PRACTICAL DATA SCIENCE: EXAMINING THE CORRELATIONS BETWEEN STRUCTURAL AND ELECTRONIC PROPERTIES OF DIFFERENT PHASES OF TiO_2 NANOPARTICLES

Hasan KURBAN[1, +]

[1]Computer Engineering Department, Siirt University, 56100 Siirt, Turkey

Computer Science Department, Indiana University, Bloomington, 47405 Indiana, USA

hakurban@gmail.com

Abstract

In this work, we analyze the correlations between structural and electronic properties of anatase, brookite and rutile phases TiO_2 nanoparticles (NPs) using data science techniques. For this purpose, we use the geometries of three phases TiO_2 NPs under heat treatment obtained from molecular dynamics (MD) simulations in the frame of DFTB+ code. We investigate the relationships among electronic properties of TiO_2 and order parameter (R or segregation phenomena) & nearest number contacts (n). In this architecture, the correlations among HOMO, LUMO, Energy gap (E_g), Fermi energy (E_f), R and n have been analyzed. Our results show that there is a moderate negative correlation between R_O and E_g in the brookite and rutile phases, but a strong linear correlation between these two variables in the anatase phase. Additionally, in the brookite phase, the positive linear correlation between R_{Ti} and E_g is noteworthy. Moderate linear correlation was observed in the anatase phase and positive in the rutile phase. The positive linear dependence of n_{O-O} and E_g in brookite phase is remarkable. No strong correlation was observed in any phase between n_{Ti-Ti} and E_g. In the brookite phase, n_{O-Ti} has an almost perfect negative correlation with E_g.

Keywords: Data science, Statistical learning, Materials Science, Nanoparticles, Data analytics.

[+] This paper has been presented at the ICAT'20 (9th International Conference on Advanced Technologies) held in Istanbul (Turkey), August 10-12, 2020.

PRATİK VERİ BİLİMİ: TiO_2 NANOPARTİKÜLLERİNİN FARKLI FAZLARININ YAPISAL VE ELEKTRONİK ÖZELLİKLERİ ARASINDAKİ İLİŞKİLERİN İNCELENMESİ

Özet

Bu çalışmada, veri bilimi tekniklerini kullanarak anataz, brookit ve rutil fazlar TiO_2 nanopartiküllerinin (NP) yapısal ve elektronik özellikleri arasındaki korelasyonları analiz edilmiştir. Bu amaçla DFTB+ kodu çerçevesinde moleküler dinamik (MD) simülasyonlarından elde edilen ısıl işlem altında üç faz TiO_2 NP'lerin geometrileri kullanılmıştır. TiO_2'nin elektronik özellikleri ve sıra parametresi (R or ayrışma fenomeni) & en yakın numara kontakları (n) arasındaki ilişkiler araştırıldı. Bu mimaride, HOMO, LUMO, Enerji açığı (E_g), Fermi enerjisi (E_f), R ve n arasındaki korelasyonlar analiz edilmiştir. Sonuçlarımız, brookite ve rutil fazlarda R_O ve E_g arasında orta derecede negatif bir korelasyon olduğunu, ancak anataz fazında bu iki değişken arasında güçlü bir doğrusal korelasyon olduğunu göstermektedir. Ek olarak, brookit fazında, R_{Ti} ve E_g arasındaki pozitif doğrusal korelasyon dikkate değerdir. Anataz fazında orta derecede doğrusal korelasyon, rutil fazda pozitif olarak gözlendi. Brookite fazında n_{O-O} ve E_g'nin pozitif doğrusal bağımlılığı dikkat çekicidir. n_{Ti-Ti} ve E_g arasındaki hiçbir aşamada güçlü bir korelasyon gözlenmemiştir. Brookit fazında n_{O-Ti}, E_g ile neredeyse mükemmel bir negatif korelasyona sahiptir.

Anahtar Kelimeler: Veri bilimi, İstatistiksel öğrenme, Malzeme Bilimi, Nanopartiküller, Veri analizi.

1. Introduction

Data science has recently attracted attention because it provides a better understanding of generated data obtained from experimental and computational materials science. In this context, machine learning (ML) has widely used over material science (MS) applications [1]. Especially, the increase in the amount of raw data requires using data science techniques to figure out hidden and important information over complex data sets. In MS, identifying correlations between large amounts of structural data and

electronic properties is significant and, in this work, we show that this task can efficiently be done using data science methods.

