Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Zahed Allahyari

PhD Program: Materials Science and Engineering

Title of Thesis: Coevolutionary Search for Materials with Optimal Properties in the Space of Binary Systems

Supervisor: Prof. Artem Oganov

Date of Thesis Defense: 6 April 2020

Name of the Reviewer: Sergey Levchenko

I confirm the absence of any conflict of interest	**Signature:**  **Date: 26-02-2020**

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

Brief evaluation of the thesis quality and overall structure of the dissertation

The methodological developments in the thesis by Zahed Allahyari are hard to overestimate. They are at the forefront of modern materials science. By combining first-principles calculations, evolutionary algorithms, and data analysis, Zahed achieves something that has not been achieved before: predicts new materials of practical importance with very high degree of reliability. Combination of ab initio methods and data analysis is an emerging methodology that requires training new kind of specialists, and this thesis fulfills this important goal.

Despite the above success, the thesis overall does not create a good impression. First, it remains unclear what was contribution of the candidate to the development of the methodology and calculations. Was he the one who came up with the idea, or was he the one who implemented the idea in a computer program, or was he the one who used the method for calculations, or all of this together?

Second, the organization of the thesis and the quality of writing does not pay justice to the very advanced content. Almost all figure and table references are wrong. The structure of the writing creates an impression of repetitive text, although this problem is not very severe. Also, there are many places in the text that require clarification (see the list of detailed comments below).

Finally, a scientific analysis of the results is missing. The results in the whole thesis are presented in a form of a report, without any attempt to develop some deeper understanding of these results.

If these issues are addressed, the thesis will be of outstanding quality, because, as said above, the content is extraordinary.

The relevance of the topic of dissertation work to its actual content

The topic of dissertation work exactly reflects the actual content of the thesis.

The relevance of the methods used in the dissertation

The methods are an important part of the thesis. The whole dissertation is about development and application of new methodology for materials design.

The scientific significance of the results obtained and their compliance with the international level and current state of the art

There are three aspects of the study, demonstrating different levels of scientific significance. One aspect is the overall methodology, combining first-principles calculations, evolutionary algorithms, and data analysis approaches such as Pareto optimization and species ordering (Mendeleev numbers). In this regard, the study is at the forefront of scientific research in the field, and will have a significant influence on its further development.

Another aspect is the application of the novel methodology to practical problems. Zahed predicts several new hard, superhard, and magnetic materials. In particular, the finding of some hard phases in Mn-H system is very intriguing. However, this and other exciting results are never discussed in terms of possible mechanisms of such unusual behavior.

The other aspect is the particular choice of data analysis methods. In particular, I find the way of defining the new Mendeleev numbers used in the thesis somewhat arbitrary and not convincing. In fact, it was not clearly demonstrated in the thesis that this was necessary. The claims that the newly defined species ordering is better than previously developed seem to be based on a visual analysis of some plots, without

a quantitative analysis. It is necessary to at least discuss possible systematic approaches to defining Mendeleev numbers that overcome problems of previous approaches. In this context, automated feature selection seems particularly relevant, is a developed field of data analysis, and is used more and more often in materials science, but it is never mentioned in the thesis. Moreover, the statement that a universal species ordering exists for all target properties simultaneously is too optimistic, and requires more discussion. In other words, possible problems and pitfalls of the applied methodology should be better discussed in the thesis.

The relevance of the obtained results to applications (if applicable)

Potential for exciting applications is one of the sides where the thesis shines. Several new hard, superhard, and magnetic materials are predicted with a high degree of reliability (due to the thorough exploration of the structural space). A great interest of industry in the results is easy to anticipate.

The quality of publications

The candidate published three papers in peer-reviewed journals (Q1) and one book chapter (published by Springer Nature Switzerland) on the topic. The publications are of very high quality, scientifically solid, and have a visible impact on the field (there are about 40 citations since 2017).

Detailed comments

"transitional symmetry" -> "translational symmetry"

"In a broad sense, the energy of a material is the Gibbs free energy" - there are other tehrmodynamic potentials that can describe state of a system; explain why you put emphasis on the Gibbs free energy

Eq. 1.2: explain main approximations that lead to Eq. 1.2 - non-relativistic, stationary

You call Eq. 1.4 a Schrodinger equation in a few places in the text, but this is just a Hamiltonian

"The many-body Schrodinger equation (Eq. 1.4) provides the exact wave function for studying the behavior of the system" - how can it be exact after all the approximations that had to be made to derive this equation?

"but it is not solvable analytically" - it is, but only for a few cases

"In this approach, each particle interacts with a density of electrons rather than other particles." - This is true for Kohn-Sham approach, not for the original formulation of DFT, where there are no single particles at all.

"Pauling exclusion principal" - PAULI exclusion PRINCIPLE

"These terms are due to the Pauling exclusion principal" - this is misleading and in general not true; in DFT exchange-correlation terms are just whatever is beyond kinetic energy and Hartree repulsion

"uses the density of electrons as in a homogenous electron gas" -> "uses the energy of electrons as in a homogenous electron gas"

"Although, LDA leads to good results for some properties" - certainly not for all classes of materials; please clarify

"minimums" -> minima

"in which it is usually necessary to start in a good region of the energy landscape and move toward the local minimum exploring the neighborhood regions." - it is unclear whether this statement is also relevant to metadynamics, because later you say "metadynamics, on the other hand"; please polish this part

"One of the important advantages of an evolutionary algorithm is its capability to predict several low-energy metastable structures in addition to the stable ones." - this sounds strange; what do you call "the stable ones"? it seems you mean the global minimum, then it should be "the most stable one"?

