JOURNAL of SCIENCE

SAKARYA UNIVERSITY

Sakarya University
Journal of Science

Guest Editor

English Language Editor

RESEARCH ARTICLES

Sakarya University Journal of Science

Title: Seismic Response of Anchorage Elements Used in Curtain Wall Systems

Authors: Ferhat PAKDAMAR, Özgün BOZKURT

Seismic Response of Anchorage Elements Used in Curtain Wall Systems

Ferhat PAKDAMAR[*1], Özgün BOZKURT[2]

Abstract

Utilizing curtain wall systems as outer covering of a building has been enormously increasing for the last years. Past earthquakes have revealed the wrong design of curtain wall systems connections that can seriously endangers human life. In this study, design procedures proposed by different international and national earthquake codes and specifications to compute the forces in the connection of curtain wall systems are summarized. As a real example, a regular 30-storey reinforced concrete building is considered, and a Finite Element Model is created for this structure. A detailed model for the building and the curtain wall systems and their connection anchors are considered. The axial and shear forces in the anchors are computed using the average results of seven sets of time history analyses. The results are compared with the results of building model using spectrum modal combination analyses and the forces in the connection anchors are proposed by different Specifications; Turkish Earthquake Code 2007, Turkish Earthquake Code 2018, Federal Emergency Management Agency, European Standard. As a result, it can be concluded that the seismic forces used for the design of anchors of the curtain wall systems are better estimated by the specification suggested by Turkish Earthquake Code 2007 compared to the detailed model results.

Keywords: curtain wall systems, seismic performance, facade, non-structural elements, anchorage

1. INTRODUCTION

In the world, developments and increase the variety of materials used in conjunction with the technology affects the use of the curtain wall system positively. Popularity of curtain wall systems especially on high rise buildings are increasing day by day because of being light weight, aesthetic and easy installation. Designing a system widely used correctly is quite important. There are two main principles to be considered in the design of curtain wall systems. First point is the seismic behavior of curtain wall systems under a strong ground motion. Second point is about

---
[*] Corresponding Author: pakdamar@gtu.edu.tr
[1] Gebze Technical University, Department of Architecture, Kocaeli, TURKEY.
   ORCID: https://orcid.org/0000-0002-5594-3095
[2] Gebze Technical University, Gebze, Kocaeli, TURKEY.
   ORCID: https://orcid.org/0000-0001-7145-7057

comfort like air or water impermeability [1], wind resistance [2], solar control [3], eq. resistance [4] etc. Seismic behavior of curtain wall systems are vital when ground is shaking. Any facade parts, damaged can endanger for human life. Past earthquakes confirm this danger [5]. Additionally, wrong design or assembling of a curtain wall system parts wrongly affects welfare and comfort of building residents in a negative way [6]. Designing of curtain walls properly are very important for the reasons, explained above. Seismic loads, effecting to curtain wall system design play an important role, especially in high-rise buildings [7, 8]. However, calculation and assessments of seismic loads, acting on the facade systems has not been the subject of much research around the world there are limited study. Galli [9] compared the seismic forces acting on facade systems for various codes in his doctoral dissertation and he described the differences between them. In addition, results of some tests, acted to an actual curtain wall system like wind strength test, the water-air tightness test, etc. and computer model of this system is compared. Ting [10] mentioned two approaches about response of curtain wall systems under lateral story drifts when an earthquake happening. There are two more important studies [11, 12] about behavior of curtain walls under horizontal actions experimentally. O'Brien, et al. [13] has been formulated a closed-form equation to predict a glass panel cracking failure drift. These researches rather describes steel profiles, material tests and the history of facade systems than the effect of seismic loads. In the present study, effect of seismic loads to curtain walls, suggested in codes are compared.

2. CURTAIN WALL SYSTEMS

Curtain wall system was applied to a two story bank building in Philadelphia in 1820, for the first time in the world. There are also two representative buildings, Chicago Auditorium and Chicago Monadnock were built in 1890 and 1891 respectively. Curtain wall concept continues to improve and expand their areas with the advancement in technology. Especially, aluminum metal has substantially increased the popularity of the area with the use of facade systems [14].

Curtain wall system consists of several components. These components comprise of steel profile supporting structures, anchorage elements of steel supporting structures and covering material (glass, composite, precast, ceramic etc.). These components have to successfully pass air leakage control, vapor diffusion control, rain penetration control and condensation control tests for increasing the interior comfort of buildings. Quirouette [15] and Brenden [16] cite rules of the tests and properties of curtain wall system connections.

The most important component of a curtain wall system for the load bearing is carrier profiles. Because these steel profiles have to behave as ductile and deform harmonically with story drifts of the building on assembled under an earthquake motion. Especially with usage of aluminum in sector, curtain wall systems have become more flexible behavior and adapted the story movements. As a result they can resist to seismic forces. Building covering materials like glass, precast, ceramic, etc. are more brittle elements than load bearing steel support profiles. Expectation from these elements are air leakage control, vapor diffusion control, rain penetration control and condensation control and transfer the stress acting on them to steel profiles with seismic loads. It should be avoided from very rigid connections of the carrier steel profile to mount facade systems in the building.

3. USED REGULATIONS FOR THE CURTAIN WALL SYSTEMS

Various regulations including the design rules of the curtain wall system is available in the world. The proposed design rules in the regulations are determined in accordance with a main target. Displacements and behavior of a curtain wall system applied to a structure must not pass the life safety performance target level, while an earthquake. It has also response to internal forces occurred by seismic effects without collapsing.

Some of the regulations related to a curtain wall system with design rules are Turkish Seismic Code 2007 [17], Turkish Seismic Code 2018 [18], European Codes 8 [19], American Regulations

FEMA450 [20], Standard of New Zealand NZS1170.5 [21], Japan Regulations JASS14 [22]

3.1. Turkish Seismic Code 2007 (TDY2007) [17]

There is not a clear information about the seismic force which is effect on curtain wall system in TDY [17]. However, a seismic force description for all non-structural architectural elements and for their connection parts which connect the system to the building structure is made in the *chapter 2.11*. Because the curtain wall systems are non-structural, they are considered under the definition in the TDY2007.

In accordance with this definition, earthquake force equation for design requirements in the TDY2007 is given in Eq. (1).

$$f_e = 0.5 A_0 I w_e \left(1 + 2\frac{H_i}{H_N}\right) \tag{1}$$

This calculated earthquake load is applied to the center of gravity of the curtain wall system element in the horizantal directon to get the most unfavorable internal forces. If the element is inclined half of the force is applied vertically.

- A_0 = effective ground acceleration coefficient
- I = building importance coefficient
- w_e = weight of the component
- H_i = the height of the building on the basis of relevant element
- H_N = the total height of building which is equal to H EUROCODE-8 and h in FEMA-450.
- f_e = force which is equal to F_e in TDY2018, F_a in EUROCODE-8, F_p in FEMA-450 and F_{ph} in NZS-1170.5.
- w_e = weight which is equal to m_e in TDY2018, W_a in EUROCODE-8, W_p in FEMA-450 and W_p in NZS-1170.5.

3.2. Turkish Seismic Code 2018 (TDY2018) [18]

Earthquake force, F_{ie} which is applied to anchorage parts of curtain wall systems in Turkish Seismic Code [18] is accepted as to non structural components like Turkish Seismic Code [17]. In this context, the earthquake force which is described in *Chapter 6.2* in the code is given in Eq. (2).

$$F_{ie} = (m_e A_{ie} B_e)/R_e \tag{2}$$

m_e is the mass of element, A_{ie} is the peak ground acceleration effecting the element, B_e is amplification factor, R_e is the component behaviour factor in Eq. (2) Tables are available in the code.

3.3. European Codes EUROCODE–8 (EC8) [19]

EUROCODE is the regulation that Europian countries use. Earthquake force which is applied to curtain wall systems in EUROCODE-8 is diveded to two parts. First is anchorage forces of curtain wall system and the other is the force applied to facade system. Mentioned seismic force in Eurocode 8 [19], *Chapter 4.3.5* is given in Eq. (3).

$$F_a = (S_a \cdot W_a \cdot \gamma_a)/q_a \tag{3}$$

W_a is the weight of releavent component at i'th story for the seismic force to the i'th story. γ_a is importance factor of releavent component at i'th story. q_a is the behaviour factor of the element. There is a table for the q_a in the code. S_a is the seismic coefficient applicable to non-structural elements and the formula of it is given in Eq. (4).

$$S_a = \alpha \cdot S \cdot \left[3 \cdot \frac{1+z/H}{1+(1-Ta/T1)^2} - 0.5\right] \tag{4}$$

α = is the ratio of the design ground acceleration on type A ground, a_g, to the acceleration of gravity g

- S = is the soil factor
- Ta = is the fundamental vibration period of the non-structural element;
- $T1$ = is the fundamental vibration period of the building in the relevant direction
- z, H = is the height of the non-structural element above the level of application of the seismic action and building height

S_a can not be less than $(a \cdot S)$. For the following non-structural elements the important factor γ_a shall not be less than 1.5. In all other cases the importance factor γ_a of non-structural elements

may be assumed to be $\gamma_a = 1.0$. Upper limit values of the behaviour factor qa for non-structural elements are given in a table in the code.

3.4. FEMA - 450 Regulation [20]

The seismic design force, F_p, applied in the horizontal direction shall be centered at the component's center of gravity and distributed relative to the component's mass distribution and shall be determined in accordance with Eq. (5) in FEMA-450 *Chapter 6.2.6* [20] as follows:

$$F_p = \frac{0.4 a_p S_{DS} W_p}{R_p/I_p}\left(1 + 2\frac{z}{h}\right) \qquad (5)$$

- a_p ve R_p = The component amplification and response modification factors selected, as appropriate, from Tables in the code
- S_{DS} = The short period spectral acceleration parameter (It is correspond to value of "$Ao * I * 2.5$" in TDY2007)
- W_p = Operating weight of a nonstructural component
- I_p = The component importance factor
- z = The height above the base of the point of attachment of the component, but z shall not be taken less than 0 and the value of z/h need not exceed 1.0
- h = The average roof height of structure above the base

3.5. New Zealand Standard NZS 1170.5 [21]

The Code, used in New Zealand, includes design rules for curtain wall systems. Curtain wall systems are called as "part" in this code. Unlike other codes, description of the seismic force, acting on curtain wall systems is proposed even for horizantal and vertical direction in this code. These seismic forces, described in *Chapter 8.5* in the code is limited with a value which is related to the curtain wall weight. There are seismic force formulas for each direction in the code [21]. These horizantal and vertical seismic forces is described in Eqs. (6)-(8)

$$F_{ph} = C_p(T_p)C_{ph}R_pC_f(\xi)_s W_p \leq 3.6 W_p \qquad (6)$$

$$C_p(T_p) = C(0)C_{Hi}C_i(T_p) \qquad (7)$$

$$F_{pv} = C_{vd}C_{pv}R_pW_p \leq 2.5 W_p \qquad (8)$$

- F_{ph} = Horizontal design action
- $C_p(T_p)$ = horizontal design Response Coefficient of curtain wall part
- C_{ph} = Part response factor
- R_p = Part risk factor
- $C_f(\xi)_s$ = Damping factor for support structure for (it is 1.0 for %5 damping)
- W_p = Weight of curtain wall system
- $C(0)$ = Site Hazard Coefficient for T = 0 secs
- C_{Hi} = Floor height coefficient
- $C_i(T_p)$ = Floor Spectral Shape Factor
- C_{vd} = *Vertical design factor of* curtain wall part
- C_{pv} = *Vertical design factor of* curtain wall part

There is a table for C_{ph} and C_{pv} values depended on ductility

3.6. Japanese Standard JASS14 [22]

Seismic energy is described in two ways in this Japanese standard. By the way seismic forces of a curtain wall anchorage is described also into two parts [22]. According to this standard, P and S waves are comprised of seismic energy. *P* waves are faster then *S* waves and their influence in long direction. *S* waves are slower then *P* waves and their effect in transverse. This discrimination of seismic energy is described as horizontal seismic force and vertical seismic force, similarly to NZS standard. However, there isn't given any formulas. Verbally expressed formulas are given below:

- Seismic force composed of *P* waves on the curtain wall system part is described as the multiplication of dead load of the part by vertical acceleration.
- Seismic force composed of *S* waves on the curtain wall system part is described as the multiplication of dead load of the part by horizontal acceleration.

4. MODELLING OF CURTAIN WALL SYSTEMS ON A SAMPLE BUILDING

In this section, seismic forces, effecting to anchorage elements of curtain wall system on a high rise building are acquired using various analysis method in SAP2000 software. Suggested seismic forces in various seismic codes are also calculated. Then, all the results obtained are compared on graphs.

4.1. Building, Earthquake Information and Structural System

A 30 storey regular reinforced concrete building has been choosed. Dimensions of the building are 24 x 18 m. It has one ground floor and 29 typical floors. Ground floor has 4.5 m and each storey has 3.5 m storey height. Total height of the building is 106 m. Concrete grade has been taken C40 (40 MPa) and reinforcing steel has been taken S420 (420 MPa).

A moderate service earthquake has been considered in calculations. Spectral acceleration for short periot, S_s has been taken 1.0 g and for long period S_l is taken 0.5 g and soil type has been taken B. According to these values elastic response spectrum is given as a graph in Fig. 1.
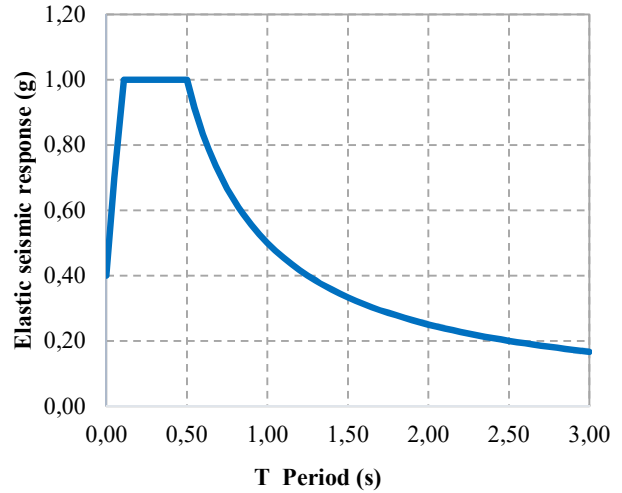
Figure 1. Used elastic response spectrum
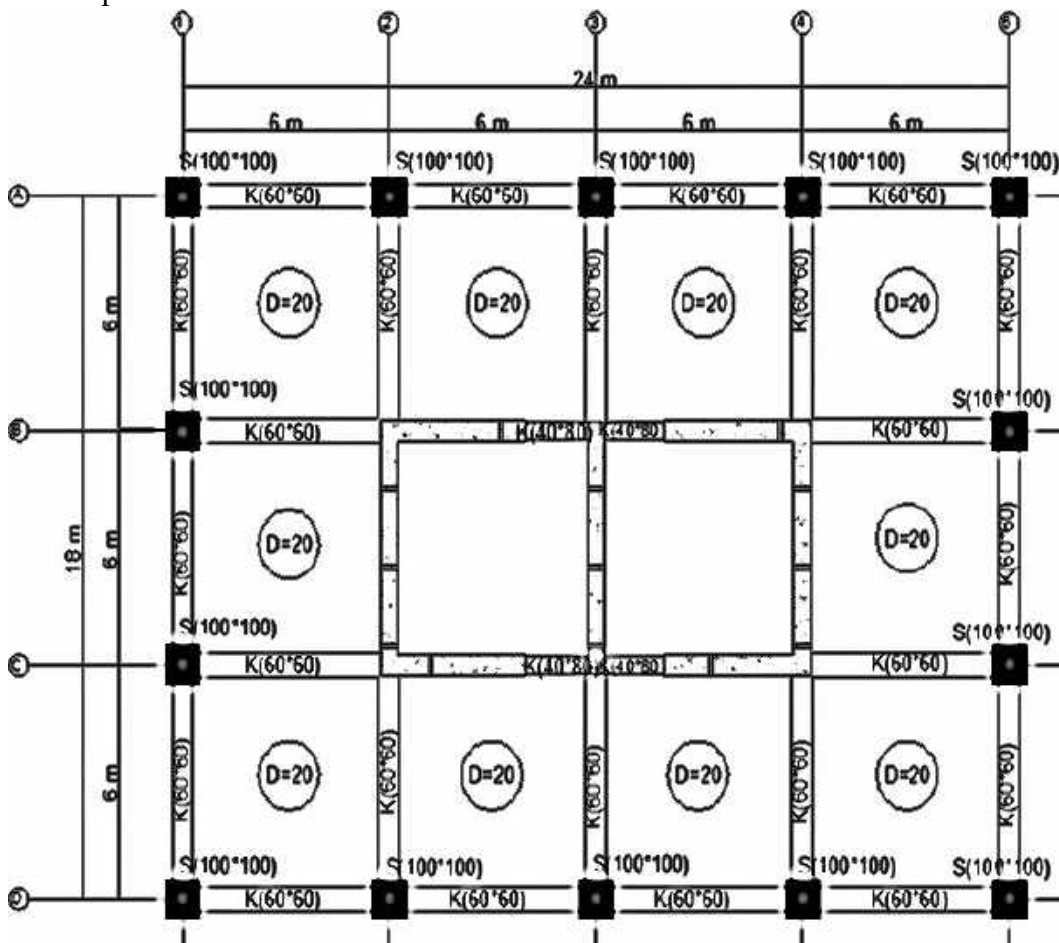
Figure 2. General floor plan (sizes are for ground)

Structural wall-frame system that has core, shear walls and columns is considered as main system. Building has 6 m axe spaces in two direction symmetrically in plan. Deck has 20 cm thickness in each floor. Core-wall at the center of building has 50 cm thickness in each floor. Beam dimensions are considered 40x80 cm at core and 60x60 cm for other beams in each floor. Column sizes vary for some storeys. Column dimensions are given in Table 1 General floor plan of sample building is given in Fig. 2.

Table 1 Column dimensions

Floors	Wide (cm)	Height (cm)
Ground - Floor 9	100	100
Floor 10 - Floor 19	90	90
Floor 20 - Floor 29	80	80

Vertical Load Acceptanses

Total dead loads, $\quad g = 7 \ kN/m^2$

Deck part $\rightarrow 0.2 \times 25 = 5 \ kN/m^2$

Plaster + cover $= 2 \ kN/m^2$

Moving loads, $q = 3 \ kN/m^2$

Curtain wall system loads $\cong 0.381 \ kN/m^2$

4.2. Information About the Applied Curtain Wall System

Classical stick curtain wall system was choosed for analysis as alluminium material. Vertical allumunim profiles called vertical mullions were anchored by 1.5 m away between two axes in plan. Horizontal allumunim profiles called horizontal mullions were anchored by 0.75 m away from top and bottom by the way middle parts were designed as 2 m height. Profiles that were used in the system have 120 x 50 mm sizes. Spandrel panels were choosed as glass. Typical anchorage detail of the system was given at Fig. 3.

Curtain wall system loads

Vertical mullions $= 22 \ N/m$

Horizontal mullions $= 1.97 \ N/m$

Stable glass parts $= 30 \ N/m^2$
Sashes $= 40 \ N/m^2$
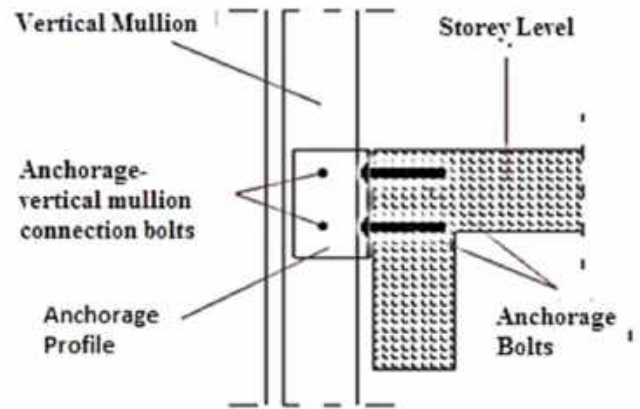
Figure 3. Typical anchorage detail of the curtain wall system

Memari, et al. [23] showed that finite-element modeling is a viable approach for analytical evaluation of curtain walls. The building was modeled using SAP2000 software. Beam and column elements were modeled as "frame", decks were modeled as "shell-thin" and wall elements were also modeled as "shell-thin". Two models were created for the analysis. In the first model, curtain wall system were applied as concantreated loads to each anchorage location. In the second model, curtain walls are modeled as "shell-thin" element and they were fixed to the structure with anchorage profiles which are modeled as "frame" elements. All material and gometric properties of these elements were described to the software. All columns are assumed as fixed support at bottom ends on basement. Results of second model are given below. In the future an enhancement of the FE model could be done as suggested by Amadio and Bedon [24], who also take into account the effects of possible initial geometrical imperfections.

4.3. Time History Loading and Code Loadings to the Structural System

This model called as Comprehensive model (CM), is the most comprehensive model created in this study because it contains curtain wall system and anchorage elements. Thus, it is a model in which the most realistic results can be obtained. For this reason, the results obtained from the other analysis

methods and different specifications, are compared with the results obtained from this full model. As anchorage elements are modelled in this model, seismic force which must be get affected on anchorage elements during design process is the maximum force created on these anchorage elements modelled for earthquake motion get affected on model. Seven earthquake motions are applied to these structure model; when taking results, it is averaged the maximum seismic forces formed in relevant anchorage element due to the effect of these seven ground motion set. Thus, seismic force is found which needs to get affected on the relevant anchorage element during design process. As ground motions are applied in both directions (x and y), seismic forces of selected anchorage element are obtained as two compounds. Faxial is obtained for the average of ground motions applied in direction of x, Fshear for the average of ground motions applied in direction of y. In full model, aluminum is selected for material of anchorage elements and glass for material of curtain walling system. Seismic forces which will come to the curtain wall system element in full structure model are given in Fig. 4.

Comparison between the force results for TDY2007, TDY2018 time history method, TDY2018 spectral analysis method, EUROCODE-8, FEMA450 loading conditions and comprehensive model (CM) are given in the Figs. 5-9 below.
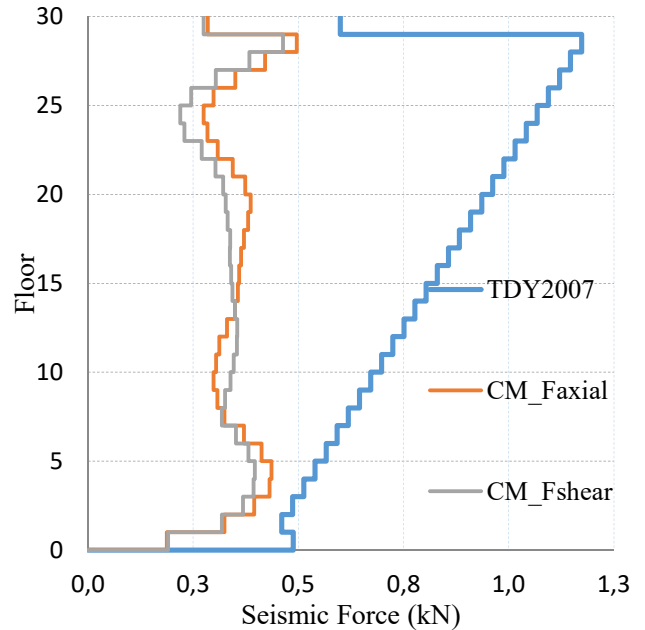
Figure 5. Comparison between the forces of TDY2007 and CM analysis

Figure 4. Seismic forces for curtain wall elements at the floor levels according to analysis methods for time history analysis.
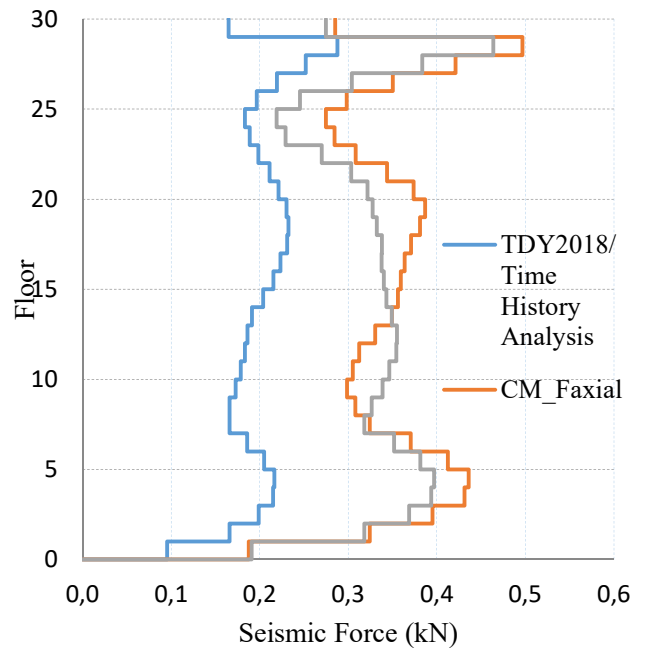
Figure 6. Comparison between the forces of TDY2018 time history analysis and CM analysis
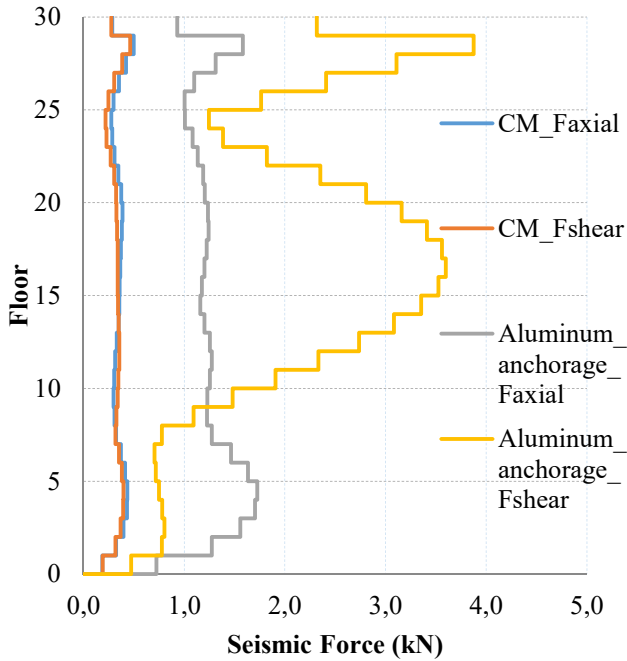
Figure 7. Comparison between the forces of TDY2018 response spectrum and CM analysis
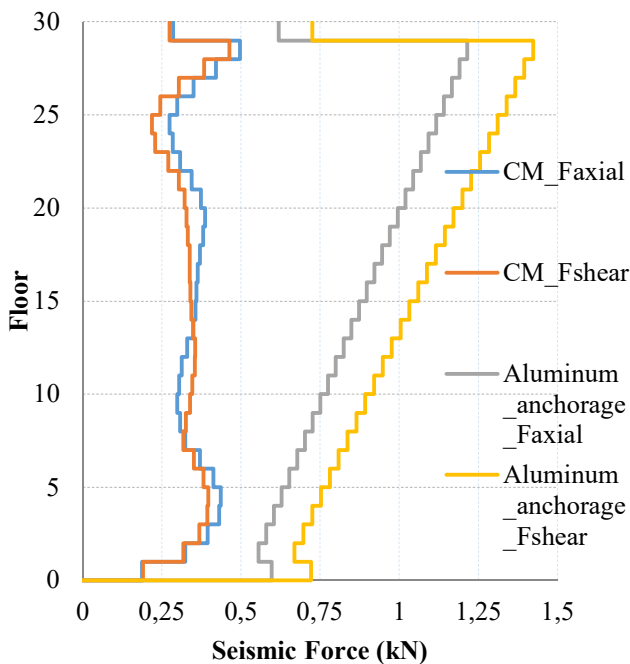
Figure 8. Comparison between the forces of EUROCODE-8 and CM analysis
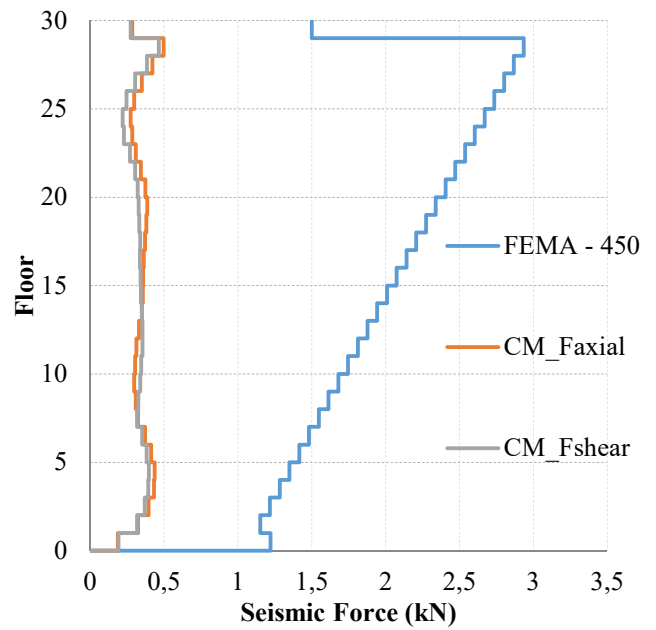
Figure 9. Comparison between the forces of FEMA-450 and CM analysis

When specifications are considered and the results are obtained, it is seen that a large part of relevant specification remains in an over secured area in relation to seismic force which needs to be get affected on curtain walling system anchorage element during design process. This mentioned security zone is defined by considering the results obtained from Comprehensive Model (CM) giving the most realistic results. It is defined as secure the specifications and methods giving bigger force values than these most realistic results. It is seen that only the results of analysis method in the field of time history situated in TDY2018, give lower values than results obtained from Comprehensive Model (CM), and curtain walling system anchorage element is not a secure method related to seismic force to be applied in design process. Besides, there are differences when it is compared the two specifications considering natural vibration period of anchorage element. The differences between especially floor 10 and 25 are big, and generally the results of TDY2018 are bigger according to Eurocode-8. When considering these two specification in themselves, pulling and cutting forces are close to each other according to Eurocode-8 Specification for two different material type. In TDY2018 Specification, this situation is valid for only pulling forces. Cutting forces results for the results

of FEMA and steel anchorage profile of TDY2018 has closeness attractively between floor 15 and 20. Besides, when considering the results obtained from Comprehensive Model as the most realistic results, it can be said that FEMA Specification is over and the most secure specification in terms of seismic force to be affected on curtain walling system anchorage element.
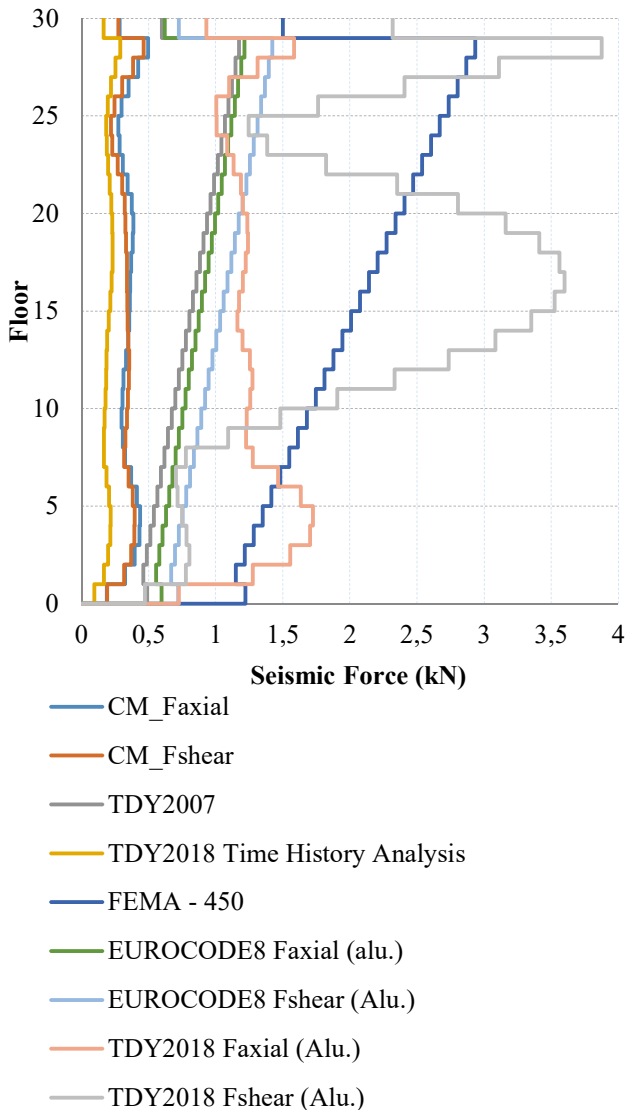
Figure 10. Comparison of the results obtained from all the regulations and CM

By remaining in security zone, TDY2007 Specification became the closest specification among used specifications to the results obtained by using Comprehensive Model (CM) accepted as the most realistic results. Because TDY2007, FEMA and Eurocode-8 Specifications do not pay attention to momentum to be occurred in floors

due to the ground motions in seismic force formulas, the obtained seismic forces increase linearly toward upper floors. Because front mass decrease in half on the top floor, seismic force also decreases in half in comparison to the previous floor. It does not increase linearly the results obtained from analysis method in time history and earthquake spectrum partaking in TDY2018 and the Comprehensive Model. Because these methods pay attention to momentums occurred by ground motion, it is obtained the results of seismic force complying with these momentums. Besides, when looking at the results for relative floor translation as well as seismic force, it is seen that curtain walling system anchorage elements does not have values to cause any problem, remain low of values recommended in TDY2018 and take place in security zone in this matter. The results showing these evaluations clearly are given in Fig. 10.

5. CONCLUSIONS

The obtained results show that TDY2018 is the closest specification to the results of Comprehensive Model (CM) having the most realistic results. However, as these results obtained according to analysis method in field of time history partaking in TDY2018 have smaller values than the most realistic results, they take place in security zone. TDY2007 is the closest specification to the most realistic results by remaining in security zone. FEMA450 is the specification having the most distant results and that we can called as over secured by remaining in security zone. TDY2007 Specification come close to the closest values to the Comprehensive Model (CM) between floors 1 and 5 in the rate of 80%. FEMA450 Specification could come close at the most between floors 1 and 5 in the rate of 30%. EUROCODE8 Specification could come close to the most realistic values in all floors in the rate of 50%. TDY2018 Specification come close between floors 1 and 8 in the rate of 50%. When looking at these results, in relation to the seismic force to be paid attention during design process for curtain walling system anchorage elements, it is concluded that TDY2007 Specification gives the best and closest proposition by remaining in security zone.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] S. G. Naggatz and S. F. Sinusas, "Water leakage testing of glass and metal curtain walls", (in English), Building Walls Subject to Water Intrusion and Accumulation: Lessons from the Past and Recommendations for the Future, vol. 1549, pp. 143-165, 2014.

[2] P. A. Hitchcock, K. C. S. Kwok, K. S. Wong, and K. M. Shum, "The effects of topography on local wind-induced pressures of a medium-rise building", (in English), Wind and Structures, vol. 13, no. 5, pp. 433-449, Sep 2010.

[3] M. F. Hossain, "Design and construction of ultra-relativistic collision PV panel and its application into building sector to mitigate total energy demand", (in English), Journal of Building Engineering, vol. 9, pp. 147-154, Jan 2017.

[4] H. Maneetes and A. M. Memari, "Development of analytical modeling for an energy-dissipating cladding panel", (in English), Structural Engineering and Mechanics, vol. 32, no. 5, pp. 587-608, Jul 30, 2009.

[5] R. P. Dhakal et al., "Seismic performance of non-structural components and contents in buildings: an overview of NZ research", (in English), Earthquake Engineering and Engineering Vibration, vol. 15, no. 1, pp. 1-17, Mar 2016.

[6] C. C. Baniotopoulos, T. N. Nikolaidis, and G. Moutsanidis, "Optimal structural design of glass curtain-wall systems," (in English), Proceedings of the Institution of Civil Engineers-Structures and Buildings, vol. 169, no. 6, pp. 450-457, Jun 2016.

[7] A. Baird, A. Palermo, and S. Pampanin, "Facade damage assessment of concrete buildings in the 2011 Christchurch earthquake", (in English), Structural Concrete, vol. 13, no. 1, pp. 3-13, Mar 2012.

[8] W. S. Lu, B. F. Huang, S. M. Chen, and K. M. Mosalam, "Acceleration demand of the outer-skin curtain wall system of the Shanghai Tower", (in English), Structural Design of Tall and Special Buildings, Article vol. 26, no. 5, p. 14, Apr 2017, Art. no. e1341.

[9] U. Galli, "Seismic behaviour of curtain wall facades: a comparison between experimental mock up test and finite element method analysis", Ph.D., VI Facoltà – Ingegneria Edile-Architettura, Politecnico Di Milano, Italy, 2011.

[10] R. Ting, "Curtain wall design against story drift", in Proceedings of the 2004 Structures Congress, Nashville, Tennessee, 2004, pp. 1-7.

[11] B. Gorenc and D. Beg, "Curtain wall facade system under lateral actions with regard to limit states", (in English), Steel Construction-Design and Research, Article vol. 9, no. 1, pp. 37-45, Feb 2016.

[12] N. Caterino, M. Del Zoppo, G. Maddaloni, A. Bonati, G. Cavanna, and A. Occhiuzzi, "Seismic assessment and finite element modelling of glazed curtain walls", (in English), Structural Engineering and

Mechanics, vol. 61, no. 1, pp. 77-90, Jan 10, 2017.

[13] W. C. O'Brien, A. M. Memari, and M. Eeri, "Prediction of seismic cracking capacity of glazing systems", (in English), Earthquakes and Structures, vol. 8, no. 1, pp. 101-132, Jan 2015.

[14] M. Colomban, "History and technical development of curtain walls", (in English), Habitat and the High-Rise, vol. 903, pp. 381-402, 1995.

[15] R. Quirouette. (2017, 01/06/2017). Glass and Aluminum Curtain Wall Systems. Available: https://www.sistemamid.com/panel/uploads/biblioteca/2014-05-25_11-46-10102938.pdf

[16] K. Brenden. (2006, 01/06/2017). Dynamic Issues Drive Curtain Wall Design. Available: http://www.structuremag.org/wp-content/uploads/2014/09/C-SD-Curtain-Wall-Aug-061.pdf

[17] (2007). Deprem bölgelerinde yapılacak binalar hakkında yönetmelik (Regulation for buildings in seismic areas).

[18] (2018). Türkiye Bina Deprem Yönetmeliği (Turkey Earthquake Building Regulations).

[19] (2004). 8: Design of structures for earthquake resistance—Part 1: General rules, seismic actions and rules for buildings (EN 1998-1: 2004). Available: ftp://ftp.norsar.no/pub/outgoing/conrad/cuba/EC8.en.1998.1.2004.pdf

[20] (2003). FEMA450, NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures Part 1: Provisions, part 2: commentary. Available: http://www.nehrp.gov/pdf/fema450provisions.pdf

[21] (2004). Structural Design Actions, Part 5: Earthquake actions–New Zealand. Available: https://shop.standards.govt.nz/catalog/1170.5%3A2004%28NZS%29/view

[22] (1996). Japanese architectural standard specification curtain wall.

[23] A. M. Memari, A. Shirazi, P. A. Kremer, and R. A. Behr, "Development of finite-element modeling approach for lateral load analysis of dry-glazed curtain walls", Journal of architectural engineering, vol. 17, no. 1, pp. 24-33, 2011.

[24] C. Amadio and C. Bedon, "A buckling verification approach for monolithic and laminated glass elements under combined in plane compression and bending (vol 52, pg 220, 2013)", (in English), Engineering Structures, vol. 57, pp. 393-393, Dec 2013.

Title: A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different
Span Lengths and Investigation on Prestressing Tendons with Comparisons

Authors: Mustafa TURAN

How to cite
Mustafa TURAN; (2020), A Case Study on Two-Span Post-Tensioned Concrete Bridge
Decks with Different Span Lengths and Investigation on Prestressing Tendons with
Comparisons. Sakarya University Journal of Science, 24(4), 575-585, DOI:
https://doi.org/10.16984/saufenbilder.544525
Access link
http://www.saujs.sakarya.edu.tr/en/pub/issue/55932/544525

New submission to SAUJS
http://dergipark.org.tr/en/journal/1115/submission/step/manuscript/new

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investigation on Prestressing Tendons with Comparisons

Mustafa TURAN[*1]

Abstract

Post-Tensioned (PT) concrete system is one of the widely used bridge superstructure types because of its better load capacity performance. By increasing load carrying capacities on conventional reinforced concrete structures, PT is considered as an advanced technology in engineering. The aim of this paper is to study on finite element analysis of two-span PT concrete bridge decks with different span lengths. For three different span lengths, two-span bridge decks were investigated. To reduce the self-weight of the deck, at mid spans, deck section was used as voided section and at diaphragm regions, section was used as filled. Two-dimensional Finite Element (FE) deck models were created by frame elements. The FE software package Midas Civil was used. Permanent, transient and time-dependent loading types were considered. Boundary conditions were defined with their real mechanical properties. To balance the deck section stresses, prestressing tendons were used. By iterations, PT tendons layout and types were studied. Interaction between span lengths and PT tendons areas was investigated. Also, different effects on PT tendons were studied and relevant comparisons were submitted. The "best" tendon using was examined for the investigated deck sections. Concluded that optimization on PT tendon areas at bridge deck sections is possible by investigating the main effects on it. Results show that a good optimization enables us to have the optimal PT tendon using with lower cost.

Keywords: bridge, deck, post-tensioned, tendon-optimization

---
[*] Corresponding Author: mustafa.turan.ce@gmail.com
[1] Gazi University, Ankara, Turkey, ORCID: https://orcid.org/0000-0002-4601-8041

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

1. INTRODUCTION

Concrete and steel materials are the two main materials used in most structure types. Using two elements in a good harmony creates a strong element which has a real load carrying capacity. In some cases, aforementioned element capacity needs to be support by external effects. Prestressing is a good way and an advanced technology on reinforced concrete elements. Application of prestressing gives section a compression stress which can be handled by concrete. In other words, prestressing aims to use high compressive strength of concrete material and reduce the tension stress on concrete section. This technology is broadly used for especially long spans like bridges, residential buildings, hotels, architectural structures.

Prestressing has two main application methods which are pre-tensioning and post tensioning. As the names imply, in pre-tensioning, tendons are stressed before concrete pouring. In post tensioning system, tendons are stressed after concrete gets initial pouring. In this paper, post-tensioning (PT) technology on bridge deck was investigated.

The use PT system allows passing long spans with more reasonable sections. This brings material savings due to reduced superstructure thickness compared to conventional slab types. By reducing deck thickness, bridge aesthetics can be emphasized. Also, reducing elements geometric properties leads to cost effectiveness which is the "greatest" point mostly in designs. There are many design variables in PT system. Hence, by designer, optimum design conditions should be investigated thoroughly.

PT concrete idea was developed in 1928 by Eugene Freyssinet and one year later, in 1929, an application was done. Since then, the idea and applications developments are still in progress.

Many articles and researches have been published addressing the behavior of PT concrete structural members [1-4]. In most studies, to understand real behavior, nonlinear models have been created. By defining time dependent parameters, not only primary but also secondary effects have been investigated.

About tendon optimization issues, many researchers have studied on different scenarios and investigated different effects on tendon usage [5-10]. Optimization technique plays a significant role in structure design. There are huge numbers of parameters that affect the design. The aim is to reach the "best" one. In mentioned structure type, optimal design refers to reach the maximum performance in different loading cases, minimum weight with reasonable concrete sections, minimum steel areas usage and minimum cost or a good combination of these.

In this paper, tendon using and optimization was investigated from different angles because one of the most important parts in design PT is deciding tendon types, tendon layout and tendon mechanical properties. Two-span bridge decks which have three different span lengths were created by FE model. All three bridges have equal geometric and mechanical properties except from span lengths. To reduce self-weight of the deck, voided sections were used at mid-spans and filled sections were used at supports. Frame elements were used for two-dimensional FE deck model. At each end, two elastomeric supports were designed. Mechanical properties of elastomeric bearings were calculated and assigned. All types of loads which are permanent, transient and time-dependent were considered. For the deck stresses at top and bottom of the section, tendons were placed in an engineering manner. At mid-spans, tendons were placed near the bottom face. At supports; tendons were placed at top face to balance tension stresses. For the same deck typical sections, tendons were compared in bridges having different span lengths and comparison results were evaluated.

2. MODELING STRATEGIES

2.1. Deck Geometries

In the scope of the study, 3 types of deck analysis models were investigated. Whole typical section properties were equal in each type but span lengths were different.

Type 1: 2 x 32.00 m.

Type 2: 2 x 34.00 m.

Type 3: 2 x 36.00 m.

Deck section is cast in situ, post-tensioned concrete voided and filled section. Deck width is 10.00 m and its depth is 1.20 m. To reduce self-weight of the deck, 6 recesses are planned in the section each has 0.70 m diameter. Flange and web thicknesses are 25 cm and 30 cm, respectively.

At pier and abutment axis region, a filled section is planned to resist shear forces by huge concrete section area. At typical section, 2 x 1.50 m width side walking is planned and at middle, 7.00 m asphalt width serves for 2 lanes. Thickness of the asphalt layer is 6.00 cm.

2.2. Design Codes

The design calculations are based on "AASHTO LRFD Bridge Design Specifications 2007 SI".

2.3. Engineering Software

Midas Civil, 2019 (v2.1) was used as FE engineering software.
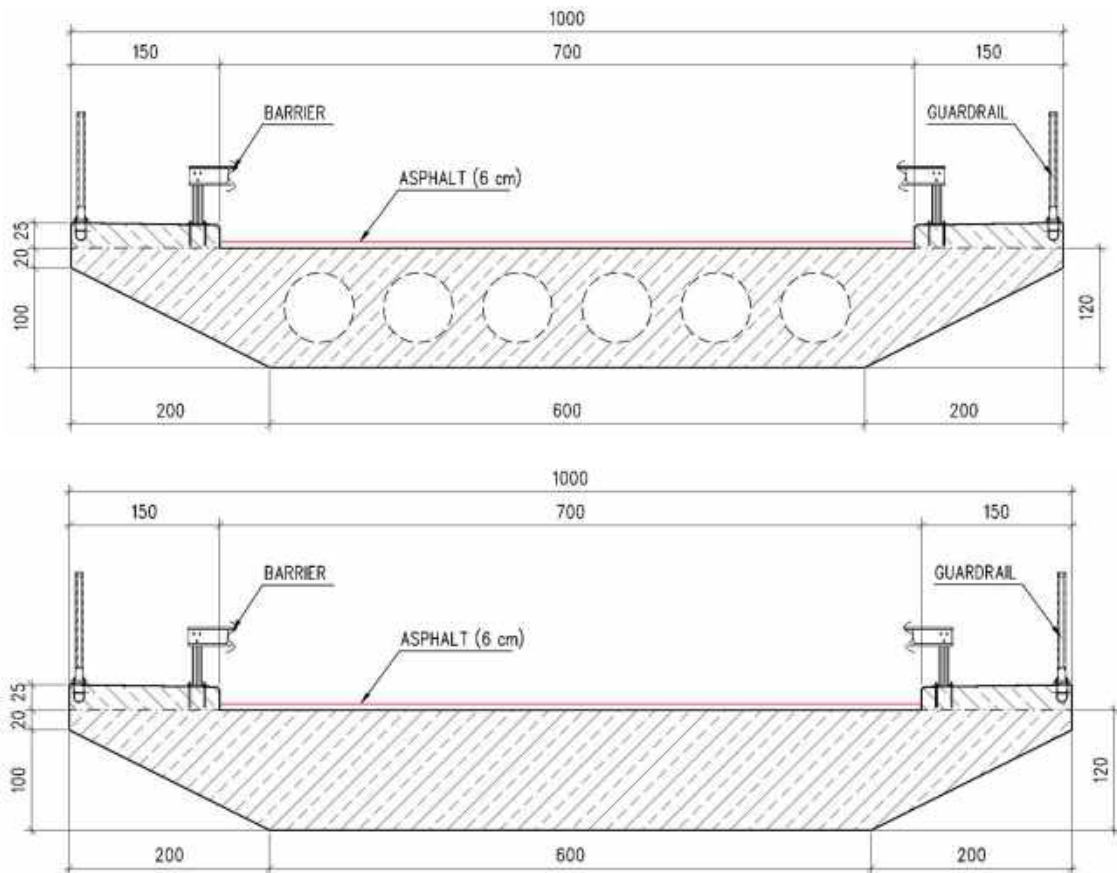
Figure 1. Voided and Filled Deck Typical Sections

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

2.4. Materials

Table 1. Concrete Properties

Class	Compressive Strength, f_{cy}	Modulus of Elasticity, E_c
-	(MPa)	(MPa)
C35/45	34000	25

Unit Weight	Poisson Ratio	Thermal Expansion Coefficient
(kN/m^3)	-	(1/^{o}C)
25	0.20	1.00 E-5

Table 2. Steel Properties

Yield Strength, f_y	Modulus of Elasticity, E_s
MPa	(MPa)
420	200000

Table 3. Post-Tension Strand Properties

Ultimate Tensile Strength, f_{pu}	Modulus of Elasticity, E_s	Yield Strength, f_{py}
MPa	(MPa)	MPa
1860	200000	1674

Prestressing Force	Friction Coefficient	Wobble Coefficient
MPa	-	(1/m)
1395	0.18	0.002

2.5. Modeling Methodology

Bridge decks were modeled by using MIDAS Civil software. Modeling method was selected as frame analysis. Two elastomeric bearings definition at each axis are included in the model as link elements. Link mechanical properties are calculated and assigned as their real values. Decks were modeled regarding to real geometries. Two main sections, voided and filled type, were defined to represent deck sections. At abutment and pier axis, a filled region is assigned to resist shear forces and the other sections are defined as voided slab to reduce

self-weight. Boundary conditions are defined with two types of supports. At one axis, translations and rotations are fixed at all directions. At the other axis, only longitudinal translation is defined as free. Node local axis was same as global axis: X along longitudinal axis, Y along Transverse axis, Z along vertical. Elements are defined from their top center points.
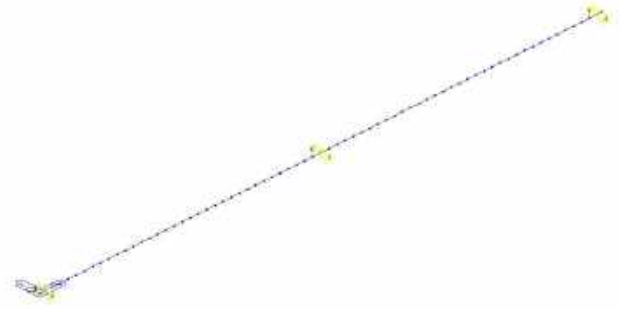
Figure 2. View of Frame Geometry
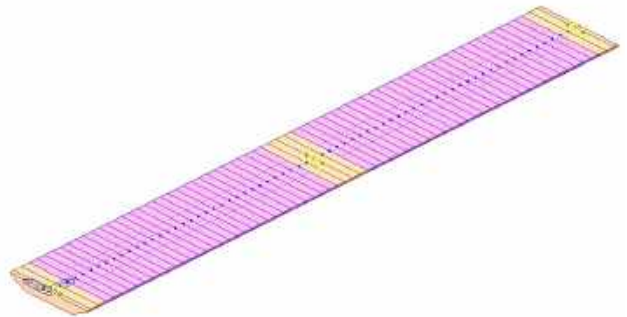
Figure 3. View of 3D Deck Geometry

2.6. Actions on Structure

2.6.1. Permanent Actions

Self-Weight, DC: Self-weight of the deck was considered in the structural loading by setting out concrete unit weight as 25.0 kN/m^3.

Additional Dead Loads, DW: Asphalt, sidewalk and railing loads were calculated as superimposed dead loads and applied to the analysis as external loads. In asphalt and sidewalk loads calculations, width and depth parameters were used with combination of accepted unit weights. Railing loads were taken from specifications as 1.50 kN/m.

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

Prestressing Force: Prestressing tendons are stressed with 1395 N/mm^2 force from two ends at the same time.

2.6.2. Transient Actions

HL-93 vehicular live loading was used in analysis for two lanes bridge deck. HL-93 loading consists of design truck / design tandem (whichever governs) plus lane load.

Design Truck, LL: Design truck loading schema was given in figure below. Axle load P1 equals to 35 kN and P2 & P3 axles equal to 145 kN each. Spacing between front axles is 4.30 m while rear axis spacing varies 4.30 to 9.00 m. Lane load, W is a uniformly distributed load over 3.00 m. widths as 9.34 kN/m.
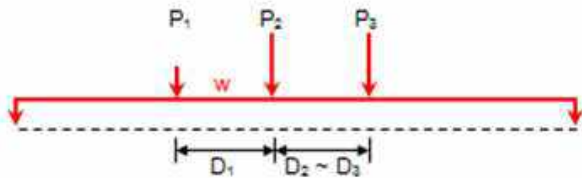
Figure 4. HL-93 Design Truck Loading

Design Tandem, LL: Design tandem consists of a pair of 110 kN axles spaced 1.20 m. apart. The transverse spacing of wheels is 1.80 m. Design lane load is also a requirement in design tandem loading.
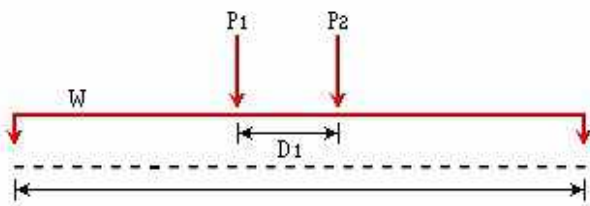
Figure 5. HL-93 Design Tandem Loading

Dynamic load allowance was selected as 33 % for deck component. Multiple presence factor was used as 1 because of two lanes loading.

Pedestrian Loads, PL: A pedestrian load 3.60 kN/m^2 was applied to all sidewalks and considered simultaneously with the vehicular design live load.

Braking Actions, BR: Braking force consists of two options mentioned below and greater of two was used.

- 25 % of the axles weights of design truck or design tandem,

- 5 % of the design truck + lane load or 5 % of the design tandem + lane load.

Temperature Actions, TU & TG: Temperature actions were applied in combination of the two loading cases which are system temperature changing and section temperature differences. In two cases, both cooling and heating actions were considered. For the first case, initial system temperature was selected as +10ºC and effects of system temperature changes up to +30ºC and -15ºC were investigated. For the second case, top and bottom faces of deck were assigned ±10ºC separately to investigate internal forces due to the section temperature differences.

Wind Actions, WS & WL: Wind actions on structure and on live load were investigated in two loading cases. On structure wind load effect, 1.90 kN/m^2 uniformly distributed load was assigned to the deck height. On live load, 1.46 kN/m^2 uniformly distributed load was assigned to deck total height of deck plus 1.80 m live load height.

2.7. Load Combinations

Service Combinations

Service 1: 1.0DC + 1.0DW + 1.0LL + 1.0BR + 1.0PL + 0.3WS + 1.0WL + 1.0 TU + 0.5TG

Service 2: 1.0DC + 1.0DW + 1.3LL + 1.3BR + 1.3PL + 1.0 TU

Service 3: 1.0DC + 1.0DW + 0.8LL + 0.8BR + 0.8PL + 1.0 TU + 0.5TG

Service 4: 1.0DC + 1.0DW + 0.7WS + 1.0 TU + 1.0TU

Strength Combinations

Strength 1: 1.25DC + 1.5DW + 1.75LL + 1.75BR + 1.75PL + 0.5 TU + 0.5TG

Strength 2: 1.25DC + 1. 5DW + 1.35LL + 1.35BR + 1.35PL + 0.5 TU + 0.5TG

Strength 3: 1.25DC + 1.5DW + 1.40WS+ 0.5 TU + 0.5TG

Strength 4: 1.50DC + 1.50DW + 0.5 TU

Strength 5: 1.25DC + 1.5DW + 1.35LL + 1.35BR + 1.35PL + 0.4WS + 1WL+ 0.5 TU + 0.5TG

Notations

DC: Dead Load of Structural Components and Attachments

DW: Dead Load of Wearing Surfaces and Utilities

SH: Force Effects due to Shrinkage

BR: Vehicular Braking Force

LL: Vehicular Live Load

PL: Pedestrian Live Load

WL: Wind Load on Live Load

WS: Wind Load on Structure

TU: Force Effect due to Uniform Temperature

TG: Force Effect due to Temperature Gradient

2.8. Construction Stages

In the first stage, it is assumed that deck concrete was poured and all prestressing tendons were stressed. Self-weight of the deck and tendon forces were activated at this stage. In order to see the creep and shrinkage activities, additional stages; 100 days, 1000 days, 5000 days and 10000 days were introduced to software. It was figured out that at 10000 days, all losses were almost completed and bridge was in service case.

2.9. Allowable Stress Limits

Table 4. Allowable Stress Limits

Stress Type	First Stage	Service Stage
Compression	**0.45fc'**	**0.60fc'**
	15.75 MPa	21.00 Mpa
Tension	**0.50√fc'**	**0.50√fc'**
	2.95 Mpa	2.95 Mpa

2.10. Modeling of Tendons

For three types of deck analysis, tendons were placed regarding to deck stresses. At mid-span, they were placed at the bottom which was 15 cm away from deck bottom face. At the support regions, they were placed at the top which was 20 cm away from deck top face. Vertical spacing between tendons was 20 cm from center to center. Due to the span length differences in deck types, tendon layouts were slightly different at each type. In plan layout, 7 tendons were located horizontally at web sections and each web had 3 row tendons. Thus, the total number of tendons was 21 at each deck section. Each tendon has a group of strands number that will be compared in this paper. Area of each strand was 150 mm^2. Duct diameter was selected as 120 mm. Tendon layout for longitudinal section, plan view and cross sections are given in figures below.

3. Row Strands
2. Row Strands
1. Row Strands
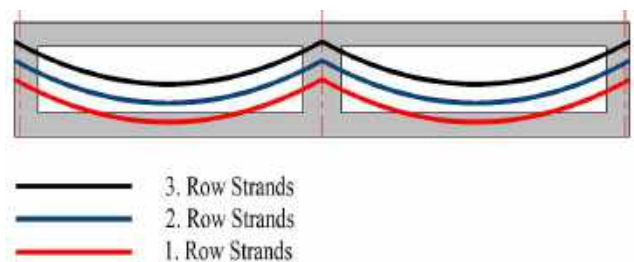
Figure 6. Strands Longitudinal Section Layout

Figure 7. Strands Plan Layout
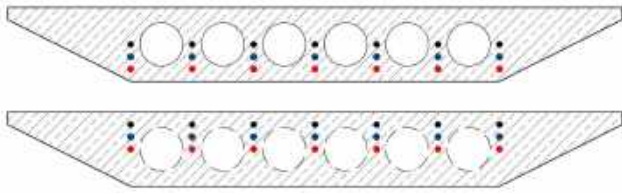
Figure 8. Mid-Span and Diaphram Sections

At each deck type analysis, 4 different strand area were controlled. Stress comparison results were given in tabular form and in graphs.

Table 5. Used Strand Numbers for Deck Types

Span Length	1. Step	2. Step	3. Step	4. Step
(m)	# of St.	# of St.	# of St.	# of St.
2 x 32	17	19	21	23
2 x 34	19	21	23	25
2 x 36	21	23	25	27

2.11. Interaction Between Deck Stresses and Strand Numbers

In stress results, (-) sign represents compression while (+) sign represents tension.

2.11.1. Deck Type 1 (2x32.00 m.)

Table 6. Deck Stresses for Type 1

Stage	Deck Face	17 St.	19 St.	21 St.	23 St.
		(Mpa)	(Mpa)	(Mpa)	(Mpa)
First	Top	-11.08	-12.48	-13.58	-14.80
	Bottom	-10.44	-12.33	-14.50	-16.50
Service	Top	3.17	1.73	0.33	0
	Bottom	-15.75	-15.43	-15.12	-14.83

At deck type 1, number of strands used 19 and 21 satisfies all stress limits for the first stage and service stage. At 17 strands check, at the top face of the pier axis, tension stress was obtained as 3.17 MPa that exceeds the tension stress capacity limit at service stage. At 23 strands check, at the bottom face of the mid-span, compression stress was obtained as 16.50 MPa which also exceeds the 15.75 MPa compression stress capacity at the first stage when self-weight and tendon forces were activated. Hence, step 1 and 4 (17 & 19 strands) are not acceptable regarding to stress capacity limits. Step 2 (19 strands) is the most reasonable choice for the design of deck type 1. Above mentioned values are given in graph below in order to be more representative.
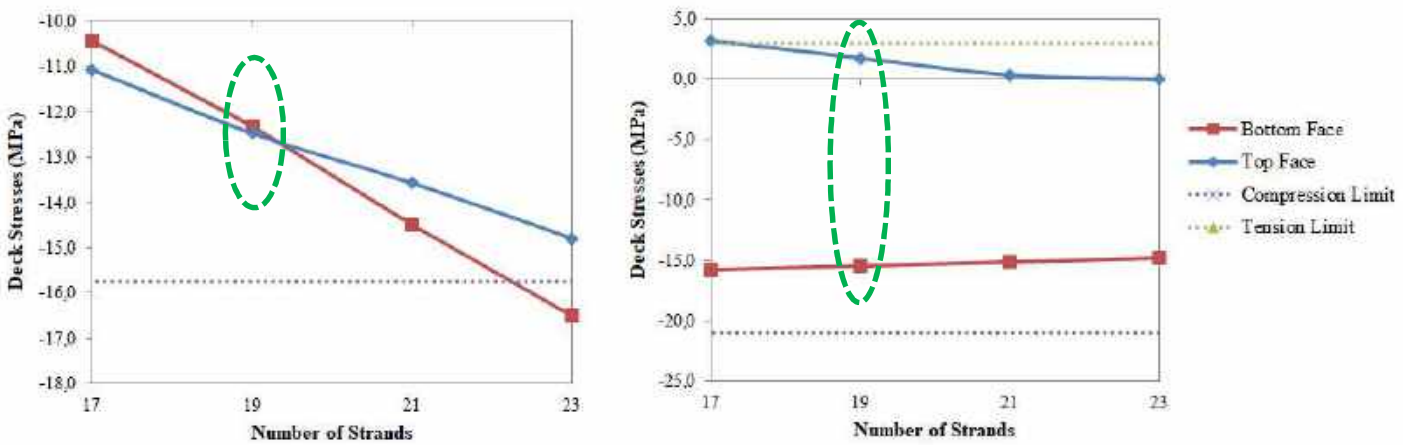
Figure 9. Deck Type 1, Stresses vs. Strands Graph (First Stage & Service Stage, Respectively)

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

2.11.2. Deck Type 2 (2x34.00 m.)

Table 7. Deck Stresses for Type 2

Stage	Deck Face	19 St.	21 St.	23 St.	25 St.
		(Mpa)	(Mpa)	(Mpa)	(Mpa)
First	Top	-12.35	-13.59	-14.83	-16.05
	Bottom	-11.55	-13.58	-15.58	-17.55
Service	Top	3.76	2.35	0.99	0.34
	Bottom	-17.69	-17.38	-17.09	-16.80

At deck type 2, number of strands used 21 and 23 satisfies all stress limits for the first stage and service stage. At 19 strands usage, top face tension stress exceeds the limits. At 25 strands usage, top and bottom face compression stress exceeds the compression stress capacity at the first stage. Hence, step 1 and 4 (19 & 25 strands usages) are not acceptable regarding to stress capacity limits. Step 2, 21 strands usage is the most reasonable choice for the design of deck type 2. Above mentioned values are given in graph below in order to be more representative.
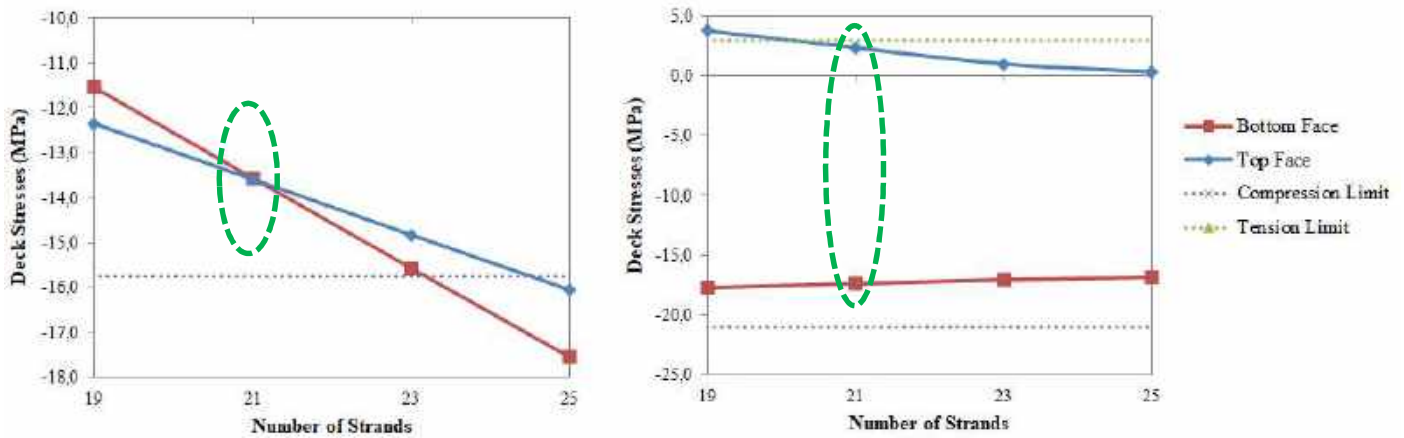
Figure 10. Deck Type 2, Stresses vs. Strands Graph (First Stage & Service Stage, Respectively)

2.11.3. Deck Type 3 (2x36.00 m.)

Table 8. Deck Stresses for Type 3

Stage	Deck Face	21 St.	23 St.	25 St.	27 St.
		(Mpa)	(Mpa)	(Mpa)	(Mpa)
First	Top	-13.74	-15.21	-15.68	-18.21
	Bottom	-12.65	-14.67	-15.20	-18.61
Service	Top	4.98	3.66	2.36	1.09
	Bottom	-20.32	-20.08	-19.84	-19.70

At deck type 3, number of strands used 25 satisfies all stress limits for the first stage and service stage. In other iterations, compression and tension stress limit problems occur. When using 21 or 23 strands, tension stress exceeds the limits for service stage. At 27 strands usage, compression stress exceeds the limits at the first stage. Hence, step 1, 2 and 4 are not acceptable regarding to stress capacity limits. Only step 3, 25 strands usage is suitable for the design of deck type 4. Above mentioned values are given in graph below in order to be more representative.

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

Figure 11. Deck Type 3, Stresses vs. Strands Graph (First Stage & Service Stage, Respectively)

2.12. Results

Decks that have three different span lengths were analyzed. The optimum number of strand for each type was decided. As indicated figure below, number of strand and span lengths are almost directly proportional.

Table 9. Calculated Optimum Strand Number for Deck Types

	Type 1	Type 2	Type 3
	2x32 m	2x34 m	2x36 m
Calculated Strand Number	19	21	25
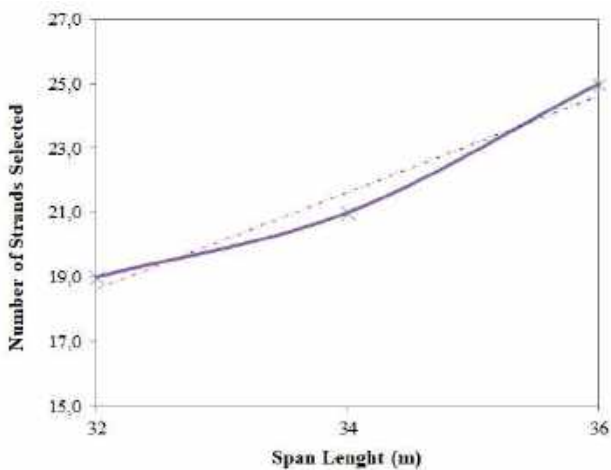
Figure 12. Span Lengths vs. Strands Graph

In type 1 analysis, obtained deck compression and tension stresses are 21 % and 42 % far from the capacity limits. Since there exists a reasonable gap between load and capacity, any optimization techniques can be investigated for this design. Changing deck elements with thinner sections, minimizing tendon diameters, decreasing material strengths can be studied on this deck type.

In type 2 analysis, ratios between loads and capacities are 14 % and 20 % for compression and tension, respectively. Design is reasonable for this type. Adequate capacity limits are available.

In type 3 analysis, mentioned ratios are 0.44 % and 19 % for compression and tension, respectively. For the given tendon layout and 36 m. span length, typical section works at the upper limits. Any improvement can be studied on design. Two important and the convenient improvement techniques are changing material properties and changing tendon layout. Increasing material strength is a good way to enhance stress limits. Tendon longitudinal profile also affects deck stress. While center of the tendons pass through deck peak stress points, this will lead to increase deck capacity. To shift tendon center to the faces, strands can be sidled by bundling.

About tendon optimization issue, three main topics were observed during calculations which are tendon longitudinal layout, initial tendon stress and tendon stress application prioritize. For the whole analysis models, above mentioned topics were investigated cautiously.

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

Tendon horizontal layout did not affect the stress distribution in analyses since decks were regular; in alignment and not skewed. However, tendon longitudinal layout affects stress distribution significantly. By much iteration, peak stress points were detected. Hence, tendon profile was decided by compound of peak values.

Initial tendon stresses were given 0.75 x f_{pu} by specification. Thus, it is all about tendon type. As the initial force increases, deck tension stresses decrease which is a preferred case mostly by the designer. However, first stage compression limits should be checked if initial tendon force increases.

Stressing tendons from both ends at the same time is a good way which causes low stresses because elastic deformation loss occurs only once at the first time.

To conclude, this research focused on PT concrete bridge decks and tendon usage. It is found out that many parameters affect the system solution which can show differences among designers. The objective procedures developed herein is to minimize the tendon number by choosing them in the most convenient way which leads to decrease the construction cost. Reaching the most suitable conditions safely and performing these conditions with lower costs can be named as the "best" design.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] G. Ranzi and A. Ostinelli, "Ultimate behaviour and design of post-tensioned composite slabs", Engineering Structures, vol. 150, pp. 711-718, 2017.

[2] J. D. Chaitanya Kumar and L. Venkat, "Genetic algorithm based optimum design of prestressed concrete beam", International Journal of Civil and Structural Engineering, vol. 3, pp. 644–654, 2013.

[3] P. Fanning, "Nonlinear models of reinforced and posttensioned concrete beams", Electronic Journal of Structural Engineering, vol. 2, pp. 111–119, 2001.

[4] M. A. Pisani, "Behaviour under long-term loading of externally prestressed concrete beams", Engineering Structures, vol. 160, pp. 24-33, 2018.

[5] O. F. Hussien, T. H. K. Elafandy, A. A. Abdelrahman, S. A. Abdel Baky, and E. A. Nasr, "Behavior of bonded and unbonded prestressed normal and high strength concrete beams", HBRC Journal, vol. 8, pp. 239–251, 2012.

[6] M. A. Utrilla and A. Samartin "Optimized design of the prestress in continuous bridge decks", Computers and Structures, vol. 64, issue 1-4, pp. 719-728, 1997.

[7] M. S. Kim, and Y. H. Lee, "Flexural behavior of posttensioned flat plates depending on tendon layout", Advances in Materials Science and Engineering, vol. 2016, p. 11, 2016.

[8] A. Ghallab, "Calculating ultimate tendon stress in externally prestressed continuous concrete beams using simplified formulas", Engineering Structures, vol. 46, pp. 417-430, 2013.

[9] W. Marks, W. Trochymiak "The selection of a system of prestressing tendons in hyperstatic beams as a problem of linear

Mustafa TURAN

A Case Study on Two-Span Post-Tensioned Concrete Bridge Decks with Different Span Lengths and Investi...

integer programming", Structural and Multidisciplinary Optimization, vol. 3, pp. 59-67, 1991.

[10] S. A. Quiroga, U. M. A. Arroyo "Optimization of prestressed concrete bridge decks", Computers & Structures, vol. 41, pp. 553-559, 1991.

Title: The Effect of Thickness on Photocatalytic Performance in MgZnO Thin Films

The Effect of Thickness on Photocatalytic Performance in MgZnO Thin Films

Mehmet KURU[*][1]

Abstract

In this study, the effect of thickness of MgZnO thin films which provide high efficiency as photocatalyst under UV light on structural and photocatalytic performance was investigated. For this reason, MgZnO thin films were produced by RF/DC magnetron sputtering technique at room temperature. MgZnO thin films at different thicknesses were deposited on Si (100) substrate and samples were subsequently annealed in the oven at 400 °C for 1 hour. Structural and morphological properties of MgZnO thin films were investigated using the Scanning Electron Microscopy (SEM), Grazing Incident X-ray diffraction (GIXRD) and Atomic Force Microscopy (AFM). All films have hexagonal-wurtzite crystal structure. Also, the crystallite size was 22.95 nm for the 400 nm film and the average crystallite size rised to 35.42 nm with increasing film thickness. The results showed that the structural properties roughness and surface morphology of the films varied depending on the thickness. Also, photocatalytic performances of MgZnO thin films at different thicknesses were measured by UV–Vis spectroscopy. The reaction rate constant (k) for MgZnO photocatalyst with a thickness of 800 nm was calculated as $27.86 \times 10^{-2} \text{s}^{-1}$. This result shows that photocatalysts of different thickness suitable with the first-order velocity law because the thin films degradation in the low concentration methylene blue (MB) solution.

Keywords: MgZnO thin films, Magnetron co-sputtering, Photocatalytic performance, Methylene Blue

---

[*] Corresponding Author: kurumehmet54@gmail.com
[1] Department of Metallurgy and Materials Engineering, Ondokuz Mayıs University, 55139, Samsun, Turkey
  ORCID: https://orcid.org/0000-0001-6030-0791

1. INTRODUCTION

The fact that textile dyestuffs do not break down in nature is one of the major factors that cause pollution of limited drinking water resources. Various organic compounds used as dyes, pesticides, insecticides, phenols and solvents are widely used in industrial applications and in daily life. These substances are mixed with water resources and cause pollution of our natural resources [1,2]. Also, the wastewater generated by the textile industry contains highly toxic aromatic amine compounds and is one of the main causes of environmental pollution [3]. These compounds having a complex aromatic structure are highly resistant to chemicals, light, temperature and other factors. Due to these properties, they cannot be treated efficiently by conventional treatment methods. Recently, Advanced Oxidation Process (AOP) is one of the promising methods for the purification of organic aromatic compounds from wastewater. In this method, exposure of the catalyst or oxidizing compound to UV light or visible light produces hydroxyl radicals on their surfaces. These radical groups attack the bonds of organic pollutants to break them into harmless compounds [4]. The photocatalysis method, which is one of the advanced oxidation processes, promises the future because it is economical and highly efficient than the methods using Ozone and Hydrogen Peroxide [5].

Many of the semiconductor metal oxides have been extensively studied because they retain their stable structure when stimulated by UV and visible light. These can be referred to as Titanium Oxide (TiO_2), Cerium Oxide (CeO_2), Zinc Oxide (ZnO), Tin Oxide (SnO_2), Tungsten Oxide (WO_3) and Zirconium Oxide (ZrO_3) [6,7]. ZnO, which is one of the semiconductor metal oxides, is preferred more than other metal oxides in the removal of organic pollutants due to its anti-bacterial, high quantum yield and environmentally friendly [8]. In addition, since ZnO absorbs much more of the light spectrum than TiO_2 [9], it can break down organic pollutants more effectively. The preparation methods, phase purity, band energy range, surface area, additive and crystal size affect the photocatalytic activity of ZnO [10]. When the band energy range increases, their photocatalytic activity increases because the semiconductor metal oxides occur high-energy electron-hole pairs [11].

Nowadays, the decomposition of dyes as a result of the stimulation of semiconductor metal oxides with UV light is one of the promising treatment methods. ZnO is widely preferred among metal oxides due to its wide band energy and oxidation ability [12]. In addition, ZnO is inexpensive, antibacterial and environmentally friendly, allowing it to stay one step ahead of other photocatalysts and attract more attention [13]. Although ZnO has a high photocatalytic potential [10], its use under high energy UV light is limited due to the narrow band energy range. This narrow band gap is one of the main factors that reduce the photocatalytic activity of ZnO and is one of the most important barriers to its use in commercial applications. In recent studies [7], it has been observed that the addition of elements with different properties to ZnO increases the photocatalytic activity of ZnO by creating band energy gap and surface defects.

The simplest and most convenient method of increasing the band energy range is to add elements to ZnO thin films [11]. Element doping may cause thermal imbalances in the crystal structure. However, it allows the formation of factors that increase the photocatalytic activity in the crystal size and surface area of the films. The doping occurs in two different groups using metal or nonmetallic ions. Metals are preferred as doping elements because they can easy transfer electrons and rise the band energy range [14]. In general, Al, Cu Mg, Ag, Ni, Mn and Co are used as doping elements. [15]. Recently Mg^{+2} (0.57Å) is more preferred than other metals [16] because it has similar radius of ion as Zn^{+2} (0.60Å) and can easily settle into the crystal structure. In addition, it is used in ZnO doping due to its low cost and harmless to the environment [17]. The increase of band energy range with the addition of Mg is due to the Moss-Burstein effect of electrons occurred by oxygen gaps. During doping, Mg^{+2} and Zn^{+2} are replaced by different ionic radius and electronegativity, increasing electron

concentration and oxygen gaps [16]. Also, Mg doping allows the extension of the absorbed wavelength from the UV-A region to the UV-B and UV-C region [17]. The absorbed wavelength and band energy ranges, which are important in photocatalytic applications, can be increased by Mg addition.

Various methods such as RF/DC magnetron sputtering, chemical vapor deposition (CVD), sol-gel, metal organic vapor phase epitaxy (MOVPE), pulsed laser deposition (PLD) and molecular beam epitaxy (MBE) can be used in the production of MgZnO films [18]. Magnetron sputtering method attracts attention due to its easy control deposition rate, chemical composition, film formation [19]. It is also possible to produce materials with different properties by changing the production parameters with this method. Since magnetron co-sputtering method takes place under high vacuum, it enables the production of homogeneous thin films of high purity and quality. Consequently, it is possible to produce films with electrical, optical and magnetic properties that are resistant to corrosion and corrosion [20]. Because of these properties, the co-sputtering method promises future to produce Mg doped ZnO thin films.

2. MATERIALS AND METHODS

MgZnO thin films were coated by RF/DC magnetron co-sputtering technique on Si (100) substrate. The Mg target material has a purity of 99.95% and ZnO has a purity of 99.999%. Substrates was cleaned in the ultrasonic bath with acetone, ethanol and isopropanol respectively for 5 min each and dried with nitrogen. The vacuum chamber was flushed several times with argon to avoid any contamination gas in the chamber, and subsequently purged down to 2×10^{-7} Torr base pressure. We pre-sputtered the targets for 90 min before starting deposition for remove the oxide layer from the surface of the targets. During pre-sputtering, the substrate shutter was kept closed to avoid any contamination of the substrates. 10 W DC power sputtering of Mg and 75 W RF power sputtering of ZnO were used for fabrication of 400 nm, 500 nm, 600 nm, 700 nm, 800 nm and 900 nm MgZnO films. The thickness of the films was measured by the Inficon©-QCM thickness measurement system during the growth process. The deposition was completed in 4mTorr deposition pressure, 0.4 Å/s coating rate and 15 sccm argon gas flow at room temperature. After the deposition of films at different thickness, annealed oxygen atmosphere in the oven at 400 °C for 1 hour.

3. RESULTS AND DISCUSSIONS

After completed fabrication, structural properties were characterized by using (GIXRD) in 2θ range of 20°–80°. Morphological and roughness properties of MgZnO thin films were examined with SEM and AFM. The photocatalytic performance of MgZnO thin films was investigated by degradation of MB solution having a concentration of 1 mmol/ L. The MB solution was irradiated by continuous stirring for 14 hours with a UV lamp of 254 nm wavelength. The differences in absorbance were monitored by UV–Vis spectrophotometer. The XRD spectrum of MgZnO films deposited on Si (100) having a thickness of 400 nm, 500 nm, 600 nm, 700 nm, 800 nm, 900 nm and annealed at 400 °C is given in Figure 1. As seen from the XRD spectrum, 2θ = 34.5° (002) and 2θ = 63 (103) peaks were observed in all films. These peaks confirm that MgZnO thin films of different thickness have hexagonal-wurtzite structure according to the JCPDS 36-1451 [21]. However, low intensity (200) and (111) MgO peaks were also observed in MgZnO films produced at different thicknesses and subjected to heat treatment at 400 °C [22]. The same crystal structure was observed for all films, but the peak full width at half maximum (FWHM) and these peak intensities varied depending on the film thickness. With the increase of film thickness, the peak intensities increased while the FWHM decreased. As the film thickness increases, it is understood that there is a growth race between adjacent crystals relative to the crystal orientation. This competition leads to the growth of crystal faces of the same type to form a free surface. The growth of these crystal surfaces and the decrease in the FWHM by increasing the peak intensity increases the crystalline quality of the films. In MgZnO films deposited at different thicknesses, peak intensities

increased as the thickness increased, while the maximum peak intensity was observed in 800 nm thickness film. After this thickness, peak intensity decreased and (111), (200) MgO peaks became more dominant.

The average crystal sizes (D) of the films were calculated with the help of the Debye-Scherrer equation given by equation 1 [7].

$$D = \frac{0.9\lambda}{\beta cos\theta} \qquad (1)$$

In here λ, θ and β are expressed as wavelength of X-ray, Bragg's angle and angular width of peaks at FWHM, respectively. According to the calculations, the crystallite size was 22.95 nm for the 400 nm film and the average crystallite size rised to 35.42 nm with increasing film thickness.
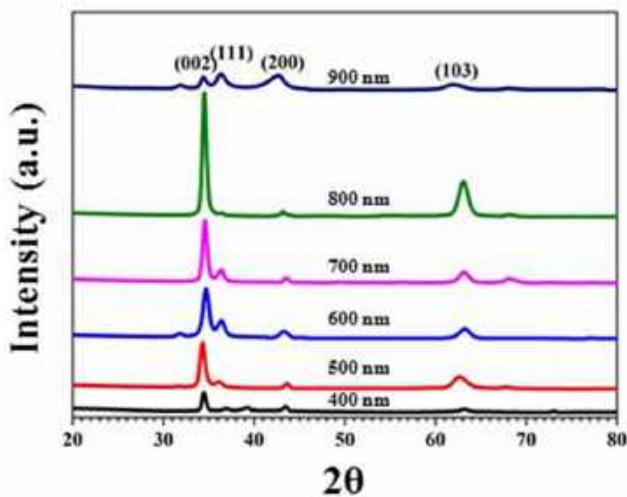
Figure 1. XRD diffraction pattern of MgZnO thin films at different thickness

a, b and c lattice parameters of the hexagonal system are calculated by the equation 2 [23].

$$\frac{1}{d^2} = \frac{4\,(h^2+k^2+hk)}{3a^2} + \frac{l^2}{c^2} \qquad (2)$$

In this equation d, (hkl) and a,c expressed as interplanar spacing, miller indices and lattice parameters, respectively. In addition, the volume of the unit cell (V) is calculated using equation 3 [23],

$$V = \frac{\sqrt{3}}{2} a^2 c \qquad (3)$$

Average crystallite size (D), lattice parameters (a=b, c) and unit cell volumes of MgZnO thin films are given in Table 1. As can be seen clearly in the Table 1, average crystal sizes are different as peak intensities and the FWHM vary depending on thickness. Also, the calculated lattice constants of thin films with different thicknesses are in agreement with the standard values (a=b=3.25Å, c=5.20Å) [24]. This indicates that the films have tensile strength along the c axis for all thicknesses.

Table 1. The peak full width at half maximum (FWHM), Average crystallite size (D), lattice parameters (a = b, c), unit cell volume and Root mean square roughness values of MgZnO thin films

Thickness (nm)	FWHM	D (nm)	a (Å)	c (Å)	V (Å)³	RMS (nm)
400	0.759	22.92	3.228	5.206	46.978	8.845
500	0.620	28.02	3.271	5.224	48.375	9.674
600	0.569	30.56	3.256	5.163	47.449	9.452
700	0.510	34.08	3.238	5.182	47.052	9.997
800	0.505	34.42	3.237	5.192	47.150	27.916
900	0.750	23.20	3.332	5.212	50.112	14.929

Surface and cross-sectional SEM images of MgZnO thin films with different thicknesses and annealed at 400 °C are given in Figure 2. When the surface SEM images are examined, it is seen that the substrate surface is completely homogeneously and uniformly coated with MgZnO particles. It is also clear that film thickness has a large effect on surface morphology. SEM results show that coating thickness can directly affect crystal growth, grain boundary amount and grain size of films, which can play an important role in photocatalytic properties. The MgZnO film, which has a thickness of 400 nm, consists of fine-grained structures, and as the thickness increases, the grains combine to form coarse-grained structures. The largest grains were observed in the 800 nm

film, while the 900 nm film reduced the grain size. The results obtained from the SEM images agree with the XRD calculations. EDX analysis of MgZnO thin films is given in Table 2.

Table 2. EDX results of MgZnO thin films

Atomic (%)	Film Thickness (nm)					
	400	500	600	700	800	900
Zn	33.74	40.02	44.08	43.12	40.79	42.08
O	54.36	49.28	45.21	46.17	45.52	47.21
Mg	11.90	10.70	10.71	10.71	13.69	10.71

When the EDX results were examined, no impurity peaks were observed for each film except Zn, O and Mg. In addition, Mg ratios in all films are very close to each other. These results show that Mg is successfully added to ZnO by RF/DC magnetron sputtering method simultaneously.
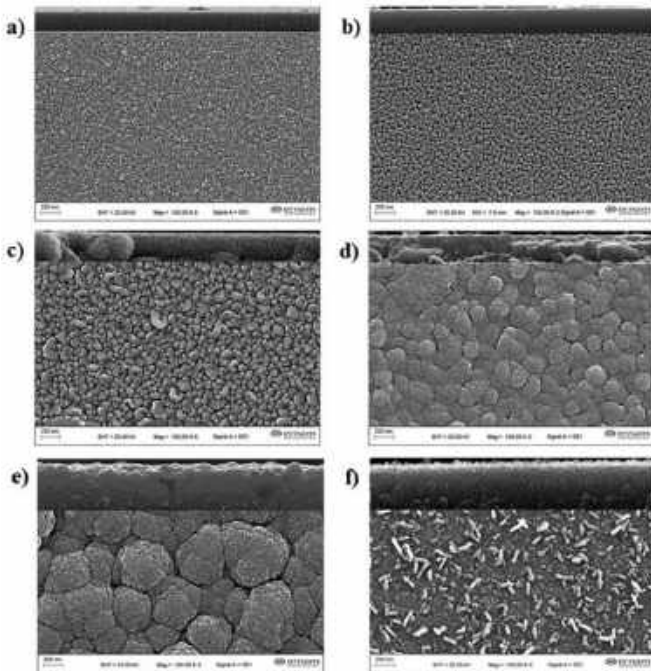
Figure 2. SEM images of MgZnO thin films. a) 400nm, b) 500nm, c) 600nm, d) 700 nm, e) 800 nm and f) 900 nm

The change of surface roughness of MgZnO thin films depending on film thickness was examined by atomic force microscopy. 3D AFM images of

5×5 μm^2 dimensions of MgZnO thin films of different thicknesses are given in Figure 3 and the root mean square (RMS) values are given in Table 1. As shown in Figure 3, 400 nm thick film is composed of small grains and RMS value is 8.845 nm. As the film thickness increased, the small grains joined to form larger grains and RMS value increased. The highest surface roughness was observed in 800 nm thickness film with 27.916 nm value. With increasing film thickness, the increase in RMS roughness is sourced from the larger grain formation as well as the increase in porosity of the films [25]. Also, RMS roughness values were increased with film thickness similar to the average crystal sizes obtained from XRD results.
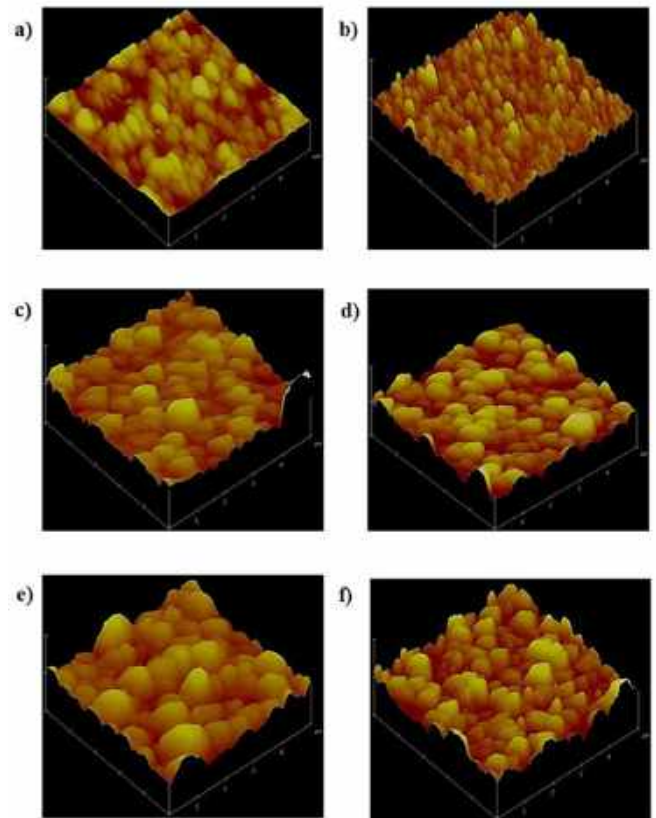
Figure 3. 3D AFM images of MgZnO thin films. a) 400nm, b) 500nm, c) 600nm, d) 700 nm, e) 800 nm and f) 900 nm.

In MgZnO film with a thickness of 900 nm, peak intensity, average crystal size and surface roughness were reduced as in XRD results. The change of grain sizes and surface fluctuations can explain the difference in RMS roughness of thin films. In films with high surface roughness, we

think that photocatalytic performance increases because the surface area increases. [26]

3.1. Photocatalytic performance

In the photocatalysis method, when the photocatalyst is irradiated with a high energy light, the electrons in the valence band (VB) are transferred to the conductivity band (CB). Thus, electron holes (h^+) are occurred in the VB. The holes obtained radical OH^- groups by oxidizing the water molecules on the photocatalyst. These OH^- groups convert organic dyes to harmless form. As the increased electron density in the conductivity band decreases the amount of negatively charged oxygen by increasing the amount of dissolved oxygen [27], the degradation rate also increases [28].
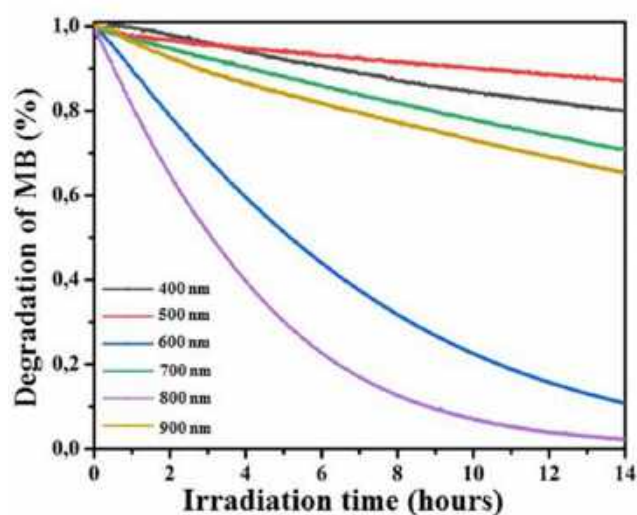
Figure 4. Comparison of time dependent degradation rate of MB for MgZnO thin films

The photocatalytic performance of the MgZnO thin films was calculated by degrading MB by stimulating with a 254 nm wavelength UV lamp. MgZnO thin films with different thickness of UV light irradiation time versus the degradation rate of MB is given in Figure 4. In the MgZnO photocatalyst with a thickness of 500 nm, the degradation rate of MB was 14%, while the 800 nm thickness photocatalyst was 99%. The degradation rate of MB for MgZnO photocatalysts is given in Table 3. As seen in Table 3 and Figure 4, there is no linear relationship between film thickness and

degradation. In other words, as the film thickness increased, there was no increase in the rate of degradation. While the maximum degradation of MB was observed in MgZnO photocatalyst having a thickness of 800 nm with 99% ratio, almost all the dyes were degraded. When the degradation rates are compared with the average grain size and surface roughness values, the results are in agreement with each other. The best results were obtained in 800 nm thickness MgZnO thin film.

Table 3. The degradation rate of MB and kinetic rate constant values for MgZnO photocatalysts

Thickness (nm)	Degradation rate of MB (%)	Kinetic rate constants (k, s^{-1}) (10^{-2})
400	21	1.52
500	14	0.86
600	89	17.34
700	30	2.43
800	99	27.86
900	35	2.92

While the maximum degradation of MB was observed in MgZnO photocatalyst having a thickness of 800 nm with 99% ratio, almost all the dyes were degraded. When the degradation rates are compared with the average grain size and surface roughness values, the results are in agreement with each other. The best results were obtained in 800 nm thickness MgZnO thin film.

In order to determine the photocatalytic performance of MgZnO thin films, reaction kinetics of degradation were investigated. To define the kinetics of photocatalytic degradation, several models have been developed such as zeroth-order velocity law, the first-order velocity law, and the parabolic diffusion model [29]. In

diluted solutions, the reaction takes place in accordance with the first order kinetics given by Eq.4,

$$ln\left(\frac{C}{C_o}\right) = -kt \qquad (4)$$

In this equation C is the concentration after reaction, C_o initial concentration, k reaction rate constant and t UV irradiation time. The time-dependent $ln(C/C_o)$ graph for MgZnO films is given in Figure 5. The kinetic rate constant (k) values calculated from the slope of Figure 5 are given in Table 3.6.
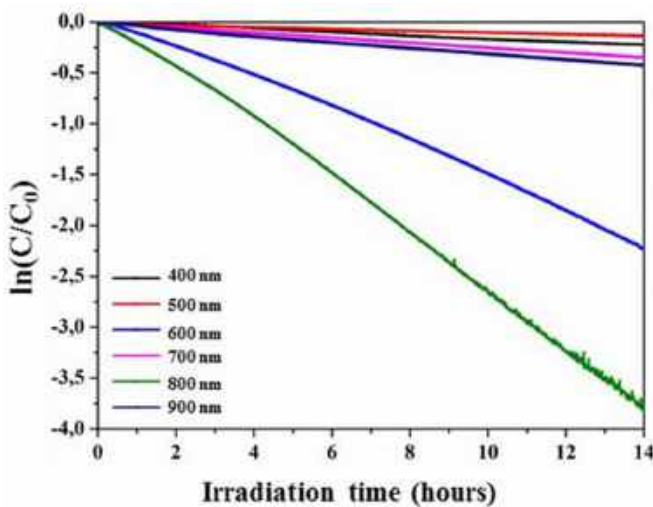
Figure 5. The time-dependent $ln(C/C_o)$ graph for MgZnO films

The kinetic rate constant (k) for the MgZnO photocatalyst with a thickness of 500nm was $0.86 \times 10^{-2} s^{-1}$, while the value for the film with a thickness of 800 nm was calculated as $27.86 \times 10^{-2} s^{-1}$, and the reaction rate constant value increased almost 30 times. Some factors such as crystallinity, crystal orientation and surface area affect the photocatalytic performance of MgZnO thin films [30]. For MgZnO photocatalysts of different thickness, degradation of MB at low concentrations indicates that it is suitable with the first-order velocity law.

