

Proceeding Paper

Machine Learning-DFT-Based Approach to Predict the Electrical Properties of Tin Oxide Materials [†]

Hichem Ferhati ^{1,2}, Tarek Berghout ³, Abderraouf Benyahia ¹ and Faycal Djéffal ^{1,*}

¹ LEA, Department of Electronics, University of Batna 2, Batna 05000, Algeria; email1@email.com (H.F.); email2@email.com (A.B.)

² ISTA, University of Larbi Ben M'hidi, Oum El Bouaghi 04000, Algeria

³ Laboratory of Automation and Manufacturing Engineering, University of Batna 2, Batna 05000, Algeria; email2@email.com

* Correspondence: faycal.djeffal@univ-batna2.dz

[†] Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: <https://ecsa-10.sciforum.net/>.

Abstract: The effects of oxygen concentration and growth technique during the deposition process on the electrical properties of tin oxide alloy (SnOx) should be investigated for developing new eco-friendly photosensors and photovoltaic devices. The present work aims to predict the electrical key governing parameters throughout the device developing processes such as the Energy level values and band-gap energy as function of the injected oxygen concentrations. For realization, over 100 data points were collected by modeling the effect of oxygen contents on the SnOx electrical properties using Density Function Theory (DFT). Through extensive Machine Learning (ML) analysis, the impact of the oxygen concentration on the electrical properties and the material type is well predicted, where the applied ML prediction model for band-gap energy showed a good correlation between predicted values and the calculated ones using DFT computations. It is revealed that the combined DFT-ML-based approach can be a powerful tool to study and accelerate the developing of new highly efficient materials for microelectronic applications.

Keywords: tin-oxide; DFT; machine learning; prediction; photosensors

1. Introduction

Tin oxide (SnOx) semiconducting alloys have been considered promising candidates for the next generation of microelectronic materials and have attracted considerable attention in developing high-performance sensing devices (e.g., photodetectors, Gas sensors, photocatalysts, photovoltaics, ... etc.) because of their scalable elaboration techniques, tunable electrical and optical properties, good light-matter interactions, adjustable electronic energy band structures, and excellent gas molecules interaction properties [1–4]. Tuning the electrical and sensing properties of tin-oxide-based alloys can be carried out using deferent experimental approaches such as chemical doping, strain engineering and changing the elemental composition (i.e., tin and oxide). The latter technique is considered as an effective approach to modulate the optical, electrical and structural parameters, where it was demonstrated that the sensing properties of the material are significantly affected by the band-gap energy value and the elemental composition ratio in the SnOx alloy [4,5]. In other words, the oxygen concentration can be varied in the SnOx film, which can modify its electronic and optical properties. Consequently, the impact of oxygen content on the electrical characteristics of SnOx should be investigated to offer new insights in developing eco-friendly and high-performance devices for sensing applications.

In this work, a new modeling framework approach is proposed to predict the electrical key governing parameters throughout the device developing processes such as the

Citation: Ferhati, H.; Berghout, T.; Benyahia, A.; Djéffal, F. Machine Learning-DFT-Based Approach to Predict the Electrical Properties of Tin Oxide Materials. *Eng. Proc.* **2023**, *56*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor(s): Name

Published: 15 November 2023



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Energy level values and band-gap energy as function of the injected oxygen concentrations. To do so, over 100 data points were collected by modeling the impact of the oxygen contents on the SnO_x electrical properties using Density Function Theory (DFT). Through extensive combined DFT-ML (Machine Learning) analysis, the effect of the oxygen concentrations on the electrical properties and the material type (metal, P-type and N-type) is well predicted, where the applied ML prediction model for band-gap energy showed a good correlation between predicted values and the calculated ones using DFT computations.

2. Modeling Frameworks

In this section, multipurpose modeling approaches are considered to predict the electrical properties of SnO_x -based alloys based on combined DFT-ML calculations. The first step will be used to build the required database to forecast the electronic properties of SnO_x thin-film using DFT computations. Secondly, the ML- based calculation will be employed to predict the impact of the oxygen concentration on the electrical properties and the material type of the tin-oxide materials.

