C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

VOLUME: 61 **Number: 1** **YEAR: 2019**

Faculy of Sciences, Ankara University
06100 Beşevler, Ankara-Turkey
ISSN 1303-6009 E-ISSN 2618-6462

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 61 **Number : 1** **Year :2019**

This Journal is published two issues in a year by the Faculty of Sciences, University of Ankara. Articles and any other material published in this journal represent the opinions of the author(s) and should not be construed to reflect the opinions of the Editor(s) and the Publisher(s).

Correspondence Address:
COMMUNICATIONS
EDITORIAL OFFICE
Ankara University, Faculty of Sciences,
06100 Tandoğan, ANKARA – TURKEY
Tel: (90) 312-212 67 20 Fax: (90) 312-223 23 95
e-mail: commun@science.ankara.edu.tr
http://communications.science.ankara.edu.tr/index.php?series=A2A3

Print:
Ankara University Press
İncitaş Sokak No:10 06510 Beşevler
ANKARA – TURKEY

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 61 **Number : 1** **Year :2019**

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 1-11 (2019)
DOI: 10.33769/aupse.533815
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

PRIVACY-PRESERVING K-NEAREST NEIGHBOUR INTERPOLATION METHOD IN AN OUTSOURCED ENVIRONMENT

Muhammad Rifthy KALIDEEN, Murat OSMANOGLU and Bulent TUGRUL

ABSTRACT. One of the most emerging computer technologies of this decade is cloud computing that allows data owners to outsource their storage and computing requirements. It enables data owners to avoid the costs of building and maintaining a private storage infrastructure. While outsourcing data to cloud promises significant benefits, it possesses substantial security and privacy concerns, especially when data stored in the cloud is sensitive and confidential, like a business plan. Encrypting the data before outsourcing can ensure privacy. However, it will be very difficult to process the cipher text created by the traditional encryption method. Considering this fact, we propose an efficient protocol that allows a query owner to retrieve the interpolation of the top k records from two different databases that are closest to a query point. Note that the databases are stored in two different cloud service providers in encrypted form. We also show that the proposed protocol ensures the privacy and the security of the data and the query point.

1. INTRODUCTION

Recent developments in networking technology and the increasing needs for computing resources have motivated many entities to outsource their storage requirements and computing needs. This new computing model is recognized as cloud computing that enables users to outsource their database and the processing functionalities to a cloud [1]. Moreover, they provide access mechanisms for the database, and allow users to make queries on the stored data. Outsourcing a database possesses substantial advantages such as avoidance of the costs of building and maintaining a private storage, and universal access to data without location dependency. On the other hand, clouds are untrustworthy in the context of preserving confidentiality and privacy of the data. Because, it is challenging for clouds to ensure the confidentiality of the stored data, i.e. a break in the cloud where

Received by the editors: November 21, 2018; Accepted: December 13, 2018.
Key word and phrases: Spatial interpolation, k-NN, privacy – preserving, homomorphic encryption.

the data is stored may cause all the data to be visible to the attackers [2-5]. Besides, an inside attacker from the cloud can reveal the data and some other related information such as the queries made for the data to a third party [6, 7]. Thus, public cloud can be considered as semi-honest (i.e. honest-but-curious) because of the presence of such attacker.

To overcome the privacy and security issues discussed above, the client should encrypt the data before outsourcing it to the cloud. The traditional encryption systems such as Advanced Encryption Standard (AES) [8], can be deployed within this aim. However, they do not allow processing the data directly. Particularly, if the client wants to do any operation on the data, he needs to download all the data from the cloud first, and to decrypt the data with its private key to be able to make the operations. Such limitation turns the cloud into a simple data storage service. On the other hand, a homomorphic encryption scheme that allows the data to be processed in the encrypted form is well suited for this type of scenarios. Within this scope, a large number of secure schemes [9-12] have been proposed that employ an homomorphic encryption to support processing on the encrypted data.

In this paper, we study the problem of k-Nearest Neighbour (k-NN) query processing over encrypted spatial data. We consider the following scenario: two data owners DO_1 and DO_2 holding two different spatial data sets D_1 and D_2 outsource their data to two different cloud service providers CSP_1 and CSP_2, respectively. Note that each data set consists of two-dimensional location vectors, and each data owner encrypts his data before outsourcing it. Assume there is a query owner that wants to extract the interpolation of the top k vectors of both data sets that are closest to a query point he provides. In the scenario, the query owner does not want to reveal the query point to the service providers, and the data owners do not want to reveal their data to the query owner or the service providers. Within this direction, we propose an efficient protocol that allows the query owner to retrieve the interpolation of the closest k records from both databases for his query point. The proposed protocol protects the confidentiality of data and the privacy of the query.

Paper overview. We discuss the related works at Section 2. In Section 3, we give the basic definitions. We propose the protocol in Section 4. In addition, we analyze the security and the complexity of our protocol in Section 5.

2. Related Work

k-NN problems have gained a lot of attention due to the wide usage of social media in recent years, and many solutions have been proposed within this direction [13]. These solutions are being used in most of the applications such as similarity check, data mining, and pattern recognition. The studies on k-NN problems can be divided into two categories: according to whether the data is encrypted or not; centralized and distributed k-NN problems.

In the former one, the database and its functionalities were outsourced to an untrusted cloud service provider (CSP) so that the cloud service provider manages the data on behalf of the data owner, and allows an authorized user to access the data. However, this may cause some security problems associated with the privacy and the confidentiality of the data. To overcome this obstacle, the data owner can use some cryptographic techniques to protect the data before outsourcing it to the cloud.

One of the techniques used to protect the confidentiality of the data is to employ a traditional encryption scheme such as AES. However, such encryption scheme does not allow the encrypted data to be processed. On the other hand, some researchers proposed several techniques to process range queries [14, 15] and aggregation queries [16, 17] on the encrypted data. However, the proposed techniques possess some limitations on the security for the k-NN queries.

In addition to these studies, different methods were proposed to solve k-NN problems for spatial data in recent years [18, 19]. Wong et al. [10] presented an encryption technique, named as Asymmetric Scalar – Product – Preserving Encryption (ASPE), in which the data and the queries are encrypted before the outsourcing. However, the encryption algorithms used to mask the data and the queries are slightly different, and the query owner knows the decryption key. On the other hand, Zhu et al. [20] proposed a better solution, named as novel SkNN, where the decryption key is not revealed to the query owner. Nevertheless, the data owner and the query owner jointly create the encryption of the queries. Other than this methods, Hu et al. [18] presented a solution for k-NN problem that utilizes a homomorphic encryption scheme. This scheme supports addition, subtraction and multiplication over encrypted data [21]. However, this solution is vulnerable to the chosen-plaintext attacks. Moreover, all the above techniques leak the access pattern of the data to the cloud servers.

In the distributed k-NN problems, data that is partitioned either vertically or horizontally is stored among the independent parties (servers). Most of the methods in this category rely on Secure Multi-Party Computation (SMC) that allows different parties to jointly compute a function over their inputs without revealing their inputs to the parties. Similar to the centralized k-NN, a number of different solutions have been proposed within this direction. Shaneck et al. [22] suggest a protocol for the horizontally distributed data, and Qi et al. [11] proposed a solution for the linear computation and communication. Besides, Vaidya et al. [23] presented an efficient method for the vertically partitioned data by. Also, a technique, called Private Information Retrieval (PIR), was proposed by Ghinita et al. [24] for the location-based services, that only protects the query privacy, but not the data confidentiality.

Note that no methods discussed above are suitable for our purpose since data are partitioned either vertically or horizontally, but stored as plaintext in the server. However, in our proposed solution, the data and the queries are in the encrypted format, and the data are stored in the server as cipher-text.

3. BACKGROUND

In this section, we give some definitions that will be used as building blocks in our proposed solution.

3.1. k –NN Algorithm

k-Nearest Neighbor algorithm is one of the most common algorithms used in the similarity matching due to its easiness, efficiency, and applicability to the big amount of data. The algorithms can be implemented for the data in which the distance between the points and the query point can be calculated using distance metrics such as Euclidean distance, Manhattan distance. Briefly, the algorithm works as follows: (i) it first calculates the distance values between the data points and the query point using the corresponding distance metric, (ii) it then finds k indexes of the k nearest distances among the distance values, and (iii) it finally outputs the k points that correspond to the k indexes found in the previous step.

3.2. Homomorphic Encryption (HE)

Homomorphic encryption is an encryption system which allows performing mathematical operations on the cipher texts. This homomorphic feature makes the scheme a functional and strong tool for cloud computing. There are several

homomorphic encryption schemes presented in the literature such as Goldwasser-Micali encryption scheme, Paillier system, ElGamal cryptosystem, Boneh-Goh-Nissim crypto scheme and so on [25].

In this paper, we utilize the Paillier encryption scheme as the homomorphic encryption scheme. Briefly, Pailler encryption is an additive homomorphic encryption scheme that allows processing addition and multiplication on the ciphertext as follows:

Addition function - $\mathrm{E_{pk}(a + b)} = \mathrm{E_{pk}(a)} * \mathrm{E_{pk}(b)} \bmod \mathrm{N}^2$ (1)

Multiplication function - $\mathrm{E_{pk}(a * b)} = \mathrm{E_{pk}(a)}^{b} \bmod \mathrm{N}^2$ (2)

where $\mathrm{a, b} \in Z_N$, pk is the public key, and N is the product of two large primes.

3.3. Basic Security Primitives

We here also present some basic security protocols that will be used in constructing our main protocol.

3.3.1 Secure Multiplication (SM) Protocol

In the protocol, a data owner (DO) with the secret key and a cloud service provider (CSP) with the inputs $\mathrm{E_{pk}(a)}$ and $\mathrm{E_{pk}(b)}$ securely compute the multiplication $\mathrm{E_{pk}(a * b)}$. During the computation, the plaintexts a and b are not revealed to any party, and the output is only revealed to CSP. Briefly, CSP first computes $\mathrm{E_{pk}}(a + r_1) = \mathrm{E_{pk}}(a) * \mathrm{E_{pk}}(r_1)$ and $\mathrm{E_{pk}}(b + r_2) = \mathrm{E_{pk}}(b) * \mathrm{E_{pk}}(r_2)$ for randomly chosen $r_1, r_2 \in Z_N$, then sends them to DO. Upon receiving, DO decrypts them as $a + r_1$ and $b + r_2$ using the corresponding secret key. DO then multiplies the decryptions as $h = (a + r_1) * (b + r_2)\ mod\ N$, encrpts h and sends it to CSP. After getting the encryption $\mathrm{E_{pk}}(h)$, CSP removes the randomness using the equation

$$a * b = (a + r_1) * (b + r_2) - a * r_1 - b * r_2 - r_1 * r_2\,, \quad (3)$$

and gets $\mathrm{E_{pk}}(a * b)$.

3.3.2 Secure Squared Euclidean Distance (SSED) Protocol [26]

In this protocol, a data owner (DO) with the secret key and a cloud service provider (CSP) with the encrypted vectors $\mathrm{E_{pk}(a)}$ and $\mathrm{E_{pk}(b)}$ securely compute $\mathrm{E_{pk}}(|a-b|^2)$, the encryption of squared Euclidean distance between m dimensional vectors a and b. During the computation, the vectors a and b are not revealed to any party, and the output is only revealed to CSP. Briefly, CSP first computes $\mathrm{E_{pk}}(a_i - b_i)$ using the homomorphic properties of the underlying encryption scheme where $1 \leq i \leq m$. Then CSP securely computes $\mathrm{E_{pk}}((a_i - b_i)^2)$ with DO using the SM protocol for each i. Finally, CSP locally adds all $\mathrm{E_{pk}}((a_i - b_i)^2)$ using the homomorphic property in order to get the encryption of squared Euclidean distance $\mathrm{E_{pk}}(|a-b|^2)$.