In our previous study, we analyzed the atomic data produced in the anatase, brookite and rutile phases TiO_2 nanoparticles (NPs) [2] at different temperatures. In this work, we do deep analysis between structural parameters and electronic properties of anatase, brookite and rutile phase TiO_2 nanoparticles (NPs) using data science methods. Herein, we used the geometrical parameters obtained from density-functional tight-binding method (DFTB) based on the DFTB+ code [3] with the hyb-0–2 [4, 5] and measured the correlations among order parameter (R) and nearest number contacts (n), HOMO, LUMO, Energy gap (E_g), Fermi energy (E_f).

2. Methodology and Experimental Results

Pearson correlation measures linear correlation between two numerical variables and takes values between -1 and 1. Value close to -1 indicates that there is a strong negative linear correlation between the two variables. Similarly, the approximation of the value to 1 indicates that the strong linear correlation between the variables is positive. Approaching the correlation value between the two variables to 0 means that the linear correlation between them weakens and the value of 0 means that the two variables are linearly independent. In Fig. 1, in the anatase, brookite and rutile phases, Pearson correlations of HOMO, LUMO, E_g, and E_f obtained at different temperatures (0-1000K), with the order parameters R_O and R_{Ti} and nearest number contact variables n_{O-O}, n_{Ti-Ti} Ti are shown. Moreover, as an example, the order parameter vs HOMO, LUMO and E_f graph is plotted to see explicitly how to correlate them under heat treatment (see Fig. 2). From the top down, the figures demonstrate the results for anatase, brookite and rutile phases, respectively. The code, which has high-resolution data visualization options and written for this study, and the data can be accessed via this GitHub link. We explain the findings in detail below.

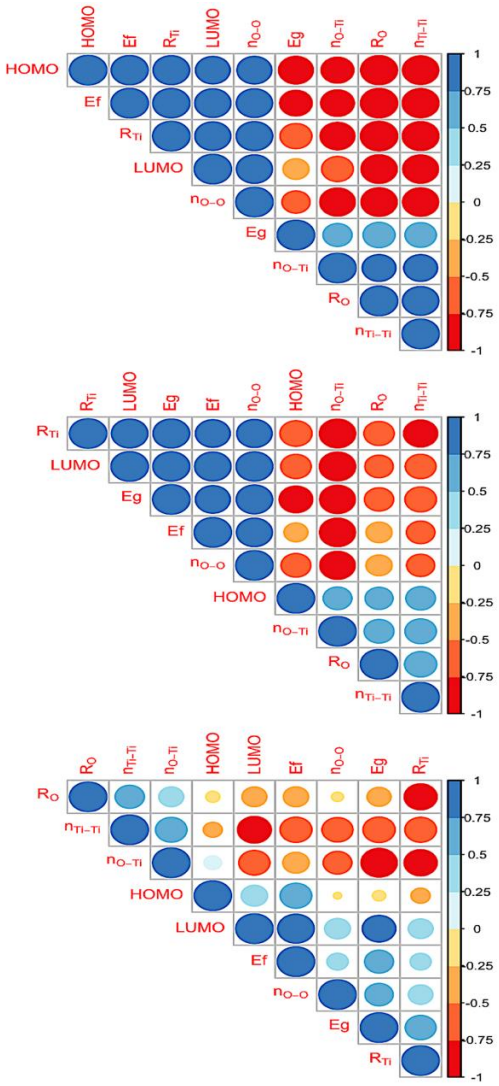

Figure 1. Correlation: The correlation among HOMO, LUMO, E_g, E_f, R and n.

2.1. *R_O vs HOMO, LUMO, E_g and E_f*

- There is a negative correlation between R_O and HOMO in the anatase phase close to perfect (-0.97). In the Brookite phase, this relationship turns into a moderately positive relationship. (0.56). In the rutile phase, these two are linearly independent.
- There is a negative correlation between R_O and LUMO in each phase. However, the strongest correlation was observed in the anatase phase whereas the weakest correlation in the rutile.
- There is a moderate negative correlation between R_O and E_g in the brookite and rutile phases. There is a strong linear correlation between these two variables in the anatase phase.

- There is a negative linear correlation between R_O and E_f in each phase, and the correlation in the anatase phase draws attention (-0.99).