4. CONCLUSIONS

In order to rise the photocatalytic performance of ZnO, MgZnO thin films with different thicknesses were produced by RF/DC magnetron

sputtering technique. When XRD results are examined, (002) and (103) peaks seen in films with different thicknesses indicate that MgZnO films are hexagonal wurtzite-type structure. The effect of thickness on photocatalytic performance was investigated by degradation of MB solution. The lowest photocatalytic performance was observed in a film having a thickness of 500 nm with a degradation rate of 14%, while the best performance was observed in a film having a thickness of 800 nm with rate 99%. As a result of the experiments, it was observed that the thickness did not have a linear effect on photocatalytic performance. But although there is no linear ratio, MgZnO photocatalysts of different thickness suitable with the first-order velocity law because the thin films degradation in the low concentration MB solution.

Research and Publication Ethics

The author declares that all the information in this paper is accurate and completely correct that he has behaved in accordance with the scientific ethic in the course of the production of the information and cited all the sources that he has used.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] M. Grassi, G. Kaykioglu, V. Belgiorno and G. Lofrano, "Removal of emerging contaminants from water and wastewater by adsorption process, in Emerging Compounds Removal from Wastewater: Natural and Solar Based Treatments," ed. by G. Lofrano, Springer, Dordrecht, 2012.

[2] K. Kümmerer, "The presence of pharmaceuticals in the environment due to human use—present knowledge and future challenges," Journal of Environmental Management, vol. 90, pp. 2354–2366, 2009.

[3] T. Robinson, G. McMullan, R. Marchant and P. Nigam, "Remediation of dyes in textile effluent: a critical review on current treatment technologies with a proposed alternative". Bioresource Technology, vol. 77, pp. 247-255, 2001.

[4] H. Lachheb, F. Dappozze, A. Houas and C. Guillard, "Adsorption and photocatalytic degradation of cysteine in presence of TiO$_2$", Journal of Photochemistry and Photobiology A: Chemistry, vol. 246, pp. 1-7, 2012.

[5] Y. Deng and R. Zhao, "Advanced oxidation processes (AOPs) in wastewater treatment," Current Pollution Reports, vol. 1, pp. 167–176, 2015.

[6] A. Ibhadon and P. Fitzpatrick, "Heterogeneous photocatalysis: recent advances and applications," Catalysts, vol. 3, pp. 189-218, 2013.

[7] M. Kuru and H. Narsat, "The effect of heat treatment temperature and Mg doping on structural and photocatalytic activity of ZnO thin films fabricated by RF magnetron co-sputtering technique," Journal of Materials Science: Materials in Electronics, vol. 30, pp. 18484–18495, 2019.

[8] W. Bolong, Y. Fucheng, L. Haishan, S. Tianyun, N. Dongmei, H. Ling, D. Hongyan, W. Shu and X. Tang, "The preparation and photocatalytic properties of Na doped ZnO porous film composited with Ag nano-sheets", Physica E: Low-dimensional Systems and Nanostructures, vol. 117, pp. 113712, 2020.

[9] O. D. M. Navarro and J. L. Sánchez-Salas, "Focus on Zinc Oxide as a Photocatalytic Material for Water Treatment", International Journal of Bioremediation & Biodegradation, vol. 116, 2018.

[10] S. Chakrabarti and B.K. Dutta, "Photocatalytic degradation of model textile dyes in wastewater using ZnO as semiconductor catalyst," Journal of Hazardous Materials, vol. 112, pp. 269–278, 2004.

[11] S. Suwanboon, P. Amornpitoksuk, P. Bangrak and N. Muensit, "Optical, photocatalytic and bactericidal properties of Zn$_{1-x}$La$_x$O and Zn$_{1-x}$Mg$_x$O nanostructures prepared by a sol–gel method," Ceramics International, vol. 39, pp. 5597–5608, 2013.

[12] D. R. Kumar, K.S. Ranjith, L.R. Nivedita, K. Asokan and R.T. Rajendra Kumar, "Swift heavy ion induced effects on structural, optical and photo-catalytic properties of Ag irradiated vertically aligned ZnO nanorod arrays", Nuclear Instruments and Methods in Physics Research B, vol. 450, pp. 95–99, 2019.

[13] A. Ziashahabi, M. Prato, Z. Dang, R. Poursalehi and N. Naser, "The efect of silver oxidation on the photocatalytic activity of Ag/ZnO hybrid plasmonic/metal-oxide nanostructures under visible light and in the dark", Scientific Reports, vol. 9, pp. 11839, 2019.

[14] M.R.D. Khaki, M.S. Shafeeyan, A.A.A. Raman and W. Daud, "Application of doped photocatalysts for organic pollutant degradation-a review," Journal of Environmental Management, vol. 198, pp. 78–94, 2017.

[15] K. Verma, B. Chaudhary, V. Kumar, V. Sharma and M. Kumar, "Investigation of structural, morphological and optical properties of Mg:ZnO thin films prepared by sol-gel spin coating method," Vacuum vol. 146, pp. 524–529,2017.

[16] T.B. Ivetić, M.R. Dimitrievska, N.L. Finčur, L.R. Đačanin, I.O. Gúth, B.F. Abramović and S.R. Lukić-Petrović, "Effect of annealing temperature on structural and optical properties of Mg-doped ZnO nanoparticles and their photocatalytic efficiency in alprazolam degradation," Ceramics International, vol. 40, pp. 1545-1552, 2014.

[17] V. Etacheri, R. Roshan, and V. Kumar, "Mg-doped ZnO nanoparticles for efficient sunlight-driven photocatalysis," ACS Applied Materials & Interfaces, vol. 4, pp. 2717-2725, 2012.

[18] C. Abed, C. Bouzidi, H. Elhouichet, B. Gelloz and M. Ferid, "Mg doping induced high structural quality of sol–gel ZnO nanocrystals: application in photocatalysis," Applied Surface Science, vol. 349, pp. 855–863, 2015.

[19] M. Kuru, A. E. Özmetin, A. Özmetin and O. Şahin, "The role of heat treatment on the structural and nano-mechanical properties of $SmCo_5$ thin films grown by RF magnetron sputtering technique," Ceramics International, vol. 43, pp. 3893-3899, 2017.

[20] S. Youssef, P. Combette, J. Podlecki, R. Al Asmar, and A. Foucaran, "Structural and optical characterization of ZnO thin films deposited by reactive rf magnetron sputtering", Crystal Growth & Design, vol. 9, pp.1088–1094, 2008.

[21] C. Liua, F. Shang, G. Pan, F. Wang, Z. Zhou, W. Gong, Z. Zi, Y. Wei, X. Chena, J. Lv, G. He, M. Zhang, X. Song and Z. Sun, "Preparation and photocatalytic activity of $Mg_xZn_{1-x}O$ thin films on silicon substrate through sol–gel process," Applied Surface Science, vol. 305, pp. 753–759, 2014.

[22] S. K. Mahadeva, J. Fan, A. Biswas, K. S. Sreelatha, L. Belova, and K. V. Rao, "Magnetism of Amorphous and Nano-Crystallized Dc-Sputter-Deposited MgO Thin Films", Nanomaterials, vol. 3, pp. 486-497, 2013.

[23] A. H. Hammad, M. Sh Abdel-Wahab, S. Vattamkandathil and A. R. Ansari, "Structural and optical properties of ZnO thin films prepared by RF sputtering at different thicknesses," Physica B, vol. 540, pp. 1–8, 2018.

[24] B.D. Cullity and C.D. Graham, "Introduction to Magnetic Materials" Wiley, Hoboken, 2009.

[25] V. Kumar, N. Singh, R.M.Mehra, A. Kapoor, L.P.Purohit and H.C.Swart, "Role of film thickness on the properties of ZnO thin films grown by sol-gel method", Thin Solid Films, vol. 539, pp. 161-165, 2013.

[26] J. Lv, W. Gong, K. Huang, J. Zhu, F. Meng, X. Song and Z. Sun, "Effect of annealing temperature on photocatalytic activity of ZnO thin films prepared by sol–gel method", Superlattices and Microstructures, vol. 50, pp. 98-106, 2011.

[27] M. A. Rauf and S. S. Ashraf, "Fundamental principles and application of heterogeneous photocatalytic degradation of dyes in solution", Chemical Engineering Journal, vol. 151, pp. 10-18, 2009.

[28] Z. Wang, C. Chen, F. Wu, B. Zou, M. Zhao, J. Wang and C. Feng, "Photodegradation of rhodamine B under visible light by bimetal co-doped TiO_2 nanocrystals," Journal of Hazardous Materials, vol. 164, pp. 615–620 2009.

[29] Zhaohui Li, "Sorption kinetics of hexadecyltrimethylammonium on natural clinoptilolite," Langmuir, vol. 15, pp. 6438–6445, 1999.

[30] N. M. Flores, U. Pal, R. Galeazzi, and A. Sandoval, "Effects of morphology, surface

area, and defect content on the photocatalytic dye degradation performance of ZnO nanostructures," RSC Advances, vol. 4, pp. 41099-41110, 2014.

Title: Non-Markovian Corrections to Quantum Regression Theorem for the Strong Coupling
Spin-Boson Model

Non-Markovian Corrections to Quantum Regression Theorem for the Strong Coupling Spin-Boson Model

Arzu KURT[*1]

Abstract

We report the results of an investigation of the effects of non-Markovian corrections to the dynamics of two-time correlation functions of the strong interaction spin-boson model. Beyond quantum regression theorem corrections are taken into account at the low environmental temperatures for a two-level system (TLS) which is in contact with a structured bath. The results indicate that the corrections lead to appreciable (small) quantitative (qualitative) differences for both biased and non-biased TLS settings.

Keywords: Quantum regression theorem, open quantum systems, spin-boson model, quantum dynamics, two-time correlation functions

---

*Corresponding Author: arzukurt@ibu.edu.tr
[1] Bolu Abant İzzet Baysal University, Department of Physics, 14030-Bolu, Turkey.
  ORCID: https://orcid.org/0000-0002-2345-3059

1. INTRODUCTION

Two-time correlation functions (TCFs) for an open quantum system are important tools that provide essential information which might not be available within the single-time averages of the physical system quantities. For instance, TCFs of an atom interacting with electromagnetic field allows to evaluate the fluorescence spectrum [1, 2]. Also, in case of current flowing through nanostructure instruments, TCFs of the current play an important role in explaining the transition properties of current deviations and noise spectrum. Quantum regression theorem (QRT) serves as a bridge between the dynamics of the single-time expectation values and that of their corresponding TCFs and is a useful method to evaluate these two-time (multi-time) correlation functions for open quantum systems when the system-environment interaction is Markovian. With the help of QRT, the knowledge of time rate of change of the reduced density matrix of the system lets one to obtain not only the single-time average values, but also two-time correlation functions in the Markovian case. However, the QRT does not work generally for non-Markovian open quantum systems. To handle this problem, there have been many studies that are concerned with finding dynamical equations for the multi-time correlation functions for non-Markovian quantum systems [3–12], such that two-time functions for non-Markovian systems can be computed using a method similar to the QRT. Alonso and de Vega [7–10] have developed a theory for non-Markovian multi-time correlation functions based on stochastic Schrodinger equation and used it to study a weakly interacting dissipative system with non-diagonal interaction and have shown that validity of quantum regression theorem depends strongly on the form of the system environment correlations. The developed approach was applied to compute the emission spectrum of a two-level system in contact with a structured non- Markovian environment, for example the radiation fields in a a photonic crystal [10]. Goan et. al. [4, 11] have derived an useful two-time correlation functions for non-Markovian dissipative quantum systems in finite temperature environments for any separable system-environment initial states (pure or mixed) by using perturbative quantum master equation, and found that calculating the two-time correlation functions of system quantities of non-Markovian dissipative open systems such as a pure-dephasing spin-boson model is not sufficient to use its single-time average, even in the second order approximation. Also, this exact model allows making the non-Markovian environment temperature arbitrary or finite, which is another important result of these studies.

Some groups have tried to develop a non-Markovianity measure based on the violation of QRT [3, 12]. Manirul Ali and coworkers [12] have investigated non-Markovianity measures based on two-time correlation functions for open quantum system by using an exact master equation based on the non-equilibrium Green's functions. They have found that the non-Markovian dynamics for the Fano-Anderson Hamiltonian significantly depends on the strength of the system-bath coupling and various physical parameters, and tend to show two different behaviors depending on time for different spectral densities. Besides, it is observed that non-Markovian memory effect on the system is always reduced by the thermal bath disturbances.

McCutcheon [5] derived a non-Markovian extension to the quantum regression theorem that gives us the facility about the calculation of two-time correlation functions and emission spectrum of weakly driven dissipative two-level system by using projection method. He has found that sideband are related to information flow from the phonon environment to the quantum dot system, which justifies true non-Markovianity and indivisibility of the dynamical map [13]. Cosacchi et. al. [6] present a practical method to calculate the multi-time correlation functions by using path integral method in the presence of the memory of the environment. Cosacchi et. al. [6] observes that the approach of McCutcheon [5] may lead to unphysical results, which presents phonon side-band appears on the wrong side of the Mollow triplet. They study it on the pure-dephasing type coupling to bosonic harmonic environments. They have shown that the resulting method that is used to characterize the emission spectrum of a quantum dot interacting with longitudinal

acoustic phonons allows to the phonon sideband appears on the correct side of the zero phonon line, which is not an agreement with the result obtained from the naive application of the quantum regression theorem.

In the literature, the studies relating to the time evolution of TCFs in the non-Markovian Dynamics considered only the weak system-bath coupling regime. One can naturally ask whether findings of those studies would hold also in the strong coupling regime? Towards an aim to answer that question, in the current work, we will analyze the contribution of non-Markovian corrections to quantum regression theorem of the two-time correlation function of the σ_z operator for the strong coupling spin-boson model with an environment that is described by a structured spectral density.

The article is organized as follows. In Sec. II, we present the strongly driven spin-boson mode in the polaron frame and give basic information about the expressions of time evolution of one and twotime correlation functions for any system operators. The important results of study are the presented in Sec. III. The paper concludes with a brief summary of the findings in Sec. IV.

2.   MODEL

We consider a two-level system (TLS) that is interacting with an harmonic environment, of which the energy splitting is ω_A and the tunneling matrix element is V. The Hamiltonian of the closed system composed of the TLS and its environment can be expressed as:

$$H = H_S + H_B + H_I, \qquad (1)$$

where H_S is the Hamiltonian of the TLS, H_B is the Hamiltonian of the reservoir which is a collection of independent harmonic oscillators with mode frequencies ω_k, and H_I is the interaction between the TLS and its environment (with $\hbar = 1$):

$$H = \frac{\omega_A}{2}\sigma_z + \frac{V}{2}\sigma_x + \sum_k \omega_k a_k^\dagger a_k$$
$$+ \sigma_z \sum_k g_k \left(a_k^\dagger + a_k\right). \qquad (2)$$

Here $a^\dagger$ (a) is a creation (annihilation) operators of the bath oscillator and g_k is the magnitude of the interaction between the TLS and the kth bath oscillator. Here, we assume that the interaction strength is in the strong coupling regime, so transforming Hamiltonian in Eq. (2) into polaron frame is very natural, which can be obtained by

$$H' = H_0' + H_I',$$
$$= \frac{\omega_A}{2}\sigma_z + V_r \sigma_x + \sum_k \omega_k a_k^\dagger a_k$$
$$+ \sigma_+ B_- + \sigma_- B_+. \qquad (3)$$

Here the superscript " ′ " implies that O' is in the polaron frame. In the following we will drop the superscript for simplicity. $\sigma_\pm$ are, then, the spin flip operators of the TLS, V_r is the reduced tunneling matrix element and $B_\pm$ refer the bath correlation operators, and given as a function of system-bath coupling strength g_k with the kth oscillator mode:

$$B_\pm = \left\langle e^{\mp \Sigma_k \frac{2g_k}{\omega_k}\left(a_k^\dagger - a_k\right)} \right\rangle_R, \qquad (4)$$

where $\langle \dots \rangle_R$ denotes averaging over the environmental degrees of freedom. It is important to note that the reduced tunneling rate V_r is equal to zero for bath spectral densities whose frequency exponent is less than 2 which is the case in the present study.

The evolution equation of the reduced density matrix of the TLS in the interaction picture can be written as

$$\frac{d}{dt}\rho_S(t) = -\int_0^t dt_1 Tr_R[H_I(t), [H_I(t_1), \rho_S \otimes \rho_R]], (5)$$

where Tr_R indicates partial trace over the environmental modes. Second-order evolution equations of any system operator A in the Schrödinger Picture can be evaluated by using

$$\frac{d}{dt_1} \langle A(t_1) \rangle = iTr_{S \otimes R}(\{[H_S, A]\}(t_1)\rho_T(0))$$
$$+ \int_0^{t_1} d\tau Tr_{S \otimes R}(\{\tilde{H}_I(\tau - t_1)[A, H_I]\}(t_1)\rho_T(0)$$
$$+ \{[H_I, A]\tilde{H}_I(\tau - t_1)\}(t_1)\rho_T(0)). \tag{6}$$

Here, $\tilde{H}_I(t)$ is the time-dependent interaction Hamiltonian:

$$\tilde{H}_I(t) = V(\sigma_-(t)B_+(t) + \sigma_+(t)B_-(t)), \tag{7}$$

which describes the time evolution in the interaction picture in the polaron frame. Curly brackets in Eqs. (6) and (8) indicate that the expression should be evaluated in the Heisenberg picture and its time should be taken as given in the post bracket. In order to establish the two time correlation functions for the non-Markovian evolution which has the optical results of some physical systems [1, 2], one can use the result of Refs. [4, 5], and its formulation is given as:

$$\frac{d}{dt_1} \langle A(t_1)B(t_2) \rangle = iTr_{S \otimes R}(\{[H_S, A]\}(t_1)B(t_2)\rho_T(0))$$
$$+ \int_0^{t_1} d\tau Tr_{S \otimes R}(\{\tilde{H}_I(\tau - t_1)[A, H_I]\}(t_1)B(t_2)\rho_T(0)$$
$$+ \{[H_I, A]\tilde{H}_I(\tau - t_1)\}(t_1)B(t_2)\rho_T(0))$$
$$+ \int_0^{t_2} d\tau Tr_{S \otimes R}\left(\{[H_I, A]\}(t_1)\{[B, \tilde{H}_I(\tau - t_2)]\}(t_2)\rho_T(0)\right). \tag{8}$$

3. RESULTS

Here, the first two terms on the right hand side of Eq. (8) are the QRT terms while the last term accounts for the corrections for the non-Markovian effects. The last integral term in Eq. (8) is the source of violation of the quantum regression theorem.

In the present study, we will investigate the dynamics of the quantity $\langle \sigma_z(t_1)\sigma_z(t_2) \rangle$ for the spinboson model in strong coupling regime within polaron frame for an arbitrary bath spectral density. Choosing $A = \sigma_z$ and $B = \sigma_z$ in Eq. (8), we obtain a set of six coupled differential equations for various combinations of TLS system operator as:

$$\frac{d}{dt_1} \langle \sigma_z(t_1)\sigma_z(t_2) \rangle = -4\Gamma_1(t_1)\langle \sigma_z(t_1)\sigma_z(t_2) \rangle - 4\Gamma_2(t_1)\langle \sigma_z(t_2) \rangle + 2(\Gamma_5(t_1, t_2)\langle \sigma_-(t_1)\sigma_+(t_2) \rangle$$
$$-\Gamma_6(t_1, t_2)\langle \sigma_+(t_1)\sigma_-t_2) \rangle - \Gamma_7(t_1, t_2)\langle \sigma_+(t_1)\sigma_+(t_2) \rangle$$
$$+\Gamma_8(t_1, t_2)\langle \sigma_-(t_1)\sigma_-(t_2) \rangle), \tag{9}$$

$$\frac{d}{dt_1} \langle \sigma_+(t_1)\sigma_-(t_2) \rangle = [i\omega_A - 2(\Gamma_1(t_1) + i\Gamma_3(t_1))]\langle \sigma_+(t_1)\sigma_-(t_2) \rangle$$
$$+2\Gamma_4(t_1)^*\langle \sigma_-(t_1)\sigma_-(t_2) \rangle + \Gamma_5(t_1, t_2)\langle \sigma_z(t_1)\sigma_z(t_2) \rangle, \tag{10}$$

$$\frac{d}{dt_1} \langle \sigma_-(t_1)\sigma_+(t_2) \rangle = -[i\omega_A + 2(\Gamma_1(t_1) - i\Gamma_3(t))]\langle \sigma_-(t_1)\sigma_+(t_2) \rangle$$
$$+2\Gamma_4(t_1)\langle \sigma_+(t_1)\sigma_+(t_2) \rangle + \Gamma_6(t_1, t_2)\langle \sigma_z(t_1)\sigma_z(t_2) \rangle, \tag{11}$$

$$\frac{d}{dt_1} \langle \sigma_+(t_1)\sigma_+(t_2) \rangle = [i\omega_A - 2(\Gamma_1(t_1) + i\Gamma_3(t_1))]\langle \sigma_+(t_1)\sigma_+(t_2) \rangle$$
$$+2\Gamma_4(t_1)^*\langle \sigma_-(t_1)\sigma_+(t_2) \rangle - \Gamma_8(t_1, t_2)\langle \sigma_z(t_1)\sigma_z(t_2) \rangle, \tag{12}$$

$$\frac{d}{dt_1}\langle\sigma_-(t_1)\sigma_-(t_2)\rangle = -[i\omega_A + 2(\Gamma_1(t_1) - i\Gamma_3(t_1))]\langle\sigma_-(t_1)\sigma_-(t_2)\rangle$$

$$+2\Gamma_4(t_1)\langle\sigma_+(t_1)\sigma_-(t_2)\rangle - \Gamma_7(t_1, t_2)\langle\sigma_z(t_1)\sigma_z(t_2)\rangle, \tag{13}$$

$$\frac{d}{dt_1}\langle\sigma_z(t_1)\rangle = -4\Gamma_1(t_1)\langle\sigma_z(t_1)\rangle - 4\Gamma_2(t_1). \tag{14}$$

Here, the first two terms in Eq.(9) (originating from the first two terms of Eq.(8)) account for the quantum regression theorem with non-Markovian effects while the rest of Eq.(9) describe the corrections to QRT for the non-Markovian

dynamics. These correction terms for QRT allow, then, us to get the last terms in Eq.(10)-Eq.(13), while the other terms in these equations come from the QRT only. Also, the time-dependent coefficients of the differential equations are:

$$\Gamma_1(t_1) = V^2\int_0^{t_1}d\tau e^{-Q_2(\tau-t_1)}\cos(Q_1(\tau-t_1))\cos(\omega_A(\tau-t_1)),$$

$$\Gamma_2(t_1) = V^2\int_0^{t_1}d\tau e^{-Q_2(\tau-t_1)}\sin(Q_1(\tau-t_1))\sin(\omega_A(\tau-t_1)),$$

$$\Gamma_3(t_1) = V^2\int_0^{t_1}d\tau e^{-Q_2(\tau-t_1)}\cos(Q_1(\tau-t_1))\sin(\omega_A(\tau-t_1)), \tag{15}$$

$$\Gamma_4(t_1) = V^2\int_0^{t_1}d\tau e^{-Q_3(\tau-t_1)}\cos(Q_1(\tau-t_1))e^{i\omega_A(\tau-t_1)},$$

$$\Gamma_5(t_1, t_2) = V^2\int_0^{t_2}d\tau e^{-Q_2(\tau-t_2-t_1)+iQ_1(\tau-t_2-t_1)}e^{i\omega_A(\tau-t_2)},$$

$$\Gamma_6(t_1, t_2) = V^2\int_0^{t_2}d\tau e^{-Q_2(\tau-t_2-t_1)+iQ_1(\tau-t_2-t_1)}e^{-i\omega_A(\tau-t_2)},$$

$$\Gamma_7(t_1, t_2) = V^2\int_0^{t_2}d\tau e^{-Q_3(\tau-t_2-t_1)-iQ_1(\tau-t_2-t_1)}e^{i\omega_A(\tau-t_2)},$$

$$\Gamma_8(t_1, t_2) = V^2\int_0^{t_2}d\tau e^{-Q_3(\tau-t_2-t_1)-iQ_1(\tau-t_2-t_1)}e^{-i\omega_A(\tau-t_2)},$$

where

$$Q_1(t) = \frac{1}{2\pi}\int_0^\infty d\omega\frac{J(\omega)}{\omega^2}\sin(\omega t),$$

$$Q_2(t) = \frac{1}{2\pi}\int_0^\infty d\omega\frac{J(\omega)}{\omega^2}\coth\left(\frac{\beta\omega}{2}\right)(1-\cos(\omega t)),$$

$$Q_3(t) = \frac{1}{2\pi}\int_0^\infty d\omega\frac{J(\omega)}{\omega^2}\coth\left(\frac{\beta\omega}{2}\right)(1+\cos(\omega t)),$$

where $Q_{1,2,3}(t)$ are the real and the imaginary parts of the bath correlation functions as a function of time, inverse temperature $\beta = 1/k_B T$. $J(\omega)$ is the bath spectral function which quantifies the frequency distribution of the bath oscillators as well as the interaction between the two level system and the bath. The reorganization energy which is defined as $E_r = \int_0^\infty d\omega\, J(\omega)/\omega$ is a rough measure of interaction strength between the TLS and its environment. In the current study, we

choose $J(\omega)$ as a structured spectral density function that defines the environment as a single harmonic oscillator whose levels are broadened by its interaction with an Ohmic environment of non-interacting harmonic oscillators [14]:

$$J(\omega) = 8\kappa^2\frac{\gamma\omega_0\omega}{(\omega^2-\omega_0^2)^2+4\gamma^2\omega^2}, \tag{16}$$

where ω_0 is the frequency of the central harmonic oscillator, κ is the strength of TLS-bosonic bath coupling, γ is the broadening term of the oscillator levels due to its interaction with the Ohmic bath. It should be noted that the reorganization energy for the chosen $J(\omega)$ is equal to κ^2/ω_0. For the structured spectral density, the bath correlation function $\exp(-Q_3(t))$ is zero which leads to vanishing of the kernels $\Gamma_4, \Gamma_7,$ and Γ_8 in Eqs. (9)-(14). Those equations in simplified form become:

$$\frac{d}{dt_1}\langle\sigma_z(t_1)\sigma_z(t_2)\rangle = -4\Gamma_1(t_1)\langle\sigma_z(t_1)\sigma_z(t_2)\rangle - 4\Gamma_2(t_1)\langle\sigma_z(t_2)\rangle$$

$$+2(\Gamma_5(t_1, t_2)\langle\sigma_-(t_1)\sigma_+(t_2)\rangle - \Gamma_6(t_1, t_2)\langle\sigma_+(t_1)\sigma_-t_2)\rangle), \tag{17}$$

$$\frac{d}{dt_1}\langle\sigma_+(t_1)\sigma_-(t_2)\rangle = [i\omega_A - 2(\Gamma_1(t_1) + i\Gamma_3(t_1))]\langle\sigma_+(t_1)\sigma_-(t_2)\rangle$$

$$+4\Gamma_5'(t_1,t_2)\langle\sigma_z(t_1)\sigma_z(t_2)\rangle, \tag{18}$$

$$\frac{d}{dt_1}\langle\sigma_-(t_1)\sigma_+(t_2)\rangle = -[i\omega_A + 2(\Gamma_1(t_1) - i\Gamma_3(t))]\langle\sigma_-(t_1)\sigma_+(t_2)\rangle$$

$$+4\Gamma_6(t_1,t_2)\langle\sigma_z(t_1)\sigma_z(t_2)\rangle, \tag{19}$$

$$\frac{d}{dt_1}\langle\sigma_z(t_1)\rangle = -4\Gamma_1(t_1)\langle\sigma_z(t_1)\rangle - 4\Gamma_2(t_1). \tag{20}$$

It is significant to note that the time rate of change of $\langle\sigma_z(t_1)\sigma_z(t_2)\rangle$ does not depend on $\langle\sigma_+(t_1)\sigma_+(t_2)\rangle$ and $\langle\sigma_-(t_1)\sigma_-(t_2)\rangle$ for the environment with the structured spectral density. As a result, six coupled differential equations of Eqs. (9)-(14) are reduced to four differential equations (Eqs. 17-20). In these equations, the terms that contain the kernel function ($\Gamma_{1,2,3}$) with a single time argument are due to non-Markovian quantum regression theorem while ($\Gamma_{5,6}$) with two time arguments account for the corrections to QRT for the non-Markovian dynamics which will be referred to as QRT+ for the rest of the paper. Non-Markovian QRT predictions reported below are obtained by neglecting $\Gamma_{5,6}$ terms in Eqs. (17)-(20).

As the numerical solutions of the dynamical equations of the two time correlation functions of the spin boson model in the current study, we will consider a low temperature environment with relatively strong system-bath coupling and investigate the difference between the non-Markovian QRT and QRT+ dynamics for two different values of level splitting of the two-level system. Initial state of the TLS+ bath closed system is taken as $\rho_0 = \rho_S(0) \otimes \rho_E(0)$ where $\rho_E(0)$ is the thermal equilibrium of the environment while $\rho_S(0)$ is the density matrix of the pure TLS state $|\psi\rangle = \frac{\sqrt{3}}{2}|e\rangle + \frac{1}{2}|g\rangle$.

First, we consider the zero-bias case for the TLS and display the kernel functions, magnetization and the two-time correlation functions obtained by numerically solving Eq. (17)-(20) in Fig. 1, respectively. The other system and the environment parameters which indicate a relatively strong system-bath interaction ($E_r = 1$) and low environmental temperature ($\beta = 100$) are provided in the caption of the figure. As can be seen from Fig. 1(a), for the non-biased TLS, only Γ_1 and Γ_6 kernels are nonzero. In the

$t \to \infty$ limit, both real and imaginary parts of Γ_6 vanishes. Behaviour of both $\langle\sigma_z(t)\rangle$ and the two time correlation functions can be understood easily based on the time dependence of these correlation functions: All the considered quantities should exponentially decay to zero at long time limit with a decay rate which is a function of the kernel Γ_1. The dynamics of two-time correlation functions obtained with non-Markovian QRT and QRT+ approaches (dashed and straight lines in Fig. 1(c) indicate that the qualitative difference between the results obtained by employing those two approaches is very small while the quantitative difference for $\langle\sigma_-(t_1)\sigma_+(t_2)\rangle$ and $\langle\sigma_+(t_1)\sigma_-(t_2)\rangle$ is larger than that for $\langle\sigma_z(t_1)\sigma_z(t_2)\rangle$.
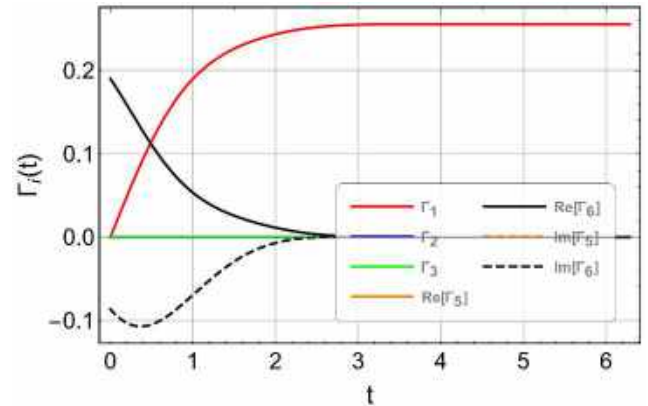
(a)
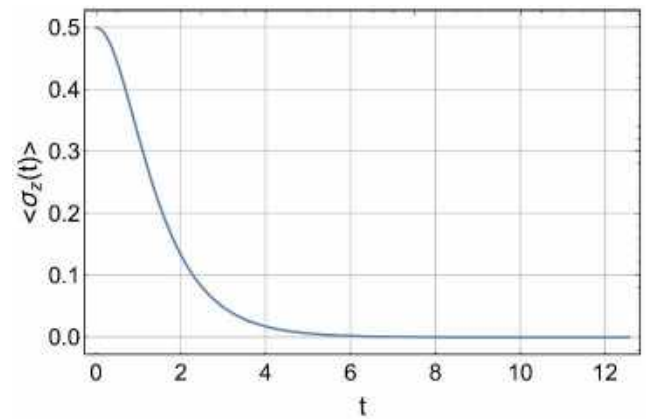
(b)
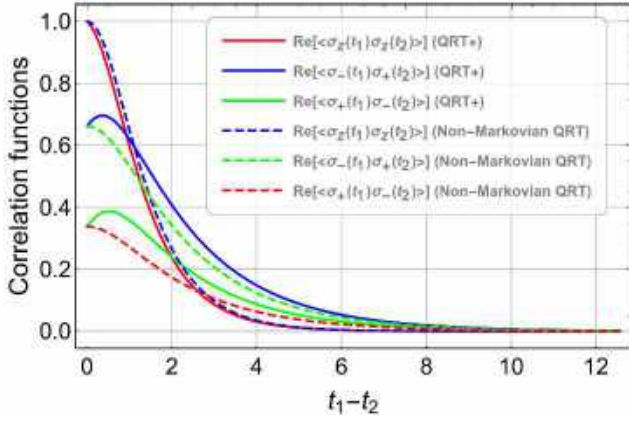
(c)

Figure 1. (a) Time-dependence of the kernels $\Gamma_{1,2,3}(t)$, and the real and the imaginary parts of $\Gamma_{5,6}(t_1, t_2)$ for the set of Eqs.(17)-(20), (b) dynamics of $\langle\sigma_z(t)\rangle$, and (c) time evolutions of the real parts of system operators for two cases: QRT (solid line) and non-Markovian QRT (dashed line). Other parameters used are $t_2 = 1$, $V = 0.5$, $\omega_0 = 1$, $\gamma = 0.1$, $\kappa = 1$, $\beta = 100$, $E_r = \kappa^2/\omega = 1$ and the TLS is considered to be non-biased ($\omega_0 = 0$).

We display, also, the same set of graphs in Figs.2(a)-(c) for the resonance case for which the splitting frequency of the TLS is equal to the central frequency of the environmental oscillator. Contrary to the non-biased case (Fig.1(a)), one can see that all kernels are non-zero at the short time while only the single time argument ones remain non-zero at the long time limit. This result causes that the average of $\sigma_z(t)$ to exponentially decay to a terminal value determined by the ratio $\Gamma_2(\infty)/\Gamma_1(\infty)$, as can be seen from Fig.2(b). It is also interesting to note that the dynamics of $\langle\sigma_z(t_1)\sigma_z(t_2)\rangle$ for the biased and non-biased TLS are very similar with somewhat different steady-state values as displayed Fig.1(c). The dynamics of $\langle\sigma_\pm(t_1)\sigma_\mp(t_2)\rangle$ for biased TLS differ from the those for the non-biased case in Fig.2(c). This result can readily be explained by the fact that both Γ_3 and ω_A are non-zero, which drives these system operators to coherent decay with decay rate related to the integrals of Γ_1, Γ_2, Γ_5, Γ_6 and oscillation frequency which is related to ω_A and Γ_3. The oscillations of two-time correlation functions are more pronounced in the resonance case. The difference between the results obtained by using QRT+ and non-Markovian QRT approaches is found to be smaller for the $\langle\sigma_\pm(t_1)\sigma_\mp(t_2)\rangle$ than that for

$\langle\sigma_z(t_1)\sigma_z(t_2)\rangle$ quantitatively, contrary to the findings in the non-biased case discussed above. In general, the results presented above show that the Dynamics for both QRT+ and the non-Markovian QRT have noticeably similar behavior for all time intervals for the strongly driven spin-boson model. This result is not valid for the weak system-bath coupling for the certain system parameters [4, 11]. Finally, we can conclude that the non-Markovian two-time correlation functions obtained in the polaron frame has somewhat different behavior depending on the TLS splitting term ω_A.
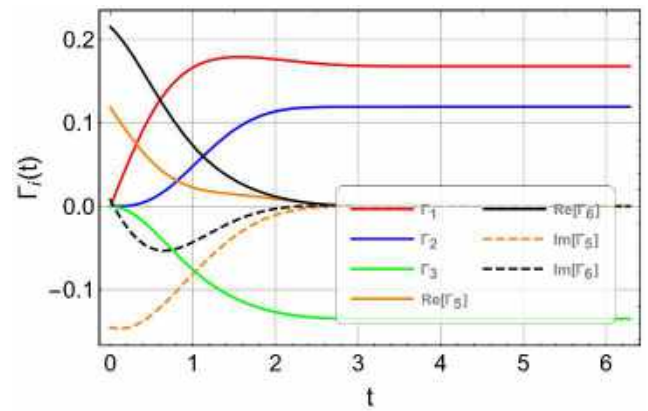
(a)
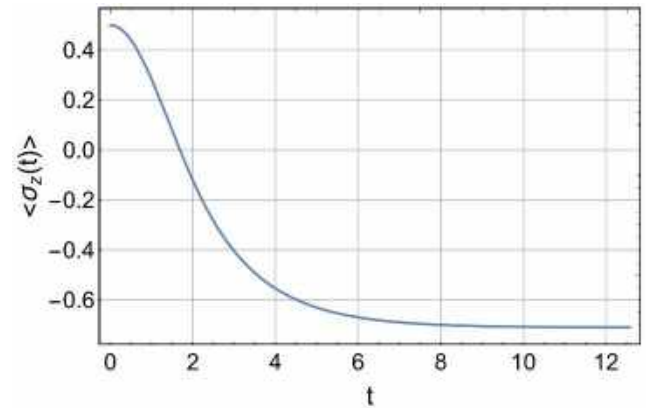
(b)
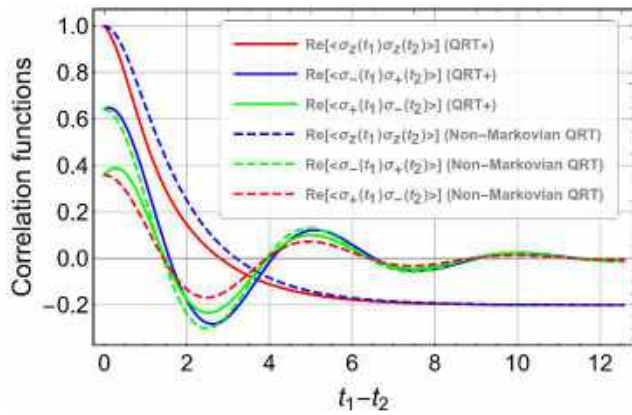
(c)

Figure 2. Same as Fig.(1) for the resonance between the TLS and the environmental oscillator ($\omega_A = \omega_0 = 1$).

4. CONCLUSIONS

In this work, we have investigated that timedependence of non-Markovian two-time correlation functions for a TLS that is in contact with a harmonic oscillator which in turn is interacting with a thermal bath of harmonic oscillators by transforming the system Hamiltonian into polaron frame when the strength of system-bath interaction is strong. We have investigated the effect of including non-Markovian corrections to the dynamics of the twocorrelation functions beyond the quantum regression theorem. In the low temperature, relatively strong system-bath interaction regime we have found that although the QRT corrections lead to small qualitative changes in the dynamics, overall qualitative time-dependence of the two time-correlation functions does not change appreciably by inclusion of the QRT corrections. This result was found to be true for both biased and non-biased system settings. If QRT violations can be considered as an indication of non-Markovianity of the dynamics, one can claim that the dynamics of the model is near Markovian for the parameter regime studied in the current work.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics**.**

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] M. O. Scully and M. S. Zubairy, "Quantum Optics", Cambridge Press, 1997.

[2] H. J. Carmichael "Statistical Methods in Quantum Optics", Springer Press, 1999.

[3] G. Guarnieri, A. Smirne, B. Vacchini, "Quantum regression theorem and non-Markovianity of quantum dynamics", Phys. Rev. A, vol. 90, pp. 022110, 2014.

[4] H.-S. Goan, P.-W. Chen, C.-C. Jian, "Non-Markovian finite-temperature two-time correlation functions of system operators: Beyond the quantum regression theorem", J. Chem. Phys., vol. 134, pp. 124112, 2011.

[5] D. P. S. McCutcheon "Optical signatures of non-Markovian behavior in open quantum systems", Phys. Rev. A, vol. 93, pp. 022119, 2016.

[6] M. Cosacchi, M. Cygorek, F. Ungar, A. M. Barth, A. Vagov, V. M. Axt, "Path-integral approach for nonequilibrium multitime correlation functions of open quantum systems coupled to Markovian and non-Markovian environments", Phys. Rev. B, vol. 98, pp. 125302, 2018.

[7] D. Alonso, I. de Vega, "Multiple-Time Correlation Functions for Non-Markovian Interaction: Beyond the Quantum Regression Theorem", Phys. Rev. Lett., vol. 94, pp. 200403, 2005.

[8] D. Alonso, I. de Vega, "Hierarchy of equations of multiple-time correlation

functions", Phys. Rev. A, vol. 75, pp. 052108, 2007.

[9] I. de Vega, D. Alonso, "Non-Markovian reduced propagator, multiple-time correlation functions, and master equations with general initial conditions in the weak-coupling limit", Phys. Rev. A, vol. 73, pp. 022102, 2006.

[10] I. de Vega, D. Alonso, "Emission spectra of atoms with non-Markovian interaction: Fluorescence in a photonic crystal", Phys. Rev. A, vol. 77, pp. 043836, 2008.

[11] H.-S. Goan, C.-C. Jian, P.-W. Chen, "Non-Markovian finite-temperature two-time correlation functions of system operators of a pure-dephasing model", Phys. Rev. A, vol. 82, pp. 012111, 2010.

[12] M. M. Ali, P.-Y. Lo, M. W.-Y. Tu, W.-M. Zhang, "Non-Markovianity measure using two-time correlation functions", Phys. Rev. A, vol. 92, pp. 062306, 2015.

[13] H.-P. Breuer, E.-M. Laine, J. Piilo, "Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems", Phys. Rev. Lett., vol. 103, pp. 210401, 2009.

[14] A. Garg, J. N. Onuchic, V. J. Ambegaokar, "Effect of Friction on Electron Transfer in Biomolecules", J. Chem. Phys., vol. 83, pp. 4491-4503, 1985.

Title: Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and
Estimating the Firing Temperature depending on the Amount of Printed Aluminum

Authors: İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temperature depending on the Amount of Printed Aluminum

İmran KANMAZ[*1], Abdullah ÜZÜM[1,2]

Abstract

Optimization of back surface field (BSF) for crystalline silicon solar cells was carried out by Afors-Het simulation software. Thickness and doping concentration parameters were optimized and electrical parameters of solar cells both with BSF and non-BSF were analyzed. The optimum BSF thickness and doping concentration for the crystalline silicon solar cell were determined as 7 µm and 1×10^{19} cm^{-3}, respectively. A special attention was given to the estimation of peak firing temperature considering the printing amount of aluminum paste in order to form an optimal BSF by the calculations using Al-Si binary phase diagram. It was concluded that the temperature of up to 950°C should be established if an amount of 3 mg/cm^2 printed aluminum was used to achieve BSF thickness of 7 µm, where 775 °C would be enough when the amount of aluminum is 8 mg/cm^2.

Keywords: Afors-Het, back surface field, fire through, silicon solar cell

1. INTRODUCTION

Crystalline silicon (c-Si) solar cells are most widely used type of solar cell in photovoltaic applications. Reducing the losses in solar cells plays a key role to improve conversion efficiency. Mainly, losses including the reflection of incoming light to the front surface, shading losses of the front contact, resistance losses of metal contacts, and recombination losses of minority carriers in the surface regions with cell structure are the main obstacles [1]. Among them, surface recombination of minority carriers occurs at the back surface of the silicon solar cells is one of the facts which is needed to be considered in cell design. Conventional c-Si solar cells mostly have a fully-screen printed aluminum (Al) on back surface which possesses a high surface recombination velocity if not a highly doped p$^+$ region is created on the rear surface of the solar cell to minimize the recombination [2]. For instance, a study of back surface field (BSF) formed by the screen-printed and sputtering methods for c-Si solar cells with thermally

---

[*]Corresponding Author: imrankanmaz@ktu.edu.tr
[1]Karadeniz Technical University, Department of Renewable Energy Resources/Technologies, Trabzon, Turkey.
ORCID: https://orcid.org/0000-0001-8827-1590; https://orcid.org/0000-0001-5324-8892
[2]Karadeniz Technical University, Department of Electrical and Electronics Engineering, Trabzon, Turkey.
ORCID: https://orcid.org/0000-0001-5324-8892. E-mail:auzum@ktu.edu.tr

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

annealing at 700°C to 1000°C, a relative increase of 6% was achieved in efficiency that summarizes the importance of BSF [3]. There are several chief parameters influences the back-side metallization process and BSF including the thickness of BSF, doping concentration of BSF, amount of printed Al paste and firing temperature. The printing amount of Al commercially varies between 3-9 mg/cm^2 which effects other properties of BSF along with the firing temperature. For instance, Al-BSF lay down for a 6-inch industrial solar cell was reported averagely at around 5.4 mg/cm^2 [4]. The importance of the properties of BSF regions was highlighted on recent studies for PERC (Passivated Emitter Rear Contact) solar cells that shows the need of optimization for better local BSF properties [5]. Firing process of Al leading to variation of concentration and the thickness of the layer is a key factor in that sense. Moreover, controlling the concentration of Al doped layers is also crucial for bifacial cell structures where they are utilized as front surface field [6]. Therefore, optimization of the printing amount of Al and the firing temperature and understanding their effects on the other parameters of BSF are important to have a quality BSF layer.

On the other hand, it is well-known that simulation and calculations can be supportive to canalize the experimental studies. Several solar simulation programs including PC1D, Opal, Afors-Het, Silvaco atlas and Matlab software are used for the simulation of solar cells where some of them free of charge and some of them costly. Afors-Het is a one-dimensional numerical computer software used for modeling multilayer homo or heterojunction solar cells and some common methods for characterization of solar cells [7]. M. Schmidt *et al.* investigated the basic properties of amorphous/crystalline heterojunction (a-Si:H/c-Si), and their effect on recombination of excess carriers. Using the surface photovoltage analysis and Afors-Het simulation software, they achieved an optimal doping for an efficient a-SiH/c-Si solar cell without an i-type a-Si:H buffer layer and a efficiency of 19.8% for the coating temperature of 210°C [8]. In another study, different parameters of heterostructure thin films were optimized with Afors-Het to show that a high performance of up to 25% efficiency can be achieved by using nanostructured surfaces for HIT silicon solar cells [9]. Neeraj Dwivedi *et al.* optimized the thicknesses of n-type a-Si:H emitter layer, front a-Si:H i-layer and p-type c-Si base as well as optimum heterojunction (HJ) and HJ with intrinsic layer (HIT) solar cells using Afors-Het and reported an efficiency of 27.02% in the bifacial HIT solar cell. Optimized thicknesses of n-type a-Si:H emitter layer and front a-Si:H i-layer were introduced as 6 nm and 3 nm respectively with 200 µm base substrate thicknesses of c-Si [10]. Afors-Het is generally used in the simulation of heterojunction solar cells effectively and not much for conventional c-Si solar cells.

In this study, optimization of BSF layer for c-Si solar cells was carried out by Afors-Het software in order to determine the optimum thickness and doping concentration of the BSF [7]. A special attention was given to the optimization of the printing amount of Al paste and the related firing temperature according to the binary phase diagram of Al-Si to achieve the optimized BSF parameters obtained by Afors-Het simulations. The role of BSF phenomenon and the correlation of the facts from processing to the design was also discussed and analyzed for better understanding of BSF structure.

2. METHODS

In this section, the structure of a BSF and its effect on the performance of a solar cell were explained. Typical firing profile for Al fire through process given with related calculations. Initial parameters for simulation and the simulation flow were also explained.

2.1. Back Surface Field (BSF) Structure

Minimizing surface recombination losses is crucial for the development of a high efficiency solar cells [11]. The presence of a potential barrier on the back of a conventional silicon solar cell can significantly improve the performance of the cell by increasing mainly the open circuit voltage *(V$_{OC}$)* and short circuit current *(I$_{SC}$)* owing to the minimized recombination. In order to achieve

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

such effect behind the $n^+/p/p^+$ structure to provide a significant improvement in cell performance, the minority carrier diffusion length in the base region should exceed the base thickness [12].
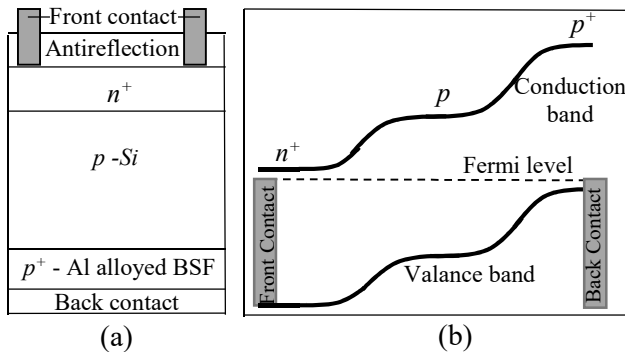
Figure 1. Schematic representations of a typical p-type c-Si solar cell with Al-alloyed BSF (a), Energy band diagram of BSF (b)

The schematic representation of the Al-alloyed BSF and the energy band diagram of the BSF are shown in the Figure 1. With the energy barrier formed between the p/p^+ junction, less of minority charge carriers formed at the base can be recombination on the rear surface. The BSF function acts like a mirror pushing carrier charges back into the cell [2]. Commonly, this effect of backside passivation is performed by alloying a screen-printed Al paste with silicon which establishes a moderate level of the BSF in commercial cell structures.

Effective back surface recombination velocity (SRV) which is a parameter to evaluate the passivation performance, can be in the range of 80-200 cm/s for boron BSF cells, while effective rear SRV of Al-BSF cells were recorded as 260-700 cm/s [13]. A back-surface recombination rate of 200 cm/s can be achieved by printing a thick layer of Al over 20 μm and firing at temperatures above 800°C for 1-5 minutes [11]. However, this process leads to a significant warpage of the wafer, which can lead to a problem with low mechanical yield when using thin substrates [14]. It is important to decrease the amount of Al in such cases which brings the necessity to optimize firing process. Therefore, it is rewarding to estimate the optimal firing temperature and effected parameters depending on the amount of printed Al paste. Furthermore, Ag is used for front

contacts and the alloying process of Ag/Si and Al/Si is commonly performed by a co-firing process. During the co-firing process, the wafers undergo a rapid treatment with a certain temperature profile similar to that of in Figure 2. This results in chemical and structural changes inside the printed pastes and substrate surfaces where Al pastes consists of Al particles with various diameter ranges, organic binders, solvents and glass frit [15]. During the firing process, Al-Si alloying takes place according to the Al-Si phase diagram [14]. An example of the temperature profile of a firing process was given in Figure 2, which can be summarized as: Solvents were removed during the drying of the screen-printed Al paste and a surface of silicon substrate with attachment of 50-70% by binders is achieved [16]. Binders burn out by increasing the temperature and alloying process of Al with silicon starts locally at around 660°C by diffusing of silicon through contacted regions. Penetration of Al into the silicon takes place with further increase of temperature and forms local spikes which expands horizontally. In addition to that molten Al-Si area acts as a sink for many contaminants and gives a perfect gettering effect. At peak temperature Al-Si liquid region is established. Al forms a eutectic alloy with silicon at a temperature of 577°C. Epitaxial recrystallization occurs by rejection of silicon during the cooling down and silicon is doped with Al at its solubility limit of Al at a given temperature, thereby producing a p^+- BSF layer [14, 16].
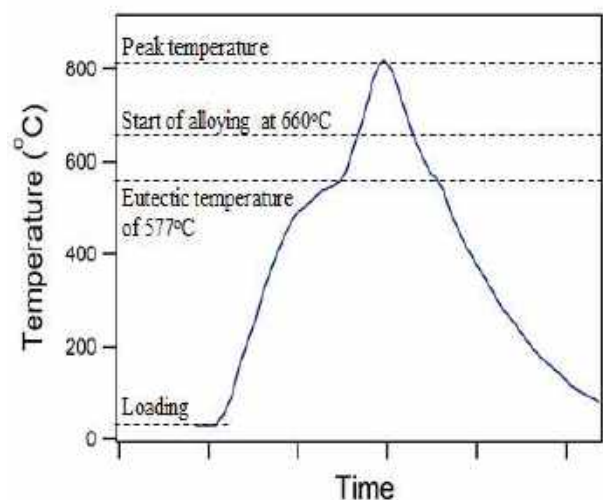
Figure 2. An example of temperature profile for the co-firing process.

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

Sufficient thickness of Al-BSF layer is necessary to obtain according to the Al-Si binary phase diagram where solubility of the silicon in the Al melt reaches 1.6 wt% at eutectic temperature of 577°C [17]. The concentration of silicon corresponding to the eutectic reaction has been reported in a variety of ranges such as 11.9 wt%, 12.2 wt%, 12.3 wt% and 12.6 wt% [18]. The BSF thickness can theoretically be calculated from the formulation of [3]:

$$W_{BSF} = \frac{m_{Si}}{A x \rho_{Si}}, \qquad (1)$$

$$m_{Si} = m_{Al}\left(\frac{F}{100-F} - \frac{E}{100-E}\right), \qquad (2)$$

where m_{Si} is the Si mass dissolved, m_{Al} is the mass of deposited Al, ρ_{Si} is the density of the silicon, A is the area of the sample, F and E are the percentage of Si mass in the liquid, at the firing temperature and at the eutectic, respectively [3].

2.2. Simulation

Optimization of the BSF layer which significantly improves the efficiency of c-Si solar cells was carried out using Afors-Het software. The influence of thickness and doping concentration on solar cell performance was specifically considered. Firstly, the effects of BSF thickness on solar cell efficiency was examined. Thickness of BSF was varied from 1 μm to 30 μm while other parameters kept constant. Table 1 shows the fixed values.

Table 1. Some initial parameter values for c-Si solar cell.

Emitter thickness (μm)	0.3
Emitter doping concentration (cm^{-3})	1×10^{20}
Reflectance (%)	8
Base Thickness (μm)	200
Base doping concentration (cm^{-3})	1.5×10^{16}
BSF doping concentration (cm^{-3})	1.0×10^{19}
Series resistance (Ω cm^2)	0.8
Parallel resistance (Ω cm^2)	10000

Secondly, the effect of BSF concentration on efficiency by varying the doping concentration from 1×10^{16} cm^{-3} to 5×10^{20} cm^{-3} was analyzed while keeping the BSF thickness constant at 7 μm.

Electrical characteristics of the cells including J_{SC}, V_{OC}, fill factor (FF) and efficiency values were calculated from I-V graphs. J_{SC}, V_{OC} and efficiency dependence on doping concentration were discussed separately. At last, theoretical BSF thicknesses depending on the peak firing temperature were calculated using Equation (1) and (2) in order to clarify the required temperatures for optimal BSF thickness.

3. RESULTS AND DISCUSSION

In this section, the results of optimization of the back-surface field parameters according to the simulations were given and discussed. Determining the optimal firing temperature for expected parameters including BSF thickness and the printed Al paste amount were explained as well.

3.1. Optimization of Back Surface Field Layer

Table 2 presents the dependence of the electrical parameters on the BSF thickness. I-V characteristics and current density dependence on BSF thickness can be seen in Figure 3. The J_{SC} of the cell increased with increasing BSF thickness from 31.95 mA/cm^2 to 32.11 mA/cm^2.

Table 2. Electrical characteristic dependence on BSF thickness.

BSF Thickness (μm)	Voc (mV)	Jsc (mA/cm^2)	FF (%)	Eff (%)
1	680.5	31.95	80.55	17.51
2	689.6	32.01	80.63	17.80
4	695.3	32.04	80.69	17.98
5	696.2	32.05	80.70	18.01
6	696.6	32.06	80.69	18.02
7	696.7	32.06	80.69	18.02
8	696.6	32.06	80.69	18.02
9	696.4	32.07	80.69	18.02
10	696.1	32.07	80.69	18.01
12	695.3	32.08	80.69	18.00
15	693.9	32.08	80.69	17.96
20	691.3	32.09	80.68	17.90
25	689.0	32.10	80.62	17.83
30	686.8	32.11	80.62	17.78

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

V_{OC} increased with increasing the BSF thickness and reached maximum value of 696.70 mV at BSF thickness of 7 μm. However, V_{OC} decreased when the BSF thickness was wider than 7 μm. Efficiency increased with increasing BSF thickness and remained constant at 18.02% when the thickness of BSF is between 6 μm and 9 μm where V_{OC} was slightly higher in case the BSF

thickness was 7 μm. These results for BSF thickness are quite consistent with experimental data. The experimental value of the BSF thickness was reported as 6.8 μm [3]. In another study, 50% Al coated contacts for best multi crystal and mono crystal cells were obtained local BSF thicknesses of 5 to 7 μm where FF values of 78% and 77% [19].
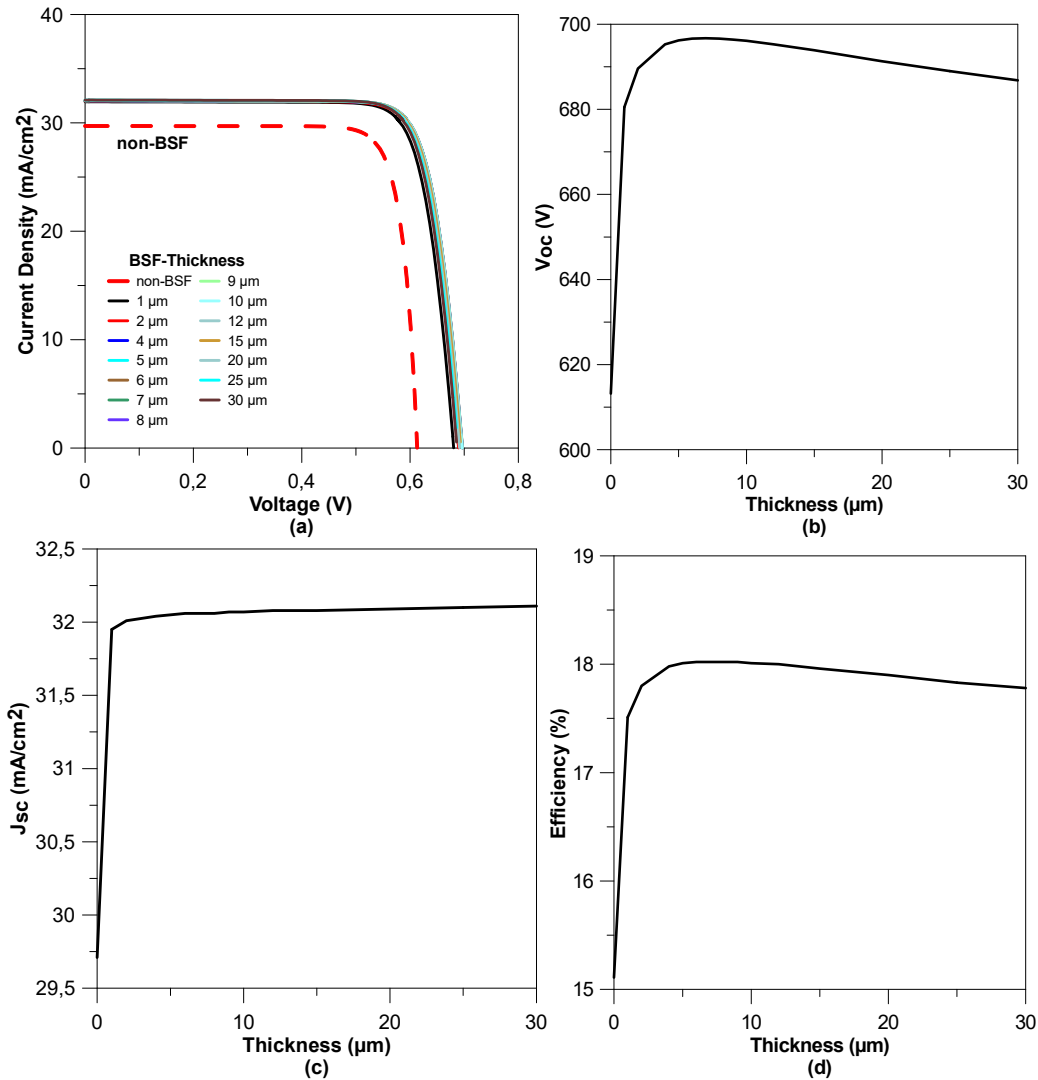
Figure 3. Characteristics of c-Si solar cells depending on BSF thickness

After these optimizations BSF thickness was fixed to 7 μm and doping concentration was varied from 1×10^{16} cm^{-3} to 5×10^{20} cm^{-3} in order to investigate the effect of the BSF doping concentration on the characteristics of the c-Si solar cell.

As shown in Table 3, an increase in V_{OC}, FF, and J_{SC} was observed as a result of increasing BSF doping concentration from 1×10^{16} cm^{-3} to 1×10^{19} cm^{-3} for c-Si solar cell. Figure 4 illustrates the electrical characteristics including $I\text{-}V$ and efficiency, V_{OC} and J_{SC} dependence on the doping concentration of the c-Si solar cell.

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

Table 3. Cell parameters varying with BSF doping concentration.

BSF Na (cm^{-3})	Voc (mV)	Jsc (mA/cm^2)	FF (%)	Eff (%)
1×10^{16}	613.8	29.76	79.32	14.49
5×10^{16}	616.6	30.00	79.37	14.68
1×10^{17}	620.6	30.29	79.39	14.92
5×10^{17}	644.3	31.39	79.89	16.16
1×10^{18}	659.5	31.73	80.18	16.78
5×10^{18}	689.9	32.03	80.62	17.82
1×10^{19}	696.7	32.06	80.69	18.02
5×10^{19}	685.4	32.06	80.69	17.99
1×10^{20}	688.1	32.02	80.63	17.77
5×10^{20}	661.5	31.76	80.23	16.86
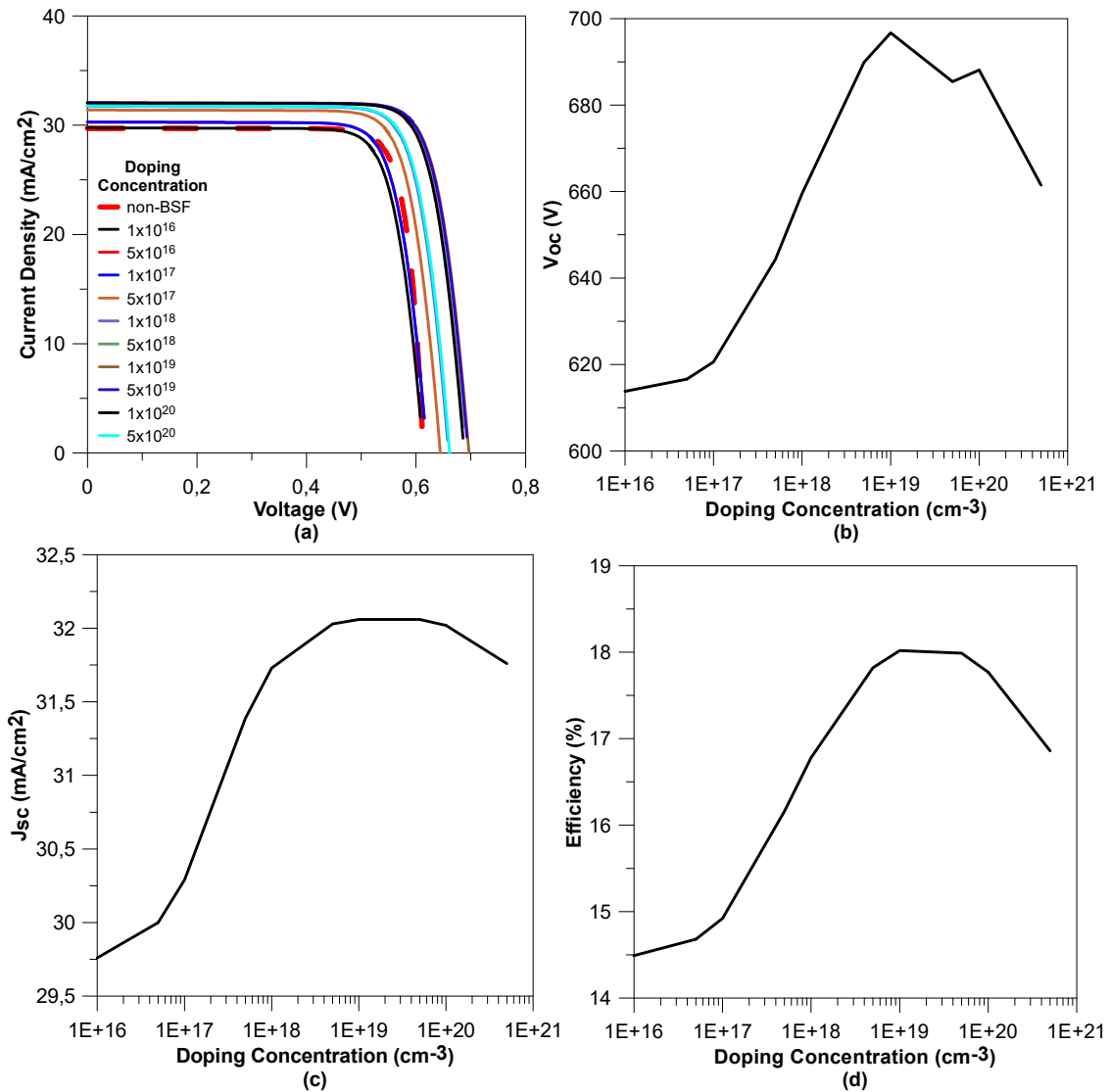
Figure 4. Electrical characteristics of c-Si solar cells depending on BSF doping concentration

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

The efficiency of the solar cell increased from 14.49% to 18.02%, respectively by increasing doping concentration of BSF from 1×10^{16} cm^{-3} to 1×10^{19} cm^{-3}. However, by increasing the doping concentration from 5×10^{19} cm^{-3} to 5×10^{20} cm^{-3} the efficiency decreased from 18.02% to down to 16.86%. The optimal doping concentration of BSF was 1×10^{19} cm^{-3} with efficiency of 18.02%. The optimum BSF doping concentration has been reported experimentally as around the similar values in 1×10^{19} cm^{-3} and 5×10^{19} cm^{-3} [20].

Table 4 summarizes the comparison of the electrical characteristics of the c-Si solar cell with optimized BSF layer to that of the solar cell without a BSF layer. The efficiency in the solar cell with optimized BSF is considerably higher than that of the cell without BSF. This clearly demonstrates the importance of BSF in c-Si solar cells. V_{OC} of the cell with optimized BSF was 696.7 mV which is significantly higher than that of the cell without BSF where V_{OC} was 613.2 mV. Efficiency of the cell with optimized BSF and the cell without BSF were resulted in 18.02% and 15.11%, respectively.

Table 4. Comparison of the solar cell parameters with optimized BSF to that of the cell without BSF.

BSF	Voc (mV)	Jsc (mA/cm²)	FF (%)	Eff (%)
Non-BSF	613.2	29.71	82.92	15.11
Optimized BSF	696.7	32.06	80.69	18.02

3.2. Determining the Optimal Firing Temperature

In addition to the optimization study of the BSF thickness and concentration using Afors-Het, BSF thickness dependence on the printed Al paste amount and the peak firing temperature was estimated using Equations (1) and (2) utilizing to the Al-Si phase diagram [17]. 12.6 wt% ratio of silicon in Al-Si alloy at eutectic temperature was used in calculations since it is mostly reported ratio in the literature [21-23]. Amount of Al was varied from 3 mg/cm^2 to 9 mg/cm^2 in the calculations of BSF thicknesses in order to determine the optimum peak firing temperature for each case.

Thickness of resulted BSF increases with either increasing the peak firing temperature or the amount of printed Al (Figure 5). In order to obtain the optimum BSF thickness of around 7 μm, the peak temperature should be adjusted by the amount of printed Al in consideration. For instance, temperature of up to 950ºC should be established if an amount of 3 mg/cm^2 printed Al was used to achieve BSF thickness of 7 μm, where 775ºC would be enough when the amount of Al is 8 mg/cm^2. Considering the average amount of printed Al is 6 or 7 mg/cm^2, a peak firing temperature of 825ºC and 800ºC need to be reached respectively, in order to form a 7 μm Al-BSF. Table 5 summarizes the calculated BSF thicknesses depending on the temperature and the amount printed Al paste.
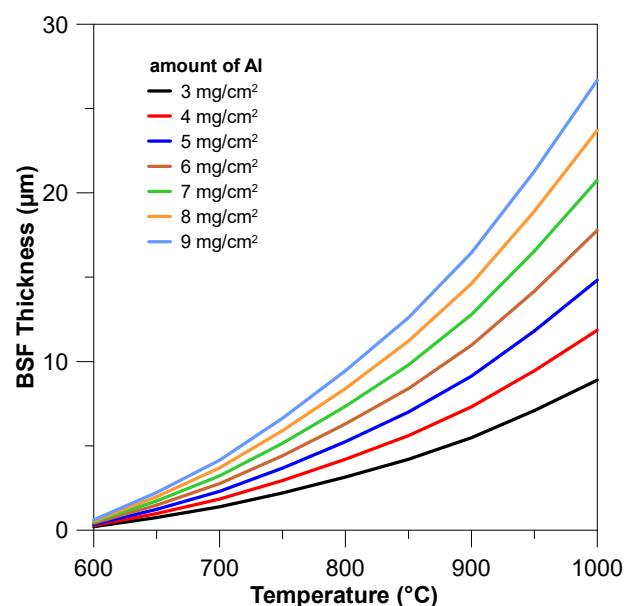
Figure 5. BSF thickness dependence to the peak firing temperature and the amount of Al.

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

Table 5. BSF thickness dependence on the peak firing temperature and the printed amount of Al paste.

		m_{Al} (mg/cm^2)							
		3	**4**	**5**	**6**	**7**	**8**	**9**	
	600	0.205	0.273	0.342	0.410	0.478	0.547	0.615	
	625	0.415	0.555	0.693	0.832	0.970	1.109	1.248	
	650	0.743	0.992	1.239	1.487	1.735	1.983	2.231	
	675	0.989	1.319	1.649	1.979	2.308	2.638	2.968	
	700	1.382	1.843	2.305	2.766	3.227	3.687	4.148	
	725	1.796	2.395	2.994	3.593	4.192	4.791	5.390	
	750	2.209	2.946	3.683	4.419	5.156	5.892	6.629	
	775	2.667	3.556	4.446	5.335	6.224	7.114	8.003	**BSF**
T (°C)	**800**	3.151	4.201	5.252	6.302	7.352	8.402	9.453	**Thickness**
	825	3.609	4.812	6.016	7.219	8.422	9.625	10.828	**(μm)**
	850	4.203	5.604	7.005	8.406	9.806	11.208	12.609	
	875	4.747	6.329	7.912	9.494	11.076	12.659	14.241	
	900	5.481	7.308	9.135	10.962	12.789	14.616	16.443	
	925	6.035	8.047	10.059	12.070	14.082	16.094	18.106	
	950	7.091	9.455	11.819	14.182	16.546	18.910	21.274	
	975	7.817	10.423	13.029	15.634	18.241	20.846	23.452	
	1000	8.893	11.857	14.822	17.786	20.750	23.715	26.679	

These optimization and results are supportive to realize the effect of BSF on the performance of c-Si solar cells. More importantly, these findings can be rewarding to understand the correlation between the printed Al paste and resulting BSF thickness which would be helpful for research of solar cells with thinner wafers.

4. CONCLUSIONS

In this article, the effects of the BSF layer on c-Si solar cells were investigated and optimized using Afors-Het simulation software. Photovoltaic parameters were determined depending on the BSF thickness of the c-Si solar cell. The maximum and close yield was obtained for the BSF thickness range of 6 μm to 8 μm. The optimum value of BSF doping concentration was in the order of 10^{19} cm^{-3}. The efficiency of c-Si solar cell without BSF (15.11%) could be improved with utilizing an optimum BSF up to 18.02%. Additionally, a special attention was given to the estimation of peak firing temperature depending on the printing amount of Al paste. It was concluded that the required peak firing temperature changes by the amount printed Al in order to form an optimal BSF. For instance, a peak firing temperature of 825°C need to be reached when the amount of printed Al is 6 mg/cm^2 to form a 7 μm Al-BSF. This study can be supportive to understand the importance of BSF structure for solar cells and to estimate the peak firing temperature depending on the amount of printed Al.

Research and Publication Ethics

Authors assure that this study was carried out by observing the international ethical rules of research and publication.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of interest

Authors declared no conflict of interest.

İmran KANMAZ, Abdullah ÜZÜM

Optimization of Back-Surface Field for Crystalline Silicon Solar Cells and Estimating the Firing Temp...

Author's contributions

IK: investigation, methodology, simulation, data analysis, writing-initial draft. AU: supervision, conceptualization, methodology, investigation, data analysis, writing-revision and finalizing

REFERENCES

[1] M. M. Desa *et al.*, "Silicon back contact solar cell configuration: A pathway towards higher efficiency," *Renewable and Sustainable Energy Reviews,* vol. 60, pp. 1516-1532, 2016.

[2] S. Tobbeche and M. N. Kateb, "Simulation and Optimization of Silicon Solar Cell Back Surface Field," *Materials Science,* vol. 21, no. 4, pp. 491-496, 2015.

[3] A. Kaminski *et al.*, "Aluminium BSF in silicon solar cells," *Solar Energy Materials and Solar Cells,* vol. 72, no. 1-4, pp. 373-379, 2002.

[4] M. C. Raval and S. M. Reddy, "Industrial Silicon Solar Cells," in *Solar Cells*: IntechOpen, 2019.

[5] N. Balaji, M. C. Raval, and S. Saravanan, "Review on Metallization in Crystalline Silicon Solar Cells," in *Solar Cells*: IntechOpen, 2019.

[6] H. Yin, K. Tang, J. Zhang, W. Shan, X. Huang, and X. Shen, "Bifacial n-type silicon solar cells with selective front surface field and rear emitter," *Solar Energy Materials and Solar Cells,* vol. 208, p. 110345, 2020.

[7] R. Varache, C. Leendertz, M. Gueunier-Farret, J. Haschke, D. Muñoz, and L. Korte, "Investigation of selective junctions using a newly developed tunnel current model for solar cell applications," *Solar Energy Materials and Solar Cells,* vol. 141, pp. 14-23, 2015.

[8] M. Schmidt *et al.*, "Physical aspects of a-Si: H/c-Si hetero-junction solar cells," *Thin Solid Films,* vol. 515, no. 19, pp. 7475-7480, 2007.

[9] M. H. Vishkasougheh and B. Tunaboylu, "Simulation of high efficiency silicon solar cells with a hetero-junction microcrystalline intrinsic thin layer," *Energy conversion and management,* vol. 72, pp. 141-146, 2013.

[10] N. Dwivedi, S. Kumar, A. Bisht, K. Patel, and S. Sudhakar, "Simulation approach for optimization of device structure and thickness of HIT solar cells to achieve~ 27% efficiency," *Solar energy,* vol. 88, pp. 31-41, 2013.

[11] T. Lauinger, J. Schmidt, A. G. Aberle, and R. Hezel, "Record low surface recombination velocities on 1 Ω cm p-silicon using remote plasma silicon nitride passivation," *Applied physics letters,* vol. 68, no. 9, pp. 1232-1234, 1996.

[12] J. G. Fossum, "Physical operation of back-surface-field silicon solar cells," *IEEE Transactions on Electron Devices,* vol. 24, no. 4, pp. 322-325, 1977.

[13] S. Riegel, S. Gloger, B. Raabe, and G. Hahn, "Comparison of the passivation quality of boron and aluminum BSF for wafers of varying thickness," in *24th European Photovoltaic Solar Energy Conference*, 2009, pp. 1596-1599.

[14] J. Szlufcik, S. Sivoththaman, J. Nlis, R. P. Mertens, and R. Van Overstraeten, "Low-cost industrial technologies of crystalline silicon solar cells," *Proceedings of the IEEE,* vol. 85, no. 5, pp. 711-730, 1997.

[15] L. J. Caballero, "Contact definition in industrial silicon solar cells," *Solar Energy,* pp. 375-398, 2010.

[16] F. Huster, "Aluminum-back surface field: bow investigation and elimination," in *20th European Photovoltaic Solar Energy*

Conference and Exhibition, Barcelona, pp. 635-638, 2005.

[17] J. Murray and A. McAlister, "The Al-Si (aluminum-silicon) system," *Bulletin of alloy phase diagrams,* vol. 5, no. 1, p. 74, 1984.

[18] M. B. Djurdjević, S. Manasijević, Z. Odanović, and N. Dolić, "Calculation of liquidus temperature for aluminum and magnesium alloys applying method of equivalency," *Advances in Materials Science and Engineering,* vol. 2013, 2013.

[19] I. Cesar *et al.*, "Industrial application of uncapped Al 2 O 3 and firing-through Al-BSF in open rear passivated solar cells," in *2011 37th IEEE Photovoltaic Specialists Conference*, pp. 001405-001410: IEEE, 2011.

[20] J. Eguren, J. Del Alamo, and A. Luque, "Optimisation of p+ doping level of n+-p-p+ bifacial bsf solar cells by ion implantation," *Electronics Letters,* vol. 16, no. 16, pp. 633-634, 1980.

[21] M. Barbes, M. Quintana, L. Verdeja, and R. Gonzalez, "Microstructures of a pressure die cast Al-8.5% Si-3.5% Cu alloy," *Kovove Mater,* vol. 55, pp. 89-96, 2017.

[22] A. Sharma and J. P. Jung, "Possibility of Al-Si Brazing Alloys for Industrial Microjoining Applications," *Journal of the Microelectronics and Packaging Society,* vol. 24, no. 3, pp. 35-40, 2017.

[23] M. Haghshenas and J. Jamali, "Assessment of circumferential cracks in hypereutectic Al-Si clutch housings," *Case Studies in Engineering Failure Analysis,* vol. 8, 12/01 2016.

JOURNAL OF SCIENCE

SAKARYA UNIVERSITY

Sakarya University Journal of Science

Title: On the Design of 2D Dynamic Drawings with Euklid DynaGeo

Authors: Engin CAN

On the Design of 2D Dynamic Drawings with Euklid DynaGeo

Engin CAN[*1]

Abstract

Nowadays in the plane geometry with the advancement of technology, the geometric constructions are designed and processed using the 2D dynamic geometry software's (DGSs). In order to use such programs, users should have the necessary information and skill. In this paper, pieces of information some of the known software are given, but more proposed Euklid DynaGeo is presented, in which users can draw and recognize the function graphs more easily. For this aim, some applications are displayed and finally, clearly presented how Euklid DynaGeo serves the users to realize the constructional geometrical drawings.

Keywords: dynamic geometry software, Euklid DynaGeo, plane geometry, 2D construction, mathematics education

1. INTRODUCTION

Sometimes geometric definitions and theorems are especially for students very abstract and difficult to understand. Because of this, interest in geometry decreases and leads to poor performance. Using computers as supporting materials on education by increasing opportunities for technology could be made abstract theories more understandable. The researchers have discovered that the results of success are quite effective in using dynamic geometry software [1-8]. Therefore, some lectures are with computer applications supported, which continuously user-friendly developed and designed. So, scientific concepts could be taught and learned by computer-aided instruction (CAI) [9]. The introduction of DGS in geometry teaching requires that instructors have

technical knowledge in dealing with DGS and in the conceptual understanding of mathematics. Thus, also teachers run into a number of new challenges [10-12].

Just before the 20th century, such DGS first by Geometric Supposer and Cabri, later on by The Geometer's Sketchpad was developed. Overtime many software programs by the advancement of computer hardware and software have been developed with the cooperation of good programmers and academics, who have knowledge of geometry. The main types of computer environments for studying school geometry are supposers, dynamic geometry environments (DGEs) and Logo-based programs [13]. Especially with drag feature by DGEs provides users to manage the geometric object into different positions. There are many free- and

---
[*] Corresponding Author: ecan@subu.edu.tr
[1] Sakarya University of Applied Sciences, Faculty of Technology, Sakarya, Turkey.
ORCID: https://orcid.org/0000-0002-4105-6460

shareware 2D DGSs frequently used today. Some of them are Cabri, Cinderella, GeoGebra, CaRMetal, GEX and DrGeo.

The rest of this paper is created as follows: In the next section, the proposed software is presented. In section 3, three application examples are demonstrated and pointed up, if a construction displayed as an application (or proof of a geometric theorem in 2D) then with the drag option could be shown that for all other positions of the construction is valid.

2. EUKLID DYNAGEO IN BRIEF

Euklid DynaGeo is a computer program about "dynamic geometry" that gives us the possibility to design drawings i.e. drawings in which (some) objects can be dragged to a new position without losing the mutual of the geometrical objects during the preceding construction. It is useful to create various geometric objects, constructions, even mappings, measure distances and angle widths, calculate and watch expressions, etc.

Moreover, it is possible creating trace lines, working with conics and affine mappings, creating macros for repetitive commands of constructions, animating drawings to demonstrate transformations and more [14].

The toolbar configurations and icons of Euklid DynaGeo look like as follows:

1. Main bar:

2. Construct:

3. Mappings:

4. Curves:

5. Shape & Colour:

6. Measure & Calculate:

In the next section, three examples are demonstrated how the software toolbars are used in practice.

3. SOME APPLICATIONS

3. 1. Desargues's Theorem

It is well known that the Desargues theorem in projective geometry is:

Two triangles are in perspective centrally if and only if they are in perspective axially.

Let the above triangles are ABC and $A^1B^1C^1$, and they are perspective by the central M. It must be shown that

$$D = AB \cap A^1B^1,$$
$$E = BC \cap B^1C^1,$$
$$F = AC \cap A^1C^1$$

are collinear. The proof is displayed in Figure 1.
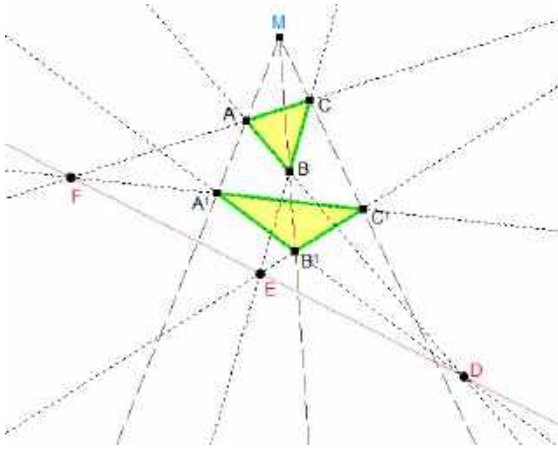
Figure 1. Displayed Desargues Theorem

All the construction steps are in the bookmark with "View >> Show construction text". This command allows users to see, how the construction has been designed, as follows:

To show that geometric construction is always valid, two dragged positions of *Desargues Theorem* are displayed in Figure 2.
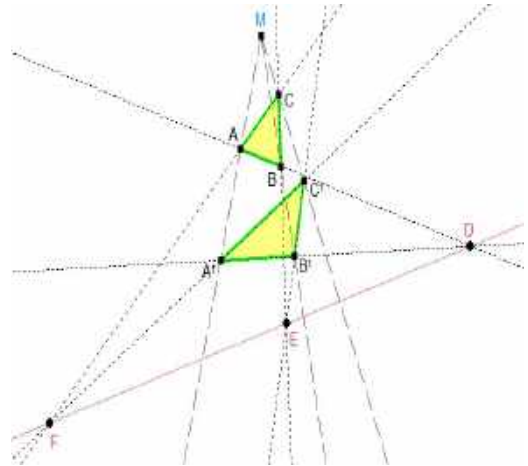
(a)
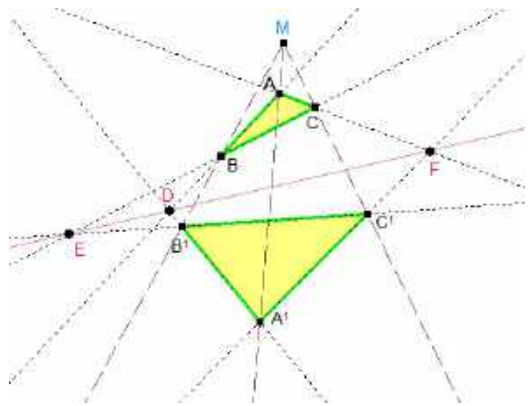
(b)

Figure 2. (a), (b), using the drag option for two different positions.

3.2. Demonstration of Similarity Ratios of a Triangle

Let g and h are two parallel lines, and choose a point A that is not on these parallel lines. We draw two different straight lines from A afterward. Let the intersection points with g and h be D, B, E, C, respectively. Finally, there is a similarity ratio between ADE and ABC triangles. Therefore,

$$\frac{|AD|}{|AB|} = \frac{|AE|}{|AC|} = \frac{|DE|}{|BC|} = \frac{|DF|}{|BG|} = \frac{|FE|}{|GC|} = \frac{|AF|}{|FG|}.$$

In Figure 3, this similarity is displayed and all similarity ratios are given.

(a)
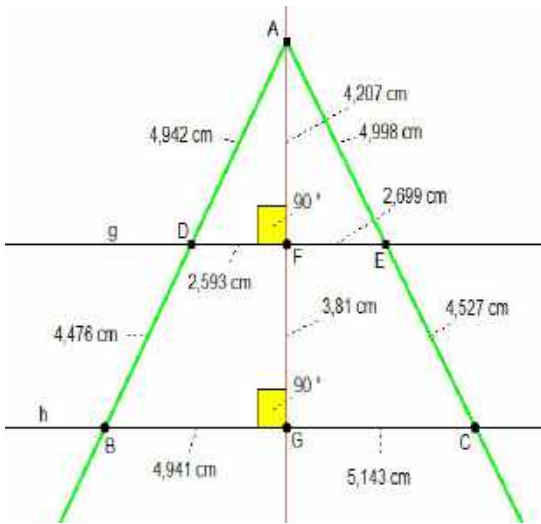
d(A;D)/d(A;B)	d(A;E)/d(A;C)
0,52	0,52
d(A;F)/d(A;G)	d(D;F)/d(B;G)
0,52	0,52
d(F;E)/d(G;C)	d(D;E)/d(B;C)
0,52	0,52

(b)

Figure 3. Displayed similarity of a triangle (a)
and the related similarity ratios (b).

(a)
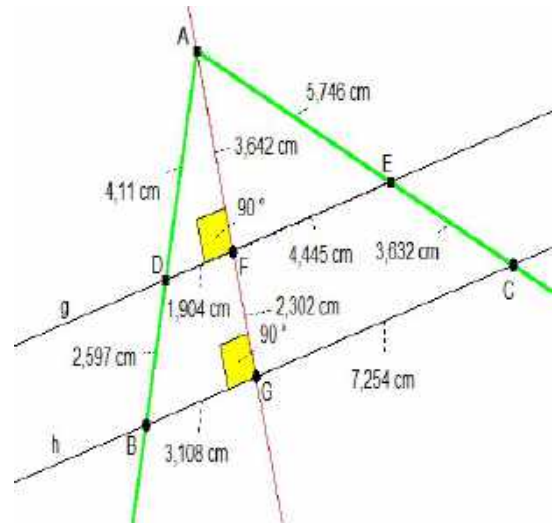
d(A;D)/d(A;B)	d(A;E)/d(A;C)
0,61	0,61
d(A;F)/d(A;G)	d(D;F)/d(B;G)
0,61	0,61
d(F;E)/d(G;C)	d(D;E)/d(B;C)
0,61	0,61

(b)

Figure 4. Using the drag option for a different
position (a) and the related similarity ratios (b).

The similarity construction steps are as follows:


```
Construction text                    —  □  ×
D is a free basis point
E is a free basis point
g is the line ( D ; E )
( P₃ is a free basis point )
h is the parallel to g through P₃
A is a free basis point
h₁ is a ray starting in point A and running through point E
a is a ray starting in point A and running through point D
B is the intersection point of the lines h and a
C is the intersection point of the lines h and h₁
g₃ is the perpendicular of A on h
F is the intersection point of the lines g₃ and g
G is the intersection point of the lines g₃ and h
α₁ is the angle ( A ; F ; D )
α₂ is the angle ( A ; G ; B )
T₄ is the expression d(D;F)/d(B;G)
T₅ is the expression d(F;E)/d(G;C)
T₁ is the expression d(A;D)/d(A;B)
T₂ is the expression d(A;E)/d(A;C)
T₃ is the expression d(A;F)/d(A;G)
T₆ is the expression d(D;E)/d(B;C)
```
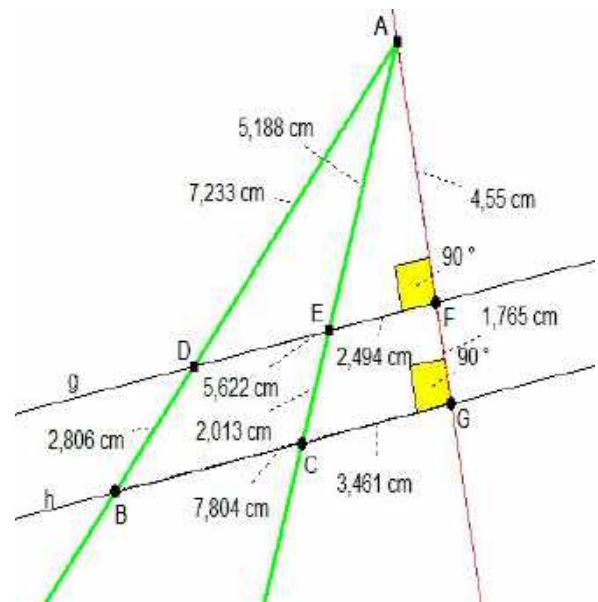

(a)

To show that geometric construction is always
valid, two dragged positions of Similarity Ratios
are displayed in Figure 4. and Figure 5.

d(A;D)/d(A;B)	d(A;E)/d(A;C)
0,72	0,72
d(A;F)/d(A;G)	d(D;F)/d(B;G)
0,72	0,72
d(F;E)/d(G;C)	d(D;E)/d(B;C)
0,72	0,72

(b)

Figure 5. Using the drag option for a another position (a) and the related similarity ratios (b).

3.3. Demonstration of Trisectrix

A straight line rotates around O and P, whereby the angle at P is three times as large as the angle at O. The path curve of the intersection S of the two straight lines is the *trisectrix*, which demonstrated in Figure 6.
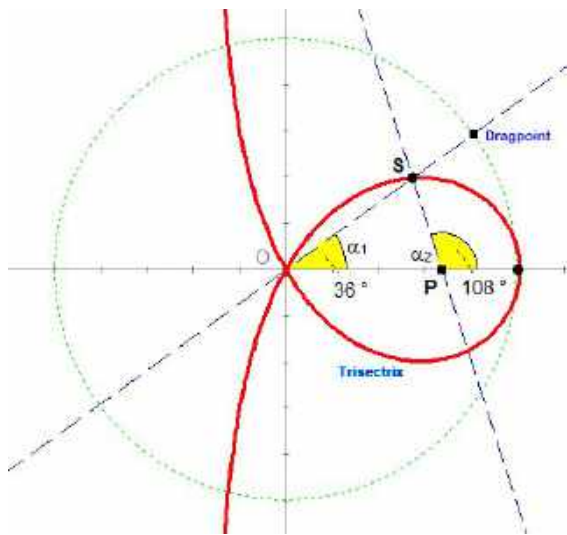
Figure 6. Trisectrix as a path curve.

The trisectrix construction steps are as follows:

Two dragged positions of trisectrix are displayed in Figure 7.
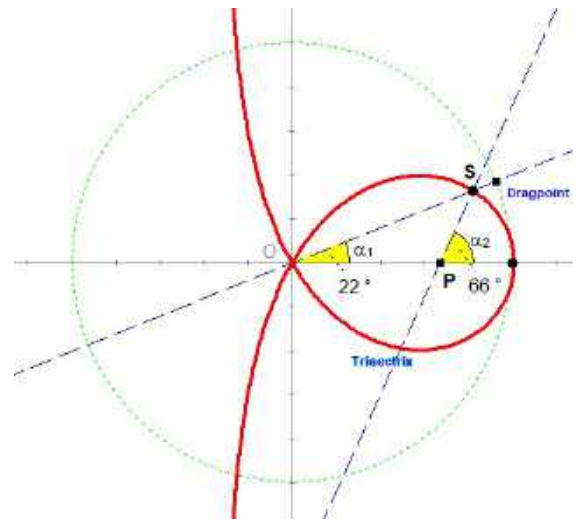
(a)
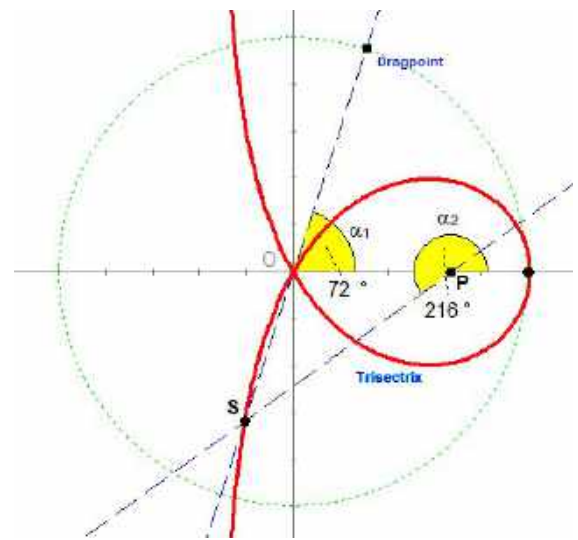
(b)

Figure 7. (a), (b), using the drag option for two different positions of trisectrix.

4. CONCLUSIONS

In this paper, how Euklid DynaGeo serves the users to realize the constructional geometrical drawings is demonstrated.

Moreover, it is clear computer users is getting increased. That means in the future, more computer programs have to use especially in lectures. The proposed software is one of them, which provides users an application solution for

the 2D geometric drawings for more creative constructions in plane geometry. Finally, if a construction displayed as proof then with the drag option could be shown that for all positions the construction is valid, which offers the users the possibility to try different positions.

Acknowledgements

Research and Publication Ethics-Ethics Committee Approval

Since this study involved a desk review, the author asserts that all procedures contributing to this study comply with the ethical standards of the relevant institutional committees. For this type of study, formal consent is not required.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

REFERENCES

[1] Cantürk-Günhan, B., & Açan, H. "The effect of using dynamic geometry software on success of geometry: A meta-analysis study", Turkish Journal of Computer and Mathematics Education, 7(1), pp. 1-23, 2016.

[2] Uğur, B., Urhan, S., Akgün Kocadere, S.,"Geometrik Cisimler Konusunun Dinamik Geometri Yazılımı ile Öğretimi", Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi, 10, S. 339-366, 2016.

[3] Baki, A., Özpınar, İ., "Logo destekli geometri öğretimi materyalinin öğrencilerin akademik başarılarına etkileri ve öğrencilerin uygulama ile ilgili görüşleri", Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 34(3), S.153-164, 2007.

[4] Demirbilek, M, Özkale, A., "GeoGebra Kullanımının Önlisans Matematik Öğretimine Etkinliğinin İncelenmesi", Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi 8, S. 98-123, 2014.

[5] Delice, A.& Karaaslan, G., "Dinamik geometri yazılımları ile çokgenler konusunda hazırlanan etkinliklerin öğrenci performansı ve öğretmen görüşlerine yansıması", Karaelmas Eğitim Bilimleri Dergisi, 3, S. 133-148, 2015.