2.1. DFT Calculations

In the present work, the DFT-based calculation technique was used to perform the band structure electrical properties of SnO super-cell [5–7]. The band structure calculations were carried out using generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) functional and the Heyd-Scuseria-Ernzerhof screened Coulomb (HSE06) hybrid-functional [5]. The experimental lattice constants are used for the initial structure of SnO. Moreover, the tetragonal system of the rutile SnO with stable crystalline phase is considered. In order to study the impact of the oxygen concentration on the SnO_x structure, additional interstitial oxygen atoms were introduced in the SnO super-cell at its lattice. In addition, the SnO semiconductor type is determined from the Fermi-level position provided by DFT calculations. It is important to note that the SnO semiconductor type obtained from DFT simulations is in good agreement with the experimental results [4].

2.2. ML Algorithm

Machine learning (ML) has demonstrated to be a powerful tool in overcoming high-cost experimental tests and practice limitations in understanding the parameters affecting material properties and their relationships [8]. Therefore, the use of ML techniques in the development of new materials for sensing applications, including SnO_x alloy, is on an upward trajectory. In this work, we explore the use of ML techniques to assess the impact of oxygen concentration on the electrical behavior of SnO_x alloy for sensing applications. The ML model has been trained using our DFT-based calculation database. Correlation analysis and machine learning algorithms have been employed to study key parameters affecting material properties and their interactions. DFT-ML predictive approach has been developed to determine the material type and the band-gap energy values associated with oxygen concentration, offering fast and crucial guidance for experimental elaboration of SnO_x-based sensing devices.

3. Results and Discussion

The obtained band structure of SnO₂ is depicted in Figure 1 using DFT calculations, it can be shown from this figure that the SnO₂ material exhibits a wide band gap of 3.56 eV with a direct transition mechanism at G symmetric point. The obtained results make SnO₂ a potential material for developing Ultra-Violet photodetectors and gas sensors. Moreover, in order to investigate the impact of the oxygen concentration on the material band-gap energy, Figure 2 plots the variation of band-gap energy values as function of oxygen concentrations. It is clearly shown that the introduction of oxygen can induce

variations on the bond angle caused by the disorder of octahedra, which leads also to increase the tin oxide band gap values. The tunability of band-gap energy values and the type of material can open up new paths in developing multispectral photodetectors and new devices for gases sensing.

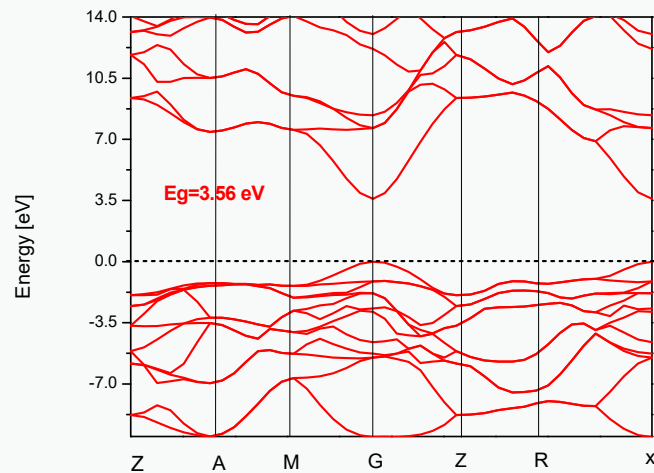


Figure 1. Band structure of thin-film SnO₂ based on DFT calculations.

