Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Nikita Stroev

PhD Program: Physics

Title of Thesis: Modelling of exciton-polariton condensates for unconventional computing

Supervisor: Prof. Natalia Berloff

The thesis document includes the following changes in answer to the external review process.

Dear Jury Members, I would like to thank you for the valuable feedback that allowed me to get an independent perspective on my work and consequently improve its quality. I appreciate your time and effort in reviewing the thesis material.

I wrote the following detailed responses to each of the comments, according to your recommendations. You can find the below together with the corresponding modifications in the final version of the text.

Reviewer: Jonathan Keeling

Substantive issues to be addressed in the thesis

- "The abstract should be revised to make clear the new research contributions in the thesis. These are all mentioned, but the abstract does not distinguish sufficiently what is new and what is background."

Answer: I agree that the current abstract does not properly highlight the novelty of the research contributions. I added several sentences to properly address these issues in the abstract and present the key results. The new abstract has the following form (the new sentences are in bold format):

"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."

- "On a similar theme, the section headings in section 3 do not always indicate the novel aspects of the work in a given section. For example, section 3.2 on tensor sum minimisation introduces the novel approach of complex coupling switching as a route to solve this problem. This key feature is not clear from the section name, thus hard to locate when considering the table of contents. The same point appears to other sections. I would suggest updating section names (and possibly adding subsections) to highlight what new ideas are introduced. "

Answer: Following the suggestion, I changed a few headings in Section 3 to highlight the key features, similarly as in the abstract updates. Specifically - 3.2 Tensor sum minimization **and complex coupling switching**; and 3.7 Engineering periodic structures **with the target velocity.** The ML section name has many detailed subsections and remained unchanged. The same works for Section 2, which gives the physical preface and Section 4, which provides the outlook on different algorithms.

- "Chapter six discusses some perspectives of work on topics somewhat related to other chapters. Part of this discussion involves poor citation practice, in that the material in this chapter is very close in structure and wording to material in the papers cited. Particularly the text of section 6.2 is very close to material in references 444 and 346. Substantial rewriting would be required to make this acceptable in a published thesis. Given the material is not required for the thesis, I would strongly suggest to remove this material (possibly this whole chapter) to avoid this problem."

Answer: Chapter 6 contains the material concerning other physical platforms acting similarly to the analog optimization platforms. Additional focus is paid to the gain-dissipative mechanism in guiding the optimization process, which was not sufficiently covered in the main text. I follow the proposed suggestion to rewrite subsection 6.2. For this purpose, the major part of the presented material that can be found in the original articles was removed. However, the subsection remained, containing only several important references for the reader and a minimal amount of information about the content.

Minor issues/typos in the thesis

- "Throughout the thesis, many figures are not referenced from the text."

Answer: More references to the figures are added to the text.

- "P20. The discussion around Eqs. 2.1 and 2.2 do not seem to be consistent: the meanings given

to various terms do not seem to be consistent with the overall expression. If rho_nu is the energy density of the radiation modes, I would assume that 2.1 is meant to be the Rayleigh-Jeans (classical) law for this, while Eq. 2.2 is the corrected Bose-Einstein quantum version. If so, there are several unclear points.

Firstly, in the Rayleigh-Jeans law, the quantity E, describes as "the mean energy of an oscillator" should presumably become kB T/2, via classical equipartition. As written, E is not defined, making this expression unclear. It may be that the idea is that E = 0.5 * (hbar nu) * nB(hbar nu), so that this equation is already the Bose-Einstein version. However that is not stated. In any case, this should be clarified.

Secondly, in Eq. 2.2, alpha_Delta is defined as a transition probability, and "Delta epsilon" is not defined. If rho_nu in 2.2. still means the radiation density (i.e. if the quantity rho_nu has the same meaning in 2.1 and 2.2), then it does not make sense to define it in terms of transition probabilities. "

Answer: I decided to remove this detailed discussion not to confuse the reader with the inconsistent equations. To get the right picture, one can refer to the original articles, mentioned in the text.

- "In several places "Schrodinger" is misspelled as Shrodinger.