4. Proposed Solution

In this section, we will give an efficient k-NN protocol that allows a query owner (QO) to extract the interpolation of the top k records, which are closest to a query point provided by QO. These k records are selected from two spatial databases stored in two different Cloud Service Providers. There are five parties involved in our proposed scheme:

- Data Owner #1 (DO_1)
- Data Owner #2 (DO_2)
- Cloud Service Provider #1 (CSP_1)
- Cloud Service Provider #2 (CSP_2)
- Query Owner (QO)

We assume that each data owner DO_u holds a spatial database D_u that consists of n records $t_1^{(u)}, \dots, t_n^{(u)}$ such that each record is a two-dimensional vector as $t_i^{(u)} = (t_{i,1}^{(u)}, t_{i,2}^{(u)})$ obtained from a specific region. Our protocol is formalized as follows:

I. Each DO_u first encrypts its database D_u using his public key pk_u, as $\left\{E_{pk_u}\left(t_1^{(u)}\right), \dots, E_{pk_u}(t_n^{(u)})\right\}$ and sends the encryption of the database to the corresponding server CSP_u.

II. When QO wants to extract the interpolation of the top k records, closest to a query point $Q = (q_1, q_2)$, she first creates two encryptions of her query point $E_{pk_1}(Q)$ and $E_{pk_2}(Q)$ under the public keys pk_1 and pk_2 of

DO_1 and DO_2, respectively. QO then sends each encryption $E_{pk_u}(Q)$ to the corresponding server CSP_u.

III. Upon receiving $E_{pk_u}(Q)$, each CSP_u executes the SSED protocol with the corresponding data owner D_u on the inputs $\left(E_{pk_u}(Q), E_{pk_u}\left(t_i^{(u)}\right)\right)$ where $1 \leq i \leq n$. As the output of the protocol, each CSP_u receives the encryption of the squared Euclidean distance between the query point Q and each record $t_i^{(u)}$ as $E_{pk_u}\left(d_i^{(u)}\right)$, and sends these encryptions to the corresponding DO_u.

IV. Upon getting the encryptions, each DO_u first decrypts them as $d_i^{(u)}$, then sorts these decryptions. At this moment, each DO_u have n sorted distance values. DO_2 sends the first k values $\left\{d_1'^{(2)}, \dots, d_k'^{(2)}\right\}$ of his sorted sequence to DO_1.

V. After that, DO_1 adds the first k values $\left\{d_1'^{(1)}, \dots, d_k'^{(1)}\right\}$ of his sorted sequence to the k values he received from DO_2. DO_1 then sorts all 2k values. Based on the first k values in this sorting, DO_1 creates the index sets S_1 and S_2 such that S_1 includes the indexes of p values from the first k that are chosen among $\left\{d_1'^{(1)}, \dots, d_k'^{(1)}\right\}$ and S_2 includes the indexes of q values from the first k that are chosen among $\left\{d_1'^{(2)}, \dots, d_k'^{(2)}\right\}$ where $p + q = k$. DO_1 sends each S_u to the corresponding server CSP_u.

VI. Upon receiving the index set $S_1 = \{i_1, \dots, i_p\}$, CSP_1 computes the sum of the n records of the corresponding indexes as $E_{pk_1}\left(t_{i_1} + \cdots + t_{i_p}\right)$ using the homomorphic properties of the encryption scheme. In a similar way, CSP_2 computes the sum of the m records of the corresponding indexes in $S_2 = \{j_1, \dots, j_q\}$ as $E_{pk_2}\left(t_{j_1} + \cdots + t_q\right)$. Then, each CSP_u masks the sum with a random vector $r_u \in Z_N \times Z_N$ as $E_{pk_u}(\gamma_u) = E_{pk_u}(T_u + r_u)$ using the homomorphic property of the encryption scheme, and sends the result to the corresponding DO_u and the random vector r_u to QO.

VII. After getting $E_{pk_u}(\gamma_u)$, each DO_u decrypts it, and sends the decryption γ_u to QO.

VIII. Upon receiving γ_1 and γ_2, QO removes the randomness r_1 and r_2, and gets the sum of the closest records as $T_1 = \gamma_1 - r_1$ and $T_2 = \gamma_2 - r_2$. Finally, QO calculates the interpolation of the closest k records of both databases to his query point Q as $(T_1 + T_1)/k$.

5. Analysis of the Proposed Solution

In this section, we analyse the computational complexity of the proposed protocol, and discuss security issues.

5.1 Complexity Analysis

At the beginning of the protocol, each data owner creates the encryption of his data that costs only O(n) encryptions. At the second step, the query owner only makes O(1) encryptions to create the encryptions of his query point. Then, each CSP_u executes O(n) instantiations of SSED protocols with the corresponding DO_u. If we implement the SSED protocol proposed by [25] which involves O(m) encryptions and O(m) exponentiations for the m-dimensional records. Since we consider the data that contains two-dimensional vectors, O(n) instantiations of SSED cost only O(n) operations for each distance set. At the fourth step, the data owners sort the n distance vectors, that costs O(nlogn) operations with an efficient sorting algorithm. Later, DO_1 collects the first k values from each distance set, and sorts them. This part costs O(klogk) operations for each. At the sixth step, CSP_1 makes O(p) multiplications, and CSP_2 makes O(q) multiplications. Since $p + q = k$, O(p) and O(q) are bounded by O(k). At the final step, QO makes just a constant number of operations to calculate the interpolation of the top k records that are closest to the query point. Therefore, the computational complexity of our protocol is bounded by O(nlogn) operations.

5.2 Security Analysis

First of all, the query owner encrypts the query point using Paillier encryption scheme before giving it to the cloud service providers. Since the Paillier encryption scheme is semantically secure, the query point is not revealed to any DO or CSP. Thus, our protocol protects the privacy of the query point. Similarly, each data owner encrypts his database using Paillier cryptosystem before outsourcing it to the cloud. Since the Paillier encryption scheme is semantically secure, no data record is revealed to any CSP and QO. Thus, our protocol preserves the confidentiality of the data.

Besides, all communications between the parties are made in the encrypted form except the one in which the DO_2 sends k distance vectors to DO_1 and the other in which the data owners DO_1 and DO_2 send the decryptions γ_1 and γ_2 to QO,

respectively. In the former one, DO_2 does not send the records directly. Instead, he sends the distance vectors that can be considered as the randomization of the records. In the latter one, each DO_u sends the decryption γ_u which is $\gamma_u = T_u + r_u$ for the random r_u. Thus, no information about the data are leaked to a third party during the protocol.

6. Conclusions

Interpolation plays a critical role in engineering problems. Obtaining measurement values across an entire region is both difficult and costly. When a measurement value is required for an unmeasured location, interpolation methods are employed to produce a prediction value. There is a variety of spatial interpolation methods. Due to its simplicity in geology and hydrology, k-NN is highly preferred as a spatial interpolation method. The accuracy and robustness of prediction models depend on the size of the data. Companies may collect data from the same region. However, no one may want to publicize its data. Therefore, they may prefer to use the methods that protects all parties' private data from each other. In this study, we proposed to execute the k-NN spatial interpolation method on outsourced data of two party without jeopardizing their private data. Our solution also protects the prediction coordinate from the data owners and cloud servers, which can be very important from the query owner's point of view. We also analysed our solution in terms of complexity and security.

References

[1] M. Armbrust, et al., A view of cloud computing, *Communications of the ACM*, 53/4 (2010) 50-58.

[2] T. Ermakova, B. Fabian, and R. Zarnekow, Security and privacy system requirements for adopting cloud computing in healthcare data sharing scenarios, *Proceedings of the Nineteenth Americas Conference on Information Systems* (2013).

[3] K. Hashizume, et al., An analysis of security issues for cloud computing, *Journal of Internet Services and Applications*, 4/1 (2013).

[4] J.J. Rodrigues, et al., Analysis of the security and privacy requirements of cloud-based electronic health records systems, *Journal of Medical Internet Research*, 15/8 (2013).

[5] M.D. Ryan, Cloud computing security: *The scientific challenge, and a survey of solutions, Journal of Systems and Software*, 86/9 (2013) 2263-2268.

[6] K. Ren, C. Wang, and Q. Wang, Security challenges for the public cloud, *IEEE Internet Computing,* 16/1 (2012) 69-73.

[7] C. Shahabi, et al., Privacy-preserving inference of social relationships from location data: a vision paper, *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 9 (2015) 1-4.

[8] J. Daemen, and V. Rijmen, The Design of Rijndael: AES - The Advanced Encryption Standard, Springer Berlin Heidelberg, (2013).
[9] D.X. Song, D. Wagner, and A. Perrig, Practical techniques for searches on encrypted data, *Proceeding 2000 IEEE Symposium on Security and Privacy,* (2000) 44-55.
[10] W.K. Wong, et al., Secure kNN computation on encrypted databases, *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data,* (2009) 139-152.
[11] Y. Qi and M.J. Atallah, Efficient privacy-preserving k-nearest neighbor search, *The 28th International Conference on Distributed Computing Systems.* (2008) 311-319.
[12] M.R. Kalideen, and B. Tugrul, Outsourcing of Secure k-Nearest Neighbours Interpolation Method, *International Journal of Advanced Computer Science and Applications,* 9/4 (2018) 319-323.
[13] J. Peng, K. K. R. Choo, and H. Ashman, Bit-level n-gram based forensic authorship analysis on social media: Identifying individuals from linguistic profiles, *Journal of Network and Computer Applications,* 70 (2016) 171-182.
[14] B. Hore, et al., Secure multidimensional range queries over outsourced data, *The VLDB Journal,* 21/3 (2012) 333-358.
[15] B. Hore, S. Mehrotra, and G. Tsudik, A privacy-preserving index for range queries, *Proceedings of the Thirtieth International Conference on Very Large Data Bases* - Volume 30, (2004) 720-731.
[16] H. Hacıgümüş, B. Iyer, and S. Mehrotra, Efficient execution of aggregation queries over encrypted relational databases, *International Conference on Database Systems for Advanced Applications,* (2004) 125-136.
[17] E. Mykletun and G. Tsudik, Aggregation queries in the database-as-a-service model, *IFIP Annual Conference on Data and Applications Security and Privacy*, (2006) 89-103.
[18] H. Hu, et al., Processing private queries over untrusted data cloud through privacy homomorphism, *IEEE 27th International Conference on Data Engineering (ICDE)*, (2011) 601-612.
[19] B. Yao, F. Li, and X. Xiao, Secure nearest neighbor revisited, *IEEE 29th International Conference on Data Engineering (ICDE).* (2013) 733-744.
[20] Y. Zhu, R. Xu, and T. Takagi, Secure k-NN computation on encrypted cloud data without sharing key with query users, *Proceedings of the 2013 International Workshop on Security in Cloud Computing.* (2013) 55-60.
[21] J. Domingo-Ferrer, A provably secure additive and multiplicative privacy homomorphism, *International Conference on Information Security.* (2002) 471-483.
[22] M. Shaneck, Y. Kim, and V. Kumar, *Privacy preserving nearest neighbor search, Machine Learning in Cyber Trust.* (2009) 247-276.
[23] J. Vaidya and C. Clifton, Privacy-preserving top-k queries, *21st International Conference on Data Engineering (ICDE'05),* (2005) 545-546.
[24] G. Ghinita, et al., Private Queries in Location Based Services: Anonymizers Are Not Necessary, *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data,* (2008) 121-132.