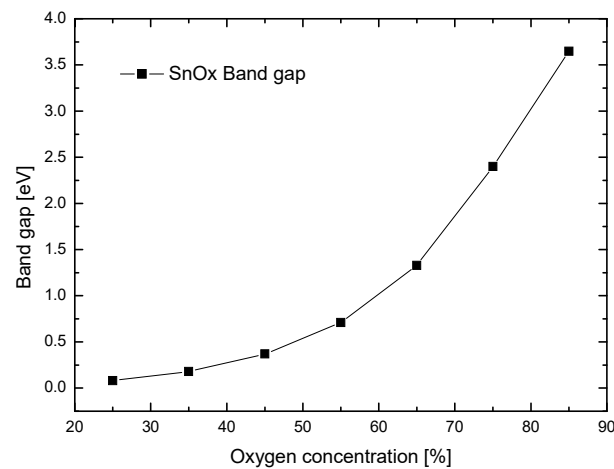


Figure 2. Variation of band gap energy as a function of the oxygen concentration for SnO_x material.

From an ML modeling perspective, this collected data can be visualized as presented by Figure 3 for a better understanding of both data change and complexity respectively. Figure 3a shows variation of energy gap with respect to oxygen concentration. Likewise, it represents differences in data linked to doping type. Regarding the former's data drift, it is a somewhat exponential variation referring to a rapid change in data characteristics. Meanwhile, the doping type data points are divided into three categories namely metallic (m) and p-type (p), n-type (n) semiconductors. The obtained data distributions show a kind of data imbalance that is perfectly revealed by class ratio calculation in Figure 3b. Besides, it is worth mentioning that data patterns in this case do not show any signs of noise or outliers in collected measurements, which requires less data processing except for normalization. Under such circumstances, ML modeling requirements face two main

challenges, particularly data drift (i.e., continuous change in data characteristics) and data complexity (i.e., class imbalance). To combat such challenges, this work proposed the following contributions [9].

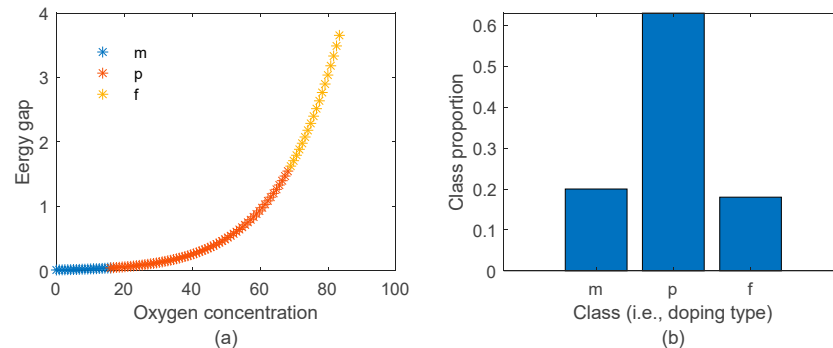


Figure 3. Visualizing data from perspective of ML modeling: (a) regression function and class scatters; (b) Class proportion.

Adaptive learning: Adaptive learning rules of the long-short term memory neural network are involved in this case, while training a single-layer neural network for both energy gap prediction and dope type classification. This ensures that the learning model is kept up-to-date by tracking only upcoming important new data.

Data sampling: To solve the class imbalance problem, the synthetic minority oversampling technique is involved in this case [10]. Such a technique helps to overcome this variation in class proportion ratio by generating synthetic examples of the minority class, thereby enabling fair representations of data points and preventing model bias towards majority class. However, an important issue could arise as a result of this contribution. First, since adaptive learning rules from long-short term memory network experience deeper representations, 100 single-dimensional points are considered a lightweight problem to solve. This can lead to a so-called underfitting problem. Second, generating data using the aforementioned synthetic minority oversampling technique may result in different drawbacks related to increased risk of misclassification due to the difficulty in generating informative samples. Therefore, an additional process of monitoring learning and validating the ML model is urgently needed. Consequently, the cross-validation technique constitutes a third contribution to this work.

Crossvalidation: cross-validation allows for efficient use of data, thereby increasing the robustness and reliability of performance estimation by dividing data into different folds and performing tests on the entire dataset [11]. In this work, the neural network used is subjected to manual tuning following simple error-trial learning rules. The following Table 1 presents the final parameters achieved for the classification and regression problems.

