pumping at the position $\mathbf{r} = \mathbf{r}_i$ and $p_i(\mathbf{r}) \equiv p(\mathbf{r} - \mathbf{r}_i)$. If the distances between the neighbouring condensates are larger than the width of $p(\mathbf{r})$, we employ the tight binding approximation and write the wavefunction of the system as a linear superposition of the wavefunctions of the individual condensates $\psi(\mathbf{r},t) \approx \sum_{i=1}^N a_i(t)\phi_i(\mathbf{r})$, where $a_i(t)$ is the time-dependent complex amplitude and $\phi_i(\mathbf{r}) \equiv \phi(\mathbf{r} - \mathbf{r}_i)$ [200, 171]. We substitute the expressions for P and ψ into Eq. (3.13) with the Taylor expansion of the saturation term given by Eq. (3.14), multiply by ϕ_j^* for $j = 1, ..., N$ and eliminate the spatial degrees of freedom by integrating in the entire plane Γ. The time evolution of the individual functions $a_i(t)$ will separate if we assume that the integrals that involve products of the wavefunctions of the separated condensates are negligible compared with the integrals of the products of the same condensates or with the integrals that involve the pumping profiles that overlap with ϕ_i, so that

$$\int_\Omega \phi_i\phi_j^*\, d\mathbf{r} \ll \int_\Omega |\phi|^2\, d\mathbf{r} = 1 \quad \text{for} \quad i \neq j, \tag{3.17}$$

$$\int_\Omega \phi_j^*\nabla^2\phi_i\, d\mathbf{r} \ll \int_\Omega \phi^*\nabla^2\phi\, d\mathbf{r} \equiv d \quad \text{for} \quad i \neq j \tag{3.18}$$

$$\int_\Omega \phi_i\phi_j^*\, d\mathbf{r} \ll \int_\Omega p_m\phi_i\phi_j^*\, d\mathbf{r} \quad \text{for} \quad i \neq j, m \in \{i,j\}, \tag{3.19}$$

$$\int_\Omega \phi_i\phi_j^*\phi_k\phi_m^*\, d\mathbf{r} \ll \int_\Omega p_m\phi_i\phi_j^*\phi_k\phi_l^*\, d\mathbf{r}, \quad m \in \{i,j,k,l\}, \quad \text{etc.}$$

The validity of these assumptions can be verified using asymptotics developed in [188] where it was shown that ϕ created with a Gaussian pump can be approximated by

$$\phi(r) = \sqrt{\frac{2}{\pi}}\beta\exp[-\beta r + ik_c r], \tag{3.20}$$

where k_c is the outflow velocity and β is the inverse characteristic width of the condensate [188]. The integrals $\chi_{ij} =$

$\int \phi_i \phi_j^* \, d\mathbf{r}$ for $i \neq j$ can be evaluated using the elliptical coordinates in terms of the Bessel functions [201]

$$\chi_{ij} = 2\beta^2 l_{ij} \left[\frac{1}{\beta} J_0(k_c l_{ij}) K_1(\beta l_{ij}) + \frac{1}{k_c} J_1(k_c l_{ij}) K_0(\beta l_{ij}) \right], \tag{3.21}$$

where $l_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. We assumed that the condensates are well separated, $l_{ij}\beta \gg 1$, so that for $i \neq j$ we have $\chi_{ij} \ll 1$ as follows from Eq. (3.21). The correctness of other assumptions can be established in a similar manner. Under these assumptions, the tight binding approximation of Eq. (3.13) leads to N equations

$$\begin{aligned} \frac{da_i}{dt} \int_\Gamma |\phi|^2 \, d\mathbf{r} &= i a_i d - iU |a_i|^2 a_i \int_\Gamma |\phi|^4 \, d\mathbf{r} + a_i (f_i \int_\Gamma p |\phi|^2 \, d\mathbf{r} - \gamma_c \int_\Gamma |\phi|^2 \, d\mathbf{r}) \\ &+ \sum_{j, j \neq i} a_j f_j \int_\Gamma p_j \phi_j \phi_i^* \, d\mathbf{r} - b^{-1} \sum_{m \in \{i,j,k,l\}, j,k,l} f_m \int_\Gamma p_m \phi_j \phi_k \phi_l^* \phi_i^* \, d\mathbf{r} \, a_j a_k a_l^* . \end{aligned}$$

We use normalization Eq. (3.16) and introduce $\Psi_i = a_i \exp(\mathrm{i}td)$ to absorb the first term on the right hand side into the phase of Ψ_i. We also denote $\gamma_i = f_i \int_\Gamma p |\phi|^2 \, d\mathbf{r} - \gamma_c$, $U = \tilde{U} \int_\Gamma |\phi|^4 \, d\mathbf{r}$, $J_{ij} = f_j \int_\Gamma p_i \phi_j \phi_i^* d\mathbf{r}$ for $j \neq i$, $\sigma_i = b f_i \int_\Gamma p |\phi|^4 \, d\mathbf{r}$ and $Q_{ijkl} = -b \sum_{m \in \{i,j,k,l\}} f_m \int_\Gamma p_m \phi_k \phi_l^* \phi_j \phi_i^* \, d\mathbf{r}$, where $i = j = k = l$ is excluded from the summation. With this notation, we obtain N equations of the form

$$\frac{d\Psi_i}{dt} = \Psi_i \left(\gamma_i - (\sigma_i + \mathrm{i}U) |\, \Psi_i \,|^2 \right) + \sum_{j, j \neq i} J_{ij} \Psi_j + \sum_{\langle j,k,l \rangle} Q_{ijkl} \Psi_j \Psi_k \Psi_l^* + D \xi_i(t) . \tag{3.22}$$

where $\langle i, j, k \rangle$ denotes the permutations of $\{j, k, l\}$ that exclude $j = k = l = i$.

In writing Eq. (3.22) we also introduced the Langevin noise $\xi_i(t)$ ($\langle \xi_i(t) \xi_i^*(t') \rangle = \delta(t - t')$) which represents intrinsic vacuum fluctuations and classical noise with a diffusion coefficient D. The rate equations on $\Psi_i(t)$ take the form similar to what we obtained for a polaritonic networks at the condensation threshold [69], but now involve higher order terms represented by the super-symmetric tensor $\mathbf{Q}$. At the condensation threshold these terms can be neglected, however, above the threshold these terms allow to minimize the higher order k-local Hamiltonians. To see this, we rewrite Eq. (3.22) in terms of the number densities ρ_i and phases θ_i using the Madelung transformation $\Psi_i = \sqrt{\rho_i} \exp[\mathrm{i}\theta_i]$ for simplicity excluding the noise:

$$\begin{aligned} \frac{1}{2} \dot{\rho}_i(t) &= (\gamma_i - \sigma_i \rho_i) \rho_i + \sum_{j, j \neq i} J_{ij} \sqrt{\rho_i \rho_j} \cos \theta_{ij} \\ &+ \sum_{\langle j,k,l \rangle} Q_{ijkl} \sqrt{\rho_i \rho_j \rho_k \rho_l} \cos \theta_{ijkl}, \end{aligned} \tag{3.23}$$

$$\begin{aligned} \dot{\theta}_i(t) &= -U \rho_i - \sum_{j, j \neq i} J_{ij} \frac{\sqrt{\rho_j}}{\sqrt{\rho_i}} \sin \theta_{ij} \\ &- \sum_{\langle j,k,l \rangle} Q_{ijkl} \frac{\sqrt{\rho_j \rho_k \rho_l}}{\sqrt{\rho_i}} \sin \theta_{ijkl}, \end{aligned} \tag{3.24}$$

where $\theta_{ij} = \theta_i - \theta_j$ and $\theta_{ijkl} = \theta_i + \theta_l - \theta_k - \theta_j$.

Equation (3.24) describes the evolution of the higher order Kuramoto oscillators. The higher order terms affect the states even in the simplest configuration of two identical oscillators pumped with $\gamma_i = \gamma$ for which the occupancy $\rho_0 = \rho_1 = \rho_2$ at the fixed point of Eqs. (3.23-3.24) reads $\rho_0 = [\gamma + J \cos \Delta\theta + \tilde{Q} \cos(2\Delta\theta)]/\sigma$ where $\Delta\theta = \theta_1 - \theta_2$ and $\tilde{Q} = \rho_0 Q$. By choosing the minimum pumping γ to reach the required occupancy, we minimize the Hamiltonian $H_{\text{two}} = -J \cos \Delta\theta - \tilde{Q} \cos(2\Delta\theta)$ while Eq. (3.24) describes the gradient descent to the local minimum of H_{two}. If $\tilde{Q}$ is negligible, we have the minimization of the XY Hamiltonian, so $\Delta\theta = 0$ or π if $J > 0$ or $J < 0$ respectively. The same minimum is realised if $\tilde{Q}$ is present but has the same sign as J. However, a different phase difference is realised when $J\tilde{Q} < 0$ and $|\tilde{Q}/J| \geq \frac{1}{4}$, namely $\Delta\theta = \arccos(-J/4\tilde{Q})$.

In the example of two oscillators the stationary state with equal occupancy of the nodes is always reached. However, in a more general system with many oscillators, unless the oscillatory network is highly symmetric (all oscillators have equal number of connections of the same strength with other oscillators) the systems breaks into subsystems characterised by different frequencies. To guarantee the full synchronisation of the network we need to choose the injection rates in such a way that all oscillators have the same occupancy ρ_{th} [69]. For instance, this can be achieved by adjusting the pumping rates dynamically, depending on the occupancy of the $i-$th oscillator at time t: $\dot{\gamma}_i = \epsilon(\rho_{\text{th}} - \rho_i)$, where the parameter ϵ characterizes how fast γ_i adjusts to changes in ρ_i. Aiming at the algorithmic implementation, we will focus on tensors of the same order k, and leave the problems with mixed order tensors for future work. In case of polariton condensates,

this means that the fourth order tensors dominate the dynamics of the second-order terms. With appropriate density adjustments, as described above, the system of N oscillators will always synchronize and achieve the stationary minimum of the Hamiltonian $H = -\sum_i \sum_{\langle j,k,l\rangle} \tilde{Q}_{ijkl} \cos\theta_{ijkl}$ with super-symmetric tensor of $k = 4$.

To replace the minimization in the space of continuous spins with binary states, one could combine nonresonant pumping with resonant at twice the frequency of the condensate which introduces the terms proportional to ψ^* (Ψ_i^*) to the right-hand side of Eq. (3.13) (Eq. (3.22)) similarly to $k = 2$ case [69]. Resonant and nonresonant excitations have been previously combined in experiments on polariton condensates using chemical etching across the sample allowing resonant excitation from the back side of the cavity. The resonant excitation laser was synchronized with the nonresonant one and with the condensate frequency [202]. Such combination of resonant and nonresonant excitations would lead to the realisation of the minimum of the k-local Hamiltonians, so to solving Eq. (3.12) with the binary spins $x_i = \cos\theta_i$, where θ_i are limited to 0 and π values due the action of the resonant forcing. However, in contrast with $k = 2$ case, the phase projections on the binary states for $k > 2$ is automatic thanks to the mixture of Ψ_{i_j} and $\Psi^*_{i_k}$ present in the tensor form, and the presence of the additional resonant field in not necessary. We elucidate the reasons for this below.

Physics-inspired optimization. The formulated principle of coherence formation at a minimum of a spin Hamiltonian formulated above inspires an efficient algorithm for finding the global minimum of HOBO. For this, we extend and simplify Eq. (3.22) to capture the mechanism of relaxation to the minimum of the HOBO but without the necessity to capture full physics of the actual system. The minimum of HOBO for N binary variables can be found by numerical integration of $2N$ equations

$$\frac{d\Psi_l}{dt} = \Psi_l(\gamma_l(t) - |\Psi_l|^2) + \sum_{\bar{\Omega}} \mathbf{A}^k_{i_1,i_2,\cdots,l,\cdots i_k} \Psi_{i_1}\Psi_{i_2}\cdots\Psi^*_{i_k}, \tag{3.25}$$

$$\frac{d\gamma_l}{dt} = \epsilon(\rho_{\rm th} - |\Psi_l|^2), \tag{3.26}$$

where $\bar{\Omega} = \Omega/l$ and the initial values for pumping strength $\gamma_l(t=0) = -\max_{1\le l\le N}\sum_{\bar{\Omega}} |\mathbf{A}^k_{i_1,i_2,\cdots,l,\cdots i_k}|$. At the fixed point, the imaginary part of Eq. (3.25) gives a set of linear equations such that the l-th equation involves superposition of $\sin(\sum_{i_j\neq\{l,i_k\}} \theta_{i_j} - \theta_{i_k} - \theta_l)$ that has to be equal to zero. In general, the only way for the system to satisfy these equations is to bring all phases θ_l to take on 0 or π. The total occupancy of the system at the fixed point is found from the real part of Eq. (3.26) and is equal to $N\rho_{\rm th}$, so that $N\rho_{\rm th} = \sum_{l=1}^N \gamma_l + \sqrt{\rho_{\rm th}}^{k-2} \sum_\Omega \mathbf{A}^k_{i_1,i_2,\cdots,i_k} \cos(\theta_{i_1})\cos(\theta_{i_2})\cdots\cos(\theta_{i_k})$. If we set the process of raising the pumping from below that guarantees that $\sum_{l=1}^N \gamma_l$ is the smallest possible injected intensity, then at the fixed point the system finds the global minimum of the $k-$local Hamiltonian $H = -\sum_\Omega \mathbf{A}^k_{i_1,i_2,\cdots,i_k} \cos(\theta_{i_1})\cos(\theta_{i_2})\cdot\cdot\cdot\cos(\theta_{i_k})$, and, therefore, solves Eq. (3.12). We will refer to the Eqs. (3.25-3.26) as the tensor gain-dissipative (TGD) method.

To illustrate the behavior of the system we first consider a toy problem: the following 3-local Hamiltonian

$$H_{\rm test}(\mathbf{x}) = -8x_1x_2x_3 - 4x_1x_2x_4 - 2x_2x_3x_4 - x_1x_3x_4, \tag{3.27}$$

with variables $x_i \in \{\pm 1\}$, while Eq. (3.25) becomes $\dot{\Psi}_l = \Psi_l(\gamma_l - |\Psi_l|^2) + \sum_{\langle j,k\rangle} K_{ljk}\Psi_j\Psi^*_k$, and K is a tensor with nonzero entries $1, 2, 4, 8$. The Hamiltonian $H_{\rm test}$ has 2^4 stationary points, among which there are three local minima: $H_1 = -9, H_2 = -11, H_3 = -13$ and the global minimum $H_4 = -15$, that all can be accessed during the time evolution of the system. To understand the basins of attraction for these stationary points we numerically integrate Eqs. (3.25-3.26) starting with initial conditions $\Psi_i(t=0) = \frac{1}{100}\exp[\mathrm{i}\theta_i(t=0)]$ where the phases $\theta_i(t=0) \in [0, 2\pi)$ are uniformly distributed in the 4-dimensional space. the corresponding details are presented at the end of this section. Figure 3-2(a) depicts the statistics of distribution of the stationary points reached and indicates that the basins of local minima combined are larger than that of the global minimum. To facilitate the search for the global minimum the algorithm needs to allow for a possibility to explore the hyperspace until the lowest lying energy state is found. This can be achieved by adding a noise (typically present in physical system as well), that shifts the trajectory from its deterministic path while allowing it to stay below any local minima. This can be ascertained by decreasing ϵ that controls the time the system spends while raising to the condensation threshold from below. We illustrate this behaviour in Fig. 3-2(b) that depicts the statistics of reaching local and global minima found by numerical integration of Eqs. (3.25-3.26) using the same initial conditions as in Fig. 3-2(a) but with the white noise added. Further decreasing the ϵ parameter allows to improve the possibility of reaching the global minima. On Fig. 3-2(d) we show one of such trajectories as it approaches the global minimum of $H_{\rm test}$ from below.

With the growth in the number of variables and concomitant growth of the system hyperspace any local noisy perturbation of the trajectory may not be sufficient to reach the global minimum basin of attraction or it would take prohibitively long time. Recent interest in heteroclinic networks - networks that exhibit saddle states that are dynamically

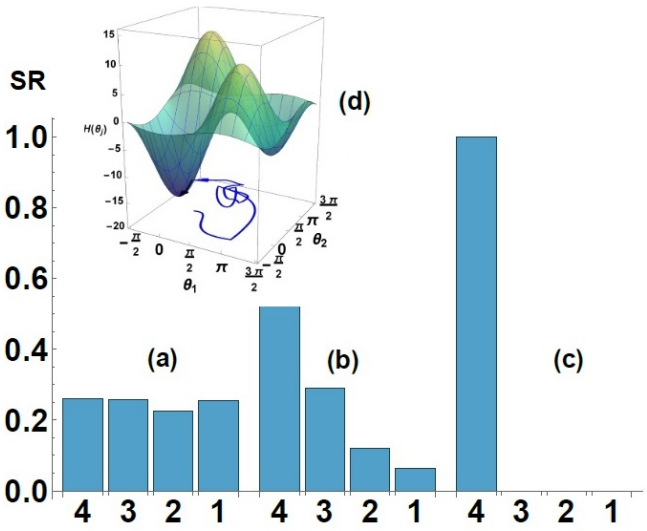


Figure 3-2: Success rates for achieving local and global minima of H_{test} given by Eq. (3.27) using numerical simulations of Eqs. (3.25-3.26) for three different controls described in the main text: fully deterministic integration without noise (a); with white noise added to the right hand side of Eq. (3.25) (b); using CC control as described in the main text (c). The indexes marked on the horizontal axis label local minima as $1, 2, 3$ that correspond to H_1, H_2 and H_3 respectively. The global minimum (H_4) is labeled 4. The insert (d) shows the two-dimensional projection ($\theta_3 = \theta_4 = 0$) of the energy landscape for H_{test} with $x_i = \cos\theta_i$. The scaled (by a factor of 3, due to the ternary terms) and shifted total injected intensity $(\sum \gamma_i - N\rho_{\text{th}})/3$ found by numerical simulation of Eqs. (3.25-3.26) and shown by the blue trajectory in the same 2D projection. The trajectory raises monotonically below the global minimum until some of the occupations exceed the threshold.

linked via heteroclinic connections – proposes a way to allow for a fast switching between the states [203]. Motivated by these ideas we introduced heteroclinic orbits into our model by engineering time-dependent complex-coupling (CC) into the network Eqs. (3.25-3.26). CCs naturally appear in polariton model if the energy shift due to a noncondensed reservoir $R(\mathbf{r}, t)$ is present in the system. This introduces $gR(\mathbf{r}, t)\psi(\mathbf{r}, t)$ term into the right-hand side of Eq. (3.13), where g parametrizes the strengths of the interactions between condensed and non-condensed particles. In the tight binding approximation and in analogy with the derivation of rate equations simulating the Ising and XY Hamiltonians [204], this term changes $A^k_{i_1,i_2,\cdots,i_k}$ to the complex coupling $A^k_{i_1,i_2,\cdots,i_k} + iB^k_{i_1,i_2,\cdots,i_k}$, where $\mathbf{B}$ represents the part of the coupling that comes from the energy shift due to reservoir. For instance, in polariton lattices such coupling can be turned on and off experimentally between the individual lattice elements and with strengths varying in time and space [205]. The presence of the complex part of the coupling introduces the phase lag in the system that leads to either shift of the stationary point of Eqs. (3.25-3.26) for small values of the complex part or destabilization of it by creating a saddle point [204]. In the latter case, if the complex part of the coupling is turned on, the system trajectory quickly leaves the neighborhood of the previous stationary point along the fastest direction. Including this switching dynamics into the system facilitates the search for the true global minimum by allowing fuller exploration of the phase space.

Complex coupling switching. To implement the complex coupling switching on Hamiltonian given by Eq. (3.27) we turn two of the real coupling coefficient into the complex ones with a significant complex part as soon as the system reaches a steady state. The system trajectory leaves the basin of attraction of that state and travels to a different part of the system hypercube $[0, 2\pi]^N$. When the complex part of the coupling is turned off another steady state may be found. By varying the coupling elements to be switched, the duration of the switching in time and the amplitude of the imaginary coupling we allow the system to efficiently search for global minimum. In our test example, implementing the switching of a coupling coefficients K_{123} and K_{124} according to $K_{123}(t) = 8(1 + 4i), K_{124}(t) = 4(1 - 10i), t \in [t_1, t_1 + 160] \cup [t_2, t_2 + 160] \cup [t_3, t_3 + 280]$ and keeping $K_{123}(t) = 8, K_{124}(t) = 4$ otherwise allows *every* trajectory irrespective of its initial state to arrive to the global minimum. Here t_1, t_2, t_3 are times at which the system settles to a steady state after switching the complex part of the couplings off. Figure 3-2(c) shows convergence of all trajectories to the global minimum with 100% probability.

Complex coupling switching for large N.

We adapt the idea of the complex couplings switching for the large scale simulations and illustrate the benefits of such approach on 20 dense and 20 sparse tensor sets K_{ijk} of 3^d rank of size 10^6 over 500 runs. The elements of the dense tensors are uniformly distributed in $[-1, 1]$. To generate sparse tensors we take the dense tensors and randomly set 9/10th of all elements to zero. To implement the complex coupling switching (TGD+CC) method on large N, as soon as the system reaches the steady state we randomly choose $N/50$ of the coupling strengths K_{ijk} (and their corresponding elements with all possible permutations of the indexes) and modify them by adding $\chi_{ijk} = 3iK_{ijk}$. This destabilises the system and forces the trajectory to leave along a certain orbit. After that we let $\chi_{ijk} = 0$ and allow the original system to relax to a new steady state. Keeping the total injected rate $\sum \gamma_l$ small forces the trajectories to explore the low

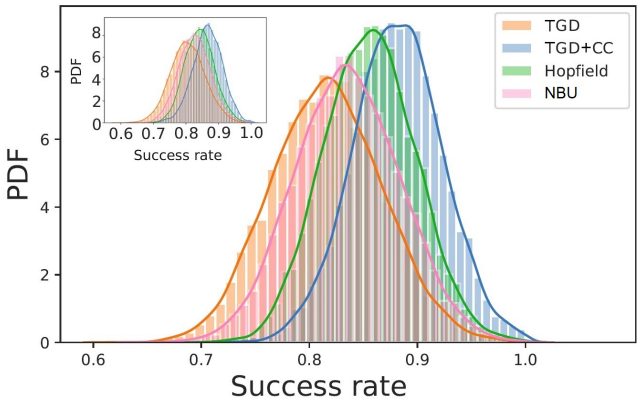


Figure 3-3: The probability density function of the final solutions in comparison with the best one found (referred as the global minimum) among 500 runs on the set of 20 dense tensors K_{ijk} in the main figure and 20 sparse tensors in the inset of the 3^d rank tensors for $N = 100$. The algorithms are TGD, NBU, Hopfield NN, and TGD+CC.

energy states of the Hamiltonian until the true global minimum is found. Further details of the numerical integration of Eqs. (3.25-3.26) are presented in end of this section.

We compared the behaviour of the TGD, TGD+CC with two network-based methods and show that TGD+CC outperforms all of these methods. The first network-based method represents the networks of bistable units (NBU) in the presence of a double-well potential derivative that forces the network elements x_l to take on ± 1 while solving Eq. (3.12):

$$\frac{dx_l}{dt} = -hx_l|x_l|^{k-1}(x_l^2-1) + \sum_{\bar{\Omega}} \mathbf{A}^k_{i_1,i_2,\cdots,l,\cdots,i_k} x_{i_1} x_{i_2} \cdots x_{i_k}, \tag{3.28}$$

where $x_l(t=0)$ are randomly distributed real numbers, and h is a control parameter. In comparison with the usual $k=2$ case [174], we balanced the degrees of polynomial between two term on the right-hand side of Eq. (3.28) by introducing $|x_l|^{k-1}$ factor.

Another efficient solver of Eq. (3.12) is given by a higher order Hopfield NNs [206]:

$$\frac{dx_l}{dt} = -x_l + \sum_{\bar{\Omega}} \mathbf{A}^k_{l,i_1,i_2,\cdots,i_k} s_{i_1} s_{i_2} \cdots s_{i_k}, \tag{3.29}$$

$$s_l = \tanh(x_l(t)/\beta), \tag{3.30}$$

where x_l are real continuous variables and β is the scaling parameter. The details of the numerical integration of Eq. (3.28) and Eqs. (3.29-3.30) are presented below respectively.

Figure 3-3 shows the advantage of TGD+CC in comparison with other methods for large scale tensor optimisation on both dense and sparse tensors. TGD+CC consistently has a better success probability of finding the global minimum.

Corresponding technical details:

Eqs. (3.25-3.26) were Euler integrated with $\epsilon = 0.001M$, $\gamma_l(t=0) = -0.2M$, $\rho_{\rm th} = 1$, with $D = \max(1-\rho/\rho_{\rm th}, 0)$, $dt = 0.01$, $T_{\rm stop} = 300$, $M = \max_{1\le l\le N} \sum_{\bar{\Omega}} |A^k_{l,i_1,i_2,\cdots,i_k}|$. Tensor elements were scaled by 0.05 to slow the dynamics down.

Eqs. (3.25-3.26) were Euler integrated with $\epsilon = 0.01M$, $\gamma_l(t=0) = -0.2M$, $\rho_{\rm th} = 0.1M$, with $D = 100\max(1-\rho/\rho_{\rm th}, 0)$, $dt = 0.001$, $M = \max_{1\le l\le N} \sum_{\bar{\Omega}} |A^k_{l,i_1,i_2,\cdots,i_k}|$. The elements of dense (sparse) tensors were multiplied by $5/6N$ $(15/6N)$.

Eq. (3.28) was Euler integrated with $dt = 0.001$ with the saem number of iterations as in other methods. The initial conditions $x_l(t=0)$ were uniformly randomly distributed in $[-0.5, 0.5]$, and $h_{j+1} = 1.2h_j$ is updated each time $\sum_l |\frac{dx_l}{dt}| < 0.001$ is satisfied.

Eqs. (3.29-3.30) were Euler integrated with $dt = 0.02$. The initial conditions $x_l(t=0)$ are randomly distributed in $[-5, 5]$, $\beta(t=0) = 10$, while $\beta_{j+1} = 0.8\beta_j$ is updated each time $\sum_l |\frac{dx_l}{dt}| < 0.001$ is satisfied.

3.3 Alternative application methods

When speaking about the EP system in the context of the applications, especially the information-processing applications, one has to understand that this condensed-matter system has many degrees of freedom, which can be experimentally manipulated. It allows one to utilize the experimental setup in many other ways than the presented methodology of pairwise/multiwise condensates interaction. Choosing other degrees of freedom mostly determines the operational setup. For the demonstrative purpose, it is helpful to present alternative approaches, which mainly deal with EPs as the macroscopic two-level system.

The first approach is based on the similarity of superfluid polariton flows [198] and superconducting electric currents. It aims at building a polariton analogue of the superconducting flux qubit [207]. The idea is to introduce a π-phase delay in the condensate ring, which forces the quantum fluid to propagate a clockwise or anticlockwise circular current to reduce the total gained phase to 0 or 2π value. This π-delay line can be provided by a dark soliton pinned to a potential well created by a C-shape nonresonant pump spot. The resulting condensate exhibit coherent oscillations, which give the artificial qubits their desired properties. The combination of different resonant/nonresonant pumping as well as their parameters allows one to manipulate the condensate given quantum state, its phase and other properties [207]. In particular, it is possible to achieve the following effective Hamiltonian in the following truncated two-level basis:

$$\hat{H} = \frac{\Delta_0}{2}\sigma_z + \frac{\delta\omega}{2}\sigma_x, \tag{3.31}$$

where Δ_0 is the energy of the basis states splitting, and $\delta\omega$ is the parameter corresponding to the gauge field-induced splitting between the circular polariton flows with opposite momenta. Based on the EP system, one can obtain familiar for the quantum information science operations, such as Pauli X and Y gates (denoted as $\mathcal{R}_x$ and $\mathcal{R}_y$, which require the electric pulses, generating a time-periodic radial electric field polarizing excitons in the cavity plane), the Z-gate $\mathcal{R}_z(\theta)$ (achieved by the appropriate time delay $\hbar\theta/\Delta_0$ between two successive single-qubit operations), the so-called iSWAP operation (permuting the states of the two qubits with the addition of $\pi/2$ phase difference), NOT-gate (simple qubit flip), CNOT-gate and Z-CNOT-gate. Moreover, using such operations, one can repeat Deutsch's algorithm [208].

Another similar complete theoretical scheme of quantum computing was introduced in [209]. It describes how to realize qubits externally controllable by applied laser pulses based on the quantum fluctuations in semiconductor micropillars. One can utilize quantum tunnelling and nonlinear interactions between the condensates to realize the quantum information domain's basic operations.

3.4 Machine learning approaches

The classical XY model, sometimes also called classical rotator or simply O(2) model, was a subject of extensive theoretical and numerical research and appears to be one of the basic blocks in the rich hierarchy of the condensed matter systems [210, 211, 212, 213, 214]. The latest advances in metrology and experimental techniques across different domains of the condensed matter field [169, 215, 216, 217] allow fabricating different devices, whose operations are based on the underlying mechanisms of the physical systems. Several activities are aiming to recreate different models on chips for both fundamental and applied purposes. It was shown before that EPs appear to be the system that gives rich opportunities for many applications [218, 219, 68, 220, 221, 222, 223, 224, 225] and in particular the reconstruction of the XY model [68, 188]. The problem arises in attempts to even start using or manipulating these systems for simple mathematical operations, not to mention more complex procedures.

We choose the classical XY model for the demonstrative purpose of a simple system, which behaves in a nonlinear way in the process of equilibration and can be used as an excellent example for possible extensions. With the motivation to translate the deep learning (DL) architecture into the classical condensed matter platform, we show how to build very complex structures based on the XY model's nonlinear blocks. The corresponding kind of Hamiltonian is quite general because it can be found in nature as magnetic materials or can be reproduced with other condensed matter systems with analog simulations [226, 227, 68].

The key points of this section are:

- I show how to realize the basic numerical operations on the small XY models, which variables tend to minimize the current Hamiltonian clusters' energy.
- The set of the obtained operations is enough to realize different combinations of mathematical operations and translate the deep NN architectures to the class of XY models possible.
- Beside the mathematical interest, the discussion of the current approach's practical significance is presented together with the engineering details and possible adjustments to the existing special-purpose hardware aimed at

reproducing the XY models.

- The general methodology of working with the simple but nonlinear systems, which can be extended to other different Hamiltonians, i.e. with different interactions or models with additional degrees of freedom (quantum effects), is shown, underlying the large scope of this work.

3.4.1 Basic XY equilibrium blocks

This part is devoted to describing the basic blocks of the XY NN with a particular emphasis on the deep learning (DL) architecture. DL is usually defined as a part of the ML methods based on the artificial NN with representation learning and proven to be effective in many scientific domains [50, 51, 52, 53]. It appears to be applicable in many areas ranging from applied fields such as chemistry and material science to fundamental ones like particle physics and cosmology [228].

The DL is typically referred to as a black box [55, 56]. When it comes to DL's application, one usually asks questions about adapting the DL architectures to the new problems, how to interpret the results and how to quantify the outcome errors reliably. Leaving these open problems behind, the focus is on a more applied task. To build DL architectures, one wants to transfer the pretrained parameters into a realization of a nonlinear computation. One of the mathematical core ideas in ML architectures is the ability to build hyperplanes on each neuron output, which taken together allows one to approximate the input data efficiently and adjust it to the output in the case of supervised learning (for example, in the classification tasks [229]). This procedure can be paraphrased as the feature engineering that before the modern DL approaches and the available computational resources was performed in a manual way [230].

Building hardware that performs the hyperplane transformation with a specific type of a nonlinear activation function with a given precision allows one to separate input data points and present building block operations for more complex tasks. Studying the hierarchical structures with such blocks leads to constructing more complex architectures capable of performing more sophisticated tasks.

Decomposing the nonlinear expressions common in the ML such as $\tanh(w_0x_0 + w_1x_1 + \ldots + w_nx_n + b)$, produces a set of mathematical operations, which one needs to approximate with our system. These are nonlinear operation tanh, which is conventionally called an activation function, the multiplication of the input variables x_i by the constant (after training) coefficients w_i (also called weights of a NN) and the summation operation (with the additional constant b, called bias).

The activation function is an essential aspect of the deep NNs, bringing nonlinearity into the learning process. The nonlinearity allows modern NNs to create complex mappings between the inputs and outputs, vital for learning and approximating complex data with high dimensionality. Moreover, the nonlinear activation functions afford backpropagation due to the smooth derivative of those functions. They normalize each neuron's output, allowing one to stack multiple layers of neurons to create a deep NN. The functional form of the nonlinear function is zero centred with the saturation effect, which leads to a certain response for the inputs that take strongly negative, neutral, or strongly positive values, not to mention the information-entropic foundation of its derivative form and other interesting properties [231, 232]. One will see that approximating this operation is straightforward with the XY networks.

Firstly, it is possible to introduce the list of simple blocks corresponding to the set of operations necessary to realize the nonlinear activation function, which can be obtained by manipulating the small clusters of spins with underlying U(1) symmetry. These clusters minimize the XY Hamiltonian:

$$H = \sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}\cos(\theta_i - \theta_j), \tag{3.32}$$

where i and j goes over N elements in the system, J_{ij} is the interaction strength between i^{th} and j^{th} spins represented by the classical phases $\theta_i \in [-\pi, \pi]$. If one takes several spins as inputs θ_i in such a system and consider the others as outputs θ_k, then it is possible to treat the whole system as a nonlinear function which returns $\arg\min_{\theta_k} H(\{\theta_i\}, \{\theta_k\})$ values due to the system equilibration into the steady-state. In some cases the ground state is unique, other cases can produce multiple equilibrium states.

It is useful to consider the analytical solution to such kind of task, describing the function with one output and several input variables. One considers the system with N spins: $\theta_i, i = 1, .., N-1$ are input spins and θ_N is the output spin coupled with the input spins by the strength coefficients $J_i \equiv J_{iN}$. The system Hamiltonian can be written as $H = \sum_i^{N-1} J_i \cos(\theta_i - \theta_N)$. By expanding H as $\sum_i^{N-1} J_i \cos(\theta_i - \theta_N) = \sum_i^{N-1} J_i \cos\theta_i \cos\theta_N + \sum_i^{N-1} J_i \sin\theta_i \sin\theta_N$ one can solve for the minimizer θ_N:

$$\theta_N \equiv F((\theta_1|J_1), (\theta_2|J_2), ..., (\theta_{N-1}|J_{N-1})) = -\operatorname{sign} B\left(\frac{\pi}{2} + \arcsin\frac{A}{\sqrt{A^2+B^2}}\right), \tag{3.33}$$

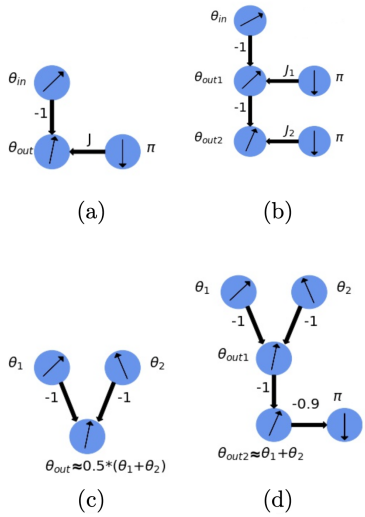


Figure 3-4: Several examples of basic blocks and their combinations used in the XY NN architectures. (a) The block performing the function $F((\theta_{in}|-1),(\pi|J))$ with one input spin and one reference/control spin with imposed π value, all coupled with the output by the ferromagnetic -1 and J ($J=1$ is used on the picture). Depending on J one can realise the operations that approximate the multiplication by the constant k so that $\theta_{out} \approx 1.5 \tanh 4\theta_{in} + 0.5\theta_{in}$ with $J=-0.9$. (b) Two blocks representing $F(F((\theta_1|-1),(\pi|J_1))|-1),(\pi|J_2))$. (c) The block $F((\theta_1|-1),(\theta_2|-1))$ for the the half sum of two variables θ_1 and θ_2. (d) Two blocks performing the function $F(F((\theta_1|-1),(\theta_2|-1))|-1),(\pi|-0.9))$. Some of the response functions for these blocks are presented in Fig.3-5.

where $A=\sum_i^{N-1} J_i \cos\theta_i$ and $B=\sum_i^{N-1} J_i \sin\theta_i$. Alternatively, this formula can be rewritten through the complex analog of the order parameter $C=\sum_i^{N-1} J_i e^{i\theta_i} = A+iB$ in the following way: $\theta_N = F(\{\theta_i|J_i\}) = -\operatorname{sign}(\operatorname{Im} C)\,(\pi - \operatorname{Arg} C)$. It is possible to present several basic blocks in Fig. 3-4 and the outcomes of the functions responses in Fig. 3-5.

The notation introduced in Eq. (3.33) will be used to describe both the activation function and the graph cluster of spins below. The recurrent notation where the output of the first block serves as the input to the next one will be used, for example $F(F((\theta_1|J_1),(\theta_2|J_2))|J_3),(\theta_4|J_4))$.

To describe the iterative implementation of many (k) identical blocks, where the input is defined in terms of the output of the previous same block, one can rewrite the recurrent formula for one argument as: $(F_1 \circ F_1 \circ ... \circ F_1)(\theta_{in}) \equiv F_1(F_1(...F_1(\theta_{in}))) \equiv F_1^k(\theta_{in})$, where $F_1(\theta_{in}) = F((\theta_{in}|J_1),(\theta_2|J_2))$ is a certain block with the predefined parameters. All possible blocks of spins are separated into several groups and consider them below in more detail.

The phase θ_i are in $[-\pi,\pi]$, however, for an efficient approximation of the operations (summation, multiplication and nonlinearity) one needs to limit the domain to $[-\pi/2,\pi/2]$ which we refer to as the *working domain*. Additionally, we need to guarantee that the values of the *working spins* (which are not fixed and are influenced by the system input, thus serving as analogue variables) are located within the limits of the working domain. This will be implemented below. Next, I consider the implementation of the elementary operations.

- *Multiplication by the constant value* $k>0$. Connecting the input spin with the output spin by the "ferromagnetic" coupling $J=-1$ will lead to the input spin's replication. In this way, one can transmit the spin value from one block to another. Changing the value of the output spin can be achieved in many ways. The addition of another spin with a different value and coupling it to the output spin with a constant coupling J is one such possibility (for example, with imposed π value, which one will refer to as a *reference/control spin*). If J is in $[0,1)$ (relative to -1 coupling between θ_{in} and θ_{out}), then the reference spin influences the output spin value with the effective "repulsion" and thus, depending on the relative coupling strength, decreases the output spin value (see Fig. 3-5(a,b) and the corresponding cluster configuration on Fig. 3-4(a)). The resulting relation between the input and output spin values can be a good approximation to the multiplication by certain values lying in the $[0,1]$ range. The block corresponding to the implementation of $F((\theta_{in}|-1),(\pi|1))$ has a peculiarity in case of $\theta_{in}=0$, which allows the output to take any value due to the degeneracy of the ground state. To overcome this degeneracy, it is possible choose $J=0.99$.

For $k>1$ one can use a ferromagnetic coupling $J<0$ (see Fig. 3-5(c) and the corresponding cluster configuration on Fig. 3-4(a)). However, the positive values of J are more reliable for the implementation since the output functions have small approximation errors (see *Supplementary information* for the exact values of this error and further clarification). One can replace the multiplication by a large factor by the multiplications by several smaller factors to reduce the final accumulated error. It is possible to guarantee the uniqueness of the output since the clusters are small, and the output

is defined by Eq. (3.33), which gives the unique solution.

- *Nonlinear activation function.*

The function $F((\theta_{in}|-1),(\pi|-0.9))$ is similar to the hyperbolic tanh function (see Fig. 3-5(c) and the *Supplementary information* for the exact difference). There are two ways of using such a transformation as an activation function.