[25] X. Yi, R. Paulet, and E. Bertino, Homomorphic Encryption and Applications, (2014) Springer.
[26] Y. Elmehdwi, B.K. Samanthula, and W. Jiang, Secure k-nearest neighbor query over encrypted data in outsourced environments, *2014 IEEE 30th International Conference on Data Engineering (ICDE)*, (2014) 664-675.

Current Address: Muhammad Rifthy KALIDEEN, Department of Islamic Studies, South Eastern University of Sri Lanka, Oluvil Park, Oluvil, Sri Lanka, 32360
E-mail: kmr@seu.ac.lk
ORCID: https://orcid.org/0000-0001-5790-1166

Current Address: Murat OSMANOGLU, Department of Computer Engineering, Ankara University, Turkey, 06100
E-mail: mosmanoglu@ankara.edu.tr
ORCID: https://orcid.org/0000-0001-8693-141X

Current Address: Bulent TUGRUL, Department of Computer Engineering, Ankara University, Turkey, 06100
E-mail: btugrul@eng.ankara.edu.tr
ORCID: https://orcid.org/0000-0003-4719-4298

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 12-23 (2019)
DOI: 10.33769/aupse.469183
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

A COMPARATIVE STUDY: VOLTAGE MULTIPLIERS FOR RF ENERGY HARVESTING SYSTEM

Filiz SARI, Yunus UZUN

ABSTRACT. Voltage multipliers are widely used for energy harvesting processes to convert the received AC signal to DC signal, also enhanced the low level received signal. In this study, Villard, Dickson and Greinacher type voltage multipliers are analyzed without impedance matching and substrate materials to decide the effective voltage multiplier type depending on the inputs of the harvester. So, load resistance, input power and input frequencies' effects are analyzed and compared with each other. Agilent Advanced Design System (ADS) is used for simulations. HSMS 2852 Schottky diode and capacitors are used for these voltage multipliers. Results show that, determining load resistance is important for evaluating high efficiency, e.g. efficiency differences are reached 33% between 2kΩ and 20 kΩ for Dickson voltage multiplier at 100 MHz input frequency. Furthermore, the best efficiency is obtained by Greinacher voltage multiplier for low input frequencies which is lower than 1 GHz but there are no significant differences are observed for high frequencies. This study shows that load resistance, input frequency and input power are important parameters for voltage multiplier selection and Greinacher voltage multiplier is the best choice to obtain high efficiency for low frequency application of RF harvesting.

1. INTRODUCTION

There has been a significant increase in the number of low power consuming devices with the development of technology. These devices are widely used especially for measurement, data recording, transfer, and so on, and they are generally supplied with chemical batteries. Energy harvesting from sources such as solar, thermal, microwave or Radio Frequency (RF) gives an opportunity to supply energy for low power devices instead of using short-lived and replacement required batteries. RF energy harvesting is one of the most promising type of energy harvesting because

Received by the editors: October 10, 2018; Accepted: February 23, 2019.
Key word and phrases: RF energy harvesting, voltage multiplier types, efficiency.

RF signal is easily and continually accessible. RF frequency bands covers frequencies from 10 kHz to 30 GHz which vary from very-low frequency (VLF) to super-high frequency (SHF) [1], [2]. Cellular base stations, RFID, Wi-Fi access points, television and radio broadcast stations can be used as a source for RF energy harvesting system. The amount of power which can be obtained from RF sources is very low. Therefore, it is very important that the obtained power must be converted to usable with the least loss. Some auxiliary circuits are used for this process.

RF energy harvesting system is shown in Figure 1 which consists of antenna, matching circuit, voltage multiplier circuit for AC to DC conversion and DC load circuit. Antenna is used for capturing RF signal from ambient air and designed for the frequency to be used. Single-band and multi-band antennas and their efficiencies were investigated in [3]–[7]. Matched circuit is used for compensating the impedance of antenna and the load, thus maximum power can be transferred [8], [9]. Voltage multiplier circuit, consisting of diodes and capacitors, is the indispensable part of energy harvesting because at this stage captured signal is enhanced and converted to DC voltage with minimum losses [10]–[19]. Actually, Voltage Multiplier (VM) devices are diode rectifier circuit, convert AC to DC, which can produce an output voltage greater than of the input voltage.

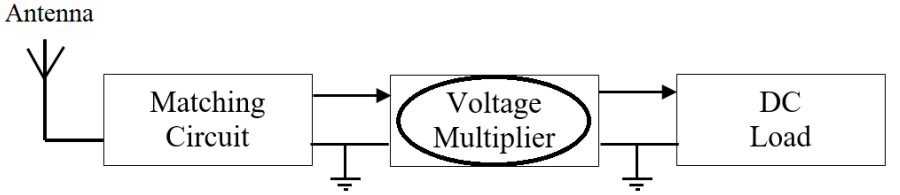


Figure 1. Basic RF energy harvesting structure

Different type and n-stage voltage multiplier circuits were investigated in literature. Villard voltage doubler were used for three band (2.1 GHz, 2.4 – 2.8 GHz, 3.3 – 3.8 GHz) in [7], results showed that maximum efficiency was obtained at the lower frequency band. Seven stage Villard voltage multiplier circuit was analyzed for 900 MHz in [20] and voltage level of 5.0 V was achieved in the measurement. Different stage Villard and Dickson voltage multiplier circuits were compared in [10] for frequency at 915 MHz and input power ranges between -20 dBm up to 20 dBm and approximately 100% improvement over other existing designs in the power range of 20 to 7 dBm was achieved. In [8], six stage Cockcroft –Walton (CWVM) voltage multiplier was used for 88 – 108 MHz frequency range to analyze load resistance and capacitance effect. In [21] Dickson and Villard type voltage multipliers were

compared depending on their efficiency for 868 MHz, results were showed that Dickson gave high efficiency and increasing number of stage provided high output power.

This paper is only focused on voltage multiplier circuits for RF energy harvester system, effect of impedance matching and substrates are ignored. The main goal of this paper is to understand the behavior of 2 stage Villard, Dickson and Greinacher voltage multiplier characteristics for different RF input frequencies, input powers and load resistances. The voltage multipliers are compared with each other and advantage and disadvantages are observed. Agilent Advanced Design System (ADS) program is used for these simulations.

2. Voltage Multiplier Circuits

Voltage multiplier circuits are used to convert the randomly changing RF signal to DC signal and enhance it according to the number of n- stage to transfer it to the load. Received power of RF signal is described by Friis transmission equation in (1)

$$P_R = P_T G_T G_R \left(\frac{c}{4\pi r f}\right)^2 \tag{1}$$

where P_R is the received power in dBm, P_T is the transmitted power in dBm, G_T is the transmit antenna gain, G_R is the receive antenna gain, c is the speed of light, f is the operating frequency, and r is the distance between the transmitter and the receiver antenna. In [7], despite different operating frequencies and transmit powers, almost same received power was obtained at the same distance for three systems. In [2], maximum RF power generated from signal generator is 20 dBm and obtained received power was very low level such as -20 dBm which generates 31.62 mV signal level for typical 50 Ω antenna. So in this study, received power, P_R, is considered fix value and assumed to change between the ranges of -20 to +20 dBm.

Received power produces an DC power at the output of the voltage multiplier. The ratio of the incident power of the RF signal to the DC power at the output of the voltage multiplier is defined as power efficiency which is given in (2)

$$\eta = \frac{P_{out}}{P_R} \tag{2}$$

η is the efficiency, P_{out} is the output power of the circuit and P_R is the received RF power defined in (1).

Especially for low received power level, enhancing stage is necessary to obtain enough power for low power consuming devices so voltage multiplier usage is required for RF energy harvesting. Voltage multipliers, consist of diodes and capacitors, are investigated because of simplicity of application and small size. At the voltage multiplier, diodes act as switches and low forward voltage is required to operate so HSMS- 2852 Schottky diodes are used in this work also these diodes have advantages of low junction capacitance and fast switching speed.

2. 1. Villard Voltage Multiplier

Villard voltage multiplier, sometimes referred to as Cockcroft-Walton voltage multipliers [10], [18], [22], is shown for single and n-stage in Figure 2(a) and (b) respectively[14].

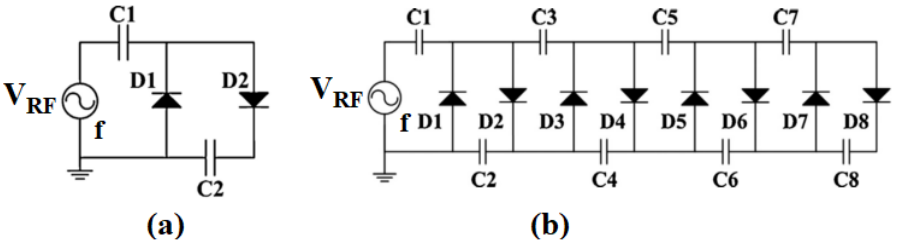


Figure 2. Basic circuit of Villard Voltage Multiplier

As shown in Figure 2 (a), the circuit consist of two diodes and two capacitors for each stage. The working principle is depending on the sign of the input voltage, thus emulating diodes behavior: specifically, considering if V_{RF} is negative, D1 is operating and D2 is not functional, while if V_{RF} is positive the two diodes will reverse their configuration [14]. The output of the n-stage VM is given in (3)

$$V_{DC} = \frac{nV_0}{nR_0+R_L} \tag{3}$$

In (3) V_0 is open circuit output voltage of a single stage and R_0 is internal resistance of the single stage and R_L is load resistance.

2.2. Dickson Voltage Multiplier

Dickson VM is given in Figure 3, the parallel configuration of capacitors in each stage reduces the circuit impedance, and hence makes the matching task simpler [4], [10], [15] when compared with the Villard VM.

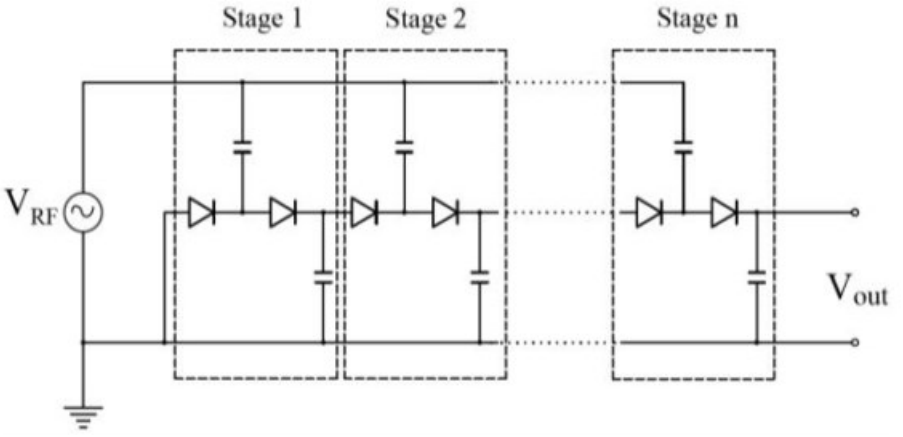


Figure 3. Basic n-stage circuit of Dickson Voltage Multiplier

Output voltage of n stage Dickson voltage multiplier is given in (4) [25].

$$V_{DC} = n(V_m - V_T) \tag{4}$$

where n is the number of stages, V_m is the peak amplitude of the input voltage and VT is the forward conduction voltage of the diodes.

