Warsaw, 5.04.2025
Dr. hab. Carmine Autieri, prof. IF PAN
e-mail: autieri@magtop.ifpan.edu.pl
Institute of Physics, Polish Academy of Science

Review of the doctoral dissertation of Dario Massa, M.Sc., entitled: "Quantum-Informed Descriptors in Materials Informatics"

The doctoral dissertation of Dario Massa, entitled: "**Quantum-Informed Descriptors in Materials Informatics**" was submitted at the the University of Warsaw (UW) under the supervision of Dr. hab. Stefanos Papanikolaou, prof. NCBJ and Dr. hab. Piotr Sankowski, prof. UW. The student conducted this study mainly in NOMATEN Center of Excellence in Multifunctional Materials research in Otwock, Poland.

The dissertation is in the form of a collection of papers with 4 manuscripts; the first and third of these manuscripts have been published in international peer-reviewed journals, while two are currently available only on the ArXiv repository at the time of the submission. The forth paper of the list is published on NPJ Computational Materials at the time that I am writing the review. The papers included in this Ph.D. thesis have garnered 2 citations on Scopus https://www.scopus.com/authid/detail.uri?authorId=584758033009 at the present date, the number is low because all papers were published in the last year.
Dario Massa's dissertation takes 121 pages. The thesis is composed by an Introduction. The second chapter contains the 4 manuscripts, while in the third and last chapter there is a short summary of the published results. The author stated his contributions to the different research articles with a few lines before the attachment of every paper.

Section 1 is an introduction on the field of the materials informatics. The authors wrote about the role of the descriptors and how the electron charge density with the help of machine learning can be used in the context of the density functional theory.

Section 2 is the core of the thesis. It consists of 4 subsections where the author reported the 4 manuscripts. Paper I study the properties of hydrogen atoms in fcc crystals from a static point of view, while paper IV study the dynamics of hydrogen atoms in bulk Mg. Papers II focuses on Transfer Learning in Materials Informatics. Paper III is the study of the properties of doped systems using crystal graph neural networks.

Subsection 2.1 reports a paper on hydrogen atoms in fcc crystals that was published in *Acta Materialia*. Dario Massa is the first and corresponding author. The authors of the paper introduced a machine learning method using electron charge density from first-principles to predict behavior of hydrogen-doped compounds. Using hydrogen in face-centered cubic crystals as an example, they show that differential charge density profiles capture key crystal and defect interaction properties. Radial Distribution Functions (RDFs) reveal screening effects and, along with hydrogen Bader charges, correlate with various atomic properties. RDFs and charge features also relate to hydrogen migration energy barriers. Unsupervised learning on RDFs helps classify materials.

Subsection 2.2 reports the second manuscript published exclusively on the arxiv repository. The Ph.D. candidate is first and corresponding author. The authors claim that the paper was accepted in a conference, but the permission to the link was not granted to to the reviewer. In this manuscript, the authors propose supervised deep learning methods to extract descriptors from ab-initio Electron Charge Density profile images, and from crystallographic text via LLMs. Their focus is on reducing architectural complexity, data dimensionality, and training costs for efficient solutions in materials design. They claim that their

findings show strong predictive performance from limited inputs, by paving the way for inverse design and low-cost AI-driven materials discovery.

Subsection 2.3 is devoted to a paper published in *AIP Advanced* where the Ph.D. candidate is first author. The authors focus on pure metallic bulk host matrices, namely, Al, Ni, Mo, and Au, they introduce defects and analyze the formation energy and the bulk modulus of the doped system. In the machine learning framework, they utilize the MEGNet description and they validate the results using density functional theory.

Subsection 2.4 includes a paper published in *npj Computational Materials volume 11, Article number: 85 (2025), DOI: 10.1038/s41524-025-01555-z* where the Ph.D. candidate is the third author. The research employed a hybrid approach that integrates Density Functional Theory calculations with Machine Learning. By implementing an active learning procedure, the model was trained through the collection of a diverse dataset of atomic configurations generated during the simulation. This dataset was then used to fine-tune advanced neural network-based interatomic potentials, achieving a balance between computational efficiency and predictive accuracy. The authors studied hydrogen diffusion coefficients and the activation energy barrier. A systematic investigation of varying hydrogen concentrations also provided novel insights into hydrogen pair formation. The study highlights the transformative potential of machine learning in materials science for investigating complex atomic dynamics in defective systems.