1) One can use the similarity between the values of the $F((\theta_{in}|-1),(\pi|-0.9))$ and the function $1.5\tanh(4\theta_{in})+0.5\theta_{in}$ (see Fig. 3-5(f)). In general, many nonlinear functions can be used in NNs. Usually, a minor modification in the functional form of the nonlinear activation function does not change the network's overall functionality, with the additional training procedures of NNs in some architectures. It is possible to train the NN initially with the $1.5\tanh(4\theta_{in})+0.5\theta_{in}$ function so that in the final transfer, it will not be necessary to adjust the spin system to approximate the given function.

2) One can use the similarity with the approximate hyperbolic tangent function within the XY spin cluster. In other words, to execute $\tanh(\theta_{in})$, one has to perform $(F((0.25\theta_{in}|-1),(\pi|-0.9))-0.5\theta_{in})/1.5$ function using the spin block operations. This option will be used below.

- *Multiplication by the constant value* $k=-1$. The main difficulty of this operation is in finding the set of parameters for the spin block where $\frac{\partial F}{\partial\theta_{in}}<0$ holds. $F((\theta_{in}|1),(\pi|J))$ is a good example of such a block. To perform the multiplication by $k=-1$, we need to embed the whole working domain into the region where the presented inequality is valid and return these values with the multiplication by $k>0$ factor. One final realization can be represented as $F_3^3(F_2(F_1^2(\theta_{in})))$ function, where $F_1(\theta_{in})=F((\theta_{in}|-1),(\pi|0.9)), F_2(\theta_{in})=F((\theta_{in}|1),(\pi|J)), F_3(\theta_{in})=F((\theta_{in}|-1),(\pi|-0.2))$.

- *Summation.* The function $F((\theta_1|-1),(\theta_2|-1))$ gives a good approximation to the half sum $(\theta_1+\theta_2)/2$. This block is presented in Fig. 3-4(c) and the cross-sections of the surface defined by the function of two variables $F((\theta_1|-1),(\theta_2|-1))$ are plotted in Fig. 3-5(e). The plots show that the spin system realizes the half sum of two spin values with a minimum discrepancy compared to the target function on a working domain. One can multiply the final result by two using previously described multiplication to achieve an ordinary summation. In general, such a type of summation can be extended on multiple spins $N>2$, in a similar way of connecting them to the output spin, with the final value of $(\theta_1+..+\theta_N)/N$.

Summarizing, the method of approximating the set of mathematical operations, necessary for performing the $\tanh(w_0x_0+w_1x_1+\ldots+w_nx_n+b)$ function, using the XY blocks described by Eq. (3.33) was presented. The output spin value of each block is formed when a global equilibrium is reached in the physical system with the speed that depends on particular system and its parameters. The universality of the model was kept, which allows one to implement this approach using various XY systems. There were no assumptions about the nature of the classical spins, their couplings, or the manipulation techniques, however, the forward propagation of information requires directional couplings. As the blocks corresponding to elementary operations are added one after another, the new output spins and new reference spins should not change the values of the output spins from the previous block. The directional couplings that affect the output spins of the next block but not the output spins of the previous block satisfy this requirement. Many systems can achieve directional couplings. For instance, in optical systems the couplings are constructed by redirecting the light with either free- space optics or optical fibres to an SLM. At the SLM, the signal from each node is multiplexed and redirected to other nodes with the desired directional coupling strengths [233].

3.4.2 One-dimensional function approximations

This section illustrates the efficiency of the proposed approximation method on one-dimensional functions of intermediate complexity by considering two examples of mathematical functions and their decomposition into the basis of nonlinear operations.

For illustration, it is possible choose two nontrivial functions (one is monotonic and another is non-monotonic). The use of the methodology for more complex functions in higher dimensions is straightforward. In the next section, I will show the method efficiency for two-dimensional data-scientific toy problems.

I consider two functions

$$\mathcal{F}_1=0.125F_t(1.2x)+0.125F_t(0.5(x+1.4)), \tag{3.34}$$

and

$$\mathcal{F}_2=0.125F_t(0.5(x-1.2))-0.03125F_t(0.5(x+1.2)), \tag{3.35}$$

where $F_t(x)=1.5\tanh(4x)$. Note, that arbitrary functions can be obtained using a linear superposition of scaled and translated basic functions $F_t(x)$.

The comparison of the XY blocks approximations and the target functions are given in Fig. 3-6 and Fig. 3-7 demonstrating a a good agreement in the working domain. I also plot the explicit structures of the corresponding XY spin clusters showing a small overhead on the number of spins used per operation.

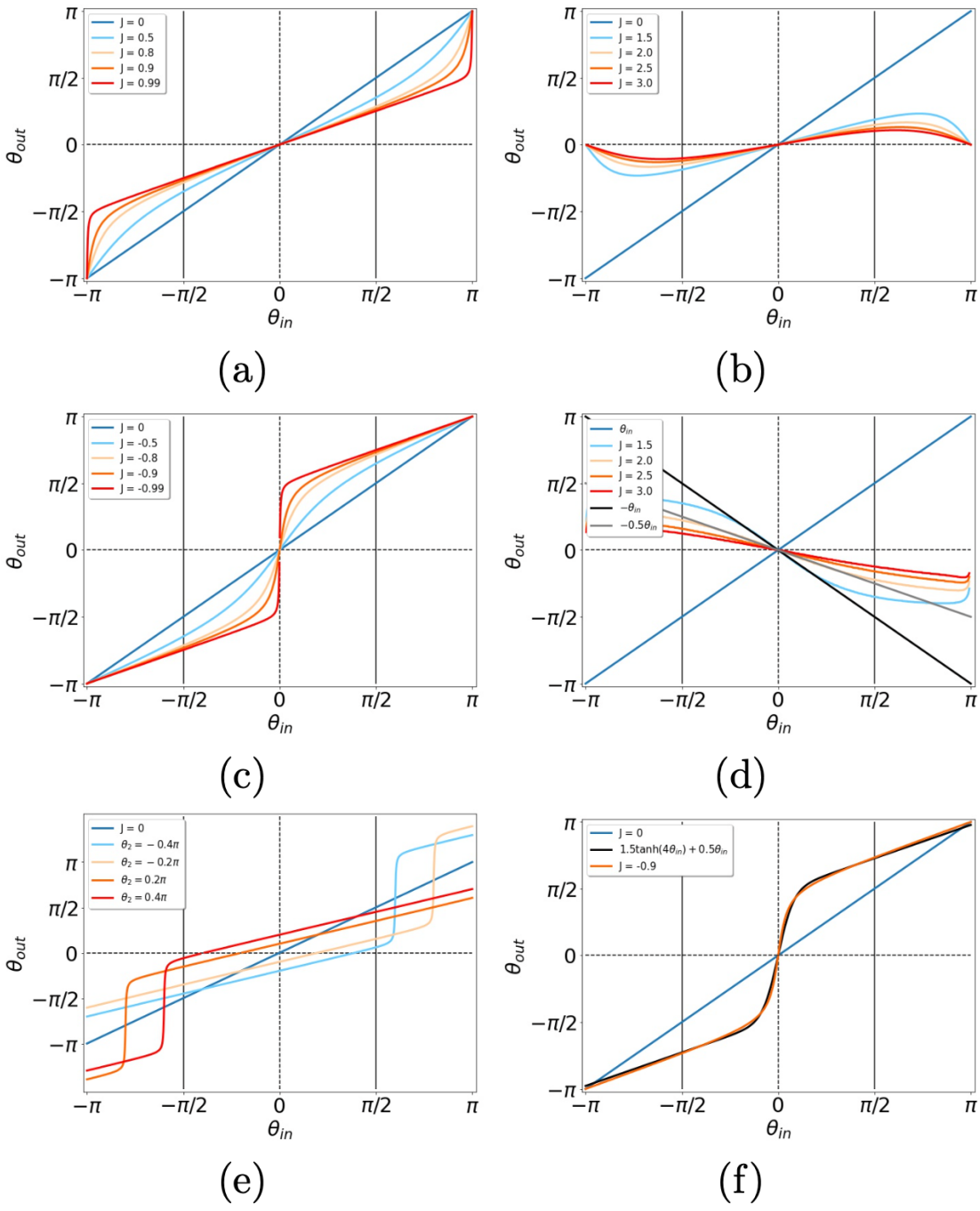


Figure 3-5: Several examples of input-output relations for the basic blocks and their combinations used in the XY NN architectures. (a) The parametrised family of $F((\theta_{in}|-1),(\pi|J))$ functions, that corresponds to the basic blocks. Depending on J coupling strength parameter one can realise the multiplication by arbitrary k. (b) The graphs of $F((\theta_{in}|-1),(\pi|J))$ functions for various values of J illustrating the the multiplication by small values of k. (c) The graphs of $F((\theta_{in}|-1),(\pi|J))$ functions for smaller values of $J<0$. (d) The graphs of $F_3^3(F_2(F_1^2(\theta_{in})|J))$ functions, where $F_1(\theta_{in}) = F((\theta_{in}|-1),(\pi|0.9)), F_2(\theta_{in}|J) = F((\theta_{in}|1),(\pi|J)), F_3(\theta_{in}) = F((\theta_{in}|-1),(\pi|-0.2))$, showing different negative outputs. (e) The graphs of $F((\theta_1|-1),(\theta_2|-1))$, implementing the block shown on Fig. 3-4(c), which approximates half sum of input variables. (d) The graphs of $1.5\tanh(4\theta_{in}) + 0.5\theta_{in}$ and $F((\theta_{in}|-1),(\pi|-0.9))$.

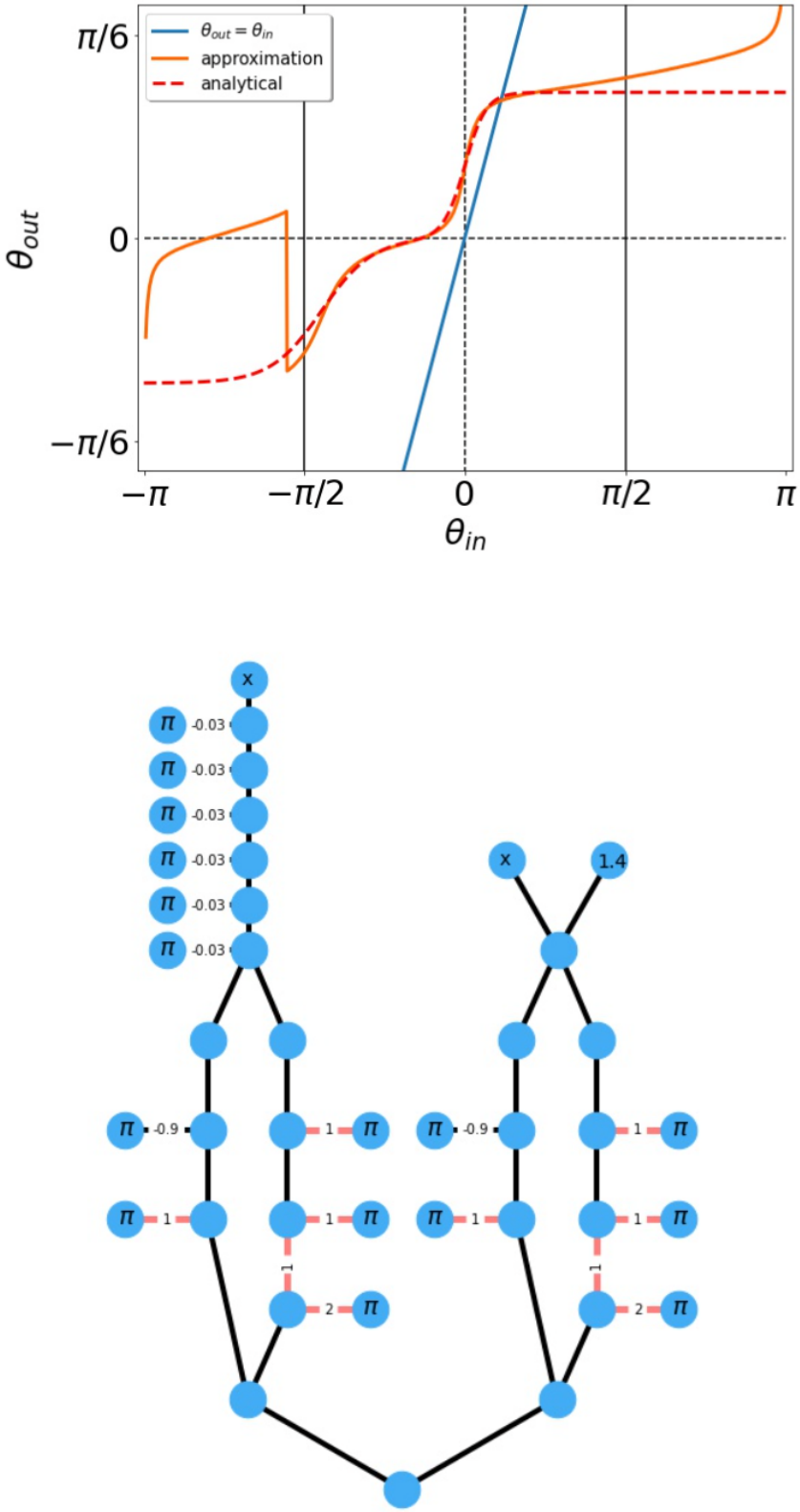


Figure 3-6: Top: The demonstration of the approximation quality, obtained by using the nonlinear XY spin clusters. The analytical monotonic function (red dashed line) is given by Eq.(3.34), the orange line is the approximation, the blue line is $\theta_{out} = \theta_{in}$. Bottom: The graph structure, representing the basic mathematical operation in the Eq.(3.34), given by the blocks, discussed in Section 3.4.1. The input variables are mapped into the top spins, after which the cluster is equilibrated before performing the next operation. The blue empty nodes are working spins that change according to the variables at higher block. The blue nodes with the fixed π-value are reference/control spins. The black edges without the notation have the fixed relative strength -1, while for others the coupling is written on the edges explicitely. The red colour of the edge represents the positive relative coupling strength J. The bottom spin gives the value of the coded function.

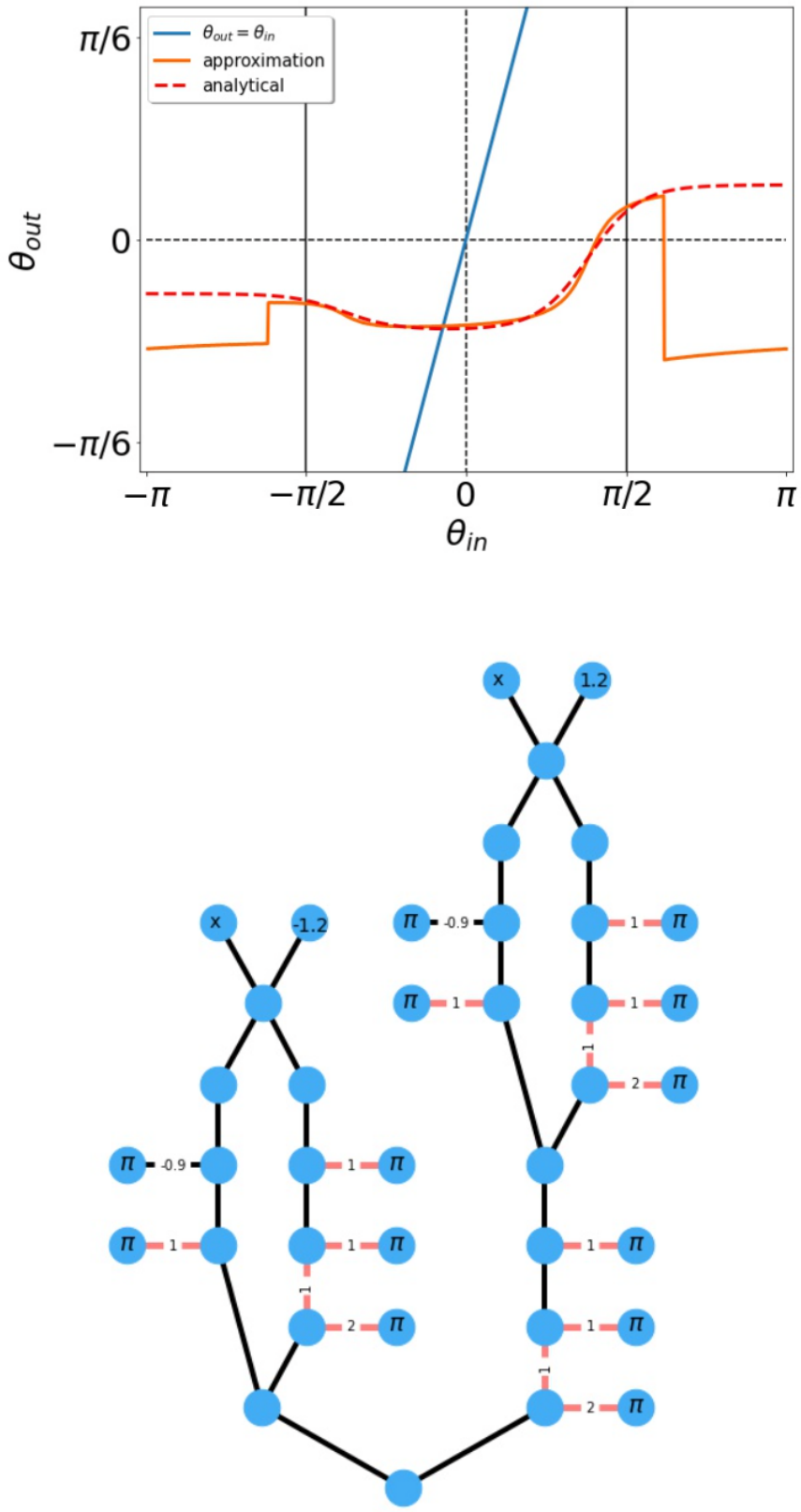


Figure 3-7: Top: The demonstration of the approximation quality, obtained by using the nonlinear XY spin clusters. The analytical non-monotonic function (red dashed line) is given by Eq.(3.35), the orange line is the approximation, the blue line is the linear identity relation. Bottom: The graph structure, representing the essential mathematical operation in Eq.(3.35), given by the blocks, discussed in Section 3.4.1. The notation used for describing the graph parameters is the same as in Fig. 3-6.

3.5 Neural networks benchmarks

In this section, the test of the XY NN architectures and check their effectiveness using typical benchmarks are presented. For simple architectures, the classification of predefined data points perfectly suits this goal. I consider standard two-dimensional datasets, which conventionally referred to as 'moons' and 'circles' and can be generated with *Scikit-learn* tools [234]. An additional useful property of such tasks is that they are easy for manual feature engineering.

First, one can train a simple NN. The parameters for the training setup are given in *Supplementary information* subsection. The final performance demonstrates perfect accuracy in both datasets. The architecture consists of two neurons' input layer, one hidden layer with three neurons for each feature and tanh activation function. The output layer consists of two neurons, which are transformed with Softmax function. The corresponding weights for both cases are given in *Supplementary information* subsection. Fig. 3-8 and Fig. 3-9 show the decision boundaries with the given pretrained architectures and the landscape of one of the final neuron visualizations together with the data points.

Since the focus is on transferring the described architectures into the XY spin cluster system, we consider the basic architecture adjustment on the example of one feature. Suppose we have the expression $\tanh(w_1x + w_2y + b)$. To repeat the chosen strategy 2) of approximating *the nonlinear activation function*, it is possible rewrite the coefficients w_1, w_2 and b as, say, $w_i \to (N/K)[(K/N)w_i]$, where K is a parameter chosen to increase the accuracy of each computation. One can approximate the square brackets' operations using the building blocks from the Section 3.4.1. The factor $N = 3$ will be cancelled by the value $1/N$ during the summation of N spins, while the factor $K = 4$ in the denominator will be taken into account during the operation of the function $F((\theta_{in}| - 1), (\pi| - 0.9))$, which is close to the $\tanh 4\theta_{in}$. The resulting procedure achieves good performance seen in Fig. 3-8, while the details are provided further in *Supplementary information* section.

The final Softmax function in the original NN serves as the comparison function of two features to achieve the final decision boundary's smooth landscape. One can omit this function and replace it with a simpler expression $x - y$. To achieve the binary decision boundaries, one can exploit the block performing $F((\theta_{in}| - 1), (\pi| - 0.9))$ several times to place the final spin value either close to $\pi/2$ or $-\pi/2$. In this way, it is possible to adjust the architecture that performs the same functions as the described simple NN on a toy model. The final decision boundary of the XY NN approximation can be seen in Fig. 3-8, which is very close to the boundary of the standard trained NN architecture.

The case of the "moons" dataset is a bit different. While the smooth functions are easy to approximate with the nonlinear XY blocks, it is quite complicated to reproduce "sharp" patterns with the high value of the function derivative. For this purpose, we adjust the NN coefficients to achieve good decision boundaries. The difference between the NN and its approximation and consequent results is shown in Fig. 3-9. Adjusted parameters of NN are given in the *Supplementary information* section. Figures 3-8 and 3-9 show the XY blocks of the spin architectures. The presented methodology allows us to upscale the XY blocks with even more complex ML tasks.

3.5.1 Transfering deep learning architecture

Deep NNs are surprisingly efficient at solving practical tasks [235, 51, 50]. The widely accepted opinion is that the key to this efficiency lies in their depth [236, 237, 238]. One can transfer the architecture's depth into our XY NN model without any significant loss of accuracy.

The transfer of the predefined architecture into the XY model was shown. This section discusses the transition of more complex deep architectures, which are considered conventional across different ML fields. To extend the method, it is possible choose two conventional image recognition task models (the architecture details are given in the *Supplementary information* section).

The focus of the architecture adjustment will be on operations, which were not previously discussed. The list of such operations are Conv2D, ReLU activation function, MaxPool2D and SoftMax [50, 239, 240]. The Conv2D is a simple convolution operation and does not present a significant difficulty, since it factorizes into the operations previously discussed. ReLU activation function can be replaced with its analog LeakyReLU. Using its similarity with the analytical expression $1.5(1 + \tanh 0.8(\theta_{in} - 1.5))$ allows one to obtain the following set of transformations through $z = F(F((\theta_{in}| - 1), (-1.5| - 1))| - 1), (\pi|1.5))$. Performing the summation of the three terms 1.5, $F((\theta_{in}| - 1), (\pi| - 0.9))$ and $-0.5\theta_{in}$ gives us a good approximation for LeakyReLU.

MaxPool2D relies on $\max(x, y)$ realisation. Therefore, it is convenient to use two similar architectures of spin value transmission, which are symmetrical with respect to variables x and y. The first one consists of $x - y$ operation, ReLU approximation (which is essentially $\max(0, x - y)$ function), and summation with the y variable, while the second architecture interchanges x and y and consists of $y - x$, ReLU and $+x$ operations. Summing the results of each architecture will give us the required value of $\max(x, y)$.

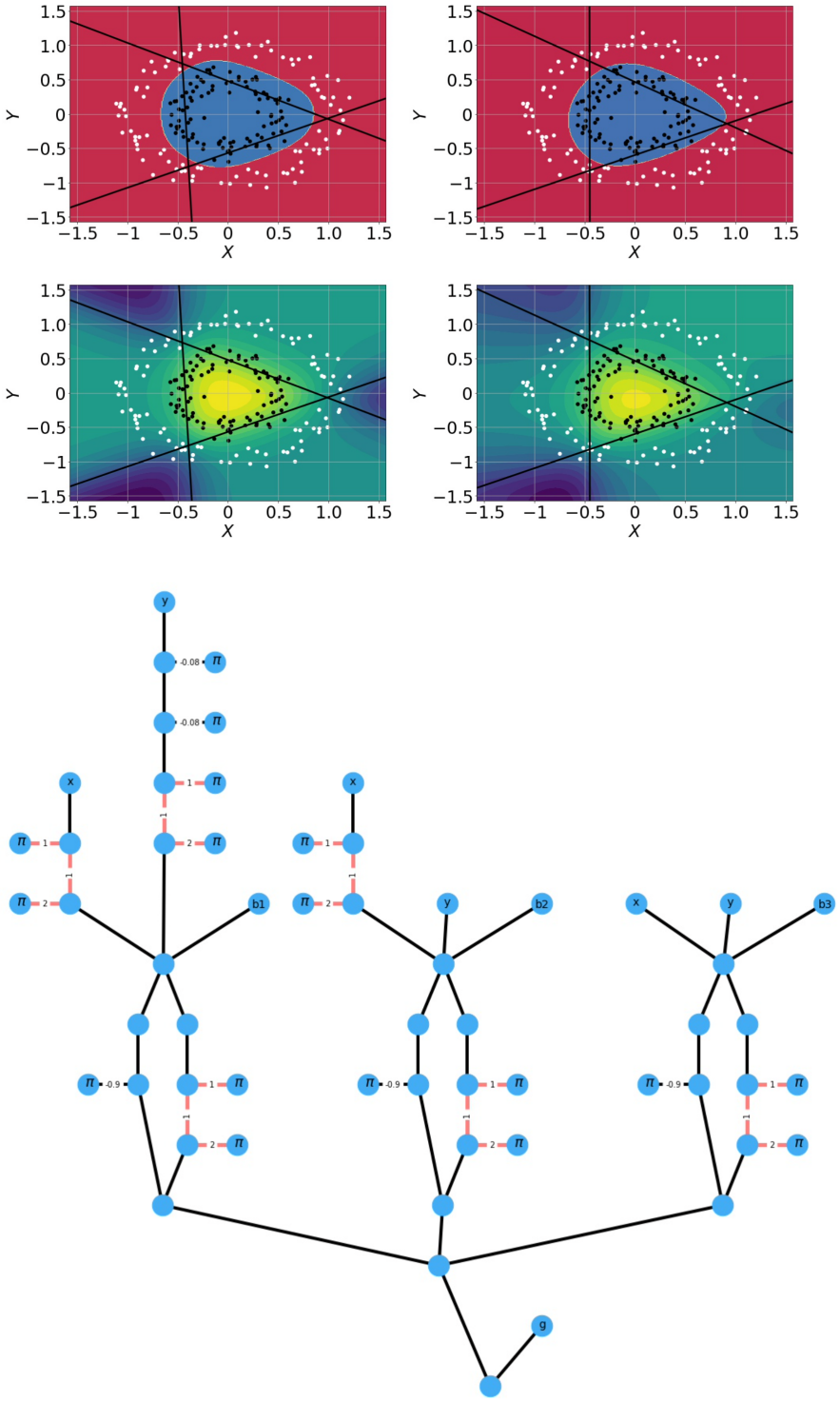


Figure 3-8: Top row: Decision boundaries of the simple (2,3,2) NN on the left and approximated ones for the XY NN on the right on toy 2-D circle data set. Black lines represent the bounds for automatically found features in the middle layer of classical NN, and the approximated features are shown on the right picture. Middle row: The isosurface for the one particular chosen feature for typical NN and its corresponding matched XY NN last variable isosurface. The parameters of both NN and XY NN architectures can be found in the *Supplementary information*. Bottom: The corresponding graph structure.

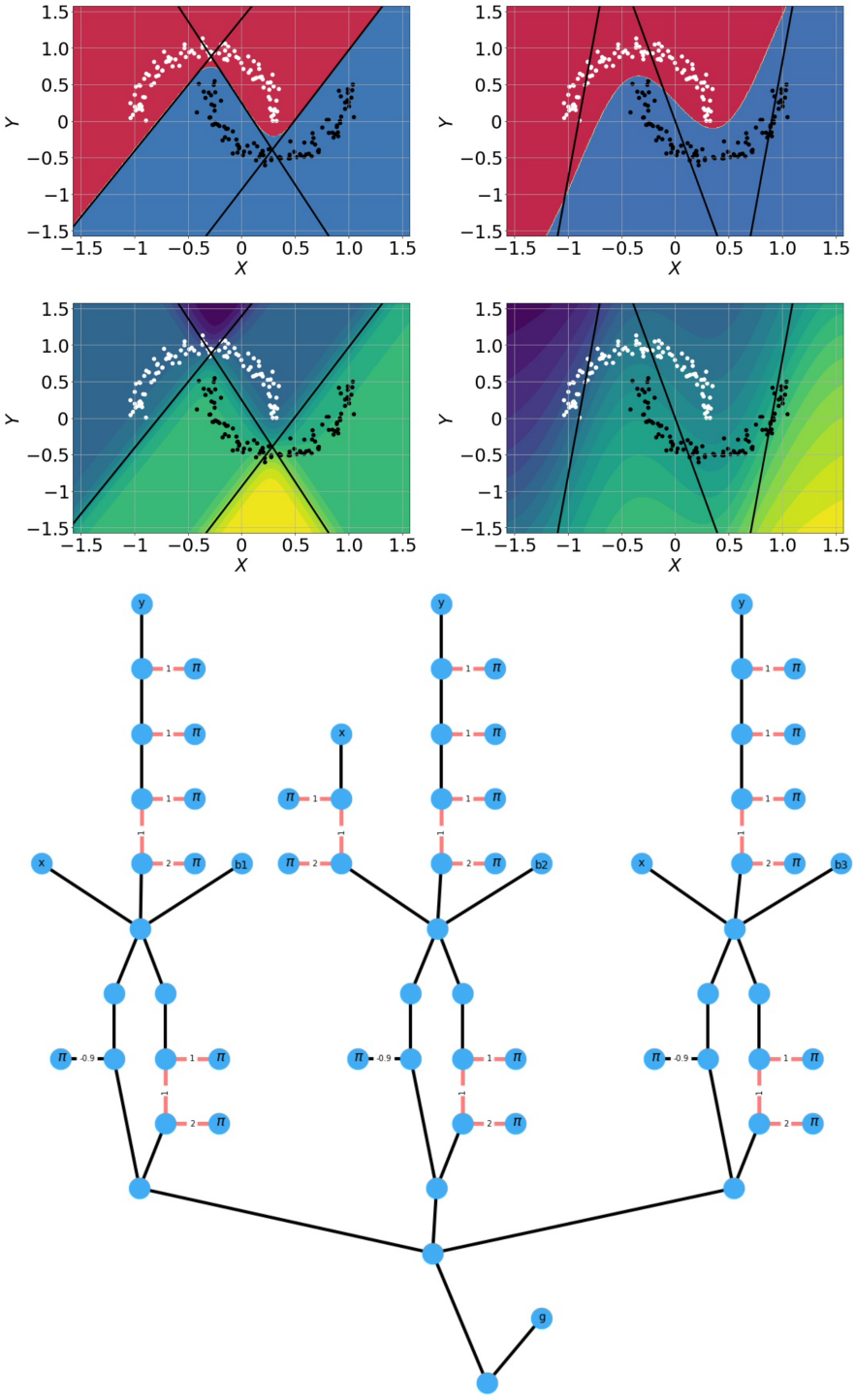


Figure 3-9: Decision boundaries of the simple (2,3,2) NN on the left and approximated ones for the XY NN on the right on toy 2-D moons data set. Black lines represent the bounds for automatically found features in the middle layer of classical NN, the approximated features adjusted for this specific task are located in the right picture. Middle row: The isosurface for the one particular chosen feature for typical NN and its corresponding matched XY NN last variable isosurface. The parameters of both NN and XY NN architectures can be found in the *Supplementary information*. One can state that the XY model can give a good approximation of the classical NN architectures with not ideally smooth representations, but requires small adjustments of their parameters, due to difficulties with representing sharp geometric figures (which in general can be multidimensional). Bottom: The corresponding graph structure.

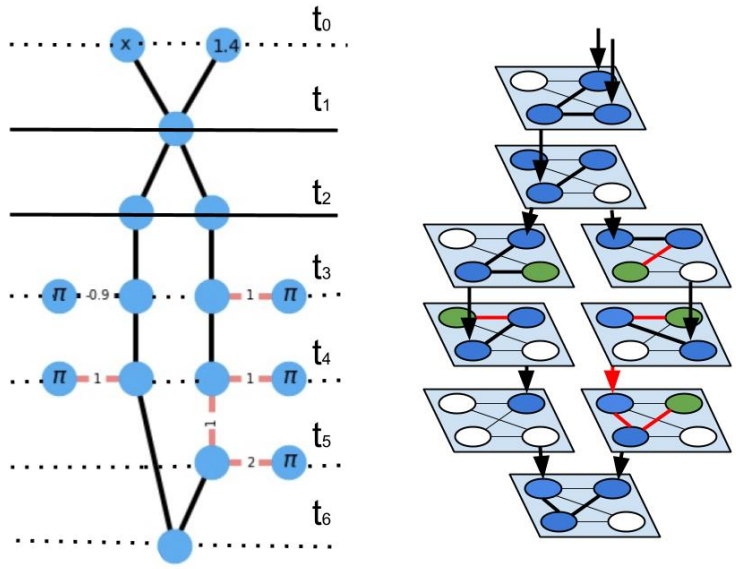


Figure 3-10: The picture's left side demonstrates the graph representation of the $F_t(0.5(x + 1.4))$ operation defined in the text. The final spin values are established through 6 time units, and the measure equals the local characteristic equilibration time of the XY spin cluster. The right side shows the alternative architecture, which performs each operation at the same cluster while saving the local output spin value and transferring it the next time through the self-locking mechanism. The colour correspondence is the same as in XY graphs: blue nodes are the working spins, and green are reference/control spins with π-values.

The SoftMax $e^{z_i}/\sum_{j=1}^{K} e^{z_j}$ with the input variables z_i can be approximated using several assumptions. Diminishing the exponential embeddings, the problem breaks down into approximating the $\frac{x}{x+y}$ function for $x > 0$ and $y > 0$. Approximating $\frac{1}{1+0.1y/x} \approx 0.9\tanh(4y/x)$, one can use the block $F((\theta_{in}| - 1), (\pi| - 0.9))$ to represent tanh function, while the scaling factor 0.1 can be controlled by an alternative architecture that depends on y.

3.5.2 Exciton-polariton setting

In this subsection, we discuss implementing the proposed technique using a system of EP condensates. As was previously shown, it is possible to reproduce XY Hamiltonian in the context of EPs [68, 188]. Each condensate is produced by the pumping sources' spatial modulation and can be treated with the reduced parameters for each unit representing the density and phase degrees of freedom, serving as the analogue variables for the minimization problem. The EP condensate is a gain-dissipative, non-equilibrium system due to the finite quasiparticle lifetimes. Polaritons decay emitting photons. Such emission carries all necessary information of the corresponding system state and can serve as the readout mechanism. Redirecting photons from one condensate to another using an SLM allows to couple the condensates in a lattice directionally [233].

The system of condensates maximizes the total occupation of condensates by arranging their relative phases so as to minimise the XY Hamiltonian [68]. The EP platform allows one to manipulate several parameters, such as coupling strengths between the condensates to tune them to the particular mathematical operation, or to fix the phase of the condensate (through the combination of resonant and non-resonant pumping, see [241]), and thus creating reference/control spins. Each input spin of the whole system can be controlled via two fixed couplings with the two reference/control spins of different values, see, for example, one of the blocks from Fig. 3-5. Fixing the coupling coefficients between other spatially located elements is required to perform the necessary operation and establish the XY network. It can be further upscaled to approximate a particular ML architecture, with the final output spin being the readout target.

One additional note is that the same system can be exploited differently by introducing the spin self-locking mechanism. It consists of saving the spin value in the system without coupling connections with the external elements. This mechanism can be achieved by coupling the local output with another element(s) with high negative coupling strength and consequent decoupling it from the previous units. The self-locking allows one to save the local output and use it for the consequent operations, without significant overhead on elements coming from the previously established operations. I

demonstrate the difference in Fig. 3-10. The initial universal methodology requires all elements, which the XY graph contains. The presented alternative, requiring a self-locking mechanism, operates with fewer spins by performing each action at the same cluster. Hence, it is volume efficient, which is noticeable in the scaling of elements per operation. Fig. 3-10 shows the same operation with 16 spins (without external nodes) and 8 total spins with the self-locking mechanism.

3.5.3 Conclusions and future directions

In this chapter, I introduced the robust and transparent approach for approximating the standard feedforward NN architectures utilizing the set of nonlinear functions arising from the classical XY spin Hamiltonian behaviour and discussed the possible extensions to other architectures. The number of additional spins required per operation scales linearly. The best-case scenario has two spin elements per multiplication and nonlinear operation (not taking into account the multiplication by a negative factor), making the general framework rather practical. Some operation approximations used in this work allow additional improvements to reduce the cumulative error between the initial architecture and its nonlinear approximation (see *Supplementary information*).

The entire spectrum of the benefits dramatically depends on a particular type of optical or condensed matter platform. The presented approach has universal applicability, and at the same time, has a certain degree of flexibility. It preserves basic blocks' simplicity, and overall structure works with the intermediate complexity architectures capable of solving toy model data scientific tasks. The upscaling to reproduce DL architectures was discussed.

Finally, the alternative of using a hybrid architecture was presented. Instead of transferring the operations used in the conventional NN model, one can introduce the nonlinear blocks coming from the presented XY model into the working functional given by a particular ML library. For example, one can change the activation function into the one that comes from the system operation and therefore is easily reproducible by that system. This would result in the architecture and the transfer process that do not require additional adjustments from the hardware perspective.

The question about the implementation of the backpropagation mechanism, i.e. computing of the gradient of the NN weights in a supervised manner, usually with the GD method, is still under consideration because of the limited scope of the work with the transfer of the predefined, pretrained architecture.

Supplementary information.

Here one can find the parameters of the NNs used in this section and details of the training and approximations for the XY graphs. Moreover, one can find two DL architectures mentioned in the main text emphasising their nonlinear operations.

For training simple classical $(2, 3, 2)$ feedforward NN architectures on "moons" and "circle" datasets the Pytorch library was used [80] with the Adam optimizer [242] with batch size 32 and learning rate 0.01 value. The expected learning procedure passed without the problems on datasets consisting of 200 points generated with the small noise of magnitude 0.1. The final performance gives perfect expected accuracy in both cases.

For the "circles" dataset, the NN parameters are presented in Table 3.1.

Table 3.1: NN (2,3,2) parameters used for the toy dataset 'circles'

(a)

w_{11}	w_{12}	b_1
-1.3465	-2.4191	1.1582
-3.5880	-0.1474	-1.5228
-1.3565	2.6776	1.5239

(b)

w_{21}	w_{22}	w_{23}	b_2
-17.8809	16.4797	-18.2794	23.6661
17.3684	-17.0227	18.4634	-23.3256

The first row of mathematical approximations gives us similar coefficients. One can rewrite in the same manner the NN parameters with minor adjustments for demonstrative purposes (see the main text for the detailed analysis of possible assumptions and the improvements such as the multiplication by the scaling coefficients).

The presented approximation architecture given in Table 3.2 was adjusted for better representation of one final feature, which is general enough to mark the decision boundaries for this particular task, while getting rid of the unnecessary parameters. Another approximation stage leads us to the final architecture, which is shown in Fig. 3-8. Let f_i denote the i-th feature and $R(x) = F((F_1(x)| - 1), (F_3(F_2(x))| - 1))$, where $F_1(x) = F((x| - 1), (\pi| - 0.9)), F_2(x) = F((x| - 1), (\pi|1)), F_3(x) = F((x|1), (\pi|2))$ represents the approximation of the activation function with the reduced accuracy, then:
$x_{11} = F((F((x_1| - 1), (\pi|0.99))|1), (\pi|2))$

Table 3.2: XY NN (2,3,1) parameters used to approximate the standard NN for the toy dataset 'circles'

(a)

w_{11}	w_{12}	b_1
−0.5	−0.75	0.35
1.0	0.0	0.45
−0.5	1.0	0.6

(b)

w_{21}	w_{22}	w_{23}	b_2
1.0	1.0	1.0	−0.31

$x_{12} = F((F((x_{11}| - 1), (\pi|0.99))|1), (\pi|2))$
$y_{11} = F((F((y_1| - 1), (\pi|0.99))|1), (\pi|2))$
$y_{12} = F((F((y_{11}| - 1), (\pi| - 0.08))| - 1), (\pi| - 0.08))$
$f_1 = F((x_{12}| - 1), (y_{12}| - 1), (b_1 = 0.35| - 1))$
$f_{11} = R(f_1);$
$x_{21} = F((F((x_2| - 1), (\pi|0.99))|1), (\pi|2))$
$f_2 = F((x_{21}| - 1), (y_2| - 1), (b_2 = 0.6| - 1))$
$f_{22} = R(f_2);$
$f_3 = F((x_3| - 1), (y_3|0), (b_3 = 0.45| - 1))$
$f_{33} = R(f_3);$
$G_0 = F((f_{11}| - 1), (f_{22}| - 1), (f_{33}| - 1))$
$G = F((G_0| - 1), (g = -0.1033| - 1))$

This structure is represented in Fig. 3-8.
The NN parameters for the "moons" dataset are presented in 3.3 Table.