2.3. Greinacher Voltage Multiplier

Greinacher VM circuit is shown in Figure 4. C1 and D1 shift the input voltage up at node B to be rectified by D2, C3 and D3 shift the voltage down at node C to be rectified by D4 and C4. After reaching balance, the circuit provides a constant output current and voltage to the load [23], [24]. Output voltage of Greinacher voltage multiplier is given in (5) [26]

$$V_{DC} = 2nV_m \tag{5}$$

where n is the number of stages and V_m is the peak amplitude of the input voltage.

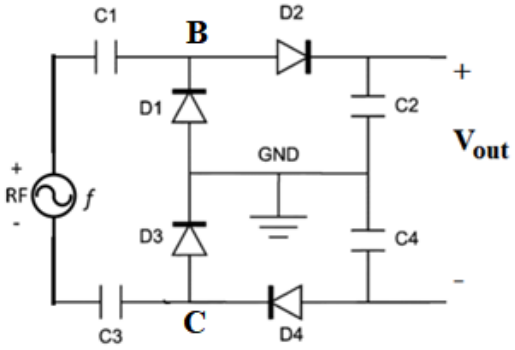


Figure 4. Basic circuit of Greinacher Voltage Multiplier

3. Simulation Results

ADS software which is an important software for RF applications is used for Villard, Dickson and Greinacher VM circuits' simulations, ADS application of Villard VM is given in Figure 5. First stage, parameters of voltage multipliers are determined, HSMS 2852 Schottky diodes and 90 pF capacitors are selected for simulations and then load resistor, input frequency and input power effects are analyzed. Smith Chart diagram of the VM multiplier circuits are given in Figure 6.

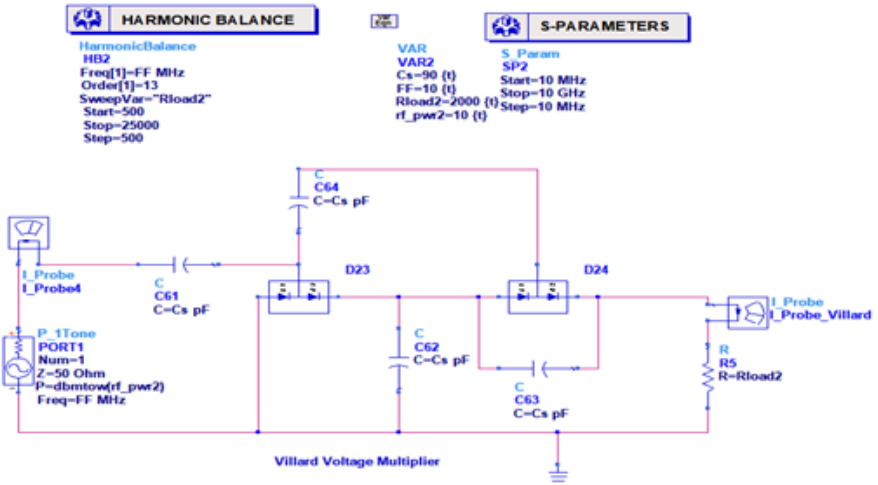


Figure 5. Villard VM circuits from ADS software

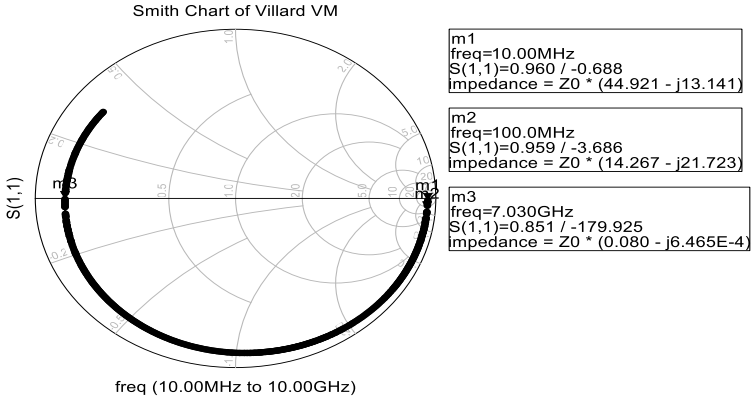


Figure 6. Smith Chart diagram for Villard VM circuit

Impedance of VM circuit changes depend on the input frequency, Figure 6, and that means these circuits are required impedance matching to increase the efficiency. But in this study, voltage multiplier circuits are analyzed depending on their output voltage, effect of input power and input frequencies without impedance matching. Figure 7 acquires output voltage levels for 10 dBm input power at 10 MHz and 100 MHz input frequencies.

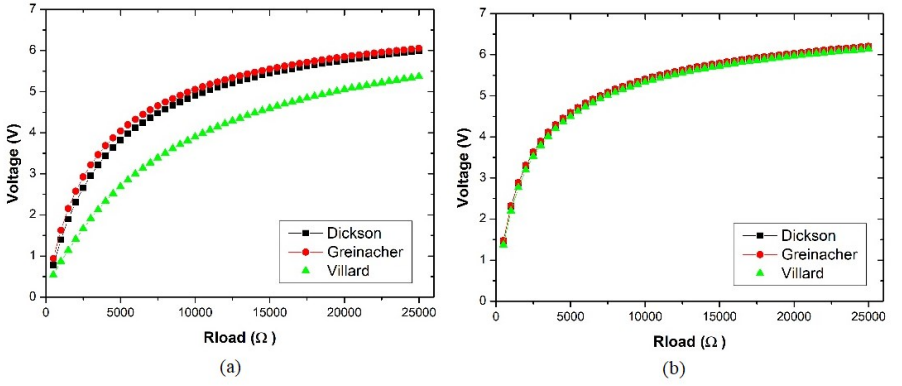


Figure 7. Output voltages versus load resistances for 10 dBm input power (a)10 MHz (b)100 MHz

Greinacher and Dickson voltage multipliers output voltages are close to each other for different input power levels at 10 MHz input power, depicted in Figure 7(a) and higher than Villard type. With the increase of frequency, in this case 100 MHz, all

voltage multipliers achieve the same output voltages, as shown in Figure 7 (b). High voltage levels can be obtained by high input power but input frequencies don't effect the output voltage levels. The voltage level is not enough parameter to compare voltage multipliers with each other so effect of the load resistance to efficiency is examined for 10 dBm input power and input frequencies of 10 MHz and 100 MHz in Figure 8.

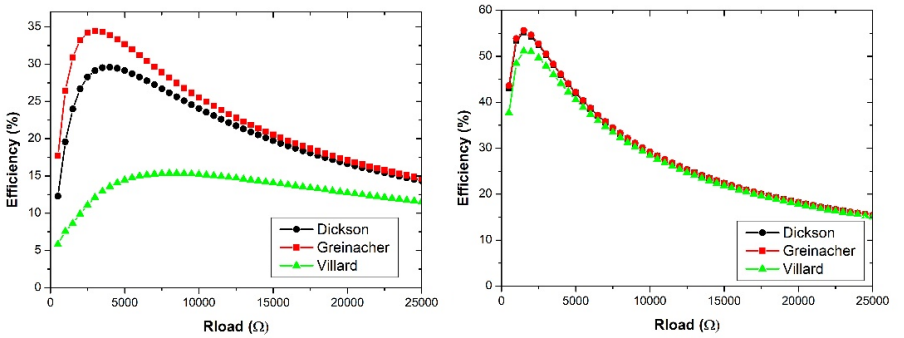


Figure 8. Efficiency vs load resistance (a) 10 MHz input frequency (b) 100 MHz input frequency

Figure 8 shows that Greinacher type voltage multiplier gives the best efficiency among Villard and Dickson types, efficiency differences are very small for 100 MHz. Also the load resistance is an important parameter for the system efficiency because the highest efficiency is obtained at 3 kΩ for 10 MHz and 2 kΩ for 100 MHz for all types of voltage multipliers. For instance, the efficiency of Greinacher voltage multiplier is 56% for 3 kΩ load resistance, but its efficiency decreases 20% for 20 kΩ load resistance. According to Figure 8, we can say that the selection of voltage multiplier is very important for low frequencies such as 10 MHz, however any voltage multiplier can be selected for high frequencies such as 100 MHz.

Input frequency is another important parameter which can help to determine the voltage multiplier for RF energy harvesting applications. So, after analyzing load resistance effect on the system efficiency for 10 dBm, input frequency effect is compared for fixed 10 dBm input power and 2 kΩ load resistance in the Figure 9.

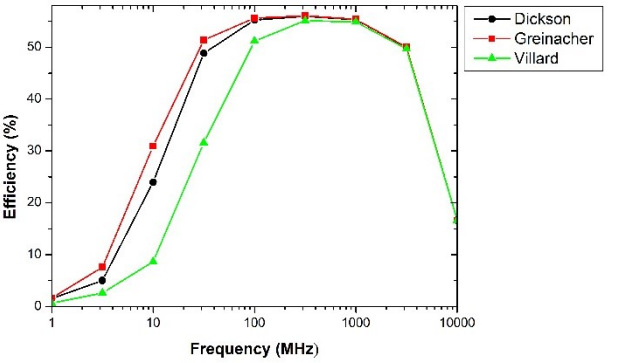


Figure 9. Input frequency effect to the efficiency

As can be seen from Figure 9, significant efficiency differences are observed up to 1 GHz. Such as Greinacher has the best and Villard has the worst efficiencies. But input frequencies higher than 1 GHz all multipliers have same efficiency furthermore, efficiencies of voltage multipliers start to diminish because of the selected diodes. If we will work above this frequency, we must change the diodes. Input power effect is also investigated. Power values of VM's for different input power is given in Figure 10, input power and load resistance are taken fix values 10 MHz and 2 kΩ respectively.

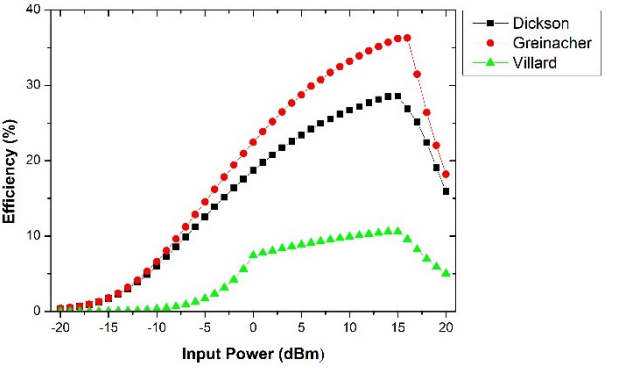


Figure 10. Efficiency vs Input power for 10 MHz input

Among the all voltage multipliers, Villard gives the worst efficiency for all input power levels. Choosing Greinacher voltage multiplier provides higher efficiencies no matter the received signal level.

6. Conclusions

The Greinacher, Villard and Dickson voltage multipliers were compared for RF energy harvester circuits. Among them Greinacher voltage multiplier have the highest efficiency for all input parameters such as input power, input frequency and load resistance. Because the output voltage is obtained from two capacitors and a diode in Villard and Dickson voltage multipliers, however, it is obtained only two capacitors in Greinacher voltage multiplier. There is no extra voltage drop on the diode for output. This is the most important reason for Greinacher voltage multiplier has higher efficiency than the other two circuits. In addition to the selection of voltage multiplier, load resistance, input power and frequency are important parameters for the system efficiency. At the low input frequencies such as 10 MHz, the selection of voltage multiplier has vital importance. However, these selection is not important at the frequencies of 1000 MHz and higher.

References

[1] E. D. Mantiply, K. R. Pohl, S. W. Poppell and J. A. Murphy, Summary of measured radiofrequency electric and magnetic fields (10 kHz to 30 GHz) in the general and work environment, *Bioelectromagnetics*, 18 /8 (1997) 563–577.