  P13, discussion of Chapter 4: "general mistake" → "general error"

  P14, "domineering"→"dominant"

  P22. I would avoid referring to authors by just their first name. As such, suggest change "Lazlo submitted" → "Tisza submitted"

  P22. "These separation" → "This separation"

  P25. "Another recent result.." This is referring to "Another result around the same time…". The word "recent" implies it being near to the present time, not near to the time about which you are writing.

  P25. "evaporate cooling" →" evaporative cooling"

  P27. "prevails the decay" → "exceeds the decay"

  P28. "If one quantize the field operator…". This is unclear, since the field operator in Eq. 2.19 is already quantized (already an operator). What you really mean is "If one writes the field operator in the momentum basis"

  P30 "called the Lindblad super-operator". This is not quite right: Eq. 2.26 is a Lindblad super-operator. Eq. 2.25 is a Liouvillian super-operator.

  P30 "in the current work" → "in that work". "Current work" refers to the document the reader is reading, i.e. the thesis.

  P33 "minimize the amount of matter". Should this not read "maximize…" ?

  P34 "Despite the spatial coordinates" → "As well as the spatial coordinates"

  P36 "The purpose of this article" → "The purpose of this section"

  P52 "sterngths" → "strengths"

  P62. Reference to Figs 2-4 probably means Fig. 3-14

P63. It would be helpful if the caption to Figure 3-14 explained the difference between panels (a,b), rather than just refer to equations.

P67 “can me mapped” → “can be mapped”

P68 “penalazing term”. If this phrase is really wanted it should be spelled “penalizing term”. However a better phrase is “cost function”.

P69 “Lagrangean” → “Lagrangian”
P71 “standartized” → “standardized”

P73 “reffers” → “refers”

P76 “it’s a more narrow”. Should this read “its more narrow”? The current phrasing does not make sense.

P76 “It is possible can” → “It is possible to” “

Answer: All of the presented misspellings, typos and mistakes were corrected according to the suggestions. I would like to express additional gratitude to the reviewer for the laborious work and effort.

- “P38 In the unnumbered equation between 3.21, 3.22, a quantity “d” is introduced (first term I a_i d) that is not defined. Please add a definition of d. “

Answer: It may seem that the value “d” is not introduced in the context. However, the definition is placed earlier in (3.18) together with all other terms.

- “P41. In Eq. 3.30, u_l(t) is not defined. Should this be x_l(t)? “

Answer: Yes, it is certainly a typo.

- “P61/62 Over these pages, a number of different scenarios are presented, and forms of rho(x) presented. Some of these appear as displayed equations (i.e. numbered equations), while some are inline equations. There does not seem an obvious reason for this different treatment. I suggest to make the treatment uniform. “

Answer: The treatment for all of the different equation types is made uniform.

- “P68, Figure 4-2. The numbers on the arrows are very hard to see due to the dark shading of the arrow. “

Answer: The color of the numbers was changed to white for the better contrast with the dark background.

- “P71 In Eq. 4.10 it is unclear why the penultimate term still involves powers of x; this may be a typo. P71 In Eq. 4.11, the quantity x^l_{ij} has not been defined. “

Answer: Eq.4.10 certainly had a typo, which was corrected. The additional description was introduced to clarify the notation in Eq. 4.11.

- “P77 Does “original binary form” mean “original quadratic form”? “

Answer: Yes, it is certainly a typo, connected with the frequent use of the “binary quadratic form”.

- “P84. There is a reference to “The next chapter will consider ..” that seems to refer to material in chapter three. If this is the case, change this to read “In chapter three, we already discussed the dynamical aspect of the minima search.” or equivalent.”

Answer: Corrected according to the suggestion.

Reviewer: Sergei K. Turitsyn

- "**Section 3.2**
  1- In page 36 additive noise shown to be effective to avoid trapping in a local minimum, what are the requirements of this additive noise? Is there a minimum value or a specific random distribution? Can optimising the variance and properties of this additive noise give rise to a system with real coupling which can perform as good as TGD+CC?"