Table 3.3: NN (2,3,2) parameters used for the toy dataset 'moons'

(a)

w_{11}	w_{12}	b_1
6.2888	−3.2930	−3.0992
−3.5880	−4.2940	5.9965
−6.1958	−2.7684	0.6882

(b)

w_{21}	w_{22}	w_{23}	b_2
−6.1143	−6.8860	−6.8151	−0.0621
6.6825	6.2848	6.8381	−0.0325

First row of the approximations parameters is given in Table 3.3 .

Table 3.4: XY NN (2,3,1) parameters used to approximate the standard NN for the toy dataset 'moons'

(a)

w_{11}	w_{12}	b_1
1.0	−0.125	−0.9
1.0	−0.125	0.9
−0.5	−0.125	0

(b)

w_{21}	w_{22}	w_{23}	b_2
1.0	1.0	1.0	0.065

The presented architecture was adjusted for better representation of one final feature. For the case of "moons," the additional adjustment has been added since the presented XY architecture has lower expressivity for the case of sharp boundaries. Another approximation stage leads us to the final architecture, which can be found in Fig. 3-9:
$y_{11} = F((F((y_1| - 1), (\pi|0.99))| - 1), (\pi|0.99))$
$y_{12} = F((F((y_{11}| - 1), (\pi|0.99))|1), (\pi|2))$
$f_1 = F((x_1| - 1), (y_{12}| - 1), (b_1 = -0.95| - 1))$
$f_{11} = R(f_1);$
$y_{21} = F((F((y_2| - 1), (\pi|0.99))| - 1), (\pi|0.99))$
$y_{22} = F((F((y_{21}| - 1), (\pi|0.99))|1), (\pi|2))$
$f_2 = F((x_1| - 1), (y_{22}| - 1), (b_1 = 0.9| - 1))$
$f_{22} = R(f_2);$
$x_{31} = F((F((x_3| - 1), (\pi|0.99))|1), (\pi|2))$
$y_{31} = F((F((y_3| - 1), (\pi|0.99))| - 1), (\pi|0.99))$
$y_{32} = F((F((y_{31}| - 1), (\pi|0.99))|1), (\pi|2))$

$f_3 = F((x_{31}| - 1), (y_{32}| - 1), (b_1 = 0| - 1))$
$f_{33} = R(f_3);$
$G_0 = F((f_{11}| - 1), (f_{22}| - 1), (f_{33}| - 1))$
$G = F((G_0| - 1), (g = 0.0216| - 1))$

The presented structure follows the graph structure given in Fig. 3-9.

The presented DL architectures are defined with the Pytorch library's help in the following Table 3.5.

Table 3.5: Examples of the simple DL architectures used to represent nonlinear/unique functions. (a) NN for simple 10 classes digit recognition. (b) NN for CIFAR10 dataset classification.

(a)

NN layer
5 × 5 Conv2D(3,6)
2 × 2 MaxPool2D
5 × 5 Conv2D(6,16)
Linear (400,120)
Linear (120,84)
Linear (84,10)

(b)

NN layer
5 × 5 Conv2D(1,10)
2 × 2 MaxPool2D
ReLu
Dropout(0.5)
5 × 5 Conv2D(10,20)
2 × 2 MaxPool2D
ReLU
Flatten
Linear (320,50)
ReLU
Linear (50,10)
Softmax

In Table 3.5, one can find the details of two DL architectures that were discussed in the main text emphasizing their nonlinear operations.

Finally, the initial task of approximating a particular set of mathematical operations by the parametrized family of nonlinear functions can be done more rigorously with a potential for the accumulated error estimation through the layers of NN.

The discrepancy between the target function and its approximation can be estimated with the $L^1([-\pi/2, \pi/2])$ norm on the working domain:

$$L^1 = \int_{-\pi/2}^{\pi/2} | - \operatorname{sign} B(x, \{J_i\}) \left(\frac{\pi}{2} + \arcsin \frac{A(x, \{J_i\})}{\sqrt{A(x, \{J_i\})^2 + B(x, \{J_i\})^2}} \right) - f(x)_{\text{target}} | dx. \tag{3.36}$$

Starting with the multiplication operation, one can calculate Eq. (3.36) with $f(x,k)_{target} = kx$ and obtain the expression (depending on the (J,k) parameters) for one block of spins. Further minimisation of Eq. (3.36) leads to the expression for $J(k)$. Evaluating Eq.(3.36) analytically can be done in simpler way by replacing the expression involving arcsin function the one with arcctg, so that initial integral (in terms of the argument of the complex parameter $C = \sum_i^{N-1} J_i e^{i\theta_i}$) contains the following expression for one input and one control/reference spin:

$$I = \int_{-\pi/2}^{\pi/2} \operatorname{arcctg} \frac{B((x|-1),(\pi|J))}{A((x|-1),(\pi|J))} dx = \int_{-\pi/2}^{\pi/2} \operatorname{arcctg} \frac{\sin(x)}{J + \cos(x)} dx. \tag{3.37}$$

We evaluate this to

$$I = x \operatorname{arcctg} \frac{\sin(x)}{J + \cos(x)} + \frac{1}{4}(x^2 + 2i \operatorname{sign}(J^2 - 1) \times \times (i[\operatorname{Li}_2(\frac{D(1-E)}{1+E}) + \operatorname{Li}_2(\frac{D^*(1-E)}{1+E})] +$$
$$+ 2x \operatorname{arctanh}(E^{-1}) - G \operatorname{arctanh}(E) + [G - 2i \operatorname{arctanh}(E)] \log \frac{2J(1+E)}{I_1(tg(x/2) - i)} + [G + 2i \operatorname{arctanh}(E)] \log \frac{2J(1+E)}{I_2(tg(x/2) + i)} +$$
$$+ \log H e^{-ix/2}(2i \operatorname{arctanh}(E) - 2i \operatorname{arctanh}(E^{-1}) + G) + \log H e^{ix/2}(2i \operatorname{arctanh}(E^{-1}) - 2i \operatorname{arctanh}(E) + G))) \Big|_{-\pi/2}^{\pi/2} \tag{3.38}$$

with variables $D(J) = (J^2 + 1 + |J^2 - 1|)/2J$, $E(x,J) = \frac{i|J^2-1|}{(J+1)^2} \operatorname{tg} x/2$, $C(J) = \arccos(-\frac{J^2+1}{2J})$, $H(J) = \frac{i|J^2-1|}{2\sqrt{J}\sqrt{J^2+2J\cos x+1}}$, $I_1(J) = 2i(J-1)$, if $J^2 > 1; 2iJ(1-J)$, otherwise, $I_2(J) = 2iJ(J-1)$, if $J^2 > 1; 2J(J-1)$, otherwise and arcctg, arctanh, arccos denoting the inverse for tangent, hyperbolic tangent, cosine functions respectively with $\operatorname{Li}_s(x) = \sum_{k=1}^{\infty} x^k/k^s$ being the polylogarithm function and $*$ denoting the complex conjugate operation. One can simplify the given formula by the

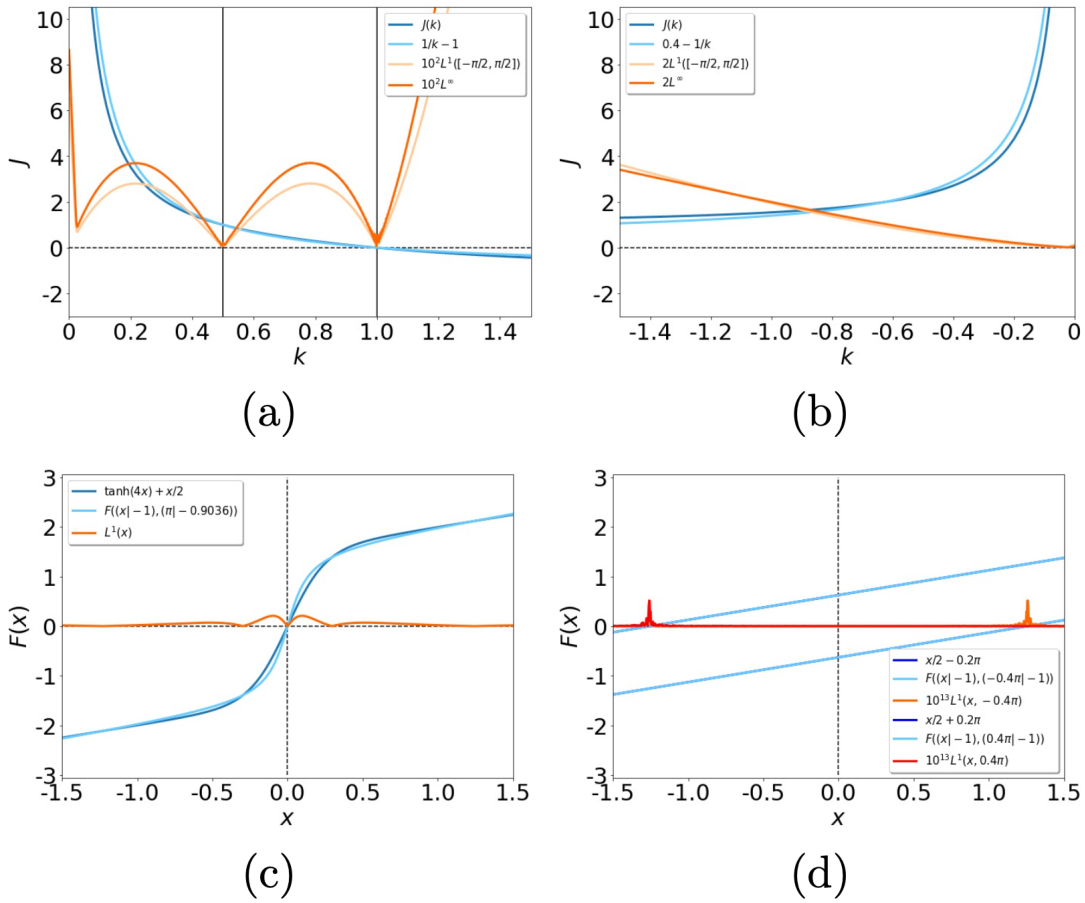


Figure 3-11: Top: (a) Function $J(k)$ that minimizes Eq. (3.36) with $F((x|-1),(\pi|J))$, $f(x)_{target} = kx$ and a positive factor k (blue line), its approximation (white blue line) and the fitted formula $\approx 1/k-1$. The supporting plots depict 10^2L^1 (light orange line) and 10^2L^∞ (orange) for each value of k. Vertical black lines denote the points with the minimal accumulated error. (b) Function $J(k)$ that minimises Eq. (3.36) with $F((x|1),(\pi|J))$, $f(x)_{target} = kx$ and a negative factor k (blue line), its approximation (white blue line) and the fitted formula $\approx 0.4-1/k$. The supporting plots depict $2L^1$ (light orange line) and $2L^\infty$ (orange) for each value of k. Bottom: (c) Function $\tanh 4x + x/2$ (blue line) and its approximation with the function $F((x|-1),(\pi|-0.9036))$ (light blue line) with the optimised parameter J. The supporting orange plot represents the error at each point on x-axis. (d) The half-sums for x and two variables $y = -0.2\pi, 0.2\pi$ (blue lines), which coincide with their approximations $F((x|-1),(y|-1))$ (white blue) giving insignificant approximating error $10^{13}L^1$ (red and orange lines).

contraction of the complex pair terms:

$$I = x\,\mathrm{arcctg}\,\frac{\sin(x)}{J+\cos(x)} + \frac{1}{4}(x^2 + 2i\,\mathrm{sign}(J^2-1)\times(i[\sum_{k=1}^{\infty}\frac{2\cos(k\phi)}{k^2}] + 2G\log H + (2x-G)\,\mathrm{arctanh}(E)+$$
$$+G\log\frac{J(1+E)^2}{E^2(J+1)^2-(J-1)^2} - 2i\,\mathrm{sign}(J^2-1)\times\mathrm{arctanh}(E)\log\frac{J(E(J+1)-(1-J))}{E(J+1)-(J-1)}))\Big|_{-\pi/2}^{\pi/2}, \qquad (3.39)$$

where a new variable $\phi = \mathrm{Arg(D(1\text{-}E)/(1+E)}$ was added.

To shorten the description of the dependence of the coupling strength J on the multiplication factor k and avoid overcomplicated analytical expressions, the plot of its approximation is presented, which alternatively can be calculated numerically and can be approximated by an expression $1/k-1$ with a good accuracy. Additionally, the $L^1([-\pi/2,\pi/2])$ was calculated according to Eq. (3.36) for each value of k. Another good measure of the approximation quality is L^∞, which is the maximal discrepancy between the functions $F((x|-1),(\pi|J))$ and $f(x)_{target}$, which has a similar behaviour as the original norm. Fig. (3-11) (a) shows all the plots corresponding to the multiplication by a positive factor $k > 0$ with the special points of the minimal error at $k = 1, 0.5$ and $k = 0$. These graphs explain why the lesser factors are more reliable for the multiplication, and why multiplying by a larger factors without factorization leads to worse performance.

The same task of multiplication, but by a negative factor $k < 0$ is illustrated in Fig. (3-11)(b). The $J(k)$ function

can be approximated with a reasonable accuracy by an expression $0.4 - 1/k$. Since the general error has a much higher factor $\approx 10^2$ for the negative values, one has to accompany this block with an additional linear embeddings to achieve a good approximation. The nonlinear function $3/2 \tanh 4x + x/2$ and its approximation with the optimised function $F((x| - 1), (\pi| - 0.9036))$ is depicted in the Fig. (3-11)(c). The lowest accuracy is observed near the origin.

The final example of the half-sum approximation is illustrated in Fig. (3-11)(d). The surprisingly good agreement (with an error of 10^{-13} of the magnitude) between the initial function and its XY representation $F((x| - 1), (y| - 1))$ can be explained with the help of the Taylor expansion of Eq.(3.33) near zeros. It gives the linear coefficient of $1/2$ accurate to the fourth order of approximation. With all the presented information, one can estimate the maximal discrepancy between an arbitrary NN and its transferred XY analog, which will be a good measure of the approximation quality and final adequacy of the transfer.

3.5.4 Revision of the results

After the general discussion, it is vital to emphasize the presented work's final results. There are many diverse outcomes of great applied interest and multidisciplinary nature (between the applied mathematics, physics, and computer science domains). The presented work solves many internal problems of the DL architecture transfer approach; however, the list of different outcomes is presented below:

It was shown that using the nonlinear operations, representing the equilibration of the spin coupled to the XY Hamiltonian cluster, we can approximate the basic mathematical operations, in other words, create the nonlinear set of operations for such logic. Besides the basic mathematical operations, we can upscale this system for DL architectures of different complexity and calculate the general accumulated error after changing the architecture by its nonlinear spin logic analog. The DL transfer was the initial motivation for such a task.

The considered system of XY spins is essentially nonlinear in the fundamental interactions between the elements. In general, nonlinearity complicates the system in the way it can be hardly manipulated. Despite these difficulties, making the nonlinear system perform an ML-like task in an approximating manner is possible.

It is possible to calculate the probability of a given particular frustrated XY model to be close to the system that realizes a specific ML architecture, given the clear similarity between the mathematical operations in the architecture and spin clusters. In general, any system with an adjustable response and wide range of output values can be treated as a black-box NN. The coupling parameters and fixed spins give the subspace of all possible configurations or the probability of realizing such configurations.

The general methodology of using some multiparametric nonlinear function to approximate the operations in the neural architecture can be extended on the Hamiltonian with different interactions or additional degrees of freedom in the same approximation manner.

The equivalence between the ML transformation and XY Hamiltonian's minimization in the gradient descent manner was shown. Pretrained ML transformation can be performed in a shallow logic by minimizing the XY model with specific parameters since each cluster was designed to give the unique output, and the cluster hierarchy was built in a consequent manner.

In the search for the best physical setup for the specialized hardware, many architectures and concepts appeared that exploit the XY models in the annealing techniques. The given correspondence allows one to adjust every specialized XY hardware to be modified to perform the ML task. Adjusting existing architecture can be of great use in some cases. The number of additional spins required per operation has linear scaling. We have two spin elements per multiplication and two spins per nonlinear operation in the best practice case.

Finally, we can stress the supporting help of working with such approximating techniques in the hybrid architecture case. Suppose the system is performing the non-related ML task. In that case, some of its elements can significantly benefit from the additional clusters, which performs the nonlinear transformations in the same approximating technique, that was proposed.

3.5.5 Reservoir computing

The section before demonstrated the particular type, close to a reservoir computing (RC) system but did not appear to be so. It was shortly mentioned about this unconventional type of computing; however, the term was not considered in great detail. Since this computational method is a subtype of the thesis topic, this section pays additional attention to this phenomenon.

RC is a term that defines the computational framework when the input variables are fed into a fixed dynamical system, which is called a reservoir and performs the nonlinear dynamical evolution, after which the readout variables are read from the system. RC can also be treated as an extension of a particular type of NNs. One can meet additional

requirements for a system to be called the RC. The essential emphasis that the considered system should be fixed and nonlinear. The training or the parameter tuning is performed only at the readout stage, i.e. the final post-processing of the output signal. The two common types of RC are liquid-state machines [243] and echo state networks [244], which also go hand in hand with their NN paradigms.

RC has additional properties that are implied when one speaks about this type of unconventional computing. One usually treats the reservoir system as a collection of recurrently connected units. These units' connectivity structure is usually supposed to be random, and the units themselves are essentially non-linear. The system's dynamic evolution should be driven by the input with the additional property of being affected by the previous states, i.e. have some analog of the internal states similar to the recurrent NNs. In most works, it is assumed that the reservoir has to realize a rich collection of dynamical input-output mapping, which is crucial and without the mathematical rigour corresponds to the available high-dimensional computational space. This transformation of the readout is adapted to the particular task using a simple regression mechanism. The initial research on the RC did not have a mathematical rigour and mostly had a heuristic approach due to the lack of the developed mathematical theory of the dynamical systems. However, the recent RC trends initialize additional interest to the mathematical description, spawn the revision, and new directions in mathematics [245, 246].

The origins of the RC ideas can be traced down to the 2000s. At that point, a new trend of using recurrent NNs started with the mentioned echo state networks [244] and liquid state machines [243]. Some of the ideas came from the ML community; however, several earlier works came from the biological domain. The biologically plausible reservoirs of the spiking-type neurons, usually associated with the liquid state machines, were first in focus. The pioneering works of Peter F. Dominey and Wolfgang Maass were first to lay the RC principle's foundations as early as 1995 [247]. Even after that moment, the author continued extending and refining the models [248, 249]. More details about the RC origins can be found in [250, 251].

The previously discussed work on translating the DL architectures into the XY Hamiltonian system can not be treated as RC since the considered translated architectures were of the feedforward type. The predefined structure does not have recurrent connections, eliminating the RC's link and the discussed concept, despite the initial closeness. In any case, the EP system can possess rich, dynamic behaviour, with the possibility of implementing the RC framework, which will be discussed further.

The computational power of the variety of available systems can reduce the effective computational cost or gain the speed benefit for a particular type of task. In such a way, the RC concept is of great practical significance for many research and industry areas with the potential of direct, helpful economic impact. There are different types of existing RC systems, implemented on the different experimental settings, which are of conventional type [252, 253].

Many various platforms are potentially suitable for implementing the RC concept. There are electron-based systems with analog circuits [254], FPGAs [255, 256], VLSI [257], memristive type of RC with neuromorphic devices and circuits [258]; optoelectronic devices [259]; photonic and optical RC type [260] with the optical node arrays [78], optoelectronic and optical feedback with gain [261], optical feedback in a laser cavity [262]; spintronic RC [263]; mechanical RC [264] and even biological type [265]. Besides, RC's framework is going beyond the traditional platforms, going into the nano-scale materials and substrates that exhibit stimulation-dependent changes. Another proposal is a nano-scale reservoir with quantum dots and chemical compounds that change their absorption spectrum depending on the pH or redox potential in their environment [266]. Each platform has its characteristics, advantages, and disadvantages, coming from the particular physical system realization and design. Part of them benefit from simpler configurations like mechanical settings, faster processes in the optical platform, or cope with imperfection and noise.

With the recent progress in quantum computing, the framework of RC has started to be implied to the quantum platforms. The intrinsic parallelism, connected with the additional degrees of freedom and the exploit of the part of the Hilbert space, promises the richer dynamics in the classical systems. Such systems' computational power can reduce the effective computational cost and memory [267, 268]. With recent advances in the so-called "extraction" of EPs' quantum nature beyond the mean-field approach, it can serve as an excellent platform for quantum RC [209].

The EP setting is the focus of researchers who want to establish it as an RC platform. It is actively investigating all the aspects of the RC concept applicability with the particular emphasis on GL system [76]. Additional research covers the extensions to quantum RC [75].

RC's field is an auspicious one with many exciting research directions, exemplary practical implementations, and inevitable potential benefits; nevertheless, it has its common problems shared between different particular platforms. One shared problem is the discrepancy between the theoretical description of a given physical system and its practical realization of its dynamic evolution. Since such a black box's evolution process is very complex and the system of differential equations describing the particular hardware is vulnerable to small perturbations, it is possible to get the significant variance in some cases. In some cases, these results can also be intractable, while simulation of such systems

may meet the difficulties of results reproducibility.

The theoretical unification is still in progress. The standard mathematical treatment of RC is the recurrent NN approaches; however, there are additional perspectives with delayed dynamical systems [269], cellular automata [270], coupled oscillators [271].

3.6 State of the art applications

Here the reader can meet the brief list of the latest experimental advances that were achieved in EP applications, gathering the outstanding works among the majority of the available papers.

The major aspects of the EP BEC state have been realized. The intriguing new physics in solid-state systems, such as the formation of a macroscopic coherent state, was observed in [272, 273] with the polariton condensation in [274]. The consequent phenomenon of superfluidity was shown in [275, 276], as well as other topological effects, like quantized vortices [277] or [278, 279, 280] together with the non-equilibrium properties [281, 282].

Shifting the attention to the technological applications, one can observe that polaritons appear to be the solid candidates for technological applications such as dissipation-less optical devices [218, 219], considered with the great details in this chapter analog simulators [68]. The information transmission using EP devices is a very fruitful direction of the technological development [220, 221, 222], since the EP complex field has various degrees of freedom that can be manipulated. Simultaneously, the overall system poses high characteristic time or operational speed, low energy consumption and volume efficiency, crucial for modern photonics.

It is possible to recreate all-optical systems, like logical switch within the EP system [283], where using an organic exciton medium with the high quantum yield at room temperature one can start building the foundation for the future all-optical logical networks. The platform's ultra-fast intrinsic speed makes the control over the macroscopic condensate wavefunction via a single photon available [284]. Significant progress is achieved, in comparison with the previous works on optical switches [223, 224, 225], which are potential carriers for fast and efficient information processing due to their photonic component.

There is much space for future research and development of practical applications since many EP BEC parameters were not exploited in terms of information transmission, like the condensate phase.

Another state of the art demonstrates the engineering spin-orbit synthetic Hamiltonians in liquid-crystal optical cavities [285], which is a part of the research aiming at recreating exotic Hamiltonian systems.

3.7 Engineering periodic structures with the target velocity

The precise implementation of well-controllable lattice potentials for bosonic condensates is an inevitable step towards the realization of advanced classical [286, 68, 287] and quantum simulators [288]. Polariton condensates in lattices could be manipulated into soliton, which is seen as a building block for polaritonic circuits, where propagation and localization are optically controlled and reconfigurable [289]. Velocity control is needed to form topologically protected stated (e.g. chiral edge states) at the surface of topological insulators that allow unidirectional transport immune to backscattering and topological edge modes [290], or to form optical delay lines with enhanced transport properties [291]. The full exploitation of polariton's potential for technological applications requires engineering microstructures with certain high quality trapping profiles with precise parameters and realising quasiparticles flow and density control. Polariton condensation has been recently achieved in a one-dimensional strong lead halide perovskite lattices [292, 293] paving the way for all-optical integrated logic circuits operating at room temperature.

However, unlike optical lattices of equilibrium ultra-cold BECs, the periodic trapping of nonequilibrium condensates provides flows between the lattice sites that are difficult to predict and challenging to control. In this section, we show how to control the velocity and density of polariton condensates by using spatially varying dissipation and pumping profiles. The spatially controlled dissipation can be achieved by many methods including proton implant technique [294], excited states absorption [295], by changing the thickness of the film [296] and other techniques [287].

Figure 3-12 shows the schematic of the experimental setup with uniform incoherent injection and nonuniform spatially varying dissipation that creates the desired flow pattern.

Methodology for a periodically varying dissipation. Our starting point is a generic laser or nonequilibrium condensate model in the form of the complex complex Ginzburg-Landau equation (cGLE) with saturable nonlinearity that results from the Maxwell-Bloch equations:

$$i\frac{\partial\Psi}{\partial t} = -\frac{1}{2}\Delta\Psi + |\Psi|^2\Psi + gn_R\Psi + i[n_R - \gamma(\mathbf{x})]\Psi, \tag{3.40}$$

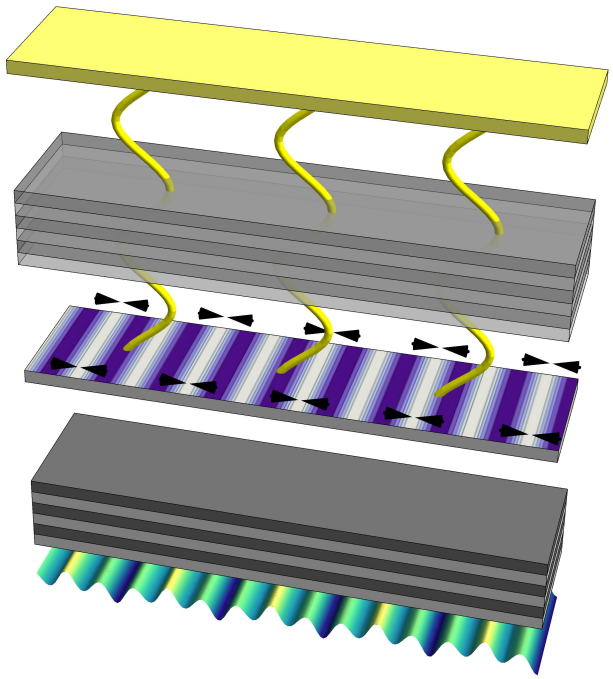


Figure 3-12: Schematics of the flow management approach presented in our work. The top yellow plate represents the source of the incoherent uniform pumping (cw excitation). Grey layers denote the distributed Bragg reflectors between which the quantum well is located and the formation of the polariton condensates takes place. On the top we show density of the polaritons with the light-blue color scheme representing the periodicity of the density regions (dark blue (white) color marks the maxima (minima) of the density). The target flow profile is shown by the black arrows indicating the direction of the flow. Below the structure we show the spatially varying dissipation that leads to such a flow profile.

where $\Psi(\mathbf{x},t)$ is the complex-valued order parameter, $\gamma(\mathbf{x})$ is the spatially varying linear losses due to imperfect confinement, n_R is the density distribution of noncondensed reservoir particles (such as hot excitons) that provides gain saturation $n_R = P(\mathbf{x},t)(1+b|\Psi|^2)^{-1}$, $P(\mathbf{x},t)$ is the gain, b characterizes the relative strengths of the gain saturation and self-interactions. Parameter g characterises the strength of the interactions between the condensate and the reservoir of noncondensed particles. This model has been quite successful in describing many aspects of lasers and polariton condensate dynamics and the resulting steady states [281, 282, 197] due to its universality [297, 298, 299].

As the pumping is increased above the threshold for condensation, the coherence across the pumping region is established. Close to the condensation threshold and with the constant pumping $P = P_c$, $n_R \approx P_c - P_c b|\Psi|^2$ and Eq. (3.40) can be replaced with a more standard form of the cGLE

$$i\frac{\partial\tilde{\Psi}}{\partial t} = -\frac{1}{2}\Delta\tilde{\Psi} + \beta|\tilde{\Psi}|^2\tilde{\Psi} + i[P_c - \gamma(\mathbf{x}) - \sigma|\tilde{\Psi}|^2]\tilde{\Psi}, \tag{3.41}$$

where $\sigma = P_c b|1 - gP_c b|^{-1}$, $\tilde{\Psi} = \Psi\exp[igP_c t]|1 - gP_c b|^{1/2}$, and $\beta = \mathrm{sgn}(1 - gbP_c)$ characterises the sign of the self-interactions: defocusing ($\beta = +1$) or focusing ($\beta = -1$). We also consider $\beta = 0$, in which case $\sigma = P_c b$ and $\tilde{\Psi} = \Psi\exp[igP_c t]$.

If the pumping strength remains constant, the system reaches the steady state characterised by the chemical potential μ, such that $\tilde{\Psi} = \psi(\mathbf{x})\exp[-i\mu t]$. The steady state written using the Madelung transformation $\psi = \sqrt{\rho(\mathbf{x},t)}e^{iS(\mathbf{x},t)}$ in terms of the density ρ and phase S satisfies

$$\mu = \frac{1}{2}\mathbf{u}^2 - \frac{1}{2}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}} + \beta\rho, \tag{3.42}$$

$$\frac{1}{2}\nabla\cdot(\rho\mathbf{u}) = (P - \gamma(\mathbf{x}) - \sigma\rho)\rho, \tag{3.43}$$

where $\mathbf{u} = \nabla S$ is the condensate outflow velocity profile.

The systems out of equilibrium such as lasers or nonequilibrium condensates are characterised by the existence of nonzero velocity profiles even at the steady states. The steady-state currents connect regions of net gain with those of net loss which form due to the combination of density-dependent gain rate and spatial inhomogeneity of either pumping or dissipation. The superflow velocities in turn affect the density profile. The main question we address in this section is how to engineer a given velocity profile using the controls available in our system such as the spatially varying pumping or dissipation. This control will be achieved by observing that nonlinear Eq. (3.42) can be solved yielding the solutions in terms of ρ for spatially constant pumping and a given velocity profiles. Such density when substituted into Eq. (3.43) allows one to specify the spatially varying dissipation that leads to the given superflow. This is a generic procedure, that can be made to works for any physically relevant target velocity profile. We will illustrate this approach using the velocity profiles that allow to integrate Eq. (3.42) exactly leading to fully analytical pumping and dissipation profiles.

We also use this method for non-periodic velocity profiles and spatially localised excitations that require numerical integration of Eq. (3.42). We consider a one-dimensional velocity profile $\mathbf{u} = u(x)$, however, the extension to higher dimensions is straightforward. Equation (3.42) is the second order nonlinear ordinary differential equation which for a given periodic expression for $u(x)$ relates to the steady state of the Gross-Pitaevskii equation (GPE) that describes trapped quasi-one-dimensional dilute gas of the Bose-Einstein condensate (BEC). The GPE for such gas reads

$$i\hat{\psi}_t = -\frac{1}{2}\hat{\psi}_{xx} + \beta|\hat{\psi}|^2\hat{\psi} + V(x)\hat{\psi}, \tag{3.44}$$

where $\hat{\psi}(x,t)$ is the macroscopic wavefunction of the Bose-Einstein condensate and $V(x)$ is the trapping potential. The steady state of Eq. (3.44) with $\hat{\psi} = \sqrt{\hat{\rho}}\exp[i\hat{S} - i\hat{\mu}t]$ and a constant phase $\hat{S}$ satisfies $\hat{\mu} = V(x) - (\sqrt{\hat{\rho}})_{xx}/2\sqrt{\hat{\rho}} + \beta\rho$, which reduces to Eq. (3.42) if we let $2V(x) = u(x)^2$, $\hat{\mu} = \mu$ and $\hat{\rho} = \rho$. So the trap used in equilibrium condensates can be thought of as the postulated expressions for the velocity profiles. The density, therefore, can be obtained similarly to finding the stationary density of equilibrium condensates in a given external potential either analytically (if such solution is known) or numerically by integrating Eq. (3.44) in imaginary time while renormalizing the wavefunction by the fixed number of particles. Once we know the density profile, Eq. (3.43) gives the expression for the spatial dissipation that leads to that density and velocity

$$\gamma(x) = P_c - \sigma\rho - \frac{1}{2}u_x - \frac{1}{2}\rho_x u/\rho. \tag{3.45}$$

For some external potentials the exact solution of the GPE could be found. An important class of such potentials are Jacobi elliptic functions, so we first consider $u(x)^2 = -2V_0\mathrm{sn}^2(x,k)$, where $\mathrm{sn}(x,k)$ denotes the Jacobi elliptic sine function with an elliptic parameter $0 < k < 1$ and $V_0 < 0$ characterizes the depth of the periodic variation. Such periodic profiles require translational invariance, that can be realised experimentally using toroidal geometry of excitation [300]. Denoting $r(x) = \sqrt{\rho(x)}$, Eq. (3.42) becomes

$$\mu r^4(x) = -\frac{r^3(x)r''(x)}{2} + \beta r^6(x) - V_0\mathrm{sn}^2(x,k)r^4(x). \tag{3.46}$$

The solutions to the nonlinear ordinary differential Eq. (3.46) can be obtained by expanding $r(x)$ in terms of the various Jacobi elliptic functions $(\mathrm{sn}, \mathrm{dn}, \mathrm{cn})$, substituting $r(x)$ into Eq. (3.46) and equating equal powers of these functions to zero to get the expressions for the parameters [301, 302].

Defocusing regime; $\beta = 1$. Firstly, we consider $r^2(x) = A\,\mathrm{sn}^2(x,k) + B$, which leads to the following conditions $\mu = \frac{1}{2}(1 + k^2 + 3B - \frac{BV_0}{k^2 - V_0})$, $0 = B(1 + \frac{B}{k^2+V_0})(k^2 + V_0 + Bk^2)$, $A = k^2 + V_0$. These conditions are satisfied in the following three cases. 1) $B = 0$ and $A = k^2 + V_0$ with $\mu = (1 + k^2)/2$ result in

$$\rho(x) = (V_0 + k^2)\mathrm{sn}^2(x,k), \quad -k^2 \le V_0 < 0; \tag{3.47}$$

2) $B = -A = -(k^2 + V_0)$ with $\mu = 1/2 - k^2 - V_0$ results in

$$\rho(x) = -(V_0 + k^2)\mathrm{cn}^2(x,k), \quad V_0 \le -k^2; \tag{3.48}$$

3) $B = -(V_0 + k^2)/k^2 = -A/k^2$ with $\mu = -1 - V_0/k^2 + k^2/2$ results in

$$\rho(x) = \frac{-(V_0 + k^2)}{k^2}\mathrm{dn}^2(x,k), \quad V_0 \le -k^2. \tag{3.49}$$

Secondly, we consider $r^2(x) = a_1\mathrm{cn}(x,k) + b_1$ which yields the following conditions on the coefficients: $V_0 = -\frac{3}{8}k^2$, $\mu = \frac{1}{8}(1 + k^2) + \frac{6a_1^2}{k^2}$, $0 = \frac{a_1^2}{4k^6}(16a_1^2 - k^4)(16a_1^2 + k^2 - k^4)$, $b_1 = \frac{4a_1^2}{k^2}$. This leads to two possibilities for amplitude a_1: $a_1 = k^2/4$, and $a_1 = -k^2/4$, resulting in

$$\rho(x) = \frac{k^2}{4}(1 \pm \mathrm{cn}(x,k)). \tag{3.50}$$

Finally we consider $r^2(x) = a_2\mathrm{dn}(x,k) + b_2$ that gives the conditions on the parameters $V_0 = -\frac{3}{8}k^2$, $\mu = \frac{1}{8}(1+k^2) + 6a_2^2$, $0 = \frac{a_2^2}{4}(16a_2^2 - 1)(16a_2^2 + k^2 - 1)$, $b_2 = 4a_2^2$. This leads to $a_2 = \pm 1/4$, $\mu = 1/2 + k^2/8$ with the expression on the amplitude

$$\rho(x) = \frac{1}{4}(1 \pm \mathrm{dn}(x,k)), \tag{3.51}$$

or 2) $a_2 = \sqrt{1 - k^2}/4$, $\mu = 1 - k^2/2$ with $k' = \sqrt{1 - k^2}$ leading to

$$\rho(x) = \frac{k'}{4}(k' + \mathrm{dn}(x,k)). \tag{3.52}$$

The dissipation profile that leads to these density and flow profiles is determined by Eq. (3.45). The representative examples of these solutions are depicted in Fig. 3-13.

Focusing regime; $\beta = -1$. The focusing case yields some different sets of solutions. Taking $r^2(x) = A\mathrm{sn}^2(x,k) + B$ leads to $\mu = \frac{1}{2}(1 + k^2 - 3B + \frac{BV_0}{V_0+k^2})$, $0 = B(\frac{B}{V_0+k^2} - 1)(V_0 + k^2 - Bk^2)$, $A = -(V_0 + k^2)$. The possible choices are 1) $B = 0$ with $\mu = (1 + k^2)/2$ leading to

$$\rho(x) = -(V_0 + k^2)\mathrm{sn}^2(x,k), \quad -k^2 \geq V_0, \tag{3.53}$$

2) $B = -(k^2 + V_0)$, with $\mu = 1/2 - k^2 - V_0$ leading to

$$\rho(x) = (V_0 + k^2)\mathrm{cn}^2(x,k), \quad -k^2 \leq V_0 < 0, \tag{3.54}$$

or 3) $B = (V_0 + k^2)/k^2$ with $\mu = -1 - V_0/k^2 + k^2/2$ leading to

$$\rho(x) = \frac{(V_0 + k^2)}{k^2}\mathrm{dn}^2(x,k), \quad -k^2 \leq V_0 < 0. \tag{3.55}$$

Taking $r^2(x) = a_1\mathrm{sn}(x,k)+b_1$ gives us a set of conditions: $V_0 = -\frac{3}{8}k^2$, $\mu = \frac{1}{8}(1+k^2)-\frac{6a_1^2}{k^2}$, $0 = -\frac{a_1^2}{4k^6}(16a_1^2-k^4)(16a_1^2-k^2)$, $b_1 = \frac{4a_1^2}{k^2}$. If 1) $a_1 = \pm k^2/4$, $\mu = 1/8 - k^2/4$, we get

$$\rho(x) = \frac{k^2}{4}(1 \pm \mathrm{sn}(x,k)), \tag{3.56}$$

or 2) $a_1 = \pm k/4$, $\mu = -1/4 + k^2/8$, so that

$$\rho(x) = \frac{1}{4}(1 \pm k\,\mathrm{sn}(x,k)). \tag{3.57}$$

Finally, we consider $r^2(x) = a_2\mathrm{dn}(x,k)+b_2$ which yields $V_0 = -\frac{3}{8}k^2$, $\mu = \frac{1}{8}(1+k^2)+6a_2^2$, $0 = \frac{a_2^2}{4}(16a_2^2-1)(16a_2^2+k^2-1)$, and $b_2 = -4a_2^2$. The only nontrivial solution exists for $a_2 = \sqrt{1-k^2}/4$. Let $k' = \sqrt{1-k^2}$, and with $\mu = 1/4 + k'^2/4$ the solution becomes

$$\rho(x) = \frac{k'}{4}(\mathrm{dn}(x,k) - k'). \tag{3.58}$$

For any given uniform pumping P_c the corresponding dissipative profile is given by Eq. (3.45) with some representative examples of these solutions shown in Fig. 3-13.