[2] X. Lu, P. Wang, D. Niyato, D.I. Kim, and Z. Han, Wireless Networks With RF Energy Harvesting : A Contemporary Survey, *IEEE Communications Surveys & Tutorials*, 17/2 (2015) 757–789.

[3] M. Piñuela, P.D. Mitcheson and S. Lucyszyn, Ambient RF energy harvesting in urban and semi-urban environments, *IEEE Trans. Microw. Theory Tech*., 61/7 (2013) 2715–2726.

[4] Y. Uzun, Design of an efficient triple band RF energy harvester, *Appl. Comput. Electromagn. Soc. J.*, 30/12 (2015) 1286–1293.

[5] B.L. Pham and A.V. Pham, Triple bands antenna and high efficiency rectifier design for RF energy harvesting at 900, 1900 and 2400 MHz*, 2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, Seattle, WA, (2013) 13–15.

[6] H. Kumar, M. Arrawatia and G. Kumar, Broadband Planar Log-Periodic Dipole Array Antenna Based RF-Energy Harvesting System, *IETE Journal of Research*, 2063 (2017) 1–5.

[7] S. Chandravanshi, S. Sen Sarma and M.J. Akhtar, Design of Triple Band Differential Rectenna for RF Energy Harvesting, *IEEE Trans. Antennas Propagation*, 66/6 (2018) 2716–2726.

[8] J.A. Leon-Gil, J.C. Perales-Cruz, L. Licea-Jimenez, S.A. PérezGarcia and A. Alvarez-Quintana, RF energy scavenging system for DC power from FM broadcasting based on an optimized Cockroft Walton voltage multiplier, *Journal of Electromagnetics Waves and Applications*, 29/11 (2015) 1440–1453.

[9] A. Nimo, D. Grgić and L.M. Reindl, Optimization of Passive Low Power Wireless Electromagnetic Energy Harvesters, *Sensors*, 12/12 (2012) 13636–13663.

[10] P. Nintanavongsa, U. Muncuk and D.R. Lewis, Design Optimization and Implementation for RF Energy Harvesting, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2/1 (2012)* 24–33.

[11] N. Akter, B. Hossain, H. Kabir, A.H. Bhuiyan, M. Yeasmin and S. Sultana, Design and Performance Analysis of 10-Stage Voltage Doublers RF Energy Harvesting Circuit for Wireless Sensor Network, *Journal of Communications Engineering and Networks*, 2/2 (2014) 84–91.

[12] K.K.A. Devi, N. Din, C.K. Chakrabarty and S. Sadasivam, Design of an RF - DC Conversion Circuit for Energy Harvesting*, 2012 IEEE International Conference on Electronics Design, Systems and Applications (ICEDSA),* Kuala Lumpur, (2012) 156–161.

[13] G.P. Tan and J.J.P. Bautista, Utilization of Low Cost RF Harvester Circuit in Harnessing Electrical Energy from Multiband RF Signals*, 2017 Asian Conference on Energy, Power and Transportation Electrification (ACEPT)*, Singapore, (2017) 1-5.

[14] F. Giusa, F. Maiorca, A. Noto, C. Trigona, B. Andò and S. Baglio, A diode-less mechanical voltage multiplier: A novel transducer for vibration energy harvesting, *Sensors Actuators A: Physical*, 212 (2014) 34–41.

[15] B. Axelrod, G. Golan, Y. Berkovich and A. Shenkman, Diode–capacitor voltage multipliers combined with boost-converters: topologies and characteristics, *IET Power Electronics*, 5/6 (2012) 873–884.

[16] J. Mucko, The Cockcroft-Walton voltage multiplier fed by an inverter in which the series resonant phenomena were used*, 2018 Innovative Materials and Technologies in Electrical Engineering (i-MITEL),* Sulecin, (2018) 1-6.

[17] Y. Berkovich, B. Axelrod, D. Shoshani, Y. Beck and T. Aviv, DC-DC Converter Based on the Bipolar Boost Converter and Dickson Voltage Multiplier*, 2018 IEEE International Energy Conference (ENERGYCON),* Limassol, (2018) 1–6.

[18] H. Jabbar, Y.S. Song and T.T. Jeong, RF energy harvesting system and circuits for charging of mobile devices, *IEEE Transaction on Consumer Electronics*, 56/1 (2010) 247–253.

[19] N.A.K.Z. Abidin, N.M. Nayan, M.M. Azizan and A. Ali, Analysis of voltage multiplier circuit simulation for rain energy harvesting using circular piezoelectric, *Mechanical Systems and Signal Processing*, 101 (2018) 211–218.

[20] K.K.A. Devi, N.M. Din and C.K. Chakrabarty, Optimization of the Voltage Doubler Stages in an RF-DC Convertor Module for Energy Harvesting, *Circuits and Systems*, 3/3 (2012) 216–222.
[21] I. Chaour, A. Fakhfakh and O. Kanoun, Enhanced passive RF-DC converter circuit efficiency for low RF energy harvesting, *Sensors*, 17/3 (2017) 1–14.
[22] H. Yan, J.G. Macias Montero, A. Akhnoukh, L.C.N. de Vreede and J.N. Burghart, An Integration Scheme for RF Power Harvesting, *8th Annual Workshop on Semiconductor Advances for Future Electronics and Sensors*, Veldhoven, the Netherlands, (2005) 64-66.
[23] M. Al-Husseini, A. Haskou, N. Rishani and K.Y. Kabalan, Textile-Based Rectennas, *WIT Transactions on State-of-the-art in Science and Engineering*, (California State University, USA, 2015).
[24] U. Olgun, C.C. Chen and J. L. Volakis, Investigation of rectenna array configurations for enhanced RF power harvesting, *IEEE Antennas and Wireless Propagagation Letter,* 10/2 (2011) 262–265.
[25] F. Yuan, CMOS circuits for passive wireless microsytems, *Springer* (New York, 2011).
[26] T. Van Breussegem and M. Steyaert, CMOS Integrated Capacitive DC–DC Converters, 39 Analog Circuits and Signal Processing, Springer (Science+Business Media New York 2013)

Current Address: Filiz SARI, Department of Electrical and Electronics Engineering, Faculty of Engineering, Aksaray University, Aksaray, Turkey
E-mail: filizsari@aksaray.edu.tr
ORCID: https://orcid.org/0000-0001-8462-175X

Current Address: Yunus UZUN, Department of Electrical and Electronics Engineering, Faculty of Engineering, Aksaray University, Aksaray, Turkey
E-mail: yunusuzun@aksaray.edu.tr
ORCID: https://orcid.org/0000-0002-1412-8473

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 24-32 (2019)
DOI: 10.33769/aupse.467095
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

SYNAPTIC TRANSMISSION: A MODEL ON THE FORMATION OF END-PLATE POTENTIAL AND A STUDY ON SIMULATION

Levent ÖZBEK, Çağri TEMUÇIN, Bülent CENGIZ and Fikri ÖZTÜRK

ABSTRACT. In this study, the emergences of postsynaptic action potentials are investigated using simulation study which is depended on the statistical model of acetylcholine-quantal (Ach) release theory at motor end-plate. It is supposed that the end-plate potentials (EPP) are sum of miniature end-plate potentials (MEPP) which are released as quantal and random in number (N). By the hypothesis that N number has a Poisson and MEPPs have a Gaussian distribution, EPPs are obtained by simulating distributions parameters. Our study suggests that the simulation of the EPPs are a convenient tool in understanding both physiological process of synaptic transmission and pathological process of neuromuscular junction.

1. INTRODUCTION

The functions of multicellular living beings demand communication among cells. A simulation study on a model regarding the formation of end-plate potential which is an important subheading of neurophysiology has been made.

Any kind of change in inner or outer environment is transformed into electrical signals in specialized cells and transmitted through neurons to the brain where it is perceived, interpreted, and preserved. Commands given by the brain to the executive organs such as muscles and glands are also transmitted as potential changes in the nerve cells.

Resting membrane potential has a characteristic value for each cell. When the nerve and membrane cells are stimulated by a proper stimulus, a temporary change happens in the membrane potential. If the value exceeds a critical level, a quick

Received by the editors: October 10, 2018; Accepted: February 23, 2019.
Key word and phrases: Synaptic transmission, EPP, MEPP, Poisson distribution.

transmission of information through the potential changes which may spread along the membrane is conducted among the organs.

Cells called *nerve cells* or *neurons* feature in transmission of information in the nervous system. The touch area in which the information exchange is conducted between two neurons is called *synapse [1]*. The information gathered from other neurons by the dendrites or synaptic connections of a neuron is transmitted as potential changes through its axone to the nerve endings. It may also be transmitted from nerve endings to other cells synaptically.

The resting state of nerve and membrane resembles a polarized condenser which has a positive charge outside, a negative charge inside and a good insulator originating from lipid (fat) bilayer at the center. Under the influence of a proper stimulus, the condenser is likely to discharge and may even depolarize for a short amount of time. After this potential change that happens in about 1 millisecond in a particular area of the membrane, the potential change continues spreading along while the membrane is turning back to the resting state. This potential change which is the fundamental unit of the information transmitted through the tendon is called *the action potential*. These potential changes may be monitored and recorded with electrical measurement and observation tools.

If the cells are too far away from each other, the transmission of information takes place by chemical factors. Transfer of information between neighbor nerve cells or between a neighbor nerve cell and a membrane cell is called *synaptic transmission*; specialized regions in which the communication between neighbor cells take place are called *synapses*. The cell called *sender* and *presynaptic cell* and the receiver *postsynaptic cell* get considerably closer to each other in these intersection points.
There are little saccules at the endings of axones. These saccules contain neurotransmitters which play an active role in the synaptic transition. When the action potential reaches the end of a nerve, molecules of neurotransmitters are released from these saccules to synaptic gap. Some of the molecules moving by the diffusion in the intercellular liquid that fills the synaptic gap are able to canalise through ionic channels (or some channels are likely to be closed) in the postsynaptic membrane. As a result of the changes of conductibility and permeability, potential changes happen in the postsynaptic membrane. If the potential change that is generally called postsynaptic potential reaches a critical value, an action potential spreading through the membrane is likely to develop.

When a stimulation spreading through a motor neuron reaches the end of a nerve, it is able to be transferred to the muscle fibre through the synaptic transmission in the neuromuscular junctions. Neuromuscular junctions are also known as end-plates. The basic chemical substance in-between the neuromuscular junction is acetylcoline (ACh). There are plenty of ACh saccules at the endings of motor neurons; in suitable conditions, they combine with the membrane and release ACh molecules they contain to the synaptic gap. While some of the released ACh molecules move away from the gap by the diffusion, some of them combine with the receptors in the postsynaptic membrane causing a local potential change. If this potential change named *End-Plate Potential (EPP)* exceeds a critical level, an action potential emerges.

The ACh saccules found at the endings of nerves are occasionally able to develop and discharge by itself, causing little amounts of depolarization about 0.5 mV in the end plate. As a result of examining these potential fluctuations named *miniature late plate potentials (MEPP)*, it is understood that the molecules of acetylcoline are oscillated in little packages or *quantums* and a normal *EPP* emerges as a result of the simultaneous oscillation of many individual quantums at the presynaptic ending. Findings that Dale and his colleagues got in 1936 indicated that the transmission of the signal from the nerves to the muscles was taking place through the molecules of acetylcholine (ACh) [2]. Studies made by Katz and Fatt in 1951 provided an opportunity to understand the mechanism of this phenomenon [3]. These researchers determined that there was a spontaneous transmission per second in a frog's neuromuscular junction. They observed that these spontaneous potentials were much smaller than potential amplitudes which are obtained through nerve stimulation. The potentials that emerge through nerve stimulation are called *End-Plate Potentials (EPPs)* and the smaller potentials that emerge spontaneously are called Miniature End-Plate Potentials (MEPPs).