In **Section 3**, the author provides a two pages summary about the presented papers, highlighting the main results of the research. **Section 4** is devoted to other research activities of the author within the same field. These include 3 manuscripts/preprints and 1 book chapter where the PhD candidate appears in the second or third position in the authorship list.

After analyzing the above result, I have some questions and remarks which are reported below with the latter Q and R in addition to their number:

Q1) In manuscript I, you studied the Mn as host. In paper 3, you study Ni as host. Both were investigated without magnetism, however, Density Functional Theory is a zero temperature theory where these compounds are magnetic. How to reconcile these two effects? Are you assuming to simulate these effects above the magnetic critical temperature? Any reason to confirm the non-magnetic phase of Ni/Mn in DFT would well represent the experimental paramagnetic phase above Curie temperature??

Q2) Related to the manuscript I, you used the Principal Component Analysis (PCA). Will be possible to slightly relate PC1 and PC2 to some physical quantities as the electronegativity, energy barriers or the size of the host atoms?

Q3) Related to manuscript III, the authors preferred to study properties related to the total energy as the bulk modulus, formation energy and Fermi level rather than simple electronic properties as DOS at the Fermi level, bandwidth, etc… The total energy is simpler to obtain but the DOS is relevant for your metallic system. Any reason why you did not investigate that?

Q4) Related the manuscript IV, once the H_2 molecule is formed in Mg, how does it break? What is the mechanism that breaks the molecules and what are the differences between these mechanisms in free molecules and molecules trapped in a crystal?

Q5) Related to paper IV, the Diffusion coefficient is linear in hydrogen concentration in Figure 6a. Why the activation energy Ea is not linear in Figure 6a??

Q6) In the first three papers where the Phd student is first author, there is not so much connection with experiments? Can you tell us something about the connection of your papers with the experiments??

Q7) In condensed matter physics, it is usually used the special quasirandom structure (Phys. Rev. Lett. 65, 353, 1990) to simulate the randomness of the impurities and their interaction. Why you do not use that??

R1) In paper I, the authors used the word alloy for small concentration of hydrogen. I found this misleading. At least in the experimental side, the researchers usually used the words doping/impurities for small concentration like the one used in paper I, while the word alloy is used for disordered systems with high-concentrations of the dopant. The term “solid solution” is used for ordered alloyed systems.

R2) On manuscript IV, section 2.2 there is a typos in the subscript of the Hydrogen. Compounds with different compositions are reported to have the same stoichiometry.

R3) On manuscript IV, figure 4a was reported without the relevant colors.

In summary, the motivation for this research is presented clearly and with precision. The thesis is well-edited, and the text is legible and grammatically accurate, making it easier to analyze and evaluate the results. All the theoretical figures (most of them are likely obtained from the Ph.D. candidate) in the four papers are very detailed, striking and well-structured. The research conducted by the Ph.D. candidate is part of the new field of the Materials Informatics, Machine learning and Neural Network applied to physics. This research trend is very popular in the scientific community of natural sciences and beyond that. The first and four publications in the paper collections have appeared in peer-reviewed international journals with a significant impact factor. The author claims that the second paper is accepted in the conference proceedings of NeurIPS2024 which had 25% acceptance rate for the submitted https://www.qualcomm.com/news/onq/2024/12/qualcomm-at-neurips-2024-ai-research-demos-papers. The Ph.D. candidate is the lead author in the first three papers and third author in the last paper.

Despite numerous remarks and questions, the doctoral dissertation of Dario Massa, M.Sc., in my opinion, meets all the quantitative and qualitative criteria for doctoral dissertations in the field of Natural sciences in the discipline of physical sciences. The doctoral dissertation presents the candidate's general theoretical knowledge in physics as well as the ability to independently conduct scientific work. The subject of the doctoral dissertation is an original solution to a scientific problem. I conclude that my evaluation of the PhD thesis is POSITIVE. In connection with the above, I request that Dario Massa, M.Sc., be admitted to the next stages of the doctoral procedure.

Signature 6 April 2025