Non-interacting case: $\beta = 0$. The case of $gP_cb = 1$ has fewer physically relevant solutions. For $r^2(x) = A\mathrm{sn}^2(x,k)+B$, we get $2\mu AB = -V_0B^2 + AB + ABk^2$, $0 = (\mu B + A/2)B$, $2\mu A^2 = -3Ak^2B + A^2(1 + k^2) - 4ABV_0$, $V_0 = -k^2$, which leads to the following possibilities. $1)B = 0$ and $\mu = (1 + k^2)/2$ gives (with $A > 0$)

$$\rho(x) = A\,\mathrm{sn}^2(x,k), \tag{3.59}$$

or 2) $2\mu = -A/B = 1$ which yields ($A < 0$)

$$\rho(x) = -A\,\mathrm{cn}^2(x,k), \tag{3.60}$$

and 3) $2\mu = -A/B = k^2$ results in ($A < 0$)

$$\rho(x) = -A\,\mathrm{dn}^2(x,k)/k^2. \tag{3.61}$$

For $r^2(x) = a_1\mathrm{cn}(x,k) + b_1$, or $r^2(x) = a_1\mathrm{dn}(x,k) + b_1$ setting the terms before the basic functions to zero imposes equations: $\mu a_1^2 = -V_0a_1^2 + V_0b_1^2 + (1 - 2k^2)a_1^2/8$, $\mu b_1^2 = -V_0b_1^2 + (1 - k^2)a_1^2/8$, for cn-type and $\mu a_1^2 = -V_0a_1^2/k^2 + V_0b_1^2/k^2 + (k^2 - 2)a_1^2/8$, $\mu b_1^2 = -V_0b_1^2/k^2 + (k^2 - 1)a_1^2/8$, for dn-type with two common equations $V_0a_1^2 + (3a_1^2k^2)/8 = 0$, and $4V_0a_1b_1 + a_1b_1k^2 = 0$, which are compatable only with $b_1 = 0$, $V_0 = -3k^2/8$, giving the resulting values for $k = 1$, $\mu = 1/4$. The final solutions in both cases take the form of a solitary wave:

$$\rho(x) = a_1\mathrm{sech(x)}, a_1 > 0. \tag{3.62}$$

Next, we consider the stability of the presented solutions by evolving them dynamically according to Eq. (3.40). The instability takes place when the dissipative profile becomes strongly negative in some spatial regions, so starts representing the generation of quasiparticles that can not be compensated by the remaining positive dissipative regions. This leads to the uncontrollable growth of particles in the repulsive ($\beta = 1$) and non-interactive ($\beta = 0$) regimes; in attractive regime ($\beta = -1$) the system either exhibits oscillations due to the interplay between attractive forces and fast saturation response or undergoes the transition to another stationary solution. We note that when the solution is stable, the system finds it starting from any random noise, not just from the perturbed solution. This assures that the control

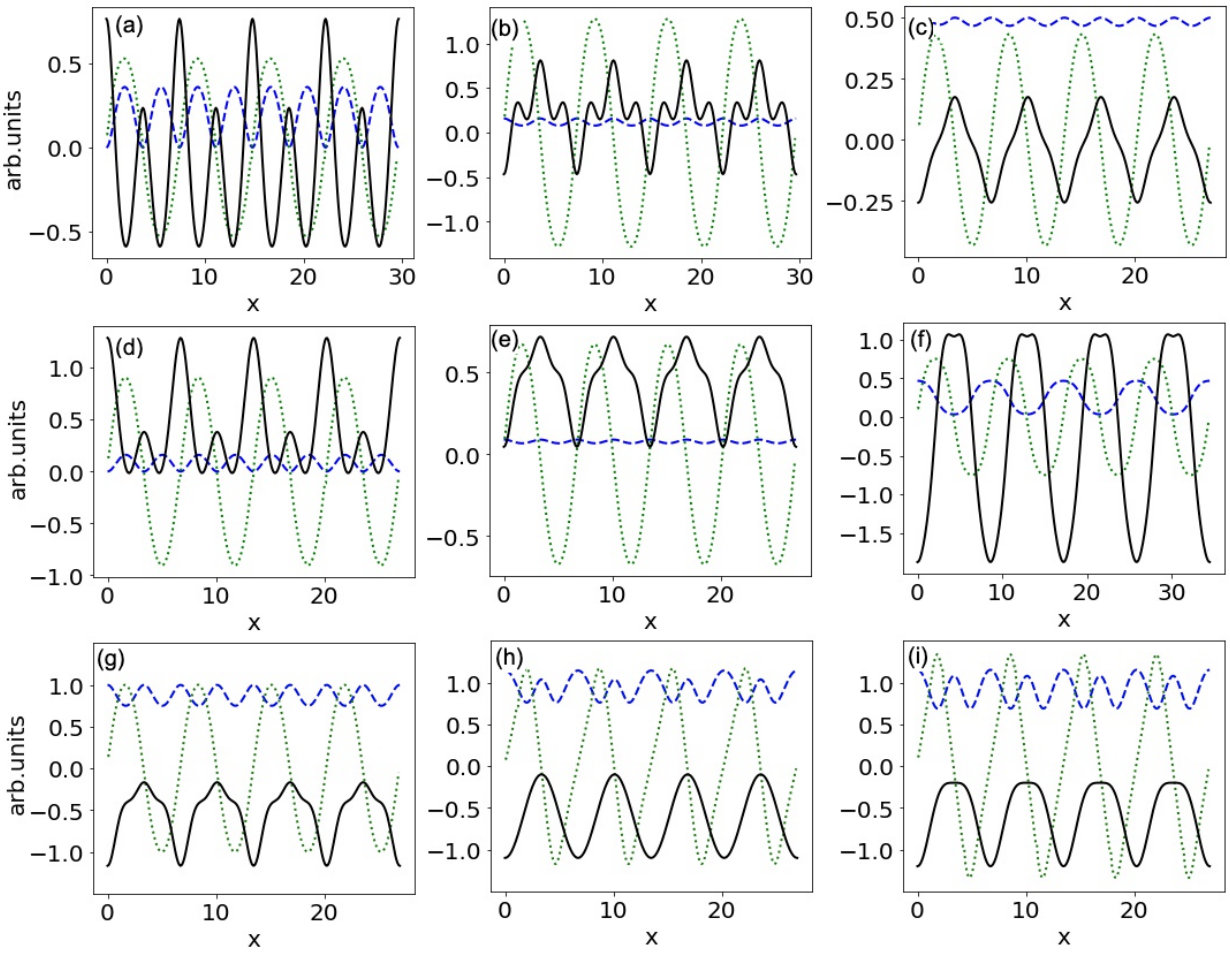


Figure 3-13: Flow velocity $u(x)$ (green dotted lines), density $\rho(x)$ (dashed blue lines) and the dissipation profile $\gamma(x)$ (black solid lines) in the repulsive $\beta = 1$ (a,b,c),(g,h,i) and attractive $\beta = -1$ (d,e,f) regimes (with $g = 1, b = 1.5$). Shown are the solutions given by Eq. (3.47) with $P_c = 0.5$, $k = 0.707$, $V_0 = -0.14$ (a), Eq. (3.49) with $P_c = 1/3$, $k = 0.707$, $V_0 = -0.58$ (b), Eq. (3.52) with $P_c = 0.5$, $k = 0.866$, $V_0 = -0.28125$ (c), Eq. (3.53) with $P_c = 5/6$, $k = 0.5$, $V_0 = -0.41$ (d), Eq. (3.55) with $P_c = 5/6$, $k = 0.5$, $V_0 = -0.2275$ (e), Eq. (3.57) with $P_c = 5/6$, $k = 0.866$, $V_0 = -0.28125$ (f), Eq. (3.49) with $P_c = 1/3$, $k = 0.5$, $V_0 = -0.5$ (g,h,i). The last row demonstrates the stable solutions with the perturbed dissipation profiles given by the equations $0.5\gamma_{max}(1 - \text{sign}(\text{cn}(x,k))|\text{cn}(x,k))|^{0.8}$ (h) and $0.5\gamma_{max}(1 - \text{sign}(\text{cn}(x,k))|\text{cn}(x,k))|^{0.6}$ (i).

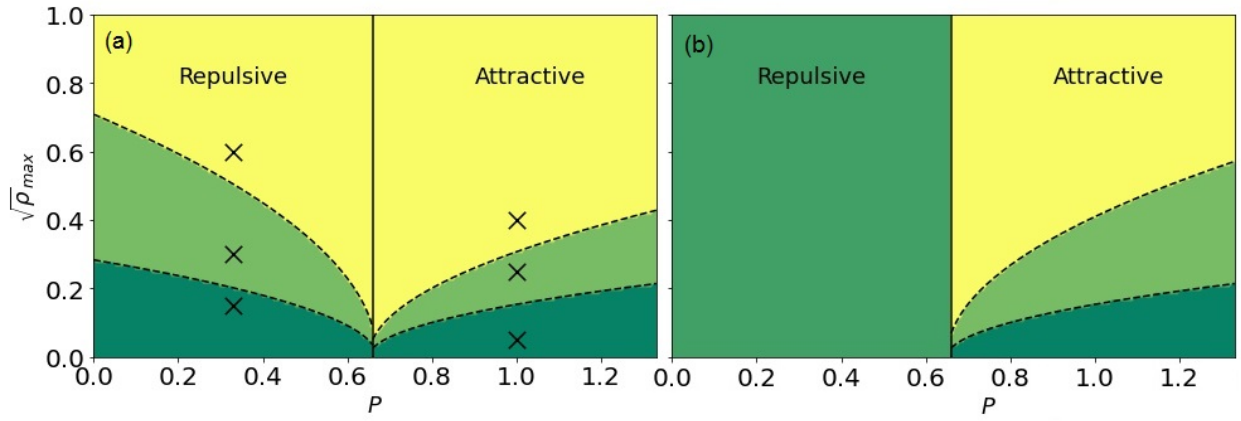


Figure 3-14: Stability diagrams for the family of dn*(x,k)* solutions given by Eq. (3.49), Eq. (3.55) and Eq. (3.61) (with $g = 1, b = 1.5$), obtained by numerical integration of Eq. (3.40)(a, without the Taylor expansion for the pumping term) and Eq. (3.41)(b, with the Taylor expansion). Green color stands for stable, yellow - for unstable and light green - for the transitional regimes to a different stable configuration. Black line depicts the solutions with $\beta = 0$, dash lines - the borders for the specified regions, while black crosses indicate the solutions also shown in Fig. 3-15.

is working and leads to the target velocity profile.

Figures 3-15 (a-c) and (d-f) demonstrate the behavior of the solutions governed by Eq. (3.49) and Eq. (3.55) respectively: stable evolution, evolution to a different stationary solution and unlimited growth of the density due to the lack of the gain saturation.

On the one hand, the negative dissipation (gain) has been experimentally realised by using injection below the threshold for the condensate formation [303]. On the other hand, one can use the linear relationship between the pumping intensity P_c and the dissipation given by Eq. (3.45) to increase P_c to make $\gamma > 0$ for all $\mathbf{x}$. This will stabilise the solution while removing the necessity to generate nonuniform gain. This will not work for analytic solution if changing P_c changes the sign of $\beta = \text{sign}(1 - P_c g b)$ that controls the type of the solution (focusing, defocusing, linear). So this approach would not work if increase in P_c changes the solution type. To overcome this problem one may choose the microcavity with lower values of g (more excitonic) and b (stronger polariton-polariton interactions)[304]. In Fig. 3-14 we used the experimentally largest values g and b.

Experimentally achievable dissipation and pump may suffer from imperfections. In Fig.(3-13) (g,h,i) we show the effect of perturbing the effective pumping/dissipation profiles.

General framework for the velocity engineering. So far we considered the velocity profiles for which the analytical solution of the governing equations exists: periodic profiles are expressed as Jacobi elliptic functions. However, *any* physically realistic velocity profile can be engineered using numerical integration of Eq. (3.44). Some of the resulting pumping/dissipation profiles are shown in Fig. 3-16 for analytical velocity profiles that correspond to a spatially localised excitations. Several subtleties arise that are not present when analytical solutions of Eq. (3.44) exist. To find the steady state of Eq. (3.42) for a given velocity profile, we used the relaxation of Eq. (3.44) (integration in imaginary time) while renormalizing the wavefunction to a fixed number of particles N at every time step. To use the spatially localised excitation, the velocity profile has to be a constant away from the excitation spot [188], so we used the velocity profiles that satisfy this condition. The resulting density decays as $\exp[-2\mu x]$ away from the excitation to satisfy Eq. (3.42). Such straightforward asymptotics allows us to evaluate the last term of Eq. (3.45) that otherwise would be problematic to find numerically for small densities.

The cGLE-type systems are pattern-forming systems with a complex interplay of gain, dissipation, nonlinearities due to self interactions and gain saturation. This makes the engineering of the flow and density profiles in such systems challenging. However, such control is necessary for implementation of various proposals from topological insulators to optical/photonic/polaritonic transistors, information processing devices and Hamiltonian simulators. In this work we proposed a method for realising periodic modulations of the density and velocity profiles in a wide variety of laser/condensate systems governed by the cGLE.

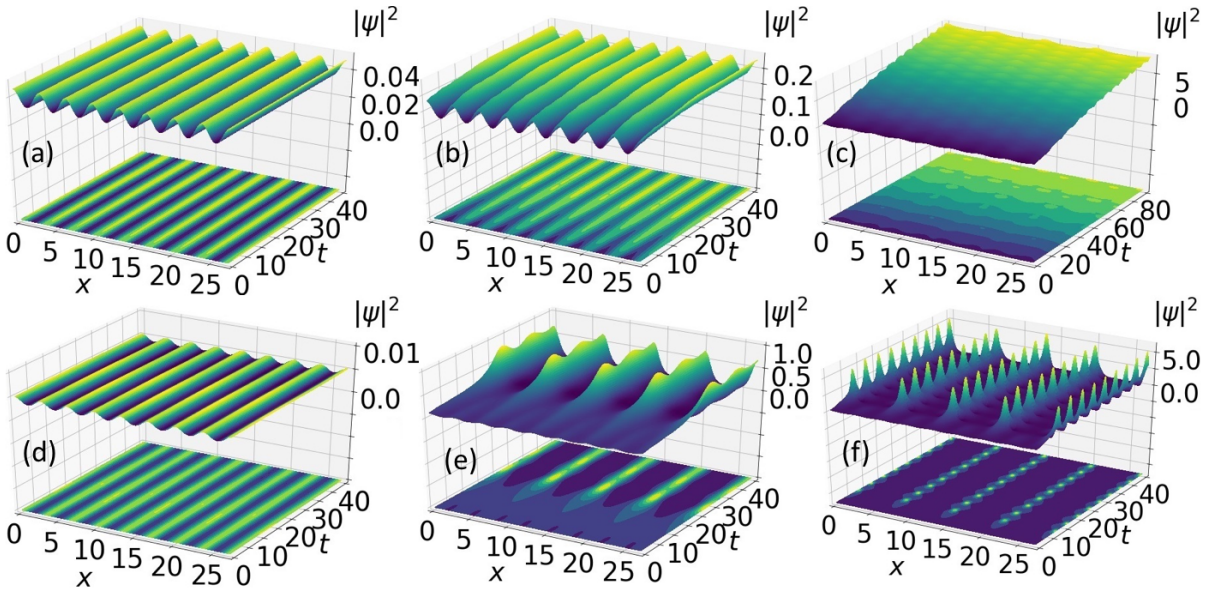


Figure 3-15: The stable (left column), transitional (middle column), unstable (right column) evolutions of dn*(x,k)* solutions given by Eq. (3.49) in a repulsive regime ($P_c = 1/3$, $\sigma = 1, \beta = 1, g = 1$) with $k = 0.5$ and $\sqrt{\rho_{max}}(t = 0) = 0.15$ (a), $\sqrt{\rho_{max}}(t = 0) = 0.3$ (b), $\sqrt{\rho_{max}}(t = 0) = 0.6$ (c) and by dn*(x,k)* solutions given by Eq. (3.55) in an attractive regime ($P_c = 1$, $\sigma = 3, \beta = -1$) with $k = 0.5$ and and $\sqrt{\rho_{\max}}(t = 0) = 0.05$ (d), $\sqrt{\rho_{\max}}(t = 0) = 0.25$ (e), $\sqrt{\rho_{\max}}(t = 0) = 0.4$ (f).

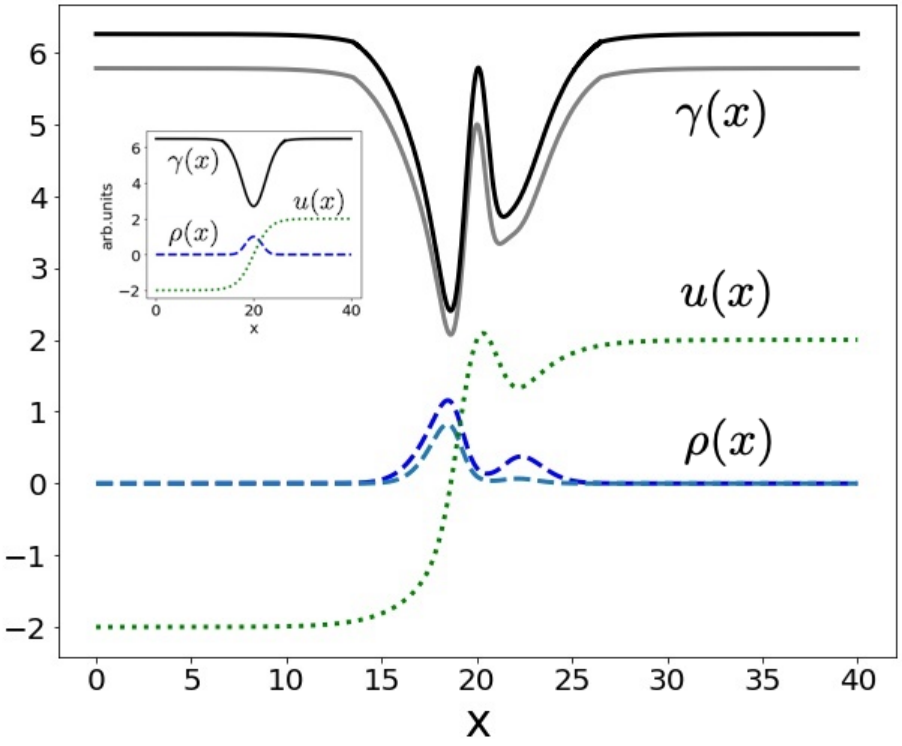


Figure 3-16: Flow velocity $u(x)$ (green dotted lines), density $\rho(x)$ (dashed dark and white blue lines) and the dissipation profile $\gamma(x)$ (black and grey solid lines) in the repulsive, $\beta = 1$, regime. The velocity profile and pumping were chosen as $u(x) = 2\tanh[0.3(x - 20)] + 2\exp[-0.5(x - 20)^2]$ and $P_c = 4$. The number of the particles as $N = \int |\Psi|^2 dx = 2$ (faint dashed-blue line for density and black line for dissipation) and $N = 4$ (dark-blue line for density and grey line for dissipation). The inset shows the velocity profile that leads to a single localised condensate peak and the velocity profile $u(x) = 2\tanh(0.3(x - 20))$.

Chapter 4

Algorithmic network

In this Chapter one can find the wide spectrum of algorithms, that can be encoded into the QUBO setup of the special-purpose hardware:

1. Optimization tasks
   (a) PUBO/QUBO (polynomial/quadratic unconstrained binary optimization)
   (b) Quadratic/polynomial programming
   (c) Nonlinear programming
2. Machine learning methods
   (a) Classical ML
      i. Regression and least squares estimation
      ii. Classification (k-NN, SVM)
      iii. Dimensionality reduction (PCA, SVD, LSA)
      iv. Clusterization (k-means, mean-shift)
   (b) Neural network architectures (Hopfield NN)
   (c) NN ensembles
   (d) Image processing tasks
   (e) Probabilistic graphical models
3. Direct encoding/decoding examples
   (a) SAT formulations
   (b) Karp NP-complete examples: MIS, set cover, treewidth computation

The additional useful links are provided on the case studies, where one can find the encoding/decoding correspondence in the context of the D-wave processor for the following tasks: graph partitioning, finding maximum cliques, nonnegative/binary matrix factorization, portfolio optimization problems, prime factorization and quantum chemistry problems.

4.1 Unified optimization picture

The primary motivation is to present the unified picture of algorithms that can be solved on the near term condensed matter and optical systems, operating in the annealer or GD-based manner and lacking the option of the proper programmability. It is essential to look at the current chapter from multiple perspectives, considering various tasks that can be heuristically solved via annealing-like techniques, algorithmic perspective on the presented equations, and the general physical view. Moreover, one can find additional information about the supplementary topics, such as constructing hard instances for the optimization tasks and their properties, their connection with the disorder physics, several helpful advice and techniques on implementing the obtained results within the hardware setup.

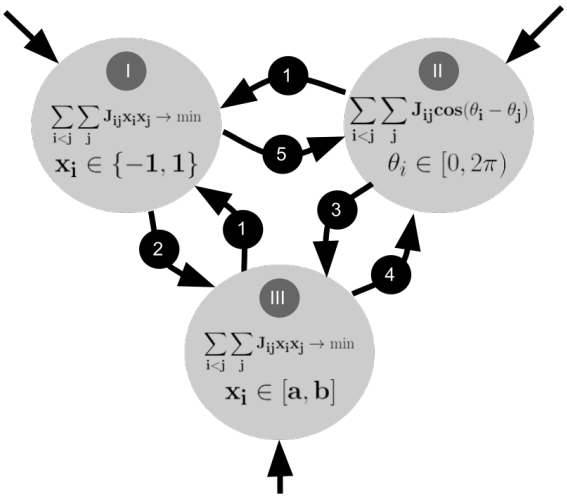


Figure 4-1: Different types of optimization problems concerning the variable domain for the functionals with the pairwise interaction. The outer black arrows denote the initial formulation of the problem. One can consider the typical QUBO under the roman number I, the classical XY-Hamiltonian II on the periodic domain, as an example of a nontrivial function, or the quadratic polynomial optimization problem on the domain $[a, b]$ III. In case of the limited functionality of the special-purpose hardware, one is usually able to solve only one of the problems. It is usefull to work with all of the presented formulations, to avoid the extensive description of efficient mapping between different assignments. For this purpose, it is necessary to present various embeddings between several types of optimization problems.1 denotes the introduction of the effective force-field (which is similar to the use of tanh term, see Hopfield NN further) to constraint the variable into the discrete domain. 2 denotes the continuous domain's discretization into the set of possible values and introduces additional variables to exploit the binary variables as the set of integers. $3, 4, 5$ denotes the unique embeddings, depending on the function type II. All details of the mappings are discussed in the main text.

Several types of optimization problems differ in either the target function or the available range of variables. Some of the discussed assignments have efficient embeddings into one of these optimization tasks but not into all of them. In order to avoid excessive description of the methodology for mapping each particular assignment into a different type of optimization problem, it is helpful to present the unified framework for all of the optimization tasks. This relieves everybody from the specification of the optimization problem in each case of practical assignment since it is enough to connect the problem with one optimization type that can be easily embedded into any other through the presented scheme.

However, the specific limitations should be imposed on the covered material that arises from the recent developments coming from the state of the art quantum information science and specific quantum systems. It is better to avoid the colossal volume of the results from quantum systems involving additional degrees of freedom governed and representing the intrinsic quantum effects, typical qubit systems or topological quantum computing. Thus, the focus is on the more robust and near-term devices, aiming to solve the three different kinds' general optimization task.

The optimization problem (I) studied in this chapter is called polynomial unconstrained binary optimization (PUBO), also known as high-order binary optimization (HOBO):

$$\min_{\mathbf{x}\in\{-1,+1\}^N} -\sum_{\Omega} \mathbf{A}^k_{i_1,\cdots,i_k} x_{i_1} x_{i_2} \cdots x_{i_k}, \tag{4.1}$$

where $\Omega = \{i_j : 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq N\}$ and $\mathbf{A}^k$ is the super-symmetric tensor of degree k. The alternative optimization problems considered in this chapter that can be mapped at the Eq. (4.1) and in the reverse way can be formulated in slightly different forms (III) are called polynomial programming (PP):

$$\min_{\mathbf{x}\in[a,b]^N} -\sum_{\Omega} \mathbf{A}^k_{i_1,\cdots,i_k} x_{i_1} x_{i_2} \cdots x_{i_k}, \tag{4.2}$$

with the same notation except, but different available domain for the arguments. The constraints in the form $x_i \in [a, b]$ can be rewritten in the following way $x_i \leq b$ and $-1x_i \leq -a$. Thus, the PP form is obtained, which is to minimize the same target function $\mathbf{A}^k_{i_1,\cdots,i_k} x_{i_1} x_{i_2} \cdots x_{i_k}$, subject to $B\mathbf{x} \preceq \mathbf{c}$, where $\mathbf{c}$ consists of $-a$ and b values and B consists of $+1, -1$ coefficients for each variable x_i respectively. The special case of $\mathbf{A}_{i_1,i_2}$ with rank 2 is known as quadratic

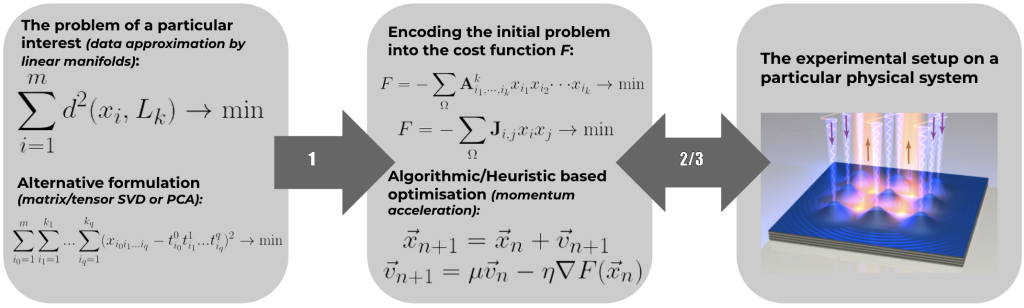


Figure 4-2: The visualization of the methodology for solving a particular task on the physical system. The left box presents a certain operation or a specific assignment. Arrow 1 direct us to the middle box, representing the encoding procedure. Following arrow 2, one can use the obtained parameters to operate a physical system to obtain a solution. Number 3 on the same arrow represents the heuristics method arising from the model description of the same physical system that can be programmed for the conventional computer architecture with all the model's corresponding benefits.

programming (QP) and can be written as:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^{\mathrm{T}} A\mathbf{x} \\ \text{subject to} & B\mathbf{x} \preceq \mathbf{c}, \end{array} \tag{4.3}$$

where one omits the linear term, that can be incorporated into the $\mathbf{x}$ vector as the additional fixed variable.

Solving the presented optimization tasks (I, II, III) with the corresponding constraints on the conventional computer architectures is currently a usual routine with the well-established methods [305, 306] facing only the problems of computational resources. However, the hardware setup tuned to perform GD-like dynamics "knows nothing" about details of a particular task in the form of constraints.

It is necessary to comment about incorporating the constraints in the formulation of task (III) on a particular hardware setup. In contrast to task (II), where the space of the feasible arguments is not limited, and task (I), where the final discrete values can be obtained by applying some sort of the effective force-field (depending on a particular physical setup), pinning the continuous variables to the discrete ones, here one has to change the target coefficients of the tensor sum. In case of the equality constraints, it is possible to add the additional cost function in the form of $C(t)(\boldsymbol{\theta}_i^T\mathbf{x} - \theta_i)^2$ to the cost function and change the coefficients of the initial minimization problem by corresponding values. This term corresponds to the constraint in the form of $\boldsymbol{\theta}_i^T\boldsymbol{x} = \theta_i$, and $C(t)$ is the time depending positive function, which reaches the threshold value parallel to the convergence of the hardware dynamics. The similar term $C(t)(\boldsymbol{\theta}_i^T\mathbf{x} - \theta_1)$ corresponds to the constraint in the form of $\boldsymbol{\theta}_i^T\boldsymbol{x} < \theta_1$ for the minimization task, similar to the regularization terms in the optimization tasks. Here $C(t)$ is slowly growing through the operational regime, until this particular constraint is satisfied. Similar ideas are expressed in more details in the context of the quantum adiabatic optimization devices [307] and repeatedly mentioned in this work.

Another interesting form of the optimization problems is nonlinear programming, which is the most general case of the considered problems. One can take into account a particular case (II), where the target function can be factorized as the sum of the pairwise periodic function cos:

$$\min_{\theta\in[0,2\pi)^N} -\sum_{\Omega} \mathbf{J}_{i_1,i_2}\cos(\theta_{i_1} - \theta_{i_2}), \tag{4.4}$$

this case is also known as the XY model in statistical physics. Further, the relationships between all these models in the context of special-purpose hardware operations will be shown. Additionally, there is no need to specify the type of optimization problem when considering a particular task since it is possible to mention only one formulation. Its connection with others is followed automatically from the presented mappings.

Here, one can find a shallow description of the general methodology of correspondence between different optimization tasks formulations based on the analytical-numerical approximations. Fig. (4-1) represents the main optimization tasks (I, II, III) together with their encoding/decoding relationships on the hardware level. Arrow (1) denotes the introduction of the effective force-field (which is similar to the use of tanh term, see Hopfield NN further, the nature of the related term depends on the particular physical system) to constraint the variable into the discrete domain. (2) can be realized in different ways. The first one is called semidefinite relaxation and used extensively in the computer science domain. One can find several approximation methods with the application to quadratic optimization problems [308], and large-scale applications with mostly image segmentation tasks [309]. A good example involves linear programming relaxation of

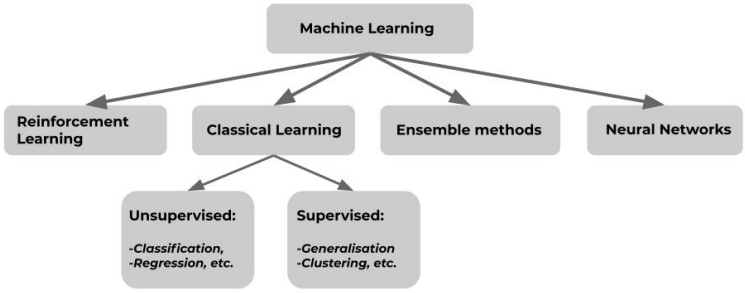


Figure 4-3: The basic scheme of the ML subfields. Part of the presented blocks are covered from the perspective of utilizing these algorithms on the hardware setup.

an integer programming problem, when one removes the integrality constraint, allowing non-integer rational solutions. On the hardware level, one can do the same with the physical sources of the constraint terms in the analogy with the computer analogue. Another method is to encode the feasibility argument space by its discrete analogue with the corresponding coefficients recalculations. (3), (4) denotes the unique embeddings, depending on the function type (II). One can see a simple correspondence, which is for sure not unique and possibly can be improved. Considering a specific type of function, as in Fig. (4-1), for the (3), one can use the analytic Taylor expansion of the cos function on the periodic domain for each pair of variables x, y. Using this relationship, one can recalculate the corresponding coefficients J_{ij}, taking into account the fixed precision of the expansion. (4) can be embeded in the similar manner. Rewriting the quadratic form (III) in the diagonal form, one has to approximate the sum of squares $\widetilde{J_1}x_1^2 + ... + \widetilde{J_n}x_n^2$ form by means of the harmonic cos basis. Using the same Taylor expansion, one can express the x_i^2 through the desired functions and recalculate the corresponding coefficients of the quadratic problem, paying attention to the additional constraints connected with both the initial formulation and embeddings. Higher order terms can be efficiently embedded into lower by introduction of auxiliary variables. (5) can be done through arrows (2) (4).

The case of Eq. (4.1) with the tensor rank 2 and the binary variables is known as quadratic unconstrained binary optimization (QUBO). It can be considered as a combinatorial optimization problem, which has many applications. The essential property of such a problem is its universality, which reflects itself in the numerous connections with other different problems and assignments, among which the most important are SAT in the computer science domain and Ising model in physics. These connections allow one to formulate more embeddings across a variety of tasks. Research on QUBO has generated a wide range of solution techniques for this essential model [310]. Among them are exact methods, such as a MAX-Cut reformulation, branch and bound algorithm, its modifications, Lagrangian decompositions, different heuristic algorithms, such as tabu search, simulated annealing or genetic algorithms.

Most of the presented formulas and equations have a specific notation typical for numerical mathematics since most of the methods and algorithms were designed specifically for conventional computer architecture. Fig. 4-2 presents the general methodology of solving a particular task using the special-purpose hardware.

4.2 Machine learning methods

Here, a brief perspective on the ML field is given before transitioning to the algorithms' list, available for the hardware platform's implementation. One of the main ML field's goal is to predict good outcomes from the given data. The richness of the data dramatically influences the methods' performance, making it easier to find patterns and predict accurate results. There are three crucial components in ML, which are data, features, and algorithms. The general structure can be seen in Fig. 4-3.

Practically speaking, one can meet the data in many places and many ways — E-mails, stock prices time-series, users databases and collection of the experimental measurements. Data can be collected in different ways, either manually, usually quite long and costly with few errors or automatically by feeding everything to some sorting algorithms. Large corporations, for example, can use their users for such kind of activity. The collected data (or datasets) can be of great value depending on the context, determining the demand for suitable rare datasets.

Features represent the properties or characteristics of the considered objects. A small amount of foundation, sorted, and significant features in most cases can guarantee the success of the ML approach to the problem. It is very time-consuming to determine the feature in the so-called 'raw' big datasets and select the right ones. Sometimes one has to avoid human-based decisions to avoid introducing subjectivity and opinion-based bias to optimize the model performance.

The latest DL success is partially tied to automatic feature engineering compared to the previous ML models.

The last part of the considered scheme is the algorithm. Choosing the method of solving a particular task depends on the context and influences such parameters as the final model's accuracy, speed, and computational complexity. In general, one problem can be solved in many different ways.

The components were presented according to their significance in the ML pipeline. Simply saying, one can not extract any useful information from the noisy and meaningless dataset. First of all, the feature extraction will be difficult, if not impossible, not to mention the last part, where even the best algorithms will be useless.

The following subsection starts the discussion with the *classical algorithms*, which are the basis of many existing applications. The reader can find the short but informative structure of the main ML methods that will be the centre of attention with the aim of transferring them into the special-purpose hardware.

4.2.1 Classical machine learning

Regression and least-squares estimation

One of the earliest methods in the area of statistical modelling is regression analysis. The general definition is statistical procedures for establishing the relationships between an output (observation or dependent variable) and one or more arguments (or independent variables). Different predictions and decision making in many applied domains benefit from the regression analysis since it is hard to have analytical expressions for every problem met.

This linear regression model assumes that the dependent variables denoted by y_i has a linear relationship depending on the m-vector of points $\{x_{i1}, \dots, x_{im}\}_{i=1}^{n}$ with an addition of the disturbance terms ϵ_i in each case. This relationship can be written in the following term:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \epsilon_i. \tag{4.5}$$

The short notation involves the matrix form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with the following variables: $\mathbf{y} = y_i, X = x_{ij}, \boldsymbol{\beta} = \beta_j, \boldsymbol{\epsilon} = \epsilon_i (i = 1, \dots, n), (j = 0, \dots, m)$, with $x_{i0} = 1$.

The task for the Eq. (4.5) is to estimate the values of the regression coefficients β_j given the data points x_{ij} and observables y_i, so that the error term $\boldsymbol{\epsilon} = \mathbf{y} - X\boldsymbol{\beta}$ is minimized. One has many procedures to deal with the estimation of the parameters. The most common estimation technique is called the least-squares estimation. In this setting, the optimum parameter is defined through the minimization of the sum of mean squared loss:

$$\min_{\beta_j} \sum_{i=1}^{n} \left(\sum_{j=0}^{m} \beta_j \cdot x_{ij} - y_i \right)^2, \tag{4.6}$$

which can be connected with the quadratic programming Eq. (4.3) for variables, where the coefficients $A_{ij} = \sum_{k=1}^{n} x_{ki} x_{kj}$, and indices i, j goes over $(j = 0, \dots, m+1)$ with $x_{i0} = 1, x_{i,m+1} = -y_i$.

The optimal solution can be obtained by differentiating the Eq. (4.6) and equating it to zero with respect to parameters β_j. Solving the matrix linear relation gives us the expression:

$$\beta = (X^T X)^{-1} X^T Y. \tag{4.7}$$

It is possible to use different modifications of the proposed procedure. It can be generalized least squares, where one introduces a certain degree of correlation between the residuals (4.6), or the weighted least squares, where the knowledge of the variance of observations is incorporated as the coefficients w_k before each of the residual. Such method is affecting the coefficients A_{ij} in a straightforward way through the sum $A_{ij} = \sum_{k=1}^{n} w_k x_{ki} x_{kj}$.

Moreover, intrinsically different techniques can be based on the maximum likelihood estimation, Bayesian methods, or the introduced regularisation. Among the forms of penalized estimations, it is helpful to point out the most general one: the Tikhonov regularization. In case of the same problem as in the Eq. (4.6) with the Euclidean norm, $\|\cdot\|_2$, to endow a particular solution with some wanted properties, one can add the regularization term in the form of $\|\Gamma\beta\|_2$, where Γ is known as Tikhonov matrix, which affects the final coefficients of the optimization problems and shift the optimal parameters β:

$$\beta = (X^T X + \Gamma^T \Gamma)^{-1} X^T Y. \tag{4.8}$$

In many cases, Tikhonov matrix is chosen as proportional to the identity matrix $\Gamma = \alpha I$, which penalizes the large values of β parameters. Other forms of the regularization include generalized Tikhonov regularization with the corresponding shift of the target parameters $\beta - \beta_\Delta$, Ridge regression, Lasso (where one introduces the constraints in the form of $\sum_{j=1}^{m} |\beta_j| < t$ into the Eq. (4.6)), etc.

The natural extension of the linear regression involves the polynomial basis. In case of one argument, it is possible to rewrite the familiar relationship Eq. (4.5) in the following term:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \epsilon_i. \tag{4.9}$$

In this case, one has the same notation $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with the following variables: $\mathbf{y} = y_i, X = x_{ij} = x_i^j, \boldsymbol{\beta} = \beta_j, \boldsymbol{\epsilon} = \epsilon_i (i = 1, \ldots, n), (j = 0, \ldots, m)$, with $x_{i0} = 1$. Given the data points, the same task of Eq. (4.6) will be used except the variables change $x_{ij} = x_i^j$. From the perspectives of regression parameters estimation, the task did not change. The only change involves the preprocessing of the QUBO coefficients. The first extension from the linear regression to the polynomial can be described as the change of the following coefficients basis $[1(= x_{i0}), x_{i1}, x_{i2}, ..., x_{im}] \rightarrow \left[1, x_i, x_i^2, \ldots, x_i^m\right]$. It is possible to extend this basis to a set of some nonlinear functions $f(x_i)_j$, not only polynomial degrees: $[1, x_i, x_i^2, \ldots, x_i^m] \rightarrow [1, x_i, f(x_i)_1, \ldots, f(x_i)_{m-1}]$.

Multiple linear regression is a generalization of linear regression to the case of more than one independent variable. The basic model for multiple linear regression can be written in a similar form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_m X_{im} + \epsilon_i, \tag{4.10}$$

where instead of one variables x_{ij} one have a set in the form of matrix X_{ij}. Depending on the chosen norm for the matrix, it is possible to formulate the task of finding the regression coefficients. Taking the square Frobenius norm of the matrix, the search is equivalent to the already mentioned task Eq.(4.6), except with the additional sum over the matrix elements (k^2):

$$\min_{\beta_j} \sum_{l=1}^{k^2} \sum_{i=1}^{n} \left(\sum_{j=0}^{m} \beta_j \cdot x_{ijl} - y_{il} \right)^2, \tag{4.11}$$

where the additional index l denotes the summation across all of the matrix elements.

This form can be extended further for multivariate linear regression or combined with the nonlinear basis with minor consequences concerning the parameters search and hardware operations, except for the much-complicated procedure of the coefficients preprocessing. Thus, the regression can be treated as the simplest form of supervised learning.