According to the Quantum Hypothesis put forward by Castillo and Catz in 1955 [4], EPPs represents the total amount of MEPPs (Catz received the Nobel prize in 1970 at University College London). This means that the synaptic transmission comprises of the reproduction and accumulation of a particular amount of little packages. It is observed with an electron microscope that saccules 500 angstrom in diameter were existing almost simultaneously. Afterwards, it is determined that these saccules were containing ACh. Thus, the quantal characteristic of the synaptic transmission has been revealed with the evidences.

This paper is organized as follows: In Section 2, a stochastic model for synaptic transmission proposed by [5] and [6] are summarized. In Section 3, simulation results are presented. Section 4 concludes.

2. Model

A simple stochastic model has been put forward to test the quantum hypothesis. The model is grounded on the anatomic finding that there were oscillation areas of transmitters which are distinguishable from each other. According to this hypothesis, *arbitrary* number of oscillation areas activate when the stimulus reaches the junction. It is a well-accepted (also tested with empirical observations) fact that the *number (n)* of the oscillation areas which are activated within a stimulus had the features of the Poisson distribution. MEPPs independent from each other take place in the postsynaptic area by the activation of these different oscillation areas.

Experimental data has shown that the amplitude of these MEPPs was distributed normally [7]. The experimental data also shows that the EPPs emerge in the postsynaptic area may be described as the total temporal and spatial amount of MEPPs. In case the amplitude of EPPs reaches a level that depolarizes a membrane over a particular threshold value, an action potential emerges. This potential spreads through the voltage-gated sodium channels in the membrane along the muscle fibre and causes muscle contraction. The amplitude of the EPPs which emerge following a presynaptic stimulation normally equals to four times the threshold value (safety factor). The mathematical expression for the hypothesis is as follows. Simple stochastic model was advanced which could test the quantum hypothesis. Based on the anatomical evidence, it was supposed that there were physically distinguishable transmitter release sites. When a nerve impulse invades the junction, a random number (say N) of these sites are activated. Let the contributions from the various release sites be independent and identically distributed random variables X_k so that the total response is

$$V = X_1 + X_2 + ... + X_N \tag{1}$$

Based on experimentation, a reasonable approximation is that X_k 's are normally distributed

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0 \qquad (2)$$

with mean μ and variance σ^2. The natural choice for N is that of a binomial random variable, but in the original model it was assumed to be Poisson (λ)

$$f(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad n = 0,1,2, \ldots \qquad (3)$$

$E(N) = \lambda$ and $Var(N) = \lambda$, so the amplitude of the EPP becomes compound Poisson. The density of V is given by

$$f(v) = e^{-\lambda}\left[\delta(v) + \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!\sqrt{k}}\right] \exp\left(\frac{-(v-k\mu)^2}{2k\sigma^2}\right) \qquad (4)$$

where $\lambda = E(N)$ (The details of the proof see Tuckwell, 1994 and Tuckwell, 1995). The mean and variance of V are

$$E(V) = \lambda\mu$$
$$Var(V) = \lambda(\mu^2 + \sigma^2)$$

The excellent fit of $f(v)$ to experimental distributions of some EPP amplitudes provided impressive evidence to support the quantum hypothesis (see Martin, 1977).

3. The Simulation for the Emergence of the End-Plate Potential

Valuing the variables in the proposed model differently, two separate simulation studies have been made.

In order to generate numbers from the density function given in Equality 4, the corresponding distribution function and its reverse form must be determined. Due to the fact that this distribution function is not easily found, Equality 1 is used to generate numbers from the distribution which is relevant to the density function given in Equality 4. Firstly, parameter λ is determined and then, the N number is generated from the Poisson distribution (Equality 3) with regard to the parameter. Afterwards, N amount of number is generated from the normal distribution

(Equation 2) in μ and σ^2 parameters set and the N amounts of numbers are added together. Same process is repeated m times and the histogram for the numbers are drawn. m was assumed as being equal to 5000 in the simulation studies made. MATLAB application has been used for the studies.

3.1 First Simulation

The simulation was generated from the Poisson distribution with an average of 2.4 N value. X_k 's were generated from the normal distribution with 0.4 mean and 0.1 standard deviation. The EPP mean was found to be 0.9601 and the standard deviation was 0.6399. The EPP minimum was found to be 0 and the maximum was 5.7179. In these circumstances, the histogram for the obtained EPPs is given Figure

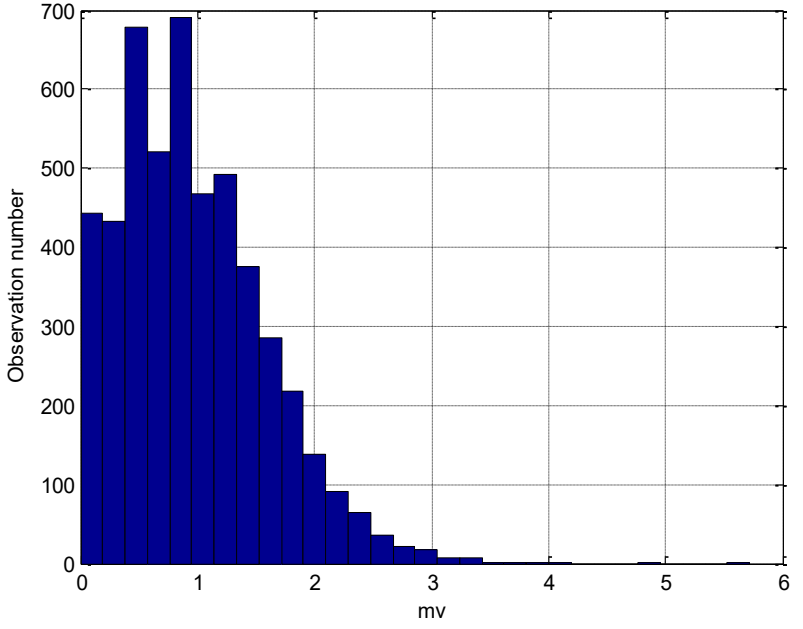


Figure 1: EPP simulation λ = 2.4, N (0.4, 0.1).

3.2 Second Simulation

It is shown that the amplitudes of MEPP decreased in *myasthenia gravis (MG)* disease. Providing that the N number is same within the simulation for the dysfunction of postsynaptic neuromuscular junction, the X_k averages which represent the average amplitude for the MEPPs are decreased. X_k were generated from the normal distribution with 0.2 average and 0.1 standard deviation. The EPP average was found to be 0.4827 and the standard deviation was 0.3488. The EPP minimum was found to be -0.1341 and the maximum was 2.7074. In these circumstances, the histogram for the obtained EPPs is given by Figure 2.

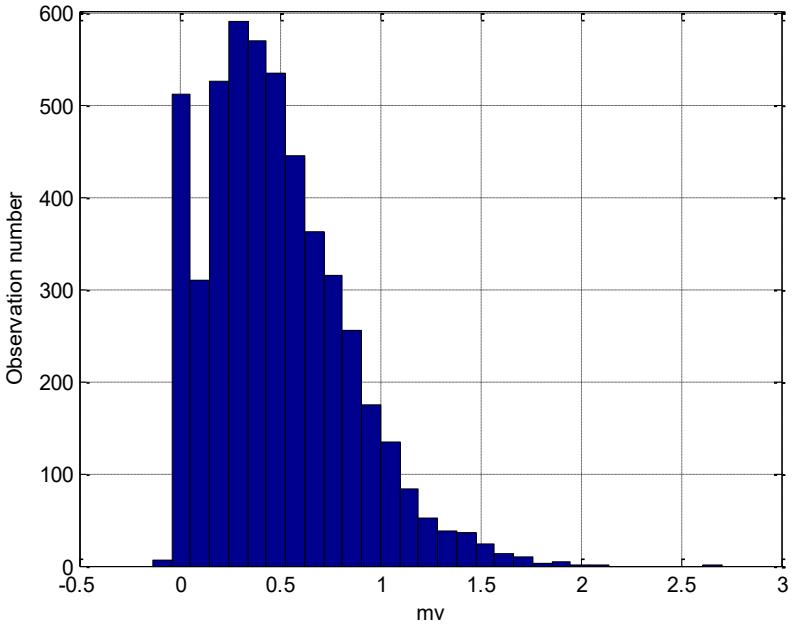


Figure 2: EPP simulation $\lambda = 2.4$, N (0.2, 0.1).

As one may see, there is a decrease in the amplitude of the EPPs in the simulation for the MG physiopathology, leading the safety factor to decrease too. These changes predicted by the simulation are coherent with the clinical and electrophysiological observations.

4. Conclusion

This study aimed to research the emergence of the postsynaptic action potentials using the simulation study based on the statistical model in the theory of acetylcholine-quantal (Ach) release at motor end-plate. End-plate potentials (EPP) have been assumed as amounting to the total for numerically (N) quantal and randomly-released miniature end-plate potentials (MEPP). EPPs have been obtained by simulating the distribution parameters based on the hypothesis that N number was in Poisson and MEPPs were in Gaussian distribution. The results of this study have suggested that the simulation of EPPs might be the appropriate means by which both the neuropathic transmission and the neuromuscular resultant's physiopathological process are understood.

By including the other variables such as the stimulation frequency, the ambient temperature, the width of the synaptic gap, the number of the presynaptic saccules, and the physiology of calcium channels in the further researches, much more realistic models and simulation studies may be made.

References

[1] Neural Signaling. In: Dale Purves, George Augustine, David Fitzpatrick, William Hall, Anthony-Samuel Lamantia, Leonard White Neuroscience. 5th Edition.
[2] Dale, HH, Feldberg W, and Vogt M, Release of acetylcholine at voluntary motor nerve endings. J Physiol., 86 (1936) 353–380
[3] Fatt P, Katz B. An analysis of the end-plate potential recorded with an intra-cellular electrode. J Physiol. 115, (1951) 320–370.
[4] del Castillo J, Katz B. Quantal components of the end-plate potential. J Physiol. 124 (1954) 560–573.
[5] Tuckwell, HC. Elemantary Aplications of Probability Theory, Chapman & Hall, 1995.
[6] Tuckwell, HC. Stochastic Processes in the neuroscience, The Society for Industrial and Applied Mathemetics, 1994.
[7] Martin, AR. Junctional Transmission II. Presynaptic mechanisms. In Hand book of Physiology, The Nervous System. Vol 1 (ed E:R: Kandel) Amer. Physiol. Soc., Bethesda, Md. P328-355.

Current Address: Levent Özbek, Ankara University, Faculty of Sciences, Department of Statistics, Ankara, Turkey
E-mail: ozbek@science.ankara.edu.tr
ORCID: https://orcid.org/0000-0003-1018-3114

Current Address: Çağrı Temuçin, Hacettepe University, Medical Faculty, Department of Neurology, Ankara, Turkey
E-mail: ctemucin@hacettepe.edu.tr
ORCID: https://orcid.org/0000-0003-0127-6861
Current Address: Bülent Cengiz, Gazi University, Medical Faculty, Department of Neurology, Ankara, Turkey
E-mail: bcengiz@gazi.edu.tr
ORCID: https://orcid.org/0000-0002-1688-9516

Current Address: Fikri Öztürk, Ankara University, Faculty of Sciences, Department of Statistics, Ankara, Turkey
E-mail: ozturkf@ankara.edu.tr
ORCID: https://orcid.org/ 0000-0002-7175-7372

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 33-56 (2019)
DOI: 10.33769/aupse.459815
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

EFFICIENT CHEBYSHEV ECONOMIZATION FOR ELEMENTARY FUNCTIONS

Esmat BEKIR

ABSTRACT. This paper presents economized power series for trigonometric and hyperbolic functions. It determines the smallest range over which a function need to be computed and scales the Chebyshev polynomials accordingly. Thus reduced degree polynomials (and hence reduced computations) can be obtained while maintaining the same accuracy as those unscaled higher degree polynomials. The paper presents the Chebyshev and the power series coefficients that enable double precision accuracy for the mathematical functions addressed herein.