[6] Akyar, K. B. "Öklid geometrisi öğretiminde dinamik geometri yazılımları kullanımının 11.sınıf öğrencilerinin geometriye yönelik tutumlarına ve akademik başarılarına etkisi", Yayınlanmamış Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, 2010.

[7] Aydoğan, A., "The effect of dynamic geometry use together with open-ended explorations in sixth grade students' performances in polygons and similarity and congruency of polygons", Yayınlanmamış Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2007.

[8] Orçanlı H. B., Orçanlı K., "Bilgisayar destekli geometri öğretimi hakkında öğrenci görüşlerinin değerlendirilmesi", Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi, 25(3), S. 55-70, 2016.

[9] Chang, C.Y, "Comparing the Impacts of a Problem-Based Computer-Assisted Instruction and the Direct-Interactive Teaching Method on Student Science Achievement", Journal of Science Educational and Technology, 10(2), pp. 147-153, 2001.

[10] A. B. Anthony & L. M. Clark, "Examining Dilemmas of Practice Associated With the Integration of Technology Into Mathematics Classrooms Serving Urban Students", Urban Education, 46(6), pp. 1300-1331, 2011.

[11] C. Laborde, "Technology as an instrument for teachers", Paper presented at the Proceedings of Working Group 4: Resources and Technology throughout the History of ICMI, Rome, Italy, 2008.

[12] Yılmaz G.K., "The Effect of Dynamic Geometry Software and Physical Manipulatives on Candidate Teachers Transformational Geometry Success", Educational Sciences: Theory & Practice, 15(5), pp. 1417-1435, 2015.

[13] M.T. Battista, "The Development of Geometric and Spatial Thinking", In Lester, Jr., F.K. (ed.). Second Handbook of Research on Mathematics Teaching and Learning. Charlotte, NC: Information Age and the National Council of Teachers of Mathematics. pp. 843–903, 2007.

[14] http://www.dynageo.com/

Title: Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of
Piezoelectric Patch Glued Tapered Beams

Authors: Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tapered Beams

Fevzi Çakmak BOLAT[*1]

Abstract

In this study, the effects of taper ratio and boundary conditions on the energy harvest performance of a beam element were examined. For these purpose, different taper ratio beams were analyzed numerically. The energy harvesting process was achieved by gluing a piezoelectric patch onto the cantilever tapered beam. Different taper ratio beams were designed and the effects of these taper ratio on stress change were investigated. In piezoelectric materials, when mechanical stress or strain is applied to the material, they generate electrical potential energy as a response. In order to increase the stresses on the tapered beam, the boundary condition was applied to be the thin edge of the tapered beam element in this study. In this way, the effect of tip mass was created and it was aimed to increase the stress magnitude due to vibration on the beam. Stress changes and displacement magnitudes of beams examined by applying load on beams having different taper ratio. The effect of these alterations on energy harvest efficiency was analyzed.

Keywords: tapered beam, piezoelectiric material, energy harvesting, vibration-induced stress.

1. INTRODUCTION

With the development of technology, sensor sizes are getting smaller each day. These advanced technology products require less energy. This energy need can also be provided with alternative renewable resources. In recent years [1-2], studies on small-scale energy harvesting studies are increasing. There are many studies in the literature where piezoelectric, electromagnetic or both are used together [3-8]. Voltage is generated when force or stress is applied to a piezoelectric material. Another energy harvesting structure is electromechanical systems. Energy is harvested from these types of structures by generating voltage by moving of a structure in which the magnet and the coil interact. By giving air flow over an aerodynamic profile, vibration can be created on the energy harvesting structure. By using these vibrational movements, mechanical energy is converted into electrical energy with using smart materials. This research study is

---
[*] Corresponding Author: fcakmakbolat@gmail.com
[1] Bolu Abant Izzet Baysal University, Department of Mechanical Engineering, 14280, Bolu, Turkey
   ORCID: https://orcid.org/0000-0003-1532-7631

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...

investigated a novel high-performance piezoelectric wind energy harvester to improve wind energy harvesting performance with Y-shaped attachments on the bluff body. A theoretical model was developed for the analysis of the energy harvesting performance of the proposed structure. Results were experimentally examined and compared with Y-geometry and without Y geometry [9].

Another one work proposes a novel geometric structure to energy harvesting from beam based harvester to gain power in low-frequency range. A gap was created on the tapered beam and the effect of the results on the energy harvest was examined. A parametric study was carried out to ensure an equal distribution of the stress distribution on the piezoelectric patch using a package program that makes numerical solutions. By conducting experimental studies, the effect of variable geometry on energy harvesting was investigated in reference [10]. In this study, energy harvesting from piezoelectric material adhered to a beam surface with tapered geometry was investigated. The result of the study that the amount of energy obtained from the beam with tapered geometry is higher than that of a normal beam [11]. This study aimed to increase the amount of energy harvest obtained by using ambient vibrations. For this purpose, the energy harvesting performance was examined by sticking piezoelectric material on the double-sided tapered beam. The energy harvesting structure was created with three different tapered beam designs, and the results obtained with the analytical model were experimentally confirmed [12]. In this research paper, different types of non-uniform based piezoelectric energy harvesters structures are analysed its enerjy harvester performance. Created a rectangular cavities are introduced into the proposed harvesters to enhance the harvester voltage. The suggested harvesters are modelled analytically and are validated with experimental results [13]. This research group are investigated the possibility of using tapered beams in piezoelectric energy harvesters. Numerical simulations did not suggest any increase in the generated output power and the lack of improvement was confirmed in practice. With the help of the numerical simulations it was further

found that the tapering does work but only for certain design configurations, namely for cantilevers with long slender beams. For cantilevers with short wide beams, the tapering has no significant effect on the output power of the harvester [14]. Zang et al. analyzed of the sea wave-based energy harvesting structure. In this study energy harvesting performance was investigated using tapered beam structure glued with piezoelectric material. For this purpose, a mathematical model of the harvester is established based on Airy linear wave theory and Bessel equations to calculate the energy harvesting magnitude [15].

In this study, different types of beams examined numerically under dynamic harmonic load in order to obtain vibration-based energy harvest. In the literature, taper beam beams are used by being supported by their thick edges. In this study, in order to create a type mass effect and increase the stress value on the piezoelectric material, fixed edge boundary condition is applied different from literature. For this aim the energy harvester beam structure was fixed by the thin edge and the amount of energy harvest to be obtained from harvester structures increased.

2. DESIGN AND ANALYSIS TAPERED BEAM

Tapered beam structure is illustrated in Figure 1. Equations of motion the beam structure is given here by using Hamilton approach.
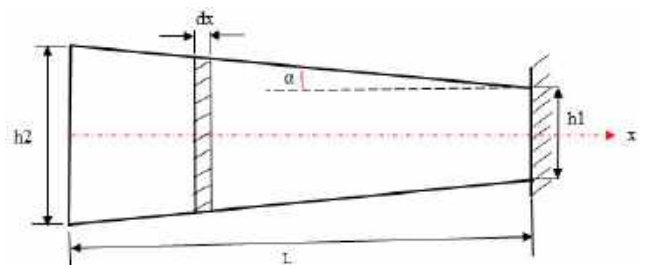
Figure 1. Tapered beam structure

$$\rho A(x)\frac{\partial^2 w(x,t)}{\partial t^2} + EI(x)\frac{\partial^2 w(x,t)}{\partial x^2} + l^2 \mu A(x)\frac{\partial^4 w(x,t)}{\partial x^4} = f(x,t) \quad (1)$$

Here $f(x,t)$ external force, μ shear modulus, $I(x)$ moment of inertia, $A(x)$ represent the area

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...

respectively. In this study, the change of both single surface and two surface taper ratios investigated. The figure below shows three different structures to be examined within the scope of this study. Here illustrates Figure 2-a represent single surface tapered beam, Figure 2-b double surface tapered beam and Figure 2-c equivalent tip mass beam. Total mass of the equivalent tip mass beam was selected to equal the single tapered beam total mass.
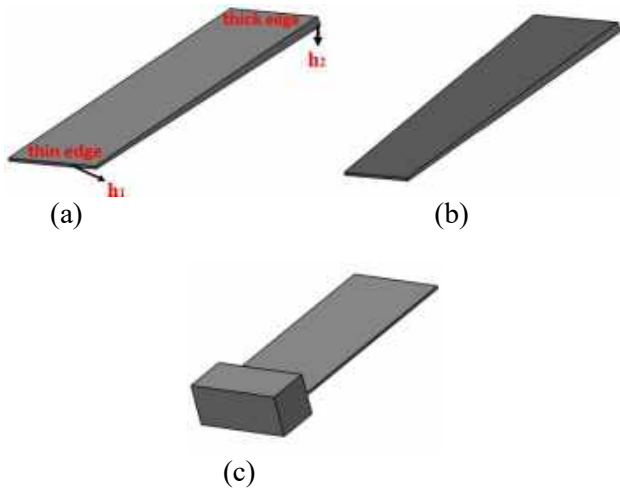
Figure 2. Taper beams a) one surface taper, b) two surface taper, c) equivalent mass beam

2.1. Nonlinear Numerical Analysis

A dynamic load is applied to the designed beam in ANSYS software to examine the under-load behavior of each proposed beam structures. As seen in Figure 3, to start the free vibration for the wing structure the dynamic harmonic load.
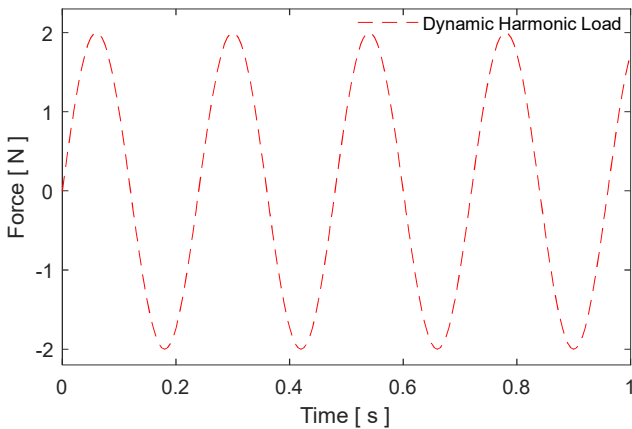
Figure 3. Dynamic harmonic load

2.2. Determining the Boundary Condition

In order to determine in which case the maximum stress occurs on the taper beam, the beam was fixed by two different boundary conditions and the results was examined. Figure 4 and Figure 5 are shows the importance of the boundary condition. The analysis results obtained show that the stress and displacement values obtained by applying the boundary condition from the thin edge are larger than the thick edge.
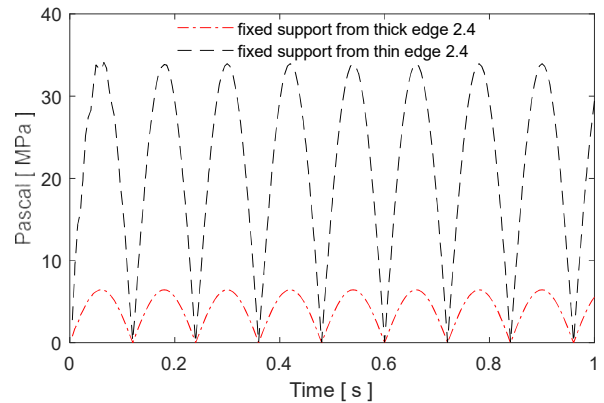
Figure 4. Stress results under the dynamic harmonic load
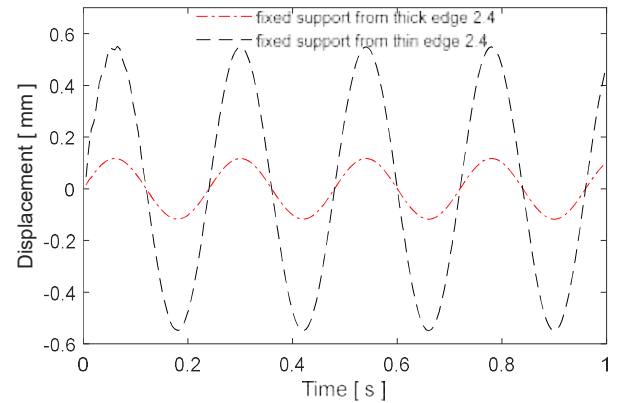
Figure 5. Deformation results under the dynamic harmonic load

2.3. Analysis for Different Types of Beam Structures

After determining the beam types of different types and thicknesses, the external force given in Figure 3 was applied to each beam structure and the effects of the results on the energy harvest were examined in detail.

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...
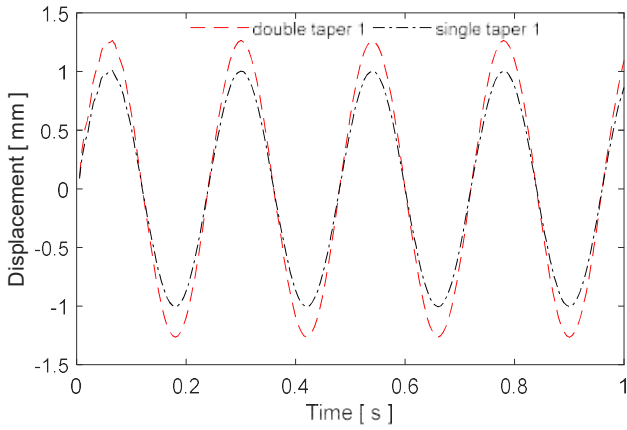
Figure.6 Single and double tapered beam displacement results
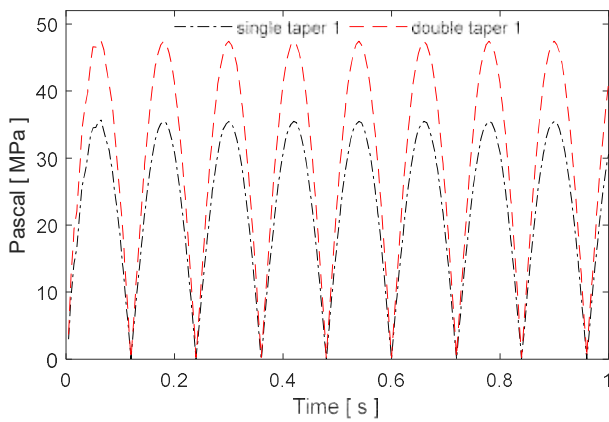
Figure.7 Single and double tapered beam stress results

Figure 6 and Figure 7 also show time-dependent displacement and stress changes of single-surface and double-surface tapered beams. The results obtained show that beside the proposed beam structure, the boundary condition is also very important on the energy harvest to be obtained.
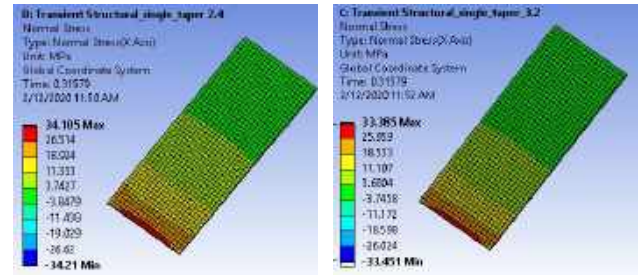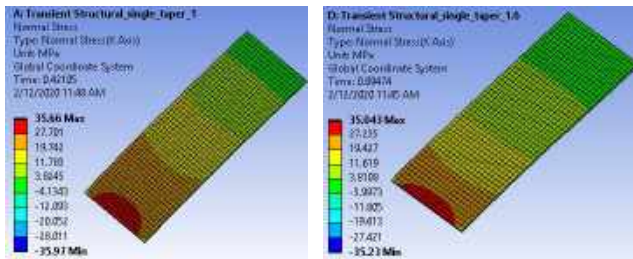
Figure 8. Tapered cases and stress changes of one surface numeric results

Table.1 One surface taper numeric result

Tapered Cases	Stress (MPa) One surface taper
$h_1 = h_2 = 1$ mm	35.66
$h_1 = 1$ and $h_2 = 1.6$ mm	35.043
$h_1 = 1$ and $h_2 = 2.4$ mm	34.105
$h_1 = 1$ and $h_2 = 3.2$ mm	33.385

As seen in Figure 8 and Table 1, maximum stress values decrease in single surface taper beams as the rate of taper increases. The stress values formed on the beam as a result of the applied dynamic load are close to each other, but are reduced. This shows that it is not very important to increase the beam thickness to increase the amount of energy. It also shows the same results in the double-sided tapered beam as seen in the results in Figure 9 and Table 2.
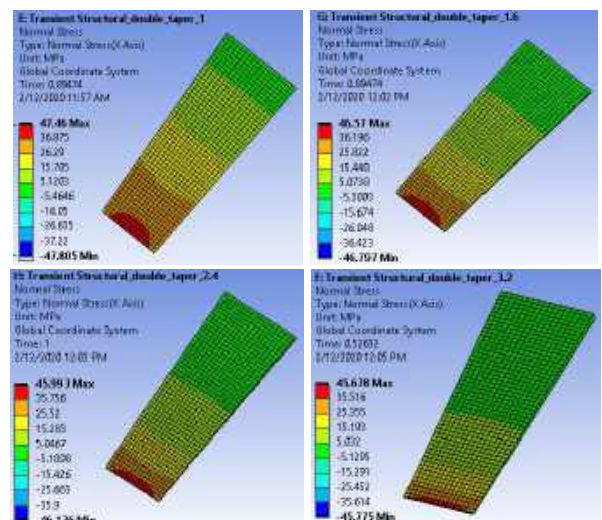
Figure 9. Tapered cases and stress changes of two surface numeric results

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...

Table.2 Two surface taper numeric results.

Tapered cases	Stress (MPa) Two surface taper
h1=h2=1 mm	47.46
h1=1 and h2=1.6 mm	46.57
h1=1 and h2=2.4 mm	45.993
h1=1 and h2=3.2 mm	45.678

The graphics in Figure 10-13 were obtained by applying the harmonic load as illustrated in Figure 3. Here, time-dependent displacement and stress results of both single tapered beams and double-sided tapered beams were obtained for different thicknesses structures.

Figure 10. Different single tapered beam displacement results

Figure 11. Different single tapered beam stress results

Figure 12. Different double tapered displacement results

Figure 13. Different single tapered beam stress results

Numerical analysis is performed for three different beams given in Figure 2 and the results are given in Figure 14 and Figure 15. As it can be seen in the graphic results, when the external disturbance force input is the same for all configurations, the most stress and deformation occurs in the double-sided tapered beam. Apart from this, a mass is added to the end point of the beam in the same weight as the single taper beam. Numerical analysis results showed beam element with the added of end mass is have more stress and displacement than the one-sided tapered beam. The meaning of this analysis result energy amount magnitude will directly affect from structure types.

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...
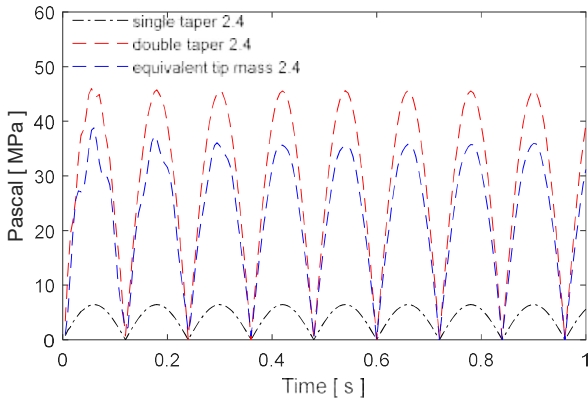
Figure 14. Different type beam stress results
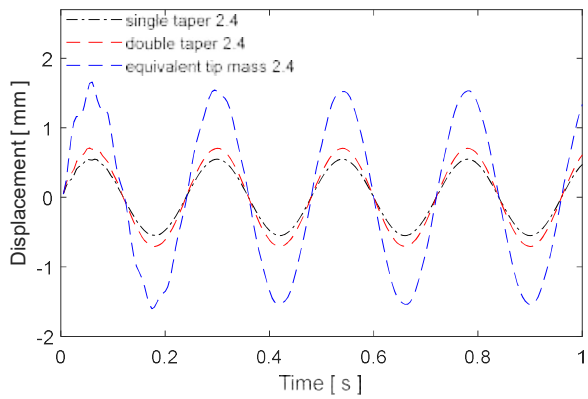
Figure 15. Different type beam displacement results

3. ENERGY HARVESTING STRUCTURES

To realize energy harvesting from piezoelectric material, it is necessary to change the shape of the piezoelectric material by giving a mechanical input. Energy harvesting from piezoelectric material can be realized against this given mechanical input. In Figure 16, a piezoelectric patch is attached to a beam element to create an energy harvesting structure. The results obtained from the previous analyzes, the beam is fixed by the thin edge to increase the amount of energy harvest. This boundary condition apllied to the other selected type beams.
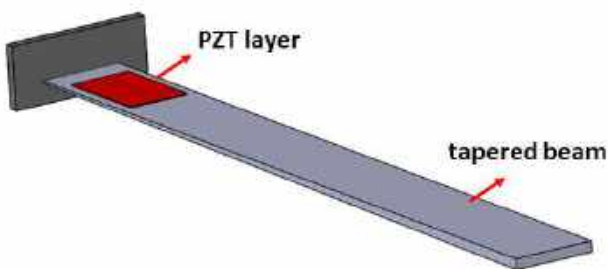
Figure 16. Energy harvester structure

Depending on the vibration of the beam element, a stress will occur on the piezoelectric patch and this mechanical energy will be converted into electrical energy.

$$\begin{bmatrix} S \\ D \end{bmatrix} = \begin{bmatrix} s^E & d^t \\ d & \varepsilon^T \end{bmatrix} \begin{bmatrix} T \\ E \end{bmatrix} \tag{2}$$

T is the stress, S^E is the strain D charge density E electric field, ε^T is the dielectric permittivity, d is the piezoelectric coefficient. For PZT, BaTiO3, PbTiO3 piezoelectric type materials these matrixs is opened as follows [16],

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & d_{31} \\ 0 & 0 & d_{31} \\ 0 & 0 & d_{33} \\ 0 & d_{15} & 0 \\ d_{15} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{4}$$

Converted electrical energy from a piezoelectric matarial calculated as follows equation [17]

$$U = QV = \frac{1}{2}(d\sigma S)(g\sigma t) = \frac{1}{2}dg\sigma^2 Volume \tag{5}$$

where surface area S and thickness t, the current constant d voltage constant g correspond to the specific co-efficients of the operational mode, σ stresss, V voltage over the piezoelectric element. In order to calculate the amount of energy harvest to be obtained from this proposed structure, the stress values obtained in section 2 are substituting in the equation (4). Properties of the piezoelectric material used here is taken from manufacturer catalog $d = -123x10^{-12}(C/N)$, and $g = -11.1x10^{-3}Vm/N$, and total surface area $V = 50*30*0.5 = 82.5482\,m^2$. Calculations were

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...

made using these values. Stress variation is given for 3 different beams in Figure 17. As can be seen in Figure 17, the stress value obtained on the double-sided taper beam is higher than the single-surface beam. The amount of energy calculated based on these values is given in Figure 18. At the same time, it was observed that the stress value increased by adding an additional mass to the end of the beam.
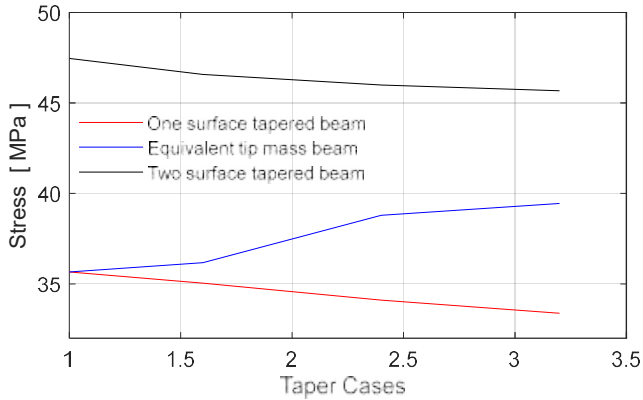
Figure 17. Stress result different beam structures
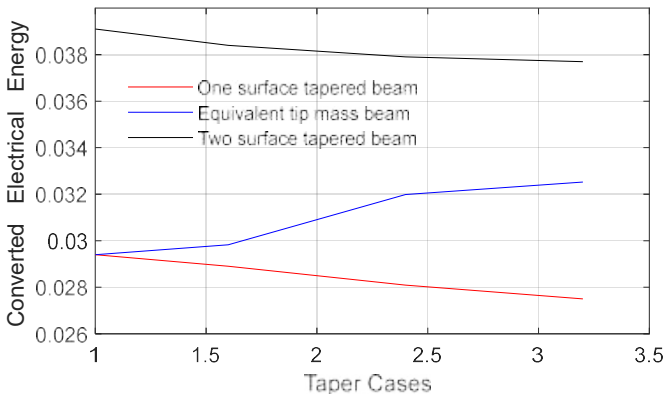
Figure 18. Converted electrical energy from different beams

4. DISCUSSION AND RESULTS

The stress value is one of the most effective factors in energy harvest from piezoelectric material. Numerical analyzes were carried out for different taper ratios, both single surface and double surface taper ratio. To increase harvest energy, in this study the boundary condition was applied to the thin edge of the beam element unlike literature. One end of the conical beam was selected as 1 mm and the other end was changed to 1.6, 2.4 and 3.2 mm, respectively. With the

numerical analysis, it was observed that when the taper was increased with one surface, the stresses on the beam increased. On the other hand, when the double surface taper is created, it has been found that these stress values increase even more. The stress change is transmitted to a directly tapered beam and piezoelectric material that is glued on the beam surface. Therefore, it has been shown that the amount of energy harvest will increase as a result of the increase in the stress value of the piezoelectric material.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] Wang, Junlei, et al. "The state-of-the-art review on energy harvesting from flow-induced vibrations", Applied Energy, 267, 114902, 2020.

[2] Zhang, Min; Hu, Guobiao; Wang, Junlei. "Bluff body with built-in piezoelectric cantilever for flow-induced energy harvesting", International Journal of Energy Research, 2020.

[3] Fevzi C., Bolat, Sinan Basaran, and Selim Sivrioglu, "Piezoelectric and electromagnetic hybrid energy harvesting with low-frequency vibrations of an aerodynamic profile under the air effect" Mechanical Systems and Signal Processing, 133, 106246, 2019.

Fevzi Çakmak BOLAT

Improving Energy Harvesting Efficiency by Vibration-Induced Stresses of Piezoelectric Patch Glued Tap...

[4] Fevzi C., Bolat, and Selim Sivrioglu, "Energy harvesting using an aerodynamic blade element at resonant frequency with air excitation" Smart Structures and Systems, 24(3), pp. 379-390, 2019.

[5] Fevzi C., Bolat, "An Experimental Analysis and Parametric Simulation of Vibration-Based Piezo-Aeroelastic Energy Harvesting Using an Aerodynamic Wing Profile" Arabian Journal for Science and Engineering, 1-8, 2020.

[6] Javed, U. Abdelkefi, A. Akhtar, I, "An improved stability characterization for aeroelastic energy harvesting applications. Commun", Nonlinear Sci. Numer. Simul., 36, pp. 252–265, 2016

[7] Erturk, A. Vieira, W.G.R. De Marqui Jr., C. Inman, D.J., "On the energy harvesting potential of piezoaeroelastic systems", Appl. Phys. Lett., 96(18), 184103, 2010.

[8] Bibo, A. Daqaq, M.F., "Energy harvesting under combined aerodynamic and base excitations", J. Sound Vib., 332(20), pp. 5086–5102, 2013.

[9] Wang, Junlei, et al., "High-performance piezoelectric wind energy harvester with Y-shaped attachments", Energy conversion and management, 181, pp. 645-652, 2019.

[10] Raju, S. Srinivasulu, et al., "An effective energy harvesting in low frequency using a piezo-patch cantilever beam with tapered rectangular cavities", Sensors and Actuators A: Physical, 297, 111522, 2019.

[11] Xie, X. D., A. Carpinteri, and Q. Wang, "A theoretical model for a piezoelectric energy harvester with a tapered shape", Engineering Structures, 144, pp. 19-25, 2017.

[12] Raju, S. Srinivasulu, M. Umapathy, and G. Uma, "High-output piezoelectric energy harvester using tapered beam with

cavity" Journal of Intelligent Material Systems and Structures, 29(5), pp. 800-815, 2018.

[13] Raju, S. Srinivasulu, M. Umapathy, and G. Uma, "Design and analysis of high output piezoelectric energy harvester using non uniform beam", Mechanics of Advanced Materials and Structures, 1-10, 2018.

[14] Matova, S. P., et al., "Effect of length/width ratio of tapered beams on the performance of piezoelectric energy harvesters", Smart Materials and Structures, 22(7), 075015, 2013.

[15] Zhang, Jinfeng, et al., "A study on a near-shore cantilevered sea wave energy harvester with a variable cross section", Energy Science & Engineering, 7(6), pp. 3174-3185, 2019.

[16] Elvin, Niell, and Alper Erturk, eds. Advances in energy harvesting methods. Springer Science & Business Media, 2013.

[17] Yang, Zhengbao, et al., "High-performance piezoelectric energy harvesters and their applications", Joule, 2(4), pp. 642-697, 2018.

Title: The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by
IRI with Observations at Magnetic Conjugate Points for Middle and High Latitudes

Authors: İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observations at Magnetic Conjugate Points for Middle and High Latitudes

İbrahim ÜNAL[*1]

ABSTRACT

In this study, the response to geomagnetic storms of the ionospheric F2 layer critical frequency (foF2) was investigated at the magnetic conjugate points. The hourly foF2 data observed at the magnetic conjugate points of middle and high latitudes for the geomagnetic stormy days around both of 1976 and foF2 data received from IRI-2016 Model for same points were used. The foF2 data observed in magnetic conjugate points and received from IRI Model were examined using the superposed epoch analysis method and the results obtained were compared. From the results of analysis, it was observed that the foF2 data observed at magnetic conjugate points and received from IRI Model simultaneously respond to geomagnetic activity changes for both the middle latitudes and the high latitudes.

Keywords: Ionosphere, conjugate points, geomagnetic activity, critical frequency, International Reference Ionosphere (IRI)

---

[*] Corresponding Author: ibrahim.unal@inonu.edu.tr
[1]İnönü University, Faculty of Education, Department of Science Education, Malatya, TURKEY.
 ORCID: https://orcid.org/ 0000- 0001-8497-4459

İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observation...

1. INTRODUCTION

The two points on Earth that connect opposite ends of the same magnetic field are called geomagnetic conjugate points. Because charged particles are trapped on the same magnetic field lines, disturbances in plasma density during geomagnetic storms can be carried between these two points. Therefore, the ionospheric and magnetospheric processes that were observed at the magnetic conjugate points were expected to occur simultaneously and symmetrically. So, the first studies on these topics were on the detection of magnetic conjugate points and the investigation of the events occurring at these points. The definition of magnetic conjugate points and the processes occurring at these points were tried to present with the results of these studies that were conducted using magnetic field models [1-5].

Later, the investigations related to magnetic conjugate have focused on polar regions, since the ionospheric and magnetospheric traces of the conjugate event during geomagnetic storms are much easier to observe in these regions [6-14]. Also, studies were conducted to determine conjugate points in the middle and low latitude regions and to examine the ionospheric and magnetospheric processes at these points. Different ionospheric and magnetospheric data were used in these studies [15-19].

However, the determination conjugate points at medium and low latitudes and observation of conjugate effects at these points are more difficult than high latitudes. Therefore, the magnetic conjugate still remains a problem, particularly at middle and low latitudes, because there are only very few possibilities to observed the magnetic conjugate with simultaneously in both hemispheres [19-21].

In our previous study, the effects of geomagnetic activity changes were investigated and magnetic conjugate points were determined for different seasons and different latitudes. For this purpose, ionospheric F2 region critical frequency (foF2) data were taken from the stations which are thought to be magnetic conjugate points and statistical methods were used [21]. In this study, similar analyzes were made by using foF2 data

from International Reference Ionosphere (IRI)-2016 Model for middle and high latitudes and the results were compared. IRI is an international project sponsored by the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI). This model has obtained using experimental data and it is not a theoretical model. It provides monthly averages of the electron density, electron temperature, ion temperature, and ion composition in the altitude range from 50 km to 2000 km for a given location, time and date. The previous studies on the determination of conjugated points using the IRI Model have not been encountered. Thus, the IRI Model has tried to be tested in determining the magnetic conjugate points.

2. DATA AND ANALYSIS METHOD

Four stations were used in this study. Akita-Brisbane and Resolute Bay-Scott Base are conjugate pair stations and two are located in the northern hemisphere and the other two are located in the southern hemisphere. The coordinates of these stations for 1976 are given in Table 1 [1].

Table 1. The locations of the stations

Station	Geographic location	Magnetic location
Brisbane	27.5° S, 152.9° E	36.7° S, 227.9° E
Akita	39.7° N, 140.1° E	32.2° N, 210.1° E
Scott Base	77.9° S, 166.8° E	79.8° S, 324.3° E
Resolute Bay	74.7° N, 265.1° E	83.9° N, 309.1° E

In this study, foF2 data and planetary activity index (K_p) data of 1976 were used. 1976 was the year in which the most complete foF2 data could be obtained in four stations, so the data for 1976 were examined. These data were obtained using the internet interface from Community Coordinated Modeling Center (CCMC-NASA) [22]. The data obtained uses the URSI option of the IRI-2016 Model and the storm mode is on.

Hourly K_p data were used in the analyzes. Hourly K_p data were calculated from 3-hour K_p data using linear interpolation method. Thus, the influence of

İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observation...

geomagnetic events on foF2 could be examined hourly. Hourly K_p data were divided into two groups according to geomagnetic activity level. $K_p \leq 2^+$ are geomagnetically quite hours, while $K_p > 2^+$ are geomagnetically active hours. Also, these analyses were conducted for geomagnetic storms that were occurred at different seasons. These seasons were given in Table 2.

The superposed epoch analysis (SEA) method was used in the analyses. This method is used to identify the effect of an event or events on the system occurring through the time series period and to measure the magnitude of response of this system against this event. The SEA describes the impact of geomagnetic storms on foF2 and indicates the magnitude of this impact [20-21, 23-24]. In this study, geomagnetically active hours ($K_p > 2^+$) were selected as events. Table 2 shows the number of events detected for the different seasons of 1976.

Table 2. The number of detected events

1976	15-29 March	15-28 June	14-28 September	16-31 December
Number of events	180	85	203	128

The analysis was applied separately to the foF2 values for all hours and for the geomagnetically quiet hours. δfoF2 values were calculated by subtracting from each other the obtained results. These values show the measure of the response of foF2 to geomagnetic storms [20-21, 23-24]. SEA was performed for all seasons and all stations and the results were compared with each other.

3. RESULTS AND DISCUSSION

Figure 1 shows the variation according to event time of δfoF2 values calculated for Akita and Brisbane. The greatest change in foF2 values for all stations and all seasons occurs at the event time. In addition, changes in foF2 values at both stations during geomagnetic storms have a very similar structure.
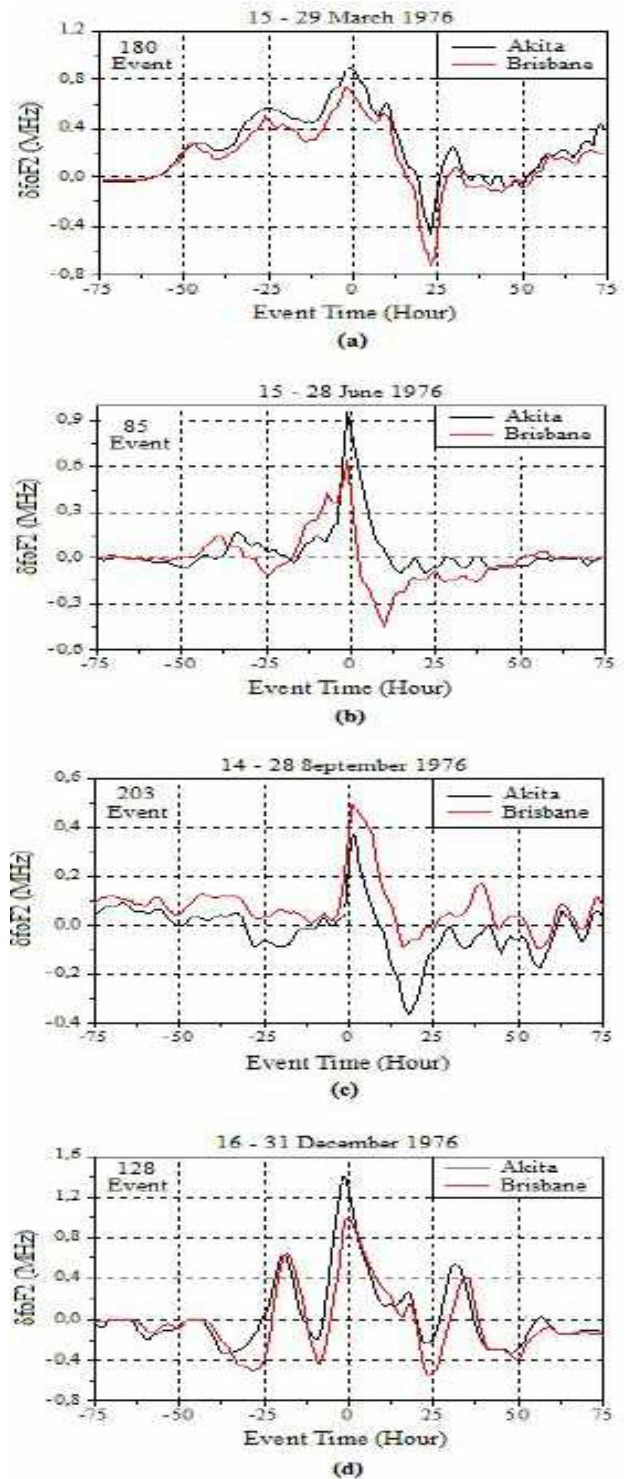
Figure 1. The changes of δfoF2 values [21]

Figure 2 shows the variation according to event time of δfoF2 values calculated for the foF2 values obtained from the IRI-2016 model for Akita and Brisbane. The greatest change in foF2 values for all stations and all seasons occurs at the event time. In addition, changes in foF2 values at both stations during geomagnetic storms have a very similar structure.

İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observation...
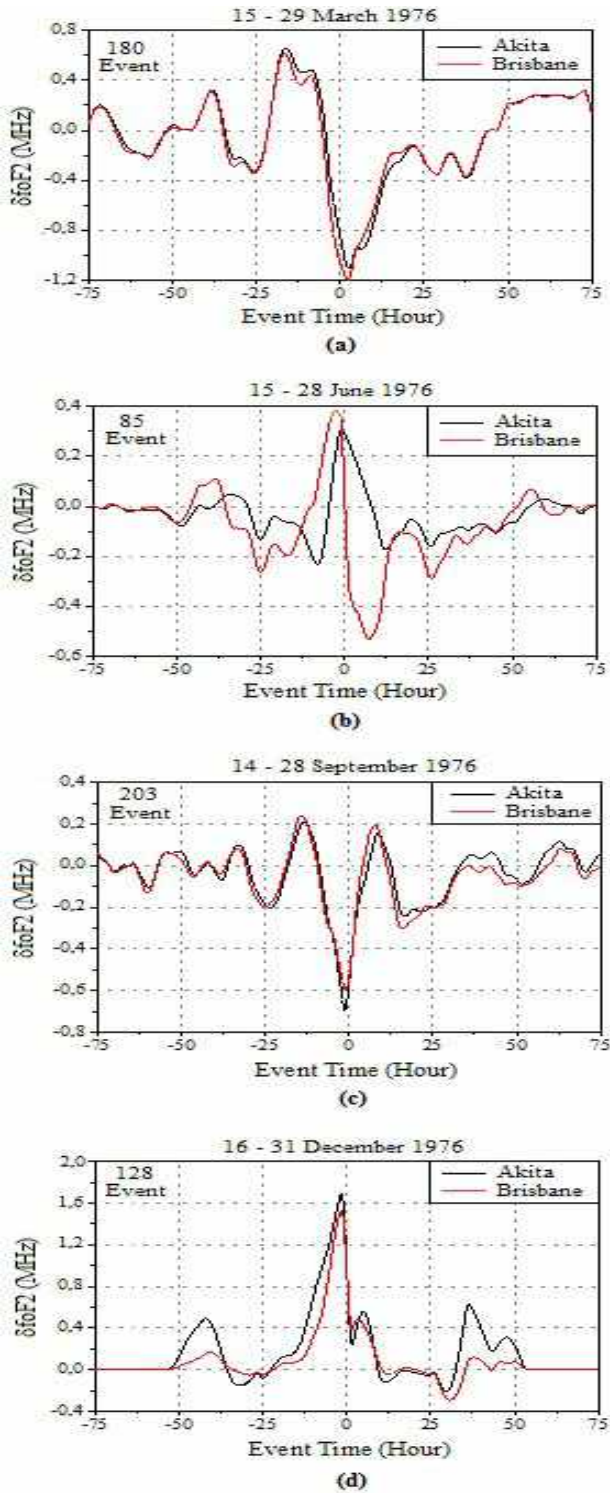
Figure 2. The changes of δfoF2 values calculated using foF2 values from IRI 2016

Figure 3 shows the variation according to event time of δfoF2 values calculated for Resolute Bay and Scott Base. The greatest change in delta fof2 values for both stations occurs at the event time. There is a very high relationship between the changes of these values.
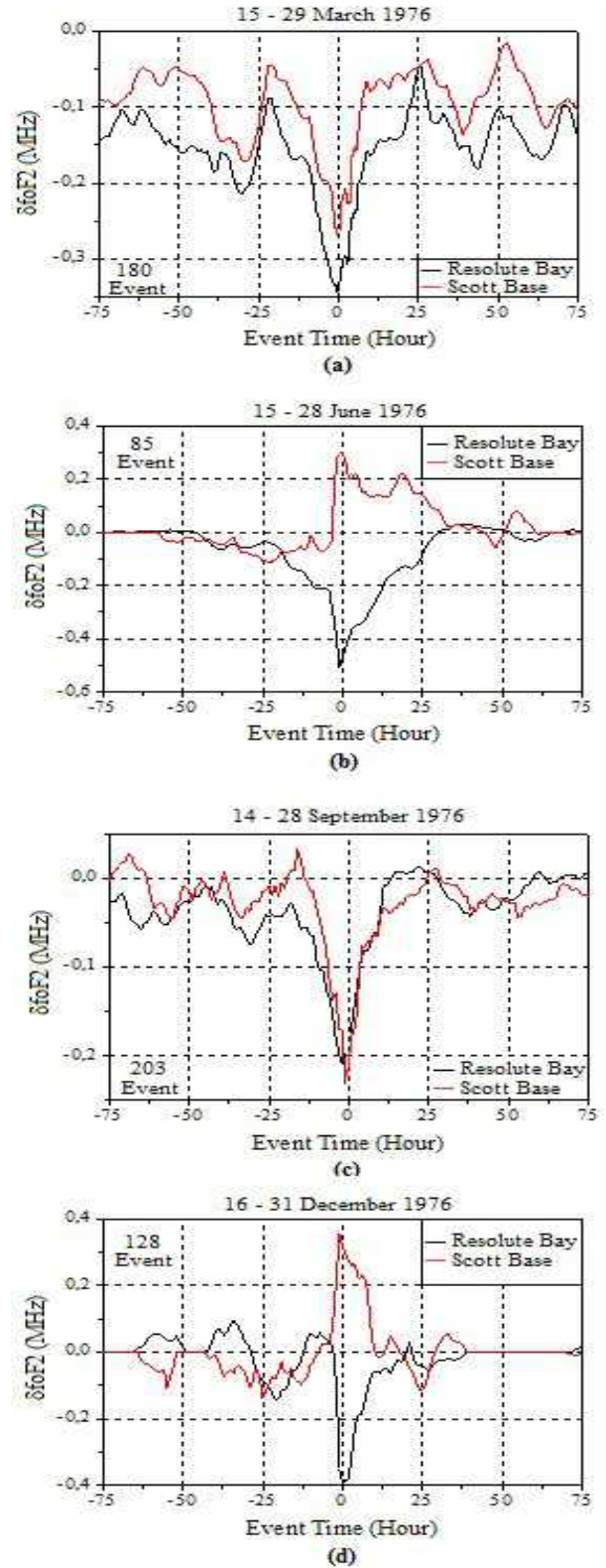
Figure 3. The changes of δfoF2 values [21]

Figure 4 shows the variation according to event time of δfoF2 values calculated for the foF2 values obtained from the IRI-2016 model for Resolute Bay and Scott Base.

İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observation...
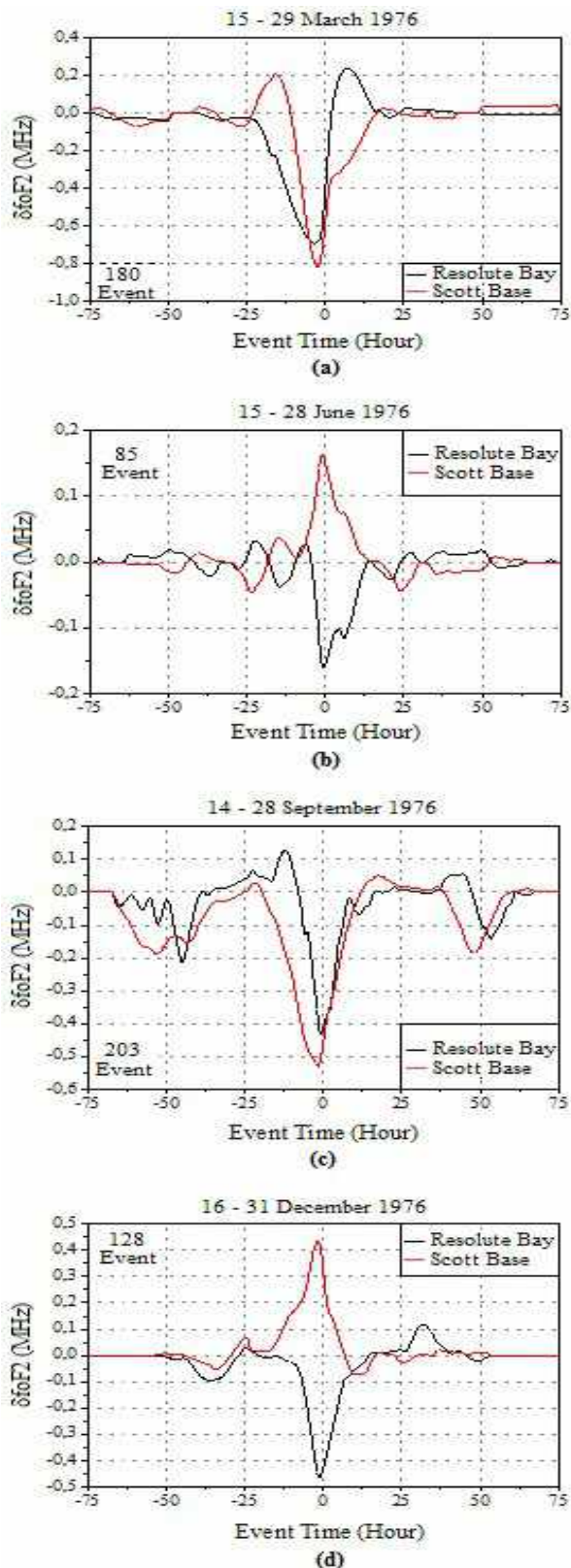
Figure 4. The changes of δfoF2 values calculated using foF2 values from IRI 2016

The greatest change in foF2 values for all stations and all seasons occurs at the event time. In addition, changes in foF2 values at both stations during geomagnetic storms have a very similar structure.

4. CONCLUSIONS

In this research, the influence of magnetic events on the foF2 values obtained from IRI-2106 for the magnetic conjugate stations located in the middle and high latitude regions was investigated using the superposed epoch analysis method. The results obtained from this study were compared with the results obtained from our previous study that were conducted for ionospheric critical frequencies obtained from same locations [21]. Thus, it was tried to determine whether the international ionosphere model can be used to determine the magnetic conjugate points.

Although the magnitude of the responses observed both from the data and the IRI model was different, they show similar characteristic properties for both the equinox and solstice.

The δfoF2 values at high latitudes have a negative relationship in June and December. This result is due to the fact that the electric fields are in the same direction at the equinoxes and opposite direction at the solstices. In other words, electric fields in different hemispheres also exhibit conjugate structure. It is take attention that is also seen from the results obtained from the IRI.

The results obtained from this research show that IRI model can be used to determine the magnetic conjugate points for medium and high latitudes.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observation...

REFERENCES

[1] T. Oguti, "Conjugate point problems", Space Science Reviews, vol. 9, pp. 745-804, 1969.

[2] K. D. Cole and J. A. Thomas, "Maps of the difference in geomagnetic field at conjugate areas", Planetary and Space Science, vol. 16, pp. 1357-1363, 1968.

[3] E. M. Wescott, "Magnetoconjugate phenomena", Space Science Reviews, vol. 5, pp. 507-561, 1966.

[4] W. H. Campbell and S. Matsushita, "World maps of conjugate coordinates and L contours", Journal of Geophysical Research, vol. 72, pp. 3518-3521, 1967.

[5] F. D. Barish and R. E. Wiley, "World contours of conjugate mirror locations", Journal of Geophysical Research: Space Physics, vol. 75, pp. 6342-6346, 1970.

[6] K. Makita, T. Hirasawa and F. Ryoichi, "Visual auroras observed at the Syowa station-Iceland conjugate pair", Polar Res. Spec., vol. 18, pp. 212-225, 1981.

[7] K. Makita, C. I. Meng and S. I. Akasofu, "Comparison of the auroral electron precipitation in the northern and southern conjugate regions by two DMSP satellites", Polar Res. Spec., vol. 26, pp. 149-159, 1983.

[8] N. Sato, R. Fujii, T. Ono, H. Fukunishi and T. Hirasawa, "Conjugacy of proton and electron auroras observed near L=6.1", Geophys. Res. Lett., vol. 13, pp. 1368-1371, 1986.

[9] T. Nagata, "Research of geomagnetically conjugate phenomena in Antarctica since the IGY", Polar Res. Spec., vol. 48, pp. 1-45, 1987.

[10] H. C. Stenbeak-Nielsen and A. Otto, "Conjugate auroras and the inter-planetary magnetic field", J. Geophys. Res., vol. 102, pp. 2223-2232, 1997.

[11] L. L. Lazutin, L. P. Borovkov, T. V. Kozelova, I. A. Kornilov, V. R. Tagirov, A. Korth, J. Stadsnes and S. Ullaland, "Investigation of the conjugacy between auroral breakup and energetic electron injection", J. Geophys. Res., vol. 105, pp. 18495-18503, 2000.

[12] R. W. Schunk and A. F. Nagy, Ionospheres: Physics, Plasma Physics and Chemistry, Cambridge University Press, pp. 366-432. 2000.

[13] N. Østgaard, S. B. Mende, H. U. Frey, T. J. Immel, L. A. Frank, J. B. Sigwarth and T. J. Stubbs, "Interplanetary magnetic field control of the location of substorm onset and auroral features in the conjugate hemispheres", J. Geophys. Res., vol. 109, A07204-A07214, 2004.

[14] N. Østgaard, S. B. Mende, H. U. Frey, J. B. Sigwarth, A. Asnes and J. M. Weygand, "Auroral conjugacy studies based on global imaging", J. Atmos. Sol-Terr. Phys., vol. 69, pp. 249-255, 2007.

[15] A.S. Besprozvannaya, "Empirical modelling of the F2 peak density at 50°–70° invariant latitude using magnetic conjugacy", Advances in Space Research, vol. 11, pp. 23-28, 1991.

[16] H. Le, L. Liu, X. Yue and W. Wan, "The ionospheric behavior in conjugate hemispheres during the 3 October 2005 solar eclipse", Ann. Geophys., vol. 27, pp. 179-184, 2009.

[17] T. L. Gulyaeva, F. Arikan and I. Stanislawska, "Inter-hemispheric imaging of the ionosphere with the upgraded IRI-Plas model during the space weather storms", Earth Planets Space, vol. 63, pp. 929–939, 2011.

[18] N. Yu, M. Ganushkina, V. Kubyshkina, N. Partamies and E. Tanskanen, "Interhemispheric magnetic conjugacy", Journal of Geophysical Research: Space Physical, vol. 118, pp. 1049-1061, 2013.

[19] E. Timoçin, "The north and south symmetry of the ionospheric storms at magnetic conjugate points for low latitudes during the March 1976 severe geomagnetic storms and the relation between daily changes of the storms with geomagnetic activity indices",

İbrahim ÜNAL

The Comparison of Responses to Geomagnetic Activity Changes of foF2 Predicted by IRI with Observation...

Advances in Space Research, vol. 63, pp. 3965-3977, 2019.

[20] E. Timoçin, "The effect of different phases of severe geomagnetic storms on the low latitude ionospheric critical frequencies", Advances in Space Research, vol. 64, pp. 2280-2289, 2019.

[21] E. Timoçin, İ. Ünal, Y. Tulunay and Ü. D. Göker, "The effect of geomagnetic activity changes on the ionospheric critical frequencies (fof2) at magnetic conjugate points", Advances in Space Research, vol. 62, no. 4, pp. 821-828, 2018.

[22] NASA Goddard Space Flight Center (GSFC), https://ccmc.gsfc.nasa.gov/model web/models/iri2016_vitmo.php, [Accessed April 15 2018].

[23] Y. Tulunay, "Variability of mid-latitude ionospheric foF2 compared to IMF-polarity inversions", Adv. Sapace Res., vol. 15, no. 2, pp. 35-44, 1995.

[24] C. J. Davis, M. N. Vild, M. Lockwood and Y. K. Tulunay, "Ionospheric and geomagnetic responses to changes in IMF B_Z: A superposed epoch study", Annali Di Geofisica, vol. 39, no. 4, pp. 853-862, 1997.

Sakarya University Journal of Science

Title: Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller Cooling System

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller Cooling System

Fatih TUFANER[*1,2]

Abstract

The aim of this study was to investigate the noise level reduction of the air cooled liquid chiller to the desired levels of environmental noise requisite. Noise control procedures are implemented with various systems to prevent environmental noise disturbances. This paper examines the application of a passive noise control approach to the control of air-cooled liquid chiller noise located in the garden of an official institution. The approach of passive noise control utilizing brick walls with acoustic sponge in the immediate vicinity of the chiller is discussed. When the chiller was covered with a brick wall, the noise level was reduced by 5.5 Leq dBA 1 m away from the nearest residence. However, this reduction could not meet the requirements of legal regulations. Therefore, the inside of brick wall was covered with acoustic sponge. As a result, the ambient noise level where 1m away from the nearest residence to the chiller has been reduced to the background noise levels with a total reduction of about 8.5 Leq dBA. It is shown that, due to the noise produced by the chiller fans and compressors, so as to achieve significantly cognizable levels of noise reduction it is necessary to isolation brick walls with acoustic sponge.

Keywords: Noise control, reduction, chiller, barrier, acoustic sponge

1. INTRODUCTION

In the past, general city noise was perceived as a city negativity that had to be accepted. However, today, with the technological developments in many areas, the types of complaints subject to noise have increased. In addition to that, necessary arrangements have been made with the laws and regulations regarding the issue of noise. As a result, the people of the city have understood that the noise is not a compulsory disorder and it can be eliminated with the necessary measures.

Especially recently, noise complaints of mechanical equipment that meet the cooling requirement such as chiller has become an important problem in city settlements. These systems are quite large and cause noise disturbance and cannot be completely isolated from their environment due to their technical requirements. Therefore, these systems need a

---

[*] Corresponding Author: ftufaner@adiyaman.edu.tr
[1] Adıyaman University, Faculty of Engineering, Department of Environmental Engineering, Adıyaman, Turkey.
 ORCID: https://orcid.org/0000-0002-1286-7846
[2] Adıyaman University, Environmental Management Application and Research Center, Adıyaman, Turkey.

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

special noise control. When the literature studies are examined, it is seen that there is a big literature gap in this regard. In most studies, there are passive and active noise control studies [1-7] for different systems producing noise. However, there are very few limited scale studies [8-11] for huge systems such as a chiller or electric generator. Therefore, in the present study, a noise control method has been investigated for the chiller cooling system that produces high noise at different frequencies.

Noise control operation should be started with the design of the facility to be made. It is generally recommended to reduce noise by engineering studies. However, generally, noise reduction efforts are started with noise complaints from the environment and are attempted by those responsible for the unit where noise occurs. Their applications generally involve the following: placing the noise source on a flexible surface; completely shut off the noisy machine; adding acoustic insulation materials to the room where the machine is located, and building a barrier between the noise source and the complainants. Installing an acoustical enclosure or barriers around the noise source usually accomplishes the reduction of airborne noise. However, it is impractical to completely enclose the noise source with an acoustic enclosure for cooling the noise source and providing other environmental requirements. In fact, completely enclosing the noise source will require the installation of an auxiliary cooling system. Indeed, industrial fans that provide the necessary air movement produce noise nearly as intense as the noise source produces. Therefore, it is not practical to completely enclosing the noise source [8].

Cooling systems such as air-cooled liquid chillers and other noisy sources are affecting noise sensitive receivers in the nearest buildings and areas. It is more difficult to guarantee a good sound environment near the dwellings and especially in the case of open windows [12]. Airborne noise control is integral to building the design and construction of indoor and outdoor as well as effective environmental management and operations.

Passive noise control is an effective way to bring environmental noise levels to appropriate conditions by adding appropriate intermediate materials or revising noise-generating systems. Noise barriers are the most effective way to reduce the roadway, railway, and industrial noise sources as well as noise control in the source [13]. Passive noise attenuation performance depends on the barrier specification. Some materials of the barrier have the ability to absorb the sound energy while some material of barrier can reflect them. Noise barriers are specifically is effective to attenuate the high-frequencies noise [14]. Also, they are particularly effective at close distances where the acoustic shadow zone is formed. The noise reduction function of the barrier decreases as the distance between the noise source or the receiver and the barrier increase [15]. Insertion loss (IL) due to a barrier is also strongly influenced by the type of ground, atmospheric turbulence and the structural features of the barrier [16-18]. In some studies, the inner face of the barriers is covered with acoustic material for better sound insulation [19-21]. Acoustic sponges are a good sound absorber due to their cavity and pore structures and open porosity [22]. Porous materials have been commonly used in numerous industrial fields to absorb sound energy [23]. Especially, in the automobiles, polyurethane acoustic sponges are applied for sound-absorbing materials in order to absorb sounds in the frequency range higher than 1600 Hz [24].

The noise level transmitted from sources such as the cooling fan, air conditioner, etc. in existing structures, inside or outside the building to noise-sensitive uses cannot exceed the background noise level more than 5 Leq dBA [25]. According to the World Health Organization (WHO) and Turkey Regulation for Assessment and Management of Environmental Noise (Çevresel Gürültünün Değerlendirilmesi ve Yönetimi Yönetmeliği (ÇGDY)) guideline, different noise levels have different influences on the dwellings near the chiller. Noise bigger than limit noise levels is a serious annoyance for dwellings in the daytime. Table 1 shows the maximum permissible noise levels for a residential environment [25, 26].

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

Table 1. Maximum permissible noise levels for residential environments by the WHO and the ÇGDY [25, 26]

Guideline	Specific environment and category	Leq dB(A)
WHO guideline values for community noise in specific environments	**Outdoor living area**	
	Serious annoyance, daytime and evening	55 (16 hours)
	Moderate annoyance, daytime and evening	50 (16 hours)
Turkey ÇGDY guideline noise limit values for specific environments	**Areas with mix use, usually dwellings**	
	Traffic noise, daytime (Existing roads)	68 (12 hours)
	Industrial noise, daytime (Among the areas where commercial buildings and noise-sensitive uses are together, the areas where houses are densely located)	65 (12 hours)

The noise of the chiller systems consists of a combination of the noise of different types of chiller units. Examples include the noise from refrigeration compressor, mechanical fans, and airflow through equipment. All of these noises are composite of many different and separate frequencies of sound that defined the number of vibrations per second and specified as Hertz [Hz]. Pure-tone sounds are more annoying than random broadband sounds in the human ear. Considering the equal dB sound level from random broadband and pure-tone sound sources, pure-tone noise is more likely to cause a noise complaint [10]. In general, to test for the presence of a dominant frequency spectrum component (tone), the sound pressure level in the 1/1 and 1/3 octave band is typically compared to the sound pressure levels on both sides and must pass both of them with a certain level difference (15 dB at low frequencies from 25 to 125 Hz, 8 dB at medium frequencies from 160 to 400 Hz, 5 dB at high frequencies from 500 to 10000 Hz) [27-29].

In this paper, the application of passive noise control to attenuation of the noise produced by an air-cooled liquid chiller was investigated in order to provide the desired levels of environmental noise requisite. In the study, the performance of a gradually constructed acoustical barrier which is designed as a semi-enclosure to reduce the noise of the air-cooled liquid chiller has been evaluated. Concerns about noise conditions and control, design, construction, and operations have been discussed with reference to the ÇGDY (2002/49/EC).

There are few studies in the literature that focus on reducing the noise of huge mechanical systems such as chillers. Therefore, this study is a pioneering field study that performs noise reduction with passive method (brick wall and acoustic sponge). While this paper is a guide in reducing noise caused by mechanical devices such as chiller, it provides a perspective on the level of noise that can be reduced in such problems.

2. METHOD

2.1. General characteristics and location of the air-cooled liquid chillers

The cooling of the official institution is provided by two chillers (YORK R-134a /YVAA0273D) (Figure 1) operating alternately. Its overall dimensions are 2403x2242x8514 mm (HxWxL). The chiller configuration is equipped with high airflow fans with variable speed. Chiller fans give vertical air release from expanded openings. Chillers have been installed on a raised concrete floor about 30 cm to avoid noise and vibration transmission. The chillers have been tied to the main building foundation with a deflection isolator as operational noise will telegraph. The side of the chillers facing the street has been surrounded by a wall with about 2.5 m high a half-roof to reduce the noise effect and the other sides have been covered with wire fences. The side of the houses is the wall and the distance from the chiller to the nearest family houses is 12 m. However, the residential area up to 100 m away actually is impacted by the chiller noise. Air-cooled liquid chillers are located in the garden of an official institution in an area where there are many residences in the city center of Adıyaman. One of the two chillers is only operated alternately during the weekdays between the hours of 08.00-17.00. Chiller has caused environmental noise

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

disturbances, especially on hot summer days. These houses have a history of complaints about chiller noise. Following the complaints, the Provincial Directorate of Environment and Urbanization requested that the noise of chillers should be reduced to the desired values in the regulation. Thereupon, official institution management desired to reduce mechanical noise in the immediate vicinity. The general structure of the air-cooled liquid chillers is shown in Figure 1.

Figure 1. Air-cooled liquid chillers

2.2. Applications to reduce the noise of the chiller

There are different noise attenuation options available to further reduce the noise level at its source thereby meeting environmental noise level regulations. In this study on these noise problems, chillers first enclosed by 18 cm brick walls for noise reduction. The dimensions of the chillers' noise isolation room built in the garden of the institution are 3.5x5x22 m (HxWxL). After that, the inner surfaces of the brick walls were covered with a 40 mm 50 dns fireproof acoustic pyramid sponge. Chillers' noise isolation room is shown in Figure 2.

Daytime sound level measurements were performed to characterize noise emissions and propagation for existing chiller and ambient conditions using a portable noise measurement device. The acoustic measurements were performed using a type 1 Cesva SC310 (Cesva instruments, SL, Barcelona, Spain) a sound-level meter and analyzer (integration time 1s, 125 ms and 35 ms LAS, LAF and LAI, respectively) which was calibrated by a sound calibrator Cesva CB006 (94 dB 1000 Hz). The octave-band frequencies ranging from 31.5 to 16,000 Hz and the one-third-octave-band frequencies ranging from 20 to 10,000 Hz was measured in all cases. The acoustical evaluations of environmental noise were performed, as described in ISO 1996-1 and ISO 1996-2 entitled ''Description, Measurement and Assessment of Environmental Noise" [28, 30] and in ISO 9613-2 entitled "Attenuation of Sound during Propagation Outdoors" [15].

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

Figure 2. Isolation room for air-cooled liquid chiller noise

In accordance with the ÇGDY, the measurement point outside the nearest house that may be uncomfortable was 1 m in front of the façade. Measurement points for the noise maps have been determined such that the difference between the measurement points does not exceed 5 Leq dBA. The noise of the chiller was measured 1 m away from the chiller. The location of the chillers with respect to neighboring family houses is shown in Figure 3. Chiller is located in an area where houses are densely located between commercial buildings and noise-sensitive uses. The background noise of the area where the study was conducted was made between the chiller and the houses that suffered from chiller and in the garden of the institution where the chiller is located. At each point, the noise level was measured at a height of 1.5 m for at least 5 minutes. IDW (Inverse Distance Weighted) interpolation method in ArcGIS 10.1 software has been used for noise mapping and analysis. Accordingly, the Leq dBA noise maps in the immediate vicinity of the chiller have been generated for the background noise of the working area and before and after control the chiller noise.

Figure 3. Location of the chillers with respect to neighboring family houses.

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

3. RESULTS AND DISCUSSIONS

If the chillers are used in noise-sensitive outside locations, the noise level must be carefully evaluated. The chiller is generally located at ground level outside the building to ensure efficient airflow with fans and not to carry the weight of the cooling system into the building. This outside location implies that noise generated by the chiller should be surveyed cautiously in terms of nearby buildings. In this case, it is very necessary that there should exist acceptable environmental noise emission standards. According to the Turkey noise regulations, the noise level transmitted from the air conditioning system located outside the building to noise-sensitive areas cannot be exceeded the background noise level more than 5 dBA in terms of Leq noise indicator [25].

The noise of a chiller is a combination of several individual sources of noise. Therefore, the noise source cannot be considered as the point noise source. The distance between the source and the receiver must be more than twice the largest dimension of a source element so that a noise source can be identified as a point source [28, 31]. In this study, at the distance of 1 meter from the chiller, which is an imaginary parallelepiped measurement surface (Figure 4a) of total surface area (S m^2), is considered around the chiller. If the sound pressure level (SPL) or sound power levels (SWL) are provided, then the following equation is true:

$$SPL(at\ 1\ m) = SWL - 10 \times logS_{(1m)} \qquad (1)$$

This equation can be used to calculate the noise levels at desired distances [32]. The fact that there is a near field and the presence of structures that will affect the sound waves around the chiller has the potential for error in the calculations. The original distance SPL calculation will variably produce a different value. This is due to the general complexity of predicting near-field noise levels, lack of directivity data and variations in the shape of the measurement plane. If sound pressure levels are provided with acceptances though, then the reduction from one distance to another (Re) can be determined by the following equation.

$$Re = 10 \times log \left(\frac{S_1}{S_2}\right) \qquad (2)$$

where S_1 is the area of parallelepiped five-sided surface which is L_1 distance away from the chiller, and S_2 is the area of parallelepiped five-sided surface which is L_2 distance away from chiller [32]. Chiller operates at different loads until idling when it first starts and depending on the weather and the desired indoor temperature. In the study, it was determined that the chiller noise at a distance of 1 m from the midpoint of the long edge of the chiller at different operating loads was between 89 and 104 dBA. Noise in the sound source and the environmental factors affecting the noise in the working area are variable. For this reason, the change of chiller noise according to distance has been calculated up to the receiving point (Figure 4b). According to Figure 4b, the SPL level is reduced by approximately 11.8 dBA at a distance of 11 m from the sound source. SPL was measured as 90.6 Leq dBA at a distance of 1 meter from the chiller and 69.9 Leq dBA at a distance of 11 m from the chiller. In total, SPL decreased by 20.7 dBA at a distance of 11 m. In this case, it is understood that 90.6-11.8-69.9 = 8.9 Leq dBA SPL decreases due to the unilateral walls built in the direction of the chiller and other environmental structures. Figure 1 shows the ground-mounted, air-cooled liquid chillers. The closest residential property line from the chiller is approximately 12 m. However, the residential area up to 100 m away actually may be uncomfortable by the chiller noise. The streets between the chiller and residential area are generally active. Moreover, the noise level of the residential area is increased with the chiller. Five-story institution-building behind the chiller and asphalt and concrete floor around the chiller are reflective surfaces for the noise so that the chiller noise has been reflected toward the residential area. A few trees and garden walls as barriers exist between the chiller and surrounding residences. The residential area is part of Adıyaman, a densely populated during the day and relatively underdeveloped city district. Background sound level of the area is moderately in the summer months, except for the occasional workplace or non-home activities. In addition, the density of traffic in the street increases at the start and end of

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

the work time due to especially the official institution where the chiller is located.
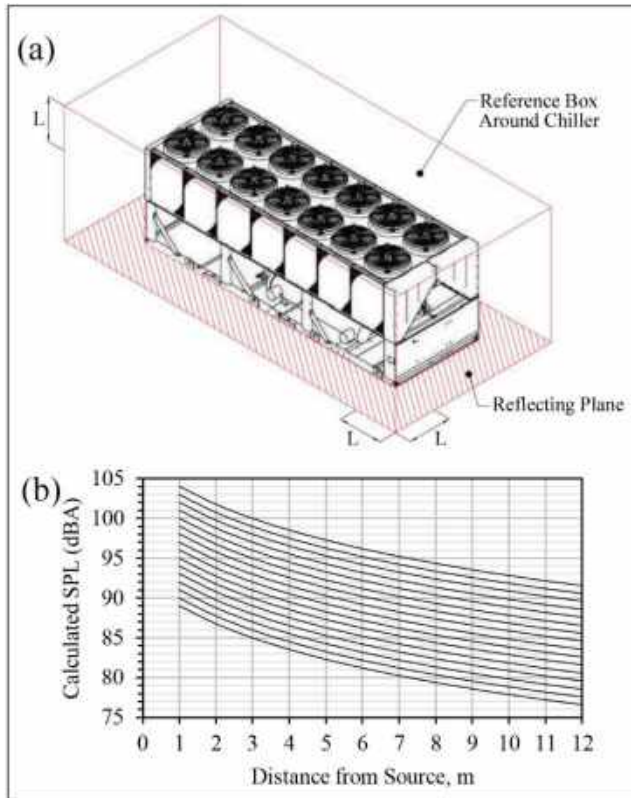
Figure 4. The parallelepiped of the imaginary five-sided constructed at a distance of L meter from the chiller faces (a), the change of chiller noise according to distance (b).

In the institution garden and the street, where the chiller is located, were measured LAS (t: 1s), Leq dBA (t: 5 s) and (t: 30 min) with a type 1 sonometer. As seen in Figure 5 (a) and (c), 30 minutes SPL values were measured as 55.7 and 61.8 Leq dBA, respectively. During the 30-minutes measurement of the background noise level, the LAS and Leq (t: 5 s) values ranged between 54-64 dBA and 54.2-61.6 dBA in the institution garden, 50.9-80.9 dBA and 51.1-78.5 in the street, respectively. A large number of noise indices have been proposed by many researchers to assess environmental noise. The A-weighted equivalent SPL shown by Leq dBA is the most commonly used of the recommended parameters. This parameter takes into account the magnitude of noise and sensitivity at different frequencies, is simple to measure and gives important information to the assessment of the noise annoyance [33-35].

A-weighted statistical noise levels (indicated by L1-99) are important because they take into account the time-dependence of noise. A-weighted statistical noise levels Ln are particularly used in community noise and road traffic assessments in order to try to better account for noise levels fluctuations in and the intermittent noise character. The level Ln defines the SPL which was exceeded for n% of the time [36]. Figure 5 (b) and (d) shows the statistical distribution of the noise levels in the institution garden and the street, respectively. For example, L5 is 66 dBA in Figure 5 (d), so the SPL is for 5% of the time for the period measured is higher than 66 dBA and the rest is less than 66 dBA. The L90 level is generally considered as an approximation of the background or residual level (also referred to as ambient level). L50 which is a useful measure of the audibility of noise from a planned facility is a mean sound level. The L10 which takes account of any annoying peaks in noise is most commonly used for road traffic noise assessment Furthermore, in many cases, L1 is used as a measure of the peak noise level [37]. In the study, L90 levels were determined as 54.8 and 51.9 dBA in the garden of the institution and on the street, respectively. The L50 noise levels were 55.5 and 56.5 dBA at the institution garden and street, respectively. While the noise level difference between L50 and Leq was 0.2 dBA in the garden of the institution, this difference was 5.3 dBA due to traffic noise in the street. Also, L10 and L5 define the noise peaks, and the difference L10–L90 defines the noise fluctuations or noise climate [38, 39]. The difference L10 – L90 and Leq represents as the sum of the noise pollution level, and defines the degree of annoyance that is caused by fluctuating noise [40]. As shown in Figures 5 (b) and (d), The L5, L10 and L10-L90 values were found as 57.5, 56.6, 1.8 dBA and 66, 63.6, 11.7 dBA at the institution garden and street, respectively. In particular, according to L10-L90 data, there is a traffic fluctuation in the street. The sound caused by traffic is often a noise disturbed roadside residents [41]. The traffic noise index (TNI) method used to estimate the annoyance response caused by traffic noise was calculated according to the following equation [42].

$$TNI = 4 \times (L10 - L90) + L90 - 30 \text{ (dBA)} \qquad (3)$$

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...
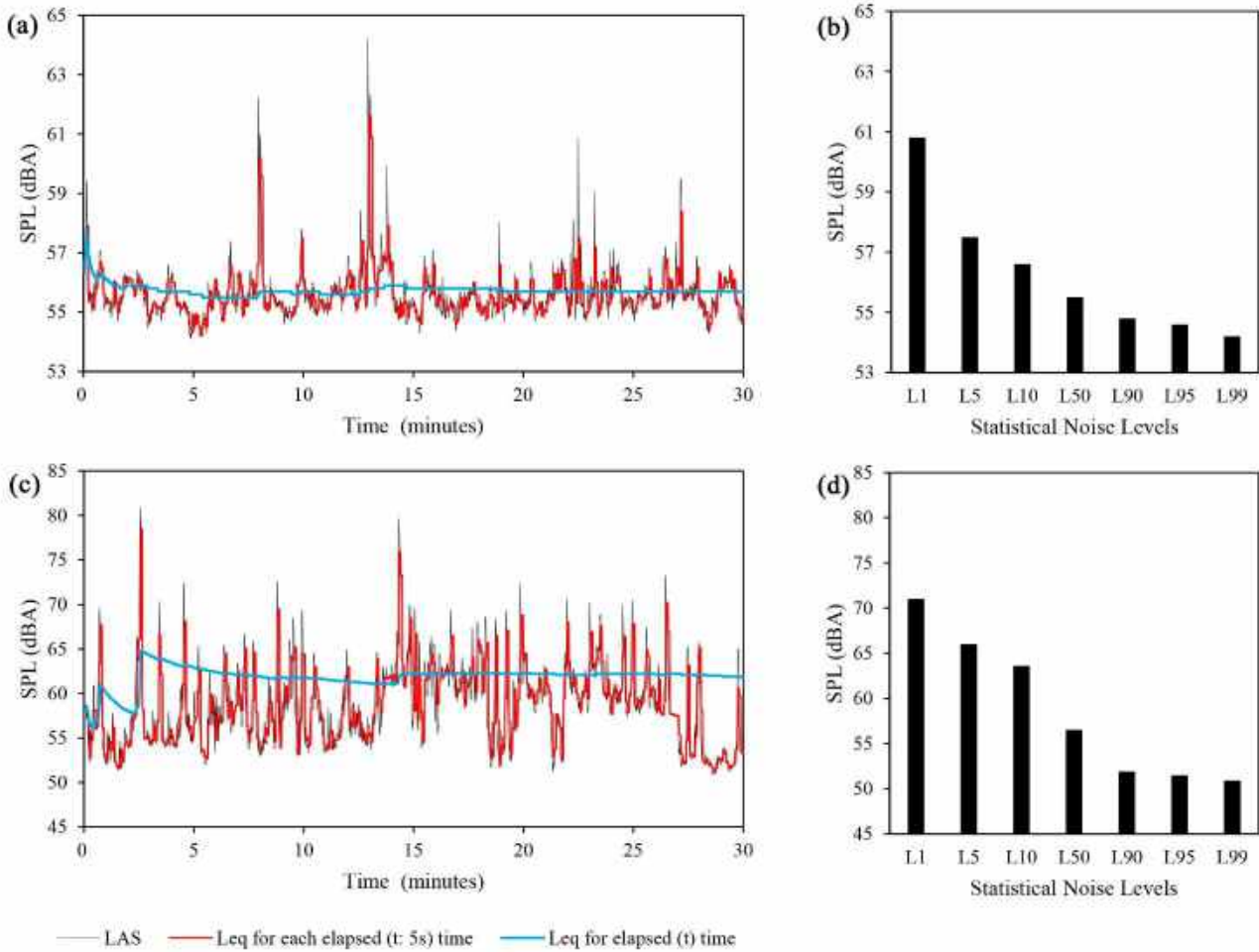
Figure 5. In the institution garden (a, b) and the street (c, d), where the chiller is located, sound level measurements and their statistical analysis.
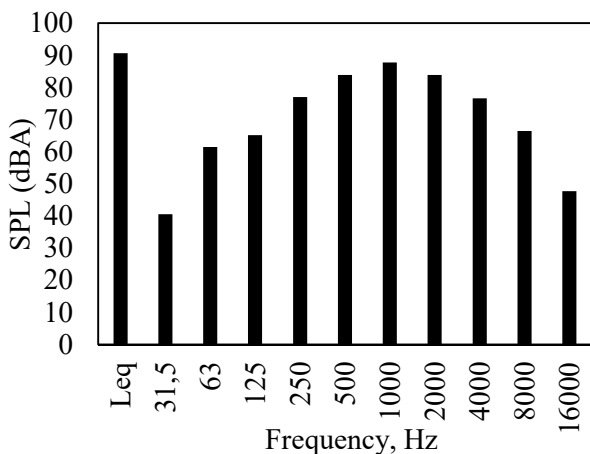
Figure 6. The chiller A-weighted octave band spectra and sound level (Leq) noise measurement with a sonometer location 1m away from the chiller

According to Figure 5 (b) and (d), the TNI in the present study were 32 and 68.7 dBA, respectively. Also, the noise pollution level (LNP), which describes the traffic noise taking into account the temporal fluctuations of noise levels, is calculated according to the following equation [43].

$$LNP = Leq + (L10 - L90) \ (dBA) \qquad (4)$$

According to Figure 5 (b) and (d), the LNP in the present study were 57.5 and 73.5 dBA, respectively.

The chiller cooling system was operated in order to assess the strength of the noise emitted from the system and the SPL emitted from the system was octave and third-octave band analyzed. Figure 6 shows the chiller A-weighted octave band spectra and sound level (Leq dBA) obtained at a central point located 1 m away from the chiller. Since the chiller fan and coolant pump cannot be operated separately, the noise emitted from these systems was calculated as a whole. The peaks of the SPL for the chiller A-weighted octave band are ascended 63 and 2000 Hz, whereas in the third-

octave band are ascended 63, 125, 250, 8000, 2000 and 6300 Hz. The peak for the octave and third-octave band A-weighted SPL is at 1000 Hz (87.8 dBA) and 800 Hz (85.4 dBA), respectively. The sound pressure octave band spectra of the chiller showed that high noise mainly in the

frequency range 250 to 4000 Hz was present at between from 76.7 to 87.8 dBA. In fact, there is no significant contribution to the increase in the total sound level of values smaller than 10 dB from the peak sound value in the spectrum of the octave band.
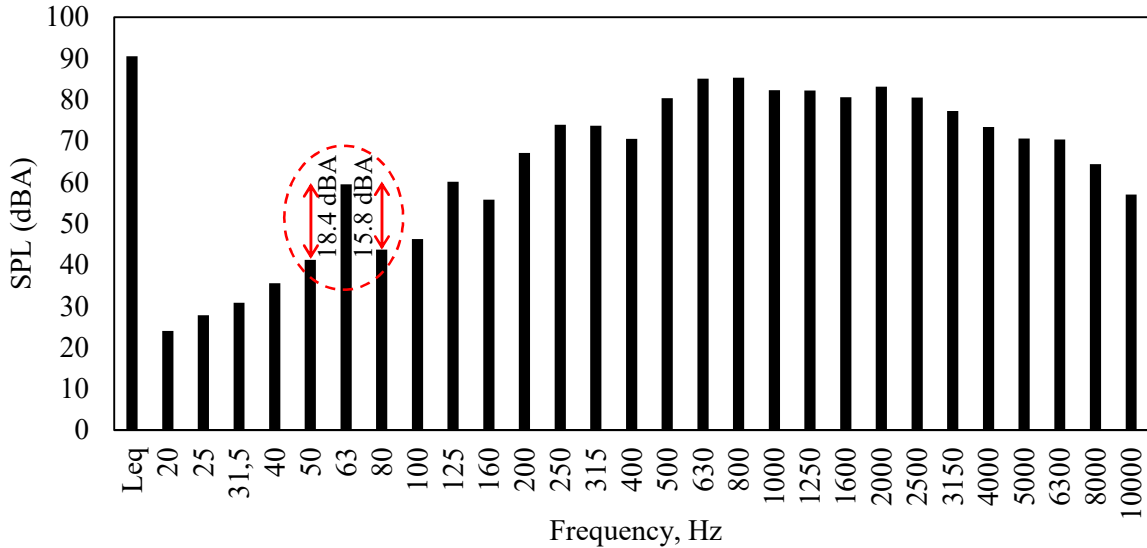
Figure 7. The chiller A-weighted one-third octave band spectra and sound level (Leq) noise measurement with a sonometer location 1 m away from the chiller

■ without control
▨ with wall barrier control
▨ with wall barrier & acoustic sponge control
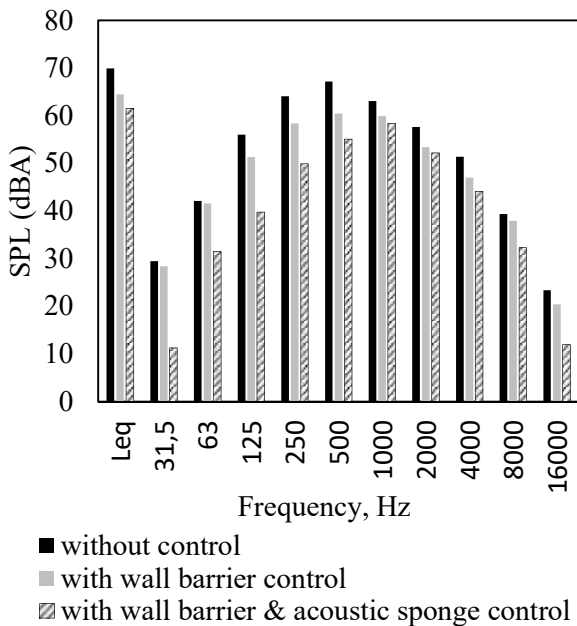
Figure 8. A-weighted octave band spectra and sound level (Leq) measurements 1 m away from the building nearest the chiller (11 m away from the chiller)

hand, in the 1/1 octave band, pure tone noise cannot be seen. As shown in Figure 8, passive noise reduction measures further reduced the noise level octave bands. However, the noise level in the 63 Hz band appears to be slightly more resistant than the other bands. Accordingly, noise reduction studies have mostly eliminated pure tone noises that may occur.

The noise level measured in terms of the Leq noise indicator 1 m away from the nearest structure was 69.9 dBA, whereas the noise level reduced to 64.5 dBA after brick wall application and reduced to 61.5 dBA after the brick wall was covered with the acoustic sponge. Avşar and Gönüllü [41] obtained that the importance of a noise reduction barrier for sensitive areas. The combination of wall barrier treatment and the acoustic sponge resulted in approximately 8.4 Leq dBA noise reduction at 1 m away from the building nearest the chiller (11 m away from the chiller). The noise level at 11 m from the chiller is reduced 5.4 Leq dBA by the brick wall while reducing 3 Leq dBA by acoustic sponge. In the uncontrolled case, the bands between 125 and 2000 Hz are predominant sounds, but the bands forming the main noise are 500 Hz (67.2 dBA)

One-third octave band, the 63 Hz band value exceeds 50 Hz band value by 18.4 dBA and the 80 Hz band value by 15.8 dBA (Figure 7). That is, the 63 Hz band is a pure tone noise because the SPL on both sides exceeds 15 dBA. On the other

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

and the bands next to it. The peak band is 500 Hz (60.5 dBA) again after the brick wall barrier control, but it has reached the approximately same level with the near bands (58.3 dBA for 250 Hz and 59.9 dBA for 1000 Hz). After the acoustic sponge application on the wall barrier, the peak sound level was determined to be in the 1000 Hz band (58.4 dBA). Furthermore, the 1000 Hz band is more resistant to noise abatement than other bands.
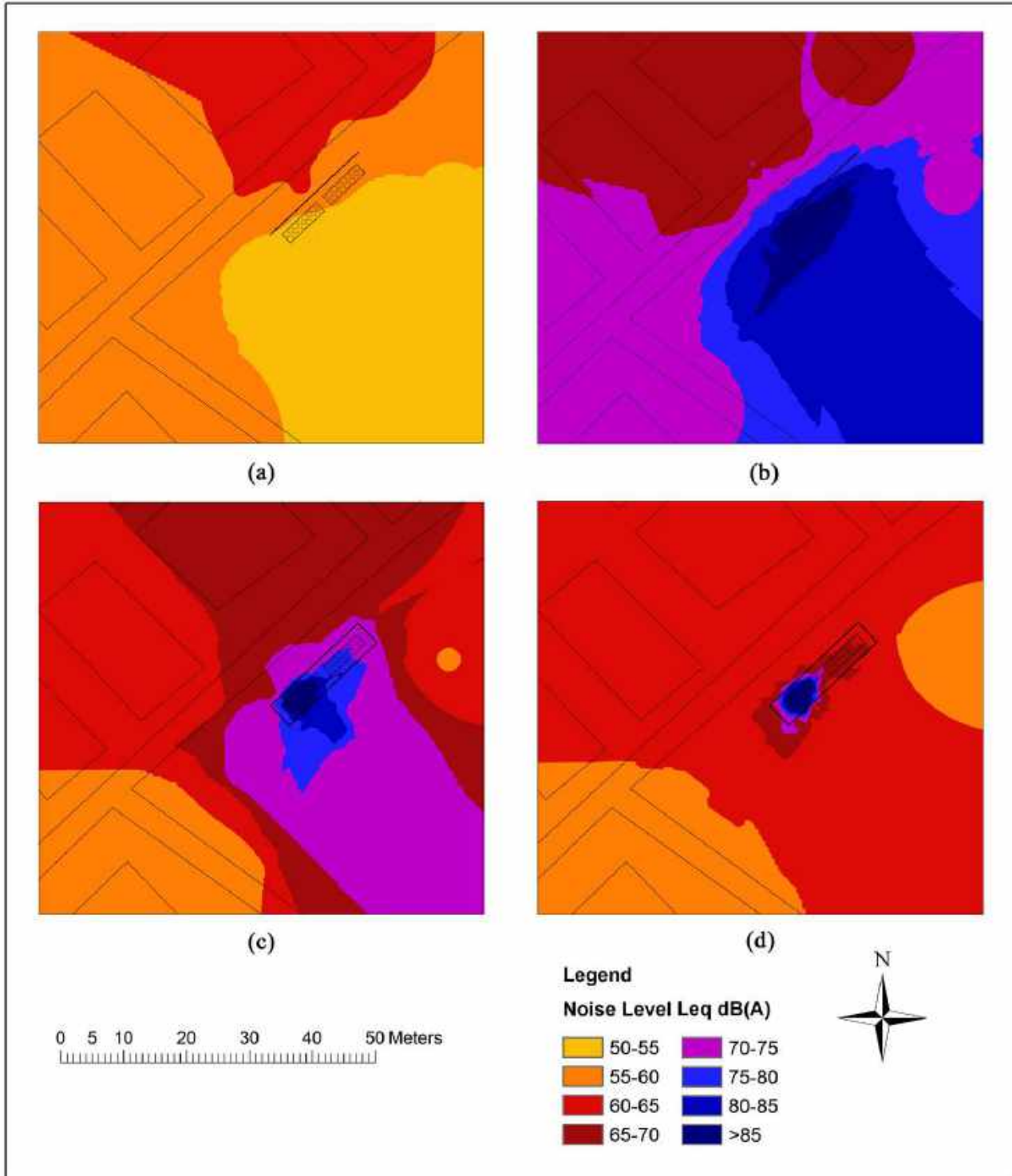
Figure 9. The noise maps in the immediate vicinity of the chiller; the background noise (a), chiller noise for the uncontrolled condition (b), chiller noise after the wall barrier is constructed (c), chiller noise after the application of the wall barrier and acoustic sponge (d)

Figure 9 shows background noise map, chiller noise maps for the uncontrolled condition after the wall barrier is constructed and after the application of the wall barrier and acoustic sponge in the immediate vicinity of the chiller. Figure 9 (a) shows that the background noise levels are between 50 and 65 Leq dBA. Figure 9 (b) shows that the noise of the chiller in the uncontrolled state is above 65 Leq dBA in the vicinity of the chiller. It is also seen that the noise levels reaching the street and nearby houses are in the range of 70-75 Leq dBA. The street-side wall reflects the noise of the chiller towards in the garden of the institution. For this reason, the noise level the garden of the institution is 75 Leq dBA and above in the uncontrolled state. As seen in Figure 9 (c), when the chiller is closed with brick walls, the noise outside the institution garden is at the level of 55-70 Leq dBA while the noise inside the institution garden is above 70 Leq dBA. It is also seen that the noise coming from the air entrance openings opened to the wall floor of the chiller room has reached more towards the garden of the institution. Figure 9 (d) shows that, except for the close distance of 5m of the chiller chamber, the environmental noise values from the chiller are reduced to between 55 to 65 Leq dBA that is background noise level. In addition, the noise from the chiller to ambient has dropped to background noise levels (55-65 Leq dBA).

In the design process, the chiller should evaluate the noise problems of the equipment. For noise reduction studies, only sound power, sound pressure levels and 1/1 - 1/3 octave band data can be obtained from the manufacturers. It is also very difficult to obtain these data for each model chiller. Therefore, if the selected equipment is to be located in a location that can generate environmental noise, the installation, and operating documentation must contain 1/3 octave band data. With this data, a distance-related decrease of the sound level of the chiller can be estimated to determine the distance and operating conditions that the near neighbors will not be affected by the noise of the chiller. According to these calculations, it is possible to decide the sound limit setpoint program in the latitude chiller [44]. Figure 10 shows that ordinary air-cooled screw chillers and latitude chillers exhibit a relative reduction in sound level when operating at off-design conditions, which occur whenever the outdoor temperature is below its design maximum. This feature of the chiller is claimed to eliminate the need for additional noise reduction, such as sound barrier walls. In Figure 10 it is seen that if the ordinary air-cooled screw and latitude chillers are operated at the lowest load, the SPL will be decreased up to 3 and 9 dBA respectively. [45]. However, on hot summer days, if the chiller is operating at full load, additional noise reduction measures become a necessity not a choice. Therefore, alternative equipment types, appropriate equipment locations, and required noise measures should be considered at the very beginning of the installation process.
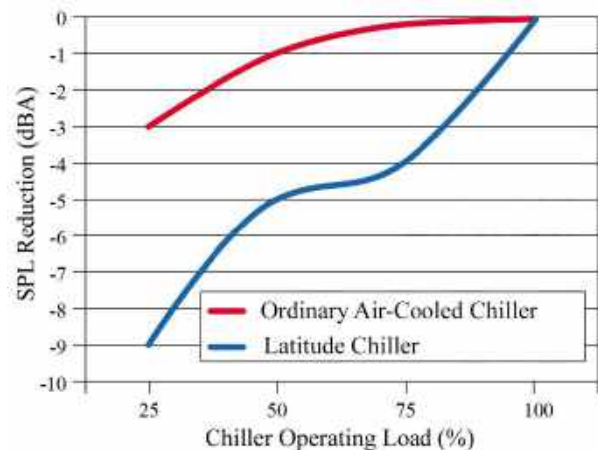
Figure 10. The SPL reduction values (dBA) when operating at off-design conditions of ordinary air-cooled screw and latitude chillers that operated for different loads.

In this study, the noise level of the surrounding residences was reduced to the background level in the general working condition of the chiller. However, chiller operation at high loads may change this situation. In addition, since the top of the chiller is not covered, it prevents full control of the noise. This means a higher risk of noise for the upper floors of buildings in the immediate vicinity. Therefore, it is recommended to consider a special noise-reducing mechanism that can cover the top of the chiller in subsequent studies. In addition, it has been observed that the closure of the chiller surrounds makes the air movement with cooling fans more stable from bottom to top. In other words, the airflow around the heated chiller equipment has become more mobile. This is thought to contribute positively to the cooling process of chiller fans. In addition, it is thought that the service life of the system is prolonged because it is protected from cold-hot and sunlight interactions by covering

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

the chiller system. However, if the chiller is covered more than necessary, the temperature of the system will increase in a short time and the system will automatically shut itself down. This will disrupt the cooling function, which is the main purpose of the chiller.

4. CONCLUSIONS

The institution chose in order to implement the two recommendations, to enclose the chillers with brick wall, and cover the inner surfaces of the brick walls with 40 mm 50 dns fireproof acoustic pyramid sponge. The combination of brick wall barrier and acoustic sponge achieved a noise reduction of approximately 8.5 Leq dBA 11 m away from the chiller (1 meter away from the nearest building to the chiller). The brick wall reduced the noise 11 m away from the chiller by approximately 5.5 Leq dBA, while acoustic sponge on brick wall as plus was reduced by approximately 3 Leq dBA.

As a result of this study, complaints in the vicinity have been significantly reduced, but not eliminated. That the complaints are not eliminated likely is owing to the chiller produces uninterrupted noise despite reducing of the noise level. This uninterrupted noise makes the presence of chiller noise more prominent. Additional noise reduction should also be achieved with a suitable structure constructed on the chiller acoustic barrier or acoustic barrier should be extended a few meters higher than the chiller.

Acknowledgements

Research and Publication Ethics

In all processes of this paper, it was acted in accordance with the research and publication ethics principles of Sakarya University Journal of Science.

Ethics Committee Approval

Ethics committee permission is not required for the data in this paper.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] J. Cheer, S. J. Elliott, "Active noise control of a diesel generator in a luxury yacht", Applied Acoustics, vol. 105, pp. 209-214, 2016.

[2] J. Landaluze, I. Portilla, J. Pagalday, A. Martínez, R. Reyero, "Application of active noise control to an elevator cabin", Control Engineering Practice, vol. 11, no. 12, pp. 1423-1431, 2003.

[3] E. Avşar, H. Erol, İ. Toröz, E. Piro, "Evaluation of environmental noise and vibration levels of dry type transformers and determination of solution alternatives: Istanbul case study", Bitlis Eren University Journal of Science vol. 8, no. 3, pp. 958-967, 2019.

[4] G. S. Papini, R. L. Pinto, E. B. Medeiros, F. B. Coelho, "Hybrid approach to noise control of industrial exhaust systems", Applied Acoustics, vol. 125, pp. 102-112, 2017.

[5] Y. J. Wong, R. Paurobally, J. Pan, "Hybrid active and passive control of fan noise", Applied Acoustics, vol. 64, no. 9, pp. 885-901, 2003.

[6] N. Tandon, B. Nakra, D. Ubhe, N. Killa, "Noise control of engine driven portable generator set", Applied acoustics, vol. 55, no. 4, pp. 307-328, 1998.