Classification

Classification is the most popular task in the ML domain. The purpose of classification is to sort the given objects among the initially defined classes. Labelled data with features is always needed for classification. There are enormous amounts of objects that can be classified. The earliest algorithms include naive Bayes and decision trees. There is no need to consider them in great detail and refer the reader to the Markov random field (MRF) encoding, which is the general case for such models and will be considered below.

The k-nearest neighbors algorithm (kNN) is a non-parametric classification method used in statistics [311, 312]. It aims to classify the objects by considering their given neighbours with the defined class, depending on the k nearest neighbours. The typical choice for the distance is the Euclidean measure. The process of consequent assignments is repeated until the convergence. There is no need to explicitly present the corresponding formulas since they are very similar to the k-means, the clusterization algorithm, presented below. The difference is that the k-NN is supervised learning, while k-means generally is not. Both are usually based on the Euclidean distances' calculations, making them a potential target for transferring for the special-purpose hardware, aiming to solve certain optimization tasks, presented at the beginning of this chapter.

support vector machine (SVM) is a supervised learning model that analyzes data for classification purposes. It aims to construct a hyperplane between the classes of training data points in a high-dimensional space, emphasizing a good separation achieved by maximizing its margin. SVM was introduced in [313] and standardized in [314]. One can treat SVM as the probabilistic binary linear classifier.

Linear SVM deals with the n points $\mathbf{x_i}$ in the m-dimensional space, where each point has an assigned a binary class $y_i = \pm 1$. The task is to construct a hyperplane that divides these two groups with the maximized distance between the nearest point(s) and the hyperplane.

The so-called "hard margin" scenario assumes that the initial data is linearly separable. Thus, one can start with constructing two parallel hyperplanes, which have the largest distance between each other and at the same time, each of the surfaces contains at least one point from different classes correspondingly. The target surface between these hyperplanes is called the maximum margin hyperplane. To mathematically describe these surfaces, one can write:

$$w_i x_i - b = \pm 1, \tag{4.12}$$

where w_i is components of the normal vector for both of the hyperplanes, x_i are m-dimensional coordinates, b defines the surface shift concerning the zero coordinates and ± 1 defines the class. Everything above the $y = 1$ is of one class, and everything below $y = -1$ is of another. The offset of the hyperplane is determined through the $b/\|\mathbf{w}\|$, while the marginal distance equals $2/\|\mathbf{w}\|$. To maximize the last one has to minimize the denominator $\|\mathbf{w}\|$.

This task can be reformulated as the optimization problem, adding the constraints that prevents data points from falling into the margin:

$$\begin{aligned} &\min \|\mathbf{w}\| \\ &\text{s.t. } y_i \left(\mathbf{w}^T \mathbf{x}_i - b\right) \geq 1 \text{ for } i = 1, ..., n \end{aligned} \tag{4.13}$$

The natural extension for the SVM is considering a so-called "soft margin" case. It is assumed that the given data points are not linearly separable. To accommodate the method to this scenario, one has to introduce a new kind of variables $\xi_i = \max(0, 1 - y_i \left(\mathbf{w}^T \mathbf{x}_i - b\right))$, which are usually referred to as the hinge loss functions. These variables have regularizers roles when dealing with real-world data. Thus it is possible to rewrite the Eq. (4.13) problem in the following way:

$$\begin{aligned} &\min \frac{1}{n} \sum_{i=1}^{n} \xi_i + C\|\mathbf{w}\|^2 \\ &\text{s.t.} y_i \left(\mathbf{w}^T \mathbf{x}_i - b\right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, \text{ for all } i, \end{aligned} \tag{4.14}$$

where the constant C regulates the interplay between the pure hard margin classifier and soft margin one. To properly address the new kind of variables and reformulate the problem to the initial scope of this chapter, exploiting the well-known Lagrange duality:

$$\begin{aligned} &\max f(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i a_i \left(\mathbf{x}_i^T \mathbf{x}_j\right) y_j a_j \\ &\text{s.t. } \sum_{i=1}^{n} a_i y_i = 0, \text{ and } 0 \leq a_i \leq \frac{1}{2nC} \text{ for all } i, \end{aligned} \tag{4.15}$$

where the $\mathbf{w}$ is expressed through the new variables a_i, so that $\mathbf{w} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{a_i y_i x_i}$, and the initial task of determining the offset of the surface is expressed via $\mathbf{b} = \mathbf{w^T x_i} - \mathbf{y_i}$. Thus, it is possible to obtain the problem, which is exactly (4.3) with $A_{ij} = y_i \left(\mathbf{x}_i^T \mathbf{x}_j\right) y_j$ with the linear field of constant $\mathbf{1}$ and B determined by constraints. This problem can be solved with the standard quadratic algorithms, and is prepared to be solved on the special-purpose hardware.

It is useful to mention the nonlinear extension of the SVM, which solves nonlinear classification and can exploit the different functional forms of kernels to deal with the problem. Depending on the particular hardware and its possible function realisation, one can modify the scalar dot product in the quadratic form in the Eq. (4.15) formulation by a certain kernel function $k(\mathbf{x_i}, \mathbf{x_j})$ to extend the initial functionality.

Let us transition to the unsupervised learning models, where the algorithms usually lack any direction in labels and target output.

Dimensionality reduction

Dimensionality reduction involves the transformation of data from the space with many dimensions into a low-dimensional space, usually preserving meaningful and valuable properties from the original data. High-dimensional data is hardly handled in practice due to the curse of dimensionality. Additionally, the practical significance lies in combining several features and working with these new feature types, also referred to as abstractions.

Dimensionality reduction is standard in data-intensive fields. It can be used in signal processing, neuroinformatics, and bioinformatics [315, 316]. One can find its applications in recommender systems [317], semantic search [318] or as a primary tool in many domains involving numerical analysis.

One of the famous methods for dimensionality reduction is the principal component analysis (PCA). It was invented in 1901 by Karl Pearson [319]. The idea behind PCA is to approximate a particular data with linear manifolds of lower dimension. PCA can be alternatively interpreted as finding subspaces of lower dimension in the orthogonal projection on which the data variation is maximum.

The initial task behind the PCA is to find the best approximation of the data points by means of lines and surfaces. Given the set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$, the aim is at finding the sequence of k k-dimensional affine spaces $L_k \subset \mathbb{R}^n$ which solve the task of:

$$\min \sum_{i=1}^{m} \mathrm{d}^2(\mathbf{x}_i, L_k) = \min \sum_{i=1}^{m} \sum_{l=1}^{n} \left(x_{il} - a_{0l} - \sum_{j=1}^{k} a_{jl} \sum_{q=1}^{n} a_{jq} \left(x_{iq} - a_{0q}\right) \right)^2, \tag{4.16}$$

for each k among L_k, where $\mathrm{d}(\mathbf{x}_i, L_k)$ is the Euclidean distance from the point $\mathbf{x}_i$ to the L_k. Affine spaces L_k are defined

as the set of linear combinations $L_k = \{\mathbf{a}_0 + \alpha_1 \mathbf{a}_1 + \cdots + \alpha_k \mathbf{a}_k\}$ with $\alpha_i \in \mathbb{R}$, while the set of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\} \subset \mathbb{R}^n$ is orthonormal.

The presented task is solved with the sequence of the same type optimization problems. The starting vector $\mathbf{a}_0$ is simply defined as

$$\mathbf{a}_0 = \underset{\mathbf{a}_0 \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{m} \mathrm{d}^2 (\mathbf{x}_i, L_0) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i. \tag{4.17}$$

The iteration loop consists of the subtraction of the resulting projection $\mathbf{x}_i := \mathbf{x}_i - \mathbf{a}_0 (\mathbf{a}_0, \mathbf{x}_i)$ (with the scalar product $(\mathbf{a}_0, \mathbf{x}_i)$) for each vector and calculating the next component. For the vectors defining the L_j:

$$\mathbf{a}_j = \underset{\|\mathbf{a}_j\|=1}{\operatorname{argmin}} \left(\sum_{i=1}^{m} \left(\mathbf{x}_i - \mathbf{a}_j (\mathbf{a}_j, \mathbf{x}_i) \right)^2 \right), \tag{4.18}$$

and following subtraction $\mathbf{x}_i := \mathbf{x}_i - \mathbf{a}_j (\mathbf{a}_j, \mathbf{x}_i)$, until the parameter k reaches the $n-1$ of the initial problem space dimension. Using the property $||\mathbf{x}_i - \mathbf{a}_j (\mathbf{a}_j, \mathbf{x}_i)||^2 = ||\mathbf{x}_i||^2 - (\mathbf{a}_j, \mathbf{x}_i)^2$ one can easily map this task into the Eq. (4.3) with the normalization constraints, where $A_{ij} = -x_i x_j$. To shorten the presented notation, the iterative procedure can be written in the following similar form with the maximization tasks:

$$\hat{\mathbf{X}}_k = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X} \mathbf{w}_{(s)} \mathbf{w}_{(s)}^{\mathrm{T}}, \tag{4.19}$$

$$\mathbf{w}_{(k)} = \underset{\|\mathbf{w}\|=1}{\arg\max} \left\{ \left\| \hat{\mathbf{X}}_k \mathbf{w} \right\|^2 \right\}, \tag{4.20}$$

where k is the number of principal component, $\mathbf{X}$ is the data matrix of size $n \times m$, w_s are the weight coefficients, similar to the $\mathbf{a_j}$ components. The maximization form is actually a Rayleigh quotient, because

$$\mathbf{w}_{(k)} = \underset{\|\mathbf{w}\|=1}{\arg\max} \left\{ \left\| \hat{\mathbf{X}}_k \mathbf{w} \right\|^2 \right\} = \arg\max \left\{ \frac{\mathbf{w}^T \hat{\mathbf{X}}_k^T \hat{\mathbf{X}}_k \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \right\}, \tag{4.21}$$

and the quotient's maximum possible value is the largest eigenvalue of the matrix $\hat{\mathbf{X}}_k^T \hat{\mathbf{X}}_k$. In case, that the sequential operation is limited on the specific hardware system, one can still use the first iteration of the PCA method to obtain the largest eigenvalues of a matrix. This is useful in some applications, like the power method (that refers to the early work [320]) and its applications, among which there is famous PageRank algorithm [321].

There are alternative formulations of the PCA task, such as covariance matrix diagonalization or singular value decomposition (SVD). SVD is a special form of a rectangular matrix decomposition in the form:

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\top}, \tag{4.22}$$

where $\mathbf{U}$ is the unitary matrix (representing the rotation as the linear transformation of the space in the geometrical interpretation), $\mathbf{\Sigma}$ is the rectangular diagonal matrix with non-negative real numbers on the diagonal (which are called the singular values, the action of the matrix has the interpretation of the corresponding scaling by diagonal elements) and $\mathbf{V}^{\top}$ is another unitary matrix (with the same additional rotation interpretation).

The correspondence between the previously mentioned PCA task and SVD decomposition is quite apparent. The calculation of the corresponding coefficients for the optimization task is straightforward. It was shown that to perform the PCA, one has to find the eigenvectors of the covariance matrix $\mathbf{X}\mathbf{X}^{\top}$ (neglecting the $\frac{1}{n-1}$ scaling factor depending on the number of points) with $\mathbf{X}$ being the data matrix. The covariance matrix is diagonalizable, and with the normalized eigenvectors, one can write:

$$\mathbf{X}\mathbf{X}^{\top} = \mathbf{W}\mathbf{D}\mathbf{W}^{\top} \tag{4.23}$$

Applying SVD to the same data matrix $\mathbf{X}$ gives:

$$\mathbf{X}\mathbf{X}^{\top} = \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top} \right) \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top} \right)^{\top} = \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top} \right) \left(\mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\top} \right), \tag{4.24}$$

which lead us to the equality:

$$\mathbf{W}\mathbf{D}\mathbf{W}^{\top} = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^{\top}, \tag{4.25}$$

where one can see the clear correspondence. Using this information, one can perform the SVD decomposition similarly to PCA on the special-purpose hardware.

Clusterization

The most detailed description of clusterization is the separation of the objects on an unknown basis. The goal can be alternatively defined as a classification, but no previously known classes have a significant difference. The number of clusters can be set in advance or defined automatically by the machine. The algorithm itself determines objects' similarity by the features that one marked and put the objects with many similar characteristics in the same class. There are currently a lot of successful applications of clusterization. Among the fields are market analysis (consumer analytics), image compressing, data analytics, and anomaly detection.

K-means clustering is a clustering method that aims to partition n observations into k clusters. Each of these observations is located in the cluster with the nearest mean, also called a centroid [322, 323, 324]. There are heuristic algorithms that deal with such an assignment; however, the initial problem is NP-hard.

Given a set of observations $\{x_1, ..., x_n\}$ in a d-dimensional space k-means algorithm aims to partition these observations into k sets $\{S_1, S_2, ..., S_k\}$ to minimize the within-cluster sum of squares (or variance):

$$\arg\min_{S_i} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_{S_i} \right\|^2, \tag{4.26}$$

where $\boldsymbol{\mu}_{S_i}$ is the mean of points in S_i.

To perform such optimization task, one usually uses an iterative technique, consisting of two steps. Starting with an initial set of k means $m_{1,(1)}, ..., m_{k,(1)}$, the first step is to assign each observation to the cluster with the nearest mean, according to the Euclidean distance: $S_{i,(t)} = \left\{ x_p : \left\| x_p - m_{i,(t)} \right\|^2 \leq \left\| x_p - m_{j,(t)} \right\|^2 \forall j, 1 \leq j \leq k \right\}$. Afterwards, one has to recalculate the centroids: $m_{i,(t+1)} = \sum_{x_j \in S_{i,(t)}} x_j$. The loop is run until the convergence.

The algorithm uses the assigning of objects to the nearest cluster by Euclidean distance, and it is a suitable method for transferring its sequential operations to the specific hardware.

Mean shift is a high-dimensional-space analysis method for locating the maximum density function given a discrete number of data sampled from this arbitrary density function. It is helpful in complex hierarchical algorithms and is used in different computer vision or image processing domains.

Given data points x_i in n-dimensional space, one can use the kernel function $k(r)$, acting on the norm value r, to determine the mean shift's value. The kernel function has to be non-negative, non-increasing and continuous. One can use the flat kernel, so that $k(r) = 1$ if $r < r_0$ and 0 outside. Each iteration is composed of calculating the function:

$$F(x) = \sum_i k\left(\frac{||x - x_i||^2}{\alpha^2} \right), \tag{4.27}$$

where α states for the scaling factor, and computing the maximum of $F(x)$. One can easily see the relationship between the Eq. (4.27) and Eq. (4.3) without the constraints.

4.2.2 Neural networks

It is common to speak about NNs starting from the biological perspective, which served as the inspiration for the so-called artificial NNs. Considering this chapter's multidisciplinary character, the focus will be devoted to the NN properties from the universal mathematical perspectives while avoiding analogies with the human brain.

Any NN can be defined as a set of neurons and connections between them. An artificial neuron's task is to take input numbers, process them in a certain way (executing a special function), and output the results. The common mathematical transformation of one NN layer can be written in the following way $\tanh(\sum_{i=0}^{N} w_i x_i + b)$, where w_i denote the weights for the input data points x_i (or independent variables) and the constant b is the shift called bias. The tanh is used as nonlinear activation function. A single layer NN that performs a similar transformation is called a perceptron, and it gives a single output. The perceptrons, assembled into multilayered structures of such units, are called multilayer perceptrons.

The reader can search the specific books for the introduction to the NNs [325, 326, 327] with more modern work [328] and the latest results after the DL breakthrough [50].

The activation function plays an important role in the NN design because the output signal would be simply a linear function in its absence. Many functional forms of activation functions, such as binary step function, sigmoid (or logistic function), hyperbolic tangent and many others. They allow a NN to map an input to the output appropriately. Thus, NN is considered a universal function approximators [329]. From the geometrical perspective, the weights of $\tanh(\sum_{i=0}^{N} w_i x_i + b)$ functions can be manually tuned to fit any set of points.

To choose the NN weights, one usually uses the backpropagation procedure [330, 331]. It consists of tuning the NN

weights according to the difference between the actual output value of the network and the predicted one, with the final goal of minimizing this discrepancy or cost function. The tuning procedure involves computing the total discrepancy gradients on each layer starting from the final one and the corresponding weight values updates. Through the extensive number of such loops, there is a chance that the weights will be corrected in the desired way.

A well-trained NN can even approximate most of the algorithms presented here, given the appropriate parameters. Any deep NN (with many layers) can be converted into a shallow recurrent one (Hopfield NN) with a significant overhead on the number of neurons in the standard layer. That means that any deep NN functionality can be performed on the device, adjusted to solving the QUBO. The only problem is to find the correct coefficients for the task.

One can match the shallow network with the simple energy model. In general, there is a conceptual approach to many ML tasks, which is called the energy-based learning [332]. This apparent correspondence will be demonstrated on the crucial model of Hopfield NNs.

Hopfield neural networks and quadratic optimization

A Hopfield network is a recurrent artificial NN considered in [333]; however one can observe earlier works based on similar Ising models. The essential elements of Hopfield NN are binary threshold nodes, which converge to a local minimum. They can store and recall multiple memories, making the Hopfield NN an attractive model for associative memory. Nevertheless, the binary nodes can sometimes converge to a wrong pattern. The balance between different patterns and the memory units' dynamical behaviour is an active investigation area with the modern research context related to artificial intelligence and cognitive sciences. There will be many more essential connections at the end of this subsection, some of which lie on the surface. There are typically two well-known formal forms of the model. The first one is discrete with respect to the time and neuron state variables and describes the state of neurons in the following way:

$$S_i(t+1) = \begin{cases} 1, & \text{if } \Sigma_j J_{ij} S_j(t) + I_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.28}$$

where S_i denotes the state of the i-th neuron, J_{ij} is the coupling strength coefficients for the influence of the neuron j to the neuron i and I_i is a direct input or bias coefficient. The structure of Eq. (4.28) is similar to the one, that have just been discussed in the beginning of the NN description $\tanh(\sum_{i=0}^{N} w_i x_i + b)$.

The second model has the form:

$$dx_i/dt = -x_i/\tau + \Sigma_j J_{ij} g\left(x_j\right) + I_i, \tag{4.29}$$

where x_i denotes the mean state of the i-th neuron that can get continuous values in the initially defined range, J_{ij} is the same coupling strength coefficients for the influence of the neuron j to the neuron i, I_i is a direct input or bias coefficient, g is a monotone function that converts continuous state into the discrete, i.e. makes the correspondence between the variables $S_i = g(x_j)$, and τ is the characteristic time of the differential equation (4.29) for the convergence to an optimal or suboptimal solution.

The analogue computation with the NN can be described as an evolution of the state space in continuous variables. One can precisely trace it following the Eq. (4.29). The vital aspect of such differential equation structure is the existence of a Lyapunov function, which is vital to stability theory of dynamical systems and control theory [334]. This Lyapunov function (that can also be referred to as an energy function) H behind the Hopfield NN can lead to the understanding of possible final states, which appear to be attractors considering the asymptotic behaviour the system. Hopfield networks are capable of universal computation in the Turing sense [335]. The explicit formula for the Lyapunov function in the discrete variant of the model:

$$H = -\frac{1}{2}\Sigma_{ij} J_{ij} S_i S_j - \Sigma_i I_i S_i, \tag{4.30}$$

which is equivalent to the initially considered problem of tensor sum minimization (4.1) with the rank 2 tensor and variables S_i being x_i, or equivalently the famous Ising model, and reduction of the non-zero field I_i by the introduction of the additional variable $S_{N+1} = 1$. In case of continuous variable, the same function has slightly different form:

$$H = -\frac{1}{2}\Sigma_{ij} J_{ij} S_i S_j - \Sigma_i I_i S_i + \frac{1}{\tau}\Sigma_i \int^{S_i} g^{-1}(Z) dZ, \tag{4.31}$$

where the last term appears due to the correspondence between the discrete and continuous state $S_i = g(x_i)$. For the g one usually pick the $\tanh(x/\beta)$ function, where the β parameter tends to zero value during the evolution of the Hopfield NN making the last term of the Eq. (4.31) to converge to the same zero value, see [25] with the additional emphasis on the optimization problems.

The associative memory was addressed by the theory of Hebbian learning [336, 337]. The basic concepts can be explained through the connection between Hopfield NN's coupling coefficients and the memory patterns. One of the

common formulaic descriptions can be presented in the following way:

$$J_{ij} = \frac{1}{p}\sum_{k=1}^{p} s_i^k s_j^k, \tag{4.32}$$

where J_{ij} is the strength of the connection between i and j neurons, p is the number of training patterns and x_i^k is the pattern input.

As mentioned before, the Hopfield model is isomorphic to the Ising model of magnetism (for zero temperature) [338], which have been extensively analyzed by the physicist community in great details. One can instantly tailor the Hopfield model with another example of the disorder systems: the Sherrington-Kirkpatrick model [339], with the difference between the coupling coefficients.

It appears that there are much more fruitful connections of the Hopfield NN with many other essential problems besides the disordered systems. The most important connections will be shortly covered, which allow one to see the correspondence between the Hopfield NN (or its equivalent formulations) with almost any possible assignment or problem of particular interest.

Finding the ground state of the Ising model is NP-hard and is the particular case of the initially defined problem of PP (4.1) with the tensor rank 2, which in some cases can be related to the QP. Another significant task is the Boolean satisfiability problem (SAT) [340], and its a more narrow analogue which is weighted MAX-2-SAT. SAT is a universal problem because many computational assignments can be reduced to some instance of the SAT problem.

The Hopfield NN can be reduced to more deep NN and reverse – every deep feedforward NN can be reduced to a shallow reecurrent one with different parameters. It is also possible to pick the appropriate parameters for choosing the deep NN and get a shallow network. Both networks will have the same approximation qualities, given the proper parameters. One can find the topic of the correspondence between deep and shallow NN being investigated in [341, 342, 343, 344]. A similar analogy can be applied to the initial problem defined by the Eq. (4.1). It is possible to reduce the tensor sum into the quadratic polynomial by introducing the auxiliary variables. The good examples of such substitutions are presented in [345]. A similar procedure can be applied to convert quadratic optimization back into the tensor-sum by reducing variables.

Recalling the Hopfield NN, one can view it from a different perspective. Changing the sign before the Lyapunov function will lead to the maximization task. It can be considered as an anti-Hopfield model. Using the Hebbian learning rule, which connects the coupling coefficients' values with the stored patterns, the anti-Hopfield model will penalize the stored patterns, which can be considered the unlearning procedure. This effect is highlighted in the context of the planted solution [346].

One of the types of Hopfield NN is called a Boltzmann machine. It is a type of stochastic recurrent NN, unlike the Hopfield NN, but with the same energy form given by Eq. (4.30). To train the Boltzmann machine, one must prepare the visible units according to the input data (in contrast to the hidden, which do not receive any information from the external sources). The next step is to run the network by choosing a unit, resetting its state, and repeating it until reaching the so-called thermal equilibrium at a particular parameter of the system T. The final state depends only on the global state's energy. One can find a review covering the main areas of use for such specific NN architectures [347], where the emphasis is made on the inverse statistical problems. The stochasticity of the Boltzmann machines allows one to trace additional significant correspondence of the Ising-like models with a special type of Markov random field (MRF) [348, 349], a class of probabilistic graphical model (PGM). Despite the theoretical interest, the Boltzmann machine is rarely used in large practical applications, even considering some modifications, like restricted Boltzmann machines [350, 351] or deep Boltzmann machines [352].

Modern Hopfield networks

Recently a new type of NN was introduced, which is called a modern Hopfield network [353]. It is the generalization of the conventional Hopfield NN with continuous states and a corresponding update rule (equivalent to the attention mechanism used in the modern transformer architectures). This type of NN can store exponentially many patterns growing with the number of variables. Moreover, it retrieves the pattern using only one update while having a small retrieval error. Such architectures can be incorporated into other, larger architectures and be applicable across various domains. For example, Hopfield layers showed greater performance on different state-of-the-art benchmark problems across various domains, including drug design datasets, compared to different ML methods.

The central part of the results from the original article [353] can be suitable for the original model of Hopfield NN with several appropriate adjustments. Moreover, since the last model can be realized on several types of hardware (including the EP platform, the topic of the current thesis), they share the same advantages of this specific architecture.

One can successfully integrate the new modern Hopfield network into DL architectures to access such functionality as storage and access specific input data or learned prototypes. Moreover, such Hopfield layers provide attention mechanisms, pooling, additional memory and other useful functions. One can find many more groups of methods, which can be realized through modern Hopfield NNs. Among them are methods mentioned in this chapter, such as SVM, logistic and multinomial regression, PCA, boosting and bagging of NNs or others, like decision trees, random forest, Bayesian methods [353]. For other connections see [354, 355].

The storage of the exponentially many patterns can be realized even for the case of binary variables [356, 357] using the extension of the original quadratic form into high-order terms. The possibility of tensor terms realization in the dynamics of the condensate degrees of freedom was previously mentioned in Chapter 3.15. This makes the discussed physical platform a prospective candidate for the realization of the functionality mentioned above. However, the topology of the possible condensate connections can be pretty sparse due to the two-dimensional physical setup or the coefficient correlations. The same concerns the topology for the possible tensor terms, which should provide good scaling for the optimization problems and memory capacity for the NN realization and should be investigated further.

4.2.3 Ensemble methods

The NN architectures relations to each other in the context of their reproduction on a particular hardware system was analyzed. In many cases, the standard approach to the NN training is to use the predefined architecture, and the backpropagation mechanism to find the suitable weights of such NN [358, 50]. There are also exciting methods of defining NN's structure and corresponding NN weights as constructive learning [359, 360, 361] that alters the network structure as learning proceeds or destructive learning, such as optimal brain damage [362, 363, 364]. However, they may not be enough or less likely to be used in practical applications. Not taking into account the current research topic about the neural architecture search [365], the focus is on more efficient ensemble methods relevant to this chapter's main topic.

Ensemble methods have an elementary underlying idea. It consists of taking several not very effective methods or NNs and combining them to correct each other's mistakes to enhance the overall quality of the resulting system [366, 367, 368, 369]. While being actively used in production, these methods give very high accuracy. The instability of the fundamental architectures or considerable variation in the outcome only improves ensemble methods' overall efficiency. Moreover, the assembly process allows exploring an immense combinatorial variability. Several critical methods were mentioned in the context of transferring it on general hardware capable of the optimization tasks defined in the beginning.

The idea of stacking [370, 371] is to train several different NNs or different algorithms on the same data and pass their results to the input of the governing one, which makes the final decision. Bagging (or Bootstrap AGGregatING) [372, 373] is an ensemble meta-algorithm designed to improve the accuracy of the basic algorithm. It can be thought of as the model averaging techniques since the idea is to use several different samples for the set of tunable algorithms' learning process. At the same time, the final answer is obtained through the averaging of the outcomes. The intrinsic capability to parallel this meta-algorithm makes it quite attractive from the engineering perspective. To utilize the Boosting [373, 374, 375] is to tune (or train) algorithms in a consequent manner so that each subsequent step is correcting the previous one. The samples for a particular step are biased concerning the important or problem sets of points. The famous variant of this meta-algorithm is called gradient boosting.

The ensemble methods are usually used in the context of working with the NNs. The vital part of this subsection is that each presented meta-algorithm can be performed on the specialized hardware, where it is set up to perform a particular task. It is possible to utilize the experimental set up sequentially (see, for example, the PCA). In the same way, one can use the hardware with the boosting or bagging methods.

4.2.4 Image processing

Several problems arising in the computer vision domain can be formulated as binary quadratic programs, which is the particular case of our target problem (4.1). The assignment which is of practical interest usually can be identified as large-scale problems. The technique called the semidefinite relaxation appeared to be very efficient [309] and was discussed at the beginning of this chapter in dealing with different types of constraints in the problem formulation.

The list of discussed problems in the later work itself includes image co-segmentation (4.33), image segmentation with different constraints and the tasks that can be related to the target problem with the significant difference in the constraints, graph matching (4.34), image deconvolution (4.35), graph bisection (4.36), and many similar related to these ones. The computational complexity of these problems is high, which leads the authors to propose an improved version of the semidefinite programming approach which is more efficient and scalable. Some of these formulations (among which several are quite obvious and will be discussed further in the Section (4.3) are listed with little corresponding details and

refer the reader to the original work:

$$\begin{array}{ll} \min_{\mathbf{x}\in\{-1,+1\}^N} & \mathbf{x}^\top \mathbf{A}\mathbf{x} \\ \text{s.t.} & \left(\mathbf{x}^\top \mathbf{t}_i\right)^2 \leq \kappa^2 n_i^2, i = 1, \ldots, s, \end{array} \tag{4.33}$$

is the image co-segmentation task with the matrix $\mathbf{A}$ that can be found in [376], s is the number of images, n_i is the number of pixels for i-th image, and $n = \sum_{i=1}^{s} n_i.\mathbf{t}_i \in \{0,1\}^n$ is the indicator vector for the i -th image. $\kappa \in (0,1]$.

$$\begin{array}{ll} \min_{\mathbf{x}\in\{0,1\}} K_L & \mathbf{h}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{H}\mathbf{x} \\ \text{s.t.} & \sum_{j=1}^{L} \mathbf{x}_{(i-1)L+j} = 1, i = 1, \ldots, K \\ & \sum_{i=1}^{K} \mathbf{x}_{(i-1)L+j} \leq 1, j = 1, \ldots, L \end{array} \tag{4.34}$$

is graph matching and $x_{(i-1)L+j} = 1$ if the i -th source point is matched to the j -th target point; otherwise it equals to $0.h_{(i-1)L+j}$ records the local feature similarity between source point i and target point j; $H_{(i-1)L+j,(k-1)L+l} = \exp\left(-\left(\mathrm{d}_{ij} - \mathrm{d}_{kl}\right)^2/\sigma^2\right)$ encodes the structural consistency of source point i, j and target point k, l. The corresponding details can be found in [377]

$$\min_{\mathbf{x}\in\{0,1\}^n} \|\mathbf{q} - \mathbf{K}\mathbf{x}\|_2^2 + \mathrm{S}(\mathbf{x}) \tag{4.35}$$

is the image deconvolution task, where $\mathbf{K}$ is the convolution matrix corresponding to the blurring kernel $\mathbf{k}$, S denotes the smoothness cost, $\mathbf{x}$ and $\mathbf{q}$ represent the input image and the blurred image respectively [309].

$$\begin{array}{ll} \min_{\mathbf{x}\in\{-1,1\}^n} & -\mathbf{x}^\top \mathbf{W}\mathbf{x} \\ \text{s.t.} & \mathbf{x}^\top \mathbf{1} = 0 \end{array} \tag{4.36}$$

is graph bisection (many analogous formulations will be discussed further) with $W_{ij} = \begin{cases} \exp\left(-\mathrm{d}_{ij}^2/\sigma^2\right) & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$ where d_{ij} denotes the Euclidean distance between i and j. All of these tasks are obviously easily mapped into the initial optimization problem defined by Eq. (4.1).

4.2.5 Probabilistic graphical models

A probabilistic graphical model (PGM) is a model in the form of a graph, where nodes represent random variables (events, conditions, etc.). At the same time, the arcs express the conditional dependence structure between these nodes. Graph structure brings with itself an important property of modularity, which allows one to build complex models from the simpler parts. Such a compositional approach and the probability theory provide a very appealing framework for describing various complex models. Many fields such as statistics, information theory, systems engineering, and statistical mechanics deal with special cases of general graphical model formalism. The PGMs provide a general view of all of these models as instances of a common formalism. Mixture models, hidden Markov models or frequently discussed Ising models are good examples of a broad and general framework. One can find good introduction to the topic of PGM in [378, 379] and comprehensive description with many rich details in [380, 381].

PGMs can provide a compact representation of joint probability distributions. Given N binary random variables, one needs 2^N parameters to describe the joint probability distribution $p(x_1, ..., x_N)$, while at the same time, a PGM may need much fewer (in some cases - exponentially fewer), depending on the connections between nodes and the model assumptions. The possible factorization of the joint probability distribution, due to the information about the model, can help effectively work with inference tasks [379].

Graphical models can be divided into undirected and directed. Undirected graphical model, also called Markov random field (MRF), is standard physics and computer vision model. Directed graphical model, or Bayesian network (BN), is popular in the ML community. The main difference between these models is that BNs are directed and acyclic, while MRFs are undirected and cyclic. The consequence of this difference is the limited ability to represent specific dependencies for each type. BNs can not represent cyclic dependencies; in contrast, MRFs can not reproduce induced dependencies.

A good example of MRF is the widely used Ising model. It is possible to obtain the corresponding state of such a model by projecting the possible state of the XY model. The last can be reproduced by a variety of condensed-matter platforms with state of the art characteristics such as superconducting Josephson junction arrays [382, 383], EP condensates [68, 188] or ultracold bosonic quantum gases in optical lattices [227, 384]. These connections emphasize the hardware perspective on the MRFs realization.

Similarly, one can realize BNs in the same setup by making the coupling coefficients asymmetrical. For example, one

can easily redirect light in the hardware optical setting using SLM to connect the elements. It is also possible to have a hybrid model with both directed and undirected connections, usually referred to as a chain graph.

The joint distribution of a MRF can be expressed in the following form:

$$p(x) = \frac{1}{Z} \prod_{K \in \mathcal{K}} \psi_K \left(X_K \right), \tag{4.37}$$

where $\mathcal{K}$ is the set of maximal cliques in the graph, $\psi_K \left(x_K \right)$ is a positive, real-valued potential function defined on the clique set of variables $X_K = \{x_i\}_{i=1}^{n}, i \in K$, and Z is the normalization factor (accounts for all possible realisation of X_K, same as in the statistical physics expression):

$$Z = \sum_{X} \prod_{K \in \mathcal{K}} \psi_K \left(X_K \right) \tag{4.38}$$

Eq. (4.37) has the form of a factor graph, which is a bipartite graph representing the factorization of a joint probability distribution function. It can be easily seen that the Ising model (4.30) is a special case of the Eq. (4.37) ($x_i \sim S_i$) with the pair potentials $\psi \left(X_K \right) = \psi \left(x_i, x_j \right) = e^{J_{ij} x_i x_j}$ for nodes i, j and individual potential functions $\psi(x_i) = e^{I_i x_i}$ for non-zero fields.

One can write similar expression for the directed graphical model:

$$\widetilde{p(x)} = \prod_{j=1}^{d} \widetilde{p} \left(x_j \mid \pi_{x_j} \right), \tag{4.39}$$

with π_{x_j} is the set of parent nodes of x_j, the index $j = 1, .., d$ represents the observable variables on the directed acyclic graph G be a directed acyclic graph with vertices $V = (x_1, \ldots, x_d)$ and the ˜ tilde denotes the unnormalized version of the probability.

Working with the PGM, we are usually interested in certain quantities, such as:

1) Calculating the normalization factor (or statistical sum), given by the Eq. (4.38).

2) Computing the marginal distribution over node or a given subset of nodes in the model (and the normalised marginal distribution):

$$\widetilde{p_i(x_i)} = \sum_{k \neq i} p(\widetilde{\{x_k\}_{k=1}^{n-1}}) = \sum_{x_k, k \neq i} \prod_{K \in \mathcal{K}} \psi_K \left(\{x_k\} \right), \tag{4.40}$$

where the expression is given for the node i in the case of unnormalised MRF.

3) The inference task is to compute the likelihood of the observed data (in models with latent variables), which is similar to the previous case, but for the family of BNs.

It is helpful to refer to all of the presented tasks as *inference* problems and describe the corresponding methods only for the first problem since other tasks share the same complexity class, which scales similarly in the worst-case scenarios. Furthermore, the inference tasks can differ in the knowledge about the underlying structure of variables with their interaction and the full/partial observability of the variables. Finally, the last factor can be easily transferred to the physical hardware by eliminating the readout mechanism from the hidden variables' elements.

There are several conventional approaches to inference tasks. Exact inference algorithms include the brute force method (counting all possible combinations of variables), the elimination of variables algorithm and the message passing (MP) (also called the belief propagation (BP)), which can be exact, given particular topology of the variables connectivity (like the tree structure). Moreover, the running time of these exact algorithms scales exponentially with the size of the largest cluster, which is called the induced width of the graph [379]. However, since the majority of the practical cases can not be treated precisely, one can use approximate inference algorithms.

There are mainly three approximate inference algorithms: variational inference (mean-field approximation), stochastic simulation (or sampling), and loopy BP. We are interested in them not from the perspective of conventional algorithms but concerning the optical hardware. Hence, it is helpful to describe the original routine methods and then go to the methods of exploiting the specific physical systems.

The simplest way to deal with the inference in intractable probabilistic models is to use sampling subroutines to gain necessary information about the configurational space. Usually, one is calculating instances of the function of interest to gain statistical distribution in the hope it closely resembles the property of the actual distribution. However, to perform the sampling on the unconventional hardware, one has to properly set up the parameters in the system (for example, set up the coupling strengths in the Ising simulator in correspondence with the pairwise probabilities of binary events) and faithfully be able to realize initial random distribution of variables to perform the dynamics of the system. In such a way, it is possible to enjoy much faster operational speed, lower energy consumption, specific intrinsically set up readout

mechanism of a specific metric or any other beneficial property that a particular physical system can offer. Moreover, by changing the operational regime of the system into finding the lowest energy values of the given functional, one can sample the low energy subspace or the most probable events in the given factor graph. The same idea works for the high energy subspace or rarest events by simply changing the sign before the coupling coefficients. It is possible to introduce post-processing techniques into the operational regime of hardware or modify it to obtain various modifications of the sampling procedure, such as importance sampling, Markov Chain Monte Carlo, or their particular cases, such as Gibbs sampling and the Metropolis-Hastings algorithm [31, 385, 386].

It is not easy to estimate how close the distribution obtained through sampling statistics is to the true one. An alternative approach is to use variational techniques. The main idea of variational methods is to reformulate the inference task in the form of a tractable model, i.e. to optimize corresponding parameters. Given an intractable probability distribution $p(x)$ variational technique is aimed to solve an optimization problem over a class of tractable distributions $\mathcal{W}$ to find a $w(x) \in \mathcal{W}$ that is similar to $p(x)$. Further working with the $w(x)$ is much simpler since its tractability, ability to analyze the convergence, bounds on the accuracy and good scalability [387, 388].

To solve the optimization problem for the parameters of an approximate distribution, one has to choose the family of models $\mathcal{W}$ and a proper cost function $Q(w(x))$, which reflects the similarity between the initial $p(x)$ and the chosen model $w(x)$. The common choice for this purpose is called the Kullback-Leibler (KL) divergence:

$$KL(w\|p) = \sum_x w(x) \log \frac{w(x)}{p(x)}, \tag{4.41}$$

which is used to measure differences in information contained within two distributions [389]. Since calculating the normalisation factor is hard, it is easier to work with the KL divergence for the unnormalised distribution:

$$KL(w\|\widetilde{p}) = \sum_x w(x) \log \frac{w(x)}{\widetilde{p(x)}} = KL(w\|p) - \log Z. \tag{4.42}$$

Due to the positivity of the KL term, one can establish the lower bound on the log partition function, which can be helpful in some cases.

The choice of approximating family $\mathcal{W}$ assumes many different methods, such as exponential families, NNs, Gaussian processes and many others. However, the practically efficient and the most widely used class of distributions is the set of fully factored functions $w(x) = w_1(x_1)...w_d(x_d)$, which is easy to optimize by solving the following optimization problem:

$$\min_{w_1,\dots,w_d} KL(w\|\widetilde{p}). \tag{4.43}$$

Such an approach of choosing $\mathcal{W}$ is called mean-field inference. To solve it, one can use coordinate descent over all of the factor functions w_i for each i, while keeping $w_j, j \neq i$ fixed for each iteration. Such iterative optimization procedure has a closed-form solution. The final optimized model is often good enough for many practical applications.