Answer: The additive noise is necessary to avoid trapping in a poor local minimum. It plays an essential role during the process of condensation in Eq.(3.22), i.e. when the densities of the condensates are close to zero, meaning that given any non-zero amplitude the noise term will guide the search in the phase space. There is no strict requirement on the amplitude of the noise, however, the good choice will be to take the value comparable with the condensed wavefunction amplitude. Empirically, the uniform probability density distribution was chosen, which can be similarly replaced with the gaussian distribution on larger time steps, due to the law of large numbers. No specific distributions are required for such a process, that is why there are few details about this term in the main text.

I believe that optimising the variance and other properties of this noise term can improve the performance of the given algorithm. Moreover, the conventional random walk in the high-dimensional space is not the best choice for the searching strategy. Strictly saying, to address the question of improving the performance by optimising the noise term one has to perform additional numerical experiments. I do not think that the improvement of the noise term alone can enhance the given algorithm, because its essential part is the interplay between the gain-dissipation process, gradient descent and noise terms. It can be demonstrated using the same similar toy model given by Eq. (3.27) and then upscaling it on the larger examples. The efficiency of the complex-coupling switching should lie in the positive correlations between the nonconvex structure of the optimization problem generated by the choice of the coupling coefficients in the tensor and the dynamical perturbation guided by the similar coefficients, however, no rigorous closed-form explanation of this fact exists.

- "**Section 3.4 (and 3.5)**
  1- It's possible to define any machine learning task (classification or regression) especially with supervised training as a minimisation problem. Having that in mind, why would someone implement a ML solution using the XY Hamiltonian blocks instead of directly translating the ML problem into a Hamiltonian using the techniques explained in Chapter 4?
  2- What is the speed of arriving at the ground state which directly impacts the data processing speed? What are the limitations? How does it change by increasing the number of spins?
  3- This is a solution for implementing the inference stage of a ML technique, what does it take to include training (adaptive change of the system parameters via SLM for example)?"

Answer: 1- Many ML tasks can be recast as optimisation problems, for which the exciton-polariton can play the role of the analog optimiser, which was discussed at the beginning of Chapter 3. To show the versatility of the exciton-polariton platform, the main topic of the thesis, I presented an alternative way of implementing neural network algorithms through a set of nonlinear operations. Moreover, this is not just another way of performing the same functions, because the presented method has its own advantages. Among them is the small overhead of the condensate elements while scaling to bigger tasks. Other benefits are discussed in subsection 3.5.2.

2-In the case of implementing the approach of analog optimiser, one can divide the evolution of the system governed by the Eqs. (3.25-3.26) into several parts. First is the process of reaching the stage when the pumping exceeds the dissipation, which depends linearly on the disbalance between the initial dissipation in the system and zeroes pumping state. Next is the process of condensation, which is characterized by the exponential growth of condensate densities. The final part of the convergence to the final state can be described by linearizing Eqs. (3.25-3.26) near the approximate solution to a problem and taking the lowest characteristic time across the spectrum of found values. Speaking about the nonlinear blocks, one can place the bounds on each operation stage, according to the solution of local cluster equilibrium and then multiply it on the effective number of stages in the architecture.

According to the simulations, close to the experimental parameters, the approximate performance takes the characteristic time near the 100-200 ps range. To improve the estimates, it is necessary to consider more detailed and specific tasks, including the information about the range of available coupling coefficients, connectivity of the spin systems, etc., which is a topic for separate consideration.

3-See the answer below.

- "**Section 3.5**
  1- How can one use this architecture to include the training stage as well? This entails performing backpropagation which includes calculating gradient. Does approximating these nonlinear activation functions in the way that is proposed here pose any issue in calculating the gradient? This is important because most often the gradient of the extremes of the nonlinear function (at small or large \theta_{in}) plays an important role and this is where the approximations shown in Figs. 3.5,3.6, and 3.7 deviate from the analytical values."

Answer: The main task that was considered in the thesis is the predefined NN architecture transfer, not building the architecture capable of learning. It was mentioned at the beginning of the corresponding Section. Furthermore, in the current context, the coupling coefficients can not be changed through the operational regime of the spin setup, which makes the subject of backpropagation not relevant. This topic requires an additional description of the tools used to manipulate the system, more details about its constituent parts and a mathematical supplement that contains the modification of the backpropagation for a different set of functions. In case that it becomes possible to introduce the feedback to control the coupling coefficient, for which there are no evident obstacles, I believe that the conventional approach to backpropagation can be applied. It implies that providing the batch of data together with the desired cost function will result in the possibility to implement gradient calculations. The complex nonlinear function does not add difficulties for such calculations, since the gradient calculations can be modified for any function composition, but the problem arises in the limited domain of its definition, which is pointed in the reviewer's question. This serious issue can be addressed in the future.

However, it is necessary to mention the existence of alternative versions for the training of the NNs, of which there are many various ways. One way is to solve the optimisation task each time given the fixed weights in the NN by fixing a small number of neuron states. In the trained stage, they will act as effective weights of pretrained NN.

- "**Section 4.1**
  1- What is the relation between the two representation of a Hopfield network in Eq. (4.34) and Eq. (4.35)? is it possible to derive (4.34) from (4.35)?
  2- In a Hopfield system explained by Eq. (4.35) what is the input? Is it Ii or the initial state, xi(0)?
  3- In page 74 paragraph 6 it says: "The Hopfield NN can reduce more deep NN because every deep feedforward NN can be reduced to a shallow one with different parameters". Could you please explain how this reduction is done? Is there a general

procedure to perform the reduction?"

Answer: 1- They are the same dynamic rules for updating the state of neurons, but Eq. (4.34,4.28 in the revised version) describes the discrete version of the NN with possible binary states and Eq. (4.35,4.29 in the revised version) describes the continuous version. The discrete version can be obtained from the continuous one by setting the "g(x)" to be the Heaviside step function (or the limit hyperbolic tangent with the effective scaling constant going towards infinity), choosing the extremely large time of convergence "\tau" in comparison to the partial sums and choosing the finite timestep "dt" in the derivative of "dx/dt". In reverse, the continuous variant is usually treated as an extension of the discrete version.

2 - "xi(0)" - the initial values for the vector are considered to be input values. The values "Ii" play the role of force-fields acting individually on "xi" values. They have the same role as the external field "hi" acting on individual components of spin vector "xi" in the Ising model. If one wants to use Eq. (4.29) as the optimisation procedure, "xi(0)" should be chosen as the random initial state with the possible positive correlation with the optimal solution of a hard problem. In the case of pattern recognition (Hopfield NN), one chooses the "xi(0)" as the string of bits representing the noisy pattern to restore or classify.

3 - There is a small typo, that is corrected. The proper sentence has the following form: "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." First of all, there is an obvious correspondence between the recurrent NNs and conventional NNs, where one can change the neuron state to be either active or "frozen" (to be constant, which makes it the effective weight analog). Second, there is a lot of literature, dedicated to the approximation capabilities of different families of NNs and the correspondence between shallow and deep NNs. To name one example, one can start with the universal approximation theorem, stating that any kind of good function defined on the multidimensional space can be approximated with wide NNs with only two layers (shallow NNs). Following this logic, one can decompose any deep enough NN into the set of output neurons, or functions, each is the target for approximation procedure, according to this theorem.

- "Few minor comments about definitions.
  1) Equation 3.1. It would be helpful to explain physical meaning of \eta_d. Why \n_r is chosen in this form? Some discussion or references could be helpful.
  2) Eqs. 3.4 и 3.5. Velocity u should be defined through derivative of S. It is easy and known, but nevertheless should be done, as not all readers are from the field."

Answer: 1) \eta_d is the energy relaxation coefficient specifying the rate at which gain decreases with increasing energy, \n_r is chosen in the form, to ensure that its derivative d\n_r/dt =0, which satisfy the condition for the quasistationary state of the excitonic reservoir. Additional details can be found in [188]. These descriptions are added to the main text.

2) Additional formula for the velocity definition is added, following the suggestion.

Reviewer: Ildar Gabitov

- "The results presented in the dissertation meet the criteria of novelty, and are published in top scientific journals. Considering that presented results were published in co-authorship with the scientific supervisor, and the fact that for defense author of the thesis has to submit his original results, I propose to clearly formulate his personal contribution to these joint work."

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

The idea was proposed by N.G.B. N.S. performed the analytical derivations and performed the computations. N.G.B verified the analytical methods and the quality of the calculations. N.S. wrote the manuscript with support from N.G.B. Both N.S. and N.G.B. authors contributed to the final version of the manuscript.