[7] H. Wang, P. Luo, M. Cai, "Calculation of noise barrier insertion loss based on varied

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

vehicle frequencies", Applied Sciences, vol. 8, no. 1, pp. 100, 2018.

[8] J. Moreland, R. Minto, "An example of in-plant noise reduction with an acoustical barrier", Applied Acoustics, vol. 9, no. 3, pp. 205-214, 1976.

[9] S. B. Knight, J. B. Evans, C. N. Himmel, "Case study: Air cooled chillers with rotary helical (screw) compressors at hospital with impact on patient rooms, residential neighborhoods, and open park", The 2001 International Congress and Exhibition on Noise Control Engineering, pp. 1-5, 2001.

[10] J. A. Paulauskis, "Addressing noise problems in screw chillers", ASHRAE Journal, vol. 41, pp. 22-27, 1999.

[11] J. B. Evans, C. N. Himmel, J. D. Leasure, "Environmental noise case studies: Air-cooled refrigeration chiller installations near residential structures", ASHRAE Transactions, vol. 118, no. 2, pp. 50-58, 2012.

[12] T. Van Renterghem, J. Forssén, K. Attenborough, P. Jean, J. Defrance, M. Hornikx, et al., "Using natural means to reduce surface transport noise during propagation outdoors", Applied Acoustics, vol. 92, pp. 86-101, 2015.

[13] T. G. Hawkins, "Studies and research regarding sound reduction materials with the purpose of reducing sound pollution", [Master Thesis]. San Luis Obispo: California Polytechnic State University; 2014.

[14] A. Shalool, N. Zainal, K. B. Gan, C. Umat, "An investigation of passive and active noise reduction using commercial and standard TDH-49 headphones", Advances in Electrical, Electronic and Systems Engineering (ICAEES), pp. 606-609, 2016.

[15] ISO_9613-2, "Acoustics – Attenuation of sound during propagation outdoors - Part 2: General method of calculation",

International Organisation for Standards, ISO, 1996.

[16] T. Isei, T. Embleton, J. Piercy, "Noise reduction by barriers on finite impedance ground", The Journal of the Acoustical Society of America, vol. 67, no. 1, pp. 46-58, 1980.

[17] Y. W. Lam, "A boundary element method for the calculation of noise barrier insertion loss in the presence of atmospheric turbulence", Applied Acoustics, vol. 65, no. 6, pp. 583-603, 2004.

[18] T. Ishizuka, K. Fujiwara, "Performance of noise barriers with various edge shapes and acoustical conditions", Applied Acoustics, vol. 65, no. 2, pp. 125-141, 2004.

[19] G. Syms, "Acoustic upgrades to wind tunnels at the national research council canada", 18th AIAA/CEAS Aeroacoustics Conference (33rd AIAA Aeroacoustics Conference) pp. 2180, 2012.

[20] P. Reiter, R. Wehr, H. Ziegelwanger, "Simulation and measurement of noise barrier sound-reflection properties", Applied Acoustics, vol. 123, pp. 133-142, 2017.

[21] B. Botterman, G. D. de la Grée, M. Hornikx, Q. Yu, H. Brouwers, "Modelling and optimization of the sound absorption of wood-wool cement boards", Applied Acoustics, vol. 129, pp. 144-154, 2018.

[22] G. Sung, J. S. Kim, J. H. Kim, "Sound absorption behavior of flexible polyurethane foams including high molecular-weight copolymer polyol", Polymers for Advanced Technologies, vol. 29, no. 2, pp. 852-859, 2018.

[23] S. Chen, Y. Jiang, J. Chen, D. Wang, "The effects of various additive components on the sound absorption performances of polyurethane foams", Advances in Materials Science and Engineering, vol. 2015, 2015.

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

[24] H. Choe, J. H. Kim, "Reactivity of isophorone diisocyanate in fabrications of polyurethane foams for improved acoustic and mechanical properties", Journal of Industrial and Engineering Chemistry, vol. 69, pp. 153-160, 2019.

[25] ÇGDY, "Regulations for Assessment and Management of Environmental Noise (ÇGDY). Official Gazette No. 27601 dated June 4, 2010 and revised in Official Gazette No. 27917 dated April 27", 2011.

[26] B. Berglund, T. Lindvall, D. H. Schwela (1995) "Guidelines for community noise", World Health Organization (WHO). https://pdfs.semanticscholar.org/a95a/a134 1e20ef356ac4c25fdfef43894b4b97e9.pdf. Accessed 6 December 2018

[27] E. Murphy, E. King, "Environmental noise pollution: Noise mapping, public health, and policy", Newnes; 2014.

[28] ISO_1996-2, "Acoustics – Attenuation of sound during propagation outdoors – Part 2: General method of calculation", International Organisation for Standards, ISO, 2007.

[29] BS_4142, "Method for rating industrial noise affecting mixed residential and industrial areas", BS: British Standards Institution, 1997.

[30] ISO_1996-1, "Acoustics – Description, measurement and assessment of environmental noise – Part 1: Basic quantities and assessment procedures", International Organisation for Standards, ISO, 2003.

[31] L. L. Beranek, I. L. Ver, "Noise and vibration control engineering-principles and applications", Noise and vibration control engineering-Principles and applications John Wiley & Sons, Inc, 814 p, 1992.

[32] ISO_3745, "Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure— Precision methods for anechoic rooms and hemi-anechoic rooms", ISO: International Organization of Standards Geneva, Switzerland, 2012.

[33] U. Ayr, E. Cirillo, I. Fato, F. Martellotta, "A new approach to assessing the performance of noise indices in buildings", Applied acoustics, vol. 64, no. 2, pp. 129-145, 2003.

[34] S. Kuwano, "Advantages and disadvantages of A-weighted sound pressure level in relation to subjective impression of environmental noises", Noise Control Eng J, vol. 33, pp. 107-115, 1989.

[35] S. Namba, S. Kuwano, "Psychological study on Leq as a measure of loudness of various kinds of noises", Journal of the Acoustical Society of Japan (E), vol. 5, no. 3, pp. 135-148, 1984.

[36] M. J. Crocker, "Handbook of noise and vibration control", John Wiley & Sons; 2007.

[37] D. A. Bies, C. Hansen, C. Howard, "Engineering noise control", CRC press; 2017.

[38] Y. Soeta, R. Shimokura, "Sound quality evaluation of air-conditioner noise based on factors of the autocorrelation function", Applied Acoustics, vol. 124, pp. 11-19, 2017.

[39] S. Tang, S. Chu, "Noise level distribution functions for outdoor applications", Journal of Sound and Vibration, vol. 248, no. 5, pp. 887-911, 2001.

[40] D. W. Robinson, "Towards a unified system of noise assessment", Journal of Sound and Vibration, vol. 14, no. 3, pp. 279-298, 1971.

[41] Y. Avşar, M. T. Gönüllü, "Determination of safe distance between roadway and school buildings to get acceptable school outdoor noise level by using noise barriers", Build

Fatih TUFANER

Reduction through Brick Wall Barrier and Acoustic Sponge of Environmental Noise Levels from Chiller C...

Environ, vol. 40, no. 9, pp. 1255-1260, 2005.

[42] F. J. Langdon, W. Scholes, "The Traffic Noise Index: A Method of Controlling Noise Nuisance", 1968.

[43] R. B. Hunashal, Y. B. Patil, "Assessment of noise pollution indices in the city of Kolhapur, India", Procedia-Social and Behavioral Sciences, vol. 37, pp. 448-457, 2012.

[44] Johnson_Controls (2015) "Model YVAA style a air-cooled screw liquid chillers with variable speed drive frame sizes", Johnson Controls. https://www.johnsoncontrols.com/en_id/~/media/jci/be/united-states/hvac-equipment/chillers/files/be_yvaa_res_maintenanceguide.pdf?la=en. Accessed 04 July 2018.

[45] Johnson_Controls (2004) "Latitude air-cooled chillers", YORK International Corporation. https://www.johnsoncontrols.com/-/media/jci/be/united-states/hvac-equipment/chillers/files/be_yciv_ycav_res_salesguide.pdf?la=en. Accessed 04 July 2018.

Sakarya University Journal of Science

Title: Some New Inequalities for (α,m1,m2)-GA Convex Functions

Authors: Mahir KADAKAL

Some New Inequalities for (α, m_1, m_2)-GA Convex Functions

Mahir KADAKAL[*1]

Abstract

In this manuscript, firstly we introduce and study the concept of (α, m_1, m_2)-Geometric-Arithmetically (GA) convex functions and some algebraic properties of such type functions. Then, we obtain Hermite-Hadamard type integral inequalities for the newly introduced class of functions by using an identity together with Hölder integral inequality, power-mean integral inequality and Hölder-İşcan integral inequality giving a better approach than Hölder integral inequality. Inequalities have been obtained with the help of Gamma function. In addition, results were obtained according to the special cases of α, m_1 and m_2.

Keywords: (α, m_1, m_2)-GA convex function, Hölder integral inequality, power-mean inequality, Hölder-İşcan inequality, Hermite-Hadamard integral inequality.

[*]Corresponding Author: mahirkadakal@gmail.com
[1]Department of Mathematics, Faculty of Arts and Sciences, Giresun University, 28200, Giresun-Turkey.
ORCID: 0000-0002-0240-918X

1. INTRODUCTION

Let $f: I \subset \mathbb{R} \to \mathbb{R}$ be a convex function defined on the interval I of real numbers and $a, b \in I$ with $a < b$. Then the following inequalities

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \int_a^b f(x)dx \leq \frac{f(a)+f(b)}{2}$$

hold. Both inequalities hold in the reversed direction if the function f is concave [4, 6]. The above inequalities were firstly discovered by the famous scientist Charles Hermite. This double inequality is well-known in the literature as Hermite-Hadamard integral inequality for convex functions. This inequality gives us upper and lower bounds for the integral mean-value of a convex function. Some of the classical inequalities for means can be derived from Hermite-Hadamard inequality for appropriate particular selections of the function f.

Convexity theory plays an important role in mathematics and many other sciences. It provides powerful principles and techniques to study a wide class of problems in both pure and applied mathematics. Readers can find more information in the recent studies [1, 5, 8, 10, 11, 15, 19, 20, 23, 24, 25] and the references therein for different convex classes and related Hermite-Hadamard integral inequalities.

Definition 1. ([17,18]) *A function* $f: I \subseteq \mathbb{R}_+ = (0, \infty) \to \mathbb{R}$ *is said to be GA-convex function on I if the inequality*

$$f\left(x^\lambda y^{1-\lambda}\right) \leq \lambda f(x) + (1-\lambda)f(y)$$

holds for all $x, y \in I$ *and* $\lambda \in [0,1]$, *where* $x^\lambda y^{1-\lambda}$ *and* $\lambda f(x) + (1-\lambda)f(y)$ *are respectively the weighted geometric mean of two positive numbers* x *and* y *and the weighted arithmetic mean of* $f(x)$ *and* $f(y)$.

Definition 2. ([22]) *A function* $f: [0, b] \to \mathbb{R}$ *is said to be* m-*convex for* $m \in (0,1]$ *if the inequality*

$$f(\alpha x + m(1-\alpha)y) \leq \alpha f(x) + m(1-\alpha)f(y)$$

holds for all $x, y \in [0, b]$ *and* $\alpha \in [0,1]$.

Definition 3. ([12]) *The function* $f: [0, b] \to \mathbb{R}$, $b > 0$, *is said to be* (m_1, m_2)-*convex, if the inequality*

$$f(m_1 t x + m_2(1-t)y) \leq m_1 t f(x) + m_2(1-t)f(y)$$

holds for all $x, y \in I$, $t \in [0,1]$ *and* $(m_1, m_2) \in (0,1]^2$.

Definition 4. ([13]) $f: [0, b] \to \mathbb{R}$, $b > 0$, *is said to be* (α, m_1, m_2)-*convex function, if the inequality*

$$f(m_1 t x + m_2(1-t)y) \leq m_1 t^\alpha f(x) + m_2(1-t^\alpha)f(y)$$

holds for all $x, y \in I$, $t \in [0,1]$ *and* $(\alpha, m_1, m_2) \in (0,1]^3$.

Definition 5. ([16]) *For* $f: [0, b] \to \mathbb{R}$ *and* $(\alpha, m) \in (0,1]^2$, *if*

$$f(tx + (1-t)y) \leq t^\alpha f(x) + m(1-t^\alpha)f(y)$$

is valid for all $x, y \in [0, b]$ *and* $t \in [0,1]$, *then we say that* $f(x)$ *is an* (α, m)-*convex function on* $[0, b]$.

Definition 6. ([17]) *The GG-convex functions (called in what follows multiplicatively convex functions) are those functions* $f: I \to J$ *(acting on subintervals of* $(0, \infty)$) *such that*

$$x, y \in I \text{ and } \lambda < \in [0,1] \Rightarrow f(x^{1-t}y^t) \leq f(x)^{1-\lambda}f(y)^\lambda$$

i.e., it is called log-convexity and it is different from the above.

Definition 7. ([9]) *Let the function* $f: [0, b] \to \mathbb{R}$ *and* $(\alpha, m) \in [0,1]^2$. *If*

$$f\left(x^t y^{m(1-t)}\right) \leq t^\alpha f(a) + m(1-t^\alpha)f(b). \quad (1.1)$$

for all $[a, b] \in [0, b]$ *and* $t \in [0,1]$, *then* $f(x)$ *is said to be* (α, m)-*geometric arithmetically convex function or, simply speaking, an* (α, m)-*GA-convex function. If (1.1) reversed, then* $f(x)$ *is*

said to be (α, m)-geometric arithmetically concave function or, simply speaking, an (α, m)-GA-concave function.

A refinement of Hölder integral inequality better approach than Hölder integral inequality can be given as follows:

Theorem 1. (Hölder-İşcan integral inequality [7]) Let $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. If f and g are real functions defined on $[a, b]$ and if $|f|^p$, $|g|^q$ are integrable functions on the interval $[a, b]$ then

$$\int_a^b |f(x)g(x)|dx$$

$$\leq \frac{1}{b-a}\left\{\left(\int_a^b (b-x)|f(x)|^p dx\right)^{\frac{1}{p}}\left(\int_a^b (b-x)|g(x)|^q dx\right)^{\frac{1}{q}}\right.$$

$$\left. + \left(\int_a^b (x-a)|f(x)|^p dx\right)^{\frac{1}{p}}\left(\int_a^b (x-a)|g(x)|^q dx\right)^{\frac{1}{q}}\right\}.$$

Definition 8. (Gamma function) *The classic gamma function is usually defined for $Rez > 0$ by*

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

The main purpose of this paper is to introduce the concept of (α, m_1, m_2)-geometric arithmetically (GA) convex functions and establish some results connected with new inequalities similar to the Hermite-Hadamard integral inequality for these classes of functions.

2. MAIN RESULTS FOR (α, m_1, m_2)-GA CONVEX FUNCTIONS

In this section, we introduce a new concept, which is called (α, m_1, m_2)-GA convex functions and we give by setting some algebraic properties for the (α, m_1, m_2)-GA convex functions, as follows:

Definition 9. *Let the function $f: [0, b] \to \mathbb{R}$ and $(\alpha, m_1, m_2) \in (0,1]^3$. If*

$$f\left(a^{m_1 t}b^{m_2(1-t)}\right) \leq m_1 t^\alpha f(a) + m_2(1 - t^\alpha)f(b) \quad (2.1)$$

for all $[a, b] \in [0, b]$ and $t \in [0,1]$, then the function f is said to be (α, m_1, m_2)-geometric arithmetically convex function, if the inequality (2.1) reversed, then the function f is said to be (α, m_1, m_2)-geometric arithmetically concave function.

Example 1. $f(x) = c, c < 0$ is a (α, m_1, m_2)-geometric arithmetically convex function.

We discuss some connections between the class of the (α, m_1, m_2)-GA convex functions and other classes of generalized convex functions.

Remark 1. *When $m_1 = m_2 = \alpha = 1$, the (α, m_1, m_2)-geometric arithmetically convex (concave) function becomes a geometric arithmetically convex (concave) function defined in [17, 18].*

Remark 2. *When $m_1 = 1$, $m_2 = m$, the (α, m_1, m_2)-geometric arithmetically convex (concave) function becomes an (α, m)-geometric arithmetically convex (concave) function defined in [9].*

Remark 3. *When $m_1 = m_2 = 1$ and $\alpha = s$, the (α, m_1, m_2)-geometric arithmetically convex (concave) function becomes a geometric arithmetically-s convex (concave) function defined in [14].*

Remark 4. *When $\alpha = 1$, the (α, m_1, m_2)-geometric arithmetically convex (concave) function becomes a (m_1, m_2)-GA convex (concave) function defined in [21].*

Proposition 1. *The function $f: I \subset (0, \infty) \to \mathbb{R}$ is (α, m_1, m_2)-GA convex function on I if and only if $f \circ exp: lnI \to \mathbb{R}$ is (α, m_1, m_2)-convex function on the interval $lnI = \{lnx | x \in I\}$.*

Proof. $(\Rightarrow)$ Let $f: I \subset (0, \infty) \to \mathbb{R}$ (α, m_1, m_2)-GA convex function. Then, we write

$$(f \circ exp)(m_1 t lna + m_2(1 - t)lnb)$$

$$\leq m_1 t^\alpha (f \circ exp)(lna) + m_2(1 - t^\alpha)(f \circ exp)(lnb).$$

From here, we get

$$f\left(a^{m_1 t} b^{m_2(1-t)}\right) \leq m_1 t^\alpha f(a) + m_2(1 - t^\alpha)f(b).$$

Hence, the function $f \circ exp$ is (α, m_1, m_2)-convex function on the interval lnI.

($\Leftarrow$) Let $f \circ exp: lnI \to \mathbb{R}$, (α, m_1, m_2)-convex function on the interval lnI. Then, we obtain

$$f\left(a^{m_1 t} b^{m_2(1-t)}\right) = f\left(e^{m_1 tlna + _2(1-t)lnb}\right)$$

$$= (f \circ exp)(m_1 tlna + m_2(1 - t)lnb)$$

$$\leq m_1 t^\alpha f(e^{lna}) + m_2(1 - t^\alpha)f(e^{lnb})$$

$$= m_1 t^\alpha f(a) + m_2(1 - t^\alpha)f(b),$$

which means that the function $f(x)$ (α, m_1, m_2)-GA convex function on I.

Theorem 2. *Let $f, g: I \subset \mathbb{R} \to \mathbb{R}$. If f and g are (α, m_1, m_2)-GA convex functions, then $f + g$ is an (α, m_1, m_2)-GA convex function and cf is an (α, m_1, m_2)-GA convex function for $c \in \mathbb{R}_+$.*

Proof. Let f, g be (α, m_1, m_2)-GA convex functions, then

$$(f + g)\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$= f\left(a^{m_1 t} b^{m_2(1-t)}\right) + g\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$\leq m_1 t^\alpha f(a) + m_2(1 - t^\alpha)f(b)$$

$$+ m_1 t^\alpha g(a) + m_2(1 - t^\alpha)g(b)$$

$$= m_1 t^\alpha (f + g)(a) + m_2(1 - t^\alpha)(f + g)(b)$$

Let f be (α, m_1, m_2)-GA convex function and $c \in \mathbb{R}(c \geq 0)$, then

$$(cf)\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$\leq c[m_1 t^\alpha f(x) + m_2(1 - t^\alpha)f(y)]$$

$$= m_1 t^\alpha (cf)(x) + m_2(1 - t^\alpha)(cf)(y).$$

This completes the proof of the theorem.

Theorem 3. *If $f: I \to J$ is a (m_1, m_2)-GG convex and $g: J \to \mathbb{R}$ is a (α, m_1, m_2)-GA convex function and nondecreasing, then $g \circ f: I \to \mathbb{R}$ is a (α, m_1, m_2)-GA convex function.*

Proof. For $a, b \in I$ and $t \in [0,1]$, we get

$$(g \circ f)\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$= g\left(f\left(a^{m_1 t} b^{m_2(1-t)}\right)\right)$$

$$\leq g\left([f(a)]^{m_1 t}[f(b)]^{m_2(1-t)}\right)$$

$$\leq m_1 t^\alpha g(f(x)) + m_2(1 - t^\alpha)g(f(y)).$$

This completes the proof of the theorem.

Theorem 4. *Let $b > 0$ and $f_\beta: [a, b] \to \mathbb{R}$ be an arbitrary family of (α, m_1, m_2)-GA convex functions and let $f(x) = sup_\beta f_\beta(x)$. If $J = \{u \in [a, b]: f(u) < \infty\}$ is nonempty, then J is an interval and f is an (α, m_1, m_2)-GA convex function on J.*

Proof. Let $t \in [0,1]$ and $x, y \in J$ be arbitrary. Then

$$f\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$= \sup_\beta f_\beta\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$\leq \sup_\beta [m_1 t^\alpha f_\alpha(x) + m_2(1 - t^\alpha)f_\beta(y)]$$

$$\leq m_1 t^\alpha \sup_\beta f_\beta(x) + m_2(1 - t^\alpha)\sup_\beta f_\beta(y)$$

$$= m_1 t^\alpha f(x) + m_2(1 - t^\alpha)f(y) < \infty.$$

This shows simultaneously that J is an interval since it contains every point between any two of its points, and that f is an (α, m_1, m_2)-GA convex function on J. This completes the proof of the theorem.

Theorem 5. *If the function* $f: [a,b] \to \mathbb{R}$ *is an* (α, m_1, m_2)-*GA convex function then* f *is bounded on the interval* $[a,b]$.

Proof. Let $K = \max\{m_1 f(a), m_2 f(b)\}$ and $x \in [a,b]$ is an arbitrary point. Then there exists a $t \in [0,1]$ such that $x = a^{m_1 t} b^{m_2(1-t)}$. Thus, since $m_1 t^\alpha \leq 1$ and $m_2(1 - t^\alpha) \leq 1$ we have

$$f(x) = f\left(a^{m_1 t} b^{m_2(1-t)}\right)$$

$$\leq m_1 t^\alpha f(a) + m_2(1 - t^\alpha)f(b) \leq 2K = M.$$

Also, for every $x \in [a^{m_1}, b^{m_2}]$ there exists a $\lambda \in \left[\sqrt{\frac{a^{m_1}}{b^{m_2}}}, 1\right]$ such that $x = \lambda \sqrt{a^{m_1} b^{m_2}}$ and $x = \frac{\sqrt{a^{m_1} b^{m_2}}}{\lambda}$. Without loss of generality we can suppose $x = \lambda \sqrt{a^{m_1} b^{m_2}}$. So, we have

$$f\left(\sqrt{a^{m_1} b^{m_2}}\right)$$

$$= f\left(\sqrt{\left[\lambda \sqrt{a^{m_1} b^{m_2}}\right] \left[\frac{\sqrt{a^{m_1} b^{m_2}}}{\lambda}\right]}\right)$$

$$\leq \sqrt{f(x)f\left(\frac{\sqrt{a^{m_1} b^{m_2}}}{\lambda}\right)}.$$

By using M as the upper bound, we obtain

$$f(x) \geq \frac{f^2\left(\sqrt{a^{m_1} b^{m_2}}\right)}{f\left(\frac{\sqrt{a^{m_1} b^{m_2}}}{\lambda}\right)} \geq \frac{f^2\left(\sqrt{a^{m_1} b^{m_2}}\right)}{M} = m.$$

This completes the proof of the theorem.

3. HERMITE-HADAMARD INEQUALITY FOR (α, m_1, m_2)-GA CONVEX FUNCTION

This section aims to establish some inequalities of Hermite-Hadamard type integral inequalities for (α, m_1, m_2)-GA convex functions. In this section, we will denote by $L[a,b]$ the space of (Lebesgue) integrable functions on the interval $[a,b]$.

Theorem 6. *Let* $f: [a,b] \to \mathbb{R}$ *be an* (α, m_1, m_2)-*GA convex function. If* $a < b$ *and* $f \in L[a,b]$,

then the following Hermite-Hadamard type integral inequalities hold:

$$f\left(\sqrt{a^{m_1} b^{m_2}}\right) \leq \frac{1}{\ln b^{m_2} - \ln a^{m_1}} \int_{a^{m_1}}^{b^{m_2}} \frac{f(u)}{u} du$$

$$\leq \frac{m_1 f(a)}{\alpha + 1} + \frac{\alpha m_2 f(b)}{\alpha + 1}. \qquad (3.1)$$

Proof. Firstly, from the property of the (α, m_1, m_2)-GA convex function of f, we can write

$$f\left(\sqrt{a^{m_1} b^{m_2}}\right) = f\left(\sqrt{a^{m_1 t} b^{m_2(1-t)} a^{m_1(1-t)} b^{m_2 t}}\right)$$

$$\leq \frac{f(a^{m_1 t} b^{m_2(1-t)}) + f(a^{m_1(1-t)} b^{m_2 t})}{2}.$$

Now, if we take integral in the last inequality with respect to $t \in [0,1]$, we deduce that

$$f\left(\sqrt{a^{m_1} b^{m_2}}\right)$$

$$\leq \frac{1}{2} \int_0^1 f\left(a^{m_1 t} b^{m_2(1-t)}\right) dt + \frac{1}{2} \left(a^{m_1(1-t)} b^{m_2 t}\right) dt$$

$$= \frac{1}{2} \frac{1}{\ln\ m_2 - \ln a^{m_1}} \int_{a^{m_1}}^{b^{m_2}} \frac{f(u)}{u} du$$

$$+ \frac{1}{2} \frac{1}{\ln b^{m_2} - \ln a^{m_1}} \int_{a^{m_1}}^{b^{m_2}} \frac{f(u)}{u} du$$

$$= \frac{1}{\ln b^{m_2} - \ln a^{m_1}} \int_{a^{m_1}}^{b^{m_2}} \frac{f(u)}{u} du.$$

Secondly, by using the property of the (α, m_1, m_2)-GA convex function of f, if the variable is changed as $u = a^{m_1 t} b^{m_2(1-t)}$, then

$$\frac{1}{\ln b^{m_2} - \ln a^{m_1}} \int_{a^{m_1}}^{b^{m_2}} \frac{f(u)}{u} du$$

$$= \int_0^1 f\left(a^{m_1 t} b^{m_2(1-t)}\right) dt$$

$$\leq \int_0^1 [m_1 t^\alpha f(a) + m_2(1 - t^\alpha)f(b)] dt$$

$$= m_1 f(a) \int_0^1 t^\alpha dt + m_2 f(b) \int_0^1 (1 - t^\alpha) dt$$

$$= \frac{m_1 f(a)}{\alpha + 1} + \frac{\alpha m_2 f(b)}{\alpha + 1}$$

This completes the proof of the theorem.

Corollary 1. *By considering the conditions of Theorem 6, if we take $m_1 = m_2 = 1$ and $\alpha = 1$ in the inequality (3.1), then we get*

$$f(\sqrt{ab}) \leq \frac{1}{lnb - lna} \int_a^b \frac{f(u)}{u} du \leq \frac{f(a) + f(b)}{2}.$$

This inequality coincides with the inequality in [2].

Corollary 2. *By considering the conditions of Theorem 6, if we take $\alpha = 1$ in the inequality (3.1), then we get*

$$f\left(\sqrt{a^{m_1} b^{m_2}}\right) \leq \frac{1}{lnb^{m_2} - ln\ m_1} \int_{a^{m_1}}^{b^{m_2}} \frac{f(u)}{u} du$$

$$\leq \frac{m_1 f(a) + m_2 f(b)}{2}.$$

This inequality coincides with the inequality in [14].

4. SOME NEW INEQUALITIES FOR (α, m_1, m_2)-GA CONVEX FUNCTIONS

The main purpose of this section is to establish new estimates that refine Hermite-Hadamard integral inequality for functions whose first derivative in absolute value, raised to a certain power which is greater than one, respectively at least one, is (α, m_1, m_2)-GA convex function. Ji et al. [9] used the following lemma. Also, we will use this lemma to obtain our results.

Lemma 1. ([3]) *Let $f: I \subseteq \mathbb{R}_+ = (0, \infty) \to \mathbb{R}$ be differentiable function and $a, b \in I$ with $a < b$. If $f' \in L([a, b])$, then*

$$\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx$$

$$= \frac{lnb - lna}{2} \int_0^1 a^{3(1-t)} b^{3t} f'(a^{1-t} b^t) dt.$$

Theorem 7. *Let the function $f: \mathbb{R}_0 = [0, \infty) \to \mathbb{R}$ be a differentiable function and $f' \in L[a, b]$ for $0 < a < b < \infty$. If $|f'|$ is (α, m_1, m_2)-GA convex on $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$ for $[\alpha, m_1, m_2] \in$*

$(0,1]^3$, *then the following integral inequalities hold*

$$\left|\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx\right| \qquad (4.1)$$

$$\leq \frac{m_1}{2} \left|f'\left(a^{\frac{1}{m_1}}\right)\right|$$

$$\left[\frac{b^3 - a^3}{3} - \frac{a^3 \Gamma(\alpha+1, 3(lna - lnb)) - a^3 \Gamma(\alpha+1, 0)}{3^{\alpha+1}(lna - lnb)^\alpha}\right]$$

$$+ \frac{m_2}{2} \left|f'\left(b^{\frac{1}{m_2}}\right)\right| \left[\frac{a^3 \Gamma(\alpha+1, 3(lna - lnb)) - a^3 \Gamma(\alpha+1, 0)}{3^{\alpha+1}(lna - lnb)^\alpha}\right],$$

where Γ is the Gamma function.

Proof. By using Lemma 1 and the inequality

$$|f'(a^{1-t} b^t)| = \left|f'\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)} f'\left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right|$$

$$\leq m_1(1 - t^\alpha) \left|f'\left(a^{\frac{1}{m_1}}\right)\right| + m_2 t^\alpha \left|f'\left(b^{\frac{1}{m_2}}\right)\right|,$$

we get

$$\left|\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx\right|$$

$$\leq \frac{ln(b/a)}{2} \int_0^1 \left|a^{3(1-t)} b^{3t}\right| |f'(a^{1-t} b^t)| dt$$

$$\leq \frac{ln(b/a)}{2} \int_0^1 a^{3(1-t)} b^{3t} \left[\begin{array}{c} m_1(1 - t^\alpha) \left|f'\left(a^{\frac{1}{m_1}}\right)\right| \\ + m_2 t^\alpha \left|f'\left(b^{\frac{1}{m_2}}\right)\right| \end{array}\right] dt$$

$$= m_1 \left|f'\left(a^{\frac{1}{m_1}}\right)\right| \frac{ln(b/a)}{2} \int_0^1 (1 - t^\alpha) a^{3(1-t)} b^{3t} dt$$

$$+ m_2 \left|f'\left(b^{\frac{1}{m_2}}\right)\right| \frac{ln(b/a)}{2} \int_0^1 t^\alpha a^{3(1-t)} b^{3t} dt$$

$$= \frac{m_1}{2} \left|f'\left(a^{\frac{1}{m_1}}\right)\right| \left[\frac{b^3 - a^3}{3}\right.$$

$$\left. - \frac{a^3 \Gamma(\alpha+1, 3(lna - lnb)) - a^3 \Gamma(\alpha+1, \ 0)}{3^{\alpha+1}(lna - lnb)^\alpha}\right]$$

$$+ \frac{m_2}{2} \left|f'\left(b^{\frac{1}{m_2}}\right)\right| \left[\frac{a^3 \Gamma(\alpha+1, \ 3(lna - l \quad)) - a^3 \Gamma(\alpha+1, \ 0)}{3^{\alpha+1}(lna - lnb)^\alpha}\right].$$

This completes the proof of the theorem.

Corollary 3. *By considering the conditions of Theorem 7, if we take $m_1 = m_2 = 1$ and $\alpha = 1$ then we get*

$$\left|\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx\right|$$

$$\leq \frac{|f'(a)|}{6}[L(a^3, b^3) - a^3] + \frac{|f'(b)|}{6}[b^3 - L(a^3, b^3)],$$

where L is the logarithmic mean.

Corollary 4. *By considering the conditions of Theorem 7, if we take $\alpha = 1$ in the inequality (4.1), then we get*

$$\left|\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx\right|$$

$$\leq \frac{m_1}{2}\left|f'\left(a^{\frac{1}{m_1}}\right)\right| [L(a^3, b^3) - a^3]$$

$$+ \frac{m_2}{2}\left|f'\left(b^{\frac{1}{m_2}}\right)\right| [b^3 - L(a^3, b^3)].$$

Theorem 8. *Let the function $f: \mathbb{R}_0 = [0, \infty) \to \mathbb{R}$ be a differentiable function and $f' \in L[a, b]$ for $0 < a < b < \infty$. If $|f'|^q$ is (α, m_1, m_2)-GA convex on $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$ for $[\alpha, m_1, m_2] \in (0,1]^3$ and $q \geq 1$ then*

$$\left|\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx\right| \tag{4.2}$$

$$\leq \frac{lnb - lna}{2} L^{1-\frac{1}{q}}(a^3, b^3)$$

$$\cdot \left[m_1\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q \left(\frac{\frac{b^3 - a^3}{3(lnb - lna)}}{-\frac{a^3\Gamma(\alpha+1, \ 3(ln \quad))-a^3\Gamma(\alpha+1, \ 0)}{3^{\alpha+1}(lnb - l \quad)(lna - lnb)^\alpha}}\right)\right.$$

$$\left. + m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q \left(\frac{a^3\Gamma(\alpha+1, \ 3(lna - l \quad))-a^3\Gamma(\alpha+1, \ 0)}{3^{\alpha+1}(lnb - lna)(lna - lnb)^\alpha}\right)\right]^{\frac{1}{q}},$$

where L is the logarithmic mean.

Proof. By using both Lemma 1, power-mean inequality and the (α, m_1, m_2)-GA convexity of

$|f'|^q$ on the interval $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$, that is, the inequality

$$|f'(a^{1-t} b^t)| = \left|f'\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)} f'\left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right|^q$$

$$\leq m_1(1-t^\alpha)\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q + m_2 t^\alpha \left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q,$$

is satisfied and we get

$$\left|\frac{b^2 f(a) - a^2 f(b)}{2} - \int_a^b x f(x) dx\right|$$

$$\leq \frac{ln\left(\frac{b}{a}\right)}{2}\left[\int_0^1 a^{3(1-t)} b^{3t} dt\right]^{1-\frac{1}{q}}$$

$$\left[\int_0^1 a^{3(1-t)} b^{3t} \left|f'\left(\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)}\left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right)\right|^q dt\right]^{\frac{1}{q}}$$

$$\leq \frac{ln\left(\frac{b}{a}\right)}{2}\left[\int_0^1 a^{3(1-t)} b^{3t} dt\right]^{1-\frac{1}{q}}$$

$$\cdot \left(\int_0^1 a^{3(1-t)} b^{3t}\left[\begin{array}{c} m_1(1-t^\alpha)\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q \\ + m_2 t^\alpha\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q \end{array}\right] dt\right)^{\frac{1}{q}}$$

$$= \frac{ln\left(\frac{b}{a}\right)}{2}\left[\int_0^1 a^{3(1-t)} b^{3t} dt\right]^{1-\frac{1}{q}}$$

$$\times \left[\begin{array}{c} m_1\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q \int_0^1 (1-t^\alpha)a^{3(1-t)} b^{3t} dt \\ + m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q \int_0^1 t^\alpha a^{3(1-t)} b^{3t} dt \end{array}\right]^{\frac{1}{q}}$$

$$= \frac{lnb - l}{2} L^{1-\frac{1}{q}}(a^3, b^3)$$

$$\cdot \left[m_1\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q \left(\frac{\frac{b^3 - a^3}{3(lnb - lna)}}{-\frac{a^3\Gamma(\alpha+1, \ 3(lna - lnb))-a^3\Gamma(\alpha+1, \ 0)}{3^{\alpha+1}(lnb - lna)(lna - lnb)^\alpha}}\right)\right.$$

$$\left. + m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q \left(\frac{a^3\Gamma(\alpha+1, \ 3(lna - lnb))-a^3\Gamma(\alpha+1, \ 0)}{3^{\alpha+1}(lnb - lna)(lna - lnb)^\alpha}\right)\right]^{\frac{1}{q}}.$$

This completes the proof of the theorem.

Corollary 5. *By considering the conditions of Theorem 8, if we take $m_1 = m_2 = 1$ and $\alpha = 1$ in the inequality (4.2), then we get*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right| \le \frac{lnb-lna}{2} L^{1-\frac{1}{q}}(a^3,b^3)$$

$$\times \left[|f'(a)|^q \frac{L(a^3,b^3)-b^3}{3(lnb-lna)} + |f'(b)|^q \frac{b^3-L(a^3,b^3)}{3(lnb-lna)}\right]^{\frac{1}{q}},$$

where L is the logarithmic mean.

Corollary 6. *By considering the conditions of Theorem 8, if we take $q = 1$, then*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right| \le$$

$$\times \left[\frac{m_1}{2}\left|f'\left(a^{\frac{1}{m_1}}\right)\right|\left(\frac{b^3 - a^3}{3}\right.\right.$$

$$\left.-\frac{a^3 \Gamma\left(\alpha+1, 3(lna-lnb)\right)-a^3\Gamma(\alpha+1,0)}{3^{\alpha+1}(lna-lnb)^\alpha}\right)$$

$$\frac{m_2}{2}\left|f'\left(b^{\frac{1}{m_2}}\right)\right|\left(\frac{a^3\Gamma(\alpha+1,\ 3(lna-lnb))-a^3\Gamma(\alpha+1,\ 0)}{3^{\alpha+1}(lna-lnb)^\alpha}\right)\right].$$

This inequality coincides with the inequality (4.1).

Corollary 7. *By considering the conditions of Theorem 8, if we take $m_1 = m_2 = 1$ and $\alpha = q = 1$ in the inequality (4.2), then we get*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\le \left[\frac{|f'(a)|}{6}(L(a^3,b^3) - b^3) + \frac{|f'(b)|}{6}(b^3 - L(a^3,b^3))\right],$$

where L is the logarithmic mean.

Corollary 8. *By considering the conditions of Theorem 8, if we take $m_1 = m$ and $m_2 = 1$ in the inequality (4.2), then we get*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right| \le \frac{lnb-lna}{2} L^{1-\frac{1}{q}}(a^3,b^3)$$

$$\cdot\left[m\left|f'\left(a^{\frac{1}{m}}\right)\right|^q \left(\frac{b^3-a^3}{3(lnb-lna)} - \frac{a^3\Gamma(\alpha+1,\ 3(lna-lnb))-a^3\Gamma(\alpha+1,\ 0)}{3^{\alpha+1}(lnb-lna)(lna-lnb)^\alpha}\right)\right.$$

$$+ |f'(b)|^q \left.\left(\frac{a^3\Gamma(\alpha+1,\ 3(lna-l\ \ ))-a^3\Gamma(\alpha+1,\ 0)}{3^{\alpha+1}(lnb-lna)(lna-lnb)^\alpha}\right)\right]^{\frac{1}{q}}.$$

This inequality coincides with the inequality in [9].

Theorem 9. *Let the function $f: \mathbb{R}_0 = [0,\infty) \to \mathbb{R}$ be a differentiable function and $f' \in L[a,b]$ for $0 < a < b < \infty$. If $|f'|^q$ is (α, m_1, m_2)-GA convex on $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$ for $[\alpha, m_1, m_2] \in (0,1]^3$ and $q > 1$, then,*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right| \le \frac{ln(b/a)}{2}$$

$$\cdot L^{\frac{1}{p}}(a^{3p},b^{3p})\left[\frac{\alpha m_1 \left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q}{\alpha+1} + \frac{m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q}{\alpha+1}\right]^{\frac{1}{q}}, (4.3)$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. By using both Lemma 1, Hölder integral inequality and the (α, m_1, m_2)-GA-convexity of the function $|f'|^q$ on the interval $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$, that is, the inequality

$$|f'(a^{1-t}b^t)| = \left|f'\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)} f'\left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right|^q$$

$$\le m_1(1-t)\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q + m_2 t\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q,$$

we get

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\le \frac{ln(b/a)}{2}\left[\int_0^1 \left(a^{3(1-t)}b^{3t}\right)^p dt\right]^{\frac{1}{p}}$$

$$\times \left[\int_0^1 \left|f'\left(\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)}\left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right)\right|^q dt\right]^{\frac{1}{q}}$$

$$\leq \frac{\ln(b/a)}{2}\left[\int_0^1 \left(a^{3(1-t)}b^{3t}\right)^p dt\right]^{\frac{1}{p}}$$

$$\cdot\left[\int_0^1 \left[m_1(1-t^\alpha)\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q +\right.\right.$$

$$\left.\left. m_2 t^\alpha \left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q\right]dt\right]^{\frac{1}{q}}$$

$$= \frac{\ln(b/a)}{2}\left[\int_0^1 a^{3p(1-t)}b^{3pt}dt\right]^{\frac{1}{p}}$$

$$\times\left[m_1\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q \int_0^1 (1-t^\alpha)dt +\right.$$

$$\left. m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q \int_0^1 t^\alpha dt\right]^{\frac{1}{q}}$$

$$= \frac{\ln(b/a)}{2}L^{\frac{1}{p}}(a^{3p},b^{3p})\left[\frac{\alpha m_1\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q}{\alpha+1}+\frac{m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q}{\alpha+1}\right]^{\frac{1}{q}}.$$

This completes the proof of the theorem.

Corollary 9. *By considering the conditions of Theorem 9, if we take $m_1 = m_2 = 1$ in the inequality (4.3), then we get*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\leq \frac{\ln(b/a)}{2}L^{\frac{1}{p}}(a^{3p},b^{3p})\left[\frac{\alpha|f'(a)|^q}{\alpha+1}+\frac{|f'(b)|^q}{\alpha+1}\right]^{\frac{1}{q}}.$$

Corollary 10. *By considering the conditions of Theorem 9, if we take $m_1 = m, m_2 = 1$ in the inequality (4.3) then we obtain*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\leq \frac{\ln(b/a)}{2}L^{\frac{1}{p}}(a^{3p},b^{3p})\left[\frac{\alpha m\left|f'\left(a^{\frac{1}{m}}\right)\right|^q}{\alpha+1}+\frac{|f'(b)|^q}{\alpha+1}\right]^{\frac{1}{q}}.$$

Corollary 11. *By considering the conditions of Theorem 9, if we take $m_1 = m_2 = 1$ in the inequality (4.3) then we obtain*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\leq \frac{\ln\left(\frac{b}{a}\right)}{2}L^{\frac{1}{p}}(a^{3p},b^{3p})A^{\frac{1}{q}}(|f'(a)|^q,|f'(b)|^q).$$

Theorem 10. *Let the function $f:\mathbb{R}_0 = [0,\infty) \to \mathbb{R}$ be a differentiable function and $f' \in L[a,b]$ for $0 < a < b < \infty$. If $|f'|^q$ is (α,m_1,m_2)-GA convex on $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$ for $[\alpha,m_1,m_2] \in (0,1]^3$ and $q > 1$, then the following integral inequalities hold*

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right| \qquad (4.4)$$

$$\leq$$

$$\frac{\ln(b/a)}{2}\left[m_1\left|f'\left(a^{\frac{1}{m_1}}\right)\right|\left(\frac{L(a^{3q},b^{3q})}{-\frac{a^{3q}\Gamma(\alpha+1,3q(\ln a-\ln b))-a^{3q}\Gamma(\alpha+1,0)}{(3q)^{\alpha+1}(\ln a-\ln b)^\alpha(\ln b-\ln a)}}\right)\right.$$

$$\left.+m_2\left|f'\left(b^{\frac{1}{m_2}}\right)\right|\left(\frac{a^{3q}\Gamma(\alpha+1,3q(\ln a-\ln b))-a^{3q}\Gamma(\alpha+1,0)}{(3q)^{\alpha+1}(\ln a-\ln b)^\alpha(\ln b-\ln a)}\right)\right]^{\frac{1}{q}},$$

where L is the logarithmic mean, Γ is the Gamma function and $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. From both Lemma 1, Hölder integral inequality and the (α,m_1,m_2)-GA-convexity of the function $|f'|^q$ on the interval $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$, we get

$$\left|\frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\leq \frac{\ln(b/a)}{2}\left(\int_0^1 1dt\right)^{\frac{1}{p}}$$

$$\cdot\left[\int_0^1 a^{3q(1-t)}b^{3qt}\left|f'\left(\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)}\left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right)\right|^q dt\right]^{\frac{1}{q}}$$

$$\leq \frac{\ln(b/a)}{2}\left(\int_0^1 a^{3(1-t)q}b^{3tq}\left[\begin{array}{c}m_1(1-t^\alpha)\left|f'\left(a^{\frac{1}{m_1}}\right)\right|^q\\+m_2 t^\alpha\left|f'\left(b^{\frac{1}{m_2}}\right)\right|^q\end{array}\right]dt\right)^{\frac{1}{q}}$$

$$= \frac{ln(b/a)}{2} \left[\begin{array}{l} m_1 \left| f'\left(a^{\frac{1}{m_1}}\right)\right|^q \int_0^1 (1-t^\alpha)a^{3q(1-t)}b^{3qt}dt \\ +m_2 \left| f'\left(b^{\frac{1}{m_2}}\right)\right|^q \int_0^1 t^\alpha a^{3q(1-t)}b^{3qt}dt \end{array} \right]^{\frac{1}{q}}$$

$$=$$

$$\frac{ln(b/a)}{2} \left[m_1 \left| f'\left(a^{\frac{1}{m_1}}\right)\right| \left(-\frac{a^{3q}\Gamma(\alpha+1,3q(lna-lnb))-a^{3q}\Gamma(\alpha+1,0)}{(3q)^{\alpha+1}(lna-lnb)^\alpha(lnb-lna)} \right) \right.$$

$$\left. +m_2 \left| f'\left(b^{\frac{1}{m_2}}\right)\right| \left(\frac{a^{3q}\Gamma(\alpha+1,3q(lna-lnb))-a^{3q}\Gamma(\alpha+1,0)}{(3q)^{\alpha+1}(lna-lnb)^\alpha(lnb-lna)}\right) \right]^{\frac{1}{q}}.$$

This completes the proof of the theorem.

Corollary 12. *By considering the conditions of Theorem 10, if we take $m_1 = m_2 = 1$ in the inequality (4.4), then we get*

$$\left|\frac{b^2f(a)-a^2f(b)}{2} - \int_a^b xf(x)dx\right| \leq \frac{ln(b/a)}{2}$$

$$\cdot \left[|f'(a)| \left(L(a^{3q},b^{3q}) - \frac{a^{3q}\Gamma(\alpha+1,3q(lna-lnb))-a^{3q}\Gamma(\alpha+1,0)}{(3q)^{\alpha+1}(lna-lnb)^\alpha(lnb-lna)} \right) \right.$$

$$\left. +|f'(b)| \left(\frac{a^{3q}\Gamma(\alpha+1,3q(lna-lnb))-a^{3q}\Gamma(\alpha+1,0)}{(3q)^{\alpha+1}(lna-lnb)^\alpha(lnb-lna)}\right) \right]^{\frac{1}{q}}.$$

Corollary 13. *By considering the conditions of Theorem 10, if we take $m_1 = m_2 = 1$ and $\alpha = 1$ in the inequality (4.4), then we get*

$$\left|\frac{b^2f(a)-a^2f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\leq \frac{ln(b/a)}{2} \left[|f'(a)| \left(\frac{L(a^{3q},b^{3q})-a^{3q}}{3q(lnb-lna)}\right) + \right.$$

$$\left. |f'(b)| \left(\frac{b^{3q}-L(a^{3q},b^{3q})}{3q(lnb-lna)}\right) \right]^{\frac{1}{q}}.$$

Theorem 11. *Let the function $f:\mathbb{R}_0 = [0,\infty) \to \mathbb{R}$ be a differentiable function and $f' \in L[a,b]$ for $0 < a < b < \infty$. If $|f'|^q$ is (α,m_1,m_2)-GA convex function on the interval $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$ for $[\alpha, m_1, m_2] \in (0,1]^3$ and $q > 1$, then the following integral inequalities hold*

$$\left|\frac{b^2f(a)-a^2f(b)}{2} - \int_a^b xf(x)dx\right| \qquad (4.5)$$

$$\leq \frac{lnb-lna}{2} \left[\frac{L(a^{3p},b^{3p})-a^{3p}}{3(lnb-lna)}\right]^{\frac{1}{p}}$$

$$\cdot \left[\left| f'\left(a^{\frac{1}{m_1}}\right)\right|^q \left(\frac{\alpha(\alpha+3)m_1}{2(\alpha^2+3\alpha+2)}\right) + \left| f'\left(b^{\frac{1}{m_2}}\right)\right|^q \left(\frac{m_2}{\alpha^2+3\alpha+2}\right) \right]^{\frac{1}{q}}$$

$$+ \frac{lnb-lna}{2} \left[\frac{b^{3p}-L(a^{3p},b^{3p})}{3(lnb-lna)}\right]^{\frac{1}{p}}$$

$$\cdot \left[m_1 \left| f'\left(a^{\frac{1}{m_1}}\right)\right|^q \left(\frac{\alpha}{2(\alpha+2)}\right) + m_2 \left| f'\left(b^{\frac{1}{m_2}}\right)\right|^q \left(\frac{1}{\alpha+2}\right) \right]^{\frac{1}{q}},$$

where L is the logarithmic mean and $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. From Lemma 1, Hölder-İşcan integral inequality and the (α, m_1, m_2)-GA convexity of the function $|f'|^q$ on the interval $\left[0, max\left\{a^{\frac{1}{m_1}}, b^{\frac{1}{m_2}}\right\}\right]$, we obtain

$$\left|\frac{b^2f(a)-a^2f(b)}{2} - \int_a^b xf(x)dx\right|$$

$$\leq \frac{lnb-lna}{2} \left[\int_0^1 (1-t)\left(a^{3(1-t)}b^{3t}\right)^p dt\right]^{\frac{1}{p}}$$

$$\cdot \left[\int_0^1 (1-t)\left| f'\left(\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)} \left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right)\right|^q dt\right]^{\frac{1}{q}}$$

$$+ \frac{lnb-lna}{2} \left[\int_0^1 t\left(a^{3(1-t)}b^{3t}\right)^p dt\right]^{\frac{1}{p}}$$

$$\times \left[\int_0^1 t\left| f'\left(\left(a^{\frac{1}{m_1}}\right)^{m_1(1-t)} \left(b^{\frac{1}{m_2}}\right)^{m_2 t}\right)\right|^q dt\right]^{\frac{1}{q}}$$

$$\leq \frac{lnb-lna}{2} \left[\int_0^1 (1-t)a^{3p(1-t)}b^{3pt}dt\right]^{\frac{1}{p}}$$

$$\times \left(\int_0^1 \left[\begin{array}{l} m_1(1-t)(1-t^\alpha)\left| f'\left(a^{\frac{1}{m_1}}\right)\right|^q \\ +m_2(1-t)t^\alpha \left| f'\left(b^{\frac{1}{m_2}}\right)\right|^q \end{array}\right] dt\right)^{\frac{1}{q}}$$

$$+ \frac{lnb-lna}{2} \left[\int_0^1 ta^{3p(1-t)}b^{3pt}dt\right]^{\frac{1}{p}}$$

$$\cdot \left[\int_0^1 \left[m_1 t(1-t^\alpha) \left| f'\left(a^{\frac{1}{m_1}}\right) \right|^q + \right. \right.$$

$$\left. \left. m_2 t t^\alpha \left| f'\left(b^{\frac{1}{m_2}}\right) \right|^q \right] dt \right]^{\frac{1}{q}}$$

$$= \frac{lnb-lna}{2} \left[\frac{L(a^{3p},b^{3p})-a^{3p}}{3(lnb-lna)} \right]^{\frac{1}{p}}$$

$$\cdot \left[\left| f'\left(a^{\frac{1}{m_1}}\right) \right|^q \left(\frac{\alpha(\alpha+3)m_1}{2(\alpha^2+3\alpha+2)} \right) + \left| f'\left(b^{\frac{1}{m_2}}\right) \right|^q \left(\frac{m_2}{\alpha^2+3\alpha+2} \right) \right]^{\frac{1}{q}}$$

$$+ \frac{lnb-lna}{2} \left[\frac{b^{3p}-L(a^{3p},b^{3p})}{3(lnb-lna)} \right]^{\frac{1}{p}}$$

$$\cdot \left[m_1 \left| f'\left(a^{\frac{1}{m_1}}\right) \right|^q \left(\frac{\alpha}{2(\alpha+2)} \right) + m_2 \left| f'\left(b^{\frac{1}{m_2}}\right) \right|^q \left(\frac{1}{\alpha+2} \right) \right]^{\frac{1}{q}}.$$

This completes the proof of the theorem.

Corollary 14. *By considering the conditions of Theorem 11, if we take $m_1 = m_2 = 1$ in the inequality (4.5), then we get*

$$\left| \frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b x f(x)dx \right|$$

$$\leq \frac{lnb-lna}{2} \left[\frac{L(a^{3p},b^{3p})-a^{3p}}{3(lnb-lna)} \right]^{\frac{1}{p}}$$

$$\cdot \left[|f'(a)|^q \left(\frac{\alpha(\alpha+3)}{2(\alpha^2+3\alpha+2)} \right) + |f'(b)|^q \left(\frac{1}{\alpha^2+3\alpha+2} \right) \right]^{\frac{1}{q}}$$

$$+ \frac{lnb-lna}{2} \left[\frac{b^{3p}-L(a^{3p},b^{3p})}{3(lnb-lna)} \right]^{\frac{1}{p}}$$

$$\times \left[|f'(a)|^q \left(\frac{\alpha}{2(\alpha+2)} \right) + |f'(b)|^q \left(\frac{1}{\alpha+2} \right) \right]^{\frac{1}{q}},$$

Corollary 15. *By considering the conditions of Theorem 11, if we take $m_1 = m_2 = 1$ and $\alpha = 1$ in the inequality (4.5), then we get*

$$\left| \frac{b^2 f(a)-a^2 f(b)}{2} - \int_a^b x f(x)dx \right|$$

$$\leq \frac{lnb-lna}{2} \left[\frac{L(a^{3p},b^{3p})-a^{3p}}{3(lnb-lna)} \right]^{\frac{1}{p}} \left[\frac{|f'(a)|^q}{3} + |f'(b)|^q \left(\frac{1}{6} \right) \right]^{\frac{1}{q}}$$

$$+ \frac{lnb-lna}{2} \left[\frac{b^{3p}-L(a^{3p},b^{3p})}{3(lnb-lna)} \right]^{\frac{1}{p}} \left[\frac{|f'(a)|^q}{6} + \frac{|f'(b)|^q}{3} \right]^{\frac{1}{q}}.$$

5. CONCLUSION

New Hermite-Hadamard type integral inequalities can be obtained by using (α, m_1, m_2)-GA convexity and different type identities.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] M.K. Bakula, M.E. Özdemir, and J. Pečarić, "Hadamard type inequalities for m-convex and (α, m)-convex functions," J. Inequal. Pure Appl. Math. 9(4), Art. 96, 12 pages, 2008.

[2] S.S. Dragomir, "Inequalities of Hermite-Hadamard type for GA-convex functions," Annales Mathematicae Silesianae. 32(1). Sciendo, 2018.

[3] S.S. Dragomir and RP. Agarwal, "Two inequalities for differentiable mappings and applications to special means of real numbers and to trapezoidal formula," Appl. Math. Lett. 11, 91-95, 1998.

[4] S.S. Dragomir and C.E.M. Pearce, "Selected Topics on Hermite-Hadamard Inequalities and Its Applications," RGMIA

Monograph, 2002.

[5] S.S. Dragomir, J. Pečarić and LE.Persson, "Some inequalities of Hadamard Type," Soochow Journal of Mathematics, 21(3), pp. 335-341, 2001.

[6] J. Hadamard, "Étude sur les propriétés des fonctions entières en particulier d'une fonction considérée par Riemann," J. Math. Pures Appl. 58, 171-215, 1893.

[7] İ. İşcan, "New refinements for integral and sum forms of Hölder inequality," 2019:304, 11 pages, 2019.

[8] İ. İşcan and M. Kunt, "Hermite-Hadamard-Fejer type inequalities for quasi-geometrically convex functions via fractional integrals, Journal of Mathematics," Volume 2016, Article ID 6523041, 7 pages, 2016.

[9] A.P. Ji, T.Y. Zhang, F. Qi, "Integral inequalities of Hermite-Hadamard type for (α, m)-GA-convex functions," arXiv preprint arXiv:1306.0852, 4 June 2013.

[10] H. Kadakal, "Hermite-Hadamard type inequalities for trigonometrically convex functions," Scientific Studies and Research. Series Mathematics and Informatics, 28(2), 19-28, 2018.

[11] H. Kadakal, "New Inequalities for Strongly r-Convex Functions," Journal of Function Spaces, Volume 2019, Article ID 1219237, 10 pages, 2019.

[12] H. Kadakal, "(m_1, m_2)-convexity and some new Hermite-Hadamard type inequalities," International Journal of Mathematical Modelling and Computations, 09(04), Fall, 297- 309, 2019.

[13] H. Kadakal, "(α, m_1, m_2)-convexity and some inequalities of Hermite-Hadamard type," Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 68(2), 2128-

2142, 2019.

[14] M. Kadakal, "(m_1, m_2)-geometric arithmetically convex functions and related inequalities," Mathematical Sciences and Applications E-Notes, (Submitted to journal), 2020.

[15] M. Kadakal, H. Kadakal and İ. İşcan, "Some new integral inequalities for n-times differentiable s-convex functions in the first sense," Turkish Journal of Analysis and Number Theory, 5(2), 63-68, 2017.

[16] V.G. Miheşan, "A generalization of the convexity," Seminar on Functional Equations, Approx. Convex, Cluj-Napoca, (Romania), 1993.

[17] C.P. Niculescu, "Convexity according to the geometric mean," Math. Inequal. Appl. 3(2), 155-167, 2000.

[18] C.P. Niculescu, "Convexity according to means," Math. Inequal. Appl. 6 (4), 571-579, 2003.

[19] S. Özcan, "Some Integral Inequalities for Harmonically (α, s)-Convex Functions, Journal of Function Spaces," 2019, Article ID 2394021, 8 pages 2019.

[20] S. Özcan, and İ. İşcan, "Some new Hermite-Hadamard type inequalities for s-convex functions and their applications," Journal of Inequalities and Applications, Article number: 2019:201, 2019.

[21] Y. Shuang, Yin, H.P. and Qi, F., "Hermite-Hadamard type integral inequalities for geometric-arithmetically s-convex functions," Analysis, 33, 197-208, 2013.

[22] G. Toader, "Some generalizations of the convexity," Proc. Colloq. Approx. Optim., Univ. Cluj Napoca, Cluj-Napoca, 329-338, 1985.

[23] F. Usta, H. Budak and M.Z. Sarıkaya, "Montgomery identities and Ostrowski type

inequalities for fractional integral operators," Revista de la Real Academia de Ciencias Exactas, Fĭ3053′fsicas y Naturales. Serie A. Matemáticas, 113(2), 1059-1080, 2019.

[24] F. Usta, H. Budak and M.Z. Sarıkaya, "Some New Chebyshev Type Inequalities Utilizing Generalized Fractional Integral Operators," AIMS Mathematics, 5(2), 1147-1161, 2020.

[25] S. Varošanec, "On h-convexity," J. Math. Anal. Appl. 326, 303-311, 2007.

Title: Novel Results based on Generalisation of Some Integral Inequalities for Trigonometrically-P Function

Authors: Sercan TURHAN

Novel Results based on Generalisation of some Integral Inequalities for Trigonometrically-*P* Function

Sercan TURHAN[*1]

Abstract

Trigonometric P-function is defined as a special case of h-convex function. In this article, we used a general lemma that gives trapezoidal, midpoint, Ostrowski, and Simpson type inequalities. With the help of this lemma, we have obtained many integral inequalities and generalisations for trigonometric P-function. We have shown that it goes down to the studies in special cases which are described in our study. Apart from that, we got new results for the trapezoidal, midpoint, Ostrowski, and Simpson type inequalities.

Keywords: Hermite-Hadamard inequality, Simpson-type inequality, Ostrowski-type inequality, Trapezoid-type inequality, Midpoint-type inequality, Trigonometrically-*P* function.

[*]Corresponding Author: sercan.turhan@giresun.edu.tr
[1]The University Of Giresun, Faculty of Science and Arts, Department of Mathematics, Giresun, TURKEY.
 ORCID: https://orcid.org/0000-0002-4392-2182

1. INTRODUCTION

In recent years, many studies on convex functions and integral inequalities have been done and investigated. Firstly, the definition of a convex function is as follows:

Definition 1. *A function* $\Upsilon: I \subset \mathbb{R} \to \mathbb{R}$ *is said to be a convex in the classical sense if, for all* $c, d \in I$ *and* $t \in [0,1]$, *we have*

$$\Upsilon(tc + (1 - t)d) \leq (1 - t)\Upsilon(c) + t\Upsilon(d).$$

In many research fields, the relationship between convexity and inequalities has always been a subject of research. The most important of these is Hermite Hadmard, Ostrowski and Simpson inequalities(see [1, 2, 4, 5, 6, 7, 12, 13, 16, 17, 18, 19, 20, 21, 22]). These famous inequalities are respectively expressed as follows:

$\Upsilon: I \subset \mathbb{R} \to \mathbb{R}$ is the convex function known on the interval I of $\mathbb{R}$ and $c, d \in I$ with $c < d$ as follows:

$$\Upsilon\left(\frac{c + d}{2}\right) \leq \frac{1}{c - d}\int_c^d \Upsilon(x)dx \leq \frac{\Upsilon(c) + \Upsilon(d)}{2}$$

holds. Both inequalities hold in the reversed direction if Υ is concave.

Let $\Upsilon: I \subseteq \mathbb{R} \to \mathbb{R}$ be a mapping differentiable in $I°$, the interior of I, and let c, $d \in I°$ with $c < d$. If $|\Upsilon'(x)| \leq M$, $x \in [c, d]$, then we the following inequality holds

$$\left|\Upsilon(x) - \frac{1}{c - d}\int_c^d \Upsilon(t)dt\right|$$
$$\leq \frac{M}{c - d}\left[\frac{(x - c)^2 + (d - x)^2}{2}\right]$$

for all $x \in [c, d]$. The best possible constant, in the sense that it cannot be replaced by a smaller one, is found $\frac{1}{4}$ in [9].

Let $\Upsilon: [c, d] \to \mathbb{R}$ be a four-times continuously differentiable mapping on (c, d) and $\left\|\Upsilon^{(4)}\right\|_\infty =$

$\sup_{x \in (c,d)} \left|\Upsilon^{(4)}(x)\right| < \infty.$ Then the following inequality holds:

$$\left|\frac{1}{3}\left[\frac{\Upsilon(c) + \Upsilon(d)}{2} + 2\Upsilon\left(\frac{c + d}{2}\right)\right]\right.$$
$$\left. - \frac{1}{d - c}\int_c^d \Upsilon(x)dx\right|$$
$$\leq \frac{1}{2880}\left\|\Upsilon^{(4)}\right\|_\infty (d - c)^2.$$

[14, 15] and therein.

After convexity became so popular, the researchers worked on new classes of convexity. Thus, they applied known integral inequalities to new convexity classes.

Definition 2. [8] *A non-negative function* $\Upsilon: I \to \mathbb{R}$ *is said to be a P-function if the inequality*

$$\Upsilon(tu + (1 - t)v) \leq \Upsilon(u) + \Upsilon(v)$$

holds for all $u, v \in I$, and $t \in [0,1]$. The set of P-functions on the interval I is denoted by $P(I)$.

Definition 3. [23] *Let* $h: J \to \mathbb{R}$ *be a non-negative function,* $h \neq 0$. *We say that* $\Upsilon: I \to \mathbb{R}$ *is an h-convex function, or that* Υ *belongs to the class* $SX(h, I)$, *if* Υ *is non-negative and for all* $u, v \in I, \alpha \in (0,1)$ *we have*

$$\Upsilon(\alpha u + (1 - \alpha)v)$$
$$\leq h(\alpha)\Upsilon(u) + h(1 - \alpha)\Upsilon(v).$$

If this inequality is reversed, then Υ is said to be h-concave, i.e. $\Upsilon \in SV(h, I)$.

In [11], Kadakal gave a different kind of trigonometrically convex function from definition of h-convex function.

Definition 4. [11] *A non-negative function* $\Upsilon: I \to \mathbb{R}$ *is called trigonometrically convex if for every* $u, v \in I$ *and* $t \in [0,1]$,

$$\Upsilon(tu + (1 - t)v) \leq$$

$$\left(sin\frac{\pi t}{2}\right)\Upsilon(u) + cos\left(\frac{\pi t}{2}\right)\Upsilon(v).$$

The class of all trigonometrically convex functions is denoted by $TC(I)$ on interval I. We note that every trigonometrically convex function is a h-convex function for $h(t) = sin\frac{\pi t}{2}$. Moreover, if $\Upsilon(u)$ is a nonnegative function, then every trigonometric convex function is a P-function.

In [3], Bekar obtained the trigonometrically P-function as follows:

Definition 5. [3] *A non-negative function $\Upsilon: I \to \mathbb{R}$ is called trigonometrically P-function if for every $u, v \in I$ and $t \in [0,1]$,*

$$\Upsilon(tu + (1-t)v)$$

$$\leq \left(sin\frac{\pi t}{2}\right) + cos\left(\frac{\pi t}{2}\right)[\Upsilon(u) + \Upsilon(v)].$$

The classes of all trigonometrically P-functions are donated by $TP(I)$ on interval I.

Remark 1. [3] *Clearly, if $\Upsilon(u)$ is a nonnegative function, then every P-function is a trigonometric P-function. Indeed, for every $u, v \in I$ and $t \in [0,1]$ we have*

$$\Upsilon(tu + (1-t)v) \leq \Upsilon(u) + \Upsilon(v)$$

$$\leq \left(sin\frac{\pi t}{2} + cos\frac{\pi t}{2}\right)[\Upsilon(u) + \Upsilon(v)].$$

Example 1. [3] *Non-negative constant functions are trigonometrically P-functions, since $sin\frac{\pi t}{2} + cos\frac{\pi t}{2} \geq 1$ for all $t \in [0,1]$.*

Lemma 1. *Every non-negative trigonometrically convex function is trigonometrically P-function* [3].

In [10], İşcan pointed out the new generalised lemma which is giving many integral inequalities as follows:

Lemma 2. [10] *Let $\Upsilon: I \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable mapping on I° such that $\Upsilon' \in L[c,d]$, where $c, d \in I$ with $c < d$ and $\theta, \lambda \in [0,1]$. Then the following equality holds:*

$$I(c,d;\theta,\lambda) =$$

$$(d-c)\left[-\lambda^2 \int_0^1 (t-\theta)\Upsilon'(tc + (1-t)A_\lambda)dt\right.$$

$$\left. +(1-\lambda)^2 \int_0^1 (t-\theta)\Upsilon'(td + (1-t)A_\lambda)dt\right]$$

where $I(c,d;\theta;\lambda) = (1-\theta)(\lambda\Upsilon(c) + (1-\lambda)\Upsilon(d)) + \theta\Upsilon((1-\lambda)c + \lambda d) - \frac{1}{d-c}\int_c^d \Upsilon(x)dx$ *and* $A_\lambda = (1-\lambda)c + \lambda d$.

We built this study on Lemma 2, where we get different types of integral inequalities. Using this generalised Lemma 2, we have obtained the generalised midpoint, trapezoidal, Simpson and Ostrowski type inequalities for trigonometrically P-function.

2. MAIN RESULTS

It will be referred to $I(c,d;\theta;\lambda) = (1-\theta)(\lambda\Upsilon(c) + (1-\lambda)\Upsilon(d)) + \theta\Upsilon(A_\lambda) - \frac{1}{d-c}\int_c^d \Upsilon(x)dx$, and $A_\lambda = (1-\lambda)c + \lambda d$ where $\theta, \lambda \in [0,1]$, as in this section.

Theorem 1. *Let $\Upsilon: I \to \mathbb{R}$ be a continuously differentiable function, let $c < d$ in I, $\lambda, \theta \in [0,1]$ and assume that $\Upsilon' \in L[c,d]$. If $|\Upsilon'|$ is a trigonometrically P-function on interval $[c,d]$, then the following inequality holds*

$$|I(c,d;\theta;\lambda)| \tag{2.1}$$

$$\leq (d-c)\left(\frac{8}{\pi^2} + \frac{2}{\pi} - \frac{8}{\pi^2}\left[sin\frac{\pi\theta}{2} + cos\frac{\pi\theta}{2}\right]\right)$$

$$[\lambda^2|\Upsilon'(c)| + (\lambda^2 + (1-\lambda)^2)|\Upsilon'((1-\lambda)c + \lambda d)| + (1-\lambda)^2|\Upsilon'(d)|].$$

Proof. If we take the absolute value of both sides of Lemma 2 and then using $|\Upsilon'|$ is trigonometrically P-function, then we have

$$|I(c, d; \theta; \lambda)| \qquad (2.2)$$

$$\leq (d - c) \left[\lambda^2 \int_0^1 |t - \theta| |\Upsilon'(tc + (1 - t)A_\lambda)| dt \right.$$

$$\left. + (1 - \lambda)^2 \int_0^1 |\theta - t| |\Upsilon'(td + (1 - t)A_\lambda)| dt \right]$$

$$\leq (d - c) \left\{ \lambda^2 \int_0^1 |t - \theta| \left(sin\frac{\pi t}{2} \right. \right.$$

$$\left. + cos\frac{\pi t}{2} \right) [|\Upsilon'(a)| + |\Upsilon'(A_\lambda)|] dt$$

$$+ (1 - \lambda)^2 \int_0^1 |\theta - t| \left(sin\frac{\pi t}{2} + cos\frac{\pi t}{2} \right) [|\Upsilon'(d)|$$

$$\left. + |\Upsilon'(A_\lambda)|] dt \right\}$$

$$= (d - c) \left\{ \lambda^2 [|\Upsilon'(c)| + |\Upsilon'(A_\lambda)|] \int_0^1 |t \right.$$

$$\left. - \theta| \left(sin\frac{\pi t}{2} + cos\frac{\pi t}{2} \right) dt \right.$$

$$+ (1 - \lambda)^2 [|\Upsilon'(d)| + |\Upsilon'(A_\lambda)|] \int_0^1 |\theta$$

$$\left. - t| \left(sin\frac{\pi t}{2} + cos\frac{\pi t}{2} \right) dt \right\}.$$

We calculate the integrals as follows:

$$\int_0^1 |t - \theta| \left(sin\frac{\pi t}{2} + cos\frac{\pi t}{2} \right) dt$$

$$= \frac{8}{\pi^2} + \frac{2}{\pi} - \frac{8}{\pi^2} \left[sin\frac{\pi\theta}{2} + cos\frac{\pi\theta}{2} \right]. \qquad (2.3)$$

When the equation (2.3) is written in the inequality, the proof is completed.

Remark 2. *If it is taken* $sin\frac{\pi\theta}{2} + cos\frac{\pi\theta}{2} \geq 1, \theta \in [0,1],$ *we get*

$$|I(c, d; \theta; \lambda)| \leq \frac{2(d - c)}{\pi} \qquad (2.4)$$

$$[\lambda^2 |\Upsilon'(c)| + (\lambda^2 + (1 - \lambda)^2) |\Upsilon'(A_\lambda)| + (1 - \lambda)^2 |\Upsilon'(d)|].$$

Corollary 1. *When θ is taken as 1 in Theorem 1, then we get generalised midpoint-type inequality as follows:*

$$\left| f((1 - \lambda)c + \lambda d) - \frac{1}{d - c} \int_c^d f(x) dx \right|$$

$$\leq \frac{2(d - c)}{\pi} [\lambda^2 |f'(c)|$$

$$+ (\lambda^2 + (1 - \lambda)^2) |f'(A_\lambda)| + (1 - \lambda)^2 |f'(d)|].$$

Corollary 2. *If θ is taken as 1 and $|\Upsilon'(u)| \leq M$, $u \in [c, d]$ in Theorem 1, then we get the following Ostrowski-type inequality*

$$\left| \Upsilon(u) - \frac{1}{d - c} \int_c^d \Upsilon(v) dv \right| \qquad (2.5)$$

$$\leq \frac{4M}{\pi} \left[\frac{(u - c)^2 + (d - u)^2}{d - c} \right]$$

for each $u \in [c, d]$.

Proof. For each $u \in [c, d]$, there exist $\lambda_u \in [0,1]$ such that $u = (1 - \lambda_u)c + \lambda_u d$. Hence we have $\lambda_u = \frac{u - c}{d - c}$ and $1 - \lambda_u = \frac{d - u}{d - c}$. Therefore for each $u \in [c, d]$, from the inequality (2.1) we obtain the inequality (2.5).

Corollary 3. *When θ is taken as 0 in Theorem 1, then we get generalised trapezoid type inequality as follows:*

$$\left| \lambda \Upsilon(c) + (1-\lambda)\Upsilon(d) - \frac{1}{d-c}\int_c^d \Upsilon(x)dx \right|$$

$$\leq \frac{2(d-c)}{\pi}$$

$[\lambda^2|\Upsilon'(c)| + (\lambda^2 + (1-\lambda)^2)|\Upsilon'(A_\lambda)| + (1-\lambda)^2|\Upsilon'(d)|].$

Corollary 4. When θ, λ are taken as $\frac{2}{3}, \frac{1}{2}$, respectively in Theorem 1, then we get Simson-type inequality as follows:

$$\left| \frac{1}{6}\left[\Upsilon(c) + 4\Upsilon\left(\frac{c+d}{2}\right) + \Upsilon(d)\right] \right.$$

$$\left. - \frac{1}{d-c}\int_c^d \Upsilon(x)dx \right|$$

$$\leq \frac{(d-c)}{2}\left(\frac{4(1-\sqrt{3})}{\pi^2} + \frac{2}{\pi}\right)$$

$$\left\{ \frac{|\Upsilon'(c)| + |\Upsilon'(d)|}{2} + \left|\Upsilon'\left(\frac{c+d}{2}\right)\right| \right\}.$$

Corollary 5. When θ, λ are taken as $1, \frac{1}{2}$, respectively in Theorem 1, then we get midpoint-type inequality as follows:

$$\left| \Upsilon\left(\frac{c+d}{2}\right) - \frac{1}{d-c}\int_c^d \Upsilon(x)dx \right|$$

$$\leq \frac{d-c}{\pi}\left\{ \frac{|\Upsilon'(c)| + |\Upsilon'(d)|}{2} + \left|\Upsilon'\left(\frac{c+d}{2}\right)\right| \right\}.$$

Corollary 6. When θ, λ are taken as $0, \frac{1}{2}$, respectively in Theorem 1, then we get Trapezoidal-type inequality as follows:

$$\left| \frac{\Upsilon(c) + \Upsilon(d)}{2} - \frac{1}{d-c}\int_c^d \Upsilon(x)dx \right|$$

$$\leq \frac{d-c}{\pi}\left\{ \frac{|\Upsilon'(c)| + |\Upsilon'(d)|}{2} + \left|\Upsilon'\left(\frac{c+d}{2}\right)\right| \right\}.$$

Theorem 2. *Let $\Upsilon: I \to \mathbb{R}$ be a continuously differentiable function, let $c < d$, $c, d \in I$ and $\lambda, \theta \in [0,1]$ assume that $q > 1$. If $|\Upsilon'|^q$ is a trigonometrically P-function on the interval $[c, d]$, then the following inequality holds*

$$|I(c, d; \theta; \lambda)| \leq \qquad (2.6)$$

$$(d-c)\left(\frac{\theta^{p+1} + (1-\theta)^{p+1}}{p+1}\right)^{\frac{1}{p}}\left(\frac{4}{\pi}\right)^{1/q}$$

$[\lambda^2(|\Upsilon'(c)|^q + |\Upsilon'(A_\lambda)|^q)^{1/q}$

$+(1-\lambda)^2(|\Upsilon'(d)|^q + |\Upsilon'(A_\lambda)|^q)^{1/q}].$

Proof. From Lemma 2 and by Hölder's integral inequality, we have

$$|I(c, d; \theta; \lambda)| \leq (d-c) \qquad (2.7)$$

$$\left[\lambda^2\left(\int_0^1 |t-\theta|^p\right)^{1/p}\left(\int_0^1 |\Upsilon'(tc + (1-t)A_\lambda)|^q dt\right)^{1/q}\right.$$

$$\left. +(1-\lambda)^2\left(\int_0^1 |t-\theta|^p\right)^{1/p}\left(\int_0^1 |\Upsilon'(td + (1-t)A_\lambda)|^q dt\right)^{1/q}\right].$$

Since $|\Upsilon'|^q$ is trigonometrically P-function on $[c, d]$, and by simple computation, we get

$$\int_0^1 |\Upsilon'(tc + (1-t)[(1-\lambda)c + \lambda d])|^q dt \quad (2.8)$$

$$\leq \int_0^1 \left(\sin\frac{\pi t}{2} + \cos\frac{\pi t}{2}\right)[|\Upsilon'(c)|^q$$

$$+ |\Upsilon'(A_\lambda)|^q]dt$$

$$= \frac{4}{\pi}[|\Upsilon'(c)|^q + |\Upsilon'(A_\lambda)|^q]$$

$$\int_0^1 |\Upsilon'(td + (1$$

$$- t)A_\lambda)|^q dt \qquad (2.9)$$

$$\leq \int_0^1 \left(\sin\frac{\pi t}{2} + \cos\frac{\pi t}{2} \right) [|\Upsilon'(d)|^q$$

$$+ |\Upsilon'(A_\lambda)|^q] dt$$

$$= \frac{4}{\pi} [|\Upsilon'(d)|^q + |\Upsilon'(A_\lambda)|^q]$$

and

$$\int_0^1 |t - \theta|^p dt = \frac{\theta^{p+1} + (1 - \theta)^{p+1}}{p + 1}. \qquad (2.10)$$

Thus, substitute (2.8)-(2.10) in (2.7), we obtain the inequality (2.6). This completes the proof.

Corollary 7. *When θ is taken as 1 in Theorem 2, then we get generalised midpoint-type inequality as follows:*

$$\left| \Upsilon((1 - \lambda)c + \lambda d) - \frac{1}{d - c} \int_a^b \Upsilon(x) dx \right|$$

$$\leq (b - a) \left(\frac{1}{p + 1} \right)^{\frac{1}{p}} \left(\frac{4}{\pi} \right)^{1/q}$$

$$[\lambda^2 (|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q)^{1/q}$$

$$+ (1 - \lambda)^2 (|\Upsilon'(b)|^q + |\Upsilon'(A_\lambda)|^q)^{1/q}].$$

Corollary 8. *When θ is taken as 0 in Theorem 2, then we get generalised trapezoidal-type inequality as follows:*

$$\left| \lambda \Upsilon(a) + (1 - \lambda)\Upsilon(b) - \frac{1}{b - a} \int_a^b \Upsilon(x) dx \right|$$

$$\leq (b - a) \left(\frac{1}{p + 1} \right)^{\frac{1}{p}} \left(\frac{4}{\pi} \right)^{1/q}$$

$$[\lambda^2 (|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q)^{1/q}$$

$$+ (1 - \lambda)^2 (|\Upsilon'(b)|^q + |\Upsilon'(A_\lambda)|^q)^{1/q}].$$

Corollary 9. *If θ is taken as 1 and $|\Upsilon'(x)| \leq M$, $x \in [a, b]$ in Theorem 2, then we get the following Ostrowski-type inequality*

$$\left| \Upsilon(x) - \frac{1}{b - a} \int_a^b \Upsilon(u) du \right| \qquad (2.11)$$

$$\leq \frac{M}{(p + 1)^{\frac{1}{p}}} \left(\frac{8}{\pi} \right)^{1/q} \left[\frac{(x - a)^2 + (b - x)^2}{b - a} \right]$$

for each $x \in [a, b]$.

Proof. For each $x \in [a, b]$, there exist $\lambda_x \in [0,1]$ such that $x = (1 - \lambda_x)a + \lambda_x b$. Hence we have $\lambda_x = \frac{x-a}{b-a}$ and $1 - \lambda_x = \frac{b-x}{b-a}$. Therefore for each $x \in [a, b]$, from the inequality (2.6) we obtain the inequality (2.11).

Corollary 10. *When θ, λ are taken as $\frac{2}{3}, \frac{1}{2}$, respectively in Theorem 2, then we get Simson-type inequality as follows:*

$$\left| \frac{1}{6} \left[\Upsilon(a) + 4\Upsilon\left(\frac{a + b}{2} \right) + \Upsilon(b) \right] \right.$$

$$\left. - \frac{1}{b - a} \int_a^b \Upsilon(x) dx \right|$$

$$\leq \frac{b - a}{4} \left(\frac{2^{p+1} + 1}{3^{p+1}(p + 1)} \right)^{1/p} \left(\frac{4}{\pi} \right)^{1/q}$$

$$\left\{\left[|\Upsilon'(a)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right.$$

$$\left.+\left[|\Upsilon'(b)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right\}.$$

Corollary 11. *When* θ, λ *are taken as 1,* $\frac{1}{2}$, *respectively in Theorem 2, then we get midpoint-type inequality as follows:*

$$\left|\Upsilon\left(\frac{a+b}{2}\right) - \frac{1}{b-a}\int_a^b \Upsilon(x)dx\right|$$

$$\leq \frac{b-a}{4}\left(\frac{1}{p+1}\right)^{\frac{1}{p}}\left(\frac{4}{\pi}\right)^{\frac{1}{q}}$$

$$\left\{\left[|\Upsilon'(a)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right.$$

$$\left.+\left[|\Upsilon'(b)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right\}.$$

Corollary 12. *When* θ, λ *are taken as 0,* $\frac{1}{2}$, *respectively in Theorem 2, then we trapezoid-type inequality as follows:*

$$\left|\frac{\Upsilon(a)+\Upsilon(b)}{2} - \frac{1}{b-a}\int_a^b \Upsilon(x)dx\right|$$

$$\leq \frac{b-a}{4}\left(\frac{1}{p+1}\right)^{\frac{1}{p}}\left\{\left[|\Upsilon'(a)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right.$$

$$\left.+\left[|\Upsilon'(b)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right\}.$$

Theorem 3. *Let* $\Upsilon: I \subseteq \mathbb{R} \to \mathbb{R}$ *be a continuously differentiable function,* $\Upsilon' \in L[a,b]$, *where* $a, b \in I°$ *with* $a < b$ *and* $\theta, \lambda \in [0,1]$. *If* $|\Upsilon'|^q$ *is trigonometrically P-function on* $[a,b]$, $q > 1$, *then the following inequality holds*

$$|I(a,b;\theta;\lambda)| \tag{2.12}$$

$$\leq (b-a)\left[\theta^2 - \theta + \frac{1}{2}\right]^{1-\frac{1}{q}}\left(\frac{8}{\pi^2}+\frac{2}{\pi}\right.$$

$$\left.-\frac{8}{\pi^2}\left(sin\frac{\pi\theta}{2} + cos\frac{\pi\theta}{2}\right)\right)^{1/q}$$

$$\{\lambda^2[|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}$$

$$+(1-\lambda)^2[|\Upsilon'(b)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}\}.$$

Proof. We proceed similarly as in the proof Theorem 2. Since $|\Upsilon'|^q$ is trigonometrically $P-$function on $[a,b]$ and using the power mean inequality, we get

$$|I(a,b;\theta;\lambda)| \leq (b-a) \tag{2.13}$$

$$\left[\lambda^2\left(\int_0^1 |t-\theta|dt\right)^{1-\frac{1}{q}}\left(\int_0^1 |t-\theta|\right.\right.$$

$$\left.|\Upsilon'(ta + (1-t)[(1-\lambda)a + \lambda b])|^q dt)^{1/q}\right.$$

$$(1-\lambda)^2\left(\int_0^1 |t-\theta|dt\right)^{1-\frac{1}{q}}\left(\int_0^1 |t-\theta|\right.$$

$$\left.\left.|\Upsilon'(tb + (1-t)[(1-\lambda)a + \lambda b])|^q dt)^{1/q}\right]\right.$$

$$\leq (b-a)\left(\theta^2 - \theta + \frac{1}{2}\right)^{1-\frac{1}{q}}\left(\frac{8}{\pi^2}+\frac{2}{\pi}\right.$$

$$\left.-\frac{8}{\pi^2}\left(sin\frac{\pi\theta}{2} + cos\frac{\pi\theta}{2}\right)\right)^{1/q}$$

$$\{\lambda^2[|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}$$

$$(1-\lambda)^2[|\Upsilon'(b)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}\}.$$

This completes the proof.

Remark 3. *If it is taken* $sin\frac{\pi\theta}{2} + cos\frac{\pi\theta}{2} \geq 1$, $\theta \in [0,1]$ *in the inequality (2.12), we get*

$$|I(a, b; \theta; \lambda)| \qquad (2.14)$$

$$\leq (b-a)\left(\theta^2 - \theta + \frac{1}{2}\right)^{1-\frac{1}{q}}\left(\frac{2}{\pi}\right)^{1/q}$$

$$\{\lambda^2[|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}$$

$$+(1-\lambda)^2[|\Upsilon'(b)| + |\Upsilon'(A_\lambda)|^q]^{1/q}\}.$$

Corollary 13. *When θ is taken as 0 in Theorem 3, then we get generalised trapezoidal-type inequality as follows:*

$$\left|\lambda\Upsilon(a) + (1-\lambda)\Upsilon(b) - \frac{1}{b-a}\int_a^b \Upsilon(x)dx\right|$$

$$\leq \frac{b-a}{2}\left(\frac{4}{\pi}\right)^{\frac{1}{q}}\{\lambda^2[|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}$$

$$+(1-\lambda)^2[|\Upsilon'(b)|^q + |\Upsilon'((A_\lambda)|^q]^{1/q}\}.$$

Corollary 14. *When θ is taken as 1 in Theorem 3, then we get generalised midpoint-type inequality as follows:*

$$\left|\Upsilon((1-\lambda)a + \lambda b) - \frac{1}{b-a}\int_a^b \Upsilon(x)dx\right|$$

$$\leq \frac{b-a}{2}\left(\frac{4}{\pi}\right)^{\frac{1}{q}}\{\lambda^2[|\Upsilon'(a)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}$$

$$+(1-\lambda)^2[|\Upsilon'(b)|^q + |\Upsilon'(A_\lambda)|^q]^{1/q}\}.$$

Corollary 15. *If θ is taken as 1 and $|\Upsilon'(x)| \leq M$, $x \in [a, b]$ in Theorem 3, then we get the following Ostrowski-type inequality*

$$\left|\Upsilon(x) - \frac{1}{b-a}\int_a^b \Upsilon(u)du\right|$$

$$\leq \frac{M}{2}\left(\frac{8}{\pi}\right)^{\frac{1}{q}}\left[\frac{(x-a)^2 + (b-x)^2}{b-a}\right]. \qquad (2.15)$$

Proof. For each $x \in [a, b]$, there exist $\lambda_x \in [0,1]$ such that $x = (1-\lambda_x)a + \lambda_x b$. Hence we have $\lambda_x = \frac{x-a}{b-a}$ and $1 - \lambda_x = \frac{b-x}{b-a}$. Therefore for each $x \in [a, b]$, from the inequality (2.12) we obtain the inequality (2.15).

Corollary 16. *When θ, λ are taken as $0, \frac{1}{2}$, respectively in Theorem 3, then we trapezoid-type inequality as follows:*

$$\left|\frac{\Upsilon(a) + \Upsilon(b)}{2} - \frac{1}{b-a}\int_a^b \Upsilon(x)dx\right|$$

$$\leq \frac{b-a}{8}\left(\frac{4}{\pi}\right)^{\frac{1}{q}}\left\{\left[|\Upsilon'(a)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{1/q}\right.$$

$$+\left.\left[|\Upsilon'(b)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{1/q}\right\}.$$

Corollary 17. *When θ, λ are taken as $1, \frac{1}{2}$, respectively in Theorem 3, then we get midpoint-type inequality as follows:*

$$\left|\Upsilon\left(\frac{a+b}{2}\right) - \frac{1}{b-a}\int_a^b \Upsilon(x)dx\right| \leq \frac{b-a}{8}\left(\frac{4}{\pi}\right)^{\frac{1}{q}}$$

$$\left\{\left[|\Upsilon'(a)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right.$$

$$+\left.\left[|\Upsilon'(b)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q\right]^{\frac{1}{q}}\right\}.$$

Corollary 18. *When θ, λ are taken as $\frac{2}{3}, \frac{1}{2}$, respectively in Theorem 3, then we get Simson-type inequality as follows:*

$$\left| \frac{1}{6}\left[\Upsilon(a) + 4\Upsilon\left(\frac{a+b}{2}\right) + \Upsilon(b) \right] \right.$$

$$\left. - \frac{1}{b-a}\int_a^b \Upsilon(x)dx \right|$$

$$\leq \frac{b-a}{4}\left(\frac{5}{18}\right)^{1-\frac{1}{q}}\left[\frac{4(1-\sqrt{3})}{\pi^2} + \frac{2}{\pi}\right]^{\frac{1}{q}}$$

$$\left\{ \left[|\Upsilon'(a)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q \right]^{\frac{1}{q}} \right.$$

$$\left. + \left[|\Upsilon'(b)|^q + \left|\Upsilon'\left(\frac{a+b}{2}\right)\right|^q \right]^{\frac{1}{q}} \right\}.$$

3. CONCLUSION

In this study, we applied the trapezoidal, midpoint, Ostrowski, and Simpson type inequalities for Trigonometrically *P*-function by using a general lemma given by İ. İşcan [10]. As a result, we obtain integral inequalities of type trapezoidal, midpoint, Ostrowski and Simpson for Trigonometrically P-function. Our results can be applied to different types of convexity.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics*.*

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] G. A. Anastassiou, "Complex Multivariate Montgomery Type Identity Leading to Complex Multivariate Ostrowski and Grüss Inequalities", Communications in Advanced Mathematical Sciences, vol. II, no. 2, pp. 161-175, 2019.

[2] W. W. Breckner, "Stetigkeitsaussagen für eine Klasse verallgemeinerter konvexer funktionen in topologischen linearen Räumen", Publ. Inst. Math., vol. 23, pp. 13–20, 1978.

[3] K. Bekar, "Hermite-Hadamard type inequalities for trigonometrically P-functions", Comptes rendus de l'Académie bulgare des Sciences, vol. 72, no. 11, pp. 1449-1457, 2019.

[4] H. Budak, F. Usta and M. Z. Sarikaya, "Refinements of the Hermite–Hadamard inequality for coordinated convex mappings", Journal of Applied Analysis, vol. 25, no. 1, pp. 73-81, 2019.

[5] H. Budak, F. Usta, M. Z. Sarikaya, M. E. Ozdemir, "On generalization of midpoint type inequalities with generalized fractional integral operators", Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas, vol. 113, no. 2, pp. 769-790, 2019.

[6] H. Budak, F. Usta, "New Upper Bounds of Ostrowski Type Integral Inequalities Utilizing Taylor Expansion", Hacettepe Journal of Mathematics and Statistics, vol. 47, no. 3, pp. 567-578, 2018.

[7] S. S. Dragomir, C. E. M. Pearce, "Selected Topics on Hermite-Hadamard Inequalities and Applications", RGMIA Monographs, Victoria University, 2000.

[8] S. S. Dragomir, J. Pecaric, L. E. Persson, "Some inequalities of Hadamard Type", Soochow Journal of Mathematics, vol. 21, no. 3, pp. 335–341, 2001.

[9]   S. S. Dragomir, Th. M. Rassias, "Ostrowski type inequalities and applications in numerical integration", Kluwer Academic Publishers, Dorcdrecht, Boston, London, 2002.

[10] İ. İşcan, New estimates on generalization of some integral inequalities for s-convex functions and their applications, Int. J. Pure Appl. Math., vol. 86, no. 4, pp. 727-746, 2013.

[11] H. Kadakal, "Hermite–Hadamard type inequalities for trigonometrically convex functions", Scientific Studies and Research. Series Mathematics and Informatics, vol. 28, no. 2, pp. 19–28, 2018.

[12] M. Kadakal, İ. İşcan, Inequalities of Hermite-Hadamard and Bullen Type for AH -Convex Functions. Universal Journal of Mathematics and Applications, vol. 2, no. 3, pp. 152-158, 2019.

[13] M. Kadakal, "Better Results for Trigonometrically Convex Functions via Hölder-İşcan and Improved Power-Mean Inequalities", Universal Journal of Mathematics and Applications, vol. 3, no. 1, pp. 38-43, 2020.

[14] M. Z. Sarikaya, E. Set, M. E. Özdemir, "On new inequalities of Simpson's type for convex functions", Computers & Mathematics with Applications, vol. 60, no. 8, pp. 2191-2199, 2010.

[15] M. Z. Sarikaya, N. Aktan, "On the generalization of some integral inequalities and their applications", Mathematical and Computer Modelling, vol. 54, no. 9-10, pp. 2175-2182, 2011.

[16] F. Usta, M. Z. Sarıkaya, "Explicit Bounds on Certain Integral Inequalities via Conformable Fractional Calculus, Cogent Mathematics, vol. 4, no. 1, 1277505, 2017.

[17] F. Usta, "On generalization of trapezoid type inequalities for s-convex functions with generalized fractional integral operators", Filomat, vol. 32, no. 6, 2018.

[18] F. Usta, M. Z. Sarıkaya, "On Generalization of Pachpatte Type Inequalities for Conformable Fractional Integral, TWMS Journal of Applied and Engineering Mathematics, vol. 8, no. 1, 106, 2018.

[19] F. Usta, M. Z. Sarıkaya, "On Bivariate Retarded Integral Inequalities and Their Applications" Facta Universitatis, Series: Mathematics and Informatics, vol. 34, no. 3, pp. 553-561, 2019.

[20] F. Usta, H. Budak, M. Z. Sarıkaya, "Montgomery Identities and Ostrowski type Inequalities for Fractional Integral Operators", Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas, vol. 113, no. 2, pp. 1059-1080, 2019.

[21] F. Usta, H. Budak, F. Ertuğral, M. Z. Sarıkaya, "The Minkowski's Inequalities Utilizing Newly Defined Generalized Fractional Integral Operators, Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, vol. 68, no. 1, pp. 686-701, 2019.

[22] F. Usta, H. Budak, M. Z. Sarıkaya, "Some New Chebyshev Type Inequalities Utilizing Generalized Fractional Integral Opetors", AIMS Mathematics, vol. 5, no. 2, 2020.

[23] S. Varošanec, "On h-convexity", J. Math. Anal. Appl., vol. 326, no. 1, pp. 303–311, 2007.

Title: On Strongly π-regular Modules

On Strongly π-regular Modules

Suat KOÇ[*1]

Abstract

In this article, we introduce the notion of strongly π-regular module which is a generalization of von Neumann regular module in the sense [13]. Let A be a commutative ring with $1 \neq 0$ and X a multiplication A-module. X is called a strongly π-regular module if for each $x \in X$, $(Ax)^m = cX = c^2X$ for some $c \in A$ and $m \in \mathbb{N}$. In addition to give many properties and examples of strongly π-regular modules, we also characterize certain class of modules such as von Neumann regular modules and second modules in terms of this new class of modules. Also, we determine when the localization of any family of submodules at a prime ideal commutes with the intersection of this family.

Keywords: von Neumann regular module, (m,n)-closed ideal, strongly π-regular module, Krull dimension, $(*)$-property, localization.