One can use the special hardware in two main ways to perform the same methodology of variational inference. First is the indirect way, when one tune the physical system parameters, corresponding to the set of independent variables and comparing it with the actual distribution. The second is direct when one solves the exact minimization problem (4.43) iteratively for each w_i over the given domain by projecting the functional (4.41) into the available optimization assignment.

bp (sum-product MP) is a message-passing algorithm for performing inference on graphical models [390, 391]. Compared to the acyclic graphs, operating on a general factor graph with dense connectivity is usually called loopy BP. It aims at calculating the same approximate inference tasks using specific functions that are called messages.

There are two types of messages that are computed differently depending on the source node on a factor graph. The first type of messages comes from a variable node x_i to a factor node h and appears to be a product of all the messages from neighboring factor nodes:

$$\mu^{(t+1)}_{i\to h}(x_i) = \prod_{H^*\setminus h} \mu^{(t)}_{h^*\to i}(x_i), \tag{4.44}$$

where H^* denotes the set of all neighboring factor nodes h^* of a x_i except the final factor node h. The second type of messages comes from a factor node to a variable node and can be calculated as the product of the potential functions from (4.37) with messages from all other nodes:

$$\mu^{(t)}_{h\to i}(x_i) = \sum_{x_j\setminus x_i} \psi_h(x_j) \prod_{k\setminus j} \mu^{(t)}_{k\to h}(x_k). \tag{4.45}$$

The indices (t) and $(t+1)$ denotes the order of calculations, however different scheduling can be used for updating the

messages. For the general case of dense graph with many loops, there is no optimal scheduling. After any number of iterations, an estimate of the max-marginals (or beliefs) is obtained as:

$$b_i^{(t)}(x_i) = \prod_{H^*} \mu_{h^* \to i}^{(t-1)}(x_i), \tag{4.46}$$

and serves as good estimators of the marginal distributions.

Direct analogue computation on the specific unconventional hardware with the limited programmability option is challenging. Nevertheless, one can attempt to decompose the mathematical operations used in calculating the messages and beliefs to the ones available on the specific hardware, as was performed in the Section 3.4 of Chapter 3, see additionally [3]. However, such an approach has many drawbacks, mainly in the enormous amount of required elements, low analogue numerical accuracy and bad scalability options.

Nevertheless, the discrete update rules of MP have a very similar mathematical structure as the dynamics of the Hopfield NN given by Eq. (4.29). This connection was investigated in [392], where one can find the close relationship between MP algorithms and models of neurodynamics. It was shown that the equations of a continuous Hopfield network could be derived from the equations of BP on a binary MRF with the assumption of sparse neuron connectivity. A similar topic was investigated further for the general MP algorithms [393]. This trick can be exploited to perform approximate inference, using the conventional dynamics of the special-purpose hardware even in the simple GD manner.

4.3 Direct encoding/decoding

Considering the previous list of methods, one can trace their derivation in different tasks of applied interest. In some cases, it is possible to pick out the part that corresponds to the task easily solvable by the GD-like algorithm. However, the big part of the classical computer science algorithms that are still important and relevant in a modern setting is missing. This subsection is devoted to the general description of the existing connections/correspondence between different computational tasks established for nearly 50 years ago [340, 394, 395]. It can be solved with the same techniques offered in this chapter. Each case of the encoding/decoding task into the polynomial optimization (4.1) can be considered uniquely; nevertheless, the present focus is on the universal approach using the SAT formulation.

The Boolean satisfiability problem (SAT) is the fundamental problem of determining whether a set of sentences in Propositional Logic is satisfiable. For better understanding, one can give a simple example of the formula "a AND b", which is satisfiable since two variables can be consistently replaced by the values TRUE that the general formula evaluates to TRUE. In case that there is no such assignment for variables, the formula would be called unsatisfiable ("a AND NOT a"). Every propositional logic formula can be transformed into an equivalent conjunctive normal form with the laws of Boolean algebra. To give an example of SAT problem and a corresponding transformation one can write:

$$(x_1 \wedge y_1) \vee (x_2 \wedge y_2) \vee ... \vee (x_n \wedge y_n) \Leftrightarrow (x_1 \vee x_2 \vee ... \vee x_n) \wedge (y_1 \vee x_2 \vee ... \vee x_n) \wedge (x_1 \vee y_2 \vee ... \vee x_n) \wedge (y_1 \vee y_2 \vee ... \vee x_n) \wedge ...$$

$$... \wedge (x_1 \vee x_2 \vee ... \vee y_n) \wedge (y_1 \vee x_2 \vee ... \vee y_n) \wedge (x_1 \vee y_2 \vee ... \vee y_n) \wedge (y_1 \vee y_2 \vee ... \vee y_n), \tag{4.47}$$

where conjunction AND is denoted by $\wedge$, disjunction OR is denoted by $\vee$, and the number of clauses corresponds to every possible combination of pair variables.

SAT is the first problem that was proven to be NP-complete [394, 340]. Currently, no known algorithm efficiently solves each SAT instance. The question of its existence is equivalent to the famous P vs NP problem [396]. Nevertheless, many heuristic SAT-algorithms can solve problem instances involving large amounts of variables, sufficient for many applications. Additionally, there are many versions of the SAT problems, like 3-SAT and the generalization k-SAT, HORN-SAT, XOR-SAT, which can better suit particular unconventional tasks.

Reformulating a particular task into the SAT notation can serve professional computer scientists as a universal tool in long-standing problem solving, like Pythagorean triples problem [397] and Schur number 5 [398]. SAT version - weighted weighted MAX-2-SAT appears to be a universal formulation since it allow one to easily reformulate the task as QUBO. Let us demonstrate the encoding procedure and the main terms more precisely. A 2-SAT has m clauses of 2 literals each. A MAX-2-SAT is the problem of assigning values that maximize the number of satisfied clauses. Weighted MAX-SAT assigns each clause a positive weight, so that the measure of violating the cost appears in the problem:

$$(\ell_{1,1} \vee \ell_{1,2}; w_1) \wedge (\ell_{2,1} \vee \ell_{2,2}; w_2) \wedge ... \wedge (\ell_{m,1} \vee \ell_{m,2}; w_m), \tag{4.48}$$

where $\ell_{i,1}$ and $\ell_{i,2}$ are the two literals of clause c_i, and w_i is its weight. To reformulate a weighted MAX-2-SAT problem as a QUBO, one has to use the fact that maximizing the weight of satisfied clauses is equivalent to minimizing the weight

of unsatisfied clauses, and using the logic $\overline{x_i \vee x_j} = \overline{x_i} \wedge \overline{x_j}$. The final form looks then:

$$\min_{x_i} \sum_i w_i \bar{\ell}_{i,1} \bar{\ell}_{i,2}, \tag{4.49}$$

which is the QUBO that has the same form as Eq.(4.1) with the tensor of rank 2. Thus, the connection between the SAT (that can be easily converted into weighted MAX-2-SAT by use of the Boolean logic) and QUBO is revealed. In such a way, one can exploit many SAT formulated tasks and convert then into the optimisation problems.

This connection, however, was lying on the surface and was explored in the vital work [307], which provided Ising formulations for many NP-complete and NP-hard problems and covered all of Karp's 21 NP-complete problems. One can find number partitioning, graph partitioning, clique existence, binary integer linear programming, exact cover, set packing (or maximal independent set), vertex cover, satisfiability (with the emphasis on 3SAT to maximum independent set reduction), minimal maximal matching, set cover, knapsack with integer weights, graph colouring, clique cover, job sequencing with integer lengths, hamiltonian cycles and paths, travelling salesman problem, minimal spanning tree with a maximal degree constraint, Steiner trees, directed and undirected feedback vertex set, feedback edge set, graph isomorphisms among the covered problems, as well as some useful tricks for the near-term quantum adiabatic optimization devices.

Here the exact encoding of three easy problems is presented, two of which are taken from the list above for the demonstrative purpose.

Maximum independent set

Considering the sets to be encoded in an undirected graph $G = (V, E)$, where each set V_i maps to a vertex $i \in V$, one seeks the largest subset of vertices of graph G. An edge $ij \in E$ exists when $V_i \cap V_j$ is non-empty. It is straightforward to see that the corresponding Hamiltonian is [307]:

$$H = A \sum_{ij \in E} x_i x_j - B \sum_i x_i, \tag{4.50}$$

where the first term is responsible for non-intersection constraint and the second is minimizing one set and maximizing the size of the opposite set. The question of what is the maximal number of vertices which may be "coloured" ($x_i = 1$) such that an edge connects no two coloured vertices, is exactly equivalent to the set packing problem, MAX-Cut or the reduction of 3SAT problems.

One can encode the binary variables $x_i \in 0, 1$ into the Ising variables, just simply using the relation $s_i = 2x_i - 1$.

Set cover

The set covering problem consists in finding the smallest possible number of sets V_is, such that the union of them is equal to U, where a set $U = \{1, ..., n\}$ consists of sets $V_i \subseteq U (i = 1, ..., N)$ so that $U = \bigcup_{i=1}^{N} V_\alpha$. This problem is NP-hard [340]. Introducing a binary variable x_i, which is 1 if the set i is included and 0 otherwise. $x_{\alpha,m}$ is another binary variable, which is 1 if the number of sets V_i with the α element is $m \geq 1$, and 0 otherwise. The first energy imposes the constraints that exactly one $x_{\alpha,m}$ must be 1 because each element of U must be included a fixed number of times. The number of times that we claimed α was included is, in fact, equal to the number of V_i have been included, with α as an element. Thus, the energy has the form [307]:

$$\begin{aligned} H_A =& A \sum_{\alpha=1}^{n} \left(1 - \sum_{m=1}^{N} x_{\alpha,m} \right)^2 \\ &+ A \sum_{\alpha=1}^{n} \left(\sum_{m=1}^{N} m x_{\alpha,m} - \sum_{i:\alpha \in V_i} x_i \right)^2 . \end{aligned} \tag{4.51}$$

It is obvious then that with the given variables one has to minimize over V_αs included in the selection:

$$H_B = B \sum_{i=1}^{N} x_i, \tag{4.52}$$

with the necessary condition $0 < B < A$ to be satisfied in any case.

Treewidth computation

To expand the methodology of encoding/decoding an arbitrary computational task, the particular assignment of finding the graph's treewidth is presented. The treewidth of a graph can be defined in several equivalent ways: the size of the largest vertex set in a tree decomposition of the graph, the size of the largest clique in a chordal completion of the graph or the maximum order of bramble, a collection of connected subgraphs that all touch each other.

The initial idea behind this task is to adjust the annealer-like hardware/heuristics into performing some dynamic programming assignment; however, it appears to be very hard to expand the initially defined scope of the setup or method if the adjustment of the particular hardware does not involve it to perform calculations in the same dynamical and not static manner (under static one assume the final state of the system which outputs the final stable set of variables, like the annealing technique converges to the suboptimal solution). Nevertheless, it is possible to get some help from such special-purpose hardware by providing insight into the considered structure and exploit this information for classical algorithms as an example.

Giving the standard definition, which will be clarified through the encoding procedure: a tree decomposition of a graph $G = (V, E)$ is a tree, T, with nodes $X_1, ..., X_n$, where each X_i is a subset of V, satisfying the following properties [399]:

- The union of all sets X_i equals V. That is, each graph vertex is contained in at least one tree node.
- If X_i and X_j both contain a vertex v, then all nodes X_k of T in the (unique) path between X_i and X_j contain v as well. Equivalently, the tree nodes containing vertex v form a connected subtree of T.
- For every edge (v, w) in the graph, there is a subset X_i that contains both v and w. Vertices are adjacent in the graph only when the corresponding subtrees have a node in common.

The *width* of a tree decomposition is the size of its largest set X_i minus one. The **treewidth** $\mathrm{tw}(G)$ of a graph G is the minimum width among all possible tree decompositions of G.

Treewidth is usually used in the parameterized complexity analysis of graph algorithms. It shows the similarity between the tree structure and the particular graph with its connections between nodes. Every complete graph has treewidth $n - 1$ with n being the number of nodes. A connected graph with at least two vertices has treewidth 1 if and only if it is a tree.

It is NP-complete to determine whether a given graph G has treewidth at most a given variable k [400]. However, when k is any fixed constant, the graphs with treewidth k can be recognized, and width k tree decomposition constructed for them, in linear time [401]. To summarizing, one can build tree decomposition of a graph, knowing its structural property, known as treewidth. Moreover, one can find the treewidth using the standard encoding into the Ising problem; however, this encoding will use the auxiliary algorithm, known as integer linear programming (ILP). Knowledge of the constant k, obtained by the annealing-technique means, can reduce the complexity of the initial NP problem and make it P class.

It is useful to use the ILP-based approach to tackle the problem of determining the treewidth of the graph that was presented in [402]. The work is based on the elimination order formulation [403, 404], which is based on the relationship between treewidth and chordalizations of graphs. Finding the treewidth of a graph G is equivalent to finding a triangulation of the graph G with minimum clique size. However, the graph can be triangulated if the perfect elimination scheme exists. The idea is to determine the best elimination order of the vertices. The maximum outdegree among the found elimination scheme's vertices should be equal to the width of the final tree decomposition of G, which is obtained after the triangulation process. To model the perfect elimination scheme, one needs to introduce the decision variables of ILP in the following way:

$$x_{ij} = \begin{cases} 1, \text{ if } i \text{ is ordered before } j \text{ in the perfect elimination scheme,} \\ 0, \text{ otherwise.} \end{cases} \tag{4.53}$$

$$y_{ij} = \begin{cases} 1, \text{ if } i \text{ is ordered before } j \text{ in the perfect elimination scheme} \\ \text{ and if } ij \text{ is an edge of the triangulation,} \\ 0, \text{ otherwise.} \end{cases} \tag{4.54}$$

Following the original article, the elimination order formulation reads:

$$\min w \tag{4.55}$$
$$w \geq \sum_{j \in V} y_{ij}, \forall i \in V \tag{4.56}$$
$$x_{ij} + x_{ji} = 1, \quad \forall \{i,j\} \subseteq V \tag{4.57}$$
$$x_{ij} + x_{jk} - x_{ik} \leq 1, \quad \forall \{i,j,k\} \subseteq V \tag{4.58}$$
$$y_{ij} \leq x_{ij}, \quad \forall \{i,j\} \subseteq V \tag{4.59}$$
$$y_{ij} = x_{ij}, \quad \forall ij \in E \tag{4.60}$$
$$x_{jk} + y_{ij} + y_{ik} - y_{jk} \leq 2, \quad \forall \{i,j,k\} \tag{4.61}$$
$$x_{ij} \in \{0,1\}, y_{ij} \in \{0,1\}, \forall i,j \in V \tag{4.62}$$

The constraints (4.58),(4.59) and (4.62) are necessary for variables x to establish a linear order. Constraints (4.60) force the y_{ij} variable to be zero as soon as i is not ordered before j. Constraints (4.60) impose the correspondence between edges of G in the triangulation. Constraints (4.62) imply the consistency among each triangle, more precisely - that if edges ij and ik are in the triangulation with i ordered before j and i ordered before k, then there must exist an edge jk in the triangulation.

By simple manually defined encoding, one can write down the following cost function terms (which have to be minimized), associated with each assignment:

$$H_{12} = \sum_{i,j} (y_{ij})^2 \tag{4.63}$$
$$H_3 = \sum_{i,j} (1 - x_{ij} - x_{ji})^2 \tag{4.64}$$
$$H_4 = \sum_{i,j,k} x_{ij} x_{jk} (1 - x_{ik}) \tag{4.65}$$
$$H_5 = \sum_{i,j} y_{ij} (1 - x_{ij}) \tag{4.66}$$
$$H_6 = \sum_{i,j} J_{ij} (y_{ij} - x_{ij})^2 \tag{4.67}$$
$$H_7 = \sum_{i,j,k} x_{jk} y_{ij} y_{ik} (1 - y_{jk}). \tag{4.68}$$

The presented terms constitute the Hamiltonian, which is intended to be minimized. It is possible to map this assignment to the (4.1) precisely working with the polynomial terms of the 4-th order. However, one needs to unfold the presented logic by introducing the additional variables to reduce the tensor's order to the matrix. It makes the encoding very hard and impractical even for the small graphs ($(2N^2)^2$ variables in the most general sense, without considering the internal symmetry).

Despite the little practical interest of the last example, it presents the methodological idea. It demonstrates the so-cold hardware-assisted computation, for which one can exploit both the conventional digital computer and the information obtained from the annealer-setup. In Chapter 3, we have already discussed the dynamical aspect of the minima search.

Chapter 5

Conclusion

The results obtained through the thesis research are pretty diverse; nevertheless, they form a unique and interconnected picture. The EP system appeared to be a promising platform, which can lead to the realization of many complex computational tasks. Setting the experimental parameters properly, one can solve the QUBO or switch to its generalization - the HOBO problem. Additionally, It was shown that complex ML tasks could be transferred to this condensed matter system, one of the best candidates for realizing different hardware due to its variety of desirable properties from the underlying physics. A rich list of possible applications was considered. Nevertheless, the analytical description of periodic structures was achieved, which is significant for both theory and applications. Moreover, an extensive analysis of the existing computational assignments was performed, which allow one to operate on the EP lattices in a purely engineering sense.

To summarise, EPs appear to be an auspicious physical system, serving as the perfect foundation for technological advancement. Nevertheless, many physical systems can be exploited in terms of computation. The best way to solve computational problems is to apply strict formalization of the task. The results show it is possible to connect much of the CS assignment with a specific platform, with the remaining question - to what degree this connection will be efficient? One has to keep in mind the tradeoff between the physical setup and engineering possibilities and the complex model/description of the system.

The conventional digital computers will be exploited for a long time. However, particular tasks of high complexity can be solved with the special purpose hardware, occupying its niche in the future, resulting in our possibilities' very rugged landscape.

Despite extensive multidisciplinary research, there are many questions one can pursue as future directions. To what extent can we complicate the system to 'squeeze' it for the most computational efficiency or exploit its dynamical behaviour to solve even more hard problems? How do the quantum effects manifest themselves in finding a good solution to a specific problem, and is there any way to incorporate such a mathematical description to obtain an efficient heuristic algorithm? What is the main differences between different physical platforms, and how to properly construct easy/hard instances to test their performance?

Chapter 6

Additional resources

6.1 Physical optical platforms for large-scale optimization

Recent years have spawned a competition of different platforms aimed at solving classical optimization problems with the advantage in speed compared to the classical hardware for a given type of problem. To gain such an advantage, one has to exploit physical phenomena for solving NP-complete problems, which is the reverse idea of trying to model natural phenomena. Hence the new multidisciplinary field appeared at the intersection of laser and condensed matter physics, engineering and complexity theories. Its main goal is to develop quantum devices to simulate spin Hamiltonians. This section contains information about such simulations for a range of physical systems and is largely based on the work [405].

6.1.1 Complex laser networks

A new generation of complex lasers such as degenerate cavity lasers, multimode fibre amplifiers, large-aperture vertical cavity surface emitting lasers, random lasers have many advantages in comparison with the relatively simple traditional laser resonators in terms of their computing properties [406]. Such networks have many advantages, including a large number of spatial degrees of freedom, nonlinear interactions that can be controlled by adjusting the spatial structures of lasing modes, the spatial coherence of emission can be tuned over a wide range, and the output beams may have arbitrary profiles. The presented list allows one to exploit the complex lasers to be used for reservoir computing [407] or perform an efficient mapping of hard computational problems into the hardware setup.

The coupling between the lasers can be engineered by mutual light injection from one laser to another, which is the reason for appearing of the losses, or in other words, dissipative coupling, which depends on the relative phases between the lasers. Such connections drive the system to a phase-locking and hence to a steady-state solution of quadratic constrained optimization, which is the same as finding the minimum of the XY Hamiltonian [408]. Degenerate cavity lasers are beneficial as solvers as all their transverse modes have nearly identical Q. This implies that a large number of transverse modes lase simultaneously since they all have similar lasing thresholds [406].

The evolution of the N single transverse and longitudinal modes class–B lasers can be described by the rate equations [409, 410] on the amplitude A_i, phase θ_i, and gain G_i of the i-th laser

$$\frac{dA_i}{dt} = (G_i - \alpha_i)\frac{A_i}{\tau_p} + \sum_j J_{ij}\frac{A_j}{\tau_p}\cos(\theta_i - \theta_j), \tag{6.1}$$

$$\frac{d\theta_i}{dt} = \Omega_i - \sum_j J_{ij}\frac{A_j}{\tau_p A_i}\sin(\theta_i - \theta_j), \tag{6.2}$$

$$\frac{dG_i}{dt} = \frac{1}{\tau_c}[P_i - G_i(1 + |A_i|^2)], \tag{6.3}$$

where P_i, α_i, Ω_i represent the pump strength, loss, frequency detuning of laser i, respectively, whereas τ_p and τ_c denote the cavity round trip time and the carrier lifetime, respectively. The coupling strengths between i-th and j-th lasers are represented by J_{ij}. If the amplitudes of all lasers are equal, Eq. (6.2) reduces to the Kuramoto equation of coupled phase oscillators

$$\frac{d\theta_i}{dt} = \Omega_i - \frac{1}{\tau_p}\sum_j J_{ij}\sin(\theta_i - \theta_j). \tag{6.4}$$

Equation (6.4) is a celebrated Kuramoto model of identical oscillators which is widely used to describe the emergence of coherent behaviour in complex systems [411, 412]. By LaSalle Invariance Principle [413] every trajectory of the Kuramoto model converges to a minimum of the XY Hamiltonian.

The probability of finding the global minimum of the XY Hamiltonian on the experimental setup of the laser array was shown to agree with the numerical simulation of Eqs. (6.1-6.3). However, simulating the Kuramoto model Eq. (6.4) on the same matrix of coupling strength gives a much lower probability of finding the global minimum. One can conclude the amplitude dynamics described by Eq. (6.1), which can reach the global minimum [410] by pumping from below. Such a statement means that the cavity lasers can be used as an efficient physical simulator for finding the XY Hamiltonian's global minimum and, therefore, for solving phase retrieval problems. The phase retrieval problem can be solved using a digital degenerate cavity laser [414]. It is an all-optical system that uses a nonlinear lasing process to find a solution that best satisfies the constraint on the Fourier magnitudes of the light scattered from an object. Furthermore, the compact support aperture is introduced inside the cavity that ensures that different configurations of laser phases compete to find the one with the minimal losses, ensuring that the solution to the phase retrieval problem is found correctly. The advantages of such a system are short round-trip times of the order of 20ns and high parallelism in selecting the advantageous mode.

6.1.2 Coherent Ising machine

The coupled optical parametric oscillator (OPO) network represents an alternative physical system for solving the Ising problem. Each OPO is a nonlinear oscillator with phases above the threshold that can take a discrete value and be interpreted as the Ising spins. These artificial spins are encoded by the optical phase of short laser pulses generated by a nonlinear optical process, i.e. optical parametric amplification. Based on the OPO elements, coherent Ising machine (CIM) is a gain-dissipative system, which exploits the correspondence between the ground state of the Ising Hamiltonian and the lowest loss configuration. The idea of such a device is to drive the system close to the near-threshold regime, where energy minima are not stable yet, except the one responsible for the optimal solution.

The latest most successful implementations of CIMs used fibre-based degenerate OPOs and a measurement-based feedback coupling. The matrix matrix-vector multiplication has to be performed on an FPGA embedded in the feedback loop to establish the coupling coefficient. Such a platform enjoys scalability, which is bounded by the electronic feedback, and the performance was demonstrated for various large-scale Ising problems [415, 416, 417]. The possible CIM's speedup over classical algorithms is the question under consideration [418, 419]. The significant advantage of this device is the ability to implement arbitrary coupling connections [415] between any two spins, so that the connectivity topology is not restricted by specific factors, compared to the quantum annealer, i.e. D-Wave machine [416].

In CIM, each Ising spin corresponds to a degenerate OPO that is described by a stochastic equation for the complex amplitude of the signal field a_i:

$$\frac{da_i}{dt} = pa_i^* - a_i - |a_i|^2 a_i + \sum_j J_{ij} a_j, \tag{6.5}$$

where p is a linear pump term, normalized linear and Nnonlinear nonlinear losses, and J_{ij} are mutual couplings. A portion of the light is extracted from the cavity after each round trip to realize these couplings experimentally. That light is then homodyned against a reference pulse to produce a_i that is next supplied to FPGA. A feedback signal is computed for each pulse afterwards. An optical modulator is applied to convert the signal back so that the resulting light can be used for the next round trip. It is convenient to reformulate the equations (6.5) in terms of the in-phase and quadrature components $a_i = c_i + is_i$ which gives us:

$$\frac{dc_i}{dt} = \Big(p - 1 - (c_i^2 + s_i^2)\Big)c_i + \sum_j J_{ij} c_j \tag{6.6}$$

$$\frac{ds_i}{dt} = \Big(-p - 1 - (c_i^2 + s_i^2)\Big)s_i + \sum_j J_{ij} s_j. \tag{6.7}$$

The computational effectiveness of these equations has been demonstrated [420] by tackling small size Ising type problems. These equations lack an individual pump variation p_i for equalizing all signal amplitudes $|a_i|$, which is crucial for achieving the global minimum.

The degenerate OPOs system is highly susceptible to external perturbations that can affect performance [416]. The requirements of the nonlinear degenerate OPO generation process are powerful laser systems and temperature-controlled nonlinear materials. This makes the experimental optical setup large and very complex, leading to other physical platforms' proposals for the same operational model. Another platform is a CIM based on optoelectronic oscillators with self-feedback, which is stable and cheaper. It operates with 100 spin elements to solve Ising optimization problems on

regular and frustrated graphs with similar or better performance compared to the original degenerate OPO-based CIM [421]. One can use a network of injection-locked multicore fibre lasers for analogue all-optical implementation of a CIM [422]. The dynamics of a network of injection-locked lasers is described with the nonlinear coupled photon rate equations. The coupling coefficients were implemented using SLMs and were reported to depend on the photon numbers. However, these numbers are unknown in advance, representing a particular difficulty for solving a given Ising Hamiltonian with the proposed device. These issues can be efficiently resolved with the gain variation methods in the future [423]. Considering the large scale machines, one can find experiments with optical Ising machine based on the use of an SLM and binary phases in separated spatial points of the optical wavefront of an amplitude-modulated laser beam with thousands of spins with tunable all-to-all pairwise interactions [424].

6.1.3 Photon and Polariton networks

Microcavity EPs were extensively discussed in the main chapters of this work. The steady states in these nonequilibrium systems are set by the balance between the pumping from the exciton's reservoir into polaritons and losses in the form of photons leakage. A similar system of photon gas confined in a dye-filled optical microcavity closely resembles the physics of polariton condensates with the macroscopic occupation of the lowest mode [425, 426, 427, 428]. The accumulation of low-energy photons happens due to the rapid thermalization of rovibrational modes of the dye molecules by their collisions with the solvent and phonon dressing of the absorption and emission by the dye molecules.

There are many ways to realize the lattices of polariton or photon condensates realized in experiments. Injecting polaritons in the specific spots of the sample using an SLM is one method of creating optical periodic structure [429, 430, 431, 432, 433]. Similarly, there are many ways of engineering potential landscapes to confine polariton or photons [434, 435, 436]. To describe the evolution of gain-dissipative condensates in a lattice considering only reduced macroscopic degrees of freedom of each structural unit, one has to derive the rate equations [423, 437] (see also the derivation in Chapter 3), which take a form of the Stuart-Landau equations:

$$\dot{\Psi}_i = -iU|\Psi_i|^2\Psi_i + (\gamma_i - |\Psi_i|^2)\Psi_i + \sum_{j\neq i}\mathcal{C}_{ij}\Psi_j, \tag{6.8}$$

where $\Psi_i = \sqrt{\rho_i}\exp[i\theta_i]$ is the complex amplitude of the $i-$th condensate, U is the strength of self-interactions between the quasi-particles, γ_i is the effective injection rate (the difference between the pumping of the quasi-particles into the system and linear losses). The coupling strength $\mathcal{C}_{ij} = J_{ij} + iG_{ij}$ is generally a complex number and consists of the Heisenberg coupling J_{ij} mediated by the injection reservoir and the Josephson part G_{ij} that comes from exchange interactions between the condensates. The system described by Eq. (6.8) reaches the fixed point when $J_{ij} \gg G_{ij}$ and the pumping feedback is introduced in the system [423]. The feedback on the pumping intensity ensures that all the occupations are the same at the fixed point by adjusting the pumping if the occupation exceeds the set threshold value $|\Psi_i|^2 = \rho_{\text{th}}$. The total injection of the particles in the system of N condensates at the fixed point is given by

$$\sum_{i=1}^{N}\gamma_i = N\rho_{\text{th}} - \sum_{i=1}^{N}\sum_{j<i}^{N} J_{ij}\cos(\theta_i - \theta_j). \tag{6.9}$$

To guarantee that the found solution is the minimum of the XY Hamiltonian, the current expression should be chosen with the lowest possible total particle injection $\sum\gamma_i$ that leads to the specific occupation ρ_{th}. In order to find the true global minimum, the system has to slowly be brought to the condensation threshold while spending enough time to explore the phase space to check different phase configurations driven by the system noise. After reaching such phase configuration, the system quickly converges to the global minimum by the GD given by the imaginary part of Eq. (6.8):

$$\dot{\theta}_i = -U\rho_{\text{th}} - \sum_{j\neq i}^{N} J_{ij}\sin(\theta_i - \theta_j). \tag{6.10}$$

This idea has been theoretically justified [423] and experimentally realized for simple polariton graphs [433] with further extensions to discrete optimization problems such as QUBO (minimizing the Ising Hamiltonian) or n-states Potts Hamiltonians [438]. Combining resonant excitation with a non-resonant one forces the spins to take the discrete values aligning with the directions set by the resonant excitation. If the resonant drive with $n:1$ ratio is added to the system, the dynamics of condensates will acquire additional term:

$$\dot{\Psi}_i = -iU|\Psi_i|^2\Psi_i + (\gamma_i - |\Psi_i|^2)\Psi_i + \sum_{j\neq i} J_{ij}\Psi_j + h(t)\Psi_i^{*(n-1)}, \tag{6.11}$$

where $h(t)$ is an increasing function that reaches some value $H > \max_i \sum_j |J_{ij}|$ at the threshold. At the fixed point, Eq. (6.9) is replaced with

$$\sum_{i=1}^{N} \gamma_i = N\rho_{\text{th}} - \sum_{i=1}^{N}\sum_{j<i}^{N} J_{ij}\cos(\theta_i - \theta_j) - H\rho_{\text{th}}^{n/2-1}\cos(n\theta_i). \tag{6.12}$$

In the case of $n = 2$, the additional term on the right-hand side provides the penalty to phases deviating from 0 or π, which reduce the possible available phase and project the initial optimization problem to QUBO. For $n > 2$, the n-state Potts Hamiltonian is minimized. The minimization of HOBO may be achieved when the system operates much above the threshold, and higher-order terms can not be neglected [2], see also the Chapter 3.

The system of N interacting coherent centers is better described by the following equations [437] in case of slow time evolution of the exciton reservoir:

$$\dot{\Psi}_i = -iU|\Psi_i|^2\Psi_i + (R_i - \gamma_c)\Psi_i + \sum_{j\neq i} J_{ij}\Psi_j, \tag{6.13}$$

$$\dot{R}_i = \Gamma_i - \gamma_R R_i - R_i|\Psi_i|^2, \tag{6.14}$$

where R_i is the occupation of the $i-$th reservoir, Γ_i, γ_R and γ_c characterize the particle injection rate into the reservoir and the linear losses of the reservoir and condensate, respectively. This is a well-known form of Lang-Kobayashi equations, used to describe the dynamical behaviour of coupled lasers from Lamb's semiclassical laser theory [439, 440], where Ψ_i is replaced by the electric field and R_i by the population inversion of the $i-$th laser. The total injection of the particles in the system of N condensates at the fixed point is given by

$$\sum_{i=1}^{N}\Gamma_i = (\gamma_R + \rho_{\text{th}})[N\gamma_c - \sum_{i=1}^{N}\sum_{j<i}^{N} J_{ij}\cos(\theta_i - \theta_j)]. \tag{6.15}$$

Similar to Eq. (6.9), if the total injection into the system is minimal, the phases of coherent centres minimize the XY Hamiltonian.

6.2 Constructing problem instances

There are many ways to construct difficult instances for the optimization algorithms, including GD-based, listed at the end of this section. Although not the product of this thesis, two noticeable approaches are presented due to their appealing mathematical formulation and solid numerical analysis.

The first one is the *matrix-tensor model* introduced in [441] in the context of generalization in deep learning. Although it's original purpose is to investigate the performance of different algorithms in the task of recovering the noisy signal, one can exploit it for the purpose of creating hard instances for the optimization. The initial task for the matrix-tensor model is posed as a statistical inference problem with the ground truth signal $x^* \in \mathbb{R}^N$ which is sampled uniformly on the $N-1$-dimensional sphere, $x \in S^{N-1}(\sqrt{N})$. Two observations about the signal, a symmetric matrix Y and an order p symmetric tensor T are given:

$$Y_{ij} = \frac{x_i^* x_j^*}{\sqrt{N}} + \xi_{ij} \tag{6.16}$$

$$T_{i_1,\ldots,i_p} = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1}^* \ldots x_{i_p}^* + \xi_{i_1,\ldots,i_p} \tag{6.17}$$

for ordered indices $1 \leq i < j \leq N$ and $1 \leq i_1 < \ldots < i_p \leq N$, and the non-diagonal components equal their symmetric terms. ξ_{ij} and $\xi_{i_1,\ldots,i_p}$ denote independent Gaussian random variables $\mathcal{N}(0, \Delta_2)$ and $\mathcal{N}(0, \Delta_p)$ respectively. Given Y and T one aims at estimating the signal x^*. The case of tensor absence comes down to the low-rank perturbation of a random symmetric matrix or the spiked covariance model [442]. The original work [441] and the corresponding consequent ones [443, 444] are focused at the algorithmic tractability of obtaining good performance for the matrix-tensor model. The inference problem can be reformulated as an optimization task with the cost function of the form:

$$\mathcal{L} = \sum_{i<j} \frac{1}{2\Delta_2}\left(Y_{ij} - \frac{x_i x_j}{\sqrt{N}}\right)^2 + \sum_{i_1<\cdots<i_p} \frac{1}{2\Delta_p}\left(T_{i_1\ldots i_p} - \frac{\sqrt{(p-1)!}}{N(p-1)/2} x_{i_1}\ldots x_{i_p}\right)^2, \tag{6.18}$$

which can be also called as log-likelihood, loss function or Hamiltonian depending on the context and can be related to the initially considered model (4.1). This model lays the foundation for the question of tractability and properties of

non-convex and high-dimensional optimization problems.

Roughly speaking, two algorithms (GD and the maximum-likelihood approximate MP) were analyzed in the context of the matrix-tensor model. The energy landscape becomes trivial for the model's specific parameters, where two algorithms perform well enough and obtain the same error; however, the maximum-likelihood approximate MP is strictly better than GD and has a broader region of better performance.

Another interesting way of constructing a problem with a tunable algorithmic hardness is called the Wishart planted ensemble (WPE) [346]. It is a class of zero-field Ising models with a specifiable ground state. The key idea of designing hard instances for QP is to use a simple procedure to generate a set of random integer programming problems with specific statistical symmetry properties. Such a way of problem construction imposes the coupler matrix to follow a type of Wishart distribution.

One can recall the main logic behind the methodology of the work, which aim is the construction of ensembles of zero-field Ising Hamiltonians over the N-spin complete graph with planted ground state t_i (where index i goes over all variables) in the form of:

$$H(s_i) = -\frac{1}{2}\sum_{i\neq j} J_{ij}s_i s_j, \tag{6.19}$$

where variable s_i and the ground state t_i belong to the configuration space $\mathbb{S}^N \triangleq \{\pm 1\}^N$, so that their relationship is $H(\pm t_i) = \min_{s_i\in\mathbb{S}^N} H(s_i)$. t_i will be taken to have the ground state $t_i = (+1, +1, ..., +1)$ and its $\mathbb{Z}_2$ image, since its variation can be subsequently concealed by gauge randomization.

The key information about the solution and its correspondent hardness is contained in the W_{ij}, which is the $N \times M$ real-valued matrix $W_{ij} \in \mathbb{R}^{N\times M}$. To denote the column of such matrix one uses w_i^μ for $\mu = 1, ..., M$. The ratio between the M and N represents the scaling factor $\alpha > 0$: $M = \alpha N$. With the given ground state t_i, the procedure is to construct a consistent homogeneous Ising-constrained linear system with $s_i = \pm t_i$ as a solution, or to obtain the specific W_{ij} such that $W_{ji}^T t_i = 0_j$. It imposes that the positive semidefinite quadratic form:

$$G(s_i) = \frac{1}{2}s_i^T W_{ij}W_{ji}^T s_i = \frac{1}{2}\left\|W_{ji}^T s_i\right\|_2^2 \tag{6.20}$$

would get its minimum zero value at $s_i = t_i$ and the coupling coefficients for the initial problem can be defined as:

$$\tilde{J_{ij}} = -\frac{1}{N}W_{i\mu}W_{\mu j}^T, \tag{6.21}$$

with a small addition about the diagonal values, which have to be zero $J_{ij} = \tilde{J_{ij}} - \mathrm{diag}(\tilde{J_{ij}})$.

From the computational perspective, the task of finding the ground state of Ising model with the predefined matrix W_{ij} is equivalent to finding a solution to the integer programming feasibility (which is an NP-hard problem):

$$\begin{aligned} &\text{solve } W_{ji}^T s_i = 0_j \\ &\text{subject to } s_i \in \mathbb{S}^N. \end{aligned} \tag{6.22}$$

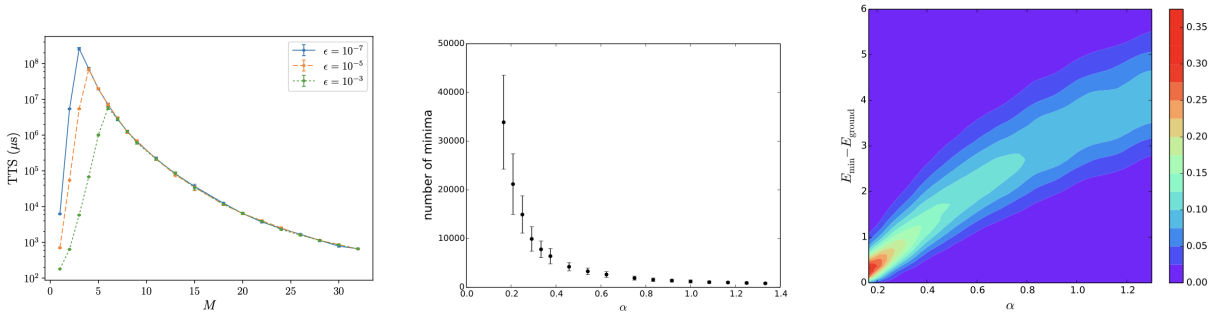
With the W_{ij} of size $N \times M$, where $M < N$ $\dim\left(\mathrm{null}\left(W_{ij}^T\right)\right) = N - M$. Large M results in a ferromagnetic system. Due to the law of large numbers:

$$\tilde{J_{ij}} = -\frac{1}{N}W_{im}W_{mj}^T = -\frac{\alpha}{M}W_{im}W_{mj}^T \to -\alpha\Sigma_{ij}, \tag{6.23}$$

with the couplers uniformly approach $J_{ij} = \alpha\frac{t_i t_j}{N-1}$. The system reduces to a gauge-transformed and rescaled Curie-Weiss ferromagnet, whose ground state is easy to find because of the absense of the frustration. Thus, the parameter M impacts problem difficulty with the hardest instances being generated for some intermediate values of M. This easy-hard-easy profile is investigated in more details in the original work [346]. Moreover, the article inspects the other connections with more computer science assignments, like the special case of $M = 1$ for Eq.(6.22), which is the subset sum problem; its further specialization that is the number partitioning problem (with many interesting properties, in particular an algorithmic easy-hard phase transition [445, 446, 447]), etc.