1. INTRODUCTION

The Chebyshev polynomials possess useful properties that render them proper for economizing transcendental functions, specifically trigonometric and hyperbolic functions, [1]-[3]. Commonly these functions use truncated Taylor series expansion. In this truncation method, the more the number of the retained terms the higher the accuracy of the approximation. However, this method suffers from the uneven distribution of errors in the approximation. The closer the evaluated point to the origin of expansion the higher the accuracy and vice versa. This means that for a desired level of accuracy the points far from the origin will need substantially more terms than those close to the origin of expansion. This problem could be alleviated by using minimization methods such as the least square (LS) algorithm. In this case the function $f(x)$ is approximated with a finite degree polynomial $\sum_{k=0}^{N-1} c_k x^k$ whose coefficients c_k are selected such that

Received by the editors: September 13, 2018; Accepted: March 22, 2019.
Key word and phrases: Chebyshev polynomial, economization, trigonometric functions, hyperbolic functions.

$$J = \int_{-1}^{1} w(x)\left(f(x) - \sum_{k=0}^{N-1} c_k x^k \right)^2 dx \tag{1.1}$$

is minimum, where $w(x)$ is an arbitrary weighting function. With no loss of generality the [-1,1] is the interval in which the function is approximated. The minimization in Eq. (1.1) yields

$$\sum_{k=0}^{N-1} c_k \int_{-1}^{1} w(x) x^k x^n dx = \int_{-1}^{1} w(x) f(x) x^n dx, \quad n = 0,1,\cdots \tag{1.2}$$

From its construction, Eq. (1.2) is impractical to solve as it requires the computation of a full two dimensional matrix. The reason is that the function $f(x)$ is approximated with the unorthogonal power series basis $(1, x, x^2, \cdots$
This could be avoided if the function is approximated instead with an orthogonal basis. That is, if the orthogonal basis is given by $T_0, T_1, T_2, \cdots$, then the coefficients c_k are determined by

$$c_k \int_{-1}^{1} w(x) T_k^2(x)\, dx = \int_{-1}^{1} w(x) f(x) T_k(x) dx, \quad k = 0,1,2,\cdots \tag{1.3}$$

Using an orthogonal basis, will cause the off diagonal terms be null; thus the coefficients become the projections of the function over the members of the basis.

Our objective is to distribute the errors uniformly over the given interval. Chebyshev basis is considered to be the best choice. To demonstrate the reason for this choice consider the simple function

$$f(x) = x^n, \quad n \text{ is integer} \geq 1$$

which we desire to approximate with a polynomial p_m of degree m (less than n). The interval over which the approximation will be carried is [-1, 1]. The error is then

$$e(x) = f(x) - p_m(x), \quad -1 \leq x \leq 1$$

In general, the error is not uniform; therefore there will be a point in the interval at which the error is a maximum. It is desired that p_m is selected such that this maximum

error is the least possible. It has been established, [4]-[6], that there is no polynomial $e(x)$ of the same degree and the same leading coefficient that has a smaller magnitude in [-1, 1] than the Chebyshev polynomial $T_n(x)$. Therefore if the error $e(x) = cT_n(x)$ then the desired approximation will be $p_m(x) = f(x) - e(x) = f(x) - cT_n(x)$. The constant c is selected to annul the coefficient of x^n. In our example since the leading coefficient of $f(x)$ is 1 and (as will be seen in the next section) the leading coefficient of $T_n(x)$ is 2^{n-1}, then $c=2^{-n+1}$. To generalize, let the above function be

$$f(x) = a_0 + a_1x^1 + a_2x^2 + \ldots + a_nx^n$$

Following the same steps and reasoning as above we present the above equation as

$$f(x) = a_0 + a_1x^1 + a_2x^2 + \ldots + a_n\left(x^n - cT_n(x)\right) + a_ncT_n(x)$$

Dropping the last term in the above equation, (after selecting c as before), will yield a lower degree polynomial with the least maximum error. After dropping the last term $a_ncT_n(x)$, we expand $T_n(x)$ into a power series to recover the approximated function. This example is cited to show that the lower powers had no effect on the minimization process. This process can be repeated to get a lower degree approximation as desired. Later we show the approximation process for an arbitrary function. But first let's discuss some properties of the Chebyshev polynomials.

2. The Chebyshev Polynomials

In closed form, the Chebyshev polynomials are given by the set of equations

$$T_n(x) = \cos(n\cos^{-1}x), \quad n = 0, 1, 2, \cdots \tag{2.1}$$

and with the transformation

$$x = \cos\theta \tag{2.2}$$

they could be alternatively represented by

$$T_n(\theta) = \cos(n\theta), \quad n = 0, 1, 2, \cdots \tag{2.3}$$

The Chebyshev polynomials can be generated recursively, shown in Appendix A, by

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x) \end{aligned} \tag{2.4}$$

Eq. (2.3) shows that the magnitudes of the basis polynomials are ≤ 1 and are equal ±1 at the extreme points (x=±1). Further, Chebyshev polynomials are orthogonal [2]

$$\begin{aligned} &\int_{-1}^{1} w(x)T_k^2(x)\,dx = \frac{\pi}{2H(k)}, \quad k = 0,1,2,\cdots \\ &\int_{-1}^{1} w(x)T_k T_n(x)\,dx = 0, \quad k \neq n \end{aligned} \tag{2.5}$$

where $H(k)$ is defined in (A.11) and $w(x)$ is the weighting function given by

$$w(x) = \frac{1}{\sqrt{1-x^2}} \tag{2.6}$$

The inverse relations of Eq. (2.1), proved in Appendix A, are given by

$$x^{2n} = \frac{1}{2^{2n-1}} \sum_{k=0}^{n} \binom{2n}{n-k} H(k) T_{2k}, \quad n = 0,1,2,\cdots \tag{2.7}$$

$$x^{2n+1} = \frac{1}{2^{2n}} \sum_{k=0}^{n} \binom{2n+1}{n-k} T_{2k+1}, \quad n = 0,1,2,\cdots \tag{2.8}$$

From Eq. (1.3), the projections of $f(x)$ onto the Chebyshev basis results in

$$c_k = \frac{2H(k)}{\pi} \int_{-1}^{1} \frac{f(x)T_k(x)}{\sqrt{1-x^2}} dx, \quad k = 0,1,2,\cdots \tag{2.9}$$

The above equation is not easily amenable to digital computations. Alternatively we consider $f(x)$ to be an analytic function represented by

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \tag{2.10}$$

and rather than projecting the entire function over the basis as mandated by Eq. (2.9), each of the polynomial terms is projected individually onto the Chebyshev polynomial through the use of Eqs. (2.7) and (2.8). Initially we shall consider the special cases of even and odd functions.

Expanding an even function $f(x) = f(-x)$ in terms of Chebyshev polynomials yields

$$f(x) = \sum_{n=0}^{\infty} a_{2n} x^{2n} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{2^{2n-1}} a_{2n} \binom{2n}{n-k} H(k) T_{2k} \tag{2.11}$$

Reversing the summation order results in

$$f(x) = \sum_{k=0}^{\infty} c_{2k} T_{2k}, \quad f(x) \text{ is even} \tag{2.12}$$

$$c_{2k} = \sum_{n=k}^{\infty} \frac{1}{2^{2n-1}} \binom{2n}{n-k} a_{2n} H(k), \quad k = 0, 1, 2, \ldots \tag{2.13}$$

Likewise, an odd function $f(x) = -f(-x)$ in terms of Chebyshev polynomials is

$$f(x) = \sum_{n=0}^{\infty} a_{2n+1} x^{2n+1} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{2^{2n}} a_{2n+1} \binom{2n+1}{n-k} T_{2k+1} \tag{2.14}$$

Reversing the summation order results in

$$f(x) = \sum_{k=0}^{\infty} c_{2k+1} T_{2k+1}, \quad f(x) \text{ is odd} \tag{2.15}$$

$$c_{2k+1} = \sum_{n=k}^{\infty} \frac{1}{2^{2n}} \binom{2n+1}{n-k} a_{2n+1}, \quad k = 0,1,2,\ldots \tag{2.16}$$

Unsymmetric function can be represented as a combination of Eqs. (2.13) and (2.16), albeit with double the amount of computations needed for a symmetric function. Implementing the economization procedure presumes that the range of the argument x of interest is between ±1. If this range is scaled by a factor of s, (range=±s), then Eqs. (2.11) –(2.16) would be modified as follows.

$$f(sx) = \sum_{k=0}^{\infty} c_{2k} T_{2k}, \qquad f(sx) \text{ is even} \tag{2.17}$$

$$c_{2k} = 2\sum_{n=k}^{\infty} \left(\frac{s}{2}\right)^{2n} \binom{2n}{n-k} a_{2n} H(k), \quad k = 0,1,2,\ldots \tag{2.18}$$

and

$$f(sx) = \sum_{k=0}^{\infty} c_{2k+1} T_{2k+1}, \qquad f(sx) \text{ is odd} \tag{2.19}$$

$$c_{2k+1} = 2\sum_{n=k}^{\infty} \left(\frac{s}{2}\right)^{2n+1} \binom{2n+1}{n-k} a_{2n+1}, \quad k = 0,1,2,\ldots \tag{2.20}$$

Smaller scales as observed from Eqs. (2.18) and (2.20) result in smaller coefficients and consequently faster coefficients decay and using of fewer expansion terms.
Now we address three issues regarding numerical implementation: accuracy, accounting for the scales and carrying out the summation terms. Truncating the first N terms in Eqs. (2.17) and (2.19) as follows:

$$f(sx) \approx \sum_{k=0}^{N-1} c_k T_k \tag{2.21}$$

implies that the approximation error, δf, on the truncated function is (recall that $|T_i| \le 1$)

$$\delta f \le \sum_{k=N}^{\infty} |c_k| \tag{2.22}$$

Because of the fast decay of the c_k terms, the above error is dominated by c_N . Thus N is selected so c_{N-1} is just less than the tolerable errors. Next we account for the scale by one of these two methods. The first is to modify the argument so that

$$f(z) = \sum_{k=0}^{N-1} b_k x^k, \quad x = \frac{z}{s}, -s \le z \le s \tag{2.23}$$

or by modifying the coefficients of Eq. (2.23) so that

$$\begin{aligned} f(x) &= \sum_{k=0}^{N-1} d_k x^k, \quad -s \le x \le s \\ d_k &= \frac{b_k}{s^k}, \quad k = 0,1,...,N-1 \end{aligned} \tag{2.24}$$

Finally, the numerical process for the summation can be carried out by converting in Eq. (2.21) into a power series which yields the power series given by Eq. (2.23)or Eq. (2.24). Alternatively, Eq. (2.21) could be evaluated by the Clenshaw recurrence [7] as follows. Substituting from Eq. (2.4) in the last term of Eq. (2.21) yields

$$\begin{aligned} f(sx) &= c_0T_0 + ... + c_{N-2}T_{N-2} + c_{N-1}(2xT_{N-2} - T_{N-3}) \\ &= c_0T_0 + ... + (c_{N-3} - c_{N-1})T_{N-3} + (c_{N-2} + 2x\,c_{N-1})T_{N-2} \end{aligned} \tag{2.25}$$

Let

$$\begin{aligned} \alpha_{N-1} &= c_{N-1} \\ \alpha_{N-2} &= c_{N-2} + 2x\alpha_{N-1} \end{aligned}$$

Again, using Eq. (2.4) in the last term in Eq. (2.25) gives

$$\begin{aligned} f(sx) &= c_0T_0 + ... + (c_{N-3} - \alpha_{N-1})T_{N-3} + \alpha_{N-2}T_{N-2} \\ &= c_0T_0 + ... + (c_{N-4} - \alpha_{N-2})T_{N-4} + (c_{N-3} + 2x\,\alpha_{N-2} - \alpha_{N-1})T_{N-3} \end{aligned} \tag{2.26}$$

Using the recursion

$$\begin{aligned} \alpha_N &= 0 \\ \alpha_{N-1} &= c_{N-1} \\ \alpha_k &= c_k + 2x\alpha_{k+1} - \alpha_{k+2}, \quad k = N-2, N-3, \cdots \end{aligned} \tag{2.27}$$

equation (2.27) reduces to

$$f(sx) = c_0 + x\alpha_1 - \alpha_2 \tag{2.28}$$

3. Trigonometric Functions Economization

The Chebyshev series coefficients for the sine and cosine are given in terms of the Bessel function [3]. Herein these coefficients are given explicitly.