1. INTRODUCTION

In this article, all rings under consideration will be assumed to be commutative with $1 \neq 0$ and all modules will be nonzero unital. Let A will always denote such a ring and X will denote such an A-module. The set of all prime ideals and maximal ideals will be denoted by $Spec(A)$ and $Max(A)$, respectively. A ring A is said to be a *von Neumann regular (briefly, vn-regular) ring* if for each $c \in A$, there exists $d \in A$ such that $c = c^2d$. In that case the principal ideal $(c) = (a)$ for some idempotent element $a \in A$ [22]. It is clear that A is a vn-regular ring iff $(c) = (c)^2$ for each $c \in A$. The concept of vn-regular ring and its

generalizations have been widely studied in many papers. See, for instance, [2], [8], [9], [10], [12], [24]. One of the important generalizations of vn-regular ring is that strongly π-regular ring. A ring A is called a *strongly π-regular ring* if for each $c \in A$, $c^m = c^{m+1}d$ for some $d \in A$ and $m \in \mathbb{N}$, or equivalently, the descending chain $(c) \supseteq (c)^2 \supseteq \cdots \supseteq (c)^m \supseteq \cdots$ of principal powers terminates at some step [14]. It is clear that a ring A is a strongly π-regular iff for each $c \in A$, there exist $m \in \mathbb{N}$ and $d \in A$ such that $(c)^m = (d) = (d)^2$. In that case $(c)^m$ is generated by an idempotent element $a \in A$. It was shown, in [8], that a ring A is strongly π-regular iff its Krull dimension $\dim(A) = 0$ iff the radical commutes

with intersection of any family of ideals (See, [8, Theorem 3] and [8, Theorem 4]).

Recently, Jayaram and Tekir, in their paper [13], extended the notion of vn-regular rings to modules by defining X-vn-regular elements and weak idempotent elements. An element $c \in A$ is said to be an *X-vn-regular element* if $cX = c^2X$ and $d \in A$ is said to be a weak idempotent if $dx = d^2x$ for each $x \in X$, namely, $d - d^2 \in ann(X)$, where $ann(X) = \{c \in A : cX = 0\}$ is the annihilator of an A-module X. They defined an A-module X as a *vn-regular module* if for each $x \in X$, the cyclic submodule $Ax = cX$ for some X-vn-regular element $c \in A$. An A-module X is said to be a *multiplication module* if its each submodule Y of X is of the form $Y = JX$ for some ideal J of A [5], [11]. Note that X is a multiplication module iff $Y = (Y:X)X$ for each submodule Y of X, where $(Y:X) = ann(X/Y)$. Also it is clear that each vn-regular module is multiplication.

This article aims to study strongly π-regular modules and use them to characterize certain class of modules such as vn-regular modules, second modules, zero dimensional modules and modules satisfying $(*)$-condition. For the sake of completeness, we shall give some notions which will be frequently used throughout the paper. Let X be an A-module. A proper submodule Y of X is said to be prime if for each $c \in A$ and $x \in X$, $cx \in Y$ implies either $c \in (Y:X)$ or $x \in Y$ [16]. In this case, $q = (Y:X) \in Spec(A)$ and Y is said to be q-prime. Let Y be a submodule of X. The radical of Y, denoted by $rad(Y)$, is the intersection of all prime submodules Q of X containing Y. If there is no such a prime submodule, we say $rad(Y) = X$. Note that in a finitely generated (briefly, f.g.) or multiplication modules, there always exists a prime submodule Q of X containing a given proper submodule Y of X. In a multiplication module X, Ameri in his paper [1], defined the product of two submodules and determined the radical of any given proper submodule Y of X. Let $Y = IX$ and $W = JX$ be two submodules of a multiplication module X. Then the product of Y and W is defined by $YW = (IJ)X$. He proved in [1], that the product is well defined and the radical of a proper submodule Y of X is characterized as follows: $rad(Y) = \{x \in X : \exists m \in \mathbb{N}$ such that $(Ax)^m \subseteq Y\}$. If X is a multiplication module, it is clear that the radical commutes with finite intersection of submodules, that is, $rad(\cap_{i=1}^{n} Y_i) = \cap_{i=1}^{n} rad(Y_i)$. So it is a natural question to ask whether the radical operation commutes (or not) with the infinite intersection. This question was first studied in [6] and continued in [7]. In [6], the authors introduced the $(*)$-property as follows: a multiplication module X is said to satisfy $(*)$-condition if for each family of submodules $\{Y_i\}_{i \in \Delta}$, $x \in rad(Y_i)$ for each $i \in \Delta$ implies $(Ax)^m \subseteq \cap_{i \in \Delta} Y_i$ for some $m \in \mathbb{N}$. They also showed that X satisfies $(*)$-condition iff the radical operation commutes with the infinite intersection of a given family of submodules. Also, recall from [7] that a multiplication module X is said to satisfy descending chain condition on principal powers if the descending chain $Ax \supseteq (Ax)^2 \supseteq \cdots \supseteq (Ax)^m \supseteq \cdots$ stops. The authors in [7] showed that a f.g. multiplication module X satisfies $(*)$-property iff it satisfies descending chain condition on principal powers iff its Krull dimension $\dim(X) = 0$ (See [7, Lemma 3], [7, Corollary 3] and [7, Lemma 5]).

We say that an A-module X as a *strongly π-regular module* if for each $x \in X$, there exist $m \in \mathbb{N}$ and $c \in A$ such that $(Ax)^m = cX = c^2X$. Also, we define an A-module X as a *weak π-regular module* if for each $c \in A$, the descending chain $cX \supseteq c^2X \supseteq \cdots \supseteq c^mX \supseteq \cdots$ stops.

Among other things in this paper, in section 2, we study the relations between strongly π-regular modules (weak π-regular modules) and other certain classes of modules such as vn-regular modules, second modules and divisible modules. In particular, we investigate the behavior of strongly π-regular modules under homomorphism, in factor modules, in localization, in cartesian product of modules (See Proposition 2.16, Corollary 2.17, Proposition 2.20, Proposition 2.21 and Theorem 2.23). Also, we give a characterization of strongly π-regular modules (weak π-regular modules) in terms of strongly π-regular rings (See Proposition 2.5, Proposition 2.10, Proposition 2.14 and Proposition 2.15). Finally, we characterize second modules and vn-regular modules in terms of

strongly π-regular modules (See, Corollary 2.19 and Theorem 2.24).

In section 3, we deal with commutativity of localization at a prime ideal and arbitrary intersection of any family of submodules. Let Y be a q-prime submodule of X. Then we know that localization at q commutes with the finite intersection of a given family of submodules. Thus it is a natural problem to ask whether this property of localization is true when we replace finite intersection by infinite one. This question has a negative answer (See Example 3.1). We show that q-prime submodule Y of a f.g. multiplication module X is a strongly prime submodule iff the localization commutes with the infinite intersection of a given family of submodules (See, Theorem 3.2). Finally, we give the relations between strongly π-regular modules and other certain class of modules such as weak π-regular modules, zero dimensional modules, modules satisfying $(*)$-condition (See, Theorem 3.6).

2. CHARACTERIZATION OF STRONGLY π-REGULAR MODULES

Definition 2.1. A multiplication A-module X is called a strongly π-regular module if for all $x \in X$, $(Ax)^m = cX = c^2X$ for some $c \in A$ and $m \in \mathbb{N}$.

Example 2.2. A ring A is a strongly π-regular ring iff it is a strongly π-regular A-module.

Recall from [20] that an A-module X is said to be simple if zero and X are the only submodules of X.

Example 2.3. All simple modules are strongly π-regular. To see this take a simple A-module X and $x \in X$. Then either $Ax = 0 = 0X$ or $Ax = X = 1X$. Hence, X is a strongly π-regular module.

Example 2.4. All vn-regular modules are clearly strongly π-regular module. But the converse need not be true. For instance, consider the $\mathbb{Z}$-module $\mathbb{Z}_{q^k}$, where q is a prime number and $k \geq 1$ is an integer. Let $\bar{x} \in \mathbb{Z}_{q^k}$. If $\gcd(x, q) = 1$, then we conclude $\mathbb{Z}\bar{x} = \mathbb{Z}_{q^m} = 1\mathbb{Z}_{q^m}$. If $\gcd(x, q) \neq 1$,

then we get $(\mathbb{Z}\bar{x})^m = (\bar{0}) = 0\mathbb{Z}_{q^m}$. But $\mathbb{Z}$-module $\mathbb{Z}_{q^m}$ is not a vn-regular module.

Proposition 2.5. Suppose that X is a f.g. strongly π-regular module. Then $A/ann(X)$ is a strongly π-regular ring.

Proof. Take $q \in Spec(A)$ with $ann(X) \subseteq q$. Then there exist $m \in Max(A)$ with $ann(X) \subseteq q \subseteq m$. Suppose that $q \subsetneq m$. Then $qX \subsetneq mX$ and so there exists $x \in mX - qX$. Since X is strongly π-regular module, $(Ax)^m = cX = c^2X$ for some $c \in A$ and $m \in \mathbb{N}$. Since X is f.g. module, cX is f.g. so by [4, Corollary 2.5], we have $(1 + dc)cX = 0$ for some $d \in A$. This gives that $(1 + dc)c \in ann(X)$. Also by [11, Corollary 2.11], qX is prime and it is clear that $(Ax)^m \nsubseteq qX$ so we get $c \notin (qX:X) = q$. As q is a prime ideal and $c(1 + dc) \in ann(X) \subseteq q$, we conclude that $1 + dc \in q \subseteq m$. Also note that $cX \subseteq mX$. This yields that $c \in m$ and so $1 \in m$, a contradiction. Thus $q = m \in Max(A)$. Therefore $A/ann(X)$ is a strongly π-regular ring by [8, Theorem 3].

Definition 2.6. A-module X is called a weak π-regular module if for each $c \in A$, $c^mX = c^{m+1}X$ for some $m \in \mathbb{N}$.

Example 2.7. A ring A is a strongly π-regular ring iff it is a weak π-regular as an A-module.

Example 2.8. Let X be a f.g. vn-regular module. Then by [13, Theorem 1], every $c \in A$ is an X-vn-regular, namely, $cX = c^2X$. Therefore, every f.g. vn-regular module is weak π-regular.

A module X over an integral domain A is called a divisible module if $cX = X$ for each $0 \neq c \in A$ [17]. Also a module X over a commutative ring A (not necessarily a domain) is said to be a second module if for each $c \in X$, either $cX = 0$ or $cX = X$ [23].

Example 2.9.

(i) Every divisible module over an integral domain is a weak π-regular module.

(ii) Every second module is a weak π-regular module.

Proposition 2.10. If X is a f.g. strongly π-regular module, then X is a weak π-regular module.

Proof. Take $c \in A$. Then by Proposition 2.5, $A/ann(X)$ is a strongly π-regular ring and so $\bar{c}^m = \bar{c}^{m+1}\bar{d}$ for some $d \in A$ and $m \in \mathbb{N}$, where $\bar{c} = c + ann(X)$. This implies that $c^m - c^{m+1}d \in ann(X)$ and so $c^m X = c^{m+1}dX \subseteq c^{m+1}X$. Since the other inclusion always holds, we have $c^m X = c^{m+1}X$. Therefore, X is a weak π-regular module.

The converse of Proposition 2.10 need not be true.

Example 2.11. Consider the $\mathbb{Z}$-module $X = \mathbb{Z}_2 \times \mathbb{Z}_2$ and take any integer c. If c is even, then $cX = 0 = c^2 X$. Also if c is odd, we have $cX = X = c^2 X$. Thus X is a weak π-regular module. Since X is not a multiplication module, X can not be a strongly π-regular module.

The following Lemma is well known in [13] and for the sake of completeness, we remind it here.

Lemma 2.12. Let X be an A-module.

(i) If $c, d \in A$ are weak idempotents, then $c + d(1 - c)$ is a weak idempotent.

(ii) $cX + dX = \big(c + d(1 - c)\big)X$.

(iii) Suppose that X is a f.g. A-module. Then $c \in A$ is X-vn-regular iff $cX = dX$ for some weak idempotent $d \in A$.

Remark 2.13. Let X be a f.g. A-module and $c^m X = c^{m+1}X$ for some $c \in A$ and $m \in \mathbb{N}$. Then clearly we have $c^m X = (c^m)^2 X$ and so c^m is an X-vn-regular element. By Lemma 2.12 (iii), $c^m X = dX$ for some weak idempotent element $d \in A$.

Proposition 2.14. Suppose that X is a f.g. multiplication A-module. If $A/ann(X)$ is a strongly π-regular ring, then X is a strongly π-regular module.

Proof. Take $c \in A$. Since $A/ann(X)$ is a strongly π-regular ring, $c^m X = c^{m+1}X$ for some $m \in \mathbb{N}$ as in the Proof of Proposition 2.10. By Remark 2.13, we have $c^m X = dX$ for some weak idempotent

element $d \in A$. Let $x \in X$. Since X is f.g. multiplication, $Ax = JX$ for some f.g. ideal J of A, where $J = (c_1, c_2, \dots, c_n)$ for some $c_1, c_2, \dots, c_n \in A$. Then we have $Ax = c_1 X + c_2 X + \cdots + c_n X$. Also note that for each $c_i \in A$, there exist $t_i \geq 1$ and weak idempotent element $d_i \in A$ such that $c_i^{t_i}X = d_i X$. Now, put $t = t_1 + t_2 + \cdots + t_n$. Then we get $(Ax)^t = \{(c_1) + (c_2) + \cdots + (c_n)\}^t X = c_1^{t_1}X + c_2^{t_2}X + \cdots + c_n^{t_n}X = d_1 X + d_2 X + \cdots + d_n X = dX = d^2 X$ for some weak idempotent element $d \in A$ by Lemma 2.12 (iii). Therefore, X is a strongly π-regular module.

Proposition 2.15. Suppose that X is a f.g. weak π-regular module. Then $A/ann(X)$ is a strongly π-regular ring.

Proof. Let $\bar{c} = c + ann(X) \in A/ann(X)$ for some $c \in A$. Now, we will show that $(\bar{c})^m = (\bar{c})^{m+1}$ for some $m \in \mathbb{N}$. As X is a weak π-regular module, $c^m X = c^{m+1}X$ for some $m \in \mathbb{N}$. Since X is f.g., by [4, Corollary 2.5], we conclude that $(1 - dc)c^m X = 0$ for some $d \in A$. This implies that $c^m - dc^{m+1} \in ann(X)$ and so $(c^m) + ann(X) = (c^{m+1}) + ann(X)$. Then we deduce $(\bar{c})^m = (\bar{c})^{m+1}$ and hence $A/ann(X)$ is a strongly π-regular ring.

Proposition 2.16. Let $g: X \to X'$ be an epimorphism of A-modules and X be a strongly π-regular module. Then X' is a strongly π-regular module.

Proof. Let $x' \in X'$. Since g is surjective, $g(x) = x'$ for some $x \in X$. Since X is a multiplication module, $Ax = JX$ for some ideal J of A. As X is a strongly π-regular module, $(Ax)^m = J^m X = cX = c^2 X$ for some $c \in A, m \in \mathbb{N}$. This implies that $g(Ax) = Ag(x) = Ax' = JX'$ and also note that $(Ax')^m = (JX')^m = J^m X' = J^m g(X) = g(J^m X) = g(cX) = g(c^2 X)$. This gives $(Ax')^m = cX' = c^2 X'$ which completes the proof.

Corollary 2.17. Let X be a strongly π-regular module and Y a submodule of X. Then X/Y is a strongly π-regular module.

An element $c \in A$ is said to be a zero divisor on X if $cx = 0$ for some $0 \neq x \in X$. Also the set of all zero divisor elements on X is denoted by $z(X)$.

Lemma 2.18. Let X be a f.g. strongly π-regular module. Then $cX = X$ for each $c \in A - z(X)$.

Proof. Let $c \in A - z(X)$. Note that for each $m \in \mathbb{N}, c^m \notin z(X)$. Since X is a f.g. strongly π-regular module, by Proposition 2.10, $c^k X = c^{k+1} X$ for some $k \in \mathbb{N}$. Let $x' \in X$. As $c^k X = c^{k+1} X$, $c^k x' = c^{k+1} x''$ for some $x'' \in X$. This yields that $c^k (x' - cx'') = 0$ and so $x' = cx''$. Therefore, we have $cX = X$.

Corollary 2.19.

(i) Suppose that X is a f.g. strongly π-regular module in which $z(X) = ann(X)$. Then X is a second module.

(ii) Suppose that X is a f.g. multiplication module in which $z(X) = ann(X)$. Then X is a second module iff X is a strongly π-regular module.

(iii) Suppose that X is a f.g. faithful strongly π-regular module over an integral domain A. Then X is a divisible module.

(iv) Suppose that X is a torsion-free module, namely, $z(X) = 0$. If X is a f.g. strongly π-regular module, then X is a divisible module.

Proof. (i) Assume that X is a f.g. strongly π-regular module in which $z(X) = ann(X)$. Let $c \in A$. If $c \in z(X) = ann(X)$, then $cX = 0$. So assume that $c \notin z(X)$. Thus by Lemma 2.18, $cX = X$ and so X is a second module.

(ii): Directly from (i), Example 2.9, Proposition 2.14 and Proposition 2.15.

(iii): It is similar to (i).

(iv): It can be obtained from (iii).

Proposition 2.20. Suppose that X is an A-module and $T \subseteq A$ is a multiplicatively closed subset of A. If X is a strongly π-regular module, then $T^{-1} X$ is a strongly π-regular $T^{-1} A$-module.

Proof. Let $\frac{x}{t} \in T^{-1} X$ for some $t \in T, x \in X$. Since X is a strongly π-regular module, $(Ax)^m = cX = c^2 X$ for some $c \in A, m \in \mathbb{N}$. As X is a multiplication module, we can write $Ax = JX$. Note that $T^{-1} A \left(\frac{x}{t} \right) = T^{-1}(Ax) = T^{-1}(JX) = T^{-1}(J)T^{-1}(X)$. This implies that $\left[T^{-1} A \left(\frac{x}{t} \right) \right]^m = \left(T^{-1}(J) \right)^m T^{-1} X = T^{-1}(J^m)T^{-1} X = T^{-1}(J^m X) = T^{-1}(cX) = T^{-1}(c^2 X)$. Then we have $\left[T^{-1} A \left(\frac{x}{t} \right) \right]^m = \frac{c}{1} T^{-1} X = \left(\frac{c}{1} \right)^2 T^{-1} X$ which completes the proof.

Recall that a commutative ring A is said to be a *quasi-semi-local* if the number of its maximal ideals is finite.

Proposition 2.21. Let X be a f.g. module over a quasi-semi-local ring. The followings are equivalent.

(i) X is a strongly π-regular module.

(ii) X_q is a strongly π-regular module for each $q \in Spec(A)$.

(iii) X_q is a strongly π-regular module for each $q \in Max(A)$.

Proof. $(i) \Rightarrow (ii)$: It can be obtained from Proposition 2.20.

$(ii) \Rightarrow (iii)$: Directly from $Max(A) \subseteq Spec(A)$.

$(iii) \Rightarrow (i)$: First note that, by [5, Lemma 2], X is a multiplication module. As X is a f.g. multiplication module, it is sufficient to show that X is a weak π-regular module. Let $c \in A$. Now, we will show that $c^m X = c^{m+1} X$ for some $m \in \mathbb{N}$. Since A is quasi semi-local, A has finitely many maximal ideals $q_1, q_2, \ldots, q_t$. As X_{q_i} is a strongly π-regular module, $\left(\frac{c}{1} \right)^{m_i} X_{q_i} = \left(\frac{c}{1} \right)^{m_i+1} X_{q_i}$ for some $m_i \in \mathbb{N}$. Now, put $m = \max\{m_1, m_2, \ldots, m_t\}$. Then we have $\left(\frac{c}{1} \right)^m X_{q_i} = \left(\frac{c}{1} \right)^{m+1} X_{q_i}$ and so $(c^m X)_{q_i} = (c^{m+1} X)_{q_i}$. This implies that $c^m X = c^{m+1} X$ which completes the proof.

Recall from [19] that a proper submodule Y of X is said to be a *pure submodule* if $cX \cap Y = cY$ for each $c \in A$.

Lemma 2.22. Let X be a strongly π-regular module and Y a pure submodule of X. Then Y is a strongly π-regular module.

Proof. First, we will show that Y is a multiplication module. Take a submodule V of Y. Since X is multiplication, $V = (V:X)X$. Let $x \in V$. Then we can write $x = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$ for some $c_i \in (V:X)$ and $x_i \in X$. As Y is pure and $c_i x_i \in c_i X \cap Y = c_i Y$, we have $c_i x_i = c_i x_i'$ for some $x_i' \in Y$. This implies that $x = c_1 x_1' + c_2 x_2' + \cdots + c_n x_n' \in (V:X)Y \subseteq (V:Y)Y$. Then we have $V = (V:X)X \subseteq (V:X)Y \subseteq (V:Y)Y \subseteq V$. Thus $V = (V:Y)Y$. Now, we will show that Y is a strongly π-regular module. Let $x \in Y$. As X is a strongly π-regular module, we conclude that $(Ax)^m = dX = d^2 X$ for some $d \in A$ and $m \in \mathbb{N}$. Since Y is a pure submodule of X, we have $dY = dX \cap Y = d^2 X \cap Y = d^2 Y$. Thus we conclude that $(Ax)^m = (Ax)^m \cap Y = dX \cap Y = dY = d^2 Y$. Therefore, Y is a strongly π-regular module. $\blacksquare$

Theorem 2.23. Let X_i be a multiplication A_i-module for all $i = 1, 2, \ldots, n$. Then $A = \prod_{i=1}^{n} A_i$-module $X = \prod_{i=1}^{n} X_i$ is a strongly π-regular module iff X_i is a strongly π-regular module for all $i = 1, 2, \ldots, n$.

Proof. $\Rightarrow$: Suppose X is a strongly π-regular A-module and choose $t \in \{1, 2, \ldots, n\}$. Now, we shall show that X_t is a strongly π-regular module. To see this, take $x_t \in X_t$. Put $x = (0, 0, \ldots, x_t, 0, 0, \ldots, 0) \in X$. Since X is a strongly π-regular module, we get $(Ax)^m = cX = c^2 X$, where $c = (c_1, c_2, c_3, \ldots, c_n) \in A$. As X_t is multiplication module, we have $A_t x_t = J_t X_t$ for some ideal J_t of A_t. This implies that $Ax = JX$, where $J = \{0\} \times \{0\} \times \ldots \times J_t \times \{0\} \times \ldots \times \{0\}$. Then we conclude that $(Ax)^m = (\{0\} \times \{0\} \times \ldots \times J_t{}^m \times \{0\} \times \ldots \times \{0\})X = cX = c^2 X$. Then we get $(A_t x_t)^m = J_t{}^m X_t = c_t X_t = c_t{}^2 X_t$. Therefore, X_t is a strongly π-regular module.

$\Leftarrow$: Suppose X_i is a strongly π-regular module for every $i = 1, 2, \ldots, n$. Let $x = (x_1, x_2, \ldots, x_n) \in X$.

Note that $Ax = \prod_{i=1}^{n}(A_i x_i)$. Since X_i is a multiplication A_i-module, for all $i = 1, 2, \ldots, n, A_i x_i = J_i X_i$. Note that $Ax = \prod_{i=1}^{n}(J_i X_i) = (\prod_{i=1}^{n} J_i)X$ and also $(Ax)^t = (\prod_{i=1}^{n} J_i)^t X = \prod_{i=1}^{n}(J_i{}^t X_i)$ for all $t \in \mathbb{N}$. Since X_i is a strongly π-regular module, $(A_i x_i)^{t_i} = J_i{}^{t_i} X_i = c_i X_i = c_i{}^2 X_i$ for some $t_i \geq 1$. Let $m = \max\{t_1, t_2, \ldots, t_n\}$. Then it is easy to see that $(A_i x_i)^m = J_i{}^m X_i = c_i X_i = c_i{}^2 X_i$. Thus we conclude that $(Ax)^m = \prod_{i=1}^{n}(J_i{}^m X_i) = \prod_{i=1}^{n}(c_i X_i) = \prod_{i=1}^{n}(c_i{}^2 X_i)$. This yields that $(Ax)^m = cX = c^2 X$, where $c = (c_1, c_2, \ldots, c_n) \in A$. $\blacksquare$

Recall that an A-module X is said to be a *reduced module* if for each $c \in A, x \in X, c^2 x = 0$ implies $cx = 0$ [15]. It was shown, in [13, Lemma 10], that a f.g. vn-regular module is reduced. Also, it is well known that a ring A is a vn-regular ring iff A is reduced strongly π-regular ring. Now, we prove this fact for vn-regular modules.

Theorem 2.24. The followings are equivalent for any f.g A-module X.

(i) X is a vn-regular module.

(ii) X is a reduced strongly π-regular module.

Proof. $(ii) \Rightarrow (i)$: We will show that $cX = c^2 X$ for each $c \in A$. Since X is a f.g. strongly π-regular module, by Proposition 2.5, $A/ann(X)$ is a strongly π-regular ring. Let $c \in A$. Then we have $c^m - dc^{m+1} \in ann(X)$ for some $d \in A$ and $m \in \mathbb{N}$. This implies that $c^m(1 - dc) \in ann(X)$ and so $(c(1 - dc))^m X = 0$. Since X is a reduced module, we conclude that $c(1 - dc)X = 0$ and so $cX = c^2 X$. Therefore, by [13, Theorem 1], X is a vn-regular module.

$(i) \Rightarrow (ii)$: Directly from [13, Lemma 10] and Example 2.4.

Recall from [3] that a proper ideal J of A is said to be an (m, n)-closed ideal if $c^m \in J$, then $c^n \in J$. Now, we will characterize vn-regular modules in terms of (m, n)-closed ideals.

Theorem 2.25. Let X be a f.g. reduced multiplication module. Then X is a vn-regular module iff every proper ideal of $A/ann(X)$ is a (3,2)-closed ideal.

Proof. $\Rightarrow$: Let J' be a proper ideal of $A/ann(X)$. Then there exists a proper ideal J of A containing $ann(X)$ such that $J' = J/ann(X)$. Now, we will show that J is a (3,2)-closed ideal of A. To see this take $c^3 \in J$. Then we have $c^3 X \subseteq JX$. Since X is a vn-regular module, we conclude that $c^2 X = c^3 X \subseteq JX$. By [21, Corollary to Theorem 9], we have $c^2 \in J + ann(X) = J$. Thus J is a (3,2)-closed ideal of A. By [3, Corollary 2.11], J' is a (3,2)-closed ideal of $A/ann(X)$.

$\Leftarrow$: Let $c \in A$. Now, we will show that $c^2 X = cX$. If $c^3 X = X$, then it is easily seen that $cX = X = c^2 X$. So suppose that $c^3 X$ is a proper submodule of X. Then by assumption $(c^3 X : X)/ann(X)$ is a (3,2)-closed ideal of $A/ann(X)$. Then by [3, Corollary 2.11], $(c^3 X : X)$ is a (3,2)-closed ideal of A. Since $c^3 \in (c^3 X : X)$, we get $c^2 \in (c^3 X : X)$ and so $c^2 X = c^3 X$. Since X is f.g., by [4, Corollary 2.5], we have $(1 - dc)c^2 X = 0$ for some $d \in A$ and so $[(1 - dc)c]^2 X = 0$. As X is reduced module, we obtain that $(1 - dc)cX = 0$ and this yields that $cX = c^2 X$. By [13, Theorem 1], X is a vn-regular module.

3. WHEN DOES LOCALIZATION AT A PRIME IDEAL COMMUTE WITH INFINITE INTERSECTION?

Let $q \in Spec(A)$ and X an A-module. Then we know that $(\bigcap_{i=1}^n Y_i)_q = \bigcap_{i=1}^n (Y_i)_q$. One can naturally ask whether this property is true when we replace the finite intersection by infinite one. This question has a negative answer.

Example 3.1. Consider the $\mathbb{Z}$-module $\mathbb{Z}$ and $q = 2\mathbb{Z}$. Let $Y_i = 3^i \mathbb{Z}$ for each $i \in \mathbb{N}$. Then it is clear that $(\bigcap_{i=1}^\infty Y_i)_q = 0_q \neq \mathbb{Z}_{(2)} = \bigcap_{i=1}^\infty (Y_i)_q$.

Recall from [18] that a prime submodule Q of X is said to be a *strongly prime submodule* if $\bigcap_{i \in \Delta} Y_i \subseteq Q$, then for some $j \in \Delta$, $Y_j \subseteq Q$. In particular, a multiplication module M is said to be a *strongly zero dimensional module* if its each

prime submodule is strongly prime. In [7], the authors showed that a f.g. multiplication module X is a strongly zero dimensional module iff its Krull dimension $\dim(X) = 0$ and X is quasi-semi-local. Now, we determine when the localization commutes with infinite intersection in terms of strongly prime submodules.

Theorem 3.2. Let X be a f.g. multiplication module and Q a q-prime submodule of X. Then Q is a strongly prime submodule iff the localization at q commutes with the intersection of any family of submodules of X.

Proof. $\Rightarrow$: Let Q be a strongly prime submodule and $q = (Q:X)$. Now we will show that $(\bigcap_{i \in \Delta} Y_i)_q = \bigcap_{i \in \Delta} (Y_i)_q$ for every family of submodules $\{Y_i\}_{i \in \Delta}$ of X. We know that the inclusion $(\bigcap_{i \in \Delta} Y_i)_q \subseteq \bigcap_{i \in \Delta} (Y_i)_q$ always holds. Let $\frac{x}{t} \in \bigcap_{i \in \Delta} (Y_i)_q$. Then for each $i \in \Delta$, $t_i x \in Y_i$ for some $t_i \notin q$. This implies that $(Y_i : Ax) \nsubseteq q$. As X is a f.g. multiplication module, we have $(Y_i : Ax)X \nsubseteq Q$. As Q is a strongly prime submodule, we conclude that $(\bigcap_{i \in \Delta} (Y_i : Ax)X) = (\bigcap_{i \in \Delta} (Y_i : Ax))X \nsubseteq Q$. Thus we have $\bigcap_{i \in \Delta} (Y_i : Ax) \nsubseteq q$ and so there exists $c \in \bigcap_{i \in \Delta} (Y_i : Ax) - q$. This implies that $\frac{x}{t} \in (\bigcap_{i \in \Delta} Y_i)_q$.

$\Leftarrow$: Assume that the localization at q commutes with the infinite intersection of any family of submodules. Now, we will show that Q is a strongly prime submodules. As X is a f.g. multiplication module, X_q is a quasi-local module over a quasi-local ring A_q. Let $\bigcap_{i \in \Delta} Y_i \subseteq Q$. Then we get $(\bigcap_{i \in \Delta} Y_i)_q = \bigcap_{i \in \Delta} (Y_i)_q \subseteq Q_q$. Since Q_q is the unique maximal submodule of X_q, we have $(Y_i)_q \subseteq Q_q$ for some $i \in \Delta$. As Q is a q-prime submodule, we deduce $Y_i \subseteq Q$ for some $i \in \Delta$. Therefore, Q is a strongly prime submodule of X.

Now, we will end this section by studying (∗)-condition on modules and improving some results in [6]. The following proposition gives the relations between (∗)-condition and strongly π-regular module.

Proposition 3.3.

(i) Let X be a f.g. strongly π-regular module. Then X satisfies (∗)-condition.

(ii) Let X be a f.g. multiplication module satisfying (∗)-condition. Then X is a strongly π-regular module.

Proof. (i): Suppose that X is a f.g. strongly π-regular module. Then it is clear that X satisfies descending chain condition on principal powers. The rest follows from [7, Lemma 3].

(ii): Let X be a f.g. multiplication module satisfying (∗)-condition. Then by [7, Lemma 3] and [7, Corollary 3], $\dim(X) = 0$. Then by [7, Lemma 4], it can be seen that $\dim(A/ann(X)) = 0$. Again by [8, Theorem 3], $A/ann(X)$ is a strongly π-regular ring. The rest follows from Proposition 2.14.

Now, we will characterize strongly zero dimensional modules in terms of localization at a prime ideal.

Theorem 3.4. The followings are equivalent for every f.g. multiplication module X.

(i) X is a strongly zero dimensional module.

(ii) X is a zero dimensional quasi-semi-local module.

(iii) X satisfies (∗)-property and no maximal submodule contains the intersection of other maximal submodules.

(iv) X satisfies descending chain condition on principal powers and no maximal submodule contains the intersection of the other maximal submodules.

(v) For every q-prime submodule Q of X, the localization at q commutes with the infinite intersection of any family of submodules of X.

Proof. $(i) \Leftrightarrow (ii)$: It can be obtained from [7, Corollary 4].

$(i) \Leftrightarrow (iii) \Leftrightarrow (iv)$: It can be obtained from [7, Theorem 4] and [7, Lemma 3].

$(i) \Leftrightarrow (v)$: It can be obtained from Theorem 3.2.

Lemma 3.5. Let X be a f.g. multiplication A-module and $x \in X$ such that $Ax = JX$ for some ideal J of A. The followings are equivalent.

(i) The descending chain $Ax \supseteq (Ax)^2 \supseteq \cdots \supseteq (Ax)^m \supseteq \cdots$ stops.

(ii) $JX + \bigcup_{m=1}^{\infty}(0:_X J^m) = X$.

Proof. $(ii) \Rightarrow (i)$: Suppose that $JX + \bigcup_{m=1}^{\infty}(0:_X J^m) = X$. Since X is a multiplication module, $(0:_X J^m) = ((0:_X J^m):X)X = (0:J^m X)X$. This gives $(J + \bigcup_{m=1}^{\infty}(0:J^m X))X = X$. Since X is f.g. module, we obtain that $J + \bigcup_{m=1}^{\infty}(0:J^m X) = A$. Then we have $c + d = 1$ for some $c \in J$, $d \in (0:J^m X)$ and $m \in \mathbb{N}$. This implies that $J^m X = cJ^m X \subseteq J^{m+1}X \subseteq J^m X$. Thus the descending chain $Ax \supseteq (Ax)^2 \supseteq \cdots \supseteq (Ax)^m \supseteq \cdots$ stops at the mth step.

$(i) \Rightarrow (ii)$: Suppose that (i) holds. Then $(Ax)^m = (Ax)^{m+1}$ for some $m \in \mathbb{N}$. As X is a f.g. multiplication module, $Ax = JX$ for some f.g. ideal J of A. Then we have $J^m X = J^{m+1}X$. Since J and X are f.g., so is $J^m X$. Then by [4, Corollary 2.5], $(1 - c)J^m X = 0$ for some $c \in J$. This implies that $(1 - c)X \subseteq (0:_X J^m)$. Since $cX \subseteq JX$, we obtain that $X = cX + (1 - c)X \subseteq JX + \bigcup_{m=1}^{\infty}(0:_X J^m) \subseteq X$ which completes the proof.

Theorem 3.6. The followings are equivalent for every f.g. multiplication module X.

(i) X is a strongly π-regular module.

(ii) X is a weak π-regular module.

(iii) $A/ann(X)$ is a strongly π-regular module.

(iv) X satisfies (∗)-condition.

(v) X satisfies descending chain condition on principal powers.

(vi) $\dim(X) = 0$.

Proof. $(i) \Rightarrow (ii)$: Directly from Proposition 2.10.

$(ii) \Rightarrow (iii)$: It can be obtained from Proposition 2.15.

$(iii) \Rightarrow (i)$: It can be obtained from Proposition 2.14.

$(i) \Leftrightarrow (iv)$: It can be obtained from Proposition 3.3.

$(iv) \Leftrightarrow (v) \Leftrightarrow (vi)$: It can be obtained from [7, Lemma 3], [7, Corollary 3] and [7, Lemma 5].

Acknowledgements

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this work.

REFERENCES

[1]   R. Ameri, "On the prime submodules of multiplication modules," International journal of Mathematics and mathematical Sciences, vol. 27, pp. 1715-1724, 2003.

[2]   D. D. Anderson, S. Chun and J. R. Juett, "Module-theoretic generalization of commutative von Neumann regular rings," Communication in Algebra, vol. 47, no. 11, pp. 4713-4728, 2019.

[3]   D. F. Anderson and A. Badawi, "On (m,n)-closed ideals of commutative rings," Journal of Algebra and Its Applications, vol. 16, no. 01, 1750013 (21 page), 2017.

[4]   M. Atiyah and I. MacDonald, Introduction to Commutative Algebra, Avalon Publishing, 1994.

[5]   A. Barnard, "Multiplication modules," Journal of Algebra, vol. 71, no. 1, pp. 174-178, 1981.

[6]   Z. Bilgin and K. H. Oral, "Coprimely structured modules," Palestine Journal of Mathematics, vol. 7, Spec. Issue 1, pp. 161-169, 2018.

[7]   Z. Bilgin, S. Koç and N. A. Özkirişci, "Strongly prime submodules and strongly 0-dimensional modules," Algebra and Discrete Mathematics, vol. 28, no. 2, pp. 171-183, 2019.

[8]   J. Brewer and F. Richman, "Subrings of zero dimensional rings," In multiplicative ideal theory in commutative algebra, Springer, Boston, MA, pp. 73-88, 2006

[9]   V. Camillo and W. K. Nicholson, "Quasi-morphic rings," Journal of Algebra and Its Applications, vol. 6, no. 05, pp. 789-799, 2007.

[10]   P. V. Danchev, "A generalization of π regular rings," Turkish Journal of Mathematics, vol. 43, no. 2, pp. 702-711, 2019.

[11]   Z. A. El-Bast and P. F.Smith, "Multiplication modules," Communications in Algebra, vol. 16, no. 4, pp. 755-779, 1988.

[12]   M. Evans, "On commutative P. P. rings," Pacific Journal of mathematics, vol. 41, no. 3, pp. 687-697, 1972.

[13]   C. Jayaram and Ü. Tekir, "von Neumann regular modules," Communications in Algebra, vol. 46, no. 5, pp. 2205-2217, 2018.

[14] I. Kaplansky, Commutative rings, Allyn and Bacon, 1970.

[15] T. K. Lee and Y. Zhou, "Reduced Modules," Rings, Modules, Algebras and Abelian Groups. Lecture Notes in Pure and Appl. Math., Vol. 236. New York: Dekker, pp. 365-377, 2004.

[16] C. P. Lu, "Prime submodules of modules," Comment. Math. Univ. Sanct. Pauli, vol. 33, no. 1, pp. 61–69, 1984.

[17] E. Matlis, "Divisible modules," Proceedings of the American Mathematical Society, vol. 11, no. 3, pp. 385-391, 1960.

[18] K. H. Oral, N. A. Özkirişci and Ü. Tekir, "Strongly 0-dimensional modules, " Canadian Mathematical Bulletin, vol. 57, no. 1, pp. 159-165, 2014.

[19] P. Ribenboim, Algebraic Numbers, New York, NY, USA: Wiley, 1974.

[20] R. Y. Sharp, Steps in commutative algebra, London Math. Soc. Stud. Texts, 19, Cambridge University Press, Cambridge, 1990.

[21] P. F. Smith, "Some remark on multiplication modules", Archiv der Mathematik, vol. 50, no. 3, pp. 223-235, 1988.

[22] J. Von Neumann, "On regular rings," Proceedings of the National Academy of Sciences of the United States of America, vol. 22, no. 12, pp. 707-713, 1936.

[23] S. Yassemi, "The dual notion of prime submodules," Arch. Math.(Brno), vol. 37, no. 4, pp. 273-278, 2001.

[24] H. Zhu and N. Ding, "Generalized morphic rings and their applications, " Communications in Algebra, vol. 35, no. 9, pp. 2820-2837, 2007.

Title: Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid
Propeller-induced Vibrations During Hovering

Authors: Halil Bahadır AKYILDIZ, İlyas KACAR, Mehmet Kürşat YALÇIN

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrations During Hovering

Halil Bahadır AKYILDIZ[1], İlyas KACAR[*2], Mehmet Kürşat YALÇIN[3]

Abstract

The vibration parameters of a bi-copter-type unmanned aerial vehicle is optimized by considering operational vibration with payloads. The double electric ducted fan loads, which transmit excitations to the fuselage, are predicted and compared using optimization methods. While the minimum vibration amplitude for stress will be achieved at 7.69 Hz, it will be 9.80 Hz. for minimum deformation without sacrificing safety factor requirement. It ensures sensitive vertical acceleration. It is not seen significant differences on results from screening and genetic algorithm methods. Correlations between frequencies and structural responses are determined. It is observed that the stress and deformation amplitudes of the structure decreases at increasing frequencies up to the next natural frequency. While the highest amplitude is seen at the first frequency, it decreases in increasing modes. The airframe structural model's operational frequency must be 7.69 or 9.80 Hz to achieve sensitive vertical acceleration. Subsequently, it is aimed to develop an autonomous task by the implemented system controlled by various algorithm as a future work.

Keywords: Bi-copter, Tandem Rotor, Unmanned Aerial Vehicle, Genetic Algorithm, Optimization.

[1]Niğde Ömer Halisdemir University, Department of Mechatronics Engineering, Niğde, Turkey.
 ORCID: https://orcid.org/ 0000-0001-9558-4846, E-mail: h.bhdr01@gmail.com
[*]Corresponding Author: ikacar@gmail.com
[2]Niğde Ömer Halisdemir University, Department of Mechatronics Engineering, Niğde, Turkey.
 ORCID: https://orcid.org/ 0000-0002-5887-8807
[3]Niğde Ömer Halisdemir University, Department of Mechatronics Engineering, Niğde, Turkey,
 ORCID: https://orcid.org/ 0000-0001-9484-1422, E-mail: mkursatyalcin@gmail.com

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...

1. INTRODUCTION

Bi-copters which mean two-rotor become a new idea as an unmanned aerial vehicle (UAV) and are inspired from helicopters. Helicopters have superior abilities compared to the fixed wing air vehicles. Especially the abilities to hover in the air and fly backwards make them mandatory for difficult tasks. They are useful as a portable surveillance tool. They will be able to carry load by performing small amount of design changes. Thanks to bi-copters, "tail rotor" and "ductless main rotor" requirements of helicopters are eliminated. Instead, electric ducted fans (EDF) are used as driver. Bi-copters have just two driver units. So energy consumption is the least among multi-rotor air vehicles. They require less space during landing/take-off. Bi-coppers also have six degrees of freedom to be able to be controlled by four input signals only. Although an additional drive system is generally used to provide stability besides the main drive system to ensure a stabilized flight on multi-copters such as quad-copter and tri-copters, no additional auxiliary system is necessary to obtain a stabilized flight on desired trajectories for bi-copters. Additional systems would bring extra electrical and mechanical loads to the aircraft. While the most-seen problem in propeller aircraft is stabilization during flight, bi-copters are far away from disturbing conditions. Balance can be obtained by control strategies. Direction can be adjusted by just changing of rotor angle.

Multi-rotor vehicles are useful in the air mapping, protection of agricultural land, crime detection by law enforcement officers [1]. More flight time is obtained by bi-copter with tilt mechanism [2]. Yoon and Lee [3] investigated the aerodynamic rotor shapes by considering flight control and kinematics. An optimization method provides suitable coefficients for controller. A control strategy to balance the bi-copter is achieved in 40% overshoot and the settling time 16 seconds [4]. Elias et al. [5] performed a wirelessly controlled bi-copter by hand gestures. It is seen that thrusts are provided by brushless motors, the direction is controlled by servo motors. Hand movements are transmitted to control signals through a special glove equipped with an Arduino© controller, accelerometer and flex sensors. By combining two bi-copters, direction control is obtained by just rotor angle.

Since UAV's are constrained to have lower weight as little as possible for energy efficiency, optimizing the material properties is an important task for both its own or any attachment's stabilization such as cameras during service life [6, 7]. The stiffness of the material affects vibration characteristics. Structural resonance can cause to impose malfunctioning and thereby degrading manoeuvrability.

UAVs are unique vehicles due to their vertical take-off and landing capabilities. Although lightness is a criterion for its body designs, it may cause higher vibrations. This study focuses on vibrations due to rotation of motor-driven propellers. If any excitation frequency interferes with structures fundamental frequency, it causes resonance. To avoid resonance, the lowest fundamental frequency has to be higher than the maximum working excitation frequencies of the propellers. The fundamental frequency of the bi-copter depends on its dimensions, material (stiffness), boundary conditions mainly. Selecting lighter material and lower vibration without sacrificing safety requirement is main strategy. In this study, natural frequencies of the designed bi-copter are determined based on finite element (FE) simulations. Also an optimization is performed by following two methods so-called screening and genetic algorithm. Optimization gives the relations between frequencies and structural responses. An implementation of the bi-copter is carried out in accordance with the optimization results. It is tested for modal frequencies. The study is organized including a conceptual design of the vehicle which defines the parts based on their kinematics. Material and method section defines physical and mechanical properties of potential materials to be used for optimization. Also it includes the experimental modal test and FE simulations for structural analyses and optimization process.

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...

2. CONCEPTUAL DESIGN

Bi-copters are mainly composed of one body and two thrust fans mechanically. The body also consists of a single board computer to run the flight firmware. Two servo motors provide the tilt angle of thrust motors. As a single board computer, a Raspberry Pi© computer is assembled to body with an electronic speed controller (ESC) and lithium polymer (Li-Po) batteries. A camera card is located on the bottom side of the body as seen in Figure 1. These are payloads for the vehicle. Its weight is 225.8 gr. totally.
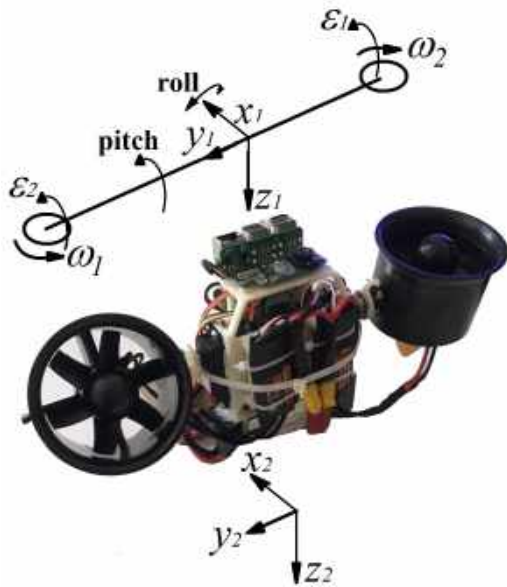
Figure 1. Kinematics of the design. ω: rotational speed, ε=pitching, z: vertical direction.

3. MATERIALS AND METHODS

Vibration analyses are performed on the UAV body experimentally. Results are compared with those from FE analyses. The subsequent sections describe how these FE results are validated by using experimental results from simulations.

Materials whose yield point is in between 25-280 MPa which correspond to polyethylene and aluminium alloys respectively are investigated due to its potential for being fuselage material. Mechanical properties are listed in Table 1.

Table 1. Material properties

	Aluminium alloy	Polyethylene
Density (kg/m^3)	2770	950
Young's Modulus (GPa)	71	1.1
Poisson Ratio	0.33	0.42
Bulk Modulus (GPa)	69	2.29
Shear Modulus (GPa)	26.7	0.38
Yield Strength (MPa)	280	25

3.1. Modal testing

Instead of surrogate model, the designed and implemented model is used for modal testing. Mass for each one of the components is measured accurately. The impulse hammer excitation test is performed on the body to determine modal frequencies. Free-free boundary connection is provided. It is elastically connected to the wall of experimental rig system as shown in Figure 2 in order to eliminate boundary effects.

Figure 2. Experimental rig system with similar to boundary conditions during hovering.

An accelerometer is used to capture the frequency responses of the body in the range of 0-200 Hz. Its weight is 0.5 grams. Measurements are done on a set of 100 excitation points along the complete structure. Up to 200 Hz, 10 mode shapes are clearly obtained.

3.2. Static Analysis

Stress distribution, deformation and safety factors are determined by means of structural analysis by applying flight loads on the model.

The model is supported by a frictionless support through the hole surface passing through its centre as it is hung in experiment. Each one of the EDF motors provides 1kgf trust loads per motor. Ansys© is used for simulations [8]. Figure 3 shows FE model and load and boundary conditions.
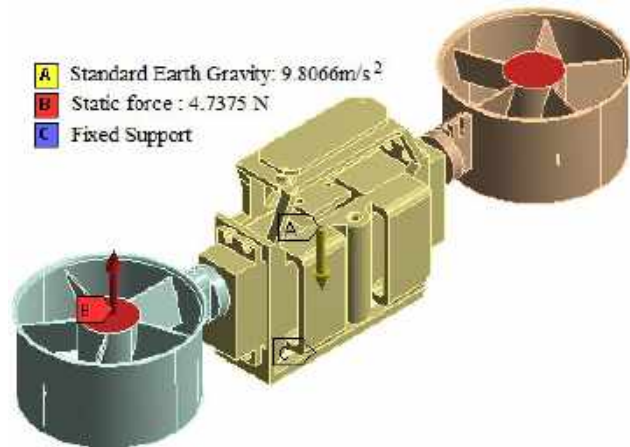
Figure 3. Load and boundary conditions

Also fatigue analysis is performed to determine the safety factor and working life in the cyclic load case. The Soderberg criterion mixed with Von Mises is used as mean stress correction theory. This is a useful criterion for materials which have ductile behaviour. Designs to be used in the aviation field must meet the requirement of fatigue strength. Stress-life (*S-N*) data from fatigue tests performed on polyethylene material are given in Table 2. The loading frequency is applied as 1 Hz. The mean stress caused by the applied periodic tension/compression load is zero (R=-1). Kf=0.87 is used in the analysis for fatigue strength reduction factor.

Table 2. *S-N* data

Cycle	Stress range (MPa)
6546	20.02369
5967	20.48856
5336	21.07367
4757	21.779
4027	22.47631
3145	23.65453
2617	24.70451
2088	25.64228
1458	27.63003
979	29.62579
577	31.84598
426	34.07418
274	36.53482
148	38.41035
98	39.69277

3.3. Forced Vibration Analysis

Forced vibration analysis gives how to be the stress and deformation responses of the structure in case of harmonic/random excitations. Prior to forced vibration analysis, a modal simulation is performed to determine the natural frequencies. The analysed frequency range is taken as 0-5000 Hz with 0.1 Hz increments in the simulations.

3.4. Optimization

The goal of the optimization is to find maximum force, without sacrificing safety factor=3 and a sensitive vertical acceleration with minimum vibration amplitude. The safety factor is calculated from both static and fatigue analyses. Stress and deformations responses are found for each frequencies from 0 to 5000 Hz with 0.1 Hz increment. Results from screening [8] and genetic algorithm (GA) [9, 10] methods are compared. Evaluated parameters are listed in Table 3. Ten thousand points are evaluated as designs of experiment.

Screening method is one of the easiest multi-objective optimization algorithms. Its approach is based on just sorting according to sampling table. It is preferred for preliminary designs.

Table 3. Input/output parameters and lower/upper limits of input variables

	Initial value	Lower limit	Upper limit
Input parameters			
Tensile yield strength (MPa)	280	10	70
Harmonic force (N)	60	0	100
Static force (N)	60	0	100
Frequency for equivalent stress (Hz)	0	0	5000
Frequency for total deformation (Hz)	0	0	5000
Output parameters			
Deformation (m) (statical)	--	--	--
Equivalent stress (MPa) (statical)	--	--	--
Safety factor (statically)	--	--	--
Life minimum (fatique)	--	--	--
Safety factor minimum (fatique)	--	--	--
Equivalent stress (MPa) (harmonic)			
Deformation (m) (harmonic)			

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...

GA sets an analogue between tabulated data and a set of solutions called population, represented by chromosomes. Yalçın et al [11] explained an application of this optimization method in detail. The GA parameters used in this study are given in Table 4.

Table 4. GA parameters used in the study

Parameters	Value
Estimated number of evaluation	2000
Number of initial samples	100
Number of samples per iteration	100
Maximum allowable Pareto percentage	70%
Convergence stability percentage	2%
Maximum number of iterations	20

The goal and applied constrains in the optimization are applied as follow to catch a sensitive vertical acceleration.

- Minimize the static deformation
- Minimize the harmonic deformation
- Minimize the static safety factor (SF) as long as SF $\geq$ 2
- Minimize the fatigue SF as long as SF $\geq$ 3
- Maximize the static force
- Maximize the harmonic force

4. RESULTS AND DISCUSSIONS

Obtained data are given in this section. The experimental modal testing was conducted to verify the structural characteristics of the developed structural model of UAV.

4.1. Natural frequencies

A comparison on the frequency responses from experiment and analysis is given in Table 5. In the table "Exp." means the experimental results. Just first ten frequencies are listed. Vibration analysis simulates the steady state structural response of body to periodical loads. So, resonance can be kept away. Mode shapes are obtained from FE simulation. The maximum working frequency of the propellers is 2570 Hz in full throttle.

Table 5. A comparison of resonance frequencies on complete body.

Mode	Modal frequencies (Hz) FE	Exp.	Mode shapes	Description
1	31.267	30.33		Rigid body pitch mode
2	37.845	38.22		Rigid body roll mode
3	55.088	56.74		1st axial mode
4	55.088	56.74		2nd axial mode
5	71.243	71.96		1st bending symmetric mode
6	82.219	84.69		1st bending antisymmetric mode
7	93.703	96.51		1st torsional symmetric mode
8	117.55	118.73		1st torsional antisymmetric mode
9	149.43	146.44		2nd bending symmetric mode
10	175.73	181.00		2nd bending antisymmetric mode

The difference between experimental and FE analysis is in the range of approximately 3%, which is also consistent with the value given in [12-14]. It is important to detect natural frequencies to avoid resonance. Thus, the driving frequency that will not stimulate the resonance can be determined. It is presented that the error between experimental results and simulations can become 7.99% for the whole models due to number of reasons such as the increased number of joints among components and the effect of the suspension rubber used to hang the bi-copter to the ceiling [7].

4.2. Optimum values

The optimized (best fitted) values are listed in Table 6, 7. For verification, analyses are repeated with optimized values. Tables include verified values. As seen in Table 6, minimum stress and deformation are obtained at frequencies 7.685

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...

Hz and 9.80 Hz. respectively. While the structure has endurance up to 19.8 N harmonic force, it is limited to 4.74 N for static load case. Yield strength of the body material must be 12.91 MPa at least. Sawalakhe and Shaaikh [15] reported that 2.4525 N will be enough for acceleration of their 1000 gr. model. So it is concluded that 4.74 N will be enough for our case which is 225.8 gr.

Table 6. The values obtained from GA optimization method and verification of the results

	Optimized value	Verified value	Error (%)
Input parameters			
Tensile yield strength (MPa)	12.91		
Harmonic force (N)	19.80		
Static force (N)	4.74		
Frequency for equivalent stress (Hz)	7.69		
Frequency for total deformation (Hz)	9.80		
Output parameters			
Deformation (m) (statical)	0.00096	0.000958	-0.31
Equivalent stress (MPa) (statical)	2958392.36	2949810.37	-0.29
Safety factor (statically)	12.21	14.58	19.41
Life minimum (fatique)	9729901.80	10000000	-
Safety factor minimum (fatique)	4.95	3.88	-21.6
Equivalent stress (MPa) (harmonic)	10.55	10.55	-
Deformation (m) (harmonic)	34.25	34.25	-0.02

Table 7. The values obtained from screening optimization method and verification of the results

	Optimized value	Verified value	Error (%)
Input parameters			
Tensile yield strength (MPa)	13.0391		
Harmonic force (N)	19.998		
Static force (N)	4.7874		
Frequency for equivalent stress (Hz)	7.7669		
Frequency for total deformation (Hz)	9.898		
Output parameters	0		
Deformation (m) (statical)	0.00097	0.000968	-0.3131
Equivalent stress (MPa) (statical)	2987976	2979308	-0.2929
Safety factor (statically)	12.3321	14.7258	19.6041
Life minimum (fatique)	9827201	10100000	-
Safety factor minimum (fatique)	4.9995	3.9188	-21.816
Equivalent stress (MPa) (harmonic)	10.6555	10.6555	-
Deformation (m) (harmonic)	34.5925	34.5925	-0.0202

Figure 4-9 shows counter plots at the last load step when optimized variables are used in analyses.
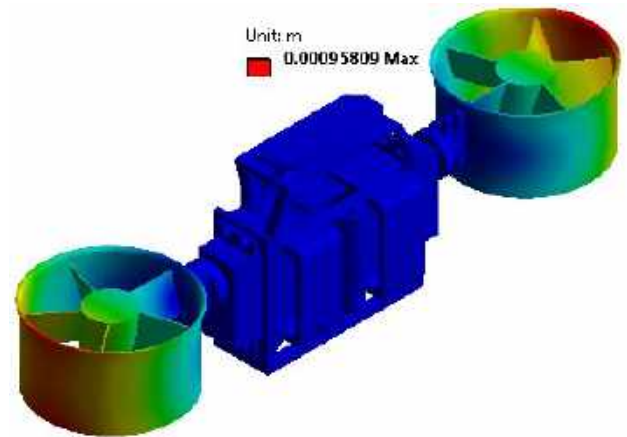
Figure 4. Deformation results
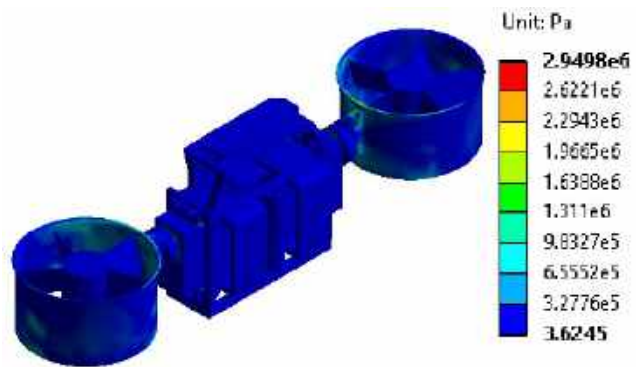
Figure 5. Stress results
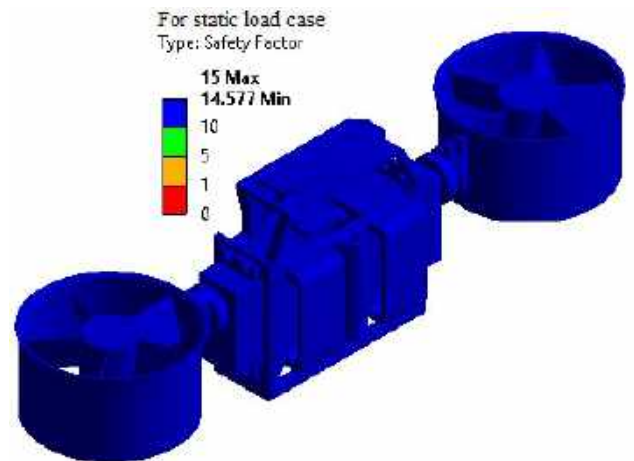
Figure 6. Safety factors (static)

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...
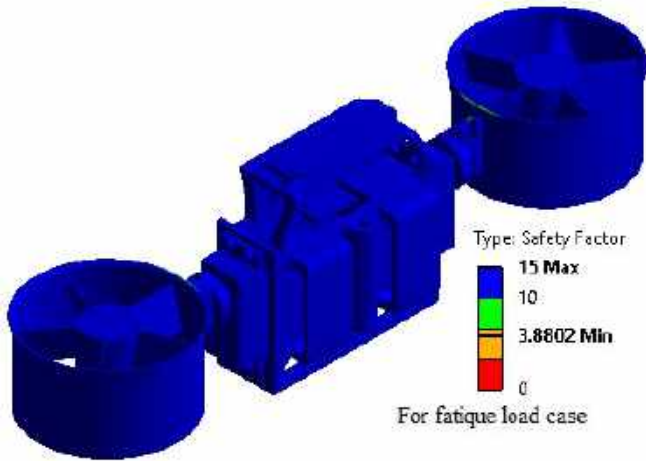
Figure 7. Safety factors (fatigue)
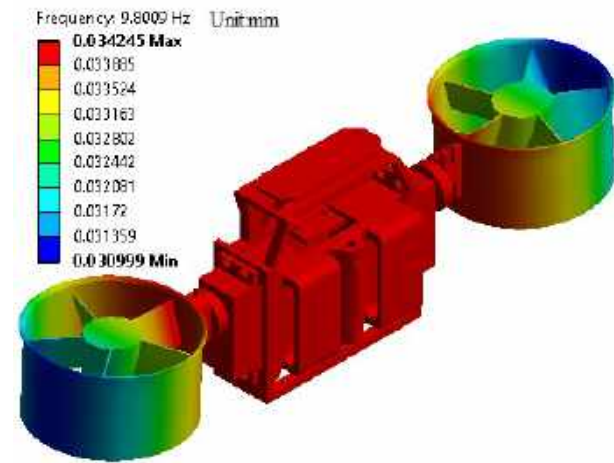
Figure 8. Displacement response at a frequency (9.8009 Hz.)
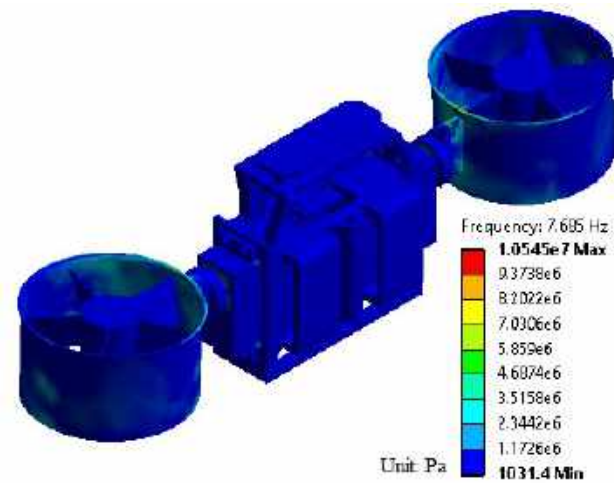
Figure 9. Stress response at a frequency (7.685 Hz.)

When the values determined by means of optimization were applied on the bi-copter, the responses shown in Figure 4-9 would be obtained. The maximum deformation value is 0.9mm, this value is much less than that of Das et. al [16] where the maximum deformation amplitude is 27.29 mm. The safety factor is 14.77, well above 3 which is the goal. Where the stress is at its maximum, its value is 2.9 MPa, and the value is well below the yield strength. So, the validation of the optimum values is done.

4.3. Relation between Parameters

The relations among all variables especially between excitation force and frequency response are obtained in the optimization procedure using Kriging method [10] as response surface investigation. Results are given in Figure 10.
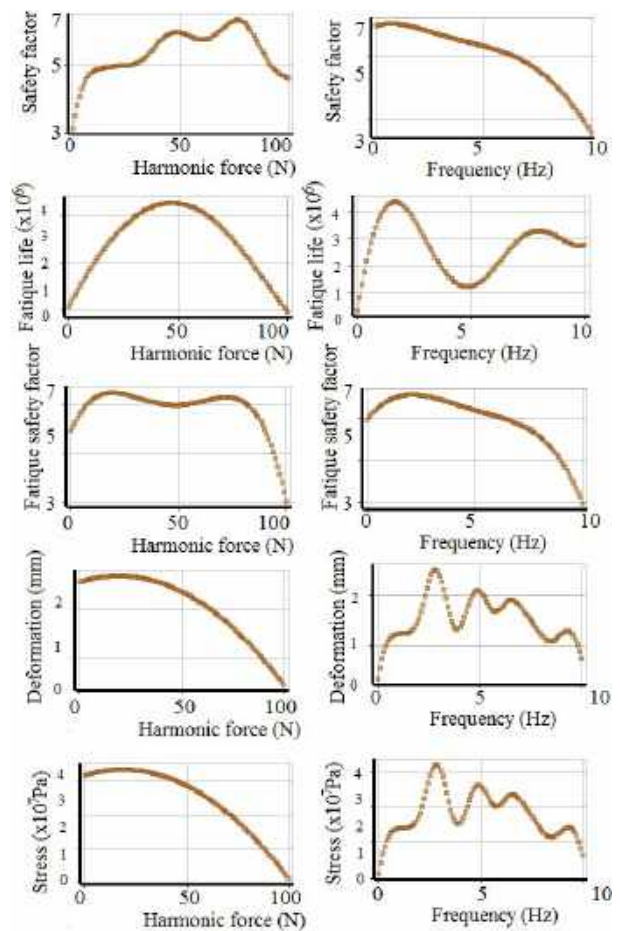
Figure 10. Relations between parameters

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...

The influence of the force and frequency is seen from the curves. It is seen that the safety factor tends to decrease in static and fatigue states with increasing force, as expected. However, with the increase of harmonic force, it is seen that the safety factor increases. The reason for this can be explained as excessive vibration in the structure and the resonance leading to increase the stress amplitude. For the same reason, it is understood that at increasing frequencies, the safety factor decreases.

5. CONCLUSIONS

In this paper, an optimization process is performed considering operational frequencies to achieve a sensitive hovering of a bi-copter type unmanned aerial vehicle to avoid propeller-induced vibrations. Structural endurance and stability of UAV is the main concern. Optimal force and frequency values are determined. So the following inferences are concluded.

- The minimum lift force 4.74 N provides the softest movement in hovering. So any attachment like camera will take images in both hovering and lifting cases with the maximum energy efficiency due to minimum propeller rotation requirement

- Any material whose yield point is bigger than 12 MPa will be enough for flight endurance.

- The structural analysis is done for calculating stresses, displacements, and safety factors. The displacements observed are negligible and within the limit. The maximum stresses produced are 2.9 MPa which are within range.

- The comparison of the FE results and experimental modal results shows that the modal characteristics are in close agreement within the of 3% error for the first then elastic global modes.

- The maximum working frequency of the propellers is 2570 Hz in full throttle which is far away from the first then fundamental frequencies.

Within the scope of this study, just optimum parameters are determined and verified. For subsequent studies, it is aimed to apply different control techniques on flight stability and equilibrium.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of the paper.

REFERENCES

[1] Wiktionary, "bicopter," Access date: 29.01.2019, Update date: 11 December 2015, https://en.wiktionary.org/wiki/bicopter#English

[2] Q. Zhang, Z. Liu, J. Zhao, and S. Zhang, "Modeling and attitude control of Bi-copter," AUS 2016 - 2016 IEEE/CSAA International Conference on Aircraft Utility Systems, art. no. 7748042, pp. 172-176, 2016.

[3] J. Yoon and J. Lee, "Altitude and roll control of a hovering quad-rotor air vehicle using the multi-objective approximate optimization of proportional–integral–differential control," Engineering Optimization, vol. 49, no. 10, pp. 1704-1718, 2017.

[4] C.M. Elias, O.M. Shehata, and E.I. Morgan, "Remote e-Lab towards an integrated cognitive experience," International

AKYILDIZ et al.

Parameter Optimization of a Bi-copter Type Unmanned Aerial Vehicle to Avoid Propeller-induced Vibrati...

Conference on Developments in eSystems Engineering, DeSE 2, art. no. 7563658, pp. 332-337, 2015.

[5] P.G. Subin, K.T Kautilya, G.S. Kumar, and A.V. Sai, "Stabilizing the bi-copter and controlling it using gesture technology," International Journal of Control Theory and Applications, vol. 9, no. 15, pp. 7235-7245, 2016.

[6] M. Verma, V. Lafarga, M. Baron, and C. Collette, "Active stabilization of unmanned aerial vehicle imaging platform," Journal of Vibration and Control, vol. 0, no. 0, pp. 1–13, 2020.

[7] A. Tullu, Y. Byun, J-N. Kim, and B-S. Kang, "Parameter optimization to avoid propeller-induced structural resonance of quadrotor type unmanned aerial vehicle," Composite Structures. vol. 193, no. 2018, pp. 63-72, 2018.

[8] G. J. DeSalvo and J.A. Swanson, "ANSYS engineering analysis system user's manual," Houston, Pa., Swanson Analysis Systems, pp. 62-78, 1985.

[9] I. Rechenberg, "Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution," Frommann-Holzboog-Verlag, Stuttgart, Broschiert, pp. 1-170, 1973.

[10] J.H. Holland, "Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence," University of Michigan Press, pp.1-183, 1975.

[11] M.K Yalçın., İ.Kacar, and H.B. Akyıldız, "Design and implementation of a bi-copter driven by dual electric ducted fans using genetic algorithm optimization technic," International conference on environment, technology and management (ICETEM), 27-29 June 2019, Niğde, Turkey, pp. 1203-1214. 2019.

[12] M. Y. Ren and A. Vipradas, "ANSYS DOE and Design Optimization Tutorial," Access date: 12.12.2019, Design Informatics Lab., School for Engineering of Matter, Transport and Energy, Arizona State University https://designinformaticslab.github.io/produ ctdesign_tutorial/2016/11/20/ansys.html

[13] J. Verbeke and S. Debruyne, "Vibration analysis of a UAV multirotor frame," Proceedings of ISMA2016 including USD2016, pp. 2329-2338, 2016.

[14] E. Camargo, N-J. Jacobsen., and D. Strafacci, "Operational modal analysis on a modified helicopter," IMAC XXIX, pp.1-9, 2011.

[15] P.V Sawalakhe and J.A. Shaaikh, "Simulation and analysis of a quadrotor UAV while landing," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6, pp. 672-680, 2020.

[16] N. Das, S. Das., D.K. Mishra, and K.M. Pandey, "Analysis of deformation and mode shape in the landing gear of light unmanned aerial vehicle," NANOMTECH 2019, IOP Conf. Series: Journal of Physics: Conf. Series 1455 (2020) 012020, pp.1-5, 2020.

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excitation

Mehmet Akif KOÇ[*1]

Abstract

In this study, dynamic analysis of Euler-Bernoulli beam and Tuned Mass Damper (TMD) interaction problem under the effect of moving load was carried out by the mode superposition method. After the differential equations of TMD are derived by Lagrange method, beams and TMD motion equations are integrated and matrices belonging to the motion equation of the whole system are obtained. The motion equation of the system is solved in the time domain using the Newmark-β algorithm. The effect of TMD on damping vibrations has been examined in terms of parameters such as frequency, damping rate, mass ratio and moving load speed. In addition, the effect of TMD on Dynamic Amplification Factor (DAF) was examined. As a result, with the TMD application carried out in this study, approximately 14% to 24% improvement was achieved in beam deformations and accelerations.

Keywords: TMD, moving load, Newmark-Beta, DAF

1. INTRODUCTION

The damping of mechanical and structural vibrations is of great importance in the machine elements industry, in the fields of construction, automotive, aerospace and robotics. In the past few decades, researchers have made a lot of effort to reduce vibrations on engineering structures [1–3]. The most traditional method of damping vibrations is passive vibration damping techniques, and this technology has been applied in the literature quite a lot [1, 4–7]. Thanks to recent tremendous advances in digital signal processing, sensors and actuator technologies, active vibration control algorithms have been

quickly applied to different engineering problems [8, 9].

Beam type structures have many applications in engineering, especially robotics, mechanics, aviation and construction. The low damping of such structures and the increasing trend of designing especially lighter structures recently cause the beam type structures to vibrate at low mode frequencies. For this reason, one of the biggest challenges engineers face is to protect these types of structures from excessive vibrations and prevent them from being damaged.

The idea of protecting the main structure from excessive vibrations in control engineering is

---

[*] Corresponding Author: makoc@subu.edu.tr
[1] Sakarya Applied Science University, Department of Mechatronics Engineering, Sakarya, Turkey.
   ORCID: https://orcid.org/0000-0001-7461-9795

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

based on Tuned Mass Damper (TMD) technology, which is connected to this main structure with spring and damping elements, which absorbs vibration energy. The frequency of this secondary structure connected to the main structure is usually adjusted to the frequency of the primary structure. Thus, the main structure is protected from the destructive effect caused by excessive vibrations. Due to the simplicity and low cost of TMD technology, it has been possible to apply it in many engineering fields [10, 11].

A good mathematical model was needed to represent the physical model of the Beam-TMD in order to effectively implement TMD in beam type structures and to optimize the position, frequency and basic parameters of this secondary structure, which will be placed in the main structure. One of the most used methods for modeling beam-TMD interaction is the Finite Elements Method (FEM). In the study [12] FEM was used in the analysis of the TMD model used to reduce vibrations in the Timoshenko beam under harmonic and random excitation force. Dynamic Vibration absorber has been used to reduce vibrations in the Timoshenko beam under the effect of harmonic distributed load [13]. Wu [14] proposed using a dynamic vibration damper to the middle of the bridge to dampen the vibrations of the beam under the action of a moving load. After obtaining the equation of motion of the system, they used FEM to determine the beam dynamics. The simple model used to obtain the optimal resistance and damping ratio of the vibration absorber with Den Hartog's approach is given in [15].Greco and Santini [16]

analyzed the beam with a rotary viscous damper placed on both ends under the effect of a moving load. Scientists have shown in their work that the performance of the damper depends on the speed of the moving load.

When we examine the literature, analytical methods, FEM, energy equations and series expansion methods are used to solve such problems. In this study, the mode superposition method was used to analyze the beam TMD interaction problem. In this way, it has been proven that the beam absorber interaction problems can be analyzed effectively with the presented method. In the study, a TMD was placed in the middle of the bridge beam to reduce vibrations in the Euler-Bernoulli beam. The dynamic response of the beam is analyzed for different moving load speeds and different mass ratios.

2. FORMULATION OF EULER-BERNOULLI BEAM COUPLED TO A DYNAMIC MASS DAMPER

In this section, initially the equations of motion in the differential form for and Euler-Bernoulli beam with simply supported boundary conditions and attached n TMDs, as illustrated Figure 1, has been derived. The parameters x and P on the figure show the time-dependent position and force value of the moving load on the beam relative to the reference point taken from the left end of the bridge, respectively.
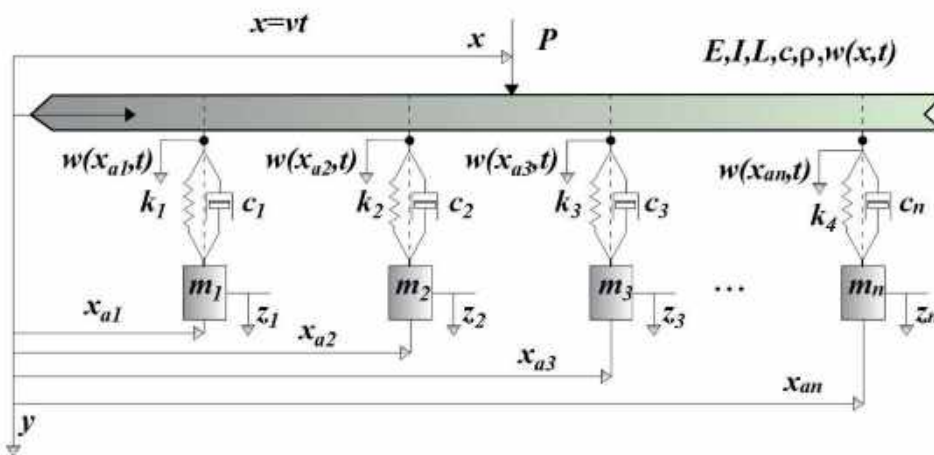
Figure 1. Euler-Bernoulli beam with a linear TMD traversed by a moving load.

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

2.1. Deriving Equation of Motion for TMD

The kinetic and potential energies for the TMDs shown in Figure 1 are written as follows:

$$E_k = \frac{1}{2}\left(\sum_{i=1}^{n} m_i z_i^2\right), (i = 1, ..., n) \tag{1a}$$

$$E_p = \frac{1}{2}\sum_{i=1}^{n}\left(k_i\left(z_i - w_i\left(x_{a,i}, t\right)\right)^2\right), (i = 1, ..., n) \tag{1b}$$

Similarly, the damping function is written as shown in Equation (2).

$$D = \frac{1}{2}\sum_{i=1}^{n} c_i\left(\dot{z}_i - \dot{w}_i\left(x_{a,i}, t\right)\right)^2, (i = 1, ..., n) \tag{2}$$

The system's Lagrange function is written taking into account the difference between kinetic energy and potential energy. The Lagrange function is written as follows:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}_i}\right) - \frac{\partial L}{\partial z_i} + \frac{\partial D}{\partial \dot{z}_i} = Q_i \tag{3}$$

Using the Lagrange function specified in Equation (3), the motion equation of TMD is written as follows:

$$m_i \ddot{z}_i + \begin{Bmatrix} k_i\left(z_i - w_i\left(x_{a,i}, t\right)\right) + \\ c_i\left(\dot{z}_i - \dot{w}_i\left(x_{a,i}, t\right)\right) \end{Bmatrix} = 0 \tag{4}$$

2.2. Deriving Equation of Motion for Beam

In the problem of interaction between moving load, Euler-Bernoulli beam and TMD, the following assumptions are made:

- The moving load is always in contact with the beam.
- Beam cross-sectional area is fixed and modeled in accordance with Euler-Bernoulli theorem.
- The speed of the moving load on the beam is constant and any accelerated movement is excluded from the scope of this study.

- Absorbers placed on the beam are considered to have linear characteristics.

With all these assumptions, according to the Euler-Bernoulli beam theory, the motion equation for the bridge beam over which the moving load passes is written as follows:

$$EI\frac{\partial w^4(x,t)}{\partial x^4} + \rho\frac{\partial w^2(x,t)}{\partial t^2} + c_b\frac{\partial w(x,t)}{\partial t} = P_b(x,t) \tag{5}$$

Parameters E, I, ρ and c_b in Equation (5) are bridge beam elasticity module, bridge beam constant cross-sectional area, mass of unit length and damping coefficient, respectively. However, parameter $w(x,t)$ represents the transverse deformation of the bridge beam at position x at any time t. In Equation (5), as is customary, the prime symbol indicates the derivative with respect to the spatial variable, i.e. $w'(x,t) = dw/dx$, and the dot symbol is derivative with respect to the time coordinate, i.e. $\dot{w}(x,t) = dw/dt$. Parameter $P_b(x,t)$ is an external load function and is expressed as follows.

$$P_b(x,t) = -\sum_{k=1}^{N_v} f_k \delta(x - vt) + \left\{k\left(z - w\left(x_{a,i}, t\right)\right) + c\left(\dot{z} - \dot{w}\left(x_{a,i}, t\right)\right)\right\} \tag{6}$$

Thee parameters k_i and c_i in Equation (6) are the spring and damping coefficients of the TMD which connected to the bridge, respectively. The parameter z_i parameter is the vertical displacement of the ith absorber connected to the beam. The parameter N_v in Equation (6) represents the number of moving loads on the beam. In this study, it is accepted that only one moving load passes over the beam. The expression $\delta(x-vt)$ in Equation (6) is the Dirac-delta function in the x direction, expressed as:

$$\int_a^b f_k \delta(x - vt) = \begin{cases} 1 & 0 \leq t \leq \dfrac{l}{v} \\ 0 & \text{else} \end{cases} \tag{7}$$

According to Galerkin's formulation, the transverse deformation of any point on the beam is expressed as follows.

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

$$w(x,t) = \sum_{i=1}^{N} \varphi_i(x)\eta_i(t) \qquad (8)$$

The expressions $\varphi_i\ (x)$ and $\eta_i\ (t)$ in Equation (8) are the mode function and modal coordinate obtained with the boundary conditions of the beam, respectively. The parameter N in Equation (8) represents the number of modes for calculation of the bridge dynamic. The mode function for the simply supported beam is written as follows:

$$\varphi_i(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{i\pi x}{L}\right), \qquad (9)$$

The orthogonality conditions between the mode functions given in Equation (9) are written as follows.

$$\int_0^L \varphi_n(x) \sum_{i=1}^{\infty} \varphi_i(x)\eta_i(t)dx = \left\{ \begin{matrix} 0 & i \neq n \\ \int_0^L \varphi_n^2(x)dx\eta_n(t) & i = n \end{matrix} \right\} \qquad (10)$$

The equation of motion given by equation (5) is written as follows with the expressions of Galerkin function, Equation (4) and orthogonality conditions given by Equation (10).

$$m_i \ddot{z}_i \varphi_n + \ddot{\eta}_n + 2\zeta_n \omega_n \dot{\eta}_n + \omega_n^2 \eta_n = -W_n \varphi_n \delta \qquad (11)$$

The expression ζ_n in equation (11) is the damping ratio corresponding to the nth mode of the beam, and is expressed as follows depending on the damping coefficient of the beam.

$$\zeta_n = \frac{c}{2\rho\omega_n} \qquad (12)$$

The parameter ω_n in Equation (12) represent nth natural frequency of beam and expressed as follows:

$$f_n = \frac{\omega_n}{2\pi} = \frac{n^2\pi}{2L^2}\left(\sqrt{\frac{EI}{\rho}}\right) (Hz.) \qquad (13)$$

2.3. Coupling Beam and TMD Equation of motions

The equation of motion given by equation (11) is written in matrix form as follows.

$$[M(t)]\{\ddot{Y}\} + [C(t)]\{\dot{Y}\} + [K(t)]\{Y\} = Q(t) \qquad (14)$$

In equation (14), $[M]$, $[C]$ and $[K]$ are the mass, damping and stiffness matrices of the beam-TMD system, respectively, written as follows:

$$[M(t)] = \begin{bmatrix} m_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & m_2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & m_n & 0 & 0 & \cdots & 0 \\ m_1\varphi_{11} & m_2\varphi_{21} & \cdots & m_1\varphi_{n1} & 1 & 0 & \cdots & 0 \\ m_1\varphi_{12} & m_2\varphi_{22} & \cdots & m_1\varphi_{n2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ m_1\varphi_{1n} & m_2\varphi_{2n} & \cdots & m_1\varphi_{1n} & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad (15a)$$

$$[C(t)] = , \begin{bmatrix} c_1 & 0 & \cdots & 0 & -c_1\varphi_{11} & -c_1\varphi_{12} & \cdots & -c_1\varphi_{1n} \\ 0 & c_2 & 0 & 0 & -c_2\varphi_{21} & -c_2\varphi_{22} & \cdots & -c_2\varphi_{2n} \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & c_n & -c_n\varphi_{n1} & -c_n\varphi_{n2} & \cdots & -c_1\varphi_{nn} \\ 0 & 0 & \cdots & 0 & 2\zeta_1\omega_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 2\zeta_2\omega_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 2\zeta_n\omega_n \end{bmatrix} \qquad (15b)$$

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

$$[K(t)] = \begin{bmatrix} k_1 & 0 & \cdots & 0 & -k_1\varphi_{11} & -k_1\varphi_{12} & \cdots & -k_1\varphi_{1n} \\ 0 & k_2 & 0 & 0 & -k_2\varphi_{21} & -k_2\varphi_{22} & \cdots & -k_2\varphi_{2n} \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & k_n & -k_n\varphi_{n1} & -k_n\varphi_{n2} & \cdots & -k_1\varphi_{nn} \\ 0 & 0 & \cdots & 0 & \omega_1^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \omega_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \omega_n^2 \end{bmatrix}, \tag{15c}$$

Similarly, the terms $\{\ddot{Y}\}, \{\dot{Y}\}$ and $\{Y\}$ are acceleration, velocity, and displacement vectors respectively, which are expressed as follows:

$$\{\ddot{Y}\} = \{\ddot{y}_1, \ddot{y}_2, \ldots, \ddot{y}_n, \ddot{\eta}_1, \ddot{\eta}_2, \ldots, \ddot{\eta}_n\}^T \tag{16a}$$

$$\{\dot{Y}\} = \{\dot{y}_1, \dot{y}_2, \ldots, \dot{y}_n, \dot{\eta}_1, \dot{\eta}_2, \ldots, \dot{\eta}_n\}^T \tag{16b}$$

$$\{Y\} = \{y_1, y_2, \ldots, y_n, \eta_1, \eta_2, \ldots, \eta_n\}^T \tag{16c}$$

$$\{Q\} = \left\{ \begin{matrix} 0, 0, \ldots, 0, \\ -W\varphi_{21}\delta, -W\varphi_{22}\delta, \ldots, -W\varphi_{2n}\delta \end{matrix} \right\}^T \tag{16d}$$

2.4. Numerically Solution Algorithm

In this study, the Newmark-β algorithm is used to solve the equation of motion given by Equation (14), and the solution algorithm is as follows.

- **Step 1:** Input bridge absorber parameters
- **Step 2:** Set the initial value X_0=0, V_0=0, t=0.
- **Step 3:** Calculate the location of the moving car
- **Step 4:** Create **M**, **C** and **K** matrix and **Q** force vector
- **Step 5:** Calculate acceleration with

$$A_0 = M\backslash(Q - CV_0 - KX_0) \tag{17a}$$

$$X_1 = X_0 + V_0 dt + 1/2 A_0 dt^2 \tag{17b}$$

$$V_1 = V_0 + A_0 dt \tag{17c}$$

$$A_1 = M\backslash(Q - CV_1 - KX_1) \tag{17d}$$

- **Step 6:** Calculate acceleration with

$$X_2 = X_0 + V_0 dt + (-1/2)A_0 dt^2 + 1/4 A_1 dt^2 \tag{17e}$$

$$V_2 = V_0 + (-1/2)A_0 dt + 1/2 A_1 dt \tag{17f}$$

$$A_2 = M\backslash(Q - CV_2 - KX_2) \tag{17g}$$

- **Step 7:** Determine A_1=A_2, V_1=V_2, X_1=X_2
- **Step 8:** if error is not small enough, go to step Step 6. İf it is enough t_i=t_i+dt.
- **Step 9:** if t_i<=T go to Step 3

3. SIMULATION RESULTS

In this section, the beam and load parameters given in Table 1 are used to analyze the beam and TMD interaction under the effect of the moving load shown in Figure 1.

Table 1. Bridge and moving load parameters used in this study.

Bridge parameters	
Elastic modulus (E), (GPa)	207
Mass per unit length (ρ), (Ton)	20
Cross-sectional inertia moment (I), (m⁴)	0.174
Beam length (L), (m)	100
Damping coefficient (c), (Ns/m)	1750
Number of modes (n)	4

3.1. Defining Mode Number, Mod Frequencies and Critical Speeds

Before starting numerical analysis, the number of beam modes to be used in the analyses was

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

determined. For this, first of all, simulation is done in different modes and the results are compared. Since the beam used in the study and the parameter values given in Table 1 is a large and bulky structure, it vibrates under the influence of low mode frequencies. Therefore, in this study, the first four modes of beam are taken into account in calculations.

The ratio of the bridge excitation frequency ω of the vehicle on the bridge to the natural frequency of the bridge $i.(i=1,2,\ldots,n)$ is called the speed parameter and is expressed as in equation (18). If $\omega = \omega i$, an event called resonance occurs, which has a negative effect for the bridge.

$$\alpha = \frac{\omega}{\omega_j} = \frac{\omega}{2\pi f_j} = \frac{vL}{j^2\pi}\left(\frac{\mu}{EI}\right)^{1/2} = \frac{v}{v_{kr}} \qquad (18)$$

Table 2. Beam frequencies used in this study.

Mode Number	1	2	3	4
Frequency (Hz.)	0.2108	0.8432	1.8972	3.3728
Critical speed (m/s)	42.15	168.63	379.43	674.55

In Table 2, vibration frequencies of the first four modes of beam used in this study and critical speed values of these frequencies are given.

3.2. Determining TMD Parameters

In this section, basic parameters are determined for damped TMD placed at the midpoint of the beam. The first is the mass ratio and is determined as follows:

$$\mu = \frac{M_{TMD}}{M_{structure}}, \qquad c_{cr,i} = 2m_i\omega_i,$$

$$\zeta_{TMD,i} = \frac{c_i}{c_{cr,i}}, \qquad \omega_{d,i} = \omega_{n,i}\sqrt{1-\zeta_i^2} \qquad (19)$$

Within equation (17), M_{TMD} represents the mass of the absorber placed in the central point of the beam, and $M_{structure}$ represents the total mass of the beam structure. Similarly, within equation (19), $c_{cr,i}$, $\zeta_{TMD,i}$, $\omega_{d,i}$ represent the critical damping value, damping rate and damped natural frequency of the absorber respectively. Table 3

shows the values used in the study for the first mode of the beam.

Table 3. Basic parameters for beam mode 1.

Mode Num	μ	$c_{cr,i}$	ζ_i	$\omega_{d,i}$	$x_{a,i}$
1	0.01	52956	0.35	1.24	L/2

In Figure 2, the effect of using TMD for the constant transition speed $v = 25$ m / s from the moving load beam on the transverse displacement at the midpoint of the beam is shown. The x axis is given the dimensionless position of the moving load on the bridge. The dimensionless position is obtained by dividing the moving load's time dependent position ($x=vt$) on the beam by the beam length (L). As shown in the figure, in the absence of TMD, the maximum deformation at the midpoint of the beam is 20.33 mm, while in the case of the TMD beam, this value is determined as 17.37 mm. As can be seen, 14.55% improvement was achieved in maximum beam mid-point displacement. This situation is explained by the fact that TMD, which is connected to the middle point of the beam and whose natural frequency is adjusted to the first mode frequency of the beam, absorbs the beam energy.
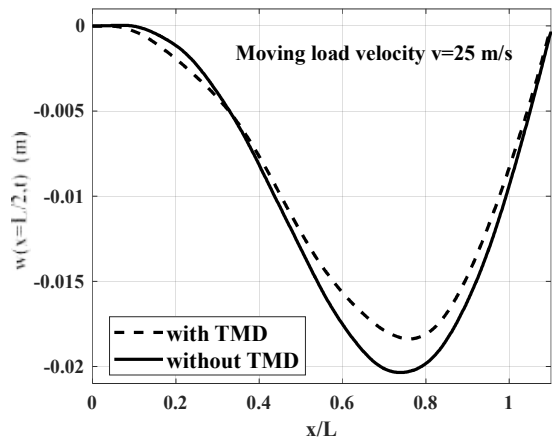
Figure 2. Beam midpoint transverse displacement.

In Figure 3, the comparison of the bridge mid-point transverse acceleration value with the case with TMD and the case without TMD is presented. As shown in the figure, in the absence of TMD, the bridge mid-point maximum transverse acceleration value is 0.025 m/s², while in the case of TMD this value is obtained as 0.019

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

m/s². As can be seen, TMD bridge mid-point has a 24% reduction in maximum acceleration value of bridge mid-point as well as maximum displacement.
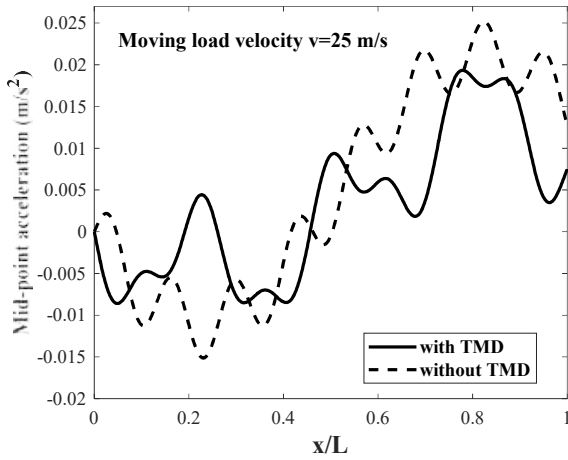
Figure 3. Beam midpoint transverse acceleration.

In Figure 4, the effect of the mass ratio given in Equation (19) on the maximum midpoint deformation of the beam is shown for different TMD damping rates. As seen in the figure, as the mass ratio increases, the maximum point deformations of the bridge midpoint decrease. Similarly, the increase in TMD damping ratio also caused this value to decrease.
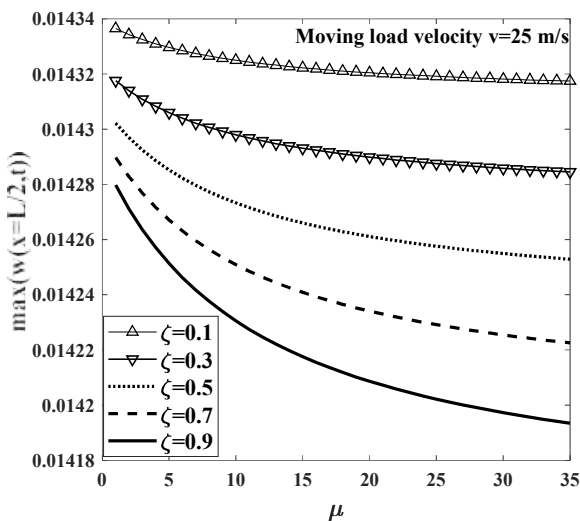
Figure 4. Maximum beam midpoint displacement for different TMD damping ratio ζ_{TMD}.

In Figure 5, the effect of the constant speed of the moving load over the beam on the Dynamic Amplification Factor (DAF) for different beam damping ratios ($\zeta = 1\%$, $\zeta = 3\%$ and $\zeta = 5\%$.) is shown.
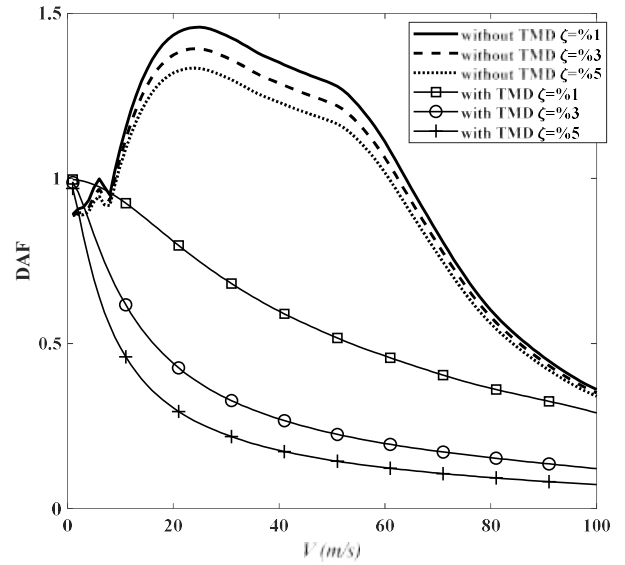
Figure 5. Beam DAF value for different beam damping ratio.

When the moving load passes on the DAF beam, the ratio of the maximum deformation (R_d (x)) occurring at the bridge midpoint to the static static deformation (R_s (x)) formed by placing the moving load on the bridge midpoint is as follows.

$$DAF = \frac{R_d(x)}{R_s(s)} \qquad (20)$$

As can be seen in Figure 5, considering the 1%, 3% and 5% beam damping ratio values, the maximum DAF values were determined as 1.45, 1.39, 1.33, respectively, for the case where there was no TMD. As can be seen, the dynamic displacement value for the beam in the case of moving load is about 50% more than static collapse. This situation proves the importance of moving load problems. In case of TMD, the maximum DAF values for the same damping rates are below 1. Also, as shown in the figure, the maximum DAF value decreases as the beam damping rate increases.

4. CONCLUTIONS

In this study, a dynamic analysis of the simply supported Euler-Bernoulli beam under the effect

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

of moving load was carried out. To reduce the maximum deformations on the beam, TMD was applied and the following results were obtained

• TMD application has been shown to be effective in reducing maximum acceleration deformations in a beam. An improvement of about 14% at maximum deformation and about 24% at maximum acceleration were achieved.

• The frequency at which the vibration absorber is adjusted is very important in TMD application. In cumbersome structures such as TMD beam, damping is more effective if it is set to a frequency corresponding to one of the low mode frequencies.

• The absorption rate and the increase in mass ratio of the absorber make damping more effective. However, although it is theoretically possible to increase the mass ratio too much, it is not possible in practice.

• TMD application has proven to be very effective in improving maximum DAF values.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] I. Esen and M.A. Koç, "Optimization of a passive vibration absorber for a barrel using the genetic algorithm.," Expert Systems with Applications. vol. 42, no. 2, p. 2015.