I think you should mention that an evolutionary algorithm cannot guarantee finding global minimum

"in which N of the best fractions of solutions in each generation" - clarify what "N of the best fractions of solutions" means

"Figure 7. In PESA, hypervolumes are defined to help select the solutions." - can the hypervolumes be shown on the figure?

"For a nondominated solution, the strength is equal to the number of other solutions it dominates divided by the size of the internal group N. For a dominated solution, the strength is obtained by dividing by N the sum of the strengths of nondominated solutions that dominate it." - I think there is a confusion of strength and fitness in the description, please check and clarify

"To improve the convergence with the best optimal solutions" -> "...convergence to the best ..."

"this method is also useful when a divergent Pareto front is obtained" - explain what a divergent Pareto front is

"he reason for this difference is that in this research, the stable Mo4N3 structure was not found, and, as a consequence, metastable structures of Mo2N and Mo3N2 appeared on the convex hull (Fig. 9a)." - the situation with Mo4N3 is unclear; explain why it was not found and whether this is important or not; is it because of the MO optimization? similar questions for the other compounds

"and is therefore much more computationally expensive and sensitive to numbers." - clarify in the text what it means "sensitive to numbers"

Figures 10 and 11: it would be good to show names of some of the best materials/structures on the plots

"In this paper we propose a universal method of generating such a one-dimensional sequence of elements at any pressure" - mentioning pressure in this context is confusing; what about temperature then? please clarify in the text

"Based on this premise, we define the chemical scale and MN from these atomic properties." - unclear who is "we" here

"A significant correlation between the Pauling electronegativity chi and atomic radius Ra (Fig. 2)" - wrong figure reference, also further in the text

"and the database was set up so as to contain neither duplicates nor very unstable structures (whose energy is more than 0.5 eV/atom above the convex hull)" - how the convex hull was determined?

"The database contains the crystal structure information for 1,591 binary and 80 unary systems." - clarify how this leads to 500,000 materials

"For magnetization, only a small number of binary systems (i.e. 446 systems which is about 15% of the whole systems) are studied in our database." -> "For magnetization, we consider only a small number of binary systems (i.e. 446 systems which is about 15% of all systems in our database)."

"the more exothermic is a material" - a material cannot be exothermic, only a reaction can

"promising materials (more exothermic - shown in orange and red)" - again, a material cannot be exothermic, and consequently more or less exothermic

table 2 is not referenced in the text

check figure numbering in chapter 4, the references are all wrong

"However, the most promising materials (i.e. the composition of different elements of the Periodic Table with heavy elements such as Cm, Bk, Cf, Es, and Fm)" - discuss how the nuclear instability and cost of these elements can affect their usefulness

"The electronegativities of many elements calculated at various pressures.58" - check the sentence

cite Ref. 58 properly

"This can help to predict new materials at arbitrary pressure, only by having a number of relevant data on other systems and plotting them onto the well-organized map produced by the USE." - can you give an example of such prediction? I guess you refer to the later chapters, but this is not clear at this point

Figure 19: why not refer to the corresponding panels in Figure 15? is there a significant difference? besides, you already call your MN USE, but here you label it "our MN"

"If this space is created by ordering the elements by their atomic numbers, we observe a periodic patchy pattern (Fig. 20a)." - wrong figure reference

"In 1984, Pettifor suggested a new quantity, " - this paragraph seems to be a repetition of what was already discussed in chapter 4

"Fig. 21a shows the overall linear correlation" - check and correct all figure references in this chapter as well

"The underlying ab initio structure relaxations and energy calculations were performed using the density functional theory with the projector augmented-wave method (PAW) as implemented in the VASP code." - tell which functional was used and why

"Many scientific and engineering problems, such as predicting novel materials that improve upon all critical properties of the known ones, involve an optimization of multiple conflicting objectives." - as far as I can tell, the paragraph that starts with this sentence is a repetition of what was said in chapter 3; please improve the integrity of the thesis

Figure 21a: "1th generation" -> "1st generation"

"In addition, our calculation revealed new, known, and potential hard and superhard binary systems" - unclear what you mean: new (previously unknown) or known materials, or both, why "and potential hard" (potentially hard?)

"using the GGA-PE functional" - GGA-PBE?

"or about one-fifth of all the systems that can be composed of 74 elements" - clarify how this is counted; I guess you mean only unary+binary systems, but even with binary systems there are many possible stoichiometries, including small concentrations (doping)

"each with about 10^2 possible compositions" - why only 100 compositions? there is practically infinite number of compositions

"2775 systems, each with about 10^2 possible compositions, each composition having an astronomically large number of possible structures" - it is unclear what you mean by systems, compositions, and structures

"In this work, new low-energy structures with high hardness were discovered for these systems (Table 5)." - wrong table number? check other references to tables

figure references in Chapter 6 are also wrong; check carefully references to figures and tables in the whole thesis

"Two thermodynamically stable compounds, P3m-1-Mn2H and P63/mmc-MnH, were discovered in this system," - explain how thermodynamic stability was evaluated

"A recent theoretical study of chromium carbides33 using the evolutionary algorithm USPEX" - I suggest to write "Our recent study..."

From the beginning of the thesis it is unclear why it is useful to search for hard materials, if all of them are not as hard as BN and C. What other properties make materials more attractive than BN and C? Please explain this somewhere in the introduction

"Diamond and cubic BN possess the best properties but are metastable at normal conditions." - At what conditions does the metastablity of diamond and BN become a problem in practice?

Provisional Recommendation

☐ *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*