3.1. Sine function

The Taylor series expansion of the sine function is given by

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \tag{3.1}$$

Scaling the argument to have a range of $\pm s$, modifies the above equation to

$$\sin(sx) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} (sx)^{2n+1} \tag{3.2}$$

Substituting from Eq. (3.2) into Eq. (2.20), the Chebyshev coefficients for the scaled sine function are given by

$$\sin(sx) = \sum_{k=0}^{\infty} c_{2k+1} T_{2k+1}, \quad -1 \le x \le 1 \tag{3.3}$$

$$\begin{aligned} c_{2k+1} &= 2\sum_{n=k}^{\infty} (s/2)^{2n+1} \binom{2n+1}{n-k} \frac{(-1)^n}{(2n+1)!} \\ &= 2\sum_{n=k}^{\infty} \frac{(s/2)^{2n+1} (-1)^n}{(n-k)!(n+k+1)!}, \quad k = 0, 1, 2, \cdots \end{aligned} \tag{3.4}$$

Periodicity of the sine implies that range of $\pm\pi/2$ will cover all the possible values. However, scaling the argument to a range of $\pm\pi/4$ allows us to economize it with fewer terms to get the desired accuracy. The rest of the range (from $\pi/4$ to $\pi/2$) is recovered using the identity $\sin(x) = \cos(\pi/2 - x)$ and computing the economized cosine function. With $s = \pi/4$, the Chebyshev coefficients for the sine function are

$$\begin{aligned} \sin(\frac{\pi}{4}x) &= \sum_{k=0}^{\infty} c_{2k+1} T_{2k+1}, \quad -1 \le x \le 1 \\ c_{2k+1} &= 2\sum_{n=k}^{\infty} \frac{(\pi/8)^{2n+1}(-1)^n}{(n-k)!(n+k+1)!}, \quad k = 0,1,2,\ldots \end{aligned} \tag{3.5}$$

Table 1 lists the Chebyshev coefficients for the sine function. Listed also are its power series coefficients for degrees up to 13 (the max for a double precision processor).

3.2. Cosine function

The Taylor series expansion of the scaled cosine function is given by

$$\cos(sx) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}(sx)^{2n} \tag{3.6}$$

Substituting from Eq. (3.6) into Eq. (2.18), the Chebyshev coefficients for the scaled cosine function, $s = \pi/4$, are given by

$$\cos(sx) = \sum_{k=0}^{\infty} c_{2k} T_{2k}, \quad -1 \le x \le 1 \tag{3.7}$$

$$\begin{aligned} c_{2k} &= 2\sum_{n=k}^{\infty} (s/2)^{2n} \binom{2n}{n-k} \frac{(-1)^n}{(2n)!} H(k) \\ &= 2\sum_{n=k}^{\infty} \frac{(\pi/8)^{2n}(-1)^n}{(n-k)!(n+k)!} H(k), \quad k = 0,1,2,\ldots \end{aligned} \tag{3.8}$$

A little disadvantage with the above economization is that the computed cosine at 0 is not exactly 1. This can be overcome by forcing the cosine function to be 1 for:

$$\frac{x^2}{2} = 2.10^{-16} \Rightarrow -2.10^{-8} \le x \le 2.10^{-8} \tag{3.9}$$

The Chebyshev coefficients for the cosine function are listed in Table 2. Listed also are the power series coefficients of the cosine function for degrees 0,2…12. On a double precision digital processor, the polynomial degree is limited to 12.

3.3. Tangent function

The tangent is computed directly via the tangent expansion as given here or indirectly via the cotangent expansion along with trigonometric identity. From [9], Sec. 4.3.67

$$\tan x = \sum_{n=0}^{\infty} \frac{(-1)^n 2^{4n+4}(1-2^{-2n-2})B_{2n+2}}{(2n+2)!} x^{2n+1} \tag{3.10}$$

Substituting from Eq. (3.10) into Eq. (2.20), the Chebyshev coefficients for the scaled tangent function are given by

$$\begin{aligned} \tan(sx) &= \sum_{k=0}^{n} c_{2k+1} T_{2k+1}; \quad -1 \le x \le 1 \\ c_{2k+1} &= 2\sum_{n=k}^{\infty} (s/2)^{2n+1} \binom{2n+1}{n-k} \frac{(-1)^n 2^{4n+4}(1-2^{-2n-2})B_{2n+2}}{(2n+2)!} \\ &= 4\sum_{n=k}^{\infty} (2s)^{2n+1} \frac{(-1)^n (1-2^{-2n-2})B_{2n+2}}{(n+1)(n-k)!(n+k+1)!}, \quad k = 0,1,2,\ldots \end{aligned} \tag{3.11}$$

In Eq. (3.11) and in the sequel, the sequence B_n denotes the Bernoulli numbers. With a scale factor $s = \pi/4$ one can use Eq. (3.11) to compute the tangent for any value in the interval $[-\pi/4, \pi/4]$. From the trigonometric identity $\tan(\pi/2-x)=1/\tan(x)$ one can compute the function in the rest of the range $[\pi/4,\pi/2]$. With this scaling, the coefficients in Eq. (3.11) are extremely slow to converge. Alternatively, by selecting $s = \pi/8$ one can have an economized function that achieves the maximum double precision accuracy of 10^{-16} with a polynomial of degree 19. In this case we can use the trigonometric identity $\tan(\pi/4 - x) = (1 - \tan x)/(1 + \tan x)$ to compute the function for an argument in the interval $[\pi/8, \pi/4]$. A more economized polynomial is given by the cotangent approach as described bellow.

3.4. Cotangent function

Here we economize $f(x) = x \cot x$. From [9], Sec. 4.3.70, the Taylor series expansion of the cotangent function is given by

$$f(x) = \sum_{n=0}^{\infty} \frac{(-1)^n 2^{2n} B_{2n}}{2n!} x^{2n} \tag{3.12}$$

As in the tangent function, we use the scale factor $s = \pi/8$. Substituting from Eq. (3.12) into Eq. (2.18), the scaled cotangent function coefficients are given by

$$f(sx) = \sum_{k=0}^{n} c_{2k} T_{2k}, \quad -1 \le x \le 1 \tag{3.13}$$

$$\begin{aligned} c_{2k} &= 2\sum_{n=k}^{\infty} (s/2)^{2n} \binom{2n}{n-k} \frac{(-1)^n 2^{2n} B_{2n}}{(2n)!} H(k) \\ &= 2\sum_{n=k}^{\infty} \frac{(-1)^n s^{2n} B_{2n}}{(n-k)!(n+k)!} H(k), \quad k = 0, 1, 2, \ldots \end{aligned} \tag{3.14}$$

The tangent function is determined by

$$\tan x = \frac{x}{f(x)} \tag{3.15}$$

The maximum error on the tangent function is determined by variation of Eq. (3.15)

$$\delta \tan x = -\frac{x \delta f}{f^2(x)} \tag{3.16}$$

The maximum value is attained at $x = s = \pi/8$ for which $f(s) = \pi/8 \cot(\pi/8)$ and

$$\left|\delta \tan s\right| = \frac{s \delta f}{f^2(s)} = \frac{\tan^2(\pi/8)}{\pi/8} \delta f = .4369\, \delta f \tag{3.17}$$

Chebyshev coefficients for the cotangent function are listed in Table 3. Listed also are the power series coefficients of the cotangent function for degrees 2,4…12. Using the cotangent approach we obtain the maximum double precision accuracy with an even polynomial of degree 12 rather than the odd polynomial of degree 19 with the tangent approach.

3.5. Arctangent function

The Chebyshev coefficients for this function have been derived in [5,8]. Herein we shall obtain these coefficients through a rather elegant approach, [10]. Let

$$u = k(x + i\sqrt{1-x^2}) \qquad v = k(x - i\sqrt{1-x^2}) \tag{3.18}$$

Substituting for

$$x = \cos\theta \tag{3.19}$$

in Eq. (3.18) gives

$$u = ke^{i\theta} \qquad v = ke^{-i\theta} \tag{3.20}$$

which implies that

$$u^n + v^n = k^n (e^{in\theta} + e^{-in\theta}) = 2k^n \cos n\theta = 2k^n \cos n \cos^{-1} x = 2k^n T_n(x) \tag{3.21}$$

The Taylor series expansion of the arctangent function is

$$\tan^{-1} u = u - \frac{1}{3}u^3 + \frac{1}{5}u^5 - \frac{1}{7}u^7 + \dots$$

from which we get

$$\tan^{-1} u + \tan^{-1} v = (u+v) - \frac{1}{3}(u^3 + v^3) + \frac{1}{5}(u^5 + v^5) - \frac{1}{7}(u^7 + v^7) + \dots$$

Substituting from Eq. (3.21) in the above equation

$$\begin{aligned} \tan^{-1} u + \tan^{-1} v &= 2kT_1(x) - \frac{2k^3}{3}T_3(x) + \frac{2k^5}{5}T_5(x) - \frac{2k^7}{7}T_7(x) + \dots \\ &= 2\sum_{n=0} \frac{(-1)^n k^{2n+1}}{2n+1} T_{2n+1}(x) \end{aligned} \tag{3.22}$$

Also substituting in the trigonometric identity

$$\tan^{-1} \frac{u+v}{1-uv} = \tan^{-1} u + \tan^{-1} v$$

for u and v from Eq. (3.18) in the LHS and from Eq. (3.22) in the RHS yields

$$\tan^{-1}\frac{2k}{1-k^2}x = 2\sum_{n=0}\frac{(-1)^n k^{2n+1}}{2n+1}T_{2n+1}(x) \tag{3.23}$$

Setting the parameter k to

$$k = \tan(\alpha/2)$$

implies that

$$\frac{2k}{1-k^2} = \tan\alpha$$

which upon substituting in Eq. (3.23) yields

$$\tan^{-1}(\tan\alpha)x = 2\sum_{n=0}\frac{(-1)^n \tan^{2n+1}(\alpha/2)}{2n+1}T_{2n+1}(x), \quad -1 \le x \le 1 \tag{3.24}$$

In Eq. (3.24) $\tan\alpha$ becomes a scale that naturally sets the domain of the economized function to $\pