Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Evgenii Tsymbalov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Machine Learning for Elastic Strain Engineering

Supervisor: Prof. Alexander Shapeev

Name of the Reviewer: Bohayra Mortazavi

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 16-09-2020**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

Provisional Recommendation

Evgenii's thesis entitled "MACHINE LEARNING FOR ELASTIC STRAIN ENGINEERING" describes the exploration of the silicon and diamond properties under elastic strain with the help of machine learning. Engineering the electronic and optical properties via mechanical straining is now a very attractive topic in academia and high tech industries. Moreover, the application of machine learning methodologies to address the complex problems in the materials science is another highly active and extending field of research. In particular, extending the first-principles modeling accuracy to solve challenging real-world problems is another rapidly-growing research area. Evgenii's thesis is a symbol of a complete and advanced work, which includes development of a combined first-principles modeling plus machine learning to solve a problem of critical importance in academia and industry. In the heart of this work, Evgenii develops and tests the neural network model for the prediction of crystal properties under strain; in which the topics of deformed crystal capabilities and uncertainty estimation for the neural networks in general are also extensively covered and provide scientific advances to state of the art.

The thesis is well-organized, with content separated into eight chapters, as follows: The first chapter is dedicated to a brief overview and motivation behind the work. The second chapter contains the gentle introduction into the necessary background knowledge about crystal structures, first-principles calculations, machine learning in general and active learning in particular. Due to the interdisciplinary nature of the work, there are many topics to describe the latest advances on, and the author focuses on the narrow fields of elastic strain engineering and machine learning applications to simulations, as the most relevant to the content of his work, in Chapter 3. The following chapter describes the methodology of ab-initio simulations used in this work, together with a brief analysis of the sampled data. In Chapter 5, the author details the algorithms for uncertainty estimation and active learning and provides the prerequisites to the next chapter, which is dedicated to numerical experiments. Here, a short part is devoted to the accuracy check of the proposed approach for the strain engineering, while the majority of the chapter reveals comprehensive checks of the developed algorithms of uncertainty estimation and active learning in the broader context of neural network models and tasks. Chapter 7 demonstrates scientific discoveries made possible with the developed models – namely, an extensive description of the band gap and related properties in the space of admissible strains for the diamond and silicon crystals. The last chapter summarizes the work and discusses possible future prospects.

As stated earlier, the employed techniques, developed methodology and acquired results described in this work contribute to state of the art in several fields, including: material modeling, extension of first-principles modeling, strain engineering of electronic and optical properties, machine learning application in materials science and uncertainty estimation. The high quality of this thesis is certainly concludable from published papers, which include five first-authored and two co-authored publications. One of the major contribution of this thesis was published in the highly renowned ***PNAS*** journal, entitled "Deep Elastic Strain Engineering of Bandgap through Machine Learning", with Evgenii as the equal contribution first author has been received 24 citations in less than two years, which indicates the remarkable impact and quality of conducted research. They also filed a patent on this topic which highlights the prospect for the practical application of this research.

Based on the review above, I recommend that Evgenii is prepared and deserves to defend his thesis by means of a formal thesis defense. However, there are a few comments and remarks to be considered for the further improvement:

- The formula for the central-difference effective mass component approximation (5.6) is wrong (possibly a typo).
- Could this model be used to predict other bandgap-related properties of the materials, such as thermal conductivity, optical absorption, or phononic properties?
- Enhancement of carrier mobilities in diamond and silicon by straining could be considered as the further extension of this work. In this case, the pattern of valance and conductance bands should be considered for the engineering. More discussions should be included with this regard.

- I also suggest to include more discussion on the effect of type of atomic lattices on the strain engineering, because by changing the lattice type the symmetry also changes.
- Please also comment on the extension of this approach for the case of 2D materials, I think that would be highly appealing and computationally more feasible.
- One promising aspect is to couple the developed approach with machine learning interatomic potentials in order to accelerate the computations, please add more discussions about such a possibility.
- A significant part of the content is related to the uncertainty estimation for the neural networks, yet it is used merely for active learning in the case of strain engineering task. I would appreciate if the author could elaborate more on the precise mechanism on uncertainty estimation and active learning for the diamond crystal and indicate whether there other uses of the produced uncertainty estimates, e.g., confidence intervals for the predicted values of band gap.
- Personally, I would suggest moving the details of numerical experiments to the appendices.

☒ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*