Istituto SPIN
Genova, L'Aquila, Napoli, Roma, Salerno

Consiglio Nazionale delle Ricerche

To the EDITOR of Journal of Applied Physics

Chieti, February 12th, 2022

Dear Editor,

We would like to submit the enclosed manuscript, entitled "*Towards machine learning for microscopic mechanisms: a formula search for crystal structure stability based on atomic properties*" for publication in the Journal of Applied Physics.

As well known, many relevant properties of functional materials are particularly difficult to predict based on simple "rules of thumbs", due to their inherent complexity. To tackle the problem, over the last decade researchers have used different machine learning (ML) methods to analyze these properties. However, not always these ML methods are characterized by a good interpretability and accuracy, the former aspect creating barriers in adopting these methods.

To overcome this drawback, we propose a method combining first-principles simulations based on density functional theory (DFT) and machine learning approaches based on regression methods to identify accurate descriptors, correlating different fundamental atomic quantities (such as electronegativity, electronic energy levels, atomic shells radii) and aimed at describing the energetics of different crystalline phases. By means of combinatorial methodologies, we systematically constructed a set of possible features and used Linear Regression (LR) to identify a final "best formula". As a test case for verification, we used our approach to predict the total energy difference between Rock-Salt (RS) and Zinc-Blende (ZB) phases of 82 binary compounds. Our results show that:

1). Formulas identified by our approach can successfully classify the stable phase between RS and ZB and can quantitatively predict the energy difference with a good accuracy.

2). From the selected formulas by LR, we also learn that atomic shell radii (in particular related to the p shell of the cation) play an important role in deciding the stable crystalline phase. This shows that ML methods can support researchers in better understanding the microscopic mechanisms underlying various phenomena, by suggesting which are the relevant ingredients at play.

3). We further studied the importance of individual atomic properties inside the ML-suggested formulas by using a procedure called "formula optimization". This procedure is shown to improve the performances of the formulas at a quite small computational cost.

We believe that our work can play an important role in bridging the gap between Artificial Intelligence and the Solid-State Physics community. This will be highly beneficial, since we expect that, upon increasing computational power, big data will be correspondingly generated and different machine learning techniques will be increasingly used.

Genova: C.so F. M. Perrone, 24 - 16152 - Genova - ☎ +39 010 6598710 - Fax: +39 010 6506302
L'Aquila: c/o Dip.to di Fisica - Via Vetoio - 67010 - Coppito (AQ) - ☎ +39 0862 433014 - Fax: +39 0862 433033
Napoli: c/o Dip.to di Fisica - Complesso di Monte S. Angelo - 80126 - Napoli - ☎ +39 081 676438 - Fax: +39 081 676446
Roma: Area della Ricerca di Tor Vergata - Via del Fosso del Cavaliere, 100 - 00133 - Roma - ☎ +39 06 45488392- Fax: +39 06 45488018
Salerno: c/o Dip.to di Fisica - Via Giovanni Paolo II, 132 - 84084 - Fisciano (SA) ☎ +39 089 969146 - Fax: +39 089 969659

Partita IVA IT 02118311006 – C.F. 80054330586

We also remark that, in the manuscript, we have shown only one, though paradigmatic, example. However, clearly the same approach can be applied to any physical problem in which a compound's property of interest can be described in terms of basic properties (i.e., atomic properties and not only) of the constituents. As such, the methodology, along with the related code that we have developed (that is freely available and Open-Source), looks very general and can pave the way to applications in many different fields.

In summary, we strongly believe that our manuscript fully meets the criteria of the highest quality research and well demonstrates the significance of our results to the broad readership of the Journal of Applied Physics.

As potential referees, we suggest;

1) Prof. Claudia Draxl, Humboldt-Universität zu Berlin, claudia.draxl@physik.hu-berlin.de.

2) Prof. Stefano Sanvito, Trinity College Dublin, sanvitos@tcd.ie

3) Prof. Chris Wolverton, Northwestern University, c-wolverton@northwestern.edu

Sincerely yours,

Udaykumar Gajera

(On behalf of all the authors)