Table 1. Parameters tuning results.

Hyperparameters	Regression	Classification
Maximum number of epochs	50	250
Mini-batch size	10	5
Neurons	20	30
Learning algorithm	Adam optimizer	RMS propagation
Initial learning rate	0.01	0.1
Gradient threshold	1	1
L2 regularization	0.0001	0.0001

Training and evaluation of the discussed ML model goes through a three-fold cross-validation process for both problems. The performance of ML modeling for the regression process is evaluated using well-known metrics including root mean square error (RMSE), root mean squares (RMS), and mean absolute error (MAE). The expected result of these measurements is to get closer to “zero” for greater accuracy of approximation and generalization. Additionally, the famous R^2 metric is also included while when its value approaches “one” it refers to better performance. Similarly, classification performance evaluations involve four different metrics well used in the literature, including accuracy, F1 score, recall, and precision. The expected result of the classification metrics is to approach the value “one”. In this work, we focused on collecting results from the validation set because they are more important than the training set. Because in this case, they make it possible to observe both generalization and approximation capacities at the same time. It is also worth mentioning that such experiments are conducted on i7 processing unit computing power of 16 GB RAM and 12 MB cache memory. Additionally, MATLAB r2023a library is used as the main programming platform for this application. Table 2 is dedicated to summarizing results obtained from the whole experiment. On the one hand, and discussing results obtained on the regression problem, the prediction models behave in a similar way. This means that they have the ability to induce stability even across different datasets/folds. This is proven by the performance evaluation results of RMSE, MSE and MAE respectively in validation folds from 1 to 3. Their mean values and standard deviation also explain the similarity of results obtained. Likewise, R^2 values show the same patterns of prediction stability and similarity between other folds, achieving an excellent performance of 0.71. On the other hand, the averaged values of the classification metrics show impressive results in terms of stability and accuracy, reaching 0.99 for all metrics with a very small standard deviation of 0.008. While most models show 100% prediction performance.

Table 2. Summary of obtained results.

Regression				
Crossvalidation Folds	RMSE	MSE	MAE	R²
1	0.51	0.26	0.28	0.72
2	0.41	0.17	0.20	0.71
3	0.56	0.31	0.38	0.69
Average	0.49	0.25	0.29	0.71
Standard deviation	0.1341			0.01
Classification				
Crossvalidation folds	Accuracy	F1 score	Recall	Precision
1	1	1	1	1
2	1	1	1	1
3	0.98	0.98	0.98	0.98
Average	0.99	0.99	0.99	0.99
Standard deviation	0.008			

Overall, obtained results are promising for application of such modeling process in predicting the electronic, optical and electrical properties of wide band gap materials, which can be effective for the design of alternative optoelectronic and gas sensing devices. However, certain points/limitations must be taken into account when generalizing such investigations to real applications. These points can be addressed as follows.

- 100 data points are somehow too small for results generalizability in terms of regression;

- 3 different types of classes with different proportions under a too small set of data can create problems of misrepresentation of data when using generative models during data balancing;
- There is a higher probability of overfitting when trying further number of cross validation folds.

4. Conclusions

In this paper, the effect of oxygen concentration on the electronic properties of SnOx material is investigated. First-principles calculations are carried out to estimate the band gap of SnOx material for various oxygen levels. It is found that the material band gap increases with the oxygen content increase to reach its highest value of 3.56 eV corresponding to oxygen-rich SnOx. Machine Learning analysis are then performed to predict the electronic properties and material type of SnOx alloy with varied oxygen containing for photodetectors and Gas sensing applications. Adaptive learning rules of long-term memory under cross-validation techniques are involved during the ML modeling process. Additionally, synthetic minority oversampling techniques are integrated into the classification process. The whole methodology turns out to be very effective for both regression and classification, achieving impressive results, especially for classification. Regarding future opportunities, and consistent with the limitations discussed in Section 3, future opportunities will revolve around: (i) targeting an even more massive and complex dataset; (ii) discuss different ML tools under different adaptive learning algorithms; (iii) discuss other generative modeling and subsampling tools to address class imbalance issues. The obtained results make the proposed approach a power tool for fast and accurate predicting the electrical properties metal oxides for sensing applications.