The analytical results were confirmed by extensive numerical calculations using the highly optimized implementation of parallel tempering Monte Carlo, including the observed hardness of finding the ground state of moderately sized instances, see Fig.6-1. The visualization of the different energy landscape is presented in Fig.6-2. The investigation of both models, the WPE and the matrix-tensor models leave the space for many questions concerning the dynamics of the algorithms and their behaviours in the multidimensional rugged landscape of a particular problem.

Other models. In addition to the matrix and matrix-tensor models, various differently defined models can describe complicated problems with tunable hardness. One can pay attention to the community detection and the stochastic block



(a) (b) (c)

Figure 6-1: (a) The mentioned easy-hard-easy profile, which is represented in the time to approximate solution vs M corresponding size of the W_{ij} matrix in Eq.(6.20) for the $N = 32$. Optimized parallel tempering Monte Carlo was used, despite the focus on the GD-like algorithms in this article. ϵ is the acceptable excess energy over the ground state. (b) Expected number of local optima for WPE instances of size $N = 24$ as a function of $\alpha = M/N$. (c) Residual locally optimal energy distributions for WPE instances of size $N = 24$ as a function of $\alpha = M/N$. For small α, the distributions are concentrated on low residual energy values. All three pictures are taken from [346].

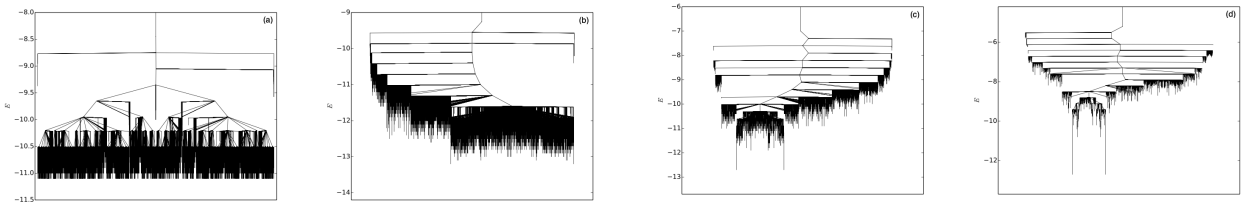


Figure 6-2: Disconnectivity graphs (which are two-dimensional representations of high-dimensional energy landscapes with corresponding energy barriers, that can be traced along a specific pathway from one minima to another) representing the energy landscape of four specific $N = 24$ WPE instances. Leaf nodes of the tree structures represent local minima, and internal nodes - barrier states. (a) $M = 4$ with a tremendous near-degeneracy of a large number of metastable states with energy very close to that of the planted solution. (b) $M = 8$, the degeneracy begins to lift with the decrease in the number of minima. As M increases to 15 (c) and 32 (d), the ground state becomes increasingly dominant and the problems computationally easier. The pictures are taken from the [346].

models [448, 47, 449, 450]. Other models include the locked constraint satisfaction problems [451], and the universal NK model [452, 453] with tunably rugged cost functions that can cover the complexity of a variety of different systems. Moreover, a physics-based approach was used for generating hard 3-SAT problems with planted solutions [454], the quiet planting [455] is the technique for devising graph q-colouring problems with known solutions.

No Free Lunch Theorem. The no free lunch theorem (NFLT) is a theorem stating that an effective general-purpose universal optimization strategy can not exist, and the difference in the performance between the strategies depends on the structure of the specific problem under consideration. NFLT appeared in 1997 in the work of David Wolpert, and William Macready [456]. Previously, Wolpert had derived the analogue theorem for ML (statistical inference) [457]. NFLT is important not only for the domain of algorithms and problem instances but also in the context of special-purpose hardware since each experimental setup represents its way of finding the solution to the specific problem. One can paraphrase the NFLT in many ways. For example - any two optimization algorithms are equivalent when their performance is averaged across all possible problems, or each algorithm has its own best and worst-case scenario—the same works for the specialized hardware. Therefore, one should pay attention to the NFLT when demonstrating performance results on the so-called best case scenario instances and carefully specify the details of the experiment. Moreover, we want to address several related questions closely connected with optimization as future research directions. Among them are:

1) Introduction of the standard indicators for the fair comparison of the hardware.

2) Methods of the hardware performance evaluation and introduction of the related specific metric (for example, time to solution, the quality of the performance, etc.)

3) How to introduce the constraints into the hardware operational regime?

4) Methodology of the hard instance generation for a particular setup (as well as easy ones).

5) How to generalize the results of the hardware operations.

Bibliography

[1] Nikita Stroev and Natalia G Berloff. Managing the flow of liquid light. *Physical Review B*, 102(20):201114, 2020.

[2] Nikita Stroev and Natalia G Berloff. Discrete polynomial optimization with coherent networks of condensates and complex coupling switching. *Physical Review Letters*, 126(5):050504, 2021.

[3] Nikita Stroev and Natalia G Berloff. Xy neural network. *arXiv preprint arXiv:2103.17244*, 2021.

[4] Nikita Stroev and Natalia G Berloff. Modern optical computing. in preparation, 2021.

[5] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011.

[6] Vijay V Vazirani. *Approximation algorithms*. Springer Science Business Media, 2013.

[7] Dorit S Hochba. Approximation algorithms for np-hard problems. *ACM Sigact News*, 28(2):40–52, 1997.

[8] David S Johnson. Approximation algorithms for combinatorial problems. *Journal of computer and system sciences*, 9(3):256–278, 1974.

[9] Daniel J Rosenkrantz, Richard Edwin Stearns, and Philip M Lewis. Approximate algorithms for the traveling salesperson problem. In *15th Annual Symposium on Switching and Automata Theory (SWAT 1974)*, pages 33–42. IEEE, 1974.

[10] EG Co man Jr, MR Garey, and DS Johnson. Approximation algorithms for bin packing: A survey. *Approximation algorithms for NP-hard problems*, pages 46–93, 1996.

[11] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[12] Rasmus Bro et al. Parafac. tutorial and applications. *Chemometrics and intelligent laboratory systems*, 38(2):149–172, 1997.

[13] Pierre Comon. Tensor decompositions. *Mathematics in Signal Processing V*, pages 1–24, 2002.

[14] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[15] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[16] Peter Baranyi, Peter Varlaki, Laszlo Szeidl, and Yeung Yam. Definition of the hosvd based canonical form of polytopic dynamic models. In *2006 IEEE International Conference on Mechatronics*, pages 660–665. IEEE, 2006.

[17] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[18] Qibin Zhao, Guoxu Zhou, Shengli Xie, Liqing Zhang, and Andrzej Cichocki. Tensor ring decomposition. *arXiv preprint arXiv:1606.05535*, 2016.

[19] Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

[20] Stephan Rabanser, Oleksandr Shchur, and Stephan Gunnemann. Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv:1711.10781*, 2017.

[21] Fengyu Cong, Qiu-Hua Lin, Li-Dan Kuang, Xiao-Feng Gong, Piia Astikainen, and Tapani Ristaniemi. Tensor decomposition of eeg signals: a brief review. *Journal of neuroscience methods*, 248:59–69, 2015.

[22] Zheng Zhang, Xiu Yang, Ivan V Oseledets, George E Karniadakis, and Luca Daniel. Enabling high-dimensional hierarchical uncertainty quantification by anova and tensor-train decomposition. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(1):63–76, 2014.

[23] Alexander Novikov, Pavel Izmailov, Valentin Khrulkov, Michael Figurnov, and Ivan V Oseledets. Tensor train decomposition on tensorflow (t3f). *Journal of Machine Learning Research*, 21(30):1–7, 2020.

[24] Alexander Novikov, Dmitrii Podoprikhin, Anton Osokin, and Dmitry P Vetrov. Tensorizing neural networks. In *Advances in neural information processing systems*, pages 442–450, 2015.

[25] John J Hopfield and David W Tank. "neural" computation of decisions in optimization problems. *Biological cybernetics*, 52(3):141–152, 1985.

[26] Norbeto M Grzywacz and Alan L Yuille. Motion correspondence and analog networks. In *AIP Conference Proceedings*, volume 151, pages 200–205. American Institute of Physics, 1986.

[27] Carsten Peterson and Bo Söderberg. A new method for mapping optimization problems onto neural networks. *International Journal of Neural Systems*, 1(01):3–22, 1989.

[28] Kenneth Rose, Eitan Gurewitz, and Geoffrey Fox. A deterministic annealing approach to clustering. *Pattern Recognition Letters*, 11(9):589–594, 1990.

[29] JC Platt and JJ Hopfield. Analog decoding using neural networks. In *AIP Conference Proceedings*, volume 151, pages 364–369. American Institute of Physics, 1986.

[30] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[31] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

[32] Ibrahim Mohammad Elfadel. From random fields to networks. 1993.

[33] Alan L Yuille and JJ Kosowsky. Statistical physics algorithms that converge. *Neural computation*, 6(3): 341–356, 1994.

[34] Dave A Bayer and Jeffrey C Lagarias. The nonlinear geometry of linear programming. i. affine and projective scaling trajectories. *Transactions of the American Mathematical Society*, 314(2):499–526, 1989.

[35] Dave A Bayer and Jeffrey C Lagarias. The nonlinear geometry of linear programming. ii. legendre transform coordinates and central trajectories. *Transactions of the American Mathematical Society*, 314(2): 527–581, 1989.

[36] Margaret H Wright. Interior methods for constrained optimization. *Acta numerica*, 1(1):341–407, 1992.

[37] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311, 1984.

[38] Leonid Faybusovich. Dynamical systems which solve optimization problems with linear constraints. *IMA Journal of Mathematical Control and Information*, 8(2):135–149, 1991.

[39] Renato DC Monteiro and Ilan Adler. Interior path following primal-dual algorithms. part i: Linear programming. *Mathematical programming*, 44(1-3):27–41, 1989.

[40] Renato DC Monteiro and Ilan Adler. Interior path following primal-dual algorithms. part ii: Convex quadratic programming. *Mathematical Programming*, 44(1-3):43–66, 1989.

[41] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

[42] Peter Grassberger and Jean-Pierre Nadal. *From statistical physics to statistical inference and back*, volume 428. Springer Science & Business Media, 2012.

[43] Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*. Number 111. Clarendon Press, 2001.

[44] Andreas Engel and Christian Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, 2001.

[45] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[46] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[47] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6): 066106, 2011.

[48] Florent Krzakała, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

[49] Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.

[50] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[51] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[52] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[53] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[54] Federica Gerace, Bruno Loureiro, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Generalisation error in learning with random features and the hidden manifold model. *arXiv preprint arXiv:2002.09339*, 2020.

[55] Lenka Zdeborová. Understanding deep learning is also a job for physicists. *Nature Physics*, pages 1–3, 2020.

[56] Li Deng and Dong Yu. Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387, 2014.

[57] Shoko Utsunomiya, Kenta Takata, and Yoshihisa Yamamoto. Mapping of ising models onto injection-locked laser systems. *Optics express*, 19(19):18091–18108, 2011.

[58] Alireza Marandi, Zhe Wang, Kenta Takata, Robert L Byer, and Yoshihisa Yamamoto. Network of time-multiplexed optical parametric oscillators as a coherent ising machine. *Nature Photonics*, 8(12):937, 2014.

[59] Takahiro Inagaki, Kensuke Inaba, Ryan Hamerly, Kyo Inoue, Yoshihisa Yamamoto, and Hiroki Takesue. Large-scale ising spin network based on degenerate optical parametric oscillators. *Nature Photonics*, 10 (6):415, 2016.

[60] Peter L McMahon, Alireza Marandi, Yoshitaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsunomiya, Kazuyuki Aihara, et al. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.

[61] Yutaka Takeda, Shuhei Tamate, Yoshihisa Yamamoto, Hiroki Takesue, Takahiro Inagaki, and Shoko Utsunomiya. Boltzmann sampling for an xy model using a non-degenerate optical parametric oscillator network. *Quantum Science and Technology*, 3(1):014004, 2017.

[62] Micha Nixon, Eitan Ronen, Asher A Friesem, and Nir Davidson. Observing geometric frustration with thousands of coupled lasers. *Physical review letters*, 110(18):184102, 2013.

[63] David Dung, Christian Kurtscheid, Tobias Damm, Julian Schmitt, Frank Vewinger, Martin Weitz, and Jan Klaers. Variable potentials for thermalized light and coupled condensates. *Nature Photonics*, 11(9): 565, 2017.

[64] Iulia Buluta, Sahel Ashhab, and Franco Nori. Natural and artificial atoms for quantum computation. *Reports on Progress in Physics*, 74(10):104401, 2011.

[65] Iulia M Georgescu, Sahel Ashhab, and Franco Nori. Quantum simulation. *Reviews of Modern Physics*, 86(1):153, 2014.

[66] Francisco Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

[67] Eugene L Lawler, Jan Karel Lenstra, AHG Rinnooy Kan, David Bernard Shmoys, et al. *The traveling salesman problem: a guided tour of combinatorial optimization*, volume 3. Wiley New York, 1985.

[68] Natalia G Berloff, Matteo Silva, Kirill Kalinin, Alexis Askitopoulos, Julian D Töpfer, Pasquale Cilibrizzi, Wolfgang Langbein, and Pavlos G Lagoudakis. Realizing the classical xy hamiltonian in polariton simulators. *Nature materials*, 16(11):1120, 2017.

[69] Kirill P Kalinin and Natalia G Berloff. Networks of non-equilibrium condensates for global optimization. *New Journal of Physics*, 20(11):113023, 2018.

[70] Eleanor Rieffel and Wolfgang Polak. An introduction to quantum computing for non-physicists. *ACM Computing Surveys (CSUR)*, 32(3):300–335, 2000.

[71] Andrew Steane. Quantum computing. *Reports on Progress in Physics*, 61(2):117, 1998.

[72] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[73] Michael Freedman, Alexei Kitaev, Michael Larsen, and Zhenghan Wang. Topological quantum computation. *Bulletin of the American Mathematical Society*, 40(1):31–38, 2003.

[74] Yvan Paquot, Francois Duport, Antoneo Smerieri, Joni Dambre, Benjamin Schrauwen, Marc Haelterman, and Serge Massar. Optoelectronic reservoir computing. *Scientific reports*, 2:287, 2012.

[75] Sanjib Ghosh, Andrzej Opala, Michał Matuszewski, Tomasz Paterek, and Timothy CH Liew. Quantum reservoir processing. *npj Quantum Information*, 5(1):1–6, 2019.

[76] Andrzej Opala, Sanjib Ghosh, Timothy CH Liew, and Michał Matuszewski. Neuromorphic computing in ginzburg-landau polariton-lattice systems. *Physical Review Applied*, 11(6):064029, 2019.

[77] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123, 2019.

[78] Guy Van der Sande, Daniel Brunner, and Miguel C Soriano. Advances in photonic reservoir computing. *Nanophotonics*, 6(3):561–576, 2017.

[79] Kirill P Kalinin and Natalia G Berloff. Gain-dissipative simulators for large-scale hard classical optimisation. *arXiv preprint arXiv:1805.01371*, 2018.

[80] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[81] Allan Griffin. A brief history of our understanding of bec: from bose to beliaev. *arXiv preprint cond-mat/9901123*, 1999.

[82] Kostas Gavroglu. *Fritz London: A scientific biography.* Cambridge University Press, 2005.

[83] Ėlevter Andronikashvili. *Reflections on liquid helium.* Amer Inst of Physics, 1989.

[84] Edward Philip Jurkowitz. Interpreting superconductivity: the history of quantum theory and the theory of superconductivity and superfluidity, 1933-1957. *PhDT*, 1995.

[85] Philip W Anderson. *A career in theoretical physics*, volume 7. World Scientific, 1994.

[86] Louis-Victor de Broglie. On the theory of quanta. *Ann. de Phys*, 3, 1924.

[87] Louis de Broglie. Xxxv. a tentative theory of light quanta. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 47(278):446–458, 1924.

[88] Dallin S Durfee and Wolfgang Ketterle. Experimental studies of bose-einstein condensation. *Optics Express*, 2(8):299–313, 1998.

[89] Mike H Anderson, Jason R Ensher, Michael R Matthews, Carl E Wieman, and Eric A Cornell. Observation of bose-einstein condensation in a dilute atomic vapor. *science*, 269(5221):198–201, 1995.

[90] Pyotr Kapitza. Viscosity of liquid helium below the λ-point. *Nature*, 141(3558):74–74, 1938.

[91] John F Allen and AD Misener. Flow of liquid helium ii. *Nature*, 141(3558):75–75, 1938.

[92] JC McLennan, HD Smith, and JO Wilhelm. Xiv. the scattering of light by liquid helium. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 14(89):161–167, 1932.

[93] Fritz London. The λ-phenomenon of liquid helium and the bose-einstein degeneracy. *Nature*, 141(3571): 643–644, 1938.

[94] Laszlo Tisza. Transport phenomena in helium ii. *Nature*, 141(3577):913–913, 1938.

[95] E Zaremba, T Nikuni, and A Griffin. Two-fluid hydrodynamics in trapped bose gases and in superfluid helium. *J. Low Temp. Phys*, 116:227, 1999.

[96] Fritz London. On the bose-einstein condensation. *Physical Review*, 54(11):947, 1938.

[97] L Tisza. Sur la théorie des liquides quantiques. application à l'hélium liquide. ii. *Journal de Physique et le Radium*, 1(8):350–358, 1940.

[98] Allan Griffin, Griffin Allan, et al. *Excitations in a Bose-condensed Liquid*, volume 4. Cambridge University Press, 1993.

[99] N Bogoliubov. On the theory of superfluidity. *J. Phys*, 11(1):23, 1947.

[100] Oliver Penrose. Cxxxvi. on the quantum mechanics of helium ii. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 42(335):1373–1377, 1951.

[101] Lev Landau. Das dämpfungsproblem in der wellenmechanik. *Zeitschrift für Physik*, 45(5-6):430–441, 1927.

[102] LD Landau. Two-fluid model of liquid helium ii. *J. Phys. Ussr*, 5(1):71–90, 1941.

[103] Liesbeth Venema, Bart Verberck, Iulia Georgescu, Giacomo Prando, Elsa Couderc, Silvia Milana, Maria Maragkou, Lina Persechini, Giulia Pacchioni, and Luke Fleet. The quasiparticle zoo. *Nature Physics*, 12 (12):1085–1089, 2016.

[104] JL Yarnell, GP Arnold, PJ Bendt, and EC Kerr. Excitations in liquid helium: neutron scattering measurements. *Physical Review*, 113(6):1379, 1959.

[105] DG Henshaw and ADB Woods. Modes of atomic motions in liquid helium by inelastic scattering of neutrons. *Physical Review*, 121(5):1266, 1961.

[106] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Quantum mechanics: non-relativistic theory*, volume 3. Elsevier, 2013.

[107] ST Beliaev. Energy spectrum of a non-ideal bose gas. *Sov. Phys. JETP*, 34(2):299, 1958.

[108] L Landau. On the theory of superfluidity. *Physical Review*, 75(5):884, 1949.

[109] John Bardeen. Ln cooper, and jr schrieffer. *Phys. Rev*, 108(1175):5, 1957.

[110] V Peshkov et al. The second sound in helium ii. *Helium*, 4:166, 2013.

[111] Oliver Penrose and Lars Onsager. Bose-einstein condensation and liquid helium. *Physical Review*, 104(3): 576, 1956.

[112] Richard Phillips Feynman. Atomic theory of the λ transition in helium. *Physical Review*, 91(6):1291, 1953.

[113] Richard P Feynman. Atomic theory of the two-fluid model of liquid helium. *Physical Review*, 94(2):262, 1954.

[114] NM Hugenholtz and David Pines. Ground-state energy and excitation spectrum of a system of interacting bosons. *Physical Review*, 116(3):489, 1959.

[115] J Gavoret and Ph Nozieres. Structure of the perturbation expansion for the bose liquid at zero temperature. *Annals of Physics*, 28(3):349–399, 1964.

[116] PC Hohenberg and PC Martin. Microscopic theory of superfluid helium. *Annals of Physics*, 34(2):291–359, 1965.

[117] Lev P Pitaevskii. Vortex lines in an imperfect bose gas. *Sov. Phys. JETP*, 13(2):451–454, 1961.

[118] Vitaly L Ginzburg and Lev D Landau. On the theory of superconductivity. In *On Superconductivity and Superfluidity*, pages 113–137. Springer, 2009.

[119] LP Gor'Kov. On the energy spectrum of superconductors. *Sov. Phys. JETP*, 7(505):158, 1958.

[120] Bose-Einstein Condensation. A. griffin, dw snoke, and s. stringari. *Cambridge University Press*, 2:281, 1995.

[121] Qijin Chen, Jelena Stajic, Shina Tan, and Kathryn Levin. Bcs–bec crossover: From high temperature superconductors to ultracold superfluids. *Physics Reports*, 412(1):1–88, 2005.

[122] Kendall B Davis, M-O Mewes, Michael R Andrews, Nicolaas J van Druten, Dallin S Durfee, DM Kurn, and Wolfgang Ketterle. Bose-einstein condensation in a gas of sodium atoms. *Physical review letters*, 75 (22):3969, 1995.

[123] Cl C Bradley, CA Sackett, JJ Tollett, and Randall G Hulet. Evidence of bose-einstein condensation in an atomic gas with attractive interactions. *Physical review letters*, 75(9):1687, 1995.

[124] Anthony J Leggett. Bose-einstein condensation in the alkali gases: Some fundamental concepts. *Reviews of Modern Physics*, 73(2):307, 2001.

[125] Jan Klaers, Julian Schmitt, Frank Vewinger, and Martin Weitz. Bose–einstein condensation of photons in an optical microcavity. *Nature*, 468(7323):545–548, 2010.

[126] Monique Combescot and Shiue-Yuan Shiau. *Excitons and Cooper pairs: two composite bosons in many-body physics*. Oxford University Press, 2015.

[127] Zachary Dutton, Naomi S Ginsberg, Christopher Slowe, and Lene Vestergaard Hau. The art of taming light: ultra-slow and stopped light. *Europhysics news*, 35(2):33–39, 2004.

[128] Barbara Goss Levi. From superfluid to insulator: Bose-einstein condensate undergoes a quantum phase transition. *Physics Today*, 55(3):18–20, 2002.

[129] Jun-Ru Li, Jeongwon Lee, Wujie Huang, Sean Burchesky, Boris Shteynas, Furkan Çağrı Top, Alan O Jamison, and Wolfgang Ketterle. A stripe phase with supersolid properties in spin–orbit-coupled bose–einstein condensates. *Nature*, 543(7643):91–94, 2017.

[130] Colin J Kennedy, William Cody Burton, Woo Chang Chung, and Wolfgang Ketterle. Observation of bose–einstein condensation in a strong synthetic magnetic field. *Nature Physics*, 11(10):859–864, 2015.

[131] Matthias Schmitt, Matthias Wenzel, Fabian Böttcher, Igor Ferrier-Barbut, and Tilman Pfau. Self-bound droplets of a dilute magnetic quantum liquid. *Nature*, 539(7628):259–262, 2016.

[132] Elizabeth A Donley, Neil R Claussen, Simon L Cornish, Jacob L Roberts, Eric A Cornell, and Carl E Wieman. Dynamics of collapsing and exploding bose–einstein condensates. *Nature*, 412(6844):295–299, 2001.

[133] Eiichi Hanamura and Hartmut Haug. Condensation effects of excitons. *Physics Reports*, 33(4):209–284, 1977.

[134] Hui Deng, Hartmut Haug, and Yoshihisa Yamamoto. Exciton-polariton bose-einstein condensation. *Reviews of Modern Physics*, 82(2):1489, 2010.

[135] JJ Hopfield. Theory of the contribution of excitons to the complex dielectric constant of crystals. *Physical Review*, 112(5):1555, 1958.

[136] Robert H Dicke. Coherence in spontaneous radiation processes. *Physical review*, 93(1):99, 1954.

[137] Klaus Hepp and Elliott H Lieb. Equilibrium statistical mechanics of matter interacting with the quantized radiation field. *Physical Review A*, 8(5):2517, 1973.

[138] Peter Kirton, Mor M Roses, Jonathan Keeling, and Emanuele G Dalla Torre. Introduction to the dicke model: From equilibrium to nonequilibrium, and vice versa. *Advanced Quantum Technologies*, 2(1-2): 1800043, 2019.

[139] Paul Roger Eastham and PB Littlewood. Bose condensation in a model microcavity. *Solid state communications*, 116(7):357–361, 2000.

[140] PR Eastham and PB Littlewood. Bose condensation of cavity polaritons beyond the linear regime: The thermal equilibrium of a model microcavity. *Physical Review B*, 64(23):235101, 2001.

[141] Jonathan Keeling, PR Eastham, MH Szymanska, and PB Littlewood. Bcs-bec crossover in a system of microcavity polaritons. *Physical Review B*, 72(11):115320, 2005.

[142] PB Littlewood, PR Eastham, JMJ Keeling, FM Marchetti, BD Simons, and MH Szymanska. Models of coherent exciton condensation. *Journal of Physics: Condensed Matter*, 16(35):S3597, 2004.

[143] Amir O Caldeira and Anthony J Leggett. Quantum tunnelling in a dissipative system. *Annals of physics*, 149(2):374–456, 1983.

[144] Goran Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, 1976.

[145] Vittorio Gorini, Andrzej Kossakowski, and Ennackal Chandy George Sudarshan. Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17(5):821–825, 1976.

[146] Artem Strashko, Francesca M Marchetti, Allan H MacDonald, and Jonathan Keeling. Crescent states in charge-imbalanced polariton condensates. *Physical Review Letters*, 125(6):067405, 2020.

[147] Artem Strashko, Peter Kirton, and Jonathan Keeling. Organic polariton lasing and the weak to strong coupling crossover. *Physical review letters*, 121(19):193601, 2018.

[148] Kristin B Arnardottir, Antti J Moilanen, Artem Strashko, Päivi Törmä, and Jonathan Keeling. Multimode organic polariton lasing. *Physical Review Letters*, 125(23):233603, 2020.

[149] Peter Kirton and Jonathan Keeling. Nonequilibrium model of photon condensation. *Physical review letters*, 111(10):100404, 2013.

[150] Peter Kirton and Jonathan Keeling. Thermalization and breakdown of thermalization in photon condensates. *Physical Review A*, 91(3):033826, 2015.

[151] Jesper Levinsen, Guangyao Li, and Meera M Parish. Microscopic description of exciton-polaritons in microcavities. *Physical Review Research*, 1(3):033120, 2019.

[152] Miguel A Bastarrachea-Magnani, Arturo Camacho-Guardian, and Georg M Bruun. Attractive and repulsive exciton-polariton interactions mediated by an electron gas. *Physical Review Letters*, 126(12):127405, 2021.

[153] Dmitry K Efimkin, Emma K Laird, Jesper Levinsen, Meera M Parish, and Allan H MacDonald. Electron-exciton interactions in the exciton-polaron problem. *Physical Review B*, 103(7):075417, 2021.

[154] A Imamog, RJ Ram, S Pau, Y Yamamoto, et al. Nonequilibrium condensates and lasers without inversion: Exciton-polariton lasers. *Physical Review A*, 53(6):4250, 1996.

[155] Stanley Pau, Hui Cao, Joseph Jacobson, Gunnar Björk, Yoshihisa Yamamoto, and Atac Imamoğlu. Observation of a laserlike transition in a microcavity exciton polariton system. *Physical Review A*, 54(3): R1789, 1996.

[156] H Cao, S Pau, JM Jacobson, G Björk, Y Yamamoto, and A Imamoglu. Transition from a microcavity exciton polariton to a photon laser. *Physical Review A*, 55(6):4632, 1997.

[157] Hui Deng, Gregor Weihs, Charles Santori, Jacqueline Bloch, and Yoshihisa Yamamoto. Condensation of semiconductor microcavity exciton polaritons. *Science*, 298(5591):199–202, 2002.

[158] Jacek Kasprzak, M Richard, S Kundermann, A Baas, P Jeambrun, JMJ Keeling, FM Marchetti, MH Szymańska, R André, JL Staehli, et al. Bose–einstein condensation of exciton polaritons. *Nature*, 443(7110): 409–414, 2006.

[159] Jacek Kasprzak. *Condensation of exciton polaritons*. PhD thesis, 2006.

[160] R Balili, V Hartwell, D Snoke, L Pfeiffer, and K West. Bose-einstein condensation of microcavity polaritons in a trap. *Science*, 316(5827):1007–1010, 2007.

[161] Esther Wertz, Lydie Ferrier, Dmitry D Solnyshkov, Pascale Senellart, Daniele Bajoni, Audrey Miard, Aristide Lemaître, Guillaume Malpuech, and Jacqueline Bloch. Spontaneous formation of a polariton condensate in a planar gaas microcavity. *Applied Physics Letters*, 95(5):051108, 2009.

[162] Alberto Amo, Jérôme Lefrère, Simon Pigeon, Claire Adrados, Cristiano Ciuti, Iacopo Carusotto, Romuald Houdré, Elisabeth Giacobino, and Alberto Bramati. Superfluidity of polaritons in semiconductor microcavities. *Nature Physics*, 5(11):805–810, 2009.

[163] Giovanni Lerario, Antonio Fieramosca, Fábio Barachati, Dario Ballarini, Konstantinos S Daskalakis, Lorenzo Dominici, Milena De Giorgi, Stefan A Maier, Giuseppe Gigli, Stéphane Kéna-Cohen, et al. Room-temperature superfluidity in a polariton condensate. *Nature Physics*, 13(9):837–841, 2017.

[164] S Christopoulos, G Baldassarri Höger Von Högersthal, AJD Grundy, PG Lagoudakis, AV Kavokin, JJ Baumberg, G Christmann, R Butté, E Feltin, J-F Carlin, et al. Room-temperature polariton lasing in semiconductor microcavities. *Physical review letters*, 98(12):126405, 2007.

[165] JJ Baumberg, AV Kavokin, S Christopoulos, AJD Grundy, R Butté, G Christmann, DD Solnyshkov, G Malpuech, G Baldassarri Höger von Högersthal, El Feltin, et al. Spontaneous polarization buildup in a room-temperature polariton laser. *Physical Review Letters*, 101(13):136409, 2008.

[166] S Kéna-Cohen and SR Forrest. Room-temperature polariton lasing in an organic single-crystal microcavity. *Nature Photonics*, 4(6):371, 2010.

[167] John Michael Kosterlitz and David James Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6(7):1181, 1973.

[168] Markus Greiner, Olaf Mandel, Tilman Esslinger, Theodor W Hänsch, and Immanuel Bloch. Quantum phase transition from a superfluid to a mott insulator in a gas of ultracold atoms. *nature*, 415(6867): 39–44, 2002.

[169] Iulia Buluta and Franco Nori. Quantum simulators. *Science*, 326(5949):108–111, 2009.

[170] Tim Byrnes, Na Young Kim, and Yoshihisa Yamamoto. Exciton–polariton condensates. *Nature Physics*, 10(11):803–813, 2014.

[171] Kirill P Kalinin and Natalia G Berloff. Simulating ising and potts models and external fields with non-equilibrium condensates. *arXiv preprint arXiv:1806.11457*, 2018.

[172] Zhe Wang, Alireza Marandi, Kai Wen, Robert L Byer, and Yoshihisa Yamamoto. Coherent ising machine based on degenerate optical parametric oscillators. *Physical Review A*, 88(6):063853, 2013.

[173] Hayato Goto, Kosuke Tatsumura, and Alexander R Dixon. Combinatorial optimization by simulating adiabatic bifurcations in nonlinear hamiltonian systems. *Science advances*, 5(4):eaav2372, 2019.

[174] Timothée Leleu, Yoshihisa Yamamoto, Peter L McMahon, and Kazuyuki Aihara. Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity. *Physical review letters*, 122(4): 040607, 2019.

[175] Ignacio Rozada, Maliheh Aramon, Jonathan Machta, and Helmut G Katzgraber. Effects of setting the temperatures in the parallel tempering monte carlo algorithm. *arXiv preprint arXiv:1907.03906*, 2019.

[176] Fuxi Cai, Suhas Kumar, Thomas Van Vaerenbergh, Rui Liu, Can Li, Shimeng Yu, Qiangfei Xia, J Joshua Yang, Raymond Beausoleil, Wei Lu, et al. Harnessing intrinsic noise in memristor hopfield neural networks for combinatorial optimization. *arXiv preprint arXiv:1903.11194*, 2019.

[177] Ryan Hamerly, Takahiro Inagaki, Peter L McMahon, Davide Venturelli, Alireza Marandi, Tatsuhiro Onodera, Edwin Ng, Carsten Langrock, Kensuke Inaba, Toshimori Honjo, et al. Experimental investigation of performance differences between coherent ising machines and a quantum annealer. *arXiv preprint arXiv:1805.05217*, 2018.

[178] David Kielpinski, Ranojoy Bose, Jason Pelc, Thomas Van Vaerenbergh, Gabriel Mendoza, Nikolas Tezak, and Raymond G Beausoleil. Information processing with large-scale optical integrated circuits. In *2016 IEEE International Conference on Rebooting Computing (ICRC)*, pages 1–4. IEEE, 2016.

[179] Nikolas Tezak, Thomas Van Vaerenbergh, Jason S Pelc, Gabriel J Mendoza, David Kielpinski, Hideo Mabuchi, and Raymond G Beausoleil. Integrated coherent ising machines based on self-phase modulation in microring resonators. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(1):1–15, 2019.

[180] Maciej Lewenstein, Anna Sanpera, Veronica Ahufinger, Bogdan Damski, Aditi Sen, and Ujjwal Sen. Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond. *Advances in Physics*, 56(2):243–379, 2007.

[181] Mark Saffman, Thad G Walker, and Klaus Mølmer. Quantum information with rydberg atoms. *Reviews of modern physics*, 82(3):2313, 2010.

[182] Jonathan Simon, Waseem S Bakr, Ruichao Ma, M Eric Tai, Philipp M Preiss, and Markus Greiner. Quantum simulation of antiferromagnetic spin chains in an optical lattice. *Nature*, 472(7343):307–312, 2011.

[183] Tilman Esslinger. Fermi-hubbard physics with atoms in an optical lattice. 2010.

[184] TE Northup and R Blatt. Quantum information transfer using photons. *Nature photonics*, 8(5):356–363, 2014.

[185] Kihwan Kim, M-S Chang, Simcha Korenblit, Rajibul Islam, Emily E Edwards, James K Freericks, G-D Lin, L-M Duan, and Christopher Monroe. Quantum simulation of frustrated ising spins with trapped ions. *Nature*, 465(7298):590–593, 2010.

[186] Ben P Lanyon, Cornelius Hempel, Daniel Nigg, Markus Müller, Rene Gerritsma, F Zähringer, Philipp Schindler, Julio T Barreiro, Markus Rambach, Gerhard Kirchmair, et al. Universal digital quantum simulation with trapped ions. *Science*, 334(6052):57–61, 2011.

[187] Antonio D Córcoles, Easwar Magesan, Srikanth J Srinivasan, Andrew W Cross, Matthias Steffen, Jay M Gambetta, and Jerry M Chow. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nature communications*, 6(1):1–10, 2015.

[188] Pavlos G Lagoudakis and Natalia G Berloff. A polariton graph simulator. *New Journal of Physics*, 19 (12):125008, 2017.

[189] Julian Struck, Christoph Ölschläger, R Le Targat, Parvis Soltan-Panahi, André Eckardt, Maciej Lewenstein, Patrick Windpassinger, and Klaus Sengstock. Quantum simulation of frustrated classical magnetism in triangular optical lattices. *Science*, 333(6045):996–999, 2011.

[190] Gemma De las Cuevas and Toby S Cubitt. Simple universal models capture all classical spin physics. *Science*, 351(6278):1180–1183, 2016.

[191] Panos M Pardalos and Stephen A Vavasis. Quadratic programming with one negative eigenvalue is np-hard. *Journal of Global optimization*, 1(1):15–22, 1991.

[192] Kirill P Kalinin and Natalia G Berloff. Toward arbitrary control of lattice interactions in nonequilibrium condensates. *Advanced Quantum Technologies*, 3(2):1900065, 2020.

[193] Francisco A Rodrigues, Thomas K DM Peron, Peng Ji, and Jürgen Kurths. The kuramoto model in complex networks. *Physics Reports*, 610:1–98, 2016.

[194] Simai He, Zhening Li, and Shuzhong Zhang. Approximation algorithms for discrete polynomial optimization. *Journal of the Operations Research Society of China*, 1(1):3–36, 2013.

[195] Bo Jiang, Zhening Li, and Shuzhong Zhang. Approximation methods for complex polynomial optimization. *Computational Optimization and Applications*, 59(1-2):219–248, 2014.

[196] Michiel Wouters and Iacopo Carusotto. Excitations in a nonequilibrium bose-einstein condensate of exciton polaritons. *Physical review letters*, 99(14):140402, 2007.

[197] Jonathan Keeling and Natalia G Berloff. Spontaneous rotating vortex lattices in a pumped decaying condensate. *Physical review letters*, 100(25):250401, 2008.

[198] Iacopo Carusotto and Cristiano Ciuti. Quantum fluids of light. *Reviews of Modern Physics*, 85(1):299, 2013.

[199] Jonathan Keeling and Natalia G Berloff. Exciton–polariton condensation. *Contemporary Physics*, 52(2): 131–151, 2011.

[200] Augusto Smerzi and Andrea Trombettoni. Nonlinear tight-binding approximation for bose-einstein condensates in a lattice. *Physical Review A*, 68(2):023613, 2003.

[201] Kirill P Kalinin, Pavlos G Lagoudakis, and Natalia G Berloff. Matter wave coupling of spatially separated and unequally pumped polariton condensates. *Physical Review B*, 97(9):094512, 2018.

[202] Hamid Ohadi, RL Gregory, T Freegarde, YG Rubo, AV Kavokin, Natalia G Berloff, and PG Lagoudakis. Nontrivial phase coupling in polariton multiplets. *Physical Review X*, 6(3):031032, 2016.

[203] Fabio Schittler Neves and Marc Timme. Computation by switching in complex networks of states. *Physical review letters*, 109(1):018701, 2012.

[204] Kirill P Kalinin and Natalia G Berloff. Polaritonic network as a paradigm for dynamics of coupled oscillators. *arXiv preprint arXiv:1902.09142*, 2019.

[205] Sergey Alyatkin, Julian D Töpfer, Alexis Askitopoulos, Helgi Sigurdsson, and Pavlos G Lagoudakis. Optical control of synchronous phases in a programmable polariton cell. *arXiv preprint arXiv:1907.08580*, 2019.

[206] Gonzalo Joya, MA Atencia, and Francisco Sandoval. Hopfield neural networks for optimization: study of the different dynamics. *Neurocomputing*, 43(1-4):219–237, 2002.

[207] Yan Xue, Igor Chestnov, Evgeny Sedov, Evgeniy Kiktenko, Aleksey K Fedorov, Stefan Schumacher, Xuekai Ma, and Alexey Kavokin. Split-ring polariton condensates as macroscopic two-level quantum systems. *Physical Review Research*, 3(1):013099, 2021.

[208] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.

[209] Sanjib Ghosh and Timothy CH Liew. Quantum computing with exciton-polariton condensates. *npj Quantum Information*, 6(1):1–6, 2020.

[210] Paul M Chaikin, Tom C Lubensky, and Thomas A Witten. *Principles of condensed matter physics*, volume 10. Cambridge university press Cambridge, 1995.

[211] JM Kosterlitz. The critical properties of the two-dimensional xy model. *Journal of Physics C: Solid State Physics*, 7(6):1046, 1974.

[212] Rajan Gupta, Jerry DeLapp, George G Batrouni, Geoffrey C Fox, Clive F Baillie, and John Apostolakis. Phase transition in the 2 d xy model. *Physical review letters*, 61(17):1996, 1988.

[213] P Gawiec and DR Grempel. Numerical study of the ground-state properties of a frustrated xy model. *Physical Review B*, 44(6):2613, 1991.

[214] JM Kosterlitz and N Akino. Numerical study of spin and chiral order in a two-dimensional xy spin glass. *Physical review letters*, 82(20):4094, 1999.

[215] Alán Aspuru-Guzik and Philip Walther. Photonic quantum simulators. *Nature physics*, 8(4):285–291, 2012.