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:

N.S. and N.G.B conceived of the presented idea. N.S. developed the theoretical formalism, performed the analytic calculations and performed the numerical simulations. N.S. wrote the manuscript with support from N.G.B. N.G.B. supervised the project.

Nikita Stroev and Natalia G Berloff. XY neural network.arXiv preprint arXiv:2103.17244, 2021:

The idea was proposed by N.S. N.S. developed the theory and performed the computations. N.G.B. verified the analytical methods. N.S. wrote the manuscript. Both N.S. and N.G.B. authors contributed to the final version of the manuscript. N.G.B. supervised the project.

- “On page 13, the author provides a key formulation of the approach to the analysis of the topic disclosed in the dissertation: "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 behavior ". I propose in the considered examples to explain in more detail the physical principles and control mechanisms that allow to implement such a mapping, as well as indicate their limits of applicability.”

Answer: The key formulation was provided after finishing the main text of the thesis to give a brief summary of the work. Shortly, the whole thesis material is devoted to the exact detailed explanation of how to convert a physical system with its corresponding principles into a computing device, capable of several types of computation, so no additional examples are required. Additionally, one can generalise the presented example to other systems. Starting with equations of motions for dynamical system description of the system’s degrees of freedom (in our case the system consists of exciton-polariton condensates and the target degrees of freedom are condensate phases or classical spins) I explore what kind of the geometrical surface is formed by their interaction types. The next step is to understand how well I can map the nonconvex landscape of hard optimisation problems or the attractor landscape of patterns to be restored on the surface, produced by the interaction potentials. In such a sense I provide the correspondence between the dynamical behavior of the physical system and possible computational tasks in the form of searching the desired phase space through navigating the given surface. The engineering capabilities determine the mechanism of navigating this surface and its quality and hence the quality of the correspondence between the task and system behavior.

- “Note also that the robustness analysis discussed on page 62 can be done analytically.”

Answer: Thank you for the suggestion, but the presented calculations have the same aim of showing the robustness and stability of certain families of equations. I did not include the analytical analysis not to overload the main text with the details. Moreover, one can find a similar robustness analysis in references [301,302].

- “There are minor inaccuracies in the thesis, such as, for example, there is no reference to FIG. 3-14. ”

Answer: More references to the figures are added to the text, including the reference to FIG. 3-14.

Reviewer: Nikolay Gippius

- "The estimation of the validity of the used equations should be addressed and discussed in the text of the thesis."

Answer: I provided additional information about Eq. (3.1) together with the corresponding references on its derivation and the validity discussion.

- "The proposed heuristic itself was shown to outperform (in theory) the conventional Hoipfiled neural networks. However it is not at all clear how these theoretical proposals can be implemented experimentally as the measurement of the key variables of the model is a rather hard task, not to say about their control."

Answer: While introducing the complex-coupling switching method, it was designed to make it as simple as possible for the sake of experimental implementation. The key steps are to choose some of the coupling coefficients and introduce the complex part proportional to the real part of the coefficient. Since given the current state of the experimental progress, one can introduce complex coupling coefficients between the interacting condensates, the idea is to turn these complex parts on without their rigorous tuning for a certain duration only for the part of the condensates (~10% of all coefficients) to achieve the new dynamical regime, which according to the calculations, should significantly improve the process of searching the phase space for better solution of the initial optimisation task.

- "P.16 "Hopfiled NNS". (In the rest of the text NNs) -> Hopfiled neural networks(NNs) "

Answer: Corrected.

- "Use of EP for exciton-polariton is a bit misleading, is it different from common LP (lower polariton) branch? Seems the LP case is considered in the thesis. What is the reason for new notation? "

Answer: The presented notation is used only locally in the thesis. There is no difference between the terms LP branch of exciton-polaritons and the EP system in this thesis. The physical preface gives a comprehensive physical picture of the presented system, where one can find the similarity of these terms. I decided to leave the notation due to its frequent use across the work.

Reviewer: Pavlos Lagoudakis

No questions to address.

Other changes:

Additional acknowledgements are made.