2.2. *R_{Ti} vs. HOMO, LUMO, E_g and E_f*

- Linear dependencies of R_{Ti} and HOMO in the anatase phase are very strong (0.96). In other phases, there is a negative correlation between these two variables. In the Rutile phase, this correlation is very weak.
- Except for the rutile phase, there is a nearly perfect positive correlation between R_{Ti} and LUMO. There is moderately positive correlation in the rutile phase.
- In brookite phase, the positive linear correlation between R_{Ti} and E_g is noteworthy (0.96). Moderate linear correlation was observed in the anatase phase and positive in the rutile phase.
- There is a positive correlation between R_{Ti} and E_f in each phase, and the correlations in the anatase and brookite phases are close to perfect (0.98, 0.96, respectively).

2.3. *n_{O-O} vs. HOMO, LUMO, E_g and E_f*

- Between n_{O-O} and HOMO, a positive correlation was observed in the anatase phase, negative in the brookite phase, at a moderate level. These two variables are linearly independent in the rutile phase.
- There is a linear correlation between n_{O-O} and Lumo close to excellent in anatase and brookite phases. This relationship is also positive in the rutile phase and it is at a medium level.
- The positive linear dependence of n_{O-O} and E_g in brookite phase is remarkable (0.95).
- The positive linear relationship between n_{O-O} and E_f was excellent in anatase and brookite phases.

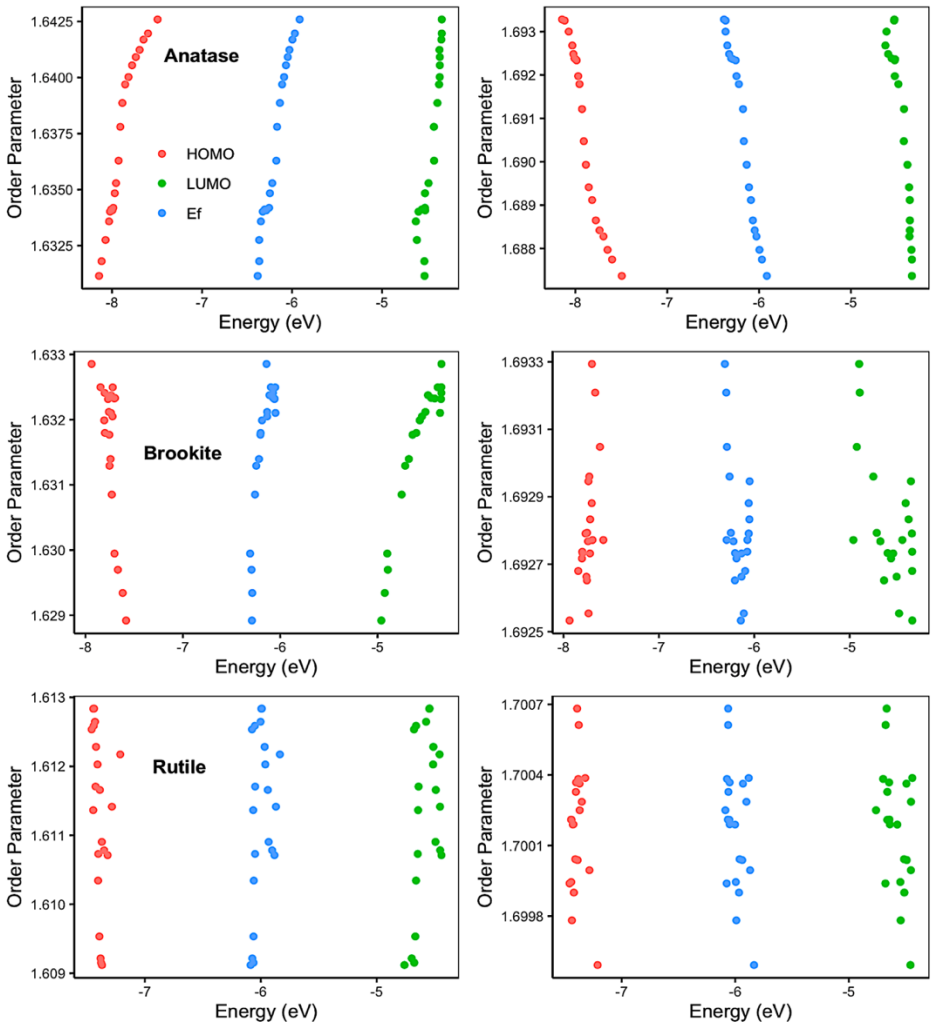

Figure 2. The order parameter R_O and R_{Ti} against HOMO, LUMO and Fermi energy (E_f) levels in the anatase, brookite and rutile phases. From top to bottom, the first two plots in the first place belong to the anatase phase, the brookite below that, and the rutile below the brookite. For each phase, the plot on the left belongs to Ti and the plot on the right belongs to O.