[2] M.A. Koç, İ. Esen, Y. Çay, et al., "Vibration Suppression of Vehicle-Bridge-Interaction System using Multiple Tuned Mass Dampers.," In: *5th International Symposium on Innovative Technologies in Engineering and Science*. pp. 1–8 (2017).

[3] H.C. Kwon, M.C. Kim, and I.W. Lee, "Vibration control of bridges under moving loads.," Computers & Structures. vol. 66, no. 4, pp. 473–480, 1998.

[4] Mahsa Moghaddas, *Finite Element Analysis and Passive Vibration Control of the Timoshenko Beam Traversed by a Moving Vehicle Using an Optimized Tuned Mass Damper.*, 2008.

[5] A.C. Mitra, G.J. Desai, S.R. Patwardhan, P.H. Shirke, W.M.H. Kurne, and N. Banerjee, "Optimization of Passive Vehicle Suspension System by Genetic Algorithm.," Procedia Engineering. vol. 144, pp. 1158–1166, 2016.

[6] M. Omar, M.M. El-kassaby, and W. Abdelghaffar, "Parametric numerical study of electrohydraulic active suspension performance against passive suspension.," Alexandria Engineering Journal. vol. 57, no. 4, pp. 3609–3614, 2018.

[7] D. Younesian, E. Esmailzadeh, and R. Sedaghati, "Passive vibration control of beams subjected to random excitations with peaked PSD.," JVC/Journal of Vibration and Control. vol. 12, no. 9, pp. 941–953, 2006.

[8] Q. Zhu, L. Li, C.J. Chen, C.Z. Liu, and G. Di Hu, "A Low-Cost Lateral Active Suspension System of the High-Speed Train for Ride Quality Based on the Resonant Control Method.," IEEE Transactions on Industrial Electronics. vol. 65, no. 5, pp. 4187–4196, 2018.

[9] I. Eski and Ş. Yildirim, "Vibration control of vehicle active suspension system using a new robust neural network control system.," Simulation Modelling Practice

Mehmet Akif KOÇ

Dynamic Response of an Euler-Bernoulli Beam Coupled with a Tuned Mass Damper under Moving Load Excita...

and Theory. vol. 17, no. 5, pp. 778–793, 2009.

[10] A.J. Yakel and A. Azizinamini, "Train-Induced Vibration Control of High-Speed Railway Bridges Equipped with Multiple Tuned Mass Dampers.," Journal of Bridge Engineering. vol. 10, no. 1, pp. 28–38, 2005.

[11] M. Moghaddas, E. Esmailzadeh, R. Sedaghati, and P. Khosravi, "Vibration control of Timoshenko beam traversed by moving vehicle using optimized tuned mass damper.," Journal of Vibration and Control. vol. 18, no. 6, pp. 757–773, 2012.

[12] F. Yang, R. Sedaghati, and E. Esmailzadeh, "Optimal vibration suppression of timoshenko beam with tuned-mass-damper using finite element method.," Journal of Vibration and Acoustics, Transactions of the ASME. vol. 131, no. 3, pp. 0310061–0310068, 2009.

[13] E. Esmailzadeh and N. Jalili, "Optimum design of vibration absorbers for structurally damped Timoshenko beams.," Journal of Vibration and Acoustics, Transactions of the ASME. vol. 120, no. 4, pp. 833–841, 1998.

[14] J.-J. Wu, "Study on the inertia effect of helical spring of the absorber on suppressing the dynamic responses of a beam subjected to a moving load.," Journal of Sound and Vibration. vol. 297, no. 3–5, pp. 981–999, 2006.

[15] J.P. DEN HARTOG, *Mechanical Vibrations*. , NEW YORK, 2008.

[16] A. Greco and A. Santini, "Dynamic response of a flexural non-classically damped continuous beam under moving loadings.," Computers and Structures. vol. 80, no. 26, pp. 1945–1953, 2002.

Authors: Tichaona Jonathan MAKOMO, Kenan ERİN, Barış BORU

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipulator

Tichaona Jonathan MAKOMO[*1], Kenan ERİN[2], Barış BORU[3]

Abstract

This work deals with the likelihood of merging a 3D sensor into a robotic manipulator, with an objective to automatically detect, track and grasp an object, placing it in another location. To enhance the flexibility and easy functionality of the robot, MATLAB, a versatile and powerful programming language is used to control the robot. For this work, a common industrial task in many factories of pick and place is implemented. A robotic system consisting of an ABB IRB120 robot equipped with a gripper and a 3D Kinect for Windows camera sensor is used. The three-dimensional data acquisition, image processing and some different parameters of the camera are investigated. The information in the image acquired from the camera is used to determine the robot's working space and to recognize workpieces. This information is then used to calculate the position of the objects. Using this information, an automatic path to grasp an object was designed and developed to compute the possible trajectory to an object in real time. To be able to detect the workpieces, object recognition techniques are applied using available algorithms in MATLAB's Computer Vision Toolbox and Image Acquisition Toolbox. These give information about the position of the object of interest and its orientation. The information is therefore sent to the robot to create a path through a server-to-client connection over a computer network in real time.

Keywords: Kinect, object recognition, 3D vision system, MATLAB, RobotStudio

1 INTRODUCTION

This work demonstrates a simple and common industrial task of manipulating objects by implementing a pick and place method using information from a vision system that is integrated with a high precision robotic manipulator. The main thrust of this

---

[*] Corresponding Author: tichaona.makomo@ogr.sakarya.edu.tr
[1]Sakarya University, Faculty of Technology, Department of Mechatronics, Sakarya, Turkey.
 ORCID: https://orcid.org/0000-0002-9860-6179
[2]Sakarya University of Applied Sciences, Faculty of Technology, Department of Mechatronics, Sakarya, Turkey.
ORCID: https://orcid.org/0000-0003-4714-1161. E-mail: kenanerin@sakarya.edu.tr
[3]Sakarya University of Applied Sciences, Faculty of Technology, Department of Mechatronics, Sakarya, Turkey.
ORCID: https://orcid.org/0000-0002-0993-3187. E-mail: barisb@sakarya.edu.tr

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...

work is to create a successful high precision grasp planning methodology using the image information from one Kinect for Windows sensor. Due to the increase in the advanced control applications, powerful hardware platforms and enhanced sensing capabilities, robotic systems have started to find place into different functionalities like mapping, exploration, entertainment industry, and wellbeing of people among others. Not a long time ago, robotic systems have been finding much use in many processing and manufacturing industries. They have been used semi-automatic, that is working side by side with humans whereas in some cases they are fully automatic, that is robots working with each other to complete a task. The objects being worked on would come with invariable physical attributes like color, size and shape. A real-time sensing system is therefore needed to be able to facilitate real time information about every object in the robot's control system. [1]. A number of technological advancements have been made in the development of flexible manufacturing systems (FMS). An FMS is described as a system consisting of one or more handling devices like robotic manipulators along with the robot controllers and machine tools, arranged so that it can handle different family of parts for which it has been designed and developed. [2]. With the lack of effective sensing abilities, many manufacturing units and assembly points cannot act intelligently in recognizing the workpieces and seeing workspace. Many of these systems have to work with predefined positions and orientations that are sent to the robot for manipulation. They however do not give room for change in those positions and orientation with which it becomes difficult for the robot to navigate to the new position. To cater for this anomaly, the robot system has to be equipped with an intelligent and flexible vision system that is able to communicate in real time with the robot to deal with imprecisely positioned objects and handle uncertainties and variations in the work environment.

This work's center of interest is to come up with a real-time sensor-based control system to applications where vital changes to varying objects is a necessity. [1] and [3], say that visual data received from the camera sensor is used for closed-loop robot control, normally known as visual servoing system. A review of the operation, characteristics and difficulties of visual servoing systems are found in [4] and [5]. In this work, object information is extracted from the acquired frame in a 2D format and then transformed to 3D pose with position of the object and its orientation for control of the robot. The control task is performed in 3D cartesian space thus the model of the camera required is that for mapping the data from 2D to 3D space. For one to be able to set up a visual servoing system, understanding of various aspects like robot modelling (dynamics and kinematics), control theory, computer vision functions including camera calibration, image processing, sensor system integration and so on is required. [6], [7], [8].

The other important capability of the vision information is enhancing the robot's ability to continuously update its field of view. This configuration is normally called eye-to-hand or eye-in-hand configuration. The architecture of the whole system is shown in figure 1 below:

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...
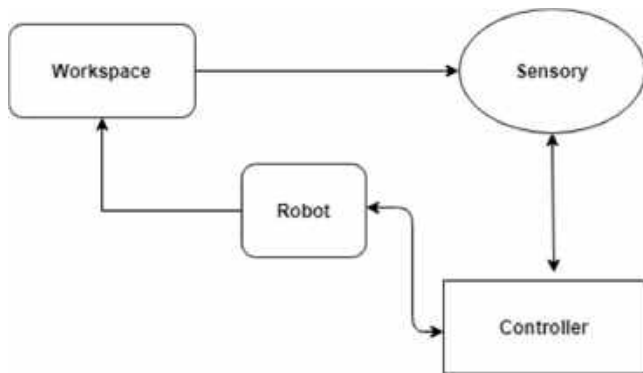
Figure 1 System of Robotic vision-based control

The diagram can be explained as follows

Workspace: includes fixtures, workpieces and tools.

Sensory system: is the vision or eyes of the robot to allow it to perceive its environment and to recognize the objects.

Control system: houses the system computer software and robot controller to organize tasks and the robot respectively.

Robot manipulator: does the action under the control of the robot controller.

With the automotive industry being traditionally the driving force and the largest consumer of automatic robot manipulators, it can be projected that as the robots are useful in the automotive industry, so shall they be for businesses with lower production capacity [9]. However, for these companies, it would cost them a lot of money in trying to automate their systems and having skilled labor for programming. Increase in production patterns will result in having cost of programming being higher than that of investment [10].

The pick and place operation in this study demonstrates how a robot is able to track, detect. and grasp an object in its field of view. The task here is to grasp a colored object placed on a table and for the task to be completed, a vision system detects the colored object by using color segmentation algorithms in real time. The object is then moved to a designated place by the robot manipulator. All this is done using communication protocols provided by TCP/IP. The image processing algorithm, the coordinate transformations, navigation algorithms and robot arm control are discussed. The validation of the system is discussed and results presented.

2. METHODOLOGY

2.1 Camera Calibration

The Kinect sensor used in this project was calibrated to bring about accuracy and effectiveness of it in the purpose for which it is being used. As is clearly highlighted in the introduction, many methods can be used to calculate the calibration parameters of the Kinect. The manufacturer's distortion parameters are not accurate so the sensor must be calibrated to correct that distortion. J.R Terven's method is used in this project to calibrate the Kinect camera. Kinect camera has an RGB camera and a depth camera which are both simultaneously calibrated using Terven's method. Terven designed a toolbox for Kinect V2 calibration in MATLAB called Kin2 Toolbox for MATLAB. The toolbox comprises of classes and functions that embeds the Microsoft Kinect 2 SDK. MATLAB is a high performance, versatile and powerful programming language developed by Mathworks which amalgamates computation and visualization to solve problems in a mathematical manner [11]. Largely based on C++ through MATLAB executable files, the version used for this project contains two classes and 30 functions put in different features such as coordinate mapping,

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...

skeleton tracking, 3D reconstruction and face gestures recognition. [11].

2.2 Experimental setup

Figure 2 below shows the experimental setup that was used in the project.

Figure 2 Schematic arrangement of work cell

The above figure is hereby explained. The Kinect camera sensor is placed above the workspace of the robot or the work cell. From there it is able to take images of objects which are placed on the work table or a conveyer. After an image is taken, object detection algorithm starts by taking a snapshot and analyzing it using MATLAB's toolboxes for computer vision and image processing. For every frame, color segmentation and region properties techniques are employed. This results in giving the coordinates of the centroid of the detected object (x, y). However, Z axis is considered to be known from the camera frame since the camera is fixed on one place vertically. That distance is measured and remains the same throughout the course of the experiment.

It should be known that in this project, the transformation matrix was not used to transform camera frame coordinates to robot frame coordinates. However, in light of this, the coordinates that we find from the camera frame are referenced to work object that is created in the robot workspace. Work object reference point is therefore used to map the found coordinates so that it makes the robot's base settle on the work object reference point. The coordinates detected in the workspace of the robot will not be referenced from the base of the robot but from the work object reference coordinate system that has been created.

The coordinates of the colored object detected are then sent to the real robot from MATLAB via TCP/IP communication. Once the robot receives the coordinates in its sockets, the robot moves to the detected object, pick it up and place it to another place in real time. A while loop was created both in the rapid program of the robot so that the robot continually listens if there is any input in its sockets. The socket continually listens and at the same time coordinates are being sent from MATLAB. The cycle continues like that as the camera keeps on taking images of the workspace and processing them to get coordinates, sends them to the robot via TCP/IP in real time and the robot continually keeps looking to receive coordinates of where it must go.

2.3 Object Detection

After the Kinect camera is calibrated and distortion corrected, it means that the camera can now acquire accurate images which can properly be processed in MATLAB without distortion. Object detection follows calibration of the camera sensor. MATLAB has toolboxes that are used in computer vision. Most popular toolboxes used in this work are Image processing toolbox and Image acquisition toolbox.

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...

Object detection starts with image acquisition and processing before we can conclude about position and orientation of the object. Image acquisition toolbox has functions that allow cameras of different types and properties to be used in MATLAB for acquiring images. Kinect that is used in this work is compatible to be used in MATLAB using 'videoinput' function. After an image is taken and a snapshot is taken then image processing starts.

2.4 Image Processing

Image processing toolbox brings in quite a number of algorithms which can be used to find properties of the detected object. When a snapshot of the object is taken, that image undergoes a series of algorithms as shown in Figure 3, to process it so that we can get what we want. Properties like area, centroid, minor axis, major axis, bounding box and so forth can be determined using 'regionprops' property in MATLAB. In this study, however, color segmentation was used to detect objects.
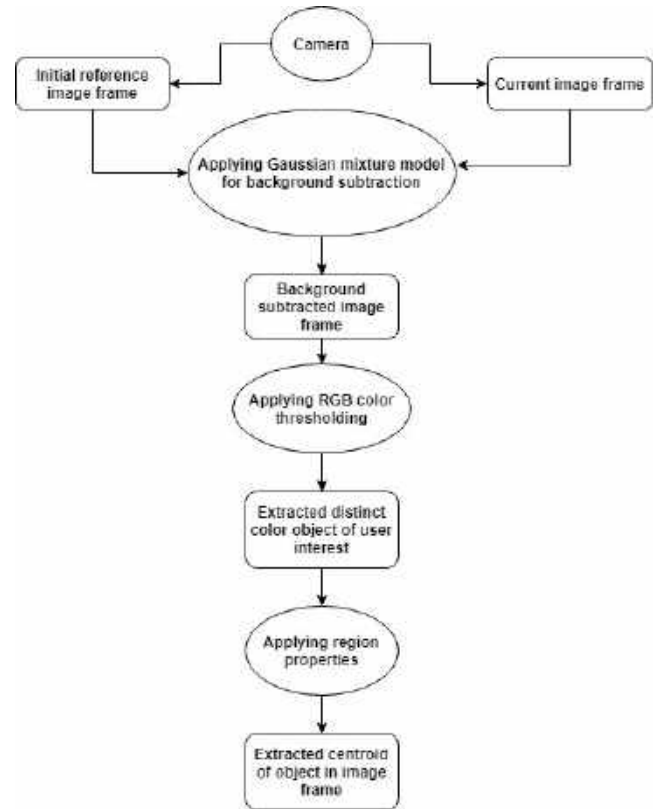
Figure 3 Flow chart involved in feature extraction

2.4.1 RegionProps and Centroid of the Object

Regionprops is a function that is very useful in image processing. It gives quite a number of properties depending on what you want to achieve. It is able to calculate 22 different properties but for this project it only focuses on three, including centroid. Centroid (x, y) is horizontal coordinate of the center of mass of the object and is represented by x-coordinate. The vertical component of the center of mass is the y-coordinate [12]. For this study, the centroid is a very important property since it is key in determining the exact location of the object. The centroid is calculated by the following equation;

$$\bar{x} = \frac{1}{n}\sum_{i-1}^{n} x_i, \ \ \bar{y} = \frac{1}{n}\sum_{i-1}^{n} y_i \qquad (1)$$

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...

where $\bar{x}, \bar{y}$ denote location of the centroid of the target. (x_i, y_i), $i = (1,2\ldots n)$ denote boundary points of the object in the x and y directions.

2.5 Getting Depth from Centroid

The Kinect toolbox of J. R. Terven was used for calibration of the Kinect camera. External and internal calibration parameters for depth and color camera were found. Compared to the manufacturers' calibration parameters, this method yielded almost the same results as the manufacturers. Calibration parameters play a major role in determining the conversion to practical distance. After capturing the depth, the object must have a distance in millimeters from the sensor. This distance is found in the OpenNI, depth (X_d, Y_d) function format. In the Kin2 toolbox for Kinect, JR Terven contains this OpenNI function, which contains the depth coordinates corresponding to the centroid of the object in the RGB frame. The coordinates used at the depth of function depth (X_d, Y_d) to find the depth or distance of this object in the sensor in millimeters. This is the Z coordinate from the camera.

2.6 TCP/IP Communication

Connection is established between a computer and robot. The robot is controlled by a programming language called RAPID in a graphical user interface called RobotStudio for ABB robots. The connection is between a computer in which MATLAB is running and the robot controller which drives the robot as shown in Figure 4 below. TCP/IP is the best choice used for this project because the robotic program RAPID used for ABB robots facilitates better support for socket communication than for any other communication protocol. The other advantage that it brings is that the controller of the ABB robot has an option for Local Area Network (LAN).
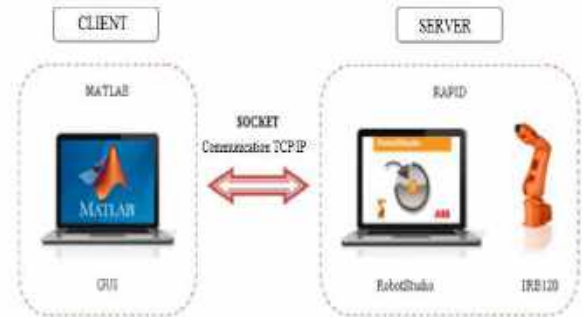
Figure 4 Relationship between client and server [13]

3. RESULTS

3.1 Experimental Setup

As shown in Figure 5 below, a cylindrical object of diameter 50mm and height of 45mm was used. It was placed on a plane surface perpendicular to the camera with zero elevation. Depth readings were taken at different distances from the camera. Depth measurement is the main parameter that we need to make sure that the robot moves to the correct height of a detected object. The actual distances were measured using a measuring tape.

The Kinect depth distance measured in this project was compared against the manufacturer's Microsoft Kinect for Windows SDK which is believed to give accurate measurements of depth and every other parameter. The SDK measurement was compared with two other measurements, the actual distance and the MATLAB program designed for the project, which in this case shall be called Kinect (project).

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...

Figure 5 Experimental setup

3.2 Results analysis

Measurements were made with the Kinect sensor and compared with the SDK which is known to give accurate measurements. Of all the measurements taken, measurements from the Kinect were close to the measurement of the SDK. The graph of the recorded measurement was compared in Figure 6 below:
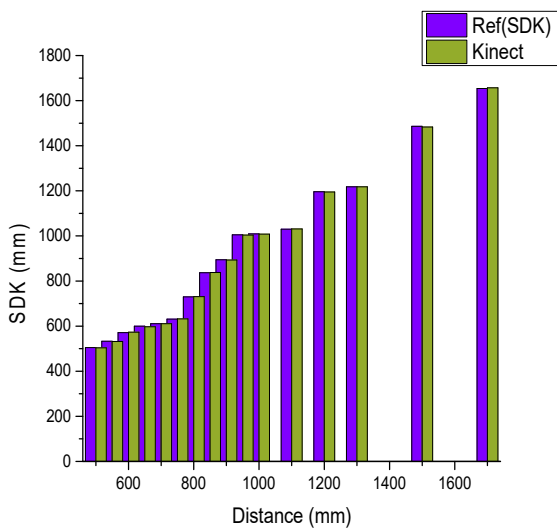
Figure 6 SDK and Kinect (Project) depth measurements

There was an average of 1.3mm of all the 16 measurements between the Kinect and the SDK. Figure 7 below shows the average differences between the Kinect (Project) and SDK.
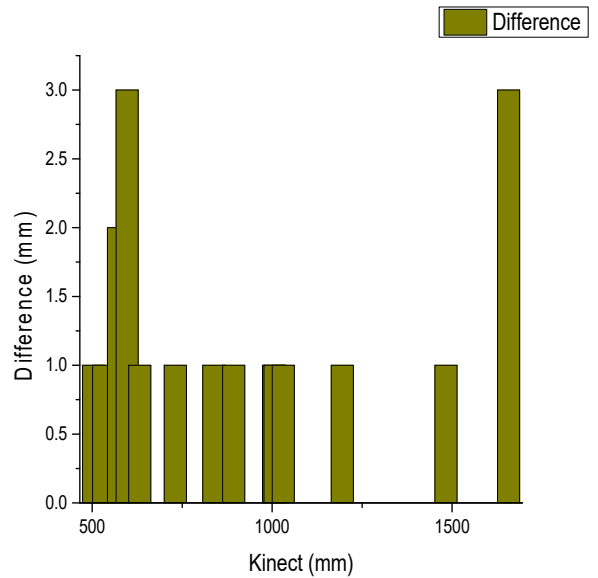
Figure 7 Average difference between Kinect and SDK

Absolute mean percentage error (AMPE), was calculated using the following formula;

$$Difference(mm) = \begin{pmatrix} SDK \\ Depth \\ Measurement \end{pmatrix} - \begin{pmatrix} Kinect \\ Dept \\ Measurement \end{pmatrix} (2)$$

$$AMPE(\%) = \left(\frac{|difference|}{\frac{SDK}{measurement}} \right) x100\% \qquad (3)$$

The graph of the distance measured by Kinect against AMPE is shown in Figure 8 below:

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...
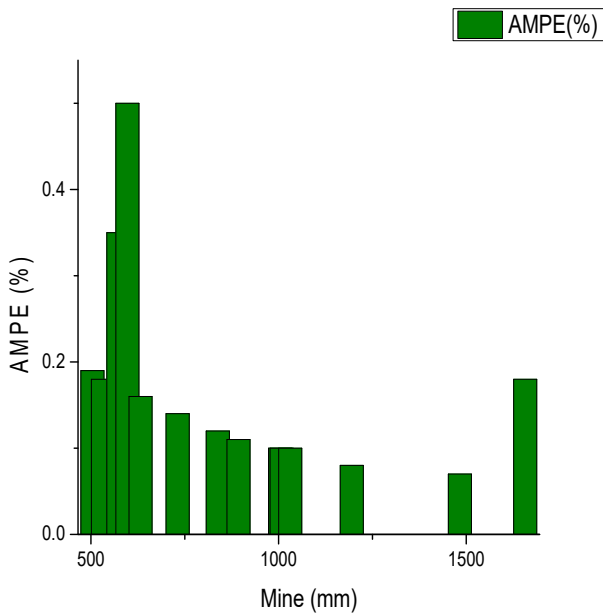
Figure 8 Distance measured by Kinect Vs AMPE

Figure 9 Comparison between SDK, Kinect and actual measured distance

The absolute mean percentage error of the measurements between the SDK and the Kinect measured depth in our project was 0.15%. This is almost a very insignificant error signifying that the code developed for this project was very accurate to SDK standard.

The difference between the SDK, Kinect (Project) and the actual distance was also analyzed. The graph in Figure 9 shows the measurement taken between the three. It can be seen from the graph that the measurement of these three gives almost the same measurement. We can therefore conclude that the actual measurement can also be used as equal to depth measurement because they are almost the same.

The average difference between the SDK and the actual measurement is just 2.8 mm. This turns the AMPE for the actual measurement with SDK to be 0.3%.

3.3 Conclusion

The algorithm for finding 3D object position according to color segmentation in real time was developed. The main focus was on how to find the depth of an object from the Kinect sensor. The height of the object was found by subtracting the distance from the depth of the plane where the objects were located to the sensor. This height was given to the robot and the robot was successfully moved to the location of the object, picked up and then placed in a specified location in real time. The average time taken to detect, select, and place an object was 8 seconds. Kinect was able to correctly distinguish color, and the robot could accurately navigate to the detected object. The algorithm developed was successful.

MAKOMO et al.

Real Time Application for Automatic Object and 3D Position Detection and Sorting with Robotic Manipul...

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The authors declared no conflict of interest.

Author Contributions

Conception:TJM-BB, Design:TJM-KE, Supervision:BB-KE-TJM, Materials:BB, Data Collection and/or Processing:TJM, Analysis and/or Interpretation:TJM, Literature Review:TJM, Writer:TJM, Critical Review:BB-KE-TJM

REFERENCES

[1] J. Hill and W. Park, "Real time control of a robot with a mobile camera," 9th International Symposium on Industrial Robots, pp. 233–246, 1979.

[2] K. Rezaie, S. Nazari Shirkouhi, and S.M. Alem, "Evaluating and selecting flexible manufacturing systems by integrating data envelopment analysis and analytical hierarchy process model," Asia International Conference on Modelling and Simulation, pp. 460–464, 2009.

[3] K. Hashimoto, "Visual Serving: Real Time Control of Robot Manipulators Based on Visual Sensory Feedback," 1993.

[4] Hutchinson, F. and Chaumette, S., "Visual servo control basic approaches," Robotics & Automation Magazine, IEEE, vol. 13, pp. 82-90, 2006.

[5] F. C. S. Hutchinson, "Visual servo control. ii. advanced approaches [tutorial]," Robotics & Automation Magazine, IEEE, vol. 14, pp. 109–118, 2007.

[6] D. Kragic, and H. I. Christensen, "Survey on visual servoing for manipulation," Computational Vision and Active Perception Laboratory Fiskartorpsv, vol. 15, 2002.

[7] H. Wu, W. Tizzano, T. Andersen, N. Andersen, and O. Ravn, "Hand-Eye Calibration and Inverse Kinematics of Robot Arm using Neural Network," Springer, pp. 581–591, 2013.

[8] H. Wu, L. Lu, C.-C. Chen, S. Hirche, and K. Khnlenz, "Cloud-based networked visual servo control," I E E E Transactions on Industrial Electronics, vol. 60, no 2, pp. 554 – 566,, 2013.

[9] Meyer, R. D. and Schraft, C., "The need for an intuitive teaching method for small and medium enterprises," ISR Robotik, Germany, 2012.

[10] B. Akan, "Human Robot Interaction Solutions for Intuitive Industrial Robot Programming," Västerås: Mälardalen University, 2012.

[11] Juan R. Terven and Diana M. Cordova, Kin2 User Guide, 2016.

[12] MathWorks, "Image Processig Toolbox User's Guide," 2014.

[13] N. B. Fernández, ""Generación de trayectorias y evitación de obstáculos para el robot IRB120 en entorno Matlab"," UNIVERSIDAD DE ALCALÁ, pp. 47, 2015.

Title: Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation:
Modelling Experimental Data Using Artificial Neural Networks

Authors: Bediha OYAR, Beytullah EREN, Abdil ÖZDEMİR

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Data Using Artificial Neural Networks

Bediha OYAR[1], Beytullah EREN[*2], Abdil ÖZDEMİR[3]

Abstract

The wide range of today's industry increases the diversity of pollutants in the wastewater characteristics. In particular, the wastewater of the textile industry is highly colored. Different techniques are used for color removal of dyes from wastewater. In this work, the removal efficiency of the textile dye (Reactive Black 5) at different current densities (48.5 A/m^2, 97.18 A/m^2, 194.36 A/m^2, 291.5 A/m^2, 388.7 A/m^2) was investigated by electrocoagulation method. The dye concentration of wastewater prepared in the laboratory scale was adjusted to 100 mg/L. Two iron electrodes and 3 g NaCl were used in the electrocoagulation system. The samples which taken periodically were measured after the centrifugal processes with the UV spectrophotometer. The experimental results were also modelled with artificial neural networks (ANNs). As a result of the experiments, approximately 90-100% color removal efficiency was obtained. According to the modelling study, the ANNs can predict the color removal efficiency with coefficient of determination (R^2) between the experimental and predicted output variable reached up to 0.99.

Keywords: Wastewater, electrocoagulation, textile dye (reactive black 5 (RB5)), color, artificial neural networks (ANNs).

[1]Sakarya University, Department of Environmental Engineering, Sakarya, Turkey.
 ORCID: https://orcid.org/0000-0003-1683-5531. E-mail: bdhoyr22@gmail.com
[*]Corresponding Author: beren@sakarya.edu.tr
[2]Sakarya University, Department of Environmental Engineering, Sakarya, Turkey.
 ORCID: https://orcid.org/ 0000-0001-6747-7004.
[3]Sakarya University, Department of Chemistry, Sakarya, Turkey.
 ORCID: https://orcid.org/ 0000-0002-0900-0221. E-mail: abdilo@sakarya.edu.tr

1. INTROCUTION

Water is one of the most important elements among naturel resources[1]. Water consumption has increased as a result of population growth and technology development. As a result of this, the amount of wastewater has also increased and so the receiving environments have been exposed to more pollutants. Polluting of the existing water resources has an important place among environmental problems. Large amounts of wastewater are produced every year from domestic and industrial areas, and this global problem continues to grow day by day[2]. The characteristic of wastewater varies according to the source of wastewater. Industrial wastewater such as textile, food production, cosmetics, paper production has a large share in the color problem that occurs in the receiving environments due to the synthetic dyes[3–5]. Textile industry is especially known as a one of the quickly growing industry which uses a large amount water in the production process[3]. Textile wastewater is contained with considerable amount of pollutant and high level color due to organic, inorganic chemicals and dye which were used in production process[6]. Approximately 50.000 tons of dye are discharged by textile industry into receiving environments every year[7]. This type of wastewater is regarded as harmful for receiving environments due to its toxicity and high organic components[8]. Removal of color in wastewater is highly difficult due to their resistance to light and oxidation[9]. However, the color in textile wastewater causes a decrease in dissolved oxygen concentration in receiving environments[10]. Due to all these adverse effects, such wastewater must be treated before discharge to the receiving environment and the pollutant concentrations must be reduced below the limit values specified in the regulations[11].Removal of color from wastewater is crucial for human health and sustainable water quality. In the literature, it was seen that many physical, chemical, biological and advanced treatment methods such as adsorption, biosorption, chemical oxidation, photolysis, chemical coagulation, flotation, ion exchange, electrodialysis, ozonation, membrane filtration were used in color removal from wastewater[12–21]. However, many of these methods require high investment and operating costs and also have various operational difficulties. In recent years, the interest in electrochemical methods has been increased considerably and is also preferred for color removal from wastewater. Electrochemical methods are frequently preferred due to the fact that they do not require chemical substances before and after wastewater treatment, produce less sludge, low space requirements, low initial investment cost and time savings[6,22].

Electroflotation, electrofloculation and electrocoagulation are some applications of electrochemical treatment methods[23]. In the literature, the electrocoagulation method, especially dye[6,11,19], selenium[24], heavy metal[25,26], polycyclic aromatic hydrocarbons (PAH)[27], phosphate[28], nitrogen[29], chemical oxygen demand (COD), oil and grease[30], drug[31] are used with high removal efficiency in the treatment of pollutants. The electrocoagulation method is also widely used because it is easily integrated into the treatment units and that simple equipment is sufficient for the treatment process.[32,33].

There are various parameters affecting the pollutant removal efficiency from wastewater by electrocoagulation method. The most important parameters effecting electrocoagulation process are pH, electrolyte concentration, current density, pollutant concentration, electrolysis time, electrode type.

Reactive dyestuffs especially used in textile industry cause black color formation in wastewater and degrade the water quality of the receiving environments. In this study, the effect of different current densities on the removal efficiency of Reactive Black 5 (RB5) by electrocoagulation was investigated. And also, the color removal efficiency was estimated by modeling the obtained experimental results with artificial neural networks (ANNs).

2. MATERIAL AND METHOD

2.1. ELECTROCOAGULATION

Electrocoagulation (EC) is a method developed to improve traditional methods in wastewater treatment[34]. This technology is a relatively low cost, easy to operate, environmentally friendly technology that requires a minimum of chemicals and produces a small amount of sludge. This method is a simultaneous operation, involving multiple chemical mechanisms[35].

The electrocoagulation process is considered to consist of 3 basic stages[7]. These stages are

1- Coagulant formation by oxidation of electrode (such as Al, Fe),
2- Stabilization of pollutants, particle suspension and emulsion breakage,
3- Collection of flocks in unstable phases[36].

Generally, direct current anode and cathode electrodes are preferred in EC process. Electrodes can vary in shape, size, number and material[37]. Various electrode types such as stainless steel, iron, aluminum, boron doped diamond, titanium coating are used as electrode material in the EC process[13,38,39]. However, it is seen that iron and aluminum electrodes are generally used due to the low cost and yield obtained[8,40–42].

If iron electrode is used in EC process, the following reactions occur[35]:

Anode:

$$Fe_{(s)} - 2e^- \rightarrow Fe^{+2}{}_{(aq)} \tag{1}$$

Cathode:

$$2H_2O(I) + 2e^- \rightarrow H_{2(g)} + 2OH^-{}_{(aq)} \tag{2}$$
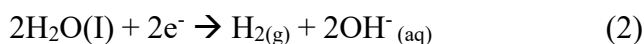
In electrocoagulation method, one of the important parameters affecting removal efficiency and treatment cost in color removal from wastewaters containing dye is current density. According to Faraday, the amount of cations dissolved from the anode is directly related to the current density passing through the wastewater[8,43].

It is observed in studies that the pollutant removal increases as the current density increases[43–46]. The mechanism of electrocoagulation is shown in Figure 1 below.
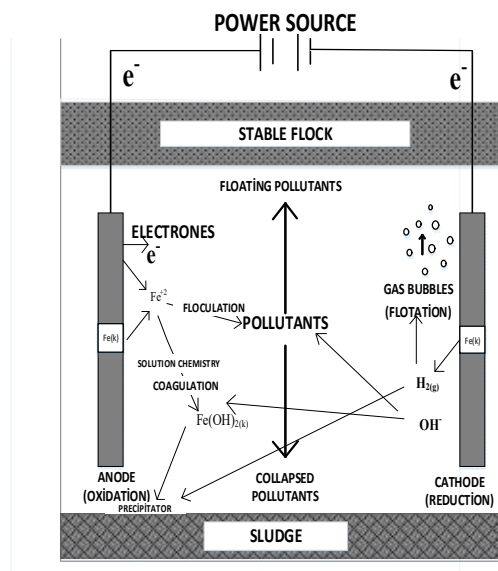
Figure 1. Mechanism of electrocoagulation[47]

2.2. EXPERIMENTAL MATERIALS

In this study, the removal of color (Reactive Black 5 (RB5) dye) at different current densities is investigated by using electrocoagulation method. The chemical bond structure of RB5 is shown in Figure 2. The solution used in experimental studies and containing dyestuff was prepared in the laboratory using ultra-pure water. For the calibration procedure, a calibration curve was created using 5 different concentrations of solution, 1 mg/L, 5 mg/L, 10 mg/L, 15 mg/L and 20 mg/L, and these is shown in Figure 3. Also, UV spectrum of absorption that is prepared for calibration is shown in Figure 4.

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...
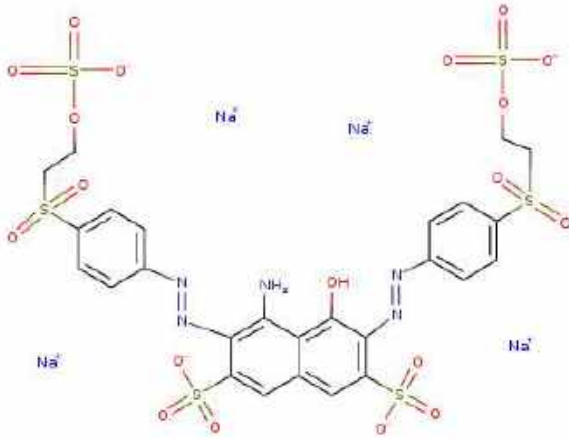
Figure 2. Chemical bond structure of Reactive Black 5
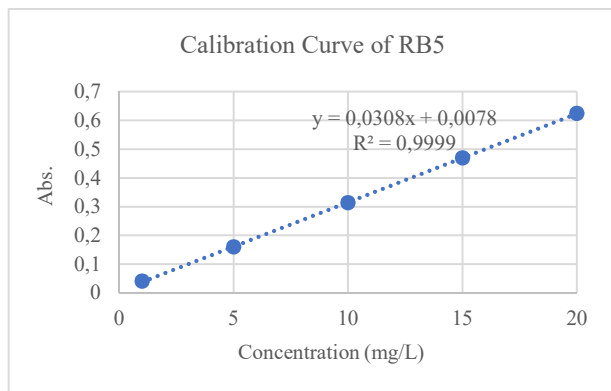
Figure 3. Calibration curve of RB5
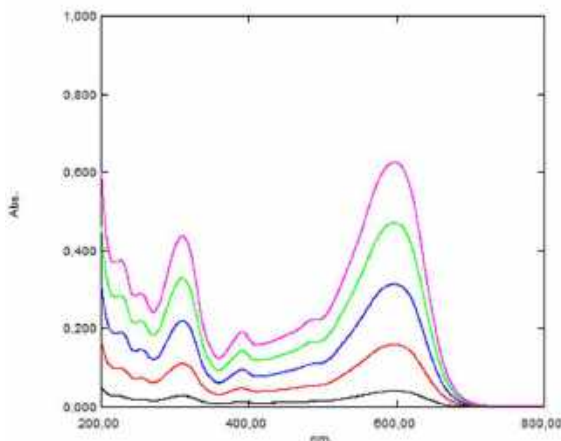
Figure 4. UV spectrum of RB5

2.3. EXPERIMENTAL METHOD

Electrocoagulation experimental setup used in this study was seen in Figure 5. The experimental setup consists of a 2000 ml glass beaker, two iron electrodes, a direct current power supply and a magnetic stirrer. The distance between the electrodes used in the experiments is 2 cm and the active surface area is 10.9 x 9.8 cm². In the experiment, firstly 100 mg/L RB5 dye concentration was prepared in glass beaker. Secondly, 3 g of NaCl was added to dye solution as electrolyte. Finally, the current densities of 48.5 A/m², 97.18 A/m², 194.36 A/m², 291.5 A/m², 388.7 A/m², respectively, were provided to contact the solution for 60 minutes for each experiment. The working environment is acidic because the RB5 dye has a natural pH of 5.8 and no pH adjustment is made to the waste water sample[48].
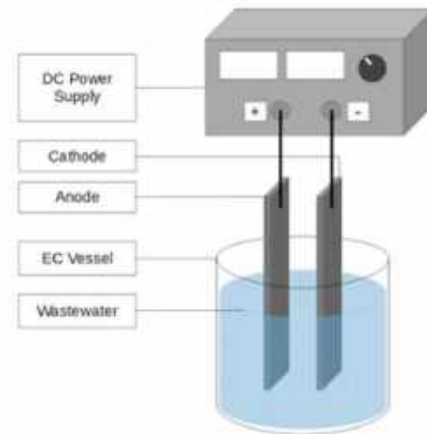
Figure 5. Electrocoagulation experimental setup[37]

5 ml samples were taken at 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 45, 60 minutes in all current densities (48.5 A/m², 97.18 A/m², 194.36 A/m², 291.5 A/m², 388.7 A/m²). After the samples were centrifuged, the measurement of all samples, including calibration, was performed at UV-Vis recording spectrophotometer at intervals of 0.5 nm between 200-800 nm. The maximum absorbance value for Reactive Black 5 was found to be 600 nm as a result of the studies.

The removal efficiency of color was calculated using to below formula.

$$C_{color}(\%) = \frac{C_0 - C}{C_0} \times 100 \tag{3}$$

$C_{color}(\%)$: The removal efficiency of color (%)

C_o : Initial dye concentration

C : Final dye concentration

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...

2.4. COST ANALYSIS

Electricity consumption is taken into account while calculating the cost of color removal from waste water with electrocoagulation process. In measurement price of unit was accepted as 0,008 $/kWh for the electricity consumption[49].

Amount of electricity consumption was calculated using equation below[33].

$$C_{Energy} = \frac{U \ x \ I \ x \ t}{v} \tag{4}$$

C_{Energy} : Consumption of electricity (kWh/m^3)

U : Volt

I : Current

t : Electrolysis time (hour)

v : Solution volume (m^3)
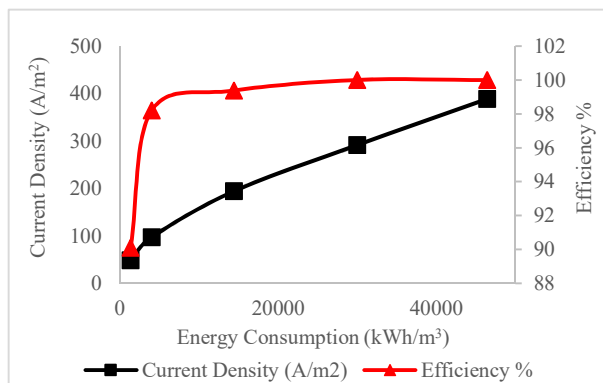
Figure 6. Effect of current density on consumption of electrical energy
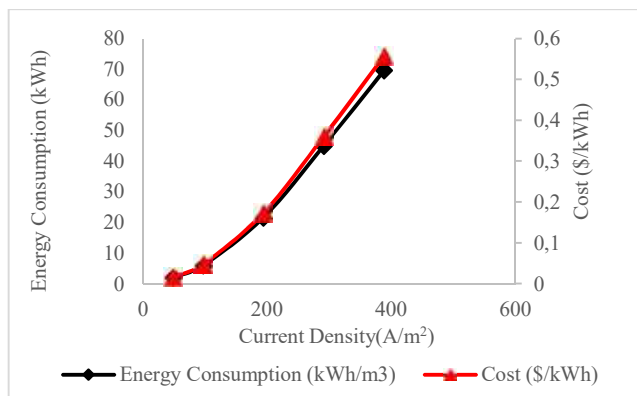
Figure 7. The relationship between energy consumption and cost

2.5. ARTIFICIAL NEURAL NETWORKS

In recent years, interest in modeling has increased in the process of estimating the status and optimum performance related to the treatment process in wastewater treatment plants. Artificial intelligence-based methods are often preferred in order to determine optimum performance in several disciplines covering water resources and environmental science. Artificial neural networks (ANNs) are the most remarkable modeling method among artificial intelligence-based methods. When compared with traditional mathematical methods, ANNs have advantages such as the lack of necessity to make mathematical definitions of events in the process, having the ability to predict using a limited number of experiments [50].

The nervous system of the human body has inspired the creation of ANNs, and these systems are important artificial intelligence systems capable of solving a number of complex problems[40,51–56].

ANNs have three layers including the input layer, the hidden layer(s) and the output layer.

In artificial neural networks, the input layer consists of independent variables, while the output layer consists of dependent variables. Between these two layers are hidden layers containing one or more neurons in layers[57]. ANNs create many connections between process elements to solve certain problems in harmony[34]. Schematic structure of the ANNs are showed in Figure 8.
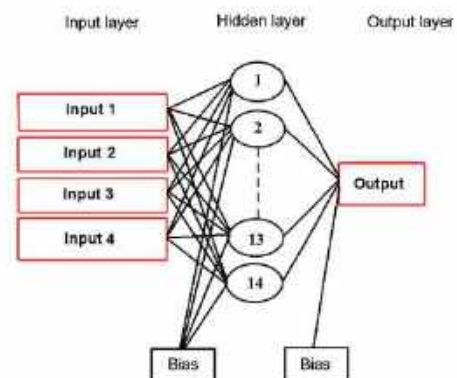
Figure 8. Structure of artificial neural network[58]

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...

The electrocoagulation method provides high efficiency in pollutant removal from wastewater, as well as the fact that it contains complex chemical events throughout the EC process, raises the idea of modeling. Recently, Artificial neural networks (ANNs) successfully applied to solve many environmental engineering problems. Modeling of complex EC systems with ANNs make it a promising modeling technique [57,59].

3. RESULT AND DISCUSSION

3.1. Determination of the effect of current density on removal efficiency of RB5 dye

There are several operating parameters that is important in electrochemical processes. Electrode properties and current density is in among these parameters. The current density can be defined as

the amount of current passing through the unit electrode area. In this study, 5 different current densities were applied to find the most suitable current density in the removal of RB5 dye by electrocoagulation process. Experimental studies were carried out using an iron electrode. Current density that effect on electricity consumption was decided as 48.5 A/m^2, 97.18 A/m^2, 194.36 A/m^2, 291.5 A/m^2, 388.7 A/m^2. During experiment current density was adjusted by direct current power supply.

The effect of current density on removal efficiency of color is shown in Figure 9.

In Figure 9, It is seen that when the current density is 48.5 A/m^2, 90% removal efficiency is obtained in the 20th minute, and when the current density is 388.7 A/m^2, the same removal efficiency is reached in the 4th minute.
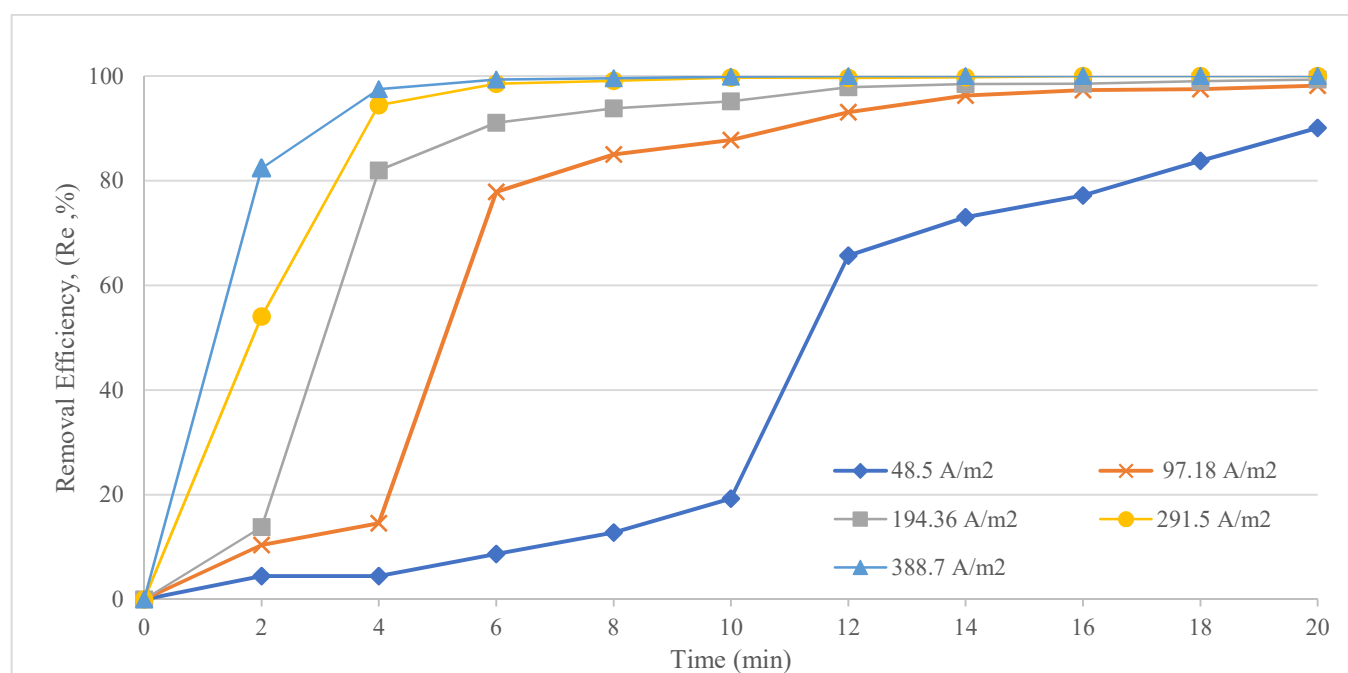
Figure 9. The effect of current density on removal efficiency of color (R$_e$, %)

3.2. ANN modeling result of removal efficiency of RB5 dye

A total of 7 ANNs model was derived from a combination of 6 different topologies (LR: linear regression; MLP: multilayer perceptron; PNN: probabilistic neural network; RBF: radial basis

function; GFF: generalized feed forward; MLPPCA: MLP with principal component analysis), up to two hidden layers (1 vs. 2), two learning algorithms (M: Momentum vs. L: Levenberg-Marquardt) and two learning modes (B: batch vs. O: online), and (N: None). In ANN modeling, experimental dataset was randomly

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...

partitioned into training, cross validation and testing datasets with the allocation order of 75%, 10% and 15%, respectively. Datasets were fed into a total of 7 ANNs to predict removal efficiency of color (Re, %) which included as inputs time (T, min) and current density (CD, A). Performance statistics used for comparing ANNs were mean square error (MSE), coefficient of correlation (R^2) and mean absolute error (MAE) as follow[60,61]:

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(yi-\hat{y})^2 \dots\dots\dots\dots\dots\dots\dots(5)$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|yi-\hat{y}| \dots\dots\dots\dots\dots\dots\dots\dots\dots(6)$$

$$R^2=1-\frac{\sum(y_i-\hat{y})^2}{\sum(y_i-\bar{y})^2} \dots\dots\dots\dots\dots\dots\dots\dots\dots(7)$$

Where,

$\hat{Y}$: Predicted value of y

$\bar{y}$: Mean value of y

Table 1. A comparison of 7 artificial neural networks (ANNs) developed as a function of time (T, min), Current Density (C, A), by Removal Efficiency (R_e, %) of Reactive Black 5 based on performance metrics of training, cross-validation and independent validation.

Artificial Neural Networks Model	Training (n=26)			Cross Validation (n=5)			Testing (n=4)		
	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE
LR-0-B-L (Linear Regression)	0.04452	0.89928	0.18346	0.07320	0.85752	0.22640	0.02877	0.97111	0.12602
MLP-1-B-L (Multilayer Perceptron)	0.00044	0.99931	0.01407	0.00931	0.98353	0.06069	0.00544	0.99674	0.05332
PNN-0-N-N (Probabilistic Neural Network)	0.05441	0.93956	0.18459	0.11850	0.87575	0.27209	0.0432	0.99803	0.14108
RBF-1-B-L (Radial Basis Function)	0.01072	0.97962	0.068955	0.08618	0.96882	0.20784	0.00212	0.99985	0.03287
GFF-1-B-L (Generalized Feedforward)	0.00103	0.99830	0.02412	0.00774	0.98780	0.07783	0.00433	0.99791	0.04745
MLPPCA-1-B-L (MLP with PCA)	0.02966	0.95300	0.13308	0.07704	0.81495	0.17828	0.01095	0.98969	0.07554
*MLP-2-B-L (Multilayer Perceptron)**	*0.00028*	*0.99963*	*0.00937*	*0.00135*	*0.99734*	*0.03112*	*0.00352*	*0.99969*	*0.03609*

*The best performing ANN is designated in italic.
Explanations of ANN Model Topologies: LR: linear regression; MLP: multilayer perceptron; PNN: probabilistic neural network; RBF: radial basis function; GFF: generalized feed forward; MLPPCA: MLP with principal component analysis), up to two hidden layers (1 vs. 2), two learning algorithms (M: Momentum vs. L: Levenberg-Marquardt) and two learning modes (B: batch vs. O: online), and (N:None)

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...
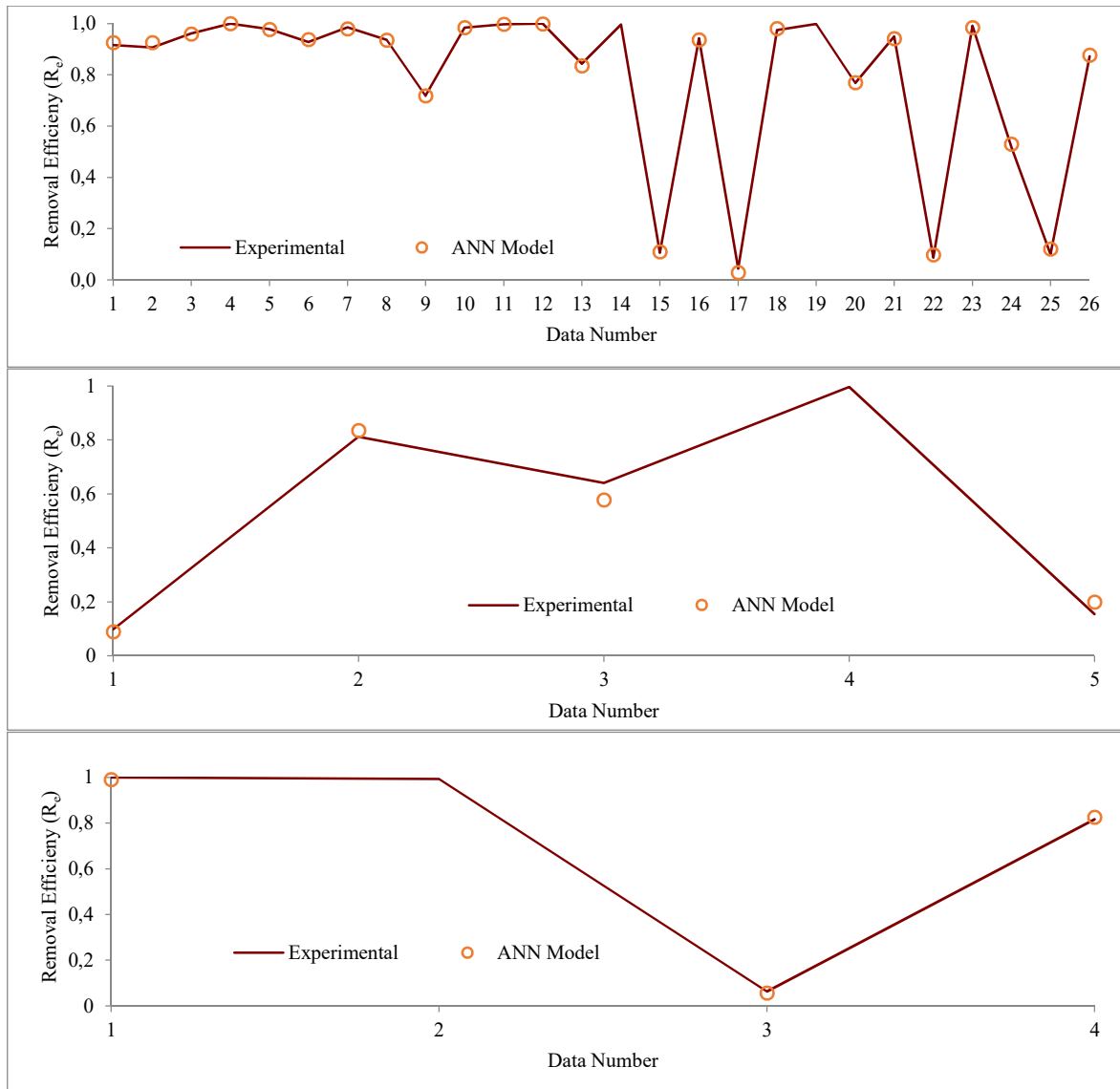
Figure 10. The comparison of Experimental and ANN model results by the best performing neural networks (MLP-2-B-L) for training, cross validation and testing. Solid and discrete lines refer to experimental and ANN model predicted values of Removal Efficiency (R_e), respectively.

The performance statistics derived from training, cross validation and testing of the 7 ANNs indicated that MLP-2-B-L performed best among the other ANNs as a function of Time and Current Density (Table 1). The predictive performance of the ANNs in terms of R^2 values based on testing ranged from 0.97111 by LR-0-B-L to 0.99969 by MLP-2-B-L. Testing-derived values of R^2, RMSE and MAE pointed to MLP topologies (MLP-2-B-L) as the best ANNs.

4. CONCLUSION

In this study, the removal efficiency of color due to Reactive Black 5 dye at different current densities using electrocoagulation method from aqueous solutions was investigated. The aqueous solution prepared in the laboratory scale was analyzed using a UV spectrophotometer and the color removal efficiencies were calculated. It has been observed that color removal increases with increasing current density. It has been calculated that the removal efficiency after 20 min experimental period, 90,09 %, 98,21%, 99,38%,

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...

100%, 100% is achieved for 48.5 A/m^2, 97.18 A/m^2, 194.36 A/m^2, 291.5 A/m^2 and 388.7 A/m^2 current densities, respectively. It is seen that the color removal efficiency increases with increasing current density. However, the electricity cost increases with the increasing of current density. Therefore, the optimum current density was determined as 97.18 A/m^2.

Also, the experimental results are modeled by artificial neural networks (ANNs). According to ANN modelling, it is seen that R^2 between experimental and predicted output variable varies in the range of 0.89-0.99. The performance statistics derived from training, cross validation and testing of the 7 different ANN models indicated that MLP-2-B-L (as in section 3.2) performed best among the other ANNs as a function of Time and Current Density. These results show that developed ANN model has good predictive ability for the removal efficiency of color.

Acknowledgements

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] Noreen, U., Ahmed, Z., Khalid, A., Di Serafino, A., Habiba, U., Ali, F., and Hussain, M., "Water pollution and occupational health hazards caused by the marble industries in district Mardan, Pakistan", Environ. Technol. Innov., 16, p. 100470, 2019.

[2] Natarajan, S., Bajaj, H. C., and Tayade, R. J., "Recent advances based on the synergetic effect of adsorption for removal of dyes from waste water using photocatalytic process", J. Environ. Sci., 65, pp. 201–222, 2018.

[3] Tavangar, T., Jalali, K., Alaei Shahmirzadi, M. A., and Karimi, M., "Toward real textile wastewater treatment: Membrane fouling control and effective fractionation of dyes/inorganic salts using a hybrid electrocoagulation – Nanofiltration process", Sep. Purif. Technol., 216, pp. 115–125, 2019.

[4] Moussavi, G. and Mahmoudi, M., "Removal of azo and anthraquinone reactive dyes from industrial wastewaters using MgO nanoparticles", J. Hazard. Mater., 168(2–3), pp. 806–812, 2009.

[5] Jayasantha Kumari, H., Krishnamoorthy, P., Arumugam, T. K., Radhakrishnan, S., and Vasudevan, D., "An efficient removal of crystal violet dye from waste water by adsorption onto TLAC/Chitosan composite: A novel low cost adsorbent", Int. J. Biol. Macromol., 96, pp. 324–333, 2017.

[6] Khemila, B., Merzouk, B., Chouder, A., Zidelkhir, R., Leclerc, J.-P., and Lapicque, F., "Removal of a textile dye using photovoltaic electrocoagulation", Sustain. Chem. Pharm., 7, pp. 27–35, 2018.

[7] Pajootan, E., Arami, M., and Mahmoodi, N. M., "Binary system dye removal by electrocoagulation from synthetic and real colored wastewaters", J. Taiwan Inst. Chem. Eng., 43(2), pp. 282–290, 2012.

[8] Keyikoglu, R., Can, O. T., Aygun, A., and Tek, A., "Comparison of the effects of various supporting electrolytes on the treatment of a dye solution by electrocoagulation process", Colloid Interface Sci. Commun., 33, p. 100210, 2019.

[9] Tian S., Xu S., Liu J., He C., Xiong Y., Feng P., "Highly efficient removal of both cationic and anionic dyes from wastewater with a water-stable and eco-friendly Fe-MOF via host-guest encapsulation", J. Clean. Prod., 239, p. 117767, 2019.

[10] Beluci, N. de C. L., Mateus, G. A. P., Miyashiro, C. S., Homem, N. C., Gomes, R. G., Fagundes-Klen, M. R., Bergamasco, R., and Vieira, A. M. S., "Hybrid treatment of coagulation/flocculation process followed by ultrafiltration in TIO2-modified membranes to improve the removal of reactive black 5 dye", Sci. Total Environ., 664, pp. 222–229, 2019.

[11] Ya V., Martin N., Choo K., Chou Y., Lee S., Le N., Li C., "High-pressure electrocoagulation system with periodic air replenishment for efficient dye wastewater treatment: Reaction dynamics and cost evaluation", J. Clean. Prod., 213, pp. 1127–1134, 2019.

[12] Santhy, K. and Selvapathy, P., "Removal of reactive dyes from wastewater by adsorption on coir pith activated carbon", Bioresour. Technol., 97(11), pp. 1329–1336, 2006.

[13] Kocaer, F. O. and Alkan, U., "Boyar madde içeren tekstil atıksularının arıtım alternatifleri", p. 9 (n.d.).

[14] Hassan, S. S. M., Awwad, N. S., and Aboterika, A. H. A., "Removal of synthetic reactive dyes from textile wastewater by Sorel's cement", J. Hazard. Mater., 162(2–3), pp. 994–999, 2009.

[15] Papic, S., "Removal of some reactive dyes from synthetic wastewater by combined Al(III) coagulation/carbon adsorption process", Dyes Pigments, 62(3), pp. 291–298, 2004.

[16] Khan, A. J., Song, J., Ahmed, K., Rahim, A., Onófrio Volpe, P. L., and Rehman, F., "Mesoporous silica MCM-41, SBA-15 and derived bridged polysilsesquioxane SBA-PMDA for the selective removal of textile reactive dyes from wastewater", J. Mol. Liq., p. 111957, 2019.

[17] Núñez, J., Yeber, M., Cisternas, N., Thibaut, R., Medina, P., and Carrasco, C., "Application of electrocoagulation for the efficient pollutants removal to reuse the treated wastewater in the dyeing process of the textile industry", J. Hazard. Mater., 371, pp. 705–711, 2019.

[18] Babu, A. N., Reddy, D. S., Sharma, P., Kumar, G. S., Ravindhranath, K., and Mohan, G. V. K., "Removal of Hazardous Indigo Carmine Dye from Waste Water Using Treated Red Mud", Mater. Today Proc., 17, pp. 198–208, 2019.

[19] Meriç, S., Kaptan, D., and Ölmez, T., "Color and COD removal from wastewater containing Reactive Black 5 using Fenton's oxidation process", Chemosphere, 54(3), pp. 435–441, 2004.

[20] Barhoumi, A., "High-rate humic acid removal from cellulose and paper industry wastewater by combining electrocoagulation process with adsorption onto granular activated carbon", Ind. Crops Prod. 140, 111715, 2019.

[21] El-Zahhar, A. A., Awwad, N. S., and El-Katori, E. E., "Removal of bromophenol blue dye from industrial waste water by synthesizing polymer-clay composite", J. Mol. Liq., 199, pp. 454–461, 2014.

[22] İlhan F., Kurt U., Apaydın Ö., Arslankaya E., M.M., "Elektrokimyasal Arıtım ve Uygulamaları: Katı Atık Sızıntı Suyu Çalışması", 18 (2-3), pp. 3–12, 2008.

[23] Bahadir, E. B., "Tekstil endüstrisi arıtılmış atıksularında renk ve öncelikli kirleticilerin ozon teknolojisi ile gideriminin araştırılması", Namık Kemal Üniversitesi, Fen Bilimleri Enstitüsü, Master Tezi, p. 140, (2012).

[24] Hansen, H. K., Peña, S. F., Gutiérrez, C., Lazo, A., Lazo, P., and Ottosen, L. M., "Selenium removal from petroleum refinery wastewater using an electrocoagulation technique", J. Hazard. Mater., 364, pp. 78–81, 2019.

[25] Kim, T., Kim, T.-K., and Zoh, K.-D., "Removal mechanism of heavy metal (Cu, Ni, Zn, and Cr) in the presence of cyanide during electrocoagulation using Fe and Al electrodes", J. Water Process Eng., 33, p. 101109, 2020.

[26] AlJaberi, F. Y., "Studies of autocatalytic electrocoagulation reactor for lead removal from simulated wastewater", J. Environ. Chem. Eng., 6(5), pp. 6069–6078, 2018.

[27] Gong, C., Shen, G., Huang, H., He, P., Zhang, Z., and Ma, B., "Removal and transformation of polycyclic aromatic hydrocarbons during electrocoagulation treatment of an industrial wastewater", Chemosphere, 168, pp. 58–64, 2017.

[28] Dura, A. and Breslin, C. B., "The removal of phosphates using electrocoagulation with Al−Mg anodes", J. Electroanal. Chem., 846, p. 113161, 2019.

[29] Li, L., Qian, G., Ye, L., Hu, X., Yu, X., and Lyu, W., "Research on the enhancement of biological nitrogen removal at low temperatures from ammonium-rich wastewater by the bio-electrocoagulation technology in lab-scale systems, pilot-scale systems and a full-scale industrial wastewater treatment plant", Water Res., 140, pp. 77–89, 2018.

[30] Priya, M. and Jeyanthi, J., "Removal of COD, oil and grease from automobile wash water effluent using electrocoagulation

technique.", Microchem. J., 150, p. 104070, 2019.

[31] Baran, W., Adamek, E., Jajko, M., and Sobczak, A., "Removal of veterinary antibiotics from wastewater by electrocoagulation", Chemosphere, 194, pp. 381–389, 2018.

[32] Hashim, K. S., "Energy efficient electrocoagulation using baffle-plates electrodes for efficient Escherichia coli removal from wastewater", J. Water Process Eng., p. 7, 2020.

[33] Bener, S., Bulca, Ö., Palas, B., Tekin, G., Atalay, S., and Ersöz, G., "Electrocoagulation process for the treatment of real textile wastewater: Effect of operative conditions on the organic carbon removal and kinetic study", Process Saf. Environ. Prot., 129, pp. 47–54, 2019.

[34] Firat, M. and Güngör, M., "Askı Madde Konsantrasyonu ve Miktarının Yapay Sinir Ağları ile Belirlenmesi", 15 (73), pp. 3267–3282, 2004.

[35] Zhang, F., Yang, C., Zhu, H., Li, Y., and Gui, W., "An integrated prediction model of heavy metal ion concentration for iron electrocoagulation process", Chem. Eng. J., p. 123628, 2019.

[36] Altınten, A. A., Demirci, Y. D., Pekel, L. C. P., and Alpbaz, M. A., "Elektrokoagülasyon reaktöründe bulanık kontrol metodu ile ph, iletkenlik ve sıcaklığın eş zamanlı kontrolü", Gazi Üniversitesi Mühendis.-Mimar. Fakültesi Derg., 31(4), p. 0, 2016.

[37] Reilly, M., Cooley, A. P., Tito, D., Tassou, S. A., and Theodorou, M. K., "Electrocoagulation treatment of dairy processing and slaughterhouse wastewaters", Energy Procedia, 161, pp. 343–351, 2019.

[38] Isik, Z., Arikan, E. B., Ozay, Y., Bouras, H. D., and Dizge, N., "Electrocoagulation and electrooxidation pre-treatment effect on

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...

fungal treatment of pistachio processing wastewater", Chemosphere, 244, p. 125383, 2020.

[39] Llanos J., Cotillas S., Caniares P., Rodrigo M.A., "Effect of bipolar electrode material on the reclamation of urban wastewater by an integrated electrodisinfection/ electrocoagulation process", Water Res., 53, pp. 329–338, 2014.

[40] da Silva Ribeiro, T., Grossi, C. D., Merma, A. G., dos Santos, B. F., and Torem, M. L., "Removal of boron from mining wastewaters by electrocoagulation method: Modelling experimental data using artificial neural networks", Miner. Eng., 131, pp. 8–13, 2019.

[41] Tchamango S.R., Darchen A., "Investigation and optimization of a new electrocoagulation reactor with horizontal bipolar electrodes: Effect of electrode structure on the reactor performances", J. Environ. Chem. Eng., 6(4), pp. 4546–4554, 2018.

[42] Aygün, A. and Eren, B., "Elektrokoagülayon Yöntemiyle Reaktif Yellow 160 Boyar Maddesinin Giderimi", Acad. Platf.-J. Eng. Sci., pp. 10–18, 2017.

[43] An, C., Huang, G., Yao, Y., and Zhao, S., "Emerging usage of electrocoagulation technology for oil removal from wastewater: A review", Sci. Total Environ., 579, pp. 537–556, 2017.

[44] Ammar, S. H., Ismail, N. N., Ali, A. D., and Abbas, W. M., "Electrocoagulation technique for refinery wastewater treatment in an internal loop split-plate airlift reactor", J. Environ. Chem. Eng., 7(6), p. 103489, 2019.

[45] Deveci, E. Ü., Akarsu, C., Gönen, Ç., and Özay, Y., "Enhancing treatability of tannery wastewater by integrated process of electrocoagulation and fungal via using RSM in an economic perspective", Process Biochem., 84, pp. 124–133, 2019.

[46] Abbasi S., Mirghorayshi M., Zinadini S., Zinatizadeh A.A., "A novel single continuous electrocoagulation process for treatment of licorice processing wastewater: Optimization of operating factors using RSM", Process Saf. Environ. Prot., 2019.

[47] Gürbüz G, Deri Endüstrisi Atıksılarının Elektrokoagülasyon ve Elektro-Fenton Yöntemleri ile Arıtımında RSM Kullanılarak Proses Optimizasyonu "T.C. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü" Master Tezi, 2015.

[48] Hadibarata T, Adnan LA, Yusoff ARM, Yuniarto A, Rubiyatno, Zubir MMFA, vd. Microbial Decolorization of an Azo Dye Reactive Black 5 Using White-Rot Fungus Pleurotus eryngii F032. Water Air Soil Pollut., 224(6),1595, 2013.

[49] AlJaberi, F. Y., "Operating cost analysis of a concentric aluminum tubes electrodes electrocoagulation reactor", Heliyon, 5(8), p. e02307, 2019.

[50] Hasani, G., Daraei, H., Shahmoradi, B., Gharibi, F., Maleki, A., Yetilmezsoy, K., and McKay, G., "A novel ANN approach for modeling of alternating pulse current electrocoagulation-flotation (APC-ECF) process: Humic acid removal from aqueous media", Process Saf. Environ. Prot., 117, pp. 111–124, 2018.

[51] Bhatti, M. S., Kapoor, D., Kalia, R. K., Reddy, A. S., and Thukral, A. K., "RSM and ANN modeling for electrocoagulation of copper from simulated wastewater: Multi objective optimization using genetic algorithm approach", Desalination, 274(1–3), pp. 74–80, 2011.

[52] Eren, B., "Prediction Of Adsorption Efficiency For The Removal Of Nickel(II) Ions By Zeolite Using Artificial Neural Network(ANN)", Fresenius Environ. Bull. 20(12), pp. 3158-3165, 2011.

[53] Eren, B., Ileri, R., Dogan, E., Caglar, N., and Koyuncu, I., "Development of artificial

OYAR et al.

Removal of Reactive Black 5 from Polluted Solutions by Electrocoagulation: Modelling Experimental Dat...

neural network for prediction of salt recovery by nanofiltration from textile industry wastewaters", Desalination Water Treat., 50(1–3), pp. 317–328, 2012.

[54] Eren, B., Yaqub, M., and Eyüpoğlu, V., "Assessment of Neural Network training algorithms for the prediction of Polymeric Inclusion Membranes Efficiency", SAÜ Fen Bilim. Enstitüsü Derg., 20(3), 2016.

[55] Eren, B., Yaqub, M., and Eyupoglu, V., "A comparative study of artificial neural network models for the prediction of Cd removal efficiency of polymer inclusion membranes", Desalination Water Treat., 143, pp. 48–58, 2019.

[56] Yaqub, M., Eren, B., and Eyupoglu, V., "Soft computing techniques in prediction Cr(VI) removal efficiency of polymer inclusion membranes", Environ. Eng. Res., 25(3), pp. 418–425, 2019.

[57] Aber, S., Amani-Ghadim, A. R., and Mirzajani, V., "Removal of Cr(VI) from polluted solutions by electrocoagulation: Modeling of experimental results using artificial neural network", J. Hazard. Mater., 171(1–3), pp. 484–490, 2009.

[58] Khataee, A. R. and Kasiri, M. B., "Artificial neural networks modeling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis", J. Mol. Catal. Chem., 331(1–2), pp. 86–100, 2010.

[59] Daneshvar, N., Khataee, A. R., and Djafarzadeh, N., "The use of artificial neural networks (ANN) for modeling of decolorization of textile dye solution containing C. I. Basic Yellow 28 by electrocoagulation process", J. Hazard. Mater., 137(3), pp. 1788–1795, 2006.

[60] Dogan, E., Ates, A., Yilmaz, E. C., and Eren, B., "Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand", Environ. Prog., 27(4), pp. 439–446, 2008.

[61] Eyupoglu, V., Eren, B., and Dogan, E., "Prediction of Ionic Cr (VI) Extraction Efficiency in Flat Sheet Supported Liquid Membrane Using Artificial Neural Networks (ANNs)", Int. J. Environ. Res., 4, pp. 463–470, 2010.

Sakarya University Journal of Science

Title: Deep Learning Based Automatic Speech Recognition For Turkish

Authors: Burak TOMBALOĞLU, Hamit ERDEM

Deep Learning Based Automatic Speech Recognition for Turkish

Burak TOMBALOĞLU[*1], Hamit ERDEM[2]

Abstract

Using Deep Neural Networks (DNN) as an advanced Artificial Neural Networks (ANN) has become widespread with the development of computer technology. Although DNN has been applied for solving Automatic Speech Recognition (ASR) problem in some languages, DNN-based Turkish Speech Recognition has not been studied extensively. Turkish language is an agglutinative and a phoneme-based language. In this study, a Deep Belief Network (DBN) based Turkish phoneme and speech recognizer is developed. The proposed system recognizes words in the system vocabulary and phoneme components of out of vocabulary (OOV) words. Sub-word (morpheme) based language modelling is implemented into the system. Each phoneme of Turkish language is also modelled as a sub-word in the model. Sub-word (morpheme) based language model is widely used for agglutinative languages to prevent excessive vocabulary size. The performance of the suggested DBN based ASR system is compared with the conventional recognition method, GMM (Gaussian Mixture Method) based Hidden Markov Model (HMM). Regarding to performance metrics, the recognition rate of Turkish language is improved in compare with previous studies.

Keywords: Deep Neural Networks, Automatic Speech Recognition, Deep Belief Networks, Turkish

---

[*]Corresponding Author: buraktombaloglu@hotmail.com
[1]Başkent University, Electrical and Electronics Engineering, Ankara, Turkey.
 ORCID: https://orcid.org/0000-0003-3994-0422
[2]Başkent University, Electrical and Electronics Engineering, Ankara, Turkey.
 ORCID: https://orcid.org/0000-0003-1704-1581. E-mail: herdem@baskent.edu.tr

1. INTRODUCTION

Automatic Speech Recognition is basically the detection and recognition of human voice by an electronic device. This system converts speech to text using various detecting and pattern recognition methods. With the developments in computers and advanced processors, area of ASR has been sprouted. Air traffic control, ticket reservations, security and biometric identification, gaming, devices control in cars and home automation, robotic are some of these applications. Moreover, this system can be used to improve life quality of disabled and elderly persons.

The most advanced ASR systems are developed for popular and widely spoken languages, e.g. Chinese, Spanish and English. These languages have extensive transcriptional speech and text data to create acoustic and language models. Pronunciation dictionaries are generated with the help of pronunciation rules and acoustic units examined by linguists [1,2].

Turkish is an agglutinative language like Finnish or Hungarian. There are suffixes which are attached to word root. Therefore as the number of the different vocabularies gets higher, the performance of speech recognizers gets lower. Using a vocabulary of sub-words (morphemes) instead of words is also a common solution. Sub-word based ASR is commonly applied for languages such as Estonian, Finnish, Turkish, Thai, Hungarian, Slovenian and Czech, which are agglutinative languages [3]. A sub-word n-gram model can assign probabilities to unseen word forms [4]. In the past, a smaller vocabulary size was important for the recognition performance of Large Vocabulary Continuous Speech recognition (LVCSR) systems, but it is now useful for the latest Neural Network Language Models (NNLMs). Excessive vocabulary leads to increased size of the input and output layers. Therefore, it is not practical. Although class-based models can reduce the layers, methods or shortlists like hierarchical softmax, the sub-word models provide an effective and natural way to lower the dimensionality [5]. In ASR applications of agglutinative languages, Hungarian Finnish and Turkish, [3-8] sub-word based language models outperform word-based language models. Therefore sub-word based language model is preferred for training the proposed ASR system.

The GMM is generally used when calculating the probability $(p (A \mid Q))$ of the acoustic feature vectors (A) conditioned by the HMM phonetic / status tags (Q) in the HMM [9]. HMM is often used as acoustic model by the ASR systems. The phonemes in the utterance are predicted by using GMM-HMM model. Then the word or set of continuous words that were spoken are determined [10]. The posterior probability of each state in HMM is estimated by these architectures which use Perceptual Linear Predictions (PLPs) and Mel Frequency Cepstral Coefficient (MFCCs) as features. ASR problem for Turkish language has been tried to be solved by conventional methods. One of them is GMM-HMM which has been mostly used in the applications [11-13]. In [11], the Phoneme error Rate (PER) is 29.3%, the Word Error Rate (WER) observed in [12],[13] are 21.46% and 32.88% respectively. The performance of Turkish ASR can be improved by using advanced methods similar to Deep learning.

In ASR applications, Gaussian mixtures for speech recognition have successfully been replaced by the Deep Learning approach. Previous studies show that, in ASR applications, DNNs work better than GMM-HMM systems [10]. DNN includes trained hidden layers. Hidden layers provide large output which is required for big number of HMM states. Each phone is modelled by triphone HMMs which cause large number of states. Many studies show that DNN outperforms GMM at acoustic modelling which is used for ASRs having large datasets and vocabularies. Instead of using GMMs, DNNs are used to produce posterior probabilities over HMM states as output. In the study, pre-trained DBN-DNN is used when calculating the probability of the acoustic feature vectors. HMM or HMM states are used to model units of speech (phones, sub-phones, phone states, syllables, words, etc.). Deep Learning is still not effectively implemented in recognition

of Turkish speech. There are few studies on DNN based ASR Applications of Turkish Language. In [9], an automatic dictation and keyword search system is designed for processing spoken lectures in Turkish. A DNN based LVCSR system is developed. The purposed system is trained by Turkish News TV program recordings and Law lecture video recordings. In [14], DNN is also used for building acoustic model. In [15], GMM and DNN based models are trained and tested using the corpus which is developed in [16]. The observed WERs are reported as 14.18%, 12.1% and 14.65% respectively. Considering the aforementioned studies the performance of Turkish ASR systems can be improved by applying sub-word (morpheme) based language models and replacing GMM acoustic modelling with DNN based models. All performance metrics are defined considering the sub-word and word recognition performances in the mentioned studies.

Sub-word (morpheme) based LM will be useful for improving the recognition rates of the agglutinative languages and prevents large vocabulary in the system. Root and suffixes are modelled separately in this model. In the proposed system, sub-word (morpheme) based language modelling is implemented. In addition to word, sub-word recognition, phoneme recognition is also supported. Each phoneme of Turkish language is also modelled as a sub-word. In this way, vocabulary coverage of the system has been enlarged.

Kaldi ASR toolbox, open speech recognition tools, is set up by Daniel Povey [17]. This tool can be used to implement DBN-HMM which builds language models and acoustic models. In the study, Kaldi handles the training and decoding for ASR of Turkish language.

The purpose of this work is suggesting a systematic approach for designing Deep Learning based ASR for Turkish language and improve the performance of the ASR by reducing effect on the PER and the WER. In this study, a standard corpus, "Turkish Microphone Speech v1.0" which was accepted by Linguistic Data Consortium (LDC) in 2005, is used to

perform the suggested system. The corpus is used in previous studies [11],[12],[13],[18]. In the similar studies [36], [37], [19], the databases which are provided by authors are used and have not been used in other studies. The databases are not standard and not confirmed by LDC. Using the standard dataset confirms the accuracy of applied method and provides an opportunity to compare the study with the previous papers which are using the same dataset. For example, the result of GMM based recognition of our study is very similar to the result which is obtained in [11].

While the previous studies Sub-word based language model, is used for training. Due to Turkish is a phonemic language, all phonemes in Turkish are also trained as a sub-word and entered to the lexicon. In case, sub-word or word recognition fails, phoneme based recognition finds out the phoneme components of unrecognized word or sub-word. The phoneme components are concatenated and unrecognized word can be formed.

In [15] and [19], the Deep Learning based ASR systems support only word-based ASR and the performance of the proposed systems are measured only by Word Error Rate (WER) metric. In our study, the performance of the applied method is also measured by Phoneme Error Rate (PER) metric. Our previous studies [20],[21], on phoneme based ASR for Turkish Language was published previously in 2016 and 2017. Since Turkish is a phoneme-based language, out of vocabulary (OOV) words can be found out by the help of recognized phoneme components. In the literature review, any deep learning-based phoneme recognizer was not encountered for Turkish language.

For training and testing, DBN in which the hidden layers are pre-trained by features is used. Considering similar studies, standard (traditional) sequential DNN and Time Delay Neural Network (TDNN) structures, have been applied in references [15],[19]. As recommended in reference [15], applying DBN will be preferred for ASR applications. The results of our study show that, the accuracy of DBN is

1.5% higher than previous DNN and TDNN based studies.

The performance of DBN-HMM based recognizer in the proposed study is compared with the performance of GMM-HMM recognizer in [11] which uses the same corpus. PER is used for comparing the results. WER performance of the methods, GMM-HMM and DBN-HMM, are also compared.

The rest of the paper is organized as follows: In Section 2, Morphological structure of Turkish language is explained, Section 3 explains the language models which are used in ASR applications as well as the preferred model. The mostly used speech recognition methods, Deep Learning, DNN algorithm and codes are explained in Section 4. Experiments and results are described in Section 5. Section 6 and 7 presents the discussions on results and conclusions.

2. TURKISH MORPHOLOGY

Turkish has an agglutinative morphology. Addition of several suffixes can derive many new words from a single stem. Prefix is not used in Turkish. The examples below show concatenated verbal and nominal inflections. The nominal inflection is less complicated [22] (See Table 1).

Table 1. Verbal and Nominal inflection examples

Nominal Inflection	araba-m-da-ki-ler (*ones in my car*).
Verbal Inflection	koş-a-ma-dı-lar (*They could not run*)

In agglutinative languages, it is possible to add morpheme one after the other. Each morpheme carries some morphological information like tense, status, agreement etc. This property of agglutinative languages causes having large vocabulary including many words with the same stem but different endings. Since there are so many words in the dictionary, there will be a large number of OOV words which are not modelled [23]. Larger lexicon size degrades the speed of decoder in word-based speech recognition [24]. Due to high OOV, speech recognition methods for English applied to Turkish give low recognition results.

Another important property of Turkish in terms of language modelling is free word order. Subject-Object-Verb word order in Turkish is a typical characteristic, but other orders are possible under certain discourse conditions. The perplexity of N-gram language models increase, since a sentence can be constructed with different order of words without changing the meaning [24].

As a consequence, base-units different from words, sub-words (morphemes) must be utilized in Turkish LVCSR to solve coverage problem and large amount of training data is needed in order to reliably train language model parameters.

2.1. Phonemes in Turkish

Turkish is a phonemic language such as Japanese or Finnish that means in the written language each phoneme is symbolized by a letter. In other words, the written text and pronunciation match exactly. However, some consonants and vowels have different voices, depending on where they are produced in the sound path. [25].

During their formation, the voices of which the oral part of the breath channel is not blocked by the teeth, tongue, or lips are called vowels. In Turkish, vowels can be classified in three groups by jaw angle, shape of lips and position of tongue [26] (see Fig. 1).

Basic speech sounds in which the breath is blocked by the teeth, tongue, or lips and which can be combined with a vowel to form a syllable are called consonants. Consonants can be classified into two groups, which are voiced and unvoiced, by the state of vocal cords. Voiced consonants are consonant sounds which are made by vibrating the vocal chords. The vocal cords do not vibrate during the formation of unvoiced consonants. Consonants can also be classified by output types and output locations [26] (see Fig. 2).
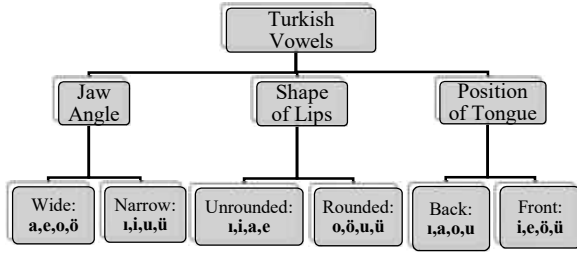
Figure 1. Turkish Vowel Classification
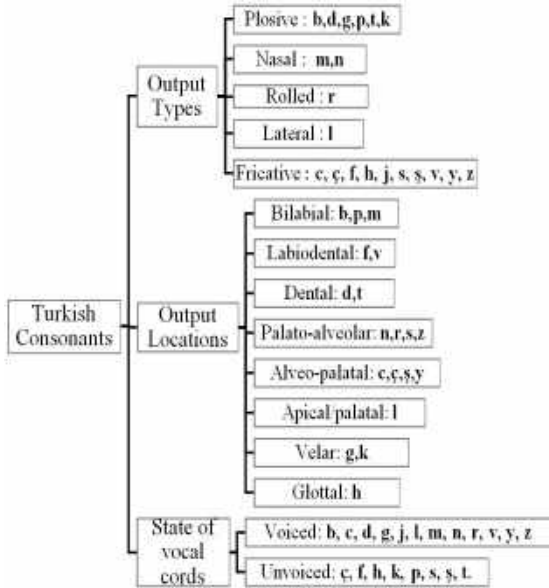
Figure 2. Turkish Consonant Classification

3. LANGUAGE MODEL (LM)

A Language Model (LM) obtains the probability of word existence based on text samples and develops the probabilistic models that can predict the following word in the chain of the given words. It contains a large list of words and their probability of occurrence. Larger models can predict sentences or paragraphs. Sub-word based model is preferred for the proposed system.

3.1. Word based LM

Word-based model is the most basic language modelling approach that uses words as recognition units. Words are the lexicon entries for speech recognition and language modelling and used as units for extracting probabilities from the training corpus [27] (see Fig. 3). The

word-based LMs are preferred when modelling analytic and isolated languages which have a low morpheme-per-word ratio. For example, English and Mandarin Chinese are analytic languages.
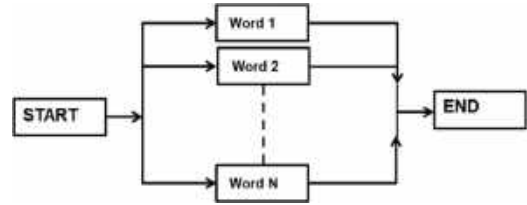
Figure 3. Word based LM

3.2. Sub-Word (Morpheme) based LM

Sub-word based LMs can improve the WER of agglutinative language ASR systems [3-8].

In Sub-word based LM, Morphemes are the basic units of the ASR system. (see Fig. 4). In Turkish language, A stem and following suffixes model the words according to the spelling rules and morphology. Bigram probabilities weight transitions between morphemes [7].

Phonetic rules are used for constructing the links between the stems and suffix lattices. The last phoneme of the stem and the last vowel determine the suffixes which may follow a particular stem. New stems can be added to the lattice automatically after this mapping is defined [29].
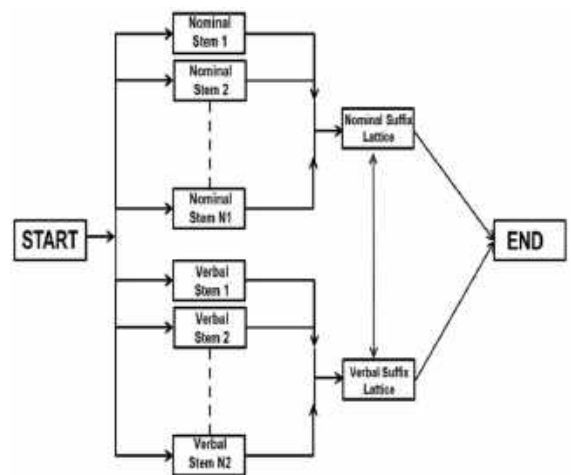
Figure 4. Sub-Word (Morpheme) based LM

4. METHODS OF SPEECH RECOGNITION

Statistical methods for ASR systems often use the HMM as an acoustic model. GMM is generally used when calculating the probability of the acoustic feature vectors for HMM. In last two decade, deep learning based speech recognition systems are developed for improving the recognition rate. Deep Learning approach has successfully replaced GMM for ASR. DNNs have been widely applied in acoustic model training and outperformed the statistical methods.

4.1. Statistical methods used in ASR

Mostly used and known ASR strategies are statistical based methods [30]. Blocks diagram of a statistical speech recognizer is shown in Fig. 5.
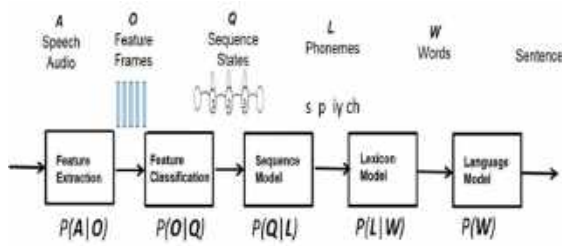
Figure 5. Statistical ASR system

The purpose of the ASR is converting speech audio to text. This procedure can be expressed statistically as below.

$O = (o_1, o_2, ..., o_n)$ (Sequence of speech vectors, o_i, the vector at time i) is the set of acoustic observations and $W = (w_1, w_2, ..., w_n)$ is the sequence of words. The maximum likelihood can be calculated as:

$$\widehat{W} = arg_w max P(W|O) = arg_w max \frac{P(W)P(O|W)}{P(O)} \quad (1)$$

Equation (1) uses Bayes rule and specifies the most likely word order. $P(O)$ represents the probability of the speech utterance. It is independent of the W sequence and can be neglected. Thus, (1) is simplified as shown below [30]:

$$\widehat{W} = arg_w max P(W)P(O|W) \quad (2)$$

Equation (2) contains two main factors:

The word sequence prior probability $P(W)$.

The probability of the acoustic data for the word sequence $P(O|W)$.

$P(W)$ value depends on the LM and $P(O|W)$ is calculated on the basis of the acoustic model. Both models can be generated separately, but they work together while recognizing a speech signal. The basis of the acoustical modelling is represented by HMMs.

4.1.1 Hidden Markov Models (HMM) for ASR

HMM method is known as state of art for speech recognition. GMM functions and state transition probabilities are used for classifying in HMM. HMMs are stochastic finite state machines and builds acoustic models and LMs in ASR applications [31]. They include a sequence of states related by transitions (see Fig. 6). The Markov process is named "hidden" because the observer cannot view the state sequence directly. A speech vector sequence which is generated from a Probability Density Function (PDF) for each state is observed.

As given in Fig. 6, A HMM assigns probabilities to a sequence of states. Consequently, the parameters below characterize an HMM: [30]

- a set of states $S = (s_1, s_2, ..., s_N)$ State at time $t : q_t$
- States transition probabilities: $A = (a_{11}, a_{12}, ..., a_{NN})$, each a_{ij} represents the transition probability from state i to state j;
- observation probabilities, each $B = b_i(o_t)$ defines the probability of an observation o_t which is built from the state i;
- the initial state distributions:

$$\pi = \{\pi_i = P[q_1 = s_i], i = 1, ..., N\}. \quad (3)$$
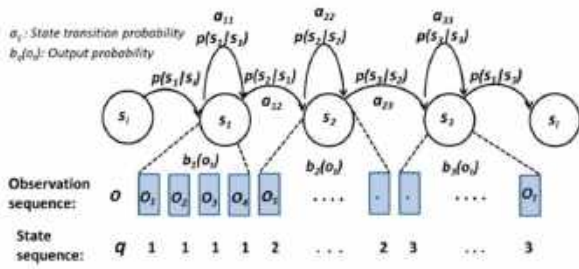
Figure 6. HMM representation

An HMM is indicated by $\lambda = (A, B, \Pi)$. HMMs suppose that the speech signal is stable in small time periods. HMMs are able to manage the speech signal variability well and they are good at speech modelling. In speech recognition models, speech is modeled by the states which are in order from left to right.

4.1.2 Gaussian Mixture Models (GMM)

Most widely used extension for standard HMMs is the mixture model of the state-output distributions. In the HMM Based Recognizer, Gauss distribution is used for modelling the state output distribution, assuming that the observed feature vectors are unimodal and symmetric [32].

In practice, speaker, gender and accent differences have tendency to create multiple modes in the data. Replacing the single Gaussian state-to-output distribution with a GMM which can model multi-modal and asymmetric data solves this problem (see Fig. 7).
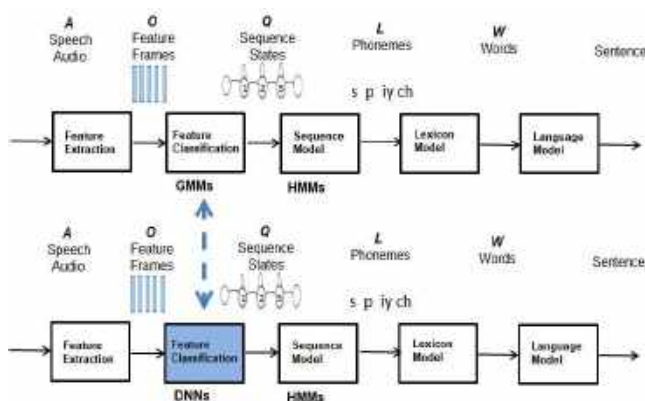
Figure 7. GMM-HMM and DNN-HMM based ASR systems

The GMM used for ASR, assumes that the spectral shape can be represented by a M component mixture model having parameters,

mixture components $p_m(x)$ and component weights w_m. The general form of a univariate mixture model is:

$$p(x) = \sum_{m=1}^{M} w_m p_m(x) \qquad (4)$$

where $p_m(x)$ are mixture components and w_m are weights. The mixture components in this case are Gaussian [33].

4.1.3 GMM and DNN in speech recognition:

The use of Gaussian mixtures in speech recognition applications has been replaced by the Deep Learning Approach. In ASR applications, DNNs work better than GMM based systems and they also left GMM's performance behind a large margin in some tasks [10] (see Fig. 7).

4.2. Deep Neural Networks (DNN)

Deep learning is an advanced architecture of ANN. More than one hidden layer is used for learning about the pattern and the features. Many complex signal patterns, for example, videos, images and speech, is successfully learned by means of many layers. These layers have nodes with nonlinear processing functions. Pre-training steps allow training large networks having millions of nodes.

DNNs with many hidden layers have been shown to perform better than GMMs in various speech recognition criteria. [34]. Hidden layers are used to detect patterns in data. More layers provide working on more complex data.

With nonlinear activation functions, DNN is able to model an arbitrary nonlinear function (projection from inputs to outputs). In each hidden unit, j, total input from the layer below, x_j, is mapped to the scalar state, y_j and sent to the layer above.

$$x_j = b_j + \sum_i y_i w_{ij} \quad (5) \qquad y_j = logistic(x_j) = \frac{1}{1 + e^{-x_j}} \quad (6)$$

b_j is the unit j bias, i is an index of units in the layer below, and w_{ij} is the weight for the link from i to j in the layer below.

For multiclass classification, the "softmax" linearity is used to get a class probability p_j

$$p_j = \frac{\exp(x_j)}{\sum_k \exp(x_k)} \qquad (7)$$

where k is an index all over classes [34].

In the training phase of DNN networks, the derivative of the mismatch between the actual and target outputs is back-propagated. The initial weights can be set to small random values. Generatively, pre-training the DNN as a Deep Boltzmann Machine (DBM) or as a DBN is a better way of initialization then the enrolment samples are used to fine-tune [35].