Author Contributions: Conceptualization, H.F., T.B. and F.D.; methodology, F.D.; validation, H.F., T.B. and F.D.; formal analysis, H.F., T.B., A.B. and F.D.; investigation, H.F., T.B., A.B. and F.D.; data curation, H.F., T.B. and F.D.; writing—original draft preparation, F.D. and T.B.; writing—review and editing, H.F., T.B., A.B. and F.D.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, G.; Lu, W.; Li, J.; Choi, J.; Jeong, Y.; Choi, S.Y.; Park, J.B.; Ryu, M.K.; Lee, K. V-Shaped Tin Oxide Nanostructures Featuring a Broad Photocurrent Signal: An Effective Visible Light-Driven Photocatalyst. *Small* **2006**, *2*, 1436–1439.
2. Dalapati, G.K.; Sharma, H.; Guchhait, A.; Chakrabarty, N.; Bamola, P.; Liu, Q.; Saianand, G.; Krishna, A.M.; Mukhopadhyay, S.; Dey, A.; Wong, T.K. Tin oxide for optoelectronic, photovoltaic and energy storage devices: A review. *J. Mater. Chem. A* **2021**, *9*, 16621–16684.
3. Ferhati, H.; Djeflal, F.; AbdelMalek, F. Towards improved efficiency of SnS solar cells using back grooves and strained-SnO2 buffer layer: FDTD and DFT calculations. *J. Phys. Chem. Solids* **2023**, *178*, 111353.
4. Djeflal, F.; Martin, N.; Ferhati, H.; Benhaya, A. Tunable band-selective photodetector based on sputter-deposited SnOx thin-films: Effect of reactive gas pulsing process. *J. Alloys Compd.* **2023**, *968*, 171851.
5. Kumar, M.; Askari, S.S.A.; Pandey, P.S.; Singh, Y.; Singh, R.; Raghuwanshi, S.K.; Singh, G.K.; Kumar, S. Experimental Investigation and DFT Study of Tin-Oxide for Its Application as Light Absorber Layer in Optoelectronic Devices. *IEEE Access* **2023**, *11*, 23347–23354.
6. Djeflal, F.; Ferhati, H.; Benyahia, A.; Dibi, Z. Performance analysis of SnS photodetector using strained SnO2 stacked layer: Numerical simulation and DFT calculations. *Microelectron. Eng.* **2023**, *273*, 111961.
7. Farah, S.E.; Dibi, Z.; Ferhati, H.; Djeflal, F. DFT-FDTD modeling of a new broadband mid-infrared IGZO thin-film phototransistor based on black phosphorus capping layer incorporating intermediate metallic film. *J. Phys. Chem. Solids vol.* **2022**, *162*, 110528.
8. Bandaru, N.; Enduri, M.K.; Reddy, C.V.; Kakarla, R.R. Aspects of effectiveness and significance: The use of machine learning methods to study CuIn1-xGaxSe2 solar cells. *Sol. Energy* **2023**, *263*, 111941.
9. Berghout, T.; Benbouzid, M.; Amirat, Y. Towards Resilient and Secure Smart Grids against PMU Adversarial Attacks: A Deep Learning-Based Robust Data Engineering Approach. *Electronics* **2023**, *12*, 2554. <https://doi.org/10.3390/electronics12122554>.

10. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *Ecol. Appl.* **2011**, *30*, e02043. <https://doi.org/10.1613/jair.953>.
11. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: Boston, MA, USA, 2009. https://doi.org/10.1007/978-0-387-39940-9_565.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.