[216] Morten Kjaergaard, Mollie E Schwartz, Jochen Braumüller, Philip Krantz, Joel I-J Wang, Simon Gustavsson, and William D Oliver. Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics*, 11:369–395, 2020.

[217] Philipp Hauke, Fernando M Cucchietti, Luca Tagliacozzo, Ivan Deutsch, and Maciej Lewenstein. Can one trust quantum simulators? *Reports on Progress in Physics*, 75(8):082401, 2012.

[218] M De Giorgi, D Ballarini, E Cancellieri, FM Marchetti, MH Szymanska, C Tejedor, R Cingolani, E Giacobino, A Bramati, G Gigli, et al. Control and ultrafast dynamics of a two-fluid polariton switch. *Physical review letters*, 109(26):266407, 2012.

[219] Félix Marsault, Hai Son Nguyen, Dimitrii Tanese, Aristide Lemaître, Elisabeth Galopin, Isabelle Sagnes, Alberto Amo, and Jacqueline Bloch. Realization of an all optical exciton-polariton router. *Applied Physics Letters*, 107(20):201115, 2015.

[220] Tingge Gao, PS Eldridge, Timothy Chi Hin Liew, SI Tsintzos, George Stavrinidis, G Deligeorgis, Z Hatzopoulos, and PG Savvidis. Polariton condensate transistor switch. *Physical Review B*, 85(23):235102, 2012.

[221] Dario Ballarini, Milena De Giorgi, Emiliano Cancellieri, Romuald Houdré, Elisabeth Giacobino, Roberto Cingolani, Alberto Bramati, Giuseppe Gigli, and Daniele Sanvitto. All-optical polariton transistor. *Nature communications*, 4:1778, 2013.

[222] Anton V Zasedatelev, Anton V Baranikov, Darius Urbonas, Fabio Scafirimuto, Ullrich Scherf, Thilo Stöferle, Rainer F Mahrt, and Pavlos G Lagoudakis. A room-temperature organic polariton transistor. *Nature Photonics*, 13(6):378, 2019.

[223] Alberto Amo, TCH Liew, Claire Adrados, Romuald Houdré, Elisabeth Giacobino, AV Kavokin, and A Bramati. Exciton–polariton spin switches. *Nature Photonics*, 4(6):361, 2010.

[224] Regine Frank. Coherent control of floquet-mode dressed plasmon polaritons. *Physical Review B*, 85(19): 195463, 2012.

[225] Regine Frank. Non-equilibrium polaritonics-non-linear effects and optical switching. *Annalen der Physik*, 525(1-2):66–73, 2013.

[226] Andrew N Pargellis, Susannah Green, and Bernard Yurke. Planar xy-model dynamics in a nematic liquid crystal system. *Physical Review E*, 49(5):4250, 1994.

[227] Julian Struck, Malte Weinberg, Christoph Ölschläger, Patrick Windpassinger, Juliette Simonet, Klaus Sengstock, Robert Höppner, Philipp Hauke, André Eckardt, Maciej Lewenstein, et al. Engineering ising-xy spin-models in a triangular lattice using tunable artificial gauge fields. *Nature Physics*, 9(11):738–743, 2013.

[228] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

[229] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

[230] C Reid Turner, Alfonso Fuggetta, Luigi Lavazza, and Alexander L Wolf. A conceptual basis for feature engineering. *Journal of Systems and Software*, 49(1):3–15, 1999.

[231] John G Taylor and Mark D Plumbley. Information theory and neural networks. In *North-Holland Mathematical Library*, volume 51, pages 307–340. Elsevier, 1993.

[232] Colin Fyfe. Artificial neural networks and information theory. *Department of Computing and Information System, The university of Paisley*, 2000.

[233] Kirill P Kalinin, Alberto Amo, Jacqueline Bloch, and Natalia G Berloff. Polaritonic xy-ising machine. *Nanophotonics*, 9(13):4127–4138, 2020.

[234] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[235] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, pages 843–852, 2017.

[236] Kenji Kawaguchi, Jiaoyang Huang, and Leslie Pack Kaelbling. Effect of depth and width on local minima in deep learning. *Neural computation*, 31(7):1462–1498, 2019.

[237] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In *international conference on machine learning*, pages 2847–2854. PMLR, 2017.

[238] Ronen Eldan and Ohad Shamir. The power of depth for feedforward neural networks. In *Conference on learning theory*, pages 907–940, 2016.

[239] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[240] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.

[241] Hamid Ohadi, Y del Valle-Inclan Redondo, A Dreismann, YG Rubo, F Pinsker, SI Tsintzos, Z Hatzopoulos, PG Savvidis, and JJ Baumberg. Tunable magnetic alignment between trapped exciton-polariton condensates. *Physical Review Letters*, 116(10):106403, 2016.

[242] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[243] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[244] Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34):13, 2001.

[245] Vasileios Athanasiou and Zoran Konkoli. On mathematics of universal computation with generic dynamical systems., 2019.

[246] Zoran Konkoli. On reservoir computing: from mathematical foundations to unconventional applications. In *Advances in unconventional computing*, pages 573–607. Springer, 2017.

[247] Peter F Dominey. Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning. *Biological cybernetics*, 73(3):265–274, 1995.

[248] Peter Ford Dominey. From sensorimotor sequence to grammatical construction: Evidence from simulation and neurophysiology. *Adaptive Behavior*, 13(4):347–361, 2005.

[249] Peter Ford Dominey and Franck Ramus. Neural network processing of natural language: I. sensitivity to serial, temporal and abstract structure of language in the infant. *Language and Cognitive Processes*, 15 (1):87–127, 2000.

[250] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[251] Mantas Lukoševičius, Herbert Jaeger, and Benjamin Schrauwen. Reservoir computing trends. *KI-Künstliche Intelligenz*, 26(4):365–371, 2012.

[252] Benjamin Schrauwen, David Verstraeten, and Jan Van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th european symposium on artificial neural networks. p. 471-482 2007*, pages 471–482, 2007.

[253] David Verstraeten, Benjamin Schrauwen, Michiel d'Haene, and Dirk Stroobandt. An experimental unification of reservoir computing methods. *Neural networks*, 20(3):391–403, 2007.

[254] Miguel C Soriano, Daniel Brunner, Miguel Escalona-Morán, Claudio R Mirasso, and Ingo Fischer. Minimal approach to neuro-inspired information processing. *Frontiers in computational neuroscience*, 9:68, 2015.

[255] Jihan Zhu and Peter Sutton. Fpga implementations of neural networks–a survey of a decade of progress. In *International Conference on Field Programmable Logic and Applications*, pages 1062–1066. Springer, 2003.

[256] Amos R Omondi and Jagath Chandana Rajapakse. *FPGA implementations of neural networks*, volume 365. Springer, 2006.

[257] Felix Schürmann, Karlheinz Meier, and Johannes Schemmel. Edge of chaos computation in mixed-mode vlsi-a hard liquid. In *Advances in neural information processing systems*, pages 1201–1208, 2005.

[258] Florian Walter, Florian Röhrbein, and Alois Knoll. Neuromorphic implementations of neurobiological learning algorithms for spiking neural networks. *Neural Networks*, 72:152–167, 2015.

[259] Laurent Larger, Miguel C Soriano, Daniel Brunner, Lennert Appeltant, Jose M Gutiérrez, Luis Pesquera, Claudio R Mirasso, and Ingo Fischer. Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing. *Optics express*, 20(3):3241–3249, 2012.

[260] François Duport, Bendix Schneider, Anteo Smerieri, Marc Haelterman, and Serge Massar. All-optical reservoir computing. *Optics express*, 20(20):22783–22795, 2012.

[261] Lennert Appeltant, Miguel Cornelles Soriano, Guy Van der Sande, Jan Danckaert, Serge Massar, Joni Dambre, Benjamin Schrauwen, Claudio R Mirasso, and Ingo Fischer. Information processing using a single dynamical node as complex system. *Nature communications*, 2(1):1–6, 2011.

[262] Daniel Brunner, Miguel Cornelles Soriano, and Ingo Fischer. High-speed optical vector and matrix operations using a semiconductor laser. *IEEE Photonics Technology Letters*, 25(17):1680–1683, 2013.

[263] Jacob Torrejon, Mathieu Riou, Flavio Abreu Araujo, Sumito Tsunegi, Guru Khalsa, Damien Querlioz, Paolo Bortolotti, Vincent Cros, Kay Yakushiji, Akio Fukushima, et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature*, 547(7664):428–431, 2017.

[264] Helmut Hauser, Auke J Ijspeert, Rudolf M Füchslin, Rolf Pfeifer, and Wolfgang Maass. Towards a theoretical foundation for morphological computation with compliant bodies. *Biological cybernetics*, 105 (5-6):355–370, 2011.

[265] Dean V Buonomano and Wolfgang Maass. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2):113–125, 2009.

[266] Oliver Obst, Adrian Trinchi, Simon G Hardin, Matthew Chadwick, Ivan Cole, Tim H Muster, Nigel Hoschke, Diet Ostry, Don Price, Khoa N Pham, et al. Nano-scale reservoir computing. *Nano Communication Networks*, 4(4):189–196, 2013.

[267] Kohei Nakajima, Keisuke Fujii, Makoto Negoro, Kosuke Mitarai, and Masahiro Kitagawa. Boosting computational power through spatial multiplexing in quantum reservoir computing. *Physical Review Applied*, 11(3):034021, 2019.

[268] Sanjib Ghosh, Tanjung Krisnanda, Tomasz Paterek, and Timothy CH Liew. Universal quantum reservoir computing. *arXiv preprint arXiv:2003.09569*, 2020.

[269] S Lepri, G Giacomelli, A Politi, and FT Arecchi. High-dimensional chaos in delayed dynamical systems. *Physica D: Nonlinear Phenomena*, 70(3):235–249, 1994.

[270] Stephen Wolfram. *Cellular automata and complexity: collected papers*. CRC Press, 2018.

[271] Jean C Coulombe, Mark CA York, and Julien Sylvestre. Computing with networks of nonlinear mechanical oscillators. *PloS one*, 12(6):e0178663, 2017.

[272] I Bloch, Th W Hänsch, and T Esslinger. Measurement of the spatial coherence of a trapped bose gas at the phase transition. *Nature*, 403(6766):166, 2000.

[273] EA Burt, RW Ghrist, CJ Myatt, MJ Holland, Eric A Cornell, and CE Wieman. Coherence, correlations, and collisions: What one learns about bose-einstein condensates from their decay. *Physical Review Letters*, 79(3):337, 1997.

[274] R Balili, V Hartwell, D Snoke, L Pfeiffer, and K West. Bose-einstein condensation of microcavity polaritons in a trap. *Science*, 316(5827):1007–1010, 2007.

[275] Alberto Amo, Jérôme Lefrère, Simon Pigeon, Claire Adrados, Cristiano Ciuti, Iacopo Carusotto, Romuald Houdré, Elisabeth Giacobino, and Alberto Bramati. Superfluidity of polaritons in semiconductor microcavities. *Nature Physics*, 5(11):805, 2009.

[276] A Amo, D Sanvitto, FP Laussy, D Ballarini, E Del Valle, MD Martin, A Lemaitre, J Bloch, DN Krizhanovskii, MS Skolnick, et al. Collective fluid dynamics of a polariton condensate in a semiconductor microcavity. *Nature*, 457(7227):291, 2009.

[277] Konstantinos G Lagoudakis, Michiel Wouters, Maxime Richard, Augustin Baas, Iacopo Carusotto, Regis André, Le Si Dang, and B Deveaud-Plédran. Quantized vortices in an exciton–polariton condensate. *Nature Physics*, 4(9):706, 2008.

[278] Torsten Karzig, Charles-Edouard Bardyn, Netanel H Lindner, and Gil Refael. Topological polaritons. *Physical Review X*, 5(3):031001, 2015.

[279] AV Nalitov, DD Solnyshkov, and G Malpuech. Polariton z topological insulator. *Physical review letters*, 114(11):116401, 2015.

[280] Andreas Lubatsch and Regine Frank. Behavior of floquet topological quantum states in optically driven semiconductors. *Symmetry*, 11(10):1246, 2019.

[281] Michiel Wouters, Iacopo Carusotto, and Cristiano Ciuti. Spatial and spectral shape of inhomogeneous nonequilibrium exciton-polariton condensates. *Physical Review B*, 77(11):115340, 2008.

[282] MH Szymańska, J Keeling, and PB Littlewood. Nonequilibrium quantum condensation in an incoherently pumped dissipative system. *Physical review letters*, 96(23):230602, 2006.

[283] Zheng Sun and David W Snoke. Optical switching with organics. *Nature Photonics*, 13(6):370–371, 2019.

[284] Anton Zasedatelev, Anton V Baranikov, Denis Sannikov, Darius Urbonas, Fabio Scafirimuto, Vladislav Yu Shishkov, Evgeny S Andrianov, Yurii E Lozovik, Ullrich Scherf, Thilo Stöferle, et al. Organic single-photon switch. *arXiv preprint arXiv:2005.05811*, 2020.

[285] Katarzyna Rechcińska, Mateusz Król, Rafał Mazur, Przemysław Morawiak, Rafał Mirek, Karolina Łempicka, Witold Bardyszewski, Michał Matuszewski, Przemysław Kula, Wiktor Piecek, et al. Engineering spin-orbit synthetic hamiltonians in liquid-crystal optical cavities. *Science*, 366(6466):727–730, 2019.

[286] Tomoki Ozawa, Hannah M Price, Alberto Amo, Nathan Goldman, Mohammad Hafezi, Ling Lu, Mikael C Rechtsman, David Schuster, Jonathan Simon, Oded Zilberberg, et al. Topological photonics. *Reviews of Modern Physics*, 91(1):015006, 2019.

[287] Kirill P Kalinin and Natalia G Berloff. Toward arbitrary control of lattice interactions in nonequilibrium condensates. *Advanced Quantum Technologies*, 3(2):1900065, 2020.

[288] Neil Na and Yoshihisa Yamamoto. Massive parallel generation of indistinguishable single photons via the polaritonic superfluid to mott-insulator quantum phase transition. *New Journal of Physics*, 12(12): 123001, 2010.

[289] D Tanese, H Flayac, D Solnyshkov, A Amo, A Lemaitre, E Galopin, R Braive, P Senellart, I Sagnes, G Malpuech, et al. Polariton condensation in solitonic gap states in a one-dimensional periodic potential. *Nature communications*, 4(1):1–9, 2013.

[290] P St-Jean, V Goblot, E Galopin, A Lemaître, T Ozawa, L Le Gratiet, I Sagnes, J Bloch, and A Amo. Lasing in topological edge states of a one-dimensional lattice. *Nature Photonics*, 11(10):651–656, 2017.

[291] S Mittal, J Fan, S Faez, A Migdall, JM Taylor, and M Hafezi. Topologically robust transport of photons in a synthetic gauge field. *Physical review letters*, 113(8):087403, 2014.

[292] Rui Su, Jun Wang, Jiaxin Zhao, Jun Xing, Weijie Zhao, Carole Diederichs, Timothy CH Liew, and Qihua Xiong. Room temperature long-range coherent exciton polariton condensate flow in lead halide perovskites. *Science advances*, 4(10):eaau0244, 2018.

[293] Rui Su, Sanjib Ghosh, Sheng Liu, Carole Diederichs, Timothy CH Liew, and Qihua Xiong. Observation of exciton polariton condensation in a perovskite lattice at room temperature. *arXiv preprint arXiv:1906.11566*, 2019.

[294] Ch Schneider, K Winkler, MD Fraser, M Kamp, Y Yamamoto, EA Ostrovskaya, and Sven Höfling. Exciton-polariton trapping and potential landscape engineering. *Reports on Progress in Physics*, 80(1): 016503, 2016.

[295] Daniele Sanvitto and Stéphane Kéna-Cohen. The road towards polaritonic devices. *Nature materials*, 15 (10):1061–1073, 2016.

[296] Sai Kiran Rajendran, Mengjie Wei, Hamid Ohadi, Arvydas Ruseckas, Graham A Turnbull, and Ifor DW Samuel. Low threshold polariton lasing from a solution-processed organic semiconductor in a planar microcavity. *Advanced Optical Materials*, page 1801791, 2019.

[297] Alan C Newell, Thierry Passot, and Joceline Lega. Order parameter equations for patterns. *Annual review of fluid mechanics*, 25(1):399–453, 1993.

[298] Len M Pismen and Len M Pismen. *Vortices in nonlinear fields: from liquid crystals to superfluids, from non-equilibrium patterns to cosmic strings*, volume 100. Oxford University Press, 1999.

[299] Mark C Cross and Pierre C Hohenberg. Pattern formation outside of equilibrium. *Reviews of modern physics*, 65(3):851, 1993.

[300] P Cristofolini, A Dreismann, G Christmann, G Franchetti, NG Berloff, P Tsotsis, Z Hatzopoulos, PG Savvidis, and JJ Baumberg. Optical superfluid phase transitions and trapping of polariton condensates. *Physical review letters*, 110(18):186403, 2013.

[301] Jared C Bronski, Lincoln D Carr, Ricardo Carretero-González, Bernard Deconinck, J Nathan Kutz, and Keith Promislow. Stability of attractive bose-einstein condensates in a periodic potential. *Physical Review E*, 64(5):056615, 2001.

[302] Jared C Bronski, Lincoln D Carr, Bernard Deconinck, J Nathan Kutz, and Keith Promislow. Stability of repulsive bose-einstein condensates in a periodic potential. *Physical Review E*, 63(3):036612, 2001.

[303] Sergey Alyatkin, JD Töpfer, Alexis Askitopoulos, Helgi Sigurdsson, and PG Lagoudakis. Optical control of couplings in polariton condensate lattices. *Physical Review Letters*, 124(20):207402, 2020.

[304] E Estrecho, T Gao, N Bobrovska, D Comber-Todd, MD Fraser, Mark Steger, K West, LN Pfeiffer, J Levinsen, MM Parish, et al. Direct measurement of polariton-polariton interaction strength in the thomas-fermi regime of exciton-polariton condensation. *Physical Review B*, 100(3):035306, 2019.

[305] Edwin KP Chong and Stanislaw H Zak. *An introduction to optimization*. John Wiley & Sons, 2004.

[306] Philip E Gill, Walter Murray, and Margaret H Wright. *Practical optimization*. SIAM, 2019.

[307] Andrew Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2:5, 2014.

[308] Zhi-Quan Luo, Wing-Kin Ma, Anthony Man-Cho So, Yinyu Ye, and Shuzhong Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

[309] Peng Wang, Chunhua Shen, Anton van den Hengel, and Philip HS Torr. Large-scale binary quadratic optimization using semidefinite relaxation and applications. *IEEE transactions on pattern analysis and machine intelligence*, 39(3):470–485, 2016.

[310] Gary Kochenberger, Jin-Kao Hao, Fred Glover, Mark Lewis, Zhipeng Lü, Haibo Wang, and Yang Wang. The unconstrained binary quadratic programming problem: a survey. *Journal of Combinatorial Optimization*, 28(1):58–81, 2014.

[311] Evelyn Fix. *Discriminatory analysis: nonparametric discrimination, consistency properties*, volume 1. USAF school of Aviation Medicine, 1985.

[312] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[313] VN Vapnik and A Ya Lerner. Recognition of patterns with help of generalized portraits. *Avtomat. i Telemekh*, 24(6):774–780, 1963.

[314] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[315] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. *J Mach Learn Res*, 10(66-71):13, 2009.

[316] Carlos Oscar Sánchez Sorzano, Javier Vargas, and A Pascual Montano. A survey of dimensionality reduction techniques. *arXiv preprint arXiv:1403.2877*, 2014.

[317] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Application of dimensionality reduction in recommender system-a case study. Technical report, Minnesota Univ Minneapolis Dept of Computer Science, 2000.

[318] Richard Jensen and Qiang Shen. Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Transactions on knowledge and data engineering*, 16(12):1457–1471, 2004.

[319] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[320] RV Mises and Hilda Pollaczek-Geiringer. Praktische verfahren der gleichungsauflösung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 9(1):58–77, 1929.

[321] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[322] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[323] Hugo Steinhaus. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci*, 1(804):801, 1956.

[324] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[325] Kevin Gurney. *An introduction to neural networks*. CRC press, 1997.

[326] James A Anderson. *An introduction to neural networks*. MIT press, 1995.

[327] Ben Kröse, Ben Krose, Patrick van der Smagt, and Patrick Smagt. An introduction to neural networks. 1993.

[328] Bayya Yegnanarayana. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.

[329] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[330] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.

[331] Yves Chauvin and David E Rumelhart. *Backpropagation: theory, architectures, and applications*. Psychology press, 1995.

[332] Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, and F Huang. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.

[333] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[334] Joseph La Salle, Solomon Lefschetz, and RC Alverson. Stability by liapunov's direct method with applications. *Physics Today*, 15(10):59, 1962.

[335] Jiří Šíma and Pekka Orponen. Continuous-time symmetric hopfield nets are computationally universal. *Neural Computation*, 15(3):693–733, 2003.

[336] Donald Olding Hebb and Wilder Penfield. Human behavior after extensive bilateral removal from the frontal lobes. *Archives of Neurology & Psychiatry*, 44(2):421–438, 1940.

[337] Donald Olding Hebb. Distinctive features of learning in the higher animal. *Brain mechanisms and learning*, 37:46, 1961.

[338] Ernst Ising. Contribution to the theory of ferromagnetism. *Z. Phys*, 31(1):253–258, 1925.

[339] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26): 1792, 1975.

[340] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[341] Lei Jimmy Ba and Rich Caruana. Do deep nets really need to be deep? *arXiv preprint arXiv:1312.6184*, 2013.

[342] Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *International Journal of Automation and Computing*, 14(5):503–519, 2017.

[343] Hrushikesh N Mhaskar and Tomaso Poggio. Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications*, 14(06):829–848, 2016.

[344] Valentin Khrulkov, Alexander Novikov, and Ivan Oseledets. Expressive power of recurrent neural networks. *arXiv preprint arXiv:1711.00811*, 2017.

[345] JD Biamonte. Nonperturbative k-body to two-body commuting conversion hamiltonians and embedding problem instances into ising spins. *Physical Review A*, 77(5):052331, 2008.

[346] Firas Hamze, Jack Raymond, Christopher A Pattison, Katja Biswas, and Helmut G Katzgraber. Wishart planted ensemble: A tunably rugged pairwise ising model with a first-order phase transition. *Physical Review E*, 101(5):052102, 2020.

[347] H Chau Nguyen, Riccardo Zecchina, and Johannes Berg. Inverse statistical problems: from the inverse ising problem to data science. *Advances in Physics*, 66(3):197–261, 2017.

[348] Ross Kindermann. Markov random fields and their applications. *American mathematical society*, 1980.

[349] Geoffrey E Hinton. Boltzmann machine. *Scholarpedia*, 2(5):1668, 2007.

[350] Jérôme Tubiana and Rémi Monasson. Emergence of compositional representations in restricted boltzmann machines. *Physical review letters*, 118(13):138301, 2017.

[351] Simona Cocco, Rémi Monasson, Lorenzo Posani, Sophie Rosay, and Jérôme Tubiana. Statistical physics and representations in real and artificial neural networks. *Physica A: Statistical Mechanics and its Applications*, 504:45–76, 2018.

[352] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR, 2009.

[353] Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*, 2020.

[354] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1): 3133–3181, 2014.

[355] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *Proceedings of the 31st international conference on neural information processing systems*, pages 972–981, 2017.

[356] Dmitry Krotov and John J Hopfield. Dense associative memory for pattern recognition. *Advances in neural information processing systems*, 29:1172–1180, 2016.

[357] Dmitry Krotov and John Hopfield. Feature to prototype transition in neural networks. In *APS March Meeting Abstracts*, volume 2017, pages P5–009, 2017.

[358] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[359] Rajesh Parekh, Jihoon Yang, and Vasant Honavar. Constructive neural-network learning algorithms for pattern classification. *IEEE Transactions on neural networks*, 11(2):436–451, 2000.

[360] Sudhir Kumar Sharma and Pravin Chandra. Constructive neural networks: A review. *International journal of engineering science and technology*, 2(12):7847–7855, 2010.

[361] Shaobo Lin, Jinshan Zeng, and Xiaoqin Zhang. Constructive neural network learning. *IEEE transactions on cybernetics*, 49(1):221–232, 2018.

[362] Yann LeCun, John S Denker, Sara A Solla, Richard E Howard, and Lawrence D Jackel. Optimal brain damage. In *NIPs*, volume 2, pages 598–605. Citeseer, 1989.

[363] Jan Gorodkin, Lars Kai Hansen, Anders Krogh, Claus Svarer, and Ole Winther. A quantitative study of pruning by optimal brain damage. *International journal of neural systems*, 4(02):159–169, 1993.

[364] Chao Liu, Zhiyong Zhang, and Dong Wang. Pruning deep neural networks by optimal brain damage. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[365] Thomas Elsken, Jan Hendrik Metzen, Frank Hutter, et al. Neural architecture search: A survey. *J. Mach. Learn. Res.*, 20(55):1–21, 2019.

[366] Michael P Perrone and Leon N Cooper. When networks disagree: Ensemble methods for hybrid neural networks. Technical report, BROWN UNIV PROVIDENCE RI INST FOR BRAIN AND NEURAL SYSTEMS, 1992.

[367] Harris Drucker, Corinna Cortes, Lawrence D Jackel, Yann LeCun, and Vladimir Vapnik. Boosting and other ensemble methods. *Neural Computation*, 6(6):1289–1301, 1994.

[368] David Opitz and Richard Maclin. Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11:169–198, 1999.

[369] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[370] Dasaratha V Sridhar, Richard C Seagrave, and Eric B Bartlett. Process modeling using stacked neural networks. *AIChE Journal*, 42(9):2529–2539, 1996.

[371] J Zhang, EB Martin, AJ Morris, and C Kiparissides. Inferential estimation of polymer quality using stacked neural networks. *Computers & Chemical Engineering*, 21:S1025–S1030, 1997.

[372] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[373] Sotiris B Kotsiantis. Bagging and boosting variants for handling classifications problems: a survey. *The Knowledge Engineering Review*, 29(1):78, 2014.

[374] Robert E Schapire and Yoav Freund. Boosting: Foundations and algorithms. *Kybernetes*, 2013.

[375] Robert E Schapire. The boosting approach to machine learning: An overview. *Nonlinear estimation and classification*, pages 149–171, 2003.

[376] Armand Joulin, Francis Bach, and Jean Ponce. Discriminative clustering for image co-segmentation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1943–1950. IEEE, 2010.

[377] Christian Schellewald and Christoph Schnörr. Probabilistic subgraph matching based on convex relaxation. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 171–186. Springer, 2005.

[378] Michael I Jordan. An introduction to probabilistic graphical models, 2003.

[379] Kevin Murphy. An introduction to graphical models. *Rap. tech*, 96:1–19, 2001.

[380] Luis Enrique Sucar. Probabilistic graphical models. *Advances in Computer Vision and Pattern Recognition. London: Springer London. doi*, 10(978):1, 2015.

[381] Michael Irwin Jordan. *Learning in graphical models*, volume 89. Springer Science & Business Media, 1998.

[382] Piero Martinoli and Chris Leemann. Two dimensional josephson junction arrays. *Journal of Low Temperature Physics*, 118(5):699–731, 2000.

[383] J Affolter, M Tesei, H Pastoriza, Ch Leemann, and Piero Martinoli. Observation of ising-like critical fluctuations in frustrated josephson junction arrays with modulated coupling energies. *Physica C: Superconductivity*, 369(1-4):313–316, 2002.

[384] Arkadiusz Kosior and Krzysztof Sacha. Simulation of frustrated classical x y models with ultracold atoms in three-dimensional triangular optical lattices. *Physical Review A*, 87(2):023602, 2013.

[385] Se Yoon Lee. Gibbs sampler and coordinate ascent variational inference: A set-theoretical review. *Communications in Statistics-Theory and Methods*, pages 1–21, 2021.

[386] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003.

[387] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[388] Marylou Gabrié. Mean-field inference methods for neural networks. *Journal of Physics A: Mathematical and Theoretical*, 53(22):223002, 2020.

[389] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[390] Judea Pearl. *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science ..., 1982.

[391] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[392] Thomas Ott and Ruedi Stoop. The neurodynamics of belief propagation on binary markov random fields. In *Advances in neural information processing systems*, pages 1057–1064. Citeseer, 2006.

[393] Marc Mézard. Mean-field message-passing equations in the hopfield model and its generalizations. *Physical Review E*, 95(2):022117, 2017.

[394] Stephen A Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158, 1971.

[395] Michael R Garey, David S Johnson, and Larry Stockmeyer. Some simplified np-complete problems. In *Proceedings of the sixth annual ACM symposium on Theory of computing*, pages 47–63, 1974.

[396] Stephen Cook. The p versus np problem. *The millennium prize problems*, pages 87–104, 2006.

[397] Marijn JH Heule, Oliver Kullmann, and Victor W Marek. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 228–245. Springer, 2016.

[398] Marijn JH Heule. Schur number five. *arXiv preprint arXiv:1711.08076*, 2017.

[399] Reinhard Diestel. Graph theory. 2005. *Grad. Texts in Math*, 101, 2005.

[400] Stefan Arnborg, Derek G Corneil, and Andrzej Proskurowski. Complexity of finding embeddings in a k-tree. *SIAM Journal on Algebraic Discrete Methods*, 8(2):277–284, 1987.

[401] Hans L Bodlaender. A linear-time algorithm for finding tree-decompositions of small treewidth. *SIAM Journal on computing*, 25(6):1305–1317, 1996.

[402] Alexander Grigoriev, Hans Ensinck, and Natalya Usotskaya. *Integer linear programming formulations for treewidth*. Maastricht Research School of Economics of Technology and Organizations, 2011.

[403] Hans L Bodlaender. Discovering treewidth. In *International Conference on Current Trends in Theory and Practice of Computer Science*, pages 1–16. Springer, 2005.

[404] Hans L Bodlaender. Treewidth: characterizations, applications, and computations. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 1–14. Springer, 2006.

[405] Kirill P Kalinin and Natalia G Berloff. Nonlinear systems for unconventional computing. In *Emerging Frontiers in Nonlinear Science*, pages 345–369. Springer, 2020.

[406] Hui Cao, Ronen Chriki, Stefan Bittner, Asher A Friesem, and Nir Davidson. Complex lasers with controllable coherence. *Nature Reviews Physics*, page 1, 2019.

[407] Daniel Brunner, Miguel C Soriano, Claudio R Mirasso, and Ingo Fischer. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature communications*, 4:1364, 2013.

[408] Ling Bao, Nam-Heon Kim, Luke J Mawst, Nikolay N Elkin, Vera N Troshchieva, Dmitry V Vysotsky, and Anatolii P Napartovich. Near-diffraction-limited coherent emission from large aperture antiguided vertical-cavity surface-emitting laser arrays. *Applied physics letters*, 84(3):320–322, 2004.

[409] Fabien Rogister, K Scott Thornburg Jr, Larry Fabiny, Michael Möller, and Rajarshi Roy. Power-law spatial correlations in arrays of locally coupled lasers. *Physical review letters*, 92(9):093905, 2004.

[410] Vishwa Pal, Chene Tradonsky, Ronen Chriki, Asher A Friesem, and Nir Davidson. Observing dissipative topological defects with coupled lasers. *Physical review letters*, 119(1):013902, 2017.

[411] Yoshiki Kuramoto. *Chemical oscillations, waves, and turbulence*. Courier Corporation, 2003.

[412] Yoshiki Kuramoto. International symposium on mathematical problems in theoretical physics. *Lecture notes in Physics*, 30:420, 1975.

[413] Hassan K Khalil. Nonlinear systems. *Upper Saddle River*, 2002.

[414] Chene Tradonsky, Oren Raz, Vishwa Pal, Ronen Chriki, Asher A Friesem, and Nir Davidson. Rapid phase retrieval by lasing. *arXiv preprint arXiv:1805.10967*, 2018.

[415] Peter L McMahon, Alireza Marandi, Yoshitaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsunomiya, Kazuyuki Aihara, et al. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.

[416] Ryan Hamerly, Takahiro Inagaki, Peter L McMahon, Davide Venturelli, Alireza Marandi, Tatsuhiro Onodera, Edwin Ng, Carsten Langrock, Kensuke Inaba, Toshimori Honjo, et al. Experimental investigation of performance differences between coherent ising machines and a quantum annealer. *Science advances*, 5 (5):eaau0823, 2019.

[417] Fabian Böhm, Takahiro Inagaki, Kensuke Inaba, Toshimori Honjo, Koji Enbutsu, Takeshi Umeki, Ryoichi Kasahara, and Hiroki Takesue. Understanding dynamics of coherent ising machines through simulation of large-scale 2d ising models. *Nature communications*, 9(1):5020, 2018.

[418] Yoshitaka Haribara, Hitoshi Ishikawa, Shoko Utsunomiya, Kazuyuki Aihara, and Yoshihisa Yamamoto. Performance evaluation of coherent ising machines against classical neural networks. *Quantum Science and Technology*, 2(4):044002, 2017.

[419] Andrew D King, William Bernoudy, James King, Andrew J Berkley, and Trevor Lanting. Emulating the coherent ising machine with a mean-field algorithm. *arXiv preprint arXiv:1806.08422*, 2018.

[420] Kenta Takata, Alireza Marandi, and Yoshihisa Yamamoto. Quantum correlation in degenerate optical parametric oscillators with mutual injections. *Physical Review A*, 92(4):043821, 2015.

[421] Fabian Böhm, Guy Verschaffelt, and Guy Van der Sande. A poor man's coherent ising machine based on opto-electronic feedback systems for solving optimization problems. *Nature communications*, 10(1):1–9, 2019.

[422] Masoud Babaeian, Dan T Nguyen, Veysi Demir, Mehmetcan Akbulut, Pierre-A Blanche, Yushi Kaneda, Saikat Guha, Mark A Neifeld, and N Peyghambarian. A single shot coherent ising machine based on a network of injection-locked multicore fiber lasers. *Nature communications*, 10(1):1–11, 2019.

[423] Kirill P Kalinin and Natalia G Berloff. Networks of non-equilibrium condensates for global optimization. *New Journal of Physics*, 20(11):113023, 2018.

[424] D Pierangeli, G Marcucci, and C Conti. Large-scale photonic ising machine by spatial light modulation. *Physical Review Letters*, 122(21):213902, 2019.

[425] Jan Klaers, Julian Schmitt, Frank Vewinger, and Martin Weitz. Bose–einstein condensation of photons in an optical microcavity. *Nature*, 468(7323):545, 2010.

[426] Jan Klaers, Frank Vewinger, and Martin Weitz. Thermalization of a two-dimensional photonic gas in "a white wall" photon box. *Nature Physics*, 6(7):512, 2010.

[427] Jan Klaers, Julian Schmitt, T Damm, F Vewinger, and M Weitz. Bose–einstein condensation of paraxial light. *Applied Physics B*, 105(1):17, 2011.

[428] Julian Schmitt, Tobias Damm, Frank Vewinger, Martin Weitz, and Jan Klaers. Thermalization of a two-dimensional photon gas in a polymeric host matrix. *New Journal of Physics*, 14(7):075019, 2012.

[429] Esther Wertz, Lydie Ferrier, DD Solnyshkov, Robert Johne, Daniele Sanvitto, Aristide Lemaître, Isabelle Sagnes, Roger Grousson, Alexey V Kavokin, Pascale Senellart, et al. Spontaneous formation and optical manipulation of extended polariton condensates. *Nature physics*, 6(11):860, 2010.

[430] Francesco Manni, Konstantinos G Lagoudakis, Timothy Chi Hin Liew, Régis André, and Benoit Deveaud-Plédran. Spontaneous pattern formation in a polariton condensate. *Physical review letters*, 107(10):106401, 2011.

[431] G Tosi, G Christmann, NG Berloff, P Tsotsis, T Gao, Z Hatzopoulos, PG Savvidis, and JJ Baumberg. Sculpting oscillators with light within a nonlinear quantum fluid. *Nature Physics*, 8(3):190, 2012.

[432] G Tosi, G Christmann, NG Berloff, P Tsotsis, Tingge Gao, Z Hatzopoulos, PG Savvidis, and JJ Baumberg. Geometrically locked vortex lattices in semiconductor quantum fluids. *Nature communications*, 3:1243, 2012.

[433] Natalia G Berloff, Matteo Silva, Kirill Kalinin, Alexis Askitopoulos, Julian D Töpfer, Pasquale Cilibrizzi, Wolfgang Langbein, and Pavlos G Lagoudakis. Realizing the classical xy hamiltonian in polariton simulators. *Nature materials*, 16(11):1120, 2017.

[434] Ch Schneider, K Winkler, MD Fraser, M Kamp, Y Yamamoto, EA Ostrovskaya, and Sven Höfling. Exciton-polariton trapping and potential landscape engineering. *Reports on Progress in Physics*, 80(1): 016503, 2016.

[435] Alberto Amo and Jacqueline Bloch. Exciton-polaritons in lattices: A non-linear photonic simulator. *Comptes Rendus Physique*, 17(8):934–945, 2016.

[436] David Dung, Christian Kurtscheid, Tobias Damm, Julian Schmitt, Frank Vewinger, Martin Weitz, and Jan Klaers. Variable potentials for thermalized light and coupled condensates. *Nature Photonics*, 11(9): 565, 2017.

[437] Kirill P Kalinin and Natalia G Berloff. Polaritonic network as a paradigm for dynamics of coupled oscillators. *arXiv preprint arXiv:1902.09142*, 2019.

[438] Kirill P Kalinin and Natalia G Berloff. Simulating ising and n-state planar potts models and external fields with nonequilibrium condensates. *Physical review letters*, 121(23):235302, 2018.

[439] Roy Lang and Kohroh Kobayashi. External optical feedback effects on semiconductor injection laser properties. *IEEE journal of Quantum Electronics*, 16(3):347–355, 1980.

[440] Juan A Acebrón, Luis L Bonilla, Conrad J Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137, 2005.

[441] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Marvels and pitfalls of the langevin algorithm in noisy high-dimensional inference. *Physical Review X*, 10(1):011057, 2020.

[442] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

[443] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, and Lenka Zdeborová. Who is afraid of big bad minima? analysis of gradient-flow in a spiked matrix-tensor model. *arXiv preprint arXiv:1907.08226*, 2019.

[444] Stefano Sarao Mannelli, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborova. Passed & spurious: Descent algorithms and local minima in spiked matrix-tensor models. In *international conference on machine learning*, pages 4333–4342. PMLR, 2019.

[445] Stephan Mertens. A physicist's approach to number partitioning. *Theoretical Computer Science*, 265(1-2): 79–108, 2001.

[446] Christian Borgs, Jennifer Chayes, and Boris Pittel. Phase transition and finite-size scaling for the integer partitioning problem. *Random Structures & Algorithms*, 19(3-4):247–288, 2001.

[447] Stephan Mertens. Number partitioning. *Computational Complexity and Statistical Physics*, page 125, 2006.

[448] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.

[449] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701, 2011.

[450] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[451] Lenka Zdeborová and Marc Mézard. Locked constraint satisfaction problems. *Physical review letters*, 101 (7):078702, 2008.

[452] Stuart Kauffman and Simon Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of theoretical Biology*, 128(1):11–45, 1987.

[453] Stuart A Kauffman and Edward D Weinberger. The nk model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of theoretical biology*, 141(2):211–245, 1989.

[454] Wolfgang Barthel, Alexander K Hartmann, Michele Leone, Federico Ricci-Tersenghi, Martin Weigt, and Riccardo Zecchina. Hiding solutions in random satisfiability problems: A statistical mechanics approach. *Physical review letters*, 88(18):188701, 2002.

[455] Florent Krzakala and Lenka Zdeborová. Hiding quiet solutions in random constraint satisfaction problems. *Physical review letters*, 102(23):238701, 2009.

[456] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

[457] David H Wolpert. The lack of a priori distinctions between learning algorithms. *Neural computation*, 8 (7):1341–1390, 1996.