4.2.1 Generative Pre-training:

In generative pre-training, feature detectors are initially designed for modelling the structure in the input data instead of designing feature detectors for discriminating between classes [34]. At first, one layer of feature detectors is trained by feature data then the trained feature detectors act as data for training the next layer. The multiple layers of feature detectors are prepared well for a discriminative "fine-tuning" phase by this generative "pre-training". During backpropagation, weights found in pre-training are slightly adjusted by DNN.

Restricted Boltzman Machine (RBM), a special Boltzman Machine, consists of a hidden unit layer and a visible unit layer having no hidden - hidden or visible-visible connections. The connections between hidden units and visible units are symmetrical and undirected. The value of hidden and visible units is often stochastic binary units. (1 or 0 depending on probability) [36].

The RBM structure is illustrated in Fig. 9. The model can be defined as follows for a set of values of (v, h) in an energy function[37]:

$$E(\boldsymbol{v}, \boldsymbol{h}) = -\sum_{i \in \text{visible}} a_i v_i - \sum_{j \in \text{hidden}} b_j h_j - \sum_{i,j} v_i h_j w_{ij} \qquad (8)$$

Where h_j and v_i are the binary states of visible unit i and hidden unit j respectively; w_{ij} is the weight between hj and vi; b_j and a_i are the biases of h_j and v_i, respectively. The joint distribution over v and h is given as follows:

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Y} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \qquad (9)$$

where Y is a partition function given as:

$$Y = \Sigma_v \Sigma_h e^{\{-E(v,h;\theta)\}} \qquad (10)$$

Learning maximum likelihood using the contrastive divergence (CD) algorithm uses the training data for estimating the RBM model parameters θ [38].

In ASR applications, DBN-DNN produces posterior probability for each state of HMM according to the features. Speech units such as words, syllables, phones, phone states, sub-phones, etc. are modelled by HMM or states of HMM [39].

4.2.2 Deep Belief Networks (DBN):

DBN has made many important achievements in image and audio classification. The purpose of DBN is that learning the typical data features by training large amounts of unlabeled data in an unsupervised way. Stacked RBMs create a DBN [40].
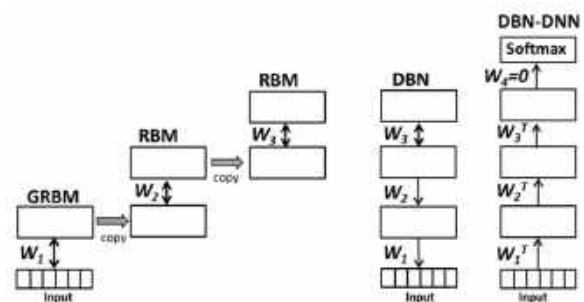
Figure 8. Deep Belief Networks [34]

The operations which are used for creating a DBN with three hidden layers and converting it to a pre-trained DBN-DNN are explained as follows (see Fig. 8). First a Gaussian-Bernoulli RBM (GRBM) is trained for modelling real-valued acoustic coefficient frames [34]. Training a RBM needs data which is the states of the binary hidden units of the GRBM. As many hidden layers as desired are created by repeating this process. The undirected connections of low-

level RBMs are replaced with top-down, directed links. Therefore The RBM stack is transformed into a single generative model and forms a DBN. A "softmax" output layer containing each possible state of each HMM is added. Therefore, a pre-trained DBN-DNN is built. The DBN-DNN is then discriminatively trained. In this way, the HMM states of the input window are estimated.

Features of speech, MFCCs, are modelled by linear variables with Gaussian noise by giving a GRBM; the RBM energy function is given as:

$$E(\boldsymbol{v}, \boldsymbol{h}) = \sum_{i \in \text{vis}} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j \in \text{hid}} b_j h_j - \sum_{i,j} \frac{v_i}{\sigma_i} h_j w_{ij} \quad (11)$$

Where b_j and a_i are the biases of h_j and v_i, respectively; and σ_i is the of the Gaussian noise standard deviation for visible unit i [34].

5. EXPERIMENTS AND ANALYSES

ASR application needs a Corpus which is a collection of texts of written or spoken language stored in a computer database. Corpora are the plural of corpus. Written texts in corpora are drawn from newspapers, books or magazines. Spoken corpora contain transcripts of spoken language.

The corpus "Turkish Microphone Speech" used in the experimental study is composed of 40 vocalized sentences from 120 speakers and 38 phonemes corresponding to 29 letters in Turkish alphabet [11].

In the experiments, GMM-HMM and DBN-HMM based two ASR systems are trained and tested. PER and WER are obtained for speaker independent situations.

In the purposed study, open source Kaldi codes are used for ASR algorithms. Voice recordings belonging to 120 people are used in the analyses. The speech data of 100 persons were used for the training and the data of 20 persons were used for the test.

5.1. Audio Corpus

The audio corpus, "Turkish Microphone Speech v1.0", used in our study was collected from 193 speakers (104 male, 89 female) in the Department of Electrical and Electronics Engineering at Middle East Technical University (METU). A set of 40 sentences among 2462 sentences are selected randomly for each speaker. This is the method which is used to form TIMIT corpus. Each sentence is uttered once by the speakers.

The corpus named "Middle East Technical University Turkish Microphone Speech v.1.0" was accepted by Linguistic Data Consortium (LDC) in 2005 [11].

Turkish language has some non-Latin vowels such as Ü, Ö, İ and non-Latin consonants such as Ç, Ş, Ğ. These special letters are symbolized as follows (see Table 2) and it has been possible to add these special letters into Kaldi speech recognition toolbox.

Table 2. Special Turkish letters in Kaldi

Special Turkish Letter	Symbols in Kaldi
Ç	CH
Ğ	GH
İ	IY
Ö	OE
Ş	SH
Ü	UE

There have been studies using "Turkish Microphone Speech Corpus", on Turkish speech recognition. These are [11], [12], [13] and [18]. In first study, the best phone recognition rate is 70.8% (PER=29.2%). In second study, word recognition rate is 78.54% (WER= 21.46%). In third study, word recognition rate is 67.12% (WER= 32.88%). In the fourth study, Common Vector Approach is used for classification. Vowel phoneme recognition rate is 48.75% and consonant phoneme recognition rate is 53.02%.

5.2. Kaldi ASR Toolbox

The Kaldi ASR toolbox is an open speech recognition tools set up by Daniel Povey. Various instructions of HMM Toolkit (HTK) are

integrated and then DNN model is introduced. The system framework is shown in Fig. 9 [17].

Kaldi speech recognition toolkit performs training and decoding operations. A monophone system builds a triphone HMM for speech recognizer. The Triphone system uses the basis of alignments which are derived from the monophone system. Therefore, the alignments supplied by the GMM system is used for training the deep learning system [39]. Kaldi provides building acoustic models and LMs. Kaldi is written in C++ programming language and it is an open source speech recognition toolkit. The Kaldi toolkit includes several shell scripts and C++ executables. The codes are easy to understand, very modern and flexible. This is available on both Linux and Microsoft Windows operating systems [41].
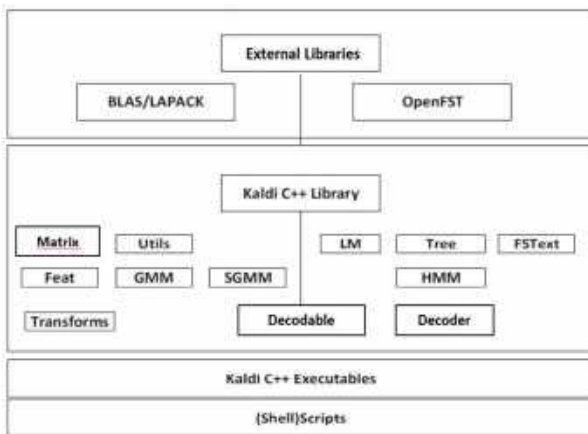
Figure 9. Kaldi system framework [17]

5.3. Data preparation (Pre-Processing)

Data preparation includes two main steps which are preparation of acoustic data and preparing language data.

Step1: Preparation of acoustic data: The acoustic data needs to be prepared before Kaldi ASR training. "Turkish Microphone Speech" corpus data which is used for training and testing are placed in the data folder (see Table 3). The data format must be converted to the desired form. Voice recordings belonging to 120 people are used in the analyses. The voice and sentence utterance of 100 people are placed in train folder. The data of 20 people for test is placed in test

folder. In training 4000 (100p x 40) sentences voiced by 100 persons are used. 3416 words are used in the training stage. After audio data is prepared, the corresponding language and acoustic models are built. Then three files, wav.scp, text and utt2spk are created manually. These files also can be created by using written scripts.

Table 3. Contents of "data" folder

Files	Explanation
wav.scp	The paths of speaker audio files
Text	Speech content of speakers
utt2spk	Tells which utterance belongs to particular speaker
spk2gender	Speaker gender
Corpus	Every single utterance transcription in ASR system (40 sentences per speaker)

The MFCC features are extracted and stored in feats.scp file after running scripts. The files which are called feats.scp, wav.scp, utt2spk, text and spk2utt are placed in the data/test and data/train folder. Corpus is placed in data/local folder (see Table 3)

The feature vectors are obtained over 20 ms frames. These frames are represented by 13 parameters consisting of 12 MFCC parameters and frame energy. 50% overlapping between frames is applied. Attributes obtained from 7 consecutive frames are added one after the other to obtain 91 dimensions (13x7) feature vector representing 80 ms frame.

Step2: Preparing language data: A folder which is named "dict" is created in data/local directory. The files related with lexicon and phonemes are created in the folder (see Table 4).

Table 4. Contents of "dict" folder

Files	Explanation
lexicon.txt	phone translations of words
nonsilence_phones.txt	Phones in the language (39 phones in Turkish)
silence_phones.txt	lists silence phones

"Lexicon.txt" file contains every word in our dictionary with its phone translations. Some word translations to phones are shown in Table 7. "Nonsilence_phones.txt" file contains all of 38

phonemes in Turkish. "Silence_phones.txt" file lists silence phones.

5.4. Model Training

In Kaldi, shell and Perl scripts are used for decoding and training. DNN model is trained based on the HMM model by Kaldi [11].

For Language Modelling SRILM Toolkit for Kaldi is used. Training and Decoding is made in three steps:

Step1: GMM-HMM model based on three phonemes (triphone) is trained using data in train folder and it is decoded by data in test folder.

Step 2: DBN-HMM model based on triphone is trained and decoded.

DBN is formed by a stack of six RBMs (see Fig. 10). In DBN-HMM, sigmoid activation function is used in hidden units having dimension of 1024 (see Table 5).

Table 5. DBN-DNN Parameters

Number of RBMs	Hidden Layer	Learning rates
6	Sigmoid dim:1024	Initial: 8×10^{-3} Final: 7.81×10^{-6}

Optimal initial learning rate for sigmoid in similar applications is 8×10^{-3}. Final learning rate is 7.81×10^{-6} when training ends (see Table 5).
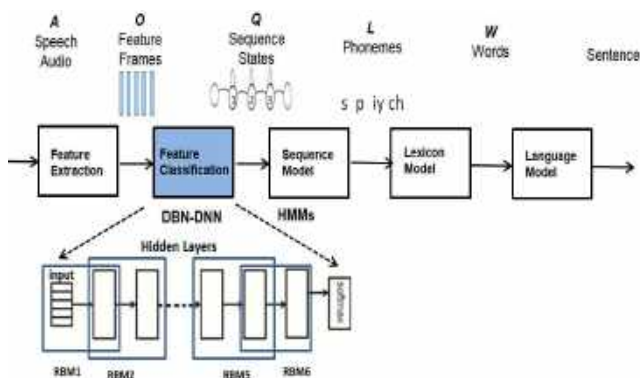
Figure 10. System Architecture

Cross-validation technique is applied for making DBN-HMM model robust against new data which are out of the training set. By this way

overfitting is prevented. The dataset is split into separate training and test subset. At every iteration, the test data subset is changed and the system is retrained. After every iteration, error function is measured. If the amount of error change is low enough, Early Stopping is applied and training process is finalized at this point.

After training and decoding process we got the following PER values. The PER of GMM-HMM, and DBN-HMM recognizers which are built on Kaldi is also given (see Table 6).

Table 6. Phoneme Error rates (PERs)

Model	PER
GMM-HMM	30.64%
DBN-HMM	24.80%

The GMM-HMM PER value which we obtain is consistent with the value of the previous study [11] in which the same corpus used.

Secondly, word recognition performance has been tested. The phoneme components of corpus vocabulary, stems and morphemes are given to the system as input. The phoneme components of some words and endings are on Table 7. The WER of the classifiers are given (see Table 8).

Table 7. Some stems, suffixes, words in corpus and their phoneme translation

Stem/Ending	Phoneme Components - Lexicon
SÖZLEŞME	S OE Z L EE SH M EE
NİN	NN IY NN
YAP	Y AA P
ILMASINI	I LL M AA S I NN I
KİM	K IY M
İSTİYOR	IY S T IY Y O RH
BUNDA	B U NN D AA
KÖŞE	K OE SH EE
GÖRÜN	G OE RR UE NN
NÜR	NN UE RH
ŞEKİL	SH EE K IY L
DE	D EE
KES	K E S
İLME	IY L M EE
MİŞ	M IY S

Table 8. Word Error rates (WERs)

Model	WER
GMM-HMM	17.21%
DBN-HMM	13.04%

6. DICUSSION OF RESULTS

In this study DBN-HMM based ASR system for Turkish language has been developed. The proposed system has been compared with GMM-HMM based ASR recognition method. In suggested system, GMM has been replaced with Deep Learning architecture. The recognition performance of the proposed method has been compared with the GMM-HMM method applied in [11]. "Turkish Microphone Speech Corpus v1.0" accepted by LDC is used for training and testing. Voice recordings belonging to 120 people are used in the analyses. The voice recording records of 100 persons were used for the training and 20 of them were used for the test.

In [11], GMM-HMM speech recognizer, trained by same corpus is included. The PER is measured as 29.3%. In the proposed study, the PER of Kaldi speech recognizer using GMM-HMM architecture is measured as 30.64%. Here we can see the results are similar and consistent.

When we train DBN-HMM speech recognizer with the same corpus, we get a degreased PER of 24.8% and improved recognition performance.

Considering word recognizer, The WER of GMM-HMM is measured as 17.21%. In the suggested Deep Learning based structure, the WER is 13.04% which is lower than in GMM-HMM.

In the similar study [15], both GMM-HMM and DNN-HMM systems are trained by the database which is recorded and prepared by authors. The WER of systems are measured as 17.40% and 14.65% respectively. The obtained results are similar to the presented study. The differences between two studies are the used speech corpus and the applied Deep Learning method. DBN is used instead of Feed Forward Neural Network in our study. As recommended in [15], applying

DBN will be preferred for ASR applications. The results of our study show that, using DBN has approximately 1.5% higher performance effect on the system. Unlike the study [15], the corpus "Turkish Microphone Speech v1.0" which is a standard dataset and accepted by LDC, has been used in the proposed study. Using an approved dataset which can be accessed by other researchers, improves accessibility, comparability and evaluability of published studies using the same database.

7. CONCLUSIONS

This paper proposed a DBN based Turkish phoneme and speech recognizer for improving the quality and performance of ASR for Turkish as an agglutinative language. The proposed system recognized words in the system vocabulary and phoneme components of OOV words. Sub-word (morpheme) LM which is commonly used for agglutinative languages is preferred. Each phoneme of Turkish language is modelled as a sub-word. Considering that Turkish language is a phoneme-based language, it is desired to reach an approximate result by concatenating the letter components of the words that cannot be recognized or are not in the dictionary. The performance of the suggested DBN based system is compared with conventional recognition method, GMM based HMM using the same dataset. In both architectures, HMM is used for sequence and language modelling. Application and comparison results show that the performance of the recognition has improved considering PER and WER criteria. Moreover, from these results it is observed that the Deep Learning models can produce deeper and more accurate feature probabilities and that the speech recognizer has a more discriminating ability. In future studies Long Short-Term Memory and Recurrent Neural Networks Language modelling with the same corpus can be performed and language-specific solutions for ASR in Turkish language can be developed.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] 'The most spoken languages worldwide (native speakers in millions) - Statistica, The Statistics portal', https:// www.statista.com / statistics / 266808 / the- most-spoken-languages-worldwide/, accessed 19 November 2018.

[2] Wang L, Tomg R, Leung C, Sivadas S, Ni C, Ma, B., 'Cloud-Based Automatic Speech recognition System for Southeast Asian Languages', International Conference on Orange Technologies (ICOT), IEEE, 2017, pp. 147-150.

[3] Varjokallio, M., Kurimo, M., Virpioja, S., 'Learning a Subword Vocabulary Based on Unigram Likelihood', IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2013, pp. 7-12.

[4] Varjokallio, M., Kurimo, 'A Word – Level Token – Passing Decoder for Subword N-gram LVCSR', IEEE Spoken Language Technology Workshop (SLT), 2014, pp. 495-500.

[5] Smit, P., Gangireddy, S., R., Enarvi, S., Virpioja, S., Kurimo, M., 'Character-Based Units for Unlimited Vocabulary Continuous Speech Recognition', IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2017, pp. 149-156.

[6] Mihajlik, P., Tüske, Z., Tárjan, B., Németh, B., Fegyó, T., 'Improved Recognition of Spontaneous Hungarian Speech-Morphological and Acoustic Modeling Techniques for a Less Resourced Task', IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 6, August 2010, pp. 1588-1600

[7] Arısoy E., Saraclar M., 'Language Modelling Approaches for Turkish Large Vocabulary Continuous Speech Recognition Based on Lattice Rescoring', 14th Signal Processing and Communications Applications, IEEE, 2006

[8] Aksungurlu T., Parlak S., Sak H, Saraçlar M., 'Comparison of Language Modelling Approaches for Turkish Broadcast News', 16th Signal Processing, Communication and Applications Conference, IEEE, 2008

[9] Arısoy, E., 'Devoloping an Automatic Transcription and Retrieval system for Spoken Lectures in Turkish', 25th Signal Processing and Communications Applications Conference (SIU), IEEE, 2017

[10] Dhankar, A., 'Study of deep Learning and CMU Sphinx in Automatic Speech Recognition', International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 2296-2301.

[11] Salor,O., Pellom, B.L., Çiloğlu, T., Demirekler, M., 'Turkish speech corpora and recognition tools developed by porting SONIC: (Towards multilingual speech recognition)', Computer Speech and Language, Elsevier, 2007, 21, pp. 580–593.

[12] Bayer, A., O., Çiloglu, T., Yondem, M., T., 'Investigation of Different Language Models for Turkish Speech Recognition', 14th Signal Processing and Communications Applications, IEEE, 2006

[13] Susman, D., Köprü, S., Yazıcı, A., 'Turkish Large Vocabulary Continuous Speech Recognition By Using Limited Audio Corpus', 20th Signal Processing and Communications Applications Conference (SIU), IEEE, 2012

[14] Arısoy E., Saraclar M., 'Compositional Neural Network Language Models for Agglutinative Languages', Interspeech 2016, San Francisco, USA, pp. 3494-3498

[15] Büyük, O., Kimanuka, U. A., 'Turkish Speech Recognition Based on Deep Neural Networks', Süleyman Demirel University Journal of Natural and Applied Sciences Volume 22, Special Issue, 2018, pp. 319-329

[16] Büyük, O., 'A new database for Turkish speech recognition on mobile devices and initial speech recognition results using the database', Pamukkale University Journal of Engineering Sciences Volume 24-2, 2018, pp. 180-184

[17] Ruan, W., Gan, Z., Liu, B., Guo Y., 'An Improved Tibetan Lhasa Speech Recognition Method Based on Deep Neural Network', 10th International Conference on Intelligent Computation Technology and Automation, IEEE, 2017, pp. 303-306

[18] Keser, S., Edizkan, R., 'Phoneme-Based Isolated Turkish Word Recognition With Subspace Classifier', 17th Signal Processing and Communications Applications Conference , IEEE, 2009.

[19] Asefisaray, B., Haznedaroğlu , A., Erden, M., Arslan, L., M., "Transfer Learning for Automatic Speech Recognition Systems", 26th Signal Processing and Communications Applications Conference (SIU), 2018

[20] Tombaloğlu, B., Erdem, H., "Development of a MFCC-SVM Based Turkish Speech Recognition system", 24th Signal Processing and Communication Application Conference (SIU), 2016

[21] Tombaloğlu, B., Erdem, H., " A SVM based speech to text converter for Turkish language", 25th Signal Processing and Communication Application Conference (SIU), 2017

[22] Arısoy E., Saraclar M., 'Lattice Extension and Vocabulary Adaptation for Turkish LVCSR', IEEE Transactıons on Audio, Speech and Language Processıng, vol. 17, no. 1, 2009

[23] Tunalı, V., 'A Speaker Dependent Large Vocabulary Isolated Word Speech Recognition System for Turkish', Msc. Thesis, Marmara University, 2005.

[24] Büyük O., 'Sub-Word Language Modelling for Turkish Speech Recognition', Msc. Thesis, Sabanci University, 2005.

[25] Salor, Ö., Pellom,B., Çiloğlu, T., Hacıoğlu, K. and Demirekler, M., 'On developing new text and audio corpora and speech recognition tools for the Turkish language, ICSLP-2002: Inter. Conf. On Spoken Language Processing, Denver, Colorado USA, pp. 349–352.

[26] Ergenç, İ., 'Konuşma Dili ve Türkçenin söyleyiş sözlüğü', Multilingual, Istanbul, 2002, p. 486.

[27] Arısoy E., Saraclar M., 'Turkish Dictation System for Broadcast News Applications', 13th European Signal Processing Conference, 2005.

[28] Arısoy E., Dutagacı H., Saraclar M., 'A unified language model for large vocabulary continuous speech recognition of Turkish', Signal Processing 86 , Elsevier, 2006, pp. 2844-2862.

[29] Dutagacı H, 'Statistical Language Models for Large Vocabulary Turkish Speech Recognition', Msc. Thesis, Boğaziçi University, 2002.

[30] Schiopu, D., 'Using Statistical Methods in a Speech Recognition System for Romanian Language', 12th IFAC Conference on Programmable Devices and Embedded Systems, 25-27 September 2013, Velke Karlovice, Czech Republic, pp. 99-103.

[31] Köklükaya, E, Coşkun, İ., "Endüktif Öğrenmeyi Kullanarak Konuşmayı Tanıma". Sakarya University Journal of Science 7, 2003, pp. 87-94.

[32] Gales, M., Young, S., 'The Application of Hidden Markov Models in Speech Recognition', Foundations and Trends in Signal Processing, Vol. 1, No. 3, 2007, pp. 195–304.

[33] Stuttle, M., N., 'A Gaussian Mixture Model Spectral Representation for Speech Recognition', Ph.D. Thesis, Cambridge University, 2003.

[34] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B., 'Deep Neural Networks for Acoustic Modelling in Speech Recognition', IEEE Signal Processing Magazine, Volume: 29 , Issue: 6 , Nov. 2012, pp. 82-97

[35] Alam, M. R., Bennamoun M., Togneri R., Sohel F., 'Deep Neural Networks for Mobile Person Recognition with Audio-Visual Signals', Mobile Biometrics, 2017, p. 97-129.

[36] Banumathi, A., C., Chandra, Dr. E., 'Deep Learning Architectures, Algorithms for Speech Recognition: An Overview', International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 1, January 2017, pp. 213-220.

[37] Siniscalchi, S., M., Svendsen, T., Lee, C., 'An artificial neural network approach to automatic speech processing', Neurocomputing, Elsevier, 2014, Vol. 140, pp. 326-338.

[38] Sharan, R. V., Moir, T. J., `An overview of applications and advancements in automatic sound recognition`, Neurocomputing, Elsevier, 2016, Vol. 200, pp. 22-34.

[39] Sustika, R., Yuliani, A. R., Zaenudin, E., Pardede, H. F., `On Comparison of Deep Learning Architectures for Distant Speech Recognition', 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE, 2017.

[40] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., `A survey of deep neural network architectures and their applications', Neurocomputing, Elsevier, 2017, Vol. 234, pp. 533-541.

[41] Yadava, G T., Jayanna, H S., `Creating Language and Acoustic Models using Kaldi to Build An Automatic Speech Recognition System for Kannada Language', 2nd IEEE International Conference On Recent Trends in Electronics Information and Communication Technology (RTEICT), May 19-20, 2017, India, IEEE, pp. 161-165

JOURNAL OF SCIENCE

SAKARYA UNIVERSITY

Sakarya University Journal of Science

Title: Structural, Microstructural and Electrochemical Characterization of Ni-YSZ
Anodes Fabricated from Pechini-Derived Composite Powders

Authors: Buse BİLBEY, Gamze EROL, Aligül BÜYÜKAKSOY

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pechini-Derived Composite Powders

Buse BİLBEY[1], Gamze EROL[2], Aligül BÜYÜKAKSOY[*3]

Abstract

The most widely used solid oxide fuel cell (SOFC) anodes, nickel-yttria stabilized zirconia (NiO-YSZ) composites, are generally fabricated by co-sintering of NiO and YSZ powders. In this study, to achieve a longer triple phase boundary length, these composites were fabricated from powders synthesized via an ethylene glycol-based Pechini method. Polymeric precursors of NiO and YSZ were prepared separately and then mixed, dried and calcined 600, 700 and 800°C for 4 hours, in air. NiO and YSZ crystals with average sizes of 26 and 7 nm, respectively were achieved upon calcination at 600 °C. With increasing heat treatment temperature, both the crystal and agglomerate sizes increased, which, in turn, an increased the anode polarization resistances. Electrochemical activities comparable to or higher than anode prepared by co-sintering of powder mixtures were achieved in the anodes prepared by sintering of the composite powders.

Keywords: Solid oxide fuel cells, polymeric precursor, NiO-YSZ, Pechini method

1. INTRODUCTION

Nickel-yttria stabilized zirconia (Ni-YSZ) composites are the most widely used anode materials in solid oxide fuel cells (SOFCs). The interest in Ni-YSZ anodes originates from the excellent electrocatalytic activity of Ni for hydrogen (or hydrocarbon) oxidation, its high electronic conductivity and the high ionic conductivity, as well as the chemical stability of YSZ [1,2,3].

---

[1]Gebze Technical University, Department of Materials Science and Engineering, 41400 Gebze, Kocaeli. ORCID: https://orcid.org/ 0000-0002-7404-866X. E-mail: bbilbey@gtu.edu.tr

[2]Gebze Technical University, Department of Materials Science and Engineering, 41400 Gebze, Kocaeli. ORCID: https://orcid.org/ 0000-0002-8619-3418. E-mail: gg.gamze.erol@gmail.com

[*]Corresponding Author: aligul@gtu.edu.tr

[3]Gebze Technical University, Department of Materials Science and Engineering, 41400 Gebze, Kocaeli
Gebze Technical University, Institute of Nanotechnology, Gebze, Kocaeli.
ORCID: https://orcid.org/0000-0003-2227-8938

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

The conventional method to fabricate Ni-YSZ anodes is based on the deposition of an ink consisting of NiO and YSZ powders, as well as solvents and binders onto a previously sintered, dense YSZ electrolyte substrate, followed by sintering at 1100-1300 °C and in-situ reduction into Ni-YSZ [4], as depicted schematically in Figure 1a. Since fuel oxidation takes place at the Ni-YSZ-gas interface, i.e., triple phase boundaries (TPBs), studies have focused on the development of fabrication techniques that would maximize the lengths of these electrocatalytic interfaces [5,6]. One alternative, which has

yielded promising results, has been the infiltration-based approach, which involves the introduction of the Ni phase into a porous YSZ scaffold via a liquid precursor through capillary action, followed by calcination [7, 8].

For example, Buyukaksoy et al obtained a polarization resistance of ca. 0.1 $\Omega.cm^2$ per electrode at 800 °C, under 3% H_2O – 97% H_2 gas mixture flow [8]. However, this method requires numerous infiltration/calcination steps, (e.g., 60 cycles in the case of 5), which necessitates significant optimization for implementation into mass production [8].
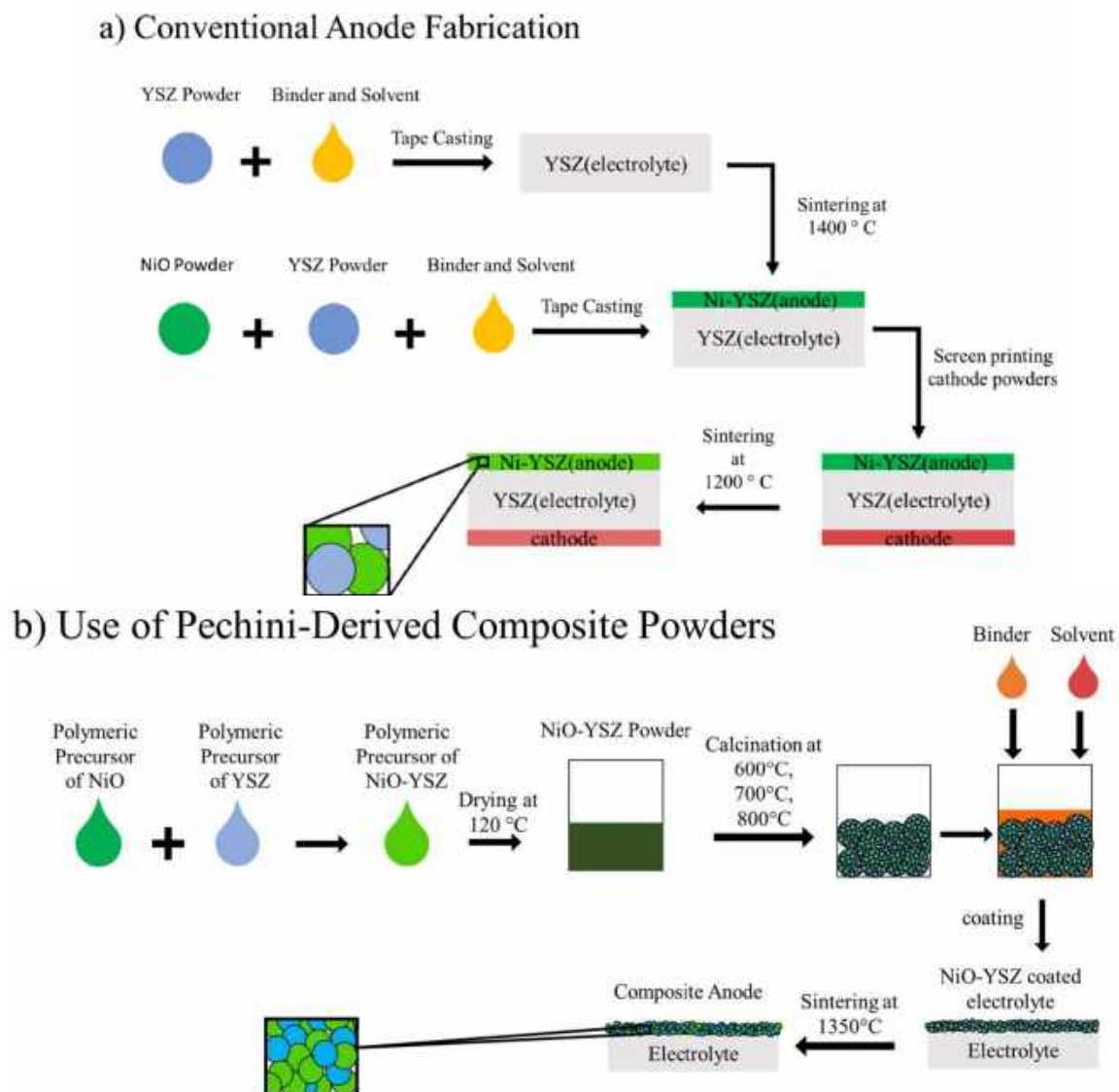

Figure 1. Schematic representation of Ni-YSZ anode fabrication via a) the conventional route involving the sintering of powder mixtures and b) the route involving the sintering of composite powders derived by a Pechini method.

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

Another approach to obtain Ni-YSZ anodes with long TPBs has been the synthesis of composite powders consisting of intimately mixed, nanoscale NiO and YSZ crystals for sintering [9]. For this purpose, Han et al calcined a mixture of commercial YSZ powder and a Ni-precursor solution [10], while Mohebbi et al tried the microwave-assisted combustion method [11]. Keech et al adopted a sol-gel based approach [12], which involved the mixing of Ni and YSZ precursor solutions. Grgicek et al [13] and Li et al [14] synthesized NiO-YSZ powders via co-precipitation, while Ringuede et al used combustion synthesis [15]. Razpotnik et al used the ethylene glycol-based Pechini technique and successfully synthesized NiO-YSZ powders [16]. Interestingly, among the above-mentioned literature, only two studies have investigated the electrochemical activity of the composite powder-derived anodes [12-15], one of which reported a ca. fifty-fold lower electrochemical performance than those fabricated by infiltration [12].

In the present work, we report on the synthesis of NiO-YSZ composite powders via an ethylene glycol-based Pechini method, due to its simplicity and thereby, scalability (Figure 1b). Long TPB lengths in the as-sintered anode, due to the molecular level mixing of the constituent cations allowed by this fabrication method is expected. In addition, for the first time in the literature, we correlate the calcination temperature used to obtain the NiO-YSZ powders from Pechni precursor, first to the microstructure, then to the electrochemical performance of the sintered Ni-YSZ anode.

2. EXPERIMENTAL

2.1. Powder synthesis and symmetrical half-cell fabrication

Polymeric precursors of NiO and YSZ were prepared separately. To obtain a polymeric NiO precursor $Ni(NO_3).6H_2O$ (Sigma-Aldrich, crystals or chunks) was first dissolved in distilled water. Then, ethylene glychol (Alfa Aesar, 99%) was added to the solution to achieve a 0.04:1 cation to ethylene glycol molar ratio. The resultant clear solution was stirred at 80°C until added water evaporated and polymerization took place. The YSZ polymeric precursor was prepared in a similar way. $Y(NO_3)_3.6H_2O$ (Alfa Aesar, 99.9%) and $ZrOCl_2.8H_2O$ (Sigma-Aldrich, ≥99.5%) were dissolved in distilled water to achieve an Y:Zr molar ratio of 0.16:0.84. Then, etyhlene glycol was added to the solution maintaining a cation to ethylene glycol molar ratio of 0.02 to 1 and the resultant solution was stirred on a hot plate at 80°C until the water evaporated. To obtain a polymeric NiO+YSZ polymeric precursor, the NiO and YSZ solution precursors were mixed in appropriate amounts that would yield NiO-YSZ powders with a 60:40 volummetric ratio. To obtain crystalline powders, the NiO/YSZ gels dried from the NiO+YSZ precursor solutions were calcined at 600, 700 and 800°C for 4 hours, in air.

The electrolyte substrate was fabricated by die-pressing 0,7 grams of commercial YSZ powder (TOSOH TZ-8Y) at 0,15 ton/m² pressure, using a die with a 1 cm diameter. For better compaction, the die-pressed pellets were subjected to cold isostatic pressure (CIP) at 200 MPa. The obtained pellets were sintered at 1400°C for 2 hours, in stagnant air.

To form the symmetrical half-cells, synthesized NiO-YSZ composite powders were mixed with α-Terpineol (Alfa-Aesar, ≥96%) and 2-Butoxyethanol (Sigma-Aldrich) to obtain inks. These inks were deposited onto the electrolyte surfaces and sintered at 1350°C for 2 hours, in air, to ensure bonding among the powders and at the powder/electrolyte interface.

2.2. Sample characterization

Particle size distributions of the synthesized powders were measured by laser diffraction method (Malvern Mastersizer 2000). Phase analyses of NiO, YSZ and NiO-YSZ powders were performed by x-ray diffraction (XRD, Rigaku D-Max 2200) using Cu Kα radiation source. Morphological and microstructural analysis of powders calcinated 600-800 for 4 hours carried out with scanning electron microscopy (SEM, Philips XL 30 SFEG).

Electrochemical impedance spectroscopy (EIS) measurements were performed at 500-750°C, in stagnant air at the $5×10^{-2}$ - 10^5 Hz frequency range, using ±10 mV excitation voltage (Gamry

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

Potentiostat, Reference 3000). NiO paste consisting of a mixture of Ni (II) oxide powder (Alfa-Aesar, 99%), α-Terpineol (Alfa-Aesar, ≥96%) and 2-Butoxyethanol (Sigma-Aldrich) was applied onto the anodes. Electrical contacs were made by attaching Nickel wires (Alfa-Aesar, 0.25 mm diameter, 99.98%) onto the electrodes using Ceramabond adhesive.

3. RESULTS AND DISCUSSION

3.1. Powder characteristic

To determine the phase evolution of NiO-YSZ composite powders, first, the crystal structure of gels dried from separate polymeric NiO and YSZ precursors after calcination at 600, 700 and 800°C for 4 hours was determined by x-ray diffraction (XRD). Peaks belonging to cubic NiO (PDF: 03-065-5745) as well as metallic Ni phase (also cubic in structure, PDF:00-004-0850) were observed, calcination upon even at 600°C (Figure 2a). With increasing calcination temperature metallic Ni transformed into cubic NiO (Figure 2a). Cubic YSZ phase (PDF: 00-030-1468) was obtained at all calcination temperatures (Figure 2b). The peak widths decreased with increasing calcination temperature, indicative of increasing crystal size (Figure 2b). The XRD patterns obtained from a mixture of polymeric NiO and YSZ precursors, showed that cubic NiO and YSZ phases crystallized separately and no solid solution formation took place, regardless of the calcination temperature (Figure 2c). The YSZ peaks typically had wider peaks in comparison to those that belonged to NiO (Figure 2b).

To determine the effect of mixing the polymeric NiO and YSZ precursors on the crystallization behavior of the respective phases, the average crystallite sizes of NiO and YSZ phases were calculated from XRD patterns using the Scherrer equation (1),

$$D_p = \frac{0.9\lambda}{B\cos\theta} \qquad (1)$$

where λ is x-ray wavelength, θ is the diffraction angle and B is the full width at half maximum intensity and Dp is the average crystallite size.

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...
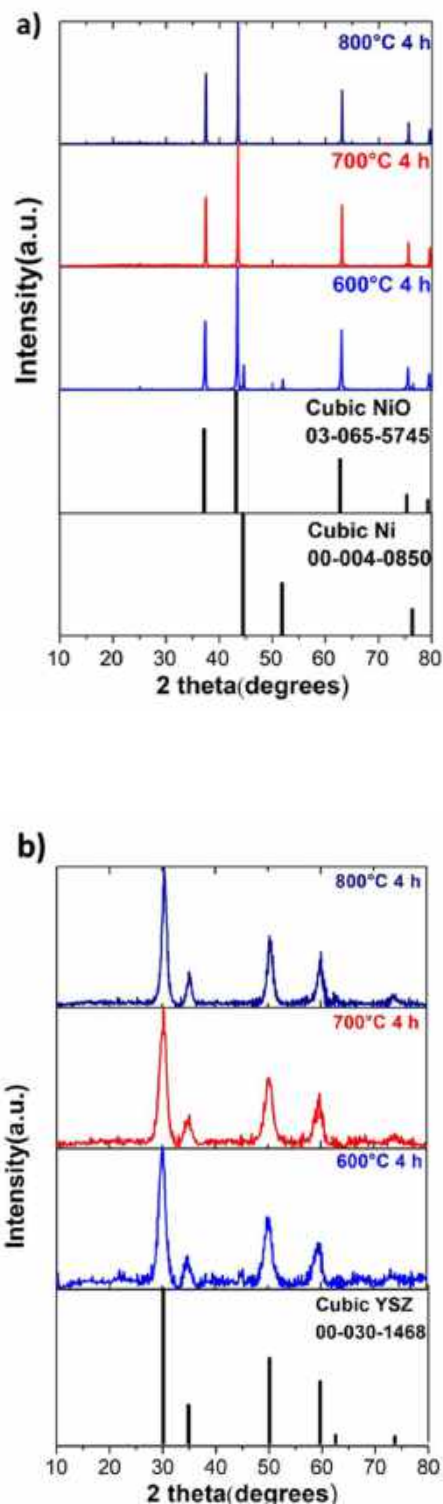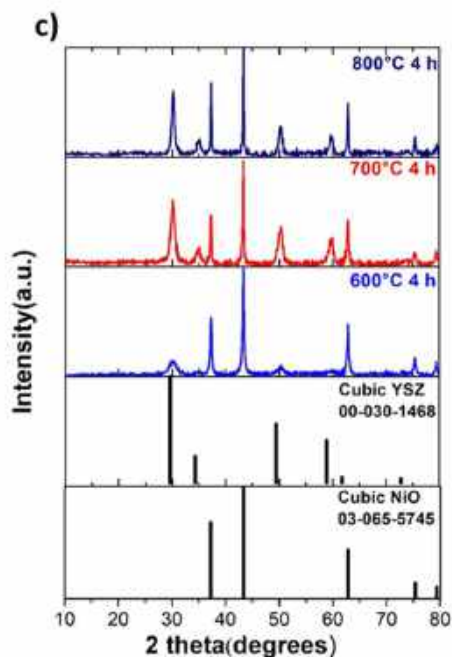
Figure 2. X-ray diffraction patterns obtained from gels dried from a) NiO, b) YSZ and c) NiO-YSZ polymeric precursors, calcined at 600, 700 and 800 °C.

Figure 3 shows the changes in the crystallite sizes of NiO and YSZ phases with increasing calcination temperature, in the cases of separate and mixed polymeric precursors. The crystallite size of NiO and YSZ phases increased with increasing calcination temperatures in both separate and mixed precursor solution cases (Figure 3). The crystallite size of the NiO phase was significantly smaller in the mixed precursor solution case in comparison to single precursor derived NiO (Figure 3). Evidently, the presence of Zr^{4+} and Y^{3+} ions in the solution hindered the clustering of Ni^{2+}, thereby resulting in smaller NiO crystals in comparison to those obtained from a separate NiO precursor. The inhibition of long-range order NiO crystal formation by the addition of Zr^{4+} and Y^{3+} ions to the solution suggests the formation of a nanocomposite powder. On the other hand, no significant effect of polymeric NiO precursor addition on the YSZ phase evolution was determined within the measurement accuracy of our equipment (Figure 3). Overall, NiO-YSZ composite powders with average NiO and YSZ crystal sizes in the ranges of 26-37 and 7 - 11 nm, respectively, were obtained using a mixture of polymeric NiO and YSZ precursors upon calcination at 600-800 °C.
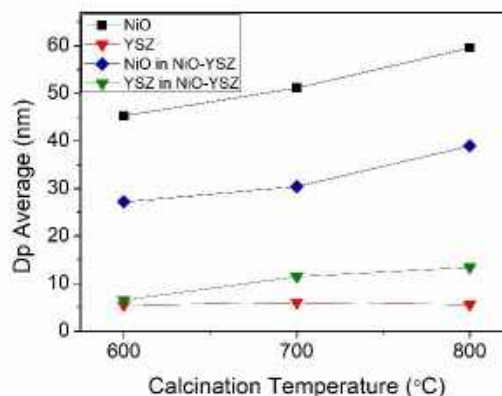
Figure 3. Average crystallite size of NiO and YSZ phases in NiO, YSZ and NiO-YSZ powders as a function of calcination temperature.

NiO-YSZ powders synthesized by microwave-assisted combustion were reported to have average crystal sizes ranging from 8 to 33 nm and from 14 to 32 nm for NiO and YSZ respectively, depending on the pH from of the solution [11]. On the other hand, Razpotnik et al. reported that average crystal sizes of NiO and YSZ varied between 42 to 48 nm and 6 to 15 nm, respectively, when a Pechini method, similar to the one adopted in the present study, was used [16]. However, the NiO and YSZ crystal sizes reported here for the Pechini-derived composite powders are somewhat smaller than those reported in Refs 11 and 16, suggesting a more intimate mixing and thus, potentially longer triple phase boundaries.

Another important characteristic of the synthesized powders that may be detrimental to the electrochemical activity is the particle size distribution. Laser diffraction measurements revealed that the amount of submicron-sized particles decreased with increasing calcination temperature (Figure 4a). The average particle size of NiO-YSZ powders was 897 ±31, 1073 ±16 and 1313 ±19 nm at 600, 700 and 800 °C, respectively (Figure 4b), which is significantly smaller than the particle size values reported for the Pechini-derived NiO-YSZ powders reported in the literature (average particle size 8.6-14.9 μm in REF-16). Particle sizes somewhat smaller than those reported in the present work have been achieved by Han et al (0.4 – 0.7 μm), who mixed a NiO precursor solution with fine, commercial YSZ powder [10] to achieve NiO-YSZ composite powders. Overall, the relatively small crystallite

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

and particle sizes were obtained in this present study via Pechini process. Thus, favourable microstructure upon sintering and thus, promising electrochemical activity can be expected.
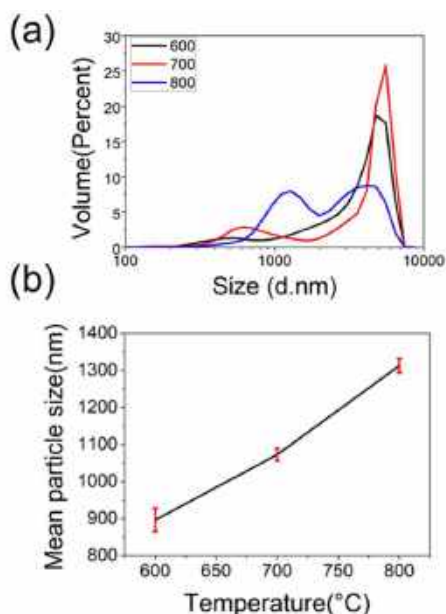
Figure 4. Effect of calcination temperature on a) particle size distribution and b) average particle size of NiO-YSZ powders.

3.2. Microstructural Analysis

Symmetrical half-cells were fabricated by depositing inks containing NiO-YSZ powders calcined at 600, 700 or 800 °C, followed by sintering at 1350 °C. The obtained NiO-YSZ composites were then reduced under 10% H_2-90%Ar flow, at 750 °C before microstructural analysis by scanning electron microscopy (SEM), to simulate electrochemical analysis conditions. Cross-sectional SEM images collected from the fracture surfaces of Ni-YSZ anode layers fabricated from NiO-YSZ powders calcined at 600-800 °C are provided in Figure 5. Ni-YSZ anode fabricated from powders calcined at 600 °C appears to be homogeneous in general, with only two larges (a few microns in size) particle chunks visible (shown with arrows in Figure 5a). With increasing powder calcination temperature, both the number and size of these large particle chunks increase significantly (Figures 5b and c), in agreement with the particle size analysis in Figure 4. Although the distribution of Ni and YSZ phases within these particle chunks may be

homogeneous and provide long triple phase boundaries (as seen in Figures 5d-f), they are hardly sinterable, thus, their extent of bonding to the electrolyte and to their surrounding particles are quite poor (Figures 5b and c). The deteriorated contact between the powders and the electrolyte substrate is also evident when the low-magnification top surface images of anodes using powders calcined at 600 and 700 °C (Figures 5g and h) to those using powders calcined at 800 °C (Figure 5i). Consequently, electrochemical activity is expected to decrease with increasing powder calcination temperature.

3.3. Electrochemical Activity

The assessment of the electrochemical activities of Ni-YSZ anodes prepared from Pechini-derived NiO-YSZ powders was performed by electrochemical impedance spectroscopy (EIS) analyses of symmetrical half-cells at 500-750 °C, under humidified 10%H_2 – 90%Ar flow. Figures 6a-c depict the Nyquist curves obtained at 750 °C from Ni-YSZ anodes fabricated from composite powders calcined at 600-800 °C, with corresponding Bode diagrams provided in the insets. At this operating temperature, the EIS responses consisted of two distinct semi-circles, indicating two separate electrochemical processes (Figures 6a-c). Therefore, an equivalent circuit consisting of one resistor (Rs) connected in series to two other resistors (R1 and R2), both of which is also connected in parallel to two constant phase elements (Q1 and Q2) was used to fit the EIS data (Figures 6a-c). Here, Rs describes the resistance to ionic conduction through the YSZ electrolyte, while R1/Q1 and R2/Q2 describes the high and low frequency processes, respectively. At lower operating temperatures, only one semi-circle was observed (data not shown), therefore fitting procedures were performed using an equivalent circuit consisting of only one R/Q element connected in series to Rs.

Table 1 provides the resistance and frequency values obtained from the equivalent fitting of the EIS data, in addition to the capacitance values calculated from these two parameters. Rs values, evidently, increased from 4.5 to 5.1 Ω.cm^2 with increasing powder calcination temperature (Table 1).

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...
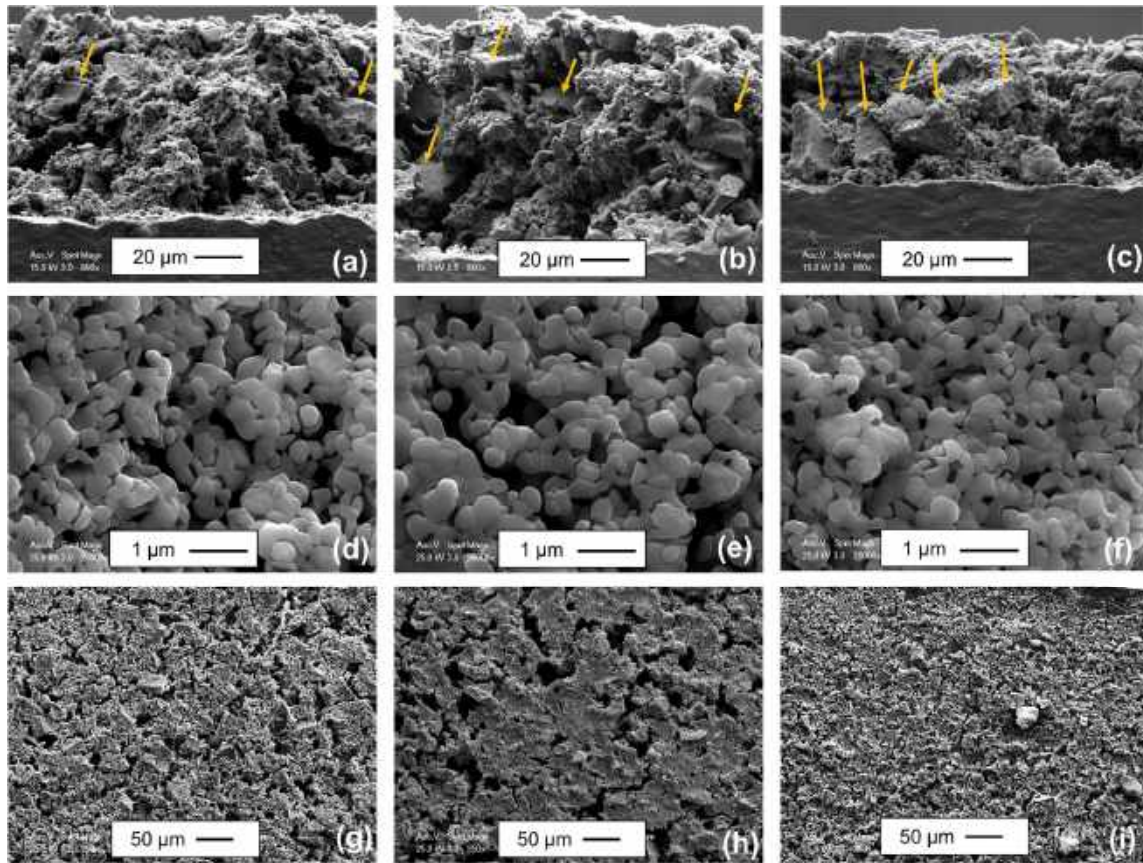
Figure 5. Scanning electron microscopy images collected from Ni-YSZ anodes fabricated from powders calcined at a, d, g) 600, b, e, g) 700 and c, f, i) 800 °C. The images in the first row are collected from the fractured cross-sections, while those in the second and third rows are collected from the top surfaces of the anodes.

Considering that i) the conductivity of YSZ electrolyte substrates are the same and ii) the thickness variation in the electrolyte substrates is below ca. 5%, the increase in ohmic resistance is considered to have originated from the poorer physical, hence electrical contact in Ni-YSZ anodes fabricated from larger particles. In a similar way, the sizes of both high and low frequency semi-circles also increased with increasing powder calcination temperature. R1 increased from 1.83 to 4.78 $\Omega.cm^2$ when the powder calcination temperature was increased from 600 to 800 °C. The capacitance values associated with the high frequency process were in the 10^{-6} F/cm^2 range, which have been ascribed to the double layer capacitance at the Ni/YSZ interface [17,18]. The low capacitance associated with the high-frequency arc suggests that the charge transfer at the triple phase boundary dominated the total polarization resistance. The

low frequency process, on the other hand had capacitance values in the 10^{-3} F/cm^2 range, which was ascribed to the hydrogen adsorption/desorption processes at the Ni surface [17]. The fact that R1 is significantly larger than R2 in all cases indicates that charge transfer at the triple phase boundary is the process limiting the hydrogen oxidation rate in all samples.

Rtotal values were determined by adding R1 and R2, which was then divided by two, to obtain the total polarization of one electrode (Ranode) in the symmetrical two-electrode cell (Table I). The lowest Ranode value obtained was 1.16 $\Omega.cm^2$, in the case of Ni-YSZ anode prepared from NiO-YSZ composite powders calcined at 600 °C. This value corresponds to a better electrode performance than reported in REF-17, in which Dasari et al obtained an Ranode value of 1.63 $\Omega.cm^2$ at 750 °C from Ni-YSZ anode fabricated by co-sintering a mixture of NiO and YSZ

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

powders. On the other hand, infiltration of Ni into porous YSZ scaffolds resulted in lower Ranode values than those reported here, i.e., 1.3 and 0.1 $\Omega.cm^2$ at 700 and 800 °C, respectively [8,19]. Despite the higher electrochemical activity of the infiltrated Ni-YSZ anodes, the necessity for numerous infiltration/drying cycles for their fabrication (as many as 60 cycles in the case of REF-8) renders them unsuited for scaling up.

Figure 6. Electrochemical impedance spectroscopy data obtained from anodes fabricated using powders calcined at a) 600, b) 700 and c) 800°C in the form of Nyquist and Bode (insets) diagrams. Note that the filled circles and the red lines show the data points and the fit curves, respectively.
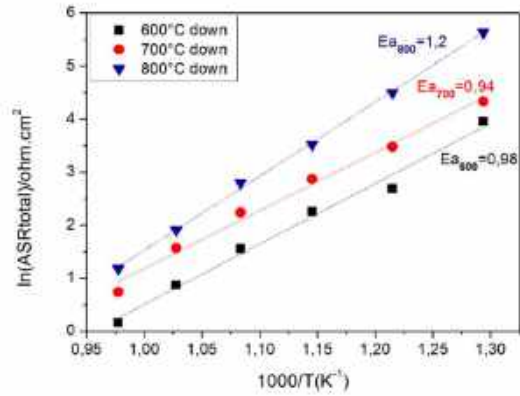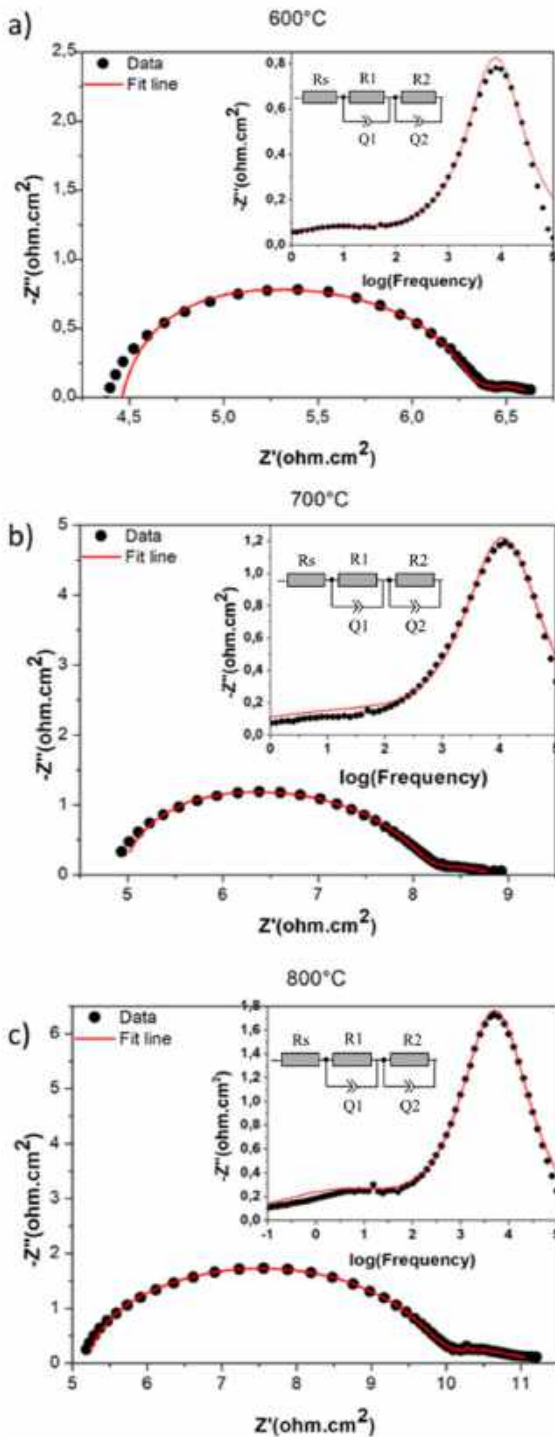
Figure 7. Temperature dependence of Ranode, measured in the cooling regime.

Figure 7 shows a comparison of the temperature dependencies of the Ni-YSZ anodes fabricated from composite NiO-YSZ powders synthesized by the Pechini method. Activation energies of anodes fabricated from NiO-YSZ composites calcined at 600, 700 and 800°C were determined as 0.98, 0.94 and 1.20 eV, respectively. The activation energies of the charge-transfer dominated polarization resistances of Ni-YSZ anodes have been reported to lie in the 1.00-1.30 eV range [8, 19], due to the combined effects of oxygen ion transport through the YSZ network (which has an activation energy of 0.9-1.1 eV [19]) and electron transfer at the triple phase boundary (which has activation energy of ca. 1.30 eV [19, 20]). The activation energies of Ranode obtained from anodes with NiO-YSZ powders calcined at 600 and 700 °C are close to that of oxygen ion transport through the YSZ network, while the activation energy of Ranode obtained from anodes with NiO-YSZ powders calcined at 800 °C is close to that of the electron transfer at the triple phase boundary. The effect of powder calcination temperature on the Ranode activation energy is likely related to the quality of the powder/electrolyte substrate contact. Specifically, coarse powders obtained by calcination at 800 °C resulted in a poor powder/electrolyte substrate

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

contact, resulting in the dominance of the resistance of the high activation energy-electron-transfer process.

4. SUMMARY

Symmetrical half-cell anodes were fabricated from NiO-YSZ composite powders, successfully synthesized via an ethylene glycol-based Pechini method. This method provided molecular level mixing of cations and hence, resulted in long triple phase boundaries. X-ray diffraction (XRD) analyses revealed that calcination of the gels dried from the polymeric NiO-YSZ precursors at 600 °C resulted in cubic Ni, NiO and YSZ phases, while the metallic Ni was absent when the powders were calcined at 700 or 800 °C. XRD measurements also revealed that mixing of the NiO and YSZ precursors yielded a NiO phase with smaller crystallites in comparison to the unmixed NiO precursor, upon calcination.

Sintering of anode layers deposited onto dense YSZ electrolytes at 1350 °C resulted in a homogeneous, mostly agglomerate-free microstructures when powders calcined at 600 °C were used. On the other hand, large agglomerates were detected in anodes fabricated from NiO-YSZ powders calcined at 700 and, especially 800 °C. Anode polarization resistances of 1.17, 2.09 and 3.10 Ω.cm² at 750 °C were achieved upon the electrochemical impedance spectroscopy measurements of anodes containing powders calcined at 600, 700 and 800 °C, respectively, which reflected the effect of powder calcination conditions on the performance of the sintered anode.

Research and Publication Ethics

The authors state that, in this work, they have complied with international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

Authors declared no conflict of interest.

Table 1. Summary of resistance, frequency and capacitance values extracted from the equivalent circuit fitting of impedance data collected at 750 °C.

Powder calcination temperature (°C)	Rs (Ω.cm²)	f_1 summit (Hz)	R_1 (Ω.cm²)	C_1 (F/ cm²)	f_2 summit (Hz)	R_2 (Ω.cm²)	C_2 (F/ cm²)	Rtotal (Ω.cm²)	Ranode (Ω.cm²)
600	4.50	17100	1.83	5.09×10^{-6}	153	0.50	2.08×10^{-3}	2.33	1.16
700	4.76	26800	3.10	1.92×10^{-6}	180	1.08	8.19×10^{-3}	4.18	2.09
800	5.06	15300	4.78	2.18×10^{-6}	25	1.42	4.56×10^{-3}	6.20	3.10

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

REFERENCES

[1]   A.H. Karim, K.-Y. Park, T.H. Lee, S.A. Muhammed Ali, S. Hossain, H.Q.H.H. Absah, J.-Y. Park, A.K. Azad, "Synthesis, structure and electrochemical performance of double perovskite oxide $Sr_2Fe_{1-x}Ti_xNbO_{6-\delta}$ as SOFC electrode," Journal of Alloys and Compounds, vol. 724, pp. 666-673, 2017.

[2]   S.C. Singhal, K. Kendall, "High temperature solid oxide fuel cells: fundamentals, design, and applications" Elsevier Science Ltd, ISBN 1-85617-387-9, Oxford, UK, 2003.

[3]   S.Tao, J. T. S. Irvine, "A redox-stable efficient anode for solid-oxide fuel cells," Nature Materials, 2(5), 320–323, 2003.

[4]   A. Sarikaya, V. Petrovsky, F. Dogan, "Effect of the anode microstructure on the enhanced performance of solid oxide fuel cells," International Journal Hydrogen Energy, vol. 37, pp. 11370-11377, 2012.

[5]   B. S. Prakash, S. S. Kumar, S.T. Aruna, "Properties and Development of Ni/YSZ as an anode material in solid oxide fuel cell: A review", Renewable and Sustainable Energy Reviews 36, 149-179, 2014.

[6]   J. Mizusaki, H. Tagawa, T. Saito, T. Yamamura, "Kinetic studies of the reaction at the nickel pattern electrode on YSZ in H_2–H_2O atmospheres," Solid State Ionics, vol. 70/71, pp. 52–58, 1994.

[7]   T. Klemenso, K. Thyden, M. Chen, H.-J. Wang, "Stability of Ni-yttria stabilized zirconia anodes based on Ni-impregnation," Journal of Power Sources, vol. 195, no 21, pp. 7295-7301, 2010.

[8]   A. Buyukaksoy, S.P. Kammampata, V. I. Birss, "Effect of porous YSZ scaffold microstructure on the long-term performance of infiltrated Ni-YSZ anodes," Journal of Power Sources, vol. 287, pp 349-358, 2015.

[9]   P. Tiwari, S. Basu, "Ni infiltrated YSZ anode stabilization by inducing strong metal support interaction between nickel and titania in solid oxide fuel cell under accelerated testing," International Journal of Hydrogen Energy, vol. 38, no. 22, 9494-9499, 2013.

[10]   K.R. Han, Y. Jeong, H. Lee, C.-S. Kim, "Fabrication of NiO/YSZ anode material for SOFC via mixed NiO precursors," Materials Letters, vol. 61, pp. 1242-1245, 2007.

[11]   H. Mohebbi, T. Ebadzadeh, F.A. Hesari, "Synthesis of nano-crystalline (Ni/NiO)–YSZ by microwave-assisted combustion synthesis method: The influence of pH of precursor solution, Journal of Power Sources, vol. 178, pp. 64-68, 2008.

[12]   P.G. Keech, D.E. Trifan, V.I. Birss, "Synthesis and Performance of Sol-Gel Prepared Ni-YSZ Cermet SOFC Anodes," Journal of The Electrochemical Society, vol. 152, no. 3, pp. A645-A651, 2005.

[13]   C.M. Grgicak, R.G. Green, W.-F. Du, J.B. Giorgi, "Synthesis and Characterization of NiO–YSZ Anode Materials: Precipitation, Calcination, and the Effects on Sintering," Journal of American Ceramic Society, vol. 88, no. 11, pp. 3081-3087, 2005.

[14]   S. Li, R. Guo, J. Li, Y. Chen, W. Liu, "Synthesis of $NiO\text{-}ZrO_2$ powders for solid oxide fuel cells," Ceramics International, vol. 29, pp. 883-886, 2003.

[15]   A. Ringuede, D. Bronine, J.R. Frade, "Assessment of Ni/YSZ anodes prepared by combustion synthesis," Solid State Ionics, vol. 146, pp. 219-224, 2002.

[16]   T. Razpotnik, J. Macek, "Synthesis of nickel oxide/zirconia powders via a modified Pechini method," Journal of

BİLBEY et al.

Structural, Microstructural and Electrochemical Characterization of Ni-YSZ Anodes Fabricated from Pec...

European Ceramic Society, vol. 27, pp. 1405-1410, 2007.

[17] H.P. Dasari, S.-Y. Park, J. Kim, J.-H. Lee, B.-K. Kim, H.-J. Je, H.-W. Lee, K. J. Yoon, "Electrochemical characterization of Ni-yttria stabilized zirconia electrode for hydrogen production in solid oxide electrolysis cells," Journal of Power Sources, vol. 240, pp. 721-728, 2013.

[18] S. Dierickx, J. Joos, A. Weber, E. Ivers-Tiffée, "Advanced impedance modelling of Ni/8YSZ cermet anodes," Electrochimica Acta, vol. 265, pp. 736-750, 2018.

[19] A. Buyukaksoy, V.I. Birss, "Comparison of the Electrochemistry of Ni Thin Film and Ni-YSZ Composite Anodes Fabricated by Polymeric Precursor Deposition," Journal of The Electrochemical Society, vol. 163, no. 13, pp. F1350-F1357, 2016.

[20] V. Sonn, A. Leonide, and E. Ivers-Tiffée, "Combined deconvolution and CNLS fitting approach applied on the impedance response of technical Ni8YSZ cermet electrodes," Journal of the Electrochemical Society, vol.155, no. 7, pp. B675-B679, 2008.

Title: Prioritizing the Factors for Customer-Oriented New Product Design in Industry 4.0

Authors: Melike ERDOĞAN, Özge Nalan BİLİŞİK

Prioritizing the Factors for Customer-Oriented New Product Design in Industry 4.0

Melike ERDOĞAN[*1], Özge Nalan BİLİŞİK[2]

Abstract

Customer-oriented new product design is one of the most important processes in the production environment to improve product quality and reliability and maximize their productivity. It is also necessary to consider customer expectations in this process for an effective design. In this paper, we present a methodology which is called Pythagorean Fuzzy Analytic Hierarchy Process (PF-AHP) for prioritizing criteria which should be considered for an efficient customer-oriented new product design in Industry 4.0 transition primarily. We use Pythagorean Fuzzy Sets (PFSs) to allow experts to make more flexible evaluations and handle the uncertain and vague information in a wider way. We determine five main and eighteen sub-criteria that affect the new product design process and after applying PF-AHP, we find that the most important main-criterion determined as "Production" and sub-criterion determined as "Return on Investment".

Keywords: Industry 4.0, Multi-Criteria Decision Making, Product Design, PFSs

---

[*]Corresponding Author: melikeerdogan@duzce.edu.tr
[1]Düzce University, Department of Industrial Engineering, Düzce, Turkey.
 ORCID: https://orcid.org/ 0000-0003-0329-8562
[2]Yıldız Technical University, Department of Industrial Engineering, Istanbul, Turkey.
 ORCID: https://orcid.org/ 0000-0002-7273-1270, E-mail: ozgenalan@gmail.com

1. INTRODUCTION

In today's competitive environment, it is a priority to meet customer demands when designing a new product. New business principles should be taken into account when considering customer requirements and companies need to design product development, purchasing, marketing, production, sales and customer service processes in line with new trends. Product lifecycle management is conducting the process from the moment a company creates the first idea for its products to the end of production. It aims to accelerate product development time, encourage innovation, reduce costs, shorten communication gaps, improve quality, and visualize product information between collaborating parts [1], [2]. This cycle begins with the development of a new product step and the success of a product life cycle comes from the success product design process. Therefore, the issue of new product design is of great importance for businesses. The new product development process' objective is to translate an opinion into a tangible physical asset. This process is structured around well-defined stages and each stage consists of lots of decision-making problems.

However, in today's developing technological conditions, the product life cycle should be realized by considering different dimensions of new product design. The latest technological developments that the industry is experiencing and expected to live in can be described as Industry 4.0. Industry 4.0 is an understanding that involves the processes of businesses and creates a global network. This approach is an innovative approach that requires the participation of internal customers, external customers and suppliers, enabling technological systems to work together and exchange information [3]. Transition to Industry 4.0 is supported by the digitalization of all areas for product generation and lifecycle management in almost all industrial sectors. To continue the competition, companies in most industrial sectors will need to digitalize their processes to gain great efficiency especially in the production environment [2]. Within the framework of the new conditions in which Industry 4.0 drives businesses, the criteria to be considered for designing a new product that meets these conditions should be revealed and the most important criteria should be determined in order to design a successful new product.

In a new product development problem, decision-makers must take into account the customers' expectations, the company's strategies, technological opportunities, and the company's capabilities. Decision-makers should be aware that any new product development process cannot be successful without taking these criteria into account [4]. Decision-making theory is one of the most fundamental issues used to find feasible solutions in decision-making situations where many factors should be taken into account. Considering the uncertainty and multi-factor inherent in new product development, applying multi-criteria decision making (MCDM) theory in the process will be useful [5].

Fuzzy logic has emerged that the mathematical methods of classical system theory are insufficient in many systems in the real world, especially in partially complex systems involving people hence uncertainty [6]. Fuzzy logic is useful for modeling uncertainty in linguistic expressions. To better address uncertainty in fuzzy sets, extended versions have been introduced on regular fuzzy sets. Among them, the PFSs are an extension of the intuitionistic fuzzy sets and give decision-makers independence in expressing their opinions in an uncertain environment [7]. PFSs are unsurpassed to other extensions, allowing membership functions to be defined more flexibly [8]. Especially in MCDM problems, membership functions need to be defined better and flexibly. PFSs facilitate representation on a broader membership and non-standard membership degrees, allowing experts to consider uncertainty more effectively than other extensions [8]–[10]. PF- AHP also appears as an approach that has been successfully applied in the literature and has proven its usability with many studies conducted in multi-criteria environment. AHP has frequently used MCDM method because of

the ability to deal with decision-making problems that are difficult or impossible to configure with traditional methods and to be able to model the problem in a hierarchical manner [11]. Besides, the ability to handle many qualitative and quantitative criteria and factors with the system approach at the same time and ease of use can be seen among the reasons for frequent reference to AHP in the literature [12]. Based on all these, we applied AHP, which is the most commonly used MCDM approach in decision making studies, in the environment of PFSs, while prioritizing the factors that should be addressed in customer-oriented new product design in Industry 4.0 for the first time unlike the other studies in the literature. At this point, our study fills the gap and contributes to the literature in terms of the combination of the proposed method and subject. In this paper, we aim to propose a methodology to prioritize the criteria which are important for the customer-oriented new product design in Industry 4.0 transition processes. In section 2, we reviewed the relevant literature to reveal the MCDM applications in new product development processes and newly released PF-AHP applications. Section 3, we presented the details of the proposed fuzzy methodology. Section 4 was included the case study for the development of the new product in Industry 4.0 transition process. In the last section, the results and discussion were given.

2. LITERATURE REVIEW

There are lots of papers in the literature about product design and design considerations. One of the most frequently approaches used in product design studies is MCDM method. In the following sub-sections, studies on product design in Industry 4.0, studies using MCDM in product design and recent studies on the proposed fuzzy extension with applied MCDM method are reviewed and summarized.

2.1. Product design in Industry 4.0

The subject of product design in Industry 4.0 applications has been addressed in some studies. Some of them can be summarized as follows.

Mourtzis et.al. [13] presented an approach to apply advanced visualization methods using Augmented Reality in product design that will be used from engineering students for visualizing product design and improving it. Wagner et.al. [14] tried to determine a consistent structure for the holistic use of digital twins in the whole process of product development. Zakoldaev et.al. [15] considered the duties of organization project activities in a digital factory and they described the components of a digital enterprise at the physical level of technological equipment and the cyber level of cloud services. Shinohara et.al. [16] aimed to propose a set of critical success factors for digital manufacturing implementation. They collected critical success factors based on academic literature review and consulting reports. They handled a conceptual framework to assist organizations in developing a strategy for digital manufacturing implementation, and map for that purpose all the required resources and capabilities. Ang et.al. [17] highlighted smart design and with smart operation as a way forward in an Industry 4.0, ranging from design to smarter ships and smart operation for energy efficiency. Albers et.al. [18] focused on identifying Industry 4.0 potentials on the product taking into account its interdependencies on the system of production. They developed an approach for assessing the impact and risk of changes on the process side or product side. Ahmed et al. [19] [20] presented a system called smart virtual product development, which can increase the product development process for industrially manufactured products and assist in decision making by using clear information about official decision events and explained how product production can be improved by using the smart virtual product development system in Industry 4.0. Bilal Ahmed et al. [20] proposed the concept of the smart virtual product development system that improves the decision making process at different phases and activities such as product design, production and inspection planning and the system they offered has a key role in Industry 4.0. Lin [21] aimed to define a product design approach based on user experience from a human-oriented viewpoint. The smart manufacturing approach proposed to strengthen

Industry 4.0 in the circular economy of the glass recycling industry investigates product decision making information systems and data-driven innovation.

2.2. MCDM studies in product design

The outstanding papers using MCDM in product design are as follows. Asmae et al. [22] presented a MCDM study based on fuzzy ontology when choosing between alternatives in the final assessment of the product design concept. Favi [23] adopted a mathematical model approach based on MCDM theory in order to determine the optimum and feasible design options in the conceptual design stage. Fatchurrohman et al. [24] used a combined decision-making approach in conceptual design selection named CoNQA which includes Concurrent Network, Quality Function Deployment and AHP. Kumar and Tandon [25] used axiomatic design taking customer requirements and design parameters into consideration in customer-driven product design and many design alternatives had developed with the help of AHP in their study. Favi [26] et al. applied MCDM to determine the applicable design options and determine the best design option for the tool holder carousel of a machine tool. Joshi and Gupta [27] used an ARTODTO system model to achieve financial, environmental and physical objectives in assessing the impact of product design in rescue operations. Zeng et al. [28] presented a fuzzy decision model for evaluating fashion-oriented industrial products for new product design that meets some selective market requirements. Song et al. [29] proposed a fuzzy AHP method to analyze and sort design parameters in the design of upper extremity rehabilitation devices. Xingli and Huchang [30] applied an improved Quality Function Deployment model for the innovative product design problem. Wang [31] proposed an AHP and DEMATEL (The Decision Making Trial and Evaluation Laboratory) under fuzzy environment approach in marketing-oriented product development realizing the practice of quality function deployment. Buchert et al. [32] integrated Life Cycle Sustainability Assessment with discrete decision trees for sustainable

product design. The benefits and barriers of the developed approach are evaluated on a bicycle framework. Lian et al. [33] adopted a MCDM approach based on layered ordinal relationship analysis in choosing the optimum scheme for product design. Liu [34] applied fuzzy Quality Function Deployment with a fuzzy MCDM approach in product design development and selection process. İç and Yıldırım [35] revealed the factors affecting the quality characteristics of a washing machine and determined the levels for each factor and an experimental design model was used in which levels of the factors optimize quality characteristics. In the analysis, Taguchi method and Gray Relational Analysis, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and VIKOR methods were used together. Guini et al. [36] proposed a multi-criteria decision support approach to select the finest design concept with the best industrial performance measures in a manufacturing company using ROC (Rank Order Centroid) and PROMETHEE (The Preference Ranking Organization Method for Enrichment Evaluation) methods. Khan et al. [37] analyzed various product design alternatives from a risk perspective for environmentally cognizant product risk modeling and assessment. Wang and Chen [38] applied a fuzzy MCDM based Quality Function Deployment approach which integrates fuzzy Delphi, fuzzy DEMATEL with linear integer programming for cooperative product design and optimal selection of module mix.

2.3. Pythagorean Fuzzy Analytic Hierarchy Process (PF-AHP)

The following studies can be given as an example of recent studies conducted using PF-AHP. Büyüközkan and Göçer [39] proposed an approach which integrated PFSs, AHP and complex proportional assessment under to evaluate alternative digital supply chain partners and they conducted a case study in Turkey to validate their proposed approach. Yıldız et al. [40] applied PF-AHP and PF-TOPSIS methods to assess 25 different districts in Istanbul for ATM location selection for a public bank serving in Turkey. Kaya et al. [41] used the MCDM

approach with PFSs to take into account a variety of contradictory factors, from qualitative to quantitative, in the site selection of the WEEE recycling facility. For this purpose, they applied to PF-AHP and PF-TOPSIS methods to find the best location alternative. Otay and Jaller [42] proposed AHP method in a Pythagorean Fuzzy environment to better model uncertainties in wind farm location selection. Karasan et al. [43] used PF-AHP method to analyze risks in the autonomous drive system. Gül [44] applied an integrated methodology consisting of PF-AHP and PF-VIKOR for risk assessment in the field of occupational health and safety. Yücesan and Kahraman [45] tried to perform a risk assessment for hydroelectric power plants using PF-AHP method. Experts identified twenty hazards and their results that could occur in the operation of the hydroelectric power plant. As a result of the study, for the three most important hazards, preventive actions had been taken and the expected result was to contribute to the safety of hydropower plants and the prevention of financial losses.

New developments for digitalization are presented and the benefits of these systems are emphasized in the presented studies related to Industry 4.0 and the new product design above. However, in our paper, we aim to present a methodology based on PF-AHP which helps decision-makers to determine and prioritize the criteria that are important and should be considered for customer-oriented new product development in Industry 4.0 era primarily. This paper is the first to implement the MCDM approach under the Pythagorean fuzzy environment for considering product design factors. With this study, it has been shown that extended fuzzy sets can be successfully applied in product design studies and the feasibility of the method in product design decision environments where uncertainty and vagueness is present.

3. Proposed Methodology

This section of the paper includes the suggested methodology and steps of the adopted application are given in the following subsections.

3.1. Pythagorean fuzzy sets

PFSs were proposed by Yager [46] derived from intuitionistic fuzzy sets which were originally proposed by Atanassov [47].

Unlike the intuitionistic fuzzy sets, the sum of membership and non-membership degrees can exceed 1, but the sum of their squares cannot in PFSs [9], [10]as defined in Definition 1.

Definition 1: Let X be a fixed set. A PFS is shown as $\tilde{P}$ [9], [10]:

$$\tilde{P} \cong \{x, \mu_{\tilde{p}}(x), \upsilon_{\tilde{p}}(x); x \in X\}$$

where the function $\mu_{\tilde{p}}(x): X \mapsto [0,1]$ describes the degree of membership and $v_{\tilde{p}}(x): X \mapsto [0,1]$ defines the degree of non-membership of the element $x \in X$ to P respectively and for every $x \in X$, it holds:

$$0 \leq \mu_{\tilde{p}}(x)^2 + v_{\tilde{p}}(x)^2 \leq 1 \qquad (1)$$

The indeterminacy ratio is obtained as in the following:

$$\mu_{\tilde{p}}(x) = \sqrt{1 - \mu_{\tilde{p}}(x)^2 - v_{\tilde{p}}(x)^2} \qquad (2)$$

A number of basic operations for PF numbers are given as in Definition 2.

Definition 2: $\beta_1 = P\left(\mu_{\beta_1}, v_{\beta_1}\right)$ and $\beta_2 = P\left(\mu_{\beta_2}, v_{\beta_2}\right)$ are two PF numbers and $\lambda > 0$. Operations on these two PF numbers are shown in the following such as [48]–[50]:

$$\beta_1 \oplus \beta_2 = P\left(\sqrt{\mu_{\beta_1}^2 + \mu_{\beta_2}^2 - \mu_{\beta_1}^2\mu_{\beta_2}^2}, v\beta_1 v\beta_2\right) \qquad (3)$$

$$\beta_1 \otimes \beta_2 = P\left(\mu\beta_1\mu\beta_2, \sqrt{v_{\beta_1}^2 + v_{\beta_2}^2 - v_{\beta_1}^2 v_{\beta_2}^2}\right) \qquad (4)$$

$$\lambda\beta_1 = P\left(\sqrt{1-\left(1-\mu_{\beta_1}^2\right)^\lambda}\,,(v\beta_1)^\lambda\right), \lambda > 0$$
(5)

$$\beta_1^\lambda = P\left((\mu\beta_1)^\lambda, \sqrt{1-\left(1-v_{\beta_1}^2\right)^\lambda}\right), \lambda > 0 \quad (6)$$

$$\beta_1 \ominus \beta_2 = P\left(\sqrt{\frac{\mu_{\beta_1}^2-\mu_{\beta_2}^2}{1-\mu_{\beta_2}^2}}, \frac{v_{\beta_1}}{v_{\beta_2}}\right), if\ \mu_{\beta_1} \geq$$

$$\mu_{\beta_2}, v_{\beta_1} \leq min\left\{v_{\beta_2}, \frac{v_{\beta_2}.\pi_{\beta_1}}{\pi_{\beta_2}}\right\} \quad (7)$$

$$\frac{\beta_1}{\beta_2} = P\left(\frac{\mu_{\beta_1}}{\mu_{\beta_2}}, \sqrt{\frac{v_{\beta_1}^2-v_{\beta_2}^2}{1-v_{\beta_2}^2}}\right), if\ \mu_{\beta_1} \leq$$

$$min\left\{\mu_{\beta_2}, \frac{\mu_{\beta_2}.\pi_{\beta_1}}{\pi_{\beta_2}}\right\}, v_{\beta_1} \geq v_{\beta_2} \quad (8)$$

π_{β_1} and π_{β_2} denote the degree of indeterminacy.

3.2. Pythagorean Fuzzy Analytic Hierarchy Process (PF-AHP)

This section of the paper includes the proposed fuzzy methodology. The stages of the PF-AHP method are shown in the following steps [43], [50], [51]:

Step 1. Compromised pairwise comparison matrix $R = (r_{jt})_{mxn}$ is constructed according to experts' evaluations. The linguistic scale that is used for decision matrices is presented in Table 1 adopted from [9].

Table 1. Scale for the interval-valued PF-AHP evaluations [9]

Linguistic Term	IVP Fuzzy Sets
Strictly Low Significance - SLS	[0,0.15][0.8,0.95]
Very Low Significance - VLS	[0.1,0.25][0.7,0.85]
Low Significance - LS	[0.2,0.35][0.6,0.75]
Below Same Significance - BSS	[0.3,0.45][0.5,0.65]
Strictly Same Significance - SSS	[0.4,0.55][0.4,0.55]
Above Same Significance - ASS	[0.5,0.65][0.3,0.45]
High Significance -HS	[0.6,0.75][0.2,0.35]
Very High Significance -VHS	[0.7,0.85][0.1,0.25]
Strictly High Significance -SHS	[0.8,0.95][0,0.15]

Step 2. The differences matrix $D = (d_{ij})_{mxm}$ is found between lower and upper points of the membership and non-membership functions using Eqs. (9) and (10):

$$d_{ij_L} = \mu_{ij_L}^2 - v_{ij_U}^2$$
(9)

$$d_{ij_U} = \mu_{ij_U}^2 - v_{ij_L}^2$$
(10)

Step 3. The interval multiplicative matrix $S = (s_{ij})_{mxm}$ is calculated via Eqs. (11) and (12):

$$S_{ij_L} = \sqrt{1000^{d_{ij_L}}}$$
(11)

$$S_{ij_U} = \sqrt{1000^{d_{ij_U}}}$$
(12)

After simulations, 1000 is found as the best value of the base, which makes all the relationships between the elements of the matrix clearer without sacrificing consistent relationship.

Step 4. The indeterminacy value $H = (h_{ij})_{mxm}$ of the r_{jt} is calculated using Eq. (13):

$$h_{ij} = 1 - \left(\mu_{ij_U}^2 - \mu_{ij_L}^2\right) - \left(v_{ij_U}^2 - v_{ij_L}^2\right)$$
(13)

Step 5. The indeterminacy degrees are multiplied with $S = (s_{ij})_{mxm}$ matrix to calculate the matrix of unnormalized weights $T = (\tau_{ij})_{mxm}$ via Eq. (14):

$$\tau_{ij} = \left(\frac{s_{ij_L} + s_{ij_U}}{2}\right) h_{ij}$$
(14)

Step 6. The priority weights w_i are obtained via Eq. (15).

$$w_i = \frac{\sum_{j=1}^m w_{ij}}{\sum_{i=1}^m \sum_{j=1}^m w_{ij}}$$
(15)

4. Case study

In this paper, we tried to prioritize the criteria to be considered in customer-oriented new product design for Industry 4.0 transition processes. For this purpose, the factors that are effective in

designing customer-oriented products were determined by conducting literature research and interviewing with experts. As a result of the literature research and interviews, it was observed that the criteria will be collected in five groups as shown in Table 2. Thus, the decision problem had been composed of five main criteria and eighteen sub-criteria. Figure 1 shows the hierarchy of considered decision problem.

Table 2. Main and sub-criteria and references

| Main Criteria	Sub-criteria	Reference
C1: Economical	C11: Return on Investment	Gupta et.al. (2015) [52]
	C12: Product Innovation	Gupta et.al. (2015) [52]
	C13: Profitability	Gupta et.al. (2015) [52]
C2: Environmental	C21: Low waste	Choi et.al. (2007) [53]
	C22: Carbon footprint	Gupta et.al. (2015) [52]
	C23: Clean development mechanism	Gupta et.al. (2015) [52]
C3: Production	C31: Energy consumption	Chandrakumar et.al. (2017) [54]
	C32: Fewer consumable	Choi et.al. (2007) [53]
	C33: Production technique	Choi et.al. (2007) [53]
	C34: Maintenance	Chandrakumar et.al. (2017) [54]
C4: Social	C41: Housing and service infrastructure	Gupta et.al. (2015) [52]
	C42: Health and education	Gupta et.al. (2015) [52]
	C43: Job opportunity	Gupta et.al. (2015) [52]
	C44: Social capital	Gupta et.al. (2015) [52]
	C45: Legislation and enforcement	Gupta et.al. (2015) [52]
C5: Competitiveness	C51: Customer demand	Choi et.al. (2007) [53]
	C52: Supplier relation	Choi et.al. (2007) [53]
	C53: Competitor trend	Choi et.al. (2007) [53]
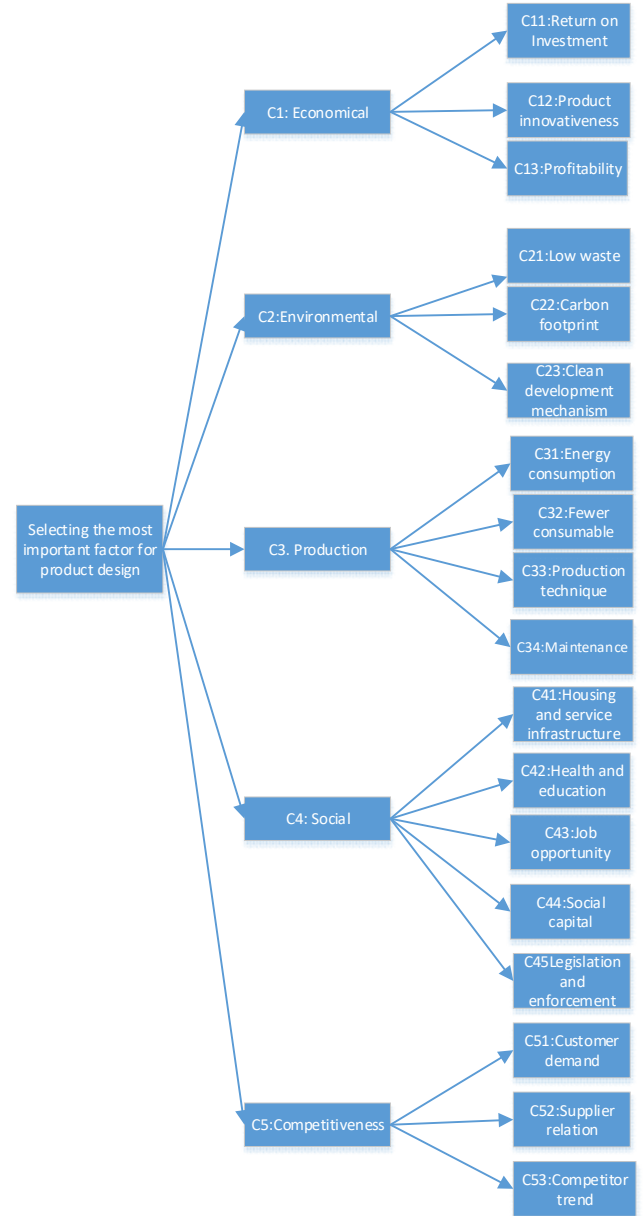
Figure 1. Hierarchy of criteria and sub-criteria for product design

Besides, the definitions of criteria are below:

- C1: Economical
  * *C11: Return on Investment:* performance measure to evaluate the efficiency of new product design industry 4.0 investments.
  * *C12: Product Innovation:* the creation or development amount of new products in industry 4.0

* *C13: Profitability:* the ability of a business to produce profit as a result of new product design in industry 4.0

- C2: Environmental
  * *C21: Low waste:* the achievement level of a low waste objective by the production of new product design in industry 4.0.
  * *C22: Carbon footprint:* the achievement level the low carbon footprint production objective by the production of new product design in industry 4.0.
  * *C23: Clean development mechanism:* to design the new product with clean development mechanism

- C3: Production
  * *C31: Energy consumption:* to design the new product with a low energy consumption feature.
  * *C32: Fewer consumable:* the low need for consumable products for the new product design in industry 4.0.
  * *C33: Production technique:* appropriate production technique for the new product design in industry 4.0.
  * *C34: Maintenance:* the easiness of determination of maintenance needs and strategies for the new product design in industry 4.0.

- C4: Social
  * *C41: Housing and service infrastructure:* easy access to housing and service infrastructure for the new product design in industry 4.0.
  * *C42: Health and education:* the effect of new product design in industry 4.0 to health and education of employees.
  * *C43: Job opportunity:* the emergence of new employment opportunities with new product design in industry 4.0.
  * *C44: Social capital:* the effect of the social relationships and structure to new product design.

* *C45: Legislation and enforcement:* the content of legislation and enforcement on new product design.

- C5: Competitiveness
  * *C51: Customer demand:* better analysis of the level of customer requirements for new products will provide a competitive advantage.
  * *C52: Supplier relation:* a better relationship with suppliers will help to handle competitive advantage.
  * *C53: Competitor trend:* the trends in the same business area will help to determine the right product features and handle competitive advantage.

After assigning the sub-criteria to the criterion groups, the scoring of the criteria was obtained. This scoring was done with the focus group work of the experts. 5 experts from university and industry are evaluated the criteria and sub-criteria. After this step, pairwise comparison matrices were formed based on these scores. The consistency of these matrices was checked and inconsistent evaluations were asked to be revised. After all matrices were made consistent, the most important factor in this decision-making process was found. Figure 2 shows the flow chart of the proposed methodology as follows.
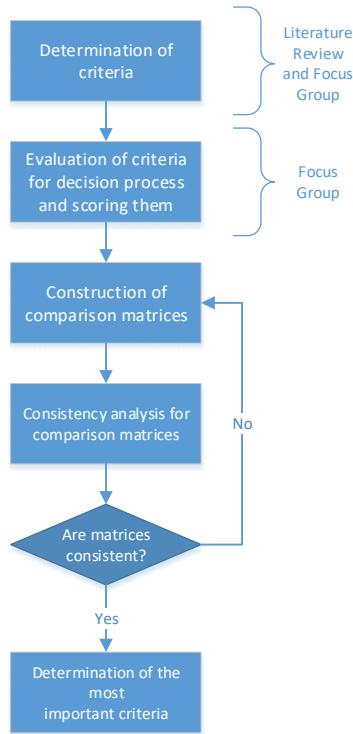
Figure 2. Flowchart of the proposed methodology

Firstly, pairwise comparisons of the main criteria were discussed and the linguistic evaluations were converted to interval-valued fuzzy numbers using the scale in Table 1. Table 3 contains the linguistic evaluations of the pairwise comparisons of the main criteria for the goal.

Table 3. Pairwise comparison of main criteria for goal

Goal	C1	C2	C3	C4	C5
C1	SSS	ASS	SSS	LS	HS
C2	BSS	SSS	BSS	BSS	SSS
C3	SSS	ASS	SSS	SSS	VHS
C4	HS	ASS	SSS	SSS	VHS
C5	LS	SSS	VLS	VLS	SSS

In the same way, pairwise comparisons of sub-criteria were obtained and matrices were formed. The pairwise comparison matrix for the social sub-criterion is shown in Table 4 below as an example.

Table 4. Pairwise comparison of sub-criteria for social criteria

C4 (Social)	C41	C42	C43	C44	C45
C41	SSS	HS	LS	LS	ASS
C42	LS	SSS	VLS	BSS	SSS
C43	HS	VHS	SSS	SSS	VHS
C44	HS	ASS	SSS	SSS	HS
C45	BSS	SSS	VLS	LS	SSS

After the pairwise comparisons were obtained for all matrices, the consistency ratios were calculated. Experts were asked to rearrange inconsistent assessments for ensuring that all matrices were consistent. Meanwhile, Table 5 was used to convert linguistic terms into crisp numbers to find consistency ratios, and then Saaty's consistency calculations were applied [55].

Table 5. Scale for the consistency ratio

Linguistic Term	Corresponded Crisp Number
Strictly Low Significance – SLS	0.11
Very Low Significance – VLS	0.14
Low Significance – LS	0.2
Below Same Significance – BSS	0.33
Strictly Same Significance – SSS	1
Above Same Significance – ASS	3
High Significance – HS	5
Very High Significance – VHS	7
Strictly High Significance – SHS	9

The consistency ratios for each matrix were calculated as in Table 6.

Table 6. Consistency degrees for pairwise comparison matrices

Based on	Consistency Ratio -CR
Goal	0,08107
Economical	0,03337
Environmental	0,03337
Production	0,05716
Social	0,09120
Competitiveness	0,02518

As shown in Table 5, the consistency ratio of all evaluations is less than 0.1. Therefore, the matrices are consistent and can be proceeded to the calculation of priority values. First, the main criteria weights were calculated. Then these weights were aggregated with the sub-criteria weights and the most important criteria can be determined.

