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:** Associate Professor Alexander Shapeev

**Name of the Reviewer: Nikolai Brilliantov**

| | |
|---|---|
| I confirm the absence of any conflict of interest<br><br>(Alternatively, Reviewer can formulate a possible conflict) | **Signature:**<br>  <br>**Date: 19-09-2020** |

*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**

Evgenii Tsymbalov's thesis is devoted to the application of the machine learning (ML) to the prediction of electronic properties of materials under large elastic deformations. It comprises eight chapters, including Introduction and Conclusion, list of references and Appendixes.

In the first chapter (Introduction) the author provides motivation of the study. In the second and third chapters a rather comprehensive description of the main elements of the density functional theory (DFT) and methods of machine learning (ML), used in the thesis, are presented. The overview of the existing approaches is also given; the new important aspects of the application of the ML are discussed in detail.

The next two chapters address the detailed description of the ML experiments and strain-induced properties of the materials. Here the author provides the results of numerical experiments and propose the models to describe the simulation results. All aspects of a practical implementation of the simulation methods, elaborated in the thesis, are explained. In the sixth chapter the simulation accuracy, which is of the primary importance for the new methodology is assessed within a couple of approaches. In the seventh chapter, the strain-induced properties of the materials, such as a band gap, the structure of the Brilluen zone, metallization of semiconductors (Si) and insulators (diamond) are discussed. Finally, the eighth chapter concludes the thesis, providing an overview of the work, and stating future problems together with the author's contribution.

The main line of the research work of the thesis, addresses the problem of reduction of the cost of density functional theory-based calculations of elastically strained crystals. This is done with the help of machine learning. Using the state-of-the-art methods from both machine learning and material science fields, the author introduces an accurate model, which, in turn, is used to generate data from DFT simulations and acquire data for silicon and carbon diamond crystals straining. The obtained results, including the map of crystal properties, within various strain regimes and corresponding experimental cost estimates, are, without a doubt, of high scientific significance. Moreover, they are not only of academic importance, but possess a high application potential. In particular, astonishing properties of the band structure of silicon and diamond, such as metallization and direct band gap existence, are revealed and explored within the proposed framework. These discoveries may pave the way to a number of practical applications. For instance, for fabrication of efficient solar elements, with a continuously varying bandgap, that may completely adsorb all spectrum of the solar light.

At the time of my review, the thesis-related publications include two machine-learning conference papers and two interdisciplinary works published in the Proceedings of the National Academy of Sciences - a high-impact journal. This proves a high scientific level of the thesis. Moreover, the author belongs to a authors team of the patent based on the research.

Although I highly esteem the work and have no doubt that the author deserves the PhD degree, I still have a couple of comments/questions, which I list below:

1) The study addresses an ideal crystal without any defects. This is correct for the idealized case of zero temperature, T=0. Each real crystal however has inevitably equilibrium point defects for non-zero temperature and generally, non-equilibrium defects like dislocations. Hence, a question arises, how the presented results would change, if these important properties of real crystals were taken into account? I do not expect quantitative estimates, but just a qualitative answer – whether the observed decrease of the bandgap persists? Will it shift for larger or smaller strain?
2) What would be the qualitative change of the results for non-zero temperature?
3) Is it possible to extend the research for the multi-component crystals or 2D materials?
4) What is the role of active learning in the developed methodology?
5) What are the limitations of the machine learning model proposed? May it benefit more from the multiple non-connected sources of data?
6) I suggest the author to comment more on the limitations of the proposed approach predictions, including surface effects, DFT-related errors, etc.

A few comments that refer to the exposition of the material:

1. In the section devoted to the DFT more theory is needed. Why did not the author mention the Slater determinant, which would naturally explain the exchange interactions? Otherwise it is not clear what is the physical meaning of these interactions.
2. In Figs. 6.21 and 6.22 the values of about 1kcal/mole are shown. Is this a large or small error? This should be discussed explicitly in the text or in the figure captions.
3. In page 126 the author makes a strange statement that the simulation results are "…more accurate, compared to the experimental values". How could it be? This sounds awkward, please reword.
4. I do not think that it is a good idea to discuss the hydraulic simulator in the main text – this is rather far from the claimed topic of the thesis. I suggest briefly mention this approach in the main text and move the according material to the Appendix.

5. I think the thesis would benefit from the proper chapter titles, in particular, the "Methodology development" chapter after the "Methodology" chapter may confuse the readers.
6. While the author introduces the density-of-states plots for the band gap as a figure of merit, it would be interesting to see similar plots for other quantities considered in the work, like the effective mass tensor components.

There are also a few typos and minor mistakes in the text, especially in the last chapters of the work. I would advise drawing the author's attention to the following pages: 20, 39, 83, 91, 93, 110, 115, as well as captions to the figures in the ML experiments chapter.

**Provisional Recommendation**

X *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*~~