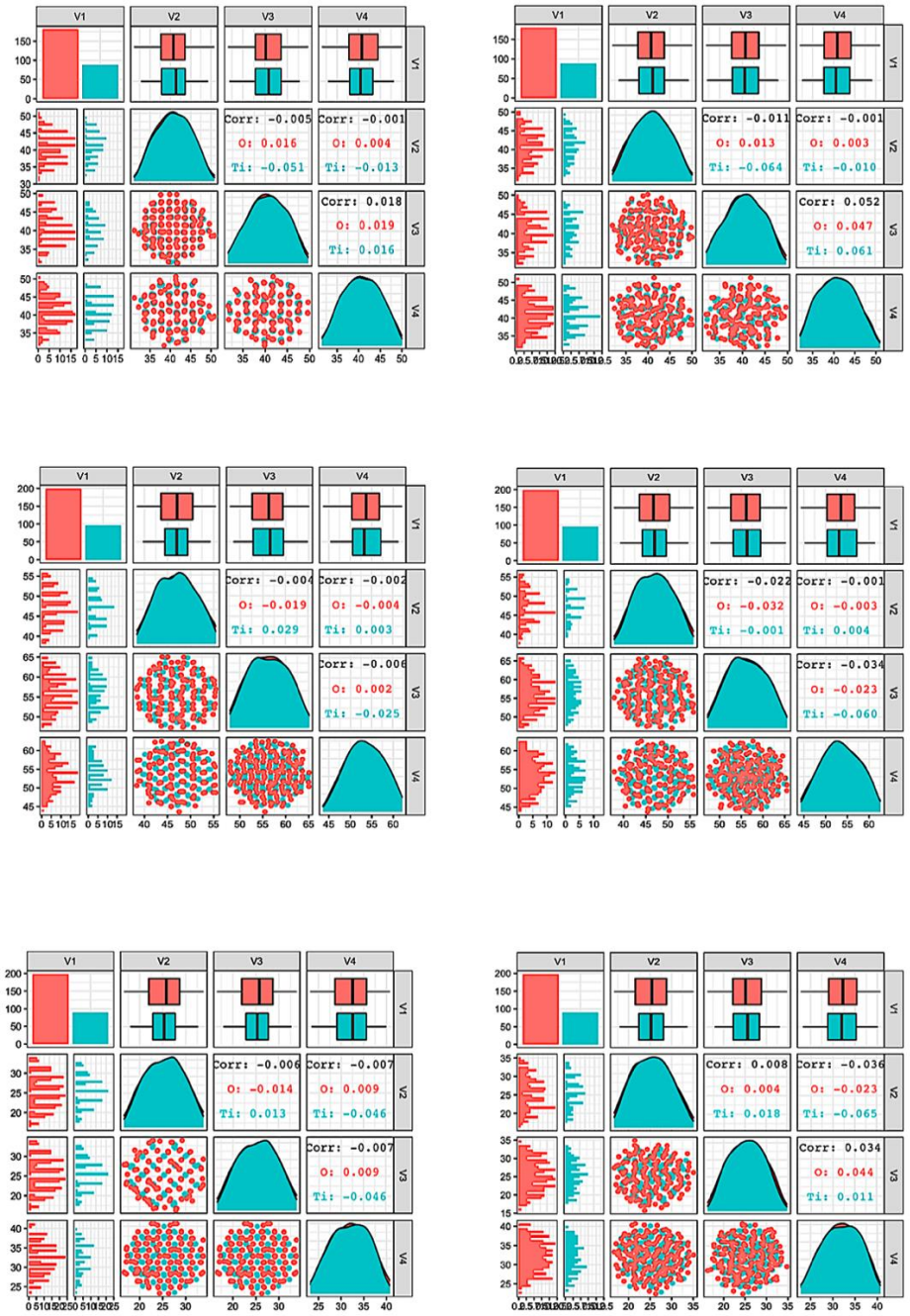

Figure 3. shows a summary of statistical properties of the data used in this study. Although we use atom geometries produced between 0K and 1000K temperatures (50 consecutive increments from 0K to 1000K), we only visualize atom geometries produced in 0K and 1000K for simplicity. The sets include 295 atoms at each temperature level and each atom is described with its 3D geometric location. Fig. 3 shows that there is no linear relationship among the variables, and each continuous variable is normally distributed. Moreover, there are more O atoms in the data sets.