The application of the adopted method started with the converting of the linguistic evaluations which are pairwise comparisons made for the goal by experts in Table 2 into the interval-valued PFSs in Table 1. Then, the differences matrix is calculated using Eq. (9) and (10) as in Table 7 for the main criteria:

Table 7. Differences matrix for goal

	C1		C2		C3		C4		C5	
	dij_L	dik_U	dij_L	dik_U	dij_L	dik_U	dij_L	dik_U	dij_L	dik_U
	-0,14	0,14	0,05	0,33	-0,14	0,14	-0,52	-0,24	0,24	0,52
	-0,33	-0,05	-0,14	0,14	-0,33	-0,05	-0,33	-0,05	-0,14	0,14
	-0,14	0,14	0,05	0,33	-0,14	0,14	-0,14	0,14	0,43	0,71
	0,24	0,52	0,05	0,33	-0,14	0,14	-0,14	0,14	0,43	0,71
	-0,52	-0,24	-0,14	0,14	-0,71	-0,43	-0,71	-0,43	-0,14	0,14

After the differences matrix is calculated, interval multiplicative matrix S is calculated via Eqs. (11) and (12) as in Table 8:

Table 8. Multiplicative matrix for goal

C1		C2		C3		C4		C5	
Sij_L	Sij_U	Sij_L	Sij_U	Sij_L	Sij_U	Sij_L	Sij_U	Sij_L	Sij_U
0,6	1,6	1,2	3,2	0,6	1,6	0,2	0,4	2,3	6,1
0,3	0,8	0,6	1,6	0,3	0,8	0,3	0,8	0,6	1,6
0,6	1,6	1,2	3,2	0,6	1,6	0,6	1,6	4,4	11,7
2,3	6,1	1,2	3,2	0,6	1,6	0,6	1,6	4,4	11,7
0,2	0,4	0,6	1,6	0,1	0,2	0,1	0,2	0,6	1,6

The indeterminacy value H of the r_{jt} is calculated using Eq. (13) and the indeterminacy degrees are multiplied with $S = (s_{ij})_{mxm}$ matrix to calculate the matrix of unnormalized weights $T = (\tau_{ij})_{mxm}$ via Eq. (14) as in Table 9:

Table 9. Unnormalized weights matrix for goal

	C1	C2	C3	C4	C5
C1	0,80	1,55	0,80	0,22	2,98
C2	0,42	0,80	0,42	0,42	0,80
C3	0,80	1,55	0,80	0,80	5,75
C4	2,98	1,55	0,80	0,80	5,75
C5	0,22	0,80	0,11	0,11	0,80

Finally, the weights of the main criteria is obtained as in Table 10 with normalizing the calculations in Table 9.

Table 10. Main criteria weights

Main Criteria	Weight
Economical	0,29
Environmental	0,20
Production	0,36
Social	0,06
Competitiveness	0,09

As can be seen from Table 10, the most important main criterion in customer-oriented product design for Industry 4.0 transition was found as "production". This criterion is followed by the "economical" and "environmental" criteria. The least important criterion in this process is the "social" criterion. After this step, the sub-criterion weights were calculated with the same steps and aggregated with the main criteria weights, and the priority value of each sub-criterion was calculated. Table 11 shows the criteria weights which aggregated with main criteria weights.

Table 11. Priority values of criteria for customer-oriented new product design

Criteria	Sub-criteria	Global Weights
Environmental	Low waste	0,031
	Carbon foot print	0,06
	Clean development mechanism	0,115
Competitiveness	Customer demand	0,061
	Supplier relation	0,019
	Competitor trend	0,005
	Return on Investment	0,163

Economical	Product innovativeness	0,044
	Profitability	0,084
Production	Energy consumption	0,111
	Production technique	0,053
	Fewer consumable	0,091
	Maintenance	0,101
Social	Housing and service infrastructure	0,01
	Health and education	0,004
	Job opportunity	0,028
	Social capital	0,016
	Legislation and enforcement	0,004

When Table 11 is examined, it is seen that the most important factor in customer-oriented new product design is "Return on Investment". This factor is followed by the "Clean development mechanism" factor under the environmental main criteria. Another factor which has a very close importance degree to this criterion is "Energy consumption". The criteria that should be given the least attention are the factors under the social main criterion. According to these results, it is confirmed that more emphasis should be given to environmental and economical factors-oriented criteria in customer-oriented new product design in Industry 4.0.

5. Comparative Analysis

In this section, a comparative analysis has been carried out to check the validity of our proposed approach. To this end, Buckley's fuzzy AHP method with triangular fuzzy numbers is utilized. The Buckley model is one of the most common approaches in studies applying to fuzzy AHP in many different areas [56]–[63]. The steps of the Buckley's fuzzy AHP method can be explained as follows [64], [65]:

Step 1: Pairwise comparison matrices are constructed. Each element ($\tilde{C}_{ij}$) of the pairwise comparison matrix (C) is a linguistic term and it is presenting the importance between two criteria. The pairwise comparison matrix is given below:

$$\tilde{C}_k = \begin{bmatrix} \tilde{1} & \tilde{C}_{12} & ... & \tilde{C}_{1n} \\ \tilde{C}_{21} & \tilde{1} & ... & \tilde{C}_{2n} \\ ... & ... & ... & ... \\ \tilde{C}_{n1} & \tilde{C}_{n2} & ... & \tilde{1} \end{bmatrix} \quad k= 1,2,...,K \qquad (16)$$

$\tilde{C}_k$ is a pairwise comparison matrix belonging to k_{th} expert. Table 12 presents the linguistic terms used for the evaluation procedure. The geometric mean is used to aggregate expert opinions.

Table 12. Linguistic Scale [64]

Linguistic Expression	Fuzzy Number
Equally important (Eq)	(1,1,3)
Weakly important (Wk)	(1,3,5)
Essentially important (Es)	(3,5,7)
Very strongly important (Vs)	(5,7,9)
Absolutely important (Ab)	(7,9,9)

Step 2: Weights are calculated. Firstly, the fuzzy weight matrix is calculated by Buckley's method.

$$\tilde{r}_i = \left[\tilde{C}_{i1} \otimes \tilde{C}_{i2} \otimes ... \otimes \tilde{C}_{in} \right]^{1/n}$$

$$(17)$$

$$\tilde{w}_i = \tilde{r}_i \otimes (\tilde{r}_1 + \tilde{r}_2 + ... + \tilde{r}_n)^{-1}$$

$$(18)$$

where $\tilde{r}_i$ is the geometric mean of fuzzy comparison value and $\tilde{w}_i$ indicated by triangular fuzzy numbers, $\tilde{w}_i$ (Lі,Mі,Uі) is fuzzy weight of ith criterion.

Step 3: Fuzzy relative weights are defuzzified. In defuzzification process, fuzzy numbers are converted into crisp values. For defuzzification, Liou and Wang's [66] total integral value method with an index of optimism ω∈[0,1] is

used. Let $\tilde{A}$ be a fuzzy number with left membership function $f_{\tilde{A}}^{L}$ and right membership function $f_{\tilde{A}}^{R}$. Then the total integral value is defined as follows:

$$E_{\omega}(\tilde{A}) = \omega E_{R}(\tilde{A}) + (1-\omega)E_{L}(\tilde{A}) \qquad (19)$$

where

$$E_{R}(\tilde{A}) = \int_{\alpha}^{\beta} x f_{\tilde{A}}^{R}(x)dx \qquad (20)$$

and

$$E_{L}(\tilde{A}) = \int_{\gamma}^{\delta} x f_{\tilde{A}}^{L}(x)dx \qquad (21)$$

where $-\infty \leq \alpha \leq \beta \leq \gamma \leq \delta \leq \infty$. For a triangular fuzzy number $\tilde{A} = (a,b,c)$, the total integral value is obtained by:

$$E_{\omega}(\tilde{A}) = \frac{1}{2}[\omega(a+b)+(1-\omega)(b+c)] \qquad (22)$$

Based on the above steps, Buckley's fuzzy AHP method was applied to the problem discussed in this study. Experts were asked to reconstruct pairwise comparison matrices according to the linguistic scale in Table 12. Table 13 presents reconsidered evaluations for goal as an example:

Table 13. Evaluations for the goal by experts

	C1	C2	C3	C4	C5
C1	1	1/Ab	Ab	Eq	Wk
C2	Ab	1	1/Ab	1/Ab	1/Vs
C3	1/Ab	Ab	1	Wk	Vs
C4	Eq	Ab	1/Wk	1	Ab
C5	1/Wk	Vs	1/Vs	1/Ab	1

After all evaluations were obtained and pairwise comparison matrices were constructed, calculations for finding criteria weights were

performed by applying Buckley's fuzzy AHP method. All the factors weights are calculated as in Table 14:

Table 14. Criteria Weights for Buckley's fuzzy AHP

Main Criteria	Sub-criteria	Global Weights
Environmental (0,22)	Low waste	0,035
	Carbon foot print	0,067
	Clean development mechanism	0,118
Competitiveness (0,07)	Customer demand	0,057
	Supplier relation	0,012
	Competitor trend	0,004
Economical (0,3)	Return on Investment	0,163
	Product innovativeness	0,048
	Profitability	0,088
Production (0,31)	Energy consumption	0,099
	Production technique	0,083
	Fewer consumable	0,071
	Maintenance	0,062
Social (0,09)	Housing and service infrastructure	0,030
	Health and education	0,019
	Job opportunity	0,013
	Social capital	0,027
	Legislation and enforcement	0,004

To understand the difference in rankings between adopted methodology and Buckley's fuzzy AHP results, we presented a bar chart as in Figure 3.
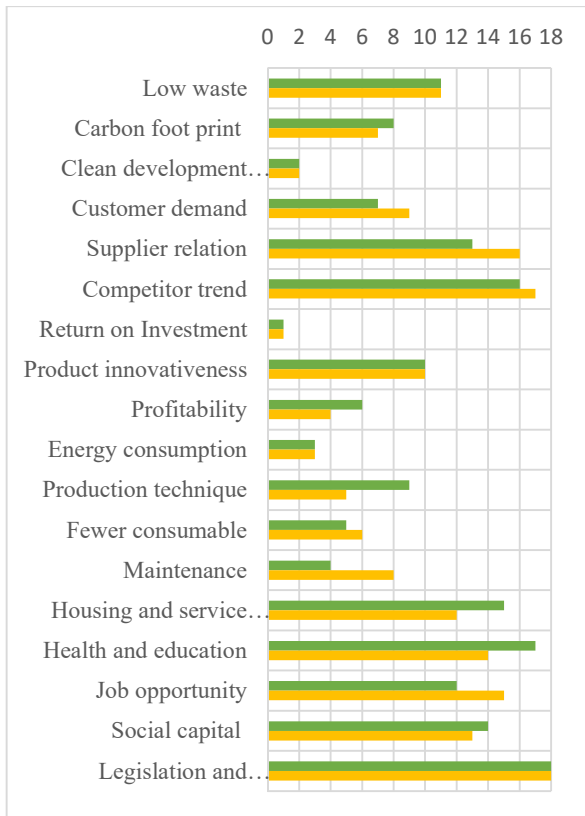
Figure 3. Ranks of criteria for PF-AHP (green line) and Buckley's fuzzy AHP (yellow line)

As can be seen from Figure 3, the criterion that should be taken in the first place in the new product design is the "Return on Investment" criterion, which is also placed on the top in the method we adopt. Likewise, the criteria placed in the second and third were in the same rankings in the comparative application. "Legislation and enforcement" and "Competitor trend" criteria, which were found in the last ranks, were found at the end of the list. In the meantime, the criteria in the ranks in intermediate levels appeared in changing rows. This comparative analysis has shown us that our results are valid; however, it has also shown that changes can be observed in the results because greater flexibility is given to decision-makers using PFs when compared to regular fuzzy sets.

6. Results and Conclusions

Making a new product design decision for a business, determining the design specifications and performing the design according to these specifications is critical for the survival of the company. It is also essential to design new products taking into account customer expectations in product development for all types of production businesses. Unless customer expectations are included in the design process, it is likely to be confused with design deficiencies. Therefore, after the need for new product design is occurred, companies need to determine the product design criteria. Within the framework of new conditions directed by Industry 4.0, the criteria that should be taken into consideration when designing a new product that meets customer expectations should be revealed and the most important criteria should be determined to design a new product successfully. For this reason, in this study, it is aimed to determine the importance ranking of the criteria to be considered for new product design in customer-oriented new product design for Industry 4.0 transition processes.

To achieve this goal, firstly, the factors that are effective in designing customer-oriented new products were determined by conducting literature research and interviewing with experts. In order to model the uncertainty in the process and to obtain the linguistic evaluations of decision-makers, a MCDM methodology was used within the fuzzy logic environment called PF-AHP. As a result of the paper, the most important main criterion in customer-oriented product design was found as "production". This criterion is followed by the "economical" and "environmental" criteria. The least important criterion in this process is the "social" criterion.

In addition, the main criteria were detailed by dividing into sub-criteria and it is observed that the most important factor in customer-oriented new product design is "Return on investment". This factor is followed by the "Clean development mechanism" factor under the environmental main criteria. The criteria that should be given the least attention are found as the factors under the social main criterion. According to these results, it is confirmed that more importance should be given to environmental and economical factors-oriented criteria in customer-oriented new product design in Industry 4.0. This study is expected to guide

companies to plan their actions according to these criteria by first determining the criteria that they should focus on.

In future studies, existing results can be compared by using different MCDM methods or by applying different fuzzy logic extensions. In addition, a prioritization study can be carried out among the products planned to be produced in the transition to Industry 4.0 using the proposed method.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] J. Stark, "Product Lifecycle Management," 2015, pp. 1–29.

[2] X. L. Liu, W. M. Wang, H. Guo, A. V. Barenji, Z. Li, and G. Q. Huang, "Industrial blockchain based framework for product lifecycle management in industry 4.0," Robot. Comput. Integr. Manuf., vol. 63, Jun. 2020, doi: 10.1016/j.rcim.2019.101897.

[3] V. Alcácer and V. Cruz-Machado, "Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems," Engineering Science and Technology, an International Journal, vol. 22, no. 3. Elsevier B.V., pp. 899–919, 01-Jun-2019, doi: 10.1016/j.jestch.2019.01.006.

[4] G. Büyüközkan and O. Feyzioğlu, "A fuzzy-logic-based decision-making approach for new product development," Int. J. Prod. Econ., vol. 90, no. 1, pp. 27–45, Jul. 2004, doi: 10.1016/S0925-5273(02)00330-4.

[5] X. T. Nguyen, V. D. Nguyen, V. H. Nguyen, and H. Garg, "Exponential similarity measures for Pythagorean fuzzy sets and their applications to pattern recognition and decision-making process," Complex Intell. Syst., vol. 5, no. 2, pp. 217–228, Jun. 2019, doi: 10.1007/s40747-019-0105-4.

[6] E. Kıyak and A. Kahvecioğlu, "Bulanık mantık ve uçuş kontrol problemine uygulanması," Havacılık ve Uzay Teknol. Derg., vol. 1, no. 2, pp. 63–72, 2003.

[7] M. Yucesan and M. Gul, "Hospital service quality evaluation: an integrated model based on Pythagorean fuzzy AHP and fuzzy TOPSIS," Soft Comput., vol. 24, no. 5, pp. 3237–3255, Mar. 2020, doi: 10.1007/s00500-019-04084-2.

[8] A. Karasan, E. Ilbahar, and C. Kahraman, "A novel pythagorean fuzzy AHP and its application to landfill site selection problem," Soft Comput., vol. 23, no. 21, pp. 10953–10968, Nov. 2019, doi: 10.1007/s00500-018-3649-0.

[9] A. Karasan, E. Ilbahar, S. Cebi, and C. Kahraman, "A new risk assessment approach: Safety and Critical Effect Analysis (SCEA) and its extension with Pythagorean fuzzy sets," Saf. Sci., vol. 108, pp. 173–187, Oct. 2018, doi: 10.1016/j.ssci.2018.04.031.

[10] E. Ilbahar, A. Karaşan, S. Cebi, and C. Kahraman, "A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP & fuzzy inference system," Saf. Sci., vol. 103, pp. 124–136, Mar. 2018, doi: 10.1016/j.ssci.2017.10.025.

[11] F. Ülengin, "Ulaşım Problemlerinde Analitik Hiyerarşi Yaklaşımı: İstanbul İçin Bir Uygulama," in TMMOB İstanbul 2. Kent İçi Ulaşım Kongresi, 1992, pp. 103–121.

[12] M. M. Jabri, "Personnel selection using INSIGHT - C: An application based on the analytic hierarchy process," J. Bus. Psychol., vol. 5, no. 2, pp. 281–285, Dec. 1990, doi: 10.1007/BF01014338.

[13] D. Mourtzis, V. Zogopoulos, and E. Vlachou, "Augmented Reality supported Product Design towards Industry 4.0: A Teaching Factory paradigm," in Procedia Manufacturing, 2018, vol. 23, pp. 207–212, doi: 10.1016/j.promfg.2018.04.018.

[14] R. Wagner, B. Schleich, B. Haefner, A. Kuhnle, S. Wartzack, and G. Lanza, "Challenges and Potentials of Digital Twins and Industry 4.0 in Product Design and Production for High Performance Products," Procedia CIRP, vol. 84, pp. 88–93, 2019, doi: 10.1016/j.procir.2019.04.219.

[15] D. A. Zakoldaev, A. V. Shukalov, I. O. Zharinov, and O. O. Zharinov, "Computer-aided design of technical documentation on the digital product models of Industry 4.0," in IOP Conference Series: Materials Science and Engineering, 2019, vol. 483, no. 1, doi: 10.1088/1757-899X/483/1/012069.

[16] S. A. Carolina, E. R. da Silva, E. P. de Lima, F. Deschamps, and S. E. G. da Costa, "Critical success factors for digital manufacturing implementation in the context of industry 4.0," Conf. Proc., pp. 199–204, 2017.

[17] J. Ang, C. Goh, A. Saldivar, and Y. Li, "Energy-Efficient Through-Life Smart Design, Manufacturing and Operation of Ships in an Industry 4.0 Environment," Energies, vol. 10, no. 5, p. 610, Apr. 2017, doi: 10.3390/en10050610.

[18] A. Albers, T. Stürmlinger, C. Mandel, J. Wang, M. B. de Frutos, and M. Behrendt, "Identification of potentials in the context of Design for Industry 4.0 and modelling of interdependencies between product and production processes," Procedia CIRP, vol. 84, pp. 100–105, 2019, doi: 10.1016/j.procir.2019.04.298.

[19] M. B. Ahmed, C. Sanin, and E. Szczerbicki, "Smart virtual product development (SVPD) to enhance product manufacturing in industry 4.0," in Procedia Computer Science, 2019, vol. 159, pp. 2232–2239, doi: 10.1016/j.procs.2019.09.398.

[20] M. Bilal Ahmed, S. Imran Shafiq, C. Sanin, and E. Szczerbicki, "Towards Experience-Based Smart Product Design for Industry 4.0," Cybern. Syst., vol. 50, no. 2, pp. 165–175, Feb. 2019, doi: 10.1080/01969722.2019.1565123.

[21] K. Y. Lin, "User experience-based product design for smart production to empower industry 4.0 in the glass recycling circular economy," Comput. Ind. Eng., vol. 125, pp. 729–738, Nov. 2018, doi: 10.1016/j.cie.2018.06.023.

[22] A. Asmae, B. A. Hussain, S. Souhail, and Z. El Moukhtar, "A fuzzy ontology-based support for multi-criteria decision-making in collaborative product development," in 2017 Intelligent Systems and Computer Vision, ISCV 2017, 2017, doi: 10.1109/ISACV.2017.8054953.

[23] C. Favi, M. Germani, and M. Mandolini, "Development of complex products and production strategies using a multi-objective conceptual design approach," Int. J. Adv. Manuf. Technol., vol. 95, no. 1–4, pp. 1281–1291, Mar. 2018, doi: 10.1007/s00170-017-1321-y.

[24] N. Fatchurrohman, S. Sulaiman, S. M. Sapuan, M. K. A. Ariffin, and B. T. H. T. Baharuddin, "A new concurrent engineering - multi criteria decision

making technique for conceptual design selection," in Applied Mechanics and Materials, 2012, vol. 225, pp. 293–298, doi: 10.4028/www.scientific.net/AMM.225.293.

[25] P. Kumar and P. Tandon, "A paradigm for customer-driven product design approach using extended axiomatic design," J. Intell. Manuf., vol. 30, no. 2, pp. 589–603, Feb. 2019, doi: 10.1007/s10845-016-1266-2.

[26] C. Favi, M. Germani, and M. Mandolini, "A Multi-objective Design Approach to Include Material, Manufacturing and Assembly Costs in the Early Design Phase," in Procedia CIRP, 2016, vol. 52, pp. 251–256, doi: 10.1016/j.procir.2016.07.043.

[27] A. D. Joshi and S. M. Gupta, "Evaluation of design alternatives of End-Of-Life products using internet of things," Int. J. Prod. Econ., vol. 208, pp. 281–293, Feb. 2019, doi: 10.1016/j.ijpe.2018.12.010.

[28] X. Zeng, Y. Zhu, L. Koehl, M. Camargo, C. Fonteix, and F. Delmotte, "A fuzzy multi-criteria evaluation method for designing fashion oriented industrial products," Soft Comput., vol. 14, no. 12, pp. 1277–1285, 2010, doi: 10.1007/s00500-009-0496-z.

[29] B. Song, Q. Peng, J. Zhang, and P. Gu, "A fuzzy number based hierarchy analytic method and application in improvement of rehabilitation devices," Comput. Aided. Des. Appl., vol. 16, no. 2, pp. 369–381, 2019, doi: 10.14733/cadaps.2019.369-381.

[30] X. Wu and H. Liao, "An approach to quality function deployment based on probabilistic linguistic term sets and ORESTE method for multi-expert multi-criteria decision making," Inf. Fusion, vol. 43, pp. 13–26, Sep. 2018, doi: 10.1016/j.inffus.2017.11.008.

[31] C. H. Wang, "An integrated fuzzy multi-criteria decision making approach for realizing the practice of quality function deployment," in IEEM2010 - IEEE International Conference on Industrial Engineering and Engineering Management, 2010, pp. 13–17, doi: 10.1109/IEEM.2010.5674597.

[32] T. Buchert, S. Neugebauer, S. Schenker, K. Lindow, and R. Stark, "Multi-criteria decision making as a tool for sustainable product development - Benefits and obstacles," in Procedia CIRP, 2015, vol. 26, pp. 70–75, doi: 10.1016/j.procir.2014.07.110.

[33] G. Lian, X. Bai, Y. Shen, D. Liu, and C. Zhu, "Multi-criteria decision making for mechanic product design schemes based on layered ordinal relation analysis," in Applied Mechanics and Materials, 2012, vol. 110–116, pp. 3990–3996, doi: 10.4028/www.scientific.net/AMM.110-116.3990.

[34] H. T. Liu, "Product design and selection using fuzzy QFD and fuzzy MCDM approaches," Appl. Math. Model., vol. 35, no. 1, pp. 482–496, Jan. 2011, doi: 10.1016/j.apm.2010.07.014.

[35] Y. T. İç and S. Yıldırım, "Çok Kriterli Karar Verme Yöntemleriyle Birlikte Taguchi Yöntemini Kullanarak Bir Ürünün Tasarımının Geliştirilmesi," Gazi Üniv. Müh. Mim. Fak. Der., vol. 2, no. 2, pp. 447–458, 2012.

[36] F. Guini, A. El Barkany, A. Jabri, and E. H. Irhirane, "An approach for the evaluation of a product's process planning during the design phase through a group multi-criteria decision-making," Int. J. Eng. Res. Africa, vol. 38, pp. 154–162, 2018, doi: 10.4028/www.scientific.net/JERA.38.154.

[37] R. A. Khan, A. Anand, and M. F. Wani, "A holistic framework for environment conscious based product risk modeling

and assessment using multi criteria decision making," J. Clean. Prod., vol. 174, pp. 954–965, Jan. 2018, doi: 10.1016/j.jclepro.2017.11.005.

[38] C. H. Wang and J. N. Chen, "Using quality function deployment for collaborative product design and optimal selection of module mix," Comput. Ind. Eng., vol. 63, no. 4, pp. 1030–1037, Dec. 2012, doi: 10.1016/j.cie.2012.06.014.

[39] G. Buyukozkan and F. Gocer, "A Novel Approach Integrating AHP and COPRAS Under Pythagorean Fuzzy Sets for Digital Supply Chain Partner Selection," IEEE Trans. Eng. Manag., 2019, doi: 10.1109/TEM.2019.2907673.

[40] A. Yildiz, E. Ayyildiz, A. T. Gumus, and C. Ozkan, "A Modified Balanced Scorecard Based Hybrid Pythagorean Fuzzy AHP-Topsis Methodology for ATM Site Selection Problem," Int. J. Inf. Technol. Decis. Mak., 2020, doi: 10.1142/S0219622020500017.

[41] A. Kaya, B. Çiçekalan, and F. Çebi, "Location selection for WEEE recycling plant by using Pythagorean fuzzy AHP," J. Intell. Fuzzy Syst., vol. 38, no. 1, pp. 1097–1106, 2020, doi: 10.3233/JIFS-179471.

[42] I. Otay and M. Jaller, "Multi-criteria and multi-expert wind power farm location selection using a pythagorean fuzzy analytic hierarchy process," in Advances in Intelligent Systems and Computing, 2020, vol. 1029, pp. 905–914, doi: 10.1007/978-3-030-23756-1_108.

[43] A. Karasan, İ. Kaya, M. Erdoğan, and A. Budak, Risk analysis of the autonomous vehicle driving systems by using pythagorean fuzzy AHP, vol. 1029. Springer International Publishing, 2020.

[44] M. Gul, "Application of Pythagorean fuzzy AHP and VIKOR methods in occupational health and safety risk

assessment: the case of a gun and rifle barrel external surface oxidation and colouring unit," Int. J. Occup. Saf. Ergon., 2018, doi: 10.1080/10803548.2018.1492251.

[45] M. Yucesan and G. Kahraman, "Risk evaluation and prevention in hydropower plant operations: A model based on Pythagorean fuzzy AHP," Energy Policy, vol. 126, pp. 343–351, Mar. 2019, doi: 10.1016/j.enpol.2018.11.039.

[46] R. R. Yager, "Pythagorean fuzzy subsets," in Proceedings of the 2013 Joint IFSA World Congress and NAFIPS Annual Meeting, IFSA/NAFIPS 2013, 2013, pp. 57–61, doi: 10.1109/IFSA-NAFIPS.2013.6608375.

[47] K. T. Atanassov, "Intuitionistic Fuzzy Sets," 1999, pp. 1–137.

[48] S. Zeng, J. Chen, and X. Li, "A hybrid method for pythagorean fuzzy multiple-criteria decision making," Int. J. Inf. Technol. Decis. Mak., vol. 15, no. 2, pp. 403–422, Mar. 2016, doi: 10.1142/S0219622016500012.

[49] X. Zhang and Z. Xu, "Extension of TOPSIS to Multiple Criteria Decision Making with Pythagorean Fuzzy Sets," Int. J. Intell. Syst., vol. 29, no. 12, pp. 1061–1078, Dec. 2014, doi: 10.1002/int.21676.

[50] M. Gul and M. F. Ak, "A comparative outline for quantifying risk ratings in occupational health and safety risk assessment," J. Clean. Prod., vol. 196, pp. 653–664, Sep. 2018, doi: 10.1016/j.jclepro.2018.06.106.

[51] B. Özkan, İ. Kaya, M. Erdoğan, and A. Karaşan, Evaluating blockchain risks by using a MCDM methodology based on pythagorean fuzzy sets, vol. 1029. Springer International Publishing, 2020.

[52] S. Gupta, G. S. Dangayach, and A. K.

Singh, "Key determinants of sustainable product design and manufacturing," in Procedia CIRP, 2015, vol. 26, pp. 99–102, doi: 10.1016/j.procir.2014.07.166.

[53] J. K. Choi, L. F. Nies, and K. Ramani, "A framework for the integration of environmental and business aspects toward sustainable product development," J. Eng. Des., vol. 19, no. 5, pp. 431–446, Oct. 2008, doi: 10.1080/09544820701749116.

[54] C. Chandrakumar, A. K. Kulatunga, and S. Mathavan, "A multi-criteria decision-making model to evaluate sustainable product designs based on the principles of design for sustainability and fuzzy analytic hierarchy process," in Smart Innovation, Systems and Technologies, 2017, vol. 68, pp. 347–354, doi: 10.1007/978-3-319-57078-5_34.

[55] A. Karasan, M. Erdogan, and E. Ilbahar, "Prioritization of production strategies of a manufacturing plant by using an integrated intuitionistic fuzzy AHP & TOPSIS approach," J. Enterp. Inf. Manag., vol. 31, no. 4, pp. 510–528, Jul. 2018, doi: 10.1108/JEIM-01-2018-0001.

[56] S. Boral, I. Howard, S. K. Chaturvedi, K. McKee, and V. N. A. Naikan, "An integrated approach for fuzzy failure modes and effects analysis using fuzzy AHP and fuzzy MAIRCA," Eng. Fail. Anal., vol. 108, p. 104195, Jan. 2020, doi: 10.1016/j.engfailanal.2019.104195.

[57] B. Das and S. C. Pal, "Combination of GIS and fuzzy-AHP for delineating groundwater recharge potential zones in the critical Goghat-II block of West Bengal, India," HydroResearch, vol. 2, pp. 21–30, Dec. 2019, doi: 10.1016/j.hydres.2019.10.001.

[58] S. Kaganski, J. Majak, and K. Karjust, "Fuzzy AHP as a tool for prioritization of key performance indicators," in Procedia CIRP, 2018, vol. 72, pp. 1227–1232, doi:

10.1016/j.procir.2018.03.097.

[59] P. Dehraj and A. Sharma, "An empirical assessment of autonomicity for autonomic query optimizers using fuzzy-AHP technique," Appl. Soft Comput. J., vol. 90, p. 106137, May 2020, doi: 10.1016/j.asoc.2020.106137.

[60] C. Acar, A. Beskese, and G. T. Temur, "Sustainability analysis of different hydrogen production options using hesitant fuzzy AHP," Int. J. Hydrogen Energy, vol. 43, no. 39, pp. 18059–18076, Sep. 2018, doi: 10.1016/j.ijhydene.2018.08.024.

[61] A. Khoshi, H. Shams Gooshki, and N. Mahmoudi, "The data on the effective qualifications of teachers in medical sciences: An application of combined fuzzy AHP and fuzzy TOPSIS methods," Data Br., vol. 21, pp. 2689–2693, Dec. 2018, doi: 10.1016/j.dib.2018.10.165.

[62] W. K. K. Hsu, S. H. S. Huang, and W. J. Tseng, "Evaluating the risk of operational safety for dangerous goods in airfreights – A revised risk matrix based on fuzzy AHP," Transp. Res. Part D Transp. Environ., vol. 48, pp. 235–247, Oct. 2016, doi: 10.1016/j.trd.2016.08.018.

[63] N. F. Pan, "Fuzzy AHP approach for selecting the suitable bridge construction method," Autom. Constr., vol. 17, no. 8, pp. 958–965, Nov. 2008, doi: 10.1016/j.autcon.2008.03.005.

[64] C. Kahraman, A. Süder, and İ. Kaya, "Fuzzy Multicriteria Evaluation of Health Research Investments," Technol. Econ. Dev. Econ., vol. 20, no. 2, pp. 210–226, Feb. 2014, doi: 10.3846/20294913.2013.876560.

[65] T. Y. Hsieh, S. T. Lu, and G. H. Tzeng, "Fuzzy MCDM approach for planning and design tenders selection in public office buildings," Int. J. Proj. Manag., vol. 22, no. 7, pp. 573–584, Oct. 2004, doi:

10.1016/j.ijproman.2004.01.002.

[66] T. S. Liou and M. J. J. Wang, "Ranking fuzzy numbers with integral value," Fuzzy Sets Syst., vol. 50, no. 3, pp. 247–255, Sep. 1992, doi: 10.1016/0165-0114(92)90223-Q.

Sakarya University Journal of Science

Title: A DLP Module Design based on Plug-in for MS Word

Authors: Hussein AL-SANABANI, Murat İSKEFİYELİ

A DLP Module Design Based on Plug-in for MS Word

Hussein AL-SANABANI[*1], Murat İSKEFİYELİ[2]

Abstract

Inadvertent Data leakage by insiders is considered a serious problem for many organizations. Organizations are increasingly implementing Data Leakage/Loss Prevention solutions also know as (DLP), to protect the confidentiality of their data. Currently, DLP solutions have difficulties to identify confidential data as well as lack the ability to allow users to distinguish confidential from non-confidential data. Moreover, they are limited to work outside organizations. In order to solve this problem, it is important to introduce a DLP-Plugins model where the data owners can identify the privacy of the files during their entire lifecycle (creating, editing, etc.) by classifying them. This model uses security measures such as data encryption and access control to prevent accidental leakage of the classified files by the insiders. The proposed model guarantees that the right user will have access to the correct files according to their security access privilege inside or outside the organization. By always keeping classified files encrypted this will protect them all the time and everywhere. The DLP-Plugins model guarantees the usability for the users, all that will be required is to simply open and close the file as they do normally. As an example of the DLP-Plugins model, we have built a DLP-Plugin for Microsoft Word.

Keywords: Data loss prevention, data leakage prevention, encryption, access control, plugin

1. INTRODUCTION

The rate at which data in digital form enters and leaves organizations today is very high. On a daily basis, a typical enterprise can send and receive millions of email messages and downloads, via various channels an enterprise saves and transfers hundreds or even thousands of files [1].

Customers, business partners, regulators and shareholders expect enterprises to protect their sensitive data that they hold [1]. Leaked data can cause serious damage to an enterprise including but not limited to loss of customer loyalty and

[*] Corresponding Author: hussein.sanabani@ogr.sakarya.edu.tr
[1] Sakarya University, Department of Computer Engineering, Sakarya, Turkey.
ORCID: https://orcid.org/0000-0001-6580-4470
[2] Sakarya University, Department of Computer Engineering, Sakarya, Turkey.
ORCID: https://orcid.org/0000-0002-8210-5070. E-mail:miskef@sakarya.edu.tr

employee confidence which can lead to lawsuits, loss of competitive advantage, political crises, and company closure among others [2], [3]. Because of this, enterprise data is one of the most important assets an enterprise has; protection of this data must therefore be given the first priority [3].

In information security, Data leakage (or data loss) is referred to as unwanted exposure of information [4]. It is one of the most serious security issues that intentionally or unintentionally expose private or sensitive data to an unauthorized entity [2], [5].

Symantec reported that in 2014 there were 1523 data breaches in total compared to 1211 in 2015 and 1209 in 2015. However, the total number of identities exposed in these breaches was 1.2 billion in 2014, 564 million in 2015 and 1.1 billion in 2016 [6]. According to Report by the Global State of Information Security Survey 2015 shows that in 2014, security breaches reached 42.8 million rising by 48% from 2013 [7]. The most-cited culprits of incidents were the employees. More so in 2016 and 2017, about 30% of breaches were reported to have come from current employees while in 2016 and 2017 about 28% and 26% of breaches respectively were reported to have come from former employees [8]. As opposed to outside crime, 32% of the respondents mentioned insider crime as the most damaging and costly breaches in the 2014 US State of Cybercrime Survey [9].

This clearly indicates the extent of the data leakage problem in all kinds of organizations and thus has to be solved. The first step towards solving is by organizations understanding what confidential data is held in the organization, how this data is managed, and how to protect it from unauthorized access [1]. As a result, various data loss prevention (DLP) solutions have been developed to cope with this problem.

According to [5] a DLP solution is defined as : "a system that is designed to detect and prevent the unauthorized access, use, or transmission of confidential information". DLP detect and prevent unauthorized access to confidential data by using deep analysis for both content and

surrounding context around confidential data [10]. [11] also defines DLP solutions as "Products that, based on central policies, identify, monitor, and protect data at rest, in motion, and in use, through deep content analysis".

DLP solutions help to identify, monitor, protect and reduce the risks of confidential data leakage. DLP solutions are used not only to detect and deter an unauthorized user from getting access to confidential data but also to protect confidential data from being inadvertently shared [3]. From the outside lots of different technologies such as Intrusion Detection and Prevention systems, anti-malware and firewalls exist to protect data, DLP systems are designed to do the job from within [3].

According to [2], [12] DLP solutions can be classified as active DLP solutions or inactive DLP solutions. Active DLP solution tracks the confidential data while a user is typing instead of parsing a file after it has been created and saved. While inactive (content-aware and/or context-aware) DLP solution perform both content inspection and/or contextual analysis to identify sensitive data while at-rest, in-use, or in-motion to prevent it from leakage, so it needs to parse all file types.

In [5] DLP solutions can handle data Leakage using two main modes:

Detective modes: The DLP solutions try to detect leakage occurrences and take the suitable corrective action to handle any leakage occurrence that happened by using Context-based inspection, Content-based inspection and Content tagging to identify the sensitive data.

Preventive modes: The DLP solutions prevent leakage before they occur by using several preventive approaches like Access control, Disabling functions, Encryption and Awareness.

The problem: Most of the current DLP solutions have difficult to identify the sensitive data using Content-based (regular expressions, statistical algorithms, pattern matching, keyword comparison, document fingerprint, etc.) or Context-based (file type/size, sender,

header/metadata information, source, destination, etc.) inspection because it is very difficult to identify the sensitive data automatically and also all these techniques need the capability to parse various file types. In addition, they are usually not able to identify the confidential data if they are compressed, encrypted, or obfuscated [3], [13]. Moreover, Most of the DLP solutions don't have the ability to allow insiders to identify the sensitive data as they are the creators of it and the most familiar to what it is containing. They also restrict users to work inside the organization's borders.

The solution: To cope with this problem, this paper introduces DLP-Plugins model that will be added to the legacy applications like Microsoft Office, pdf readers, text editor, media players, and mail applications. This model will let the data owners to identify the sensitive data by classifying it when it was created or edited according to their security level (their Privilege) that they have. The classified data will be protected by using two preventive approaches (access control and encryption). As a result, DLP-Plugins model will guarantee that the sensitive data will be standing protected all the time and everywhere at rest, in motion, and in use.

The contribution: According to the best of our knowledge DLP-Plugins model is the first DLP solution that use Plugins model to protect data against inadvertent data leakage. It doesn't need effort to parse all the files types, but on the other hand, it requires to build Plugin for the wanted application. This model also allows the data owners to identify the confidentiality of data as they are the most knowledgeable on what the data contains. It doesn't matter whether the file is encrypted, compressed, or obfuscated because the confidential data will be protected all the time by encryption. However, when the classified (protected) data need to be modified (edited), a plaintext data will be only available for the authorized users inside the authorized computers inside or outside the organization. This will provide more flexibility to expand organization workplace while guaranteeing the security of the data and without putting a burden on the user. To prove this, we have built a DLP-Plugin for

Microsoft Word as an instant of this DLP-Plugins model.

The rest of this paper is structured as follows. In Section 2 we review the related work. Section 3 presents the motivation for our approach (DLP-Plugins model), then Section 4 describes the proposed model. At last, Section 5 gives conclusions and proposes future directions.

2. RELATED WORK

Papers Of late surveys [5] have shown much interest in research concerning data loss prevention. Data leakage was addressed by current DLP solutions according to what, where and how to protect. What to protect focuses on data-at-rest (DAR), data-in-motion (DIM) and data-in-use (DIU). Where to protect concerns Endpoint and Network. While how to protect describes the leakage handling approaches [5].

A technique for data leak protection (DLP) based on monitoring confidential information as it travels inside a file system on a computer system is described in [14]. The basis of this method is the idea that every travel from confidential to non-confidential item increases the security level of the destination item to that of the source item. Therefore the system can identify the confidential information by spreading labels over all confidential items to avoid hidden passages for information leakage. [4], [5], [15], [16], [17] mention about content tagging. This technique is used to tag the file holding confidential data to identify it, and the enterprise policies will be enforced based on the assigned tag. Once the tag is assigned, the tag stays with the content as it is moved or copied or included in or attached to other files or file types even with the most extreme modification of content, like changing format, encrypting and compressing. Tags can be allocated to files in two methods: manually by the author of the confidential data or automatically by the DLP solution. This technique can identify the file but not the contained confidential data [4].

The content tagging technique is the nearest to our methodology, because it also considers the classification (tagging) of confidential data. What

distinguishes our work is that in our work the data owners identify the sensitive data by classifying it at the time of creation or during modification. And this classified data will be protected by encrypting it from the outset and always remain encrypted.

An active DLP model proposed by [2] can track the sensitive data while the user is typing as opposed to parsing a file after it has been created and saved. [18], [19] discusses Data Leakage Detection using Image and Audio Files. The Goal of this system is to find which data of the distributor (owner of data) has been leaked and if leaked, detects the agent (trusted party) who leaked that data. Basing on data-driven usage control concept, Data loss prevention solution for Microsoft Windows operating systems to allow fine-grained policy-based protection is presented in [13].

In [20] Microsoft has built a DLP solution for Office 365 this DLP solution use context-based inspection, content-based inspection and content tagging (label) to identify the sensitive data and take an action according to the predefined policy. However, our DLP-Plugins model can work together with this solution in perfect harmony.

In general, our work can be distinguished from previous work in that our DLP solution is not an independent solution. It is a plugins model directed to protect certain applications files. This plugin will prevent the potential inadvertent leakage incidents before they occur by taking proper preventive measures such as data encryption and access control in order to put the most effort into preventing potential inadvertent disclosure in the first place. Moreover, DLP-Plugins model relies on the role of the human factor for distinguishing the sensitive data. Finally, this model only focused on unintentional (accidental) data leakage.

3. MOTIVATION

The CIA triad of information security means confidentiality, integrity and availability [21]. We limit our discussion on confidentiality. By definition, Confidentiality of information is

typically assurance that sensitive information is accessed only by authorized users [22]. This task can be achieved by various mechanisms such as device control, encryption and access control [5]. To secure voluminous amounts of personal data from malicious insider and outsider attacks these simplest measures have been used [5]. However, according to [4] the easiest way to deter data leakage is by using DLP solution that relies on security policy and access rights (access control) because they have been in use long enough and follows well established foundations. All of this motivates us to focus on using encryption and access control in DLP-Plugins model to protect the confidentiality of the data.

In [23], [24] Lior Arbel, director of strategic data security solutions at Websens said that "data categorization is one of the key ways that DLP solutions use to determine which data needs heightened levels of security and what does not. Furthermore, in order to protect data they would need to classify the data first and then run discovery on that data". Based on this our DLP-Plugins model relies on classifying the data from the outset when it was created, or at any point of its lifecycle.

Christian Toon, head of information risk at Iron Mountain emphasizes that to achieve an effective DLP implementation, the human element should not be ignored: "Data loss prevention technologies are only as good as the employees using them" [24]. For that, we assigned the classification process of the sensitive data to the authorized users (data owners). However, an organization can attempt to force its employees to comply with its regulations by using control mechanisms, surveillance, and monitoring. But this approach has proven to be ineffective in several cases such as the incident of Edward Snowden [25]. Organizations can also rely on the acceptance and the cooperation of their employees because there is no never-known-to-fail method to prevent data leakage [25]. This means that intentional data leakage is harder to prevent, for that reason our method is focused only on unintentional (accidental) data leakage.

4. THE PROPOSED MODEL

In this section, we describe the classification method used by the DLP-Plugins model, the DLP-Plugins design components, how the model works and its implementation in Microsoft Word and finally we show the performance of the model.

4.1. Classification method

In any organization, corporate institution or Government department, data and information are classified according to some criteria ranging from highly to less confidential data. This criterion will define who is permitted to use and distribute the data [15]. Accordingly [26]–[28] the most common example that is used by governments and organizations is :Top Secret, Secret, Confidential and Restricted. However most governments and organizations have their own rules to state the security levels, determine the level for the data and who has permission to handle this level. Consequently, it is well known that the mostly used organizational structure type by both companies and governments is hierarchy (pyramidal). Based on this, we use hierarchy classification strategy as shown in Figure 1.

Figure 1. Hierarchy Classification Strategy

We classified the Data into five levels of security sequenced from Top Secret to Personal. The user with security Level-1 (Top Secret) has all privileges and can access to all other security levels. The user that has Level-2 can have access at this level and all levels that are bigger than Level-2, i.e. Level-3, Level-4 and Level-5. Also, the user that has Level-3 can have access at this level and all levels bigger than three, this means that he can have access to the Level-3, Level-4,

and Level-5. The same for the Level-4 and Level-5 security levels. However, the security Level-5 (Personal) is different for each user i.e. each user can see only his/her own data that classified as Level-5 but he/she cannot see the security Level-5 for the other users because it contains their personal information. This proposed classification model can be adapted or changed to fit organization's and government's requirements.

4.2. DLP-Plugins components

The DLP-Plugins model that we build has three main Components:

4.2.1. DLP-Administrator Panel

DLP-Administrator Panel is a website that provides a central control mechanism to manage users. For example (add new user or delete existing user, change the privilege of existing user in a specific computer, etc.) and for adding or deleting a new DLP-Plugin.

4.2.2. DLP-Web Service

DLP-Web Service is a web service used to check both user and computer identity to know the user security level that he/she has, then send appropriate encryption and decryption keys through the encrypted channel to DLP-Plugin according to the security level that the user has.

4.2.3. DLP-Plugin

DLP-Plugin is the plugin that will be added to the legacy application to protect their data by decrypting the files when opening them and encrypting them when closing them by the authorized users. This DLP-Plugin will communicate with DLP-Web Service to check the user identity and to get his/her security level (appropriate encryption and decryption keys). Also, it will let the users classify their data or change its security level according to the security level that they have.

The Figure 2 shows that the first and second components will be fixed to all the DLP-Plugins that will be added to the legacy applications but the third components will be varying according to

the diversity of the applications that will be added to them. As an example of these DLP-Plugins, we have developed one DLP-Plugin for Microsoft word.
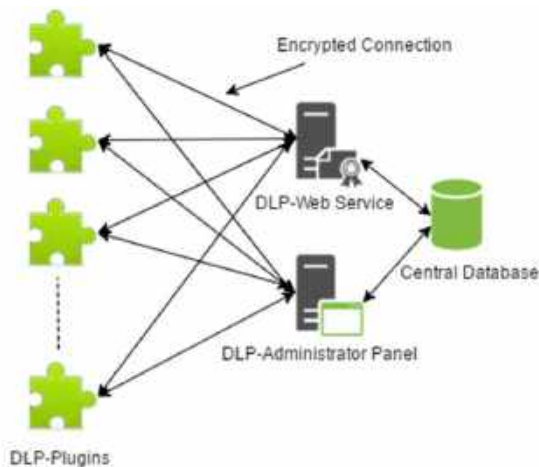

Figure 2. DLP-Plugins components

Figure 3. An example of a network topology of an organization

4.3. How the DLP-Plugins model work (ex: Microsoft Word)?

Assume we have organization that has a network topology like this in the Figure 3:

Suppose we want to protect the organization Word documents from unintentional leakage to the outside of the organization. And also we want to protect these documents from having illegal access by unauthorized insiders but at the same time, we want to provide legal access while guaranteeing the usability for the insiders from inside or outside the organization.
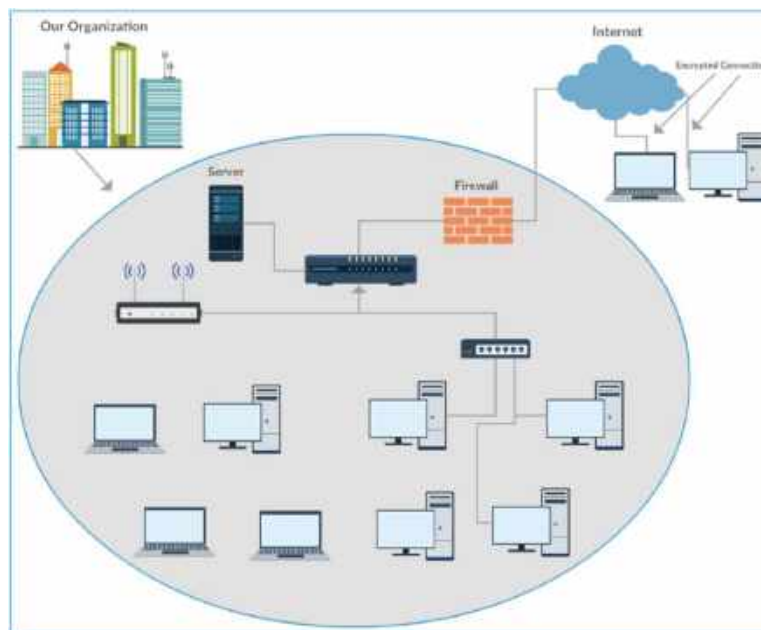
To achieve this we build DLP-Plugin for Microsoft Word application to protect Word documents from accidental leakage by internal workers or to have access by unauthorized users. This Plugin is an add-in feature for Microsoft Word that allows the DLP functions. It works with all Windows desktop versions of Microsoft Word (2010, 2013,2016, 365) except the versions that are prior to Word 2010. To understand how

the model works we divided it into three processes:

4.3.1. Word application startup process

When a word application is started the DLP-Plugin inside the word application will start to recognize the computer and user identity. After getting the identity, the DLP-Plugin will connect to the DLP-Web Service to get his security level

(encryption and decryption keys) according to his identity. Now that the computer has its privileges (security level) it means it has appropriate encryption and decryption keys.

Figure 4 below depicts how the computer request his security level by sending his identity to the DLP-Web Service and getting his security level (encryption and decryption keys) through the encrypted channel.
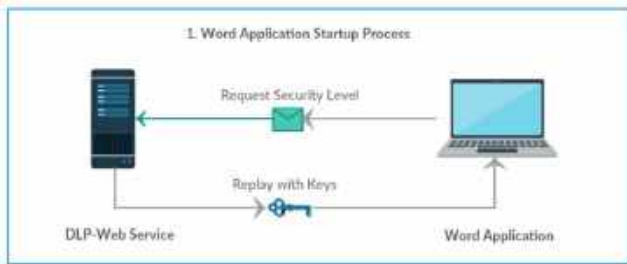

Figure 4. Word application startup process

4.3.2. Word document opening process

Every classified document is by default encrypted. From the first process described above, the user acquires his lawful decryption and encryption keys. When the user starts to open this classified document the DLP-Plugin will check the security level of the document and if this security level is legal to that user it means the DLP-Plugin has appropriate key to decrypt this document. So the DLP-Plugin will decrypt and open the document for that user. If the document security level is not legal to that user the document will remain encrypted because DLP-Plugin doesn't have appropriate key to decrypt this document. This will happen quite easily without any additional burden to the user. The user just double clicks on the document that he/she wants to open and the decryption process will be done in the background without feeling any different. This process is described in the Figure 5 below:
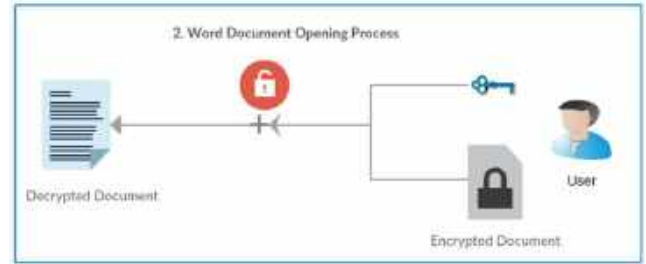

Figure 5. Word document opening process

4.3.3. Word document closing process

While the document is open, the user can change the security level of the document to any level that he has. When closing the document, the DLP-Plugin will encrypt the document according to its security level and its appropriate key. This is described in the Figure 6 below:
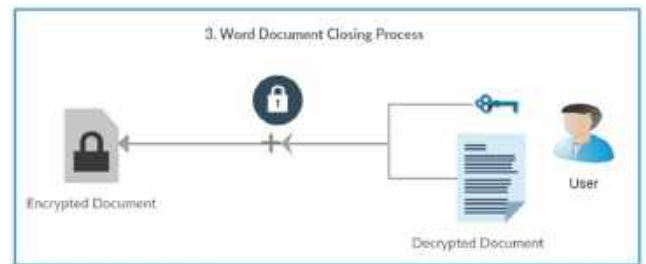

Figure 6. Word document closing process

These three processes provide a compatible protection for the sensitive data from inadvertently leaking to the outside of the organization and also preventing access by unauthorized users.

The pseudocode for these processes is shown in Algorithm 1.

Algorithm 1 Word application processes

Input: (computerID, userID)
  // *Requst SecurityLevel From DLP-Web Service*
  userSecurtiyLevel := RequstSecurityLevel(computerID,userID)
  InitializeUserInterfaceWithUserSecurtiyLevel(userSecurtiyLevel)
  While *ture*
    ListenToUserOpen_CloseDocumentEvent(event)
    If *event.Type = OPEN_Document*
      docSecurtiyLevel := GetDocumentSecurtiyLevel(event.Doc)
      if *CheckIslegalForDoc(userSecurtiyLevel, docSecurtiyLevel)*
      = true
        doc := DecryptDoc (event.Doc, userSecurtiyLevel)
        ShowDocment(doc)
      Otherwise
        ShowDocment(event.Doc) // Will show encrypted data
      end if
    end if
    if *event.Type = Close_Document*

```
docSecurtiyLevel := GetDocumentSecurtiyLevel(event.Doc)
if docSecurtiyLevel ≠ null
    encryptedDoc:= encryptDoc(event.Doc, userSecurtiyLevel)
    SaveAndCloseDocument(encryptedDoc)
Otherwise
    CloseDocument(event.Doc)
end if
  end if
 end while
```

Figure 7 shows how the document for both the authorized user and the unauthorized user will appear. The document in the figure was classified as Level-1. The user on the left side of the figure is authorized because he has this level so the document appears to him normally. But the user on the right side of the figure doesn't have Level-1 so he will see just encrypted data.

4.4. DLP-Plugins model Performance (ex: Microsoft Word)

The execution results are taken on a machine having Intel Core i3 (2.67 GHz) processor with 4 GB RAM and Windows 10 64-bit operating System and Microsoft Word 2013 32-bit. The C# 4.5 .NET Framework platform is used for implementation. The .NET Framework build-in Cryptography Dynamic-link library (DLL) for (AES, RC2, TDES) and Bouncy Castle Cryptography DLL for (RC4, Blowfish, Twofish) are used for encryption algorithm implementation. We do the three tests for each process described before in section 4.3. and all tests were run 4 times and their average was calculated, see tables for each test.
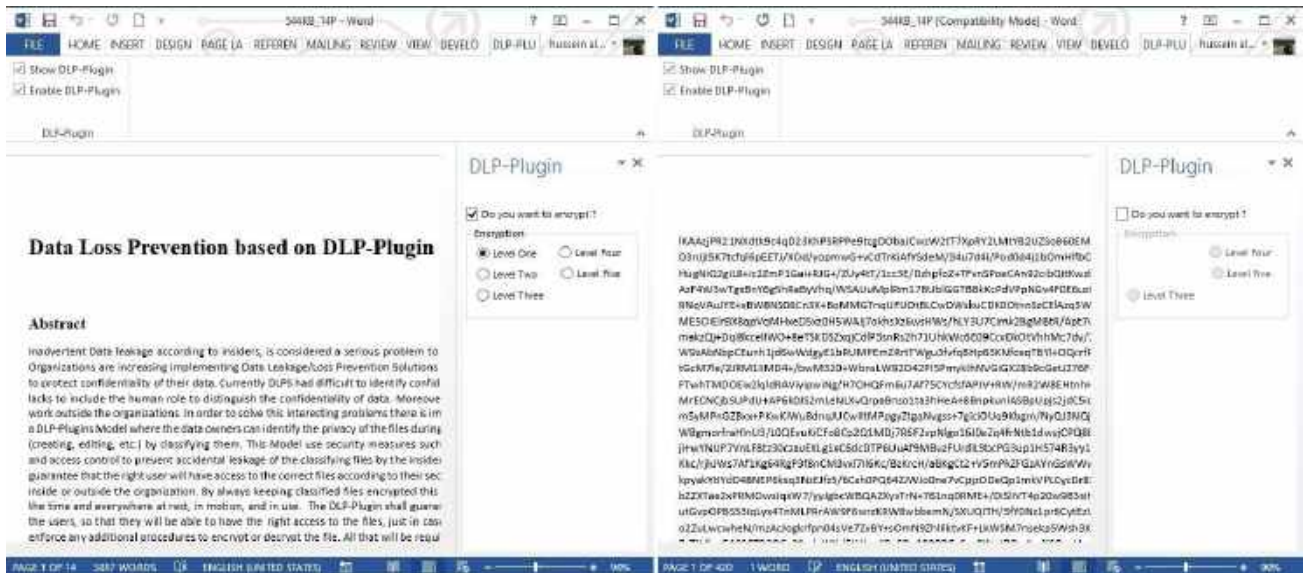

Figure 7. The appearances of the document for both authorized and unauthorized users

4.4.1. Frist test for startup the Word application

The time to get the computer identity and connect to the DLP-Web Service varies from 0.01 second to 2.0 seconds depend on the internet connection.

4.4.2. Second test for opening document

Table 1 shows the overall time of opening and decrypting the document for different file sizes and number of pages using different encryption algorithms. The overall time is a summation of three times. The first one is the time of reading the encrypted data from the document. The second one is the time of decrypting the ciphertext to get the plaintext (XML string). The third one is the time of parsing the XML string to make the document and open that document. The key size of AES, RC4, RC2, Blowfish and Twofish is 128 bits and the key size of TDES is 112 bits.

Table 1. The overall time of opening and decrypting the document for different file sizes and number of pages using different encryption algorithms

File Size	Number of Pages	AES-128	RC4-128	RC2-128	TDES-112	Blowfish-128	Twofish-128
3.45MB	9	2.695263	2.802374	2.981562	3.652868	4.309394	3.924538
544KB	14	1.204576	1.297097	1.469328	1.606416	1.703509	1.559582
847KB	50	2.050756	2.08891	2.244023	2.535857	2.759254	2.794405
1.01MB	70	2.436801	2.433337	2.696202	2.995279	3.413842	3.408304
1.60MB	100	3.377409	3.309007	3.481061	3.926474	4.511645	4.614573
1.70MB	150	4.194024	4.068522	4.567168	5.437092	5.668988	5.676459
1.77MB	200	4.501686	5.031383	5.319095	6.192539	6.639504	7.018375
2MB	300	6.088659	6.856815	7.222979	8.49772	9.526075	9.267652
Average	111.625	3.318647	3.485931	3.747677	4.355531	4.816526	4.782986

Figure 8 summarizes the results presented in Table 3 for opening document test.
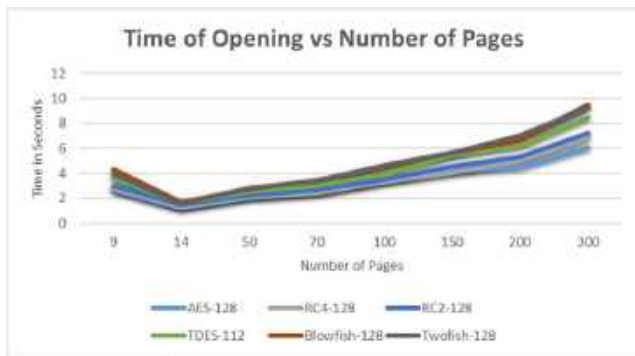

Figure 8. Time of opening vs number of pages

Observation: The results reveal that the time of opening the encrypted document depends on file size and number of pages. Increasing the number of pages or file size will increase the opening time of the document. It can be clearly seen that AES performs better followed by RC4 with key size 128 bits among the encryption algorithms tested.

4.4.3. Third test for closing document

Table 2 shows the overall time of closing and encrypting the document for different file sizes and number of pages using different encryption algorithms. The overall time is a summation of three times. The first one is the time of reading the XML string that represents the document. The second one is the time of encrypting the plaintext (XML string) to get ciphertext. The third one is the time of saving and closing the document. The key size of AES, RC4, RC2, Blowfish and Twofish is 128 bits and the key size of TDES is 112 bits.

Table 2. The overall time of closing and encrypting the document for different file sizes and number of pages using different encryption algorithms.

File Size	Number of Pages	AES-128	RC4-128	RC2-128	TDES-112	Blowfish-128	Twofish-128
3.45MB	9	2.690997	2.70322	2.564246	3.130119	3.694704	3.538402
544KB	14	0.608243	0.665121	0.771921	0.785535	0.995787	0.970647
847KB	50	1.319755	1.471099	1.49678	1.571272	1.724479	1.939608
1.01MB	70	1.520164	1.730294	1.644743	1.920754	2.057411	2.136468
1.60MB	100	2.263135	2.42154	2.467695	2.697342	3.111858	2.990853
1.70MB	150	2.922747	2.90495	3.402629	3.528995	4.196691	4.167968
1.77MB	200	2.996991	3.407393	3.750937	3.88995	4.706226	4.964365
2MB	300	3.739411	4.003328	4.3916	4.844758	5.351529	5.30996
Average	111.625	2.25768	2.413368	2.561319	2.796091	3.229835	3.252284

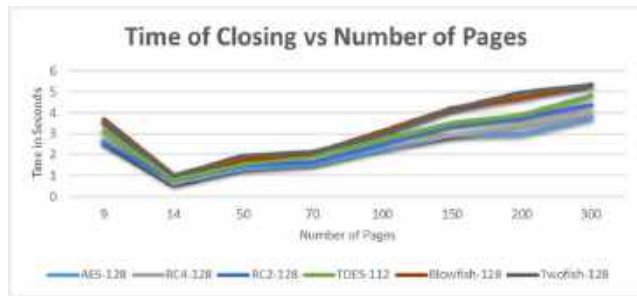Figure 9 summarizes the results presented in Table 2 for closing document.

Figure 9. Time of closing vs number of pages

Observation: The result shows that the time of closing and encrypting the document depends on file size and number of pages. Increasing the number of pages or file size will increase the closing time of the document. As with opening the document, AES performs better followed by RC4 with key size 128 bits among the encryption algorithms tested.

4.5. Discussion

For We can see that the time for closing is smaller than the time for opening this shows that the major overhead is a result of parsing the XML to make the word document. In general, we can ignore the time required to close the file because the closing processing is run in the background while the Word application can do other processes. The second and the third tests revealed that the time for opening and closing document is proportional to the number of pages and file size. As the number of pages or the file size increases, the time for opening and closing the document also increases proportionally to a number of pages or file size and vice versa. Further the results show that the AES and RC4 encryption algorithms are the fastest and suitable among the other algorithms.

5. CONCLUSIONS AND FUTURE WORK

This paper introduces a DLP-Plugins model that use two preventive approaches: Access control and Encryption to prevent the probable unintentional data leakage incidents before they occur. This DLP-Plugins model let the authorized insiders (data owners) identify the sensitive data by classifying them during its entire lifecycle rather than scanning both content and context of the data as most of the commercial DLP solutions do because it easily avoids the imposed high overhead of scanning. DLP-Plugins model provides more flexibility to work outside of the organization and guarantees the usability for the users that they have the right to access the classified documents easily, just open and close the document as they do normally.

We also realize that in our approach it is too easy for authorized users to intentionally leak data. But as it's known that intentional data leakage is impossible to prevent it and that is the problem that all the current DLP solutions faced, so the organizations should rely on the acceptance and the cooperation of their employees. Consequently, our DLP-Plugins model trusts on the approval and the cooperation of the employees and focused only on the unintentional leaking of data. However DLP-Plugins model can work together with current DLP solutions in perfect harmony.

The implementation of DLP-Plugin for Microsoft Word and the performance results show that the proposed DLP-Plugins model is feasible, easy to use and practical using current technologies. The result revealed that the AES and RC4 Encryption algorithms perform better and thus most suitable among the other algorithms.

As a future work, we suggest developing DLP-Plugins for protecting all types of office documents (Excel, PowerPoint, Access, etc.), for E-mail Program to protect all E-mail message, for PDF reader to Protect PDF file and also for images and videos player to protect all images and videos files. Also, we suggest integrating those DLP-Plugins with the current DLP solutions in the market.

Research and Publication Ethics

This paper has been prepared within the scope of international research and publication ethics.

Ethics Committee Approval

This paper does not require any ethics committee permission or special permission.

Conflict of Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

REFERENCES

[1] S. Liu and R. Kuhn, "Data loss prevention," IT Prof., vol. 12, no. 2, pp. 10–13, Mar. 2010.

[2] J.-S. Wu, Y.-J. Lee, S.-K. Chong, C.-T. Lin, and J.-L. Hsu, "Key Stroke Profiling for Data Loss Prevention," in 2013 Conference on Technologies and Applications of Artificial Intelligence, pp. 7–12, 2013.

[3] R. Tahboub and Y. Saleh, "Data Leakage/Loss Prevention Systems (DLP)," in 2014 World Congress on Computer Applications and Information Systems (WCCAIS), vol. 1, pp. 1–6, 2014.

[4] S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, "A survey on data leakage prevention systems," J. Netw. Comput. Appl., vol. 62, pp. 137–152, Feb. 2016.

[5] A. Shabtai, Y. Elovici, and L. Rokach, A Survey of Data Leakage Detection and Prevention Solutions. Boston, MA: Springer US, 2012.

[6] Symantec, "Internet Security Threat Report," vol. 22, no. April, 2017.

[7] PwC, "Managing cyber risks in an interconnected world: Key fndings from The Global State of Information Security Survey 2015," 2014.

[8] PwC, "The Global State of Information Security Survey 2018," pwc.com, 2018. [Online]. Available: https://www.pwc.com/us/en/services/cons ulting/cybersecurity/library/information-security-survey.html. [Accessed: 09-May-2020].

[9] N. Mickelberg, Kevin; Schive, Laurie; Pollard, "US cybercrime: Rising risks, reduced readiness Key fndings from the 2014 US State of Cybercrime Survey," 2014.

[10] S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, "Detecting Data Semantic: A Data Leakage Prevention Approach," in 2015 IEEE Trustcom/BigDataSE/ISPA, vol. 1, pp. 910–917, 2015.

[11] R. Mogull and M. Rothman, "Understanding and Selecting a Data Loss Prevention Solution," 2017.

[12] B. Reed and N. Wynne, "Magic Quadrant for Content-Aware Data Loss Prevention," 2016.

[13] T. Wuchner and A. Pretschner, "Data Loss Prevention Based on Data-Driven Usage Control," in 2012 IEEE 23rd International Symposium on Software Reliability Engineering, pp. 151–160, 2012.

[14] M. Petkovic, M. Popovic, I. Basicevic, and D. Saric, "A Host Based Method for Data Leak Protection by Tracking Sensitive Data Flow," in 2012 IEEE 19th International Conference and Workshops on Engineering of Computer-Based Systems, pp. 267–274, 2012.

[15] M. H. Matthee, "Tagging Data to Prevent Data Leakage (Forming Content Repositories)," SANS Inst., pp. 1–26, 2016.

[16] I. McAfee, "McAfee Host Data Loss Prevention 2.2.1 Product Guide." McAfee, Inc, pp. 1–80, 2008.

[17] I. McAfee, "McAfee Data Loss Prevention 11.0.300 Product Guide." McAfee, Inc, pp. 1–211, 2020.

[18] S. S. Dandavate, P.P.; Dhotre, "Data Leakage Detection using Image and Audio Files," Int. J. Comput. Appl., vol. 115, no. 8, pp. 1–4, 2015.

[19] P. S. V. Kale, Sandip A.; Kulkarni, "Data Leakage Detection," Int. J. Adv. Res. Comput. Commun. Eng., vol. 1, no. 9, pp. 668–678, 2012.

[20] Microsoft, "Overview of data loss prevention," microsoft.com, 2019. [Online]. Available: https://docs.microsoft.com/en-us/microsoft-365/compliance/data-loss-prevention-policies?view=o365-worldwide. [Accessed: 09-May-2020].

[21] J. Andress, "What is Information Security?," in The Basics of Information Security, Elsevier, pp. 1–22, 2014.

[22] B. Guttman and E. Roback, An Introduction to Computer Security : The NIST Handbook, vol. SP800, no. 12. 1995.

[23] L. Arbel, "Data loss prevention: the business case," Comput. Fraud Secur., vol. 2015, no. 5, pp. 13–16, May 2015.

[24] T. Caldwell, "Data loss prevention – not yet a cure," Comput. Fraud Secur., vol. 2011, no. 9, pp. 5–9, Sep. 2011.

[25] B. Hauer, "Data and Information Leakage Prevention Within the Scope of Information Security," IEEE Access, vol. 3, pp. 2554–2565, 2015.

[26] M. Diri, Mustafa; Gülçiçek, "Türkiye'de Kamu Hizmetinin Görülmesinde Kullanılmakta Olan Gizlilik Derecesi Tanımları: Uygulamadaki Sorunlar ve Çözüm Önerileri," Maliye Derg., vol. 162y, pp. 497–537, 2012.

[27] Office Cabinet UK, Government Security Classifications April 2014, pp. 1–35, 2013.

[28] Office Cabinet UK, International Classified Exchanges, no. 1–23. 2015.

Sakarya University Journal of Science

Title: On Idempotent Units in Commutative Group Rings

Authors: Ömer KÜSMÜŞ

On Idempotent Units in Commutative Group Rings

Ömer KÜSMÜŞ[*1]

Abstract

Special elements as units, which are defined utilizing idempotent elements, have a very crucial place in a commutative group ring. As a remark, we note that an element is said to be idempotent if $r^2 = r$ in a ring. For a group ring RG, idempotent units are defined as finite linear combinations of elements of G over the idempotent elements in R or formally, idempotent units can be stated as of the form $id(RG) = \{\sum_{r_g \in id(R)} r_g g : \sum_{r_g \in id(R)} r_g = 1 \text{ and } r_g r_h = 0 \text{ when } g \neq h\}$ where $id(R)$ is the set of all idempotent elements [3], [4], [5], [6]. Danchev [3] introduced some necessary and sufficient conditions for all the normalized units are to be idempotent units for groups of orders 2 and 3. In this study, by considering some restrictions, we investigate necessary and sufficient conditions for equalities:

$i. V\big(R(G \times H)\big) = id(R(G \times H))$,

$ii. V\big(R(G \times H)\big) = G \times id(RH)$,

$iii. V\big(R(G \times H)\big) = id(RG) \times H$

where $G \times H$ is the direct product of groups G and H. Therefore, the study can be seen as a generalization of [3], [4]. Notations mostly follow [12], [13].

Keywords: idempotent, unit, group ring, commutative

1. INTRODUCTION

As widely known, a group ring RG of a given group G over a ring R is defined as the set of finite sums in $\{\sum_{g \in G} r(g)g : r(g) \in R\}$. The sets of all units and normalized units in RG are denoted by $U(RG)$ and $V(RG)$ respectively [8]. Idempotent units are described as elements of the form $\sum_{r_g \in id(R)} r_g g$ such that $\sum_{r_g \in id(R)} r_g = 1$ and $r_g r_h = 0$ when $g \neq h$ [3]. Let $id_C(R)$ display a complete set of orthogonal idempotent elements. For a group G, the p-primary component of G is

---

[*]Corresponding Author: omerkusmus@yyu.edu.tr
[1]Van Yüzüncü Yıl University, Department .of Mathematics, Van, Turkey.
ORCID: https://orcid.org/0000-0001-7397-0735

generally shown by G_p and so the maximal torsion part G_0 of G is a co-product of primary components [3], [4], [11]. Since each idempotent unit can generate novel units of the same form, we can utilize the following notation for idempotent units [3], [4]:

$$id(RG) = \langle \sum_{r_g \in id_C(R)} r_g g : g \in G \rangle$$

By the way, we should recall that each element in G is said to be *trivial unit* of RG [1], [2]. Besides, we can observe that every trivial units are also idempotent units. Danchev introduced some necessary and sufficient conditions to be $V(RG) = id(RG)$ for groups of orders 2 and 3 as follows respectively [3]:

Proposition 1. Assume that $|G| = 2$. Then $V(RG) = id(RG)$ if and only if

$$1 - 2r \in U(R) \Leftrightarrow r \in id(R)$$

for all $r \in R$.

Proposition 2. Assume that $|G| = 3$. Then $V(RG) = id(RG)$ if and only if

$$1 + 3r^2 + 3f^2 + 3rf - 3r - 3f \in U(R)$$

implies that $r^2 = r$, $f^2 = f$ and $rf = 0$ (Notice that $r^2 = r$, $f^2 = f$ and $rf = 0$ directly implies that $1 + 3r^2 + 3f^2 + 3rf - 3r - 3f = 1$ is a trivial unit).

Group rings over rings of prime characteristic have been classified in terms of the equality $V(RG) = id(RG)$ as follows [3]:

Theorem 3. Let R be a unital and commutative ring of a prime characteristic p. Since G is a non-trivial and Abelian group, $V(RG) = id(RG)$ if and only if the nil-radical of R, $N(R) = 0$ and at most one of the followings is satisfied:

i. Maximal torsion part of G is trivial ($G_t = 1$),

ii. $|G| = p = 2$ and R is a Boolean ring,

iii. $|G| = 2$, $\forall r \in R$, $1 - 2r \in U(R) \Leftrightarrow r^2 = r$,

iv. $|G| = 3$ and

$$1 + 3r^2 + 3f^2 + 3rf - 3r - 3f \in U(R)$$

implies that $r^2 = r$, $f^2 = f$ and $rf = 0$.

If we consider a cyclic group G of prime order greater than 3, we can construct a unit which is not an idempotent unit using Bass cyclic unit forms as follows [3], [4]:

$$u = (1 + g)^{p-1} - \frac{2^{p-1} - 1}{p} \hat{g}$$

where

$\hat{g} = 1 + g + \cdots + g^{p-1}$ and $G = \langle g : g^p = 1 \rangle$. This means that investigating necessary and sufficient conditions to be $V(RG) = id(RG)$ for cyclic groups G of prime order ≥ 5 is meaningless.

In the next section, we consider non-cyclic groups of order ≥ 6 and construct some necessary and sufficient conditions for all normalized units to be idempotent units. Throughout the paper, R is a commutative ring with unity and D is the direct product of groups.

2. MAIN RESULTS

In this section, we should remember the following notations for the set of orthogonal idempotent elements and a complete set of orthogonal idempotent elements in R are used throughout the paper respectively:

$$id_0(R) = \{e_i \in R : e_i^2 = e_i, e_i e_j = 0 \text{ for } i \neq j\},$$

$$id_C(R) = \{e_i \in id_0(R) : \sum e_i = 1 \}.$$

Let G and H be two Abelian groups with p-primary and q-primary components G_p and H_q respectively. Using maximal torsion parts of G and H, we indicate the maximal torsion part of the direct product $D = G \times H$ as follows:

$$D_0 = \coprod_p \coprod_q G_p \times H_q = \coprod_q G_p \times \coprod_q H_q$$

where p and q are prime integers.

Owing to the fact that $G_p = 1$ means that G has no p-primary component [3], we intend by the notation $G_p \times H_q = 1$ that G or H has no p-primary or q-primary components respectively. Let $\wp$ denote the set of all prime integers.

Definition 4.

$$supp_C(G \times H) = \{pq: G_p \times H_q \neq 1\}$$

is said to be *support* of $G \times H$.

Example 5. Let $G = \mathbb{Z}_4$ and $H = \mathbb{Z}_9$. Since, $G_2 \neq 1$ and $H_3 \neq 1$, we say $6 \in supp_C(\mathbb{Z}_{36})$.

Definition 6.

$$zd_C(R) = \{pq: \exists 0 \neq r \in R, pqr = 0\}$$

and

$$inv_C(R) = \{pq: pq.1 \in U(R)\}$$

Now, using [4], we can present the first of results in this paper as follows:

Theorem 7. Let G and H be two Abelian groups. Then, since $D = G \times H$,

$$V(RD) = id(RD)$$

if and only if

$N(R) = 0$, $V(RD_0) = id(RD_0)$ and one of the following statements hold:

$a.\ D = D_0$,

$b.\ D \neq D_0$ and

$supp_C(D) \cap [\ inv_C(R) \cup zd_C(R)\ ] = \emptyset$.

Proof. Let $V(RD) = id(RD)$. In this situation,

as $D_0 \subseteq D$, we can write the embedding $V(RD_0) \hookrightarrow V(RD) = id(RD)$. Hence,

$$id(RD_0) \subseteq V(RD_0)$$

and

$$V(RD_0) \cap id(RD) = id(RD_0)$$

yield that $V(RD_0) = id(RD_0)$. Now, let r be a nilpotent element in R. Thus, $(rgh)^k = 0$ for some $g \in G, h \in H, k \in \mathbb{N}$. As $1 - (rgh)^k$ is

$$(1 - rgh)(1 + rgh \ldots + r^{k-1}g^{k-1}h^{k-1}) = 1,$$

we conclude that

$$1 + r - rgh \in V(RD) = id(RD)$$

This shows that $r \in N(R) \cap id(R)$ and so $N(R) = 0$.

If D consists only of torsion part, the proof terminates. If not, $(D \neq D_0)$ then assuming

$$supp_C(D) \cap inv_C(R) \neq \emptyset$$

we say that there exists

$$pq \in supp_C(D\ ) \cap inv_C(R)$$

$\exists p, q \in \wp$. This means that $G_p \times H_q \neq 1$ and thus $\exists g \in G_p$ and $\exists h \in H_q$. Applying these torsion elements g and h, we can generate an idempotent

$$e = e(p, q) = \frac{\widehat{(gh)}}{pq}$$

where $\widehat{(gh)} = 1 + gh + \cdots + (gh)^{p-1}$. Using [4], we can compose a unit

$$u = 1 - e + exy \in V(RD)\backslash D$$

where $\exists x \in G\backslash G_0$ ve $\exists y \in H\backslash H_0$ are torsion-free elements. Explicit form of $u = 1 - e + exy$ can be written as

$$u = 1 - p^{-1}q^{-1} - p^{-1}q^{-1}gh - \cdots$$

$$-p^{-1}q^{-1}(gh)^{pq-1} + p^{-1}q^{-1}xy$$

$$+p^{-1}q^{-1}xygh + \cdots + p^{-1}q^{-1}xy(gh)^{pq-1}$$

However, the fact that coefficients -1 and 1 are not orthogonal idempotents displays that

$$u \in V(RD)\backslash id(RD)$$

This contradiction indicates that

$$supp_C(D) \cap inv_C(R) = \emptyset$$

On the other hand, assume that

$$supp_C(D) \cap zd_C(R) \neq \emptyset$$

In this case, for $\exists pq \in supp_C(D) \cap zd_C(R)$ and $\exists 0 \neq r \in R, pqr = 0$. Then we get

$$r(1 - gh)^{pq} = r[1 - \binom{pq}{1}gh + \cdots$$
$$+ \binom{pq}{pq-1}(gh)^{pq-1} - 1]$$
$$= pqr \sum n_i(p,q)(gh)^i = 0$$

where $\exists gh \in G_p \times H_q$ and $n_i(p,q) \in \mathbb{N}$. This gives the unit

$$\omega = 1 + r - rgh \in V(RD)$$

that is not an idempotent unit which is a contradiction as well. This means that

$$supp_C(D) \cap zd_C(R) = \emptyset$$

For the converse of the proof, we assume that $N(R) = 0$ and

$$V(RD_0) = id(RD_0)$$

If $D = D_0$,

$$V(RD) = V(RD_0) = id(RD_0) = id(RD)$$

and thus the proof terminates. If $D \neq D_0$, we define the group epimorphism:

$$\phi_C: D \longrightarrow D/D_0$$
$$gh \longmapsto ghD_0$$

Extending linearly ϕ_C yields that

$$\phi_C: RD \longrightarrow R(D/D_0)$$
$$\sum_{(g,h) \in G \times H} \alpha(g,h)gh \mapsto \sum_{(g,h) \in G \times H} \alpha(g,h)gh\, D_0$$

If we restrict ϕ_C to unit groups of RD and $R(D/D_0)$, we see that

$$\phi_C(V(RD)) \subseteq V(R(D/D_0))$$

Let $V(RD_0) = id(RD_0)$. It is clear that

$$id(R(D/D_0)) \subseteq V(R(D/D_0))$$

For the converse inclusion, assume that

$$\exists u \in V(R(D/D_0)) \backslash id(R(D/D_0))$$

In this case, the augmentation map

$$\varepsilon: V(R(D/D_0)) \longrightarrow V(RD_0) = id(RD_0),$$

$$\varepsilon\left(\sum_{gh \in G \times H} \alpha(g,h)gh\, D_0\right) = \sum_{gh \in G \times H} \alpha(g,h)\, D_0$$

gives the image of u which is not an idempotent unit as

$$\varepsilon(u) \in V(RD_0) \backslash id(RD_0)$$

which contradicts with the assumption. Hence, we conclude that $V(R(D/D_0)) = id(R(D/D_0))$ by inspiring from [7],[9],[10].

It is obvious that $\phi_C(id(RD)) = id(R(D/D_0))$ and $\phi_C(id(RD)) \subseteq \phi_C(V(RD))$. Since

$$\phi_C(id(RD)) = id(R(D/D_0))$$

and

$$\phi_C(V(RD)) \subseteq V(R(D/D_0)) = id(R(D/D_0)),$$

we attain the inclusion:

$$\phi_C(V(RD)) \subseteq \phi_C(id(RD))$$

Applying the first isomorphism theorem serves that

$$\frac{V(RD)}{Ker\phi_C \subseteq V(RD_0)} \simeq \phi_C(V(RD))$$
$$= \phi_C(id(RD))$$

Remember that $\phi_C(id(RD)) = id(R(D/D_0))$. Thus,

$$V(RD) = Ker\phi_C.id(R(D/D_0))$$

$$\subseteq V(RD_0).id(R(D/D_0))$$

By the hypothesis $V(RD_0) = id(RD_0)$, we can write

$$V(RD) \subseteq id(RD_0).id(R(D/D_0))$$

so $V(RD) \subseteq id(RD)$. Thus, $V(RD) = id(RD)$. ∎

Theorem 8. Let $D = K_4$ (Klein 4-Group). Then, $V(RD) = id(RD)$ if and only if

$$1 - 4rs - 4rf - 4sf - 16rsf \in U(R)$$

implies that $r, s, f \in id_C(R)$. One can

notice that if $r, s, f \in id_C(R)$,

$$1 - 4rs - 4rf - 4sf - 16rsf = 1$$

is already a unit in R.

Proof. Since

$$D = K_4 = \langle g, h: g^2 = h^2 = 1, gh = hg \rangle$$

the group ring RD can be seen as an R-module as $RD = \langle 1, g, h, gh \rangle_R$. As the normalized units have augmentation one [12], we can state the normalized unit group as

$V(RD) =$

$\{1 - (r + s + f) + rg + sh + fgh: r, s, f \in R\}$.

Assume that $V(RD) = id(RD)$. Then, parameters of units in $V(RD)$ are idempotent elements in R. Let us consider a unit in $V(RD)$ as

$$u = 1 - (r_1 + s_1 + f_1) + r_1 g + s_1 h + f_1 gh$$

with the inverse

$$u^{-1} = 1 - (r_2 + s_2 + f_2) + r_2 g + s_2 h + f_2 gh.$$

Then,

$$uu^{-1} = 1 - X + Yg + Zh + Tgh = 1$$

so $X = Y = Z = T = 0$ where

$$X = (r_1 + s_1 + f_1 + r_2 + s_2 + f_2) - (r_1 + s_1 + f_1)(r_2 + s_2 + f_2) - r_1 r_2 - s_1 s_2 - f_1 f_2 = 0,$$

$$Y = r_2(1 - r_1 - s_1 - f_1) + r_1(1 - r_2 - s_2 - f_2) + s_1 f_2 + s_2 f_1 = 0,$$

$$Z = s_2(1 - r_1 - s_1 - f_1) + s_1(1 - r_2 - s_2 - f_2) + r_1 f_2 + r_2 f_1 = 0,$$

$$T = f_2(1 - r_1 - s_1 - f_1) + f_1(1 - r_2 - s_2 - f_2) + r_1 s_2 + r_2 s_1 = 0.$$

Arranging X, Y, Z and T, we get a system of linear equations as follows:

I. $r_2(2r_1 + s_1 + f_1 - 1) + s_2(r_1 + 2s_1 + f_1 - 1) + f_2(r_1 + s_1 + 2f_1 - 1) = r_1 + s_1 + f_1,$

II. $r_2(1 - 2r_1 - s_1 - f_1) + s_2(-r_1 + f_1) + f_2(-r_1 + s_1) = -r_1,$

III. $r_2(-s_1 + f_1) + s_2(1 - r_1 - 2s_1 - f_1) + f_2(r_1 - s_1) = -s_1,$

IV. $r_2(s_1 - f_1) + s_2(r_1 - f_1) + f_2(1 - r_1 - s_1 - 2f_1) = -f_1.$

Since $x := [r_2, s_2, f_2]^T$, $A =$

$$\begin{bmatrix} 1 - 2r_1 - s_1 - f_1 & -r_1 + f_1 & -r_1 + s_1 \\ -s_1 + f_1 & 1 - r_1 - 2s_1 - f_1 & r_1 - s_1 \\ s_1 - f_1 & r_1 - f_1 & 1 - r_1 - s_1 - 2f_1 \end{bmatrix}$$

and $B = [-r_1, -s_1, -f_1]^T$, we know that the existence of x in the system $Ax = B$ depends on whether $det(A) \in R$ is a unit.

On behalf of the simplicity, let us make the substitutions: $r_1 = r, s_1 = s$ and $f_1 = f$ while we compute $det(A)$. A straightforward computation introduces that

$$\begin{aligned} det(A) = 1 &- 4r + 4r^2 - 4s + 12rs - 8r^2 s \\ &+ 4s^2 - 8rs^2 - 4f + 12rf \\ &- 8r^2 f + 12sf - 16rsf - 8s^2 f \\ &+ 4f^2 - 8rf^2 - 8sf^2 \end{aligned}$$

Considering $r, s, f \in id(R)$ simplifies $det(A)$ as

$$det(A) = 1 - 4rs - 4rf - 4sf - 16rsf$$

which is a unit in R. This actually implies that $r, s, f \in id_C(R)$ because of the assumption $V(RD) = id(RD)$. For the reverse direction of the proof, the reader can notice that the assumption that $det(A) = 1 - 4rs - 4rf - 4sf - 16rsf$ implies that $r, s, f \in id_C(R)$ directly displays that any normalized unit in

$$V(RD) = \{1 - (r + s + f) + rg + sh + fgh$$

$$: r, s, f \in id_C(R)\}$$

is in $id(RD)$. ∎

Theorem 9. Let $D = K_4$ (Klein 4-Group). If $V(RD) = id(RD)$, then $V(RD) =$

$$\{1 - r - s - f + rg + sh + fgh : r, s, f \in R\}$$

implies that $r + f = 0$ and

$$1 - 2(r + s) \in U(R) \Leftrightarrow r + s \in id(R)$$

Proof. Let $G = \langle g \rangle, H = \langle h \rangle, o(g) = o(h) = 2$.

Define a group homomorphism as

$f: G \times H \to \langle \omega, h \rangle$ with $f(g, h) = (\omega, h)$ where $\omega = e^{i\pi}$. Extending linearly f to group rings over the ring R and restricting it to unit groups give $f: R(G \times H) \to R\langle \omega, h \rangle$ and

$$f: V\big(R(G \times H)\big) \to V(R\langle h \rangle) = V(RC_2)$$

respectively. One can easily observe that

$$Ker\ f = \langle 1 + g, h^2 \rangle_R$$

and then,

$$\frac{RD}{\langle 1 + g,\ h^2 \rangle_R} \simeq R\langle h \rangle$$

Thus, when we choose a unit u from the above definition of $V(RD)$ as

$$u = 1 - r - s - f + rg + sh + fgh,$$

we sight that

$$f(u) = 1 - 2r - s - f + (s - f)h \in V(RH)$$

Since $V(RH) = id(RH)$ and the augmentation of $f(u)$ is 1, we conclude that $r + f = 0$ and

$$1 - 2(r + s) \in U(R) \Leftrightarrow r + s \in id(R) \qquad ∎$$

Furthermore, with the help of $\exists v = f(u)^{-1}$ as $v = k + lh$, we can observe that

$$f(u)v = [1 + r - s + (s - r)h][k + lh] = 1$$

if and only if the system of linear equations

$$k(1 - s + r) + l(s - r) = 1$$
$$k(s - r) + l(1 - s + r) = 0$$

has a unique solution pair (r, s) in R. This unique solution depends on

$$(1 - s + r)^2 - (s - r)^2 = (1 - 2s)(1 + 2r)$$

Then, we can deduce that $f(u)v = 1$ if and only if $(1 - 2r)$ and $(1 - 2s)$ are units in R. Due to the fact that $V(RC_2) = id(RC_2)$ if and only if Proposition 1. hold, we can conclude that

$$V(RD) = id(RD) \Leftrightarrow V(RC_2) = id(RC_2)$$

where $D = K_4$ and C_2 is a group of order 2.

Theorem 10. Let two distinct cyclic groups be $G = \langle g: g^3 = 1 \rangle$ and $H = \langle h: h^2 = 1 \rangle$. Then,

$$V(RD) = G \times id(RH)$$

if and only if the followings hold:

i. $1 + 3(r^2 + f^2 + rf + r + f) \in U(RH)$

implies that $(r, f) \in \{(0,0), (0, -1), (-1,0)\}$,

ii. $1 - 2r \in U(R) \Leftrightarrow r \in id(R)$.

Proof. Let us define a group epimorphism

$$\rho_G: D \to H$$
$$(g, h) \mapsto h$$

Extending it to group rings as

$$\rho_G: RD \to RH$$

with $\rho_G\left(\sum_{gh\in D}\alpha_{gh}gh\right)=\sum_{gh\in D}\alpha_{gh}h$, the kernel of ρ_G is obtained as

$$\kappa_G := Ker\,\rho_G = \langle 1-g, 1-g^2\rangle_{RH}$$

One can establish a short exact sequence as

$$\kappa_G \xrightarrow{i} RD \xrightarrow{\rho_G} RH$$

with inclusion i. Moving it to unit groups, we can construct

$$K_G \xrightarrow{i} V(RD) \xrightarrow{\rho_G} V(RH)$$

with $K_G := (1+\kappa_G) \cap V(RD)$. Using the embedding $V(RH) \hookrightarrow V(RD)$, we can write

$$V(RD) = K_G \times V(RH)$$

Since $|H| = 2$, using [3] for $V(RH) = id(RH)$, we obtain the latter condition in phrase of the theorem. Now, we investigate the necessary and sufficient condition to be $K_G = G$. Due to $K_G =$

$\{u = 1 + r(1-g) + f(1-g^2): r, f \in RH\}$, choose an inverse of a unit u in K_G as

$$\exists v = 1 + r'(1-g) + f'(1-g^2)$$

Then, a straightforward computation shows that $uv = 1 + A(1-g) + B(1-g^2) = 1$ where

$$A = r + r' + 2rr' + rf' + fr' - r'f'$$

and

$$B = r' + f' - rr' + rf' + r'f + 2r'f'$$

Then the system

$$r + r' + 2rr' + rf' + fr' - r'f' = 0$$
$$r' + f' - rr' + rf' + r'f + 2r'f' = 0$$

or equivalently

$$\begin{bmatrix} 1+2r+r' & r-r' \\ -r+r' & 1+r+2r' \end{bmatrix}\begin{bmatrix} f \\ f' \end{bmatrix} = \begin{bmatrix} -r \\ -f \end{bmatrix}$$

has a unique solution if and only if

$$1 + 3(r^2 + f^2 + rf + r + f) \in V(RH)$$

and by the fact that u must be a trivial unit,

$$(r,f) \in \{(0,0), (0,-1), (-1,0)\}$$

as required. ∎

By exchanging the types of direct components of $V(RD)$ in the previous theorem, we state and prove the following one as well:

Theorem 11. Let $G = \langle g: g^3 = 1\rangle$ and $H = \langle h: h^2 = 1\rangle$. Then, $V(RD) = id(RG) \times H$ if and only if the following statements are satisfied:

a. $1 + 2r \in U(RG) \Leftrightarrow r = 0$ or $r = -1$,

b. $1 + 3(r^2 + f^2 + rf - r - f) \in U(R)$ implies that $r, f \in id_0(R)$.

Proof. Define a group and a ring epimorphisms as in the previous theorem such as

$$\rho_H: D \to G$$
$$(g,h) \mapsto g$$

and $\rho_H: RD \to RG$ with

$$\rho_H\left(\sum_{gh\in D}\alpha_{gh}gh\right) = \sum_{gh\in D}\alpha_{gh}g$$

respectively. We can view that

$$\kappa_H := Ker\,\rho_H = \langle 1-h\rangle_{RG}$$

Besides, we can construct the short exact sequence $\kappa_H \xrightarrow{i} RD \xrightarrow{\rho_H} RG$. Restricting the last sequence to unit groups, we deduce that the following short exact sequence can be established:

$$K_H \xrightarrow{i} V(RD) \xrightarrow{\rho_H} V(RG)$$

where $K_H := (1+\kappa_H) \cap V(RD)$. Because of the embedding $V(RG) \hookrightarrow V(RD)$, the last sequence splits as $V(RD) = K_H \times V(RG)$. On account of $|G| = 3$, we already know that $V(RG) = id(RG)$ if and only if

$$1 + 3(r^2 + f^2 + rf - r - f) \in U(R)$$

implies that $r, f \in id_0(R)$ [3]. On the other hand, since $|H| = 2$, one can observe that

$$K_H = \{1 + r(1 - h) : r \in RG\} = H$$

if and only if $r = 0$ or $r = -1$. ∎

3. DISCUSSIONS AND SUGGESTIONS

To sum up, we have attained some necessary and sufficient conditions for normalized unit group $V(RD)$ to be idempotent unit group $id(RD)$ or direct product of trivial unit group and idempotent unit group as $id(RG) \times H$ (or $G \times id(RH)$) in this paper. Consequently, as originality of the paper, we can say that the paper has been both extended some results in [3], [4] and defined novel types of units which are combined with both idempotent units and trivial units. As an open problem and future work, necessary and sufficient conditions for

$$V(RD) = id(RG) \times id(RH)$$

may be studied for Abelian groups.

Note: This paper, has been generated from the Ph. D. thesis of the author.

Research and Publication Ethics
The author declares that this study complies with Research and Publication Ethics.

Ethics Committee Approval
This paper does not require any ethics committee permission or special permission.

Conflicts of Interests
No potential conflict of interest was reported by the author.

Author's Contributions
ÖK performed both the theoretical results and applicational calculations with final version of the paper.

REFERENCES

[1] P. Danchev, "Trivial units in commutative group algebras," Extr. Math., vol. 23, pp. 49-60, 2008.

[2] P. Danchev, "Trivial units in abelian group algebras," Extr. Math., vol. 24, pp. 47-53, 2009.

[3] P. Danchev, "Idempotent units in commutative group rings," Kochi J. Math., vol. 4, pp. 61-68, 2009.

[4] P. Danchev, "Idempotent units of commutative group rings," Commun. Algebra, vol. 38, pp. 4649-4654, 2010.

[5] P. Danchev, "On some idempotent torsion decompositions of normed units in commutative group rings," J. Calcutta Math. Soc., vol. 6, pp. 31-34, 2010.

[6] P. Danchev, "Idempotent-torsion normalized units in abelian group rings," Bull Calcutta Math. Soc., vol. 1(103), pp. 31-34, 2011.

[7] G. Karpilovsky, "On units in commutative group rings," Arch. Math. (Basel), vol. 38, pp. 420–422, 1982.

[8] G. Karpilovsky, "On finite generation of unit groups of commutative group rings," Arch. Math. (Basel), vol. 40, pp. 503–508, 1983.

[9] G. Karpilovsky, "Unit groups of group rings," Harlow: Longman Sci. and Techn., 1989.

[10] G. Karpilovsky, "Units of commutative group algebras," Expo. Math., vol. 8, pp. 247-287, 1990.

[11] W. May, "Group algebras over finitely generated rings," J. Algebra, vol. 39 pp. 483–511, 1976.

[12] C. Polcino Milies and S. K. Sehgal, "An introduction to group rings," Kluwer, North-Holland, Amsterdam, 2002.

[13] S. K. Sehgal, "Topics in group rings," Marcel Dekker, New York, 1978.