2.4. n_{Ti-Ti} *vs. HOMO, LUMO,* E_g *and* E_f

- In the anatase phase, the correlation between n_{Ti-Ti} and HOMO is important (-0.93). In other phases, no strong correlation was observed between these two variables.
- A strong positive correlation was observed between n_{Ti-Ti} and LUMO in the anatase and brookite phases.
- No strong correlation was observed in any phase between n_{Ti-Ti} and E_g.
- In the anatase phase, positive strong correlation between n_{Ti-Ti} and Ef is remarkable.

2.5. n_{O-Ti} *vs. HOMO, LUMO,* E_g *and* E_f

In the brookite phase, n_{O-Ti} has an almost perfect negative correlation with LUMO, E_g and E_f. Additionally, there is a strong correlation between n_{O-Ti} and E_f in the anatase phase.

3. Conclusions

Data Science and mining are used in many fields to reveal important information in the data, such as astronomy [6,7], geology [8], public bus transportation [9], etc. In this work, we analyzed structural parameters and electronic properties of anatase, brookite and rutile phase TiO_2 nanoparticles (NPs) using data science techniques. We examined the correlation between structural and electronic properties of anatase, brookite and rutile phases of TiO_2 nanoparticles. The correlations among order parameter (R), nearest number contacts (n), HOMO, LUMO, Energy gap (E_g), Fermi energy (E_f) are revealed. There is a moderate negative correlation betweenR_O and E_g in the brookite and rutile phases, but a strong linear correlation between these two variables in the anatase phase. In the brookite phase, the positive linear correlations between [R_{Ti} and E_g] and [n_{O-O} and E_g] are noteworthy. In the same phase, n_{O-Ti} has an almost perfect negative correlation with E_g. No strong correlation was observed in any phase between n_{Ti-Ti} and E_g.

Acknowledgment

I would like to thank Associate Professor Mustafa Kurban for his help in data generation and valuable feedback.

References

[1] Schleder, G.R., Padilha, A.C., Acosta, C.M., Costa, M. and Fazzio, A., 2019. From DFT to machine learning: recent approaches to materials science–a review. *Journal of Physics: Materials*, *2*(3), p.032001.

[2] Kurban, H., Dalkilic, M., Temiz, S. and Kurban, M., 2020. Tailoring the structural properties and electronic structure of anatase, brookite and rutile phase TiO2 nanoparticles: DFTB calculations. *Computational Materials Science*, *183*, p.109843.

[3] Aradi, B., Hourahine, B. and Frauenheim, T., 2007. DFTB+, a sparse matrix-based implementation of the DFTB method. *The Journal of Physical Chemistry A*, *111*(26), pp.5678-5684.

[4] Luschtinetz, R., Frenzel, J., Milek, T. and Seifert, G., 2009. Adsorption of phosphonic acid at the TiO2 anatase (101) and rutile (110) surfaces. *The Journal of Physical Chemistry C*, *113*(14), pp.5730-5740.

[5] Gemming, S., Enyashin, A.N., Frenzel, J. and Seifert, G., 2010. Adsorption of nucleotides on the rutile (110) surface. *International Journal of Materials Research*, *101*(6), pp.758-764.

[6] Kurban, H., Jenne, M. and Dalkilic, M.M., 2017. Using data to build a better EM: EM* for big data. *International Journal of Data Science and Analytics*, *4*(2), pp.83-97.

[7] Jenne, M., Boberg, O., Kurban, H. and Dalkilic, M., 2014. Studying the milky way galaxy using paraheap-k. *Computer*, *47*(9), pp.26-33.

[8] Jenne, M., Zimmerman, A., Kurban, H., Johnson, C. and Dalkilic, M.M., employing software engineering principles to enhance management of climatological datasets for coral reef analysis.

[9] Zimmer, K., Kurban, H., Jenne, M., Keating, L., Maull, P. and Dalkilic, M., 2018, October. Using data analytics to optimize public transportation on a college campus. In *2018 IEEE 5th international conference on data science and advanced analytics (DSAA)* (pp. 460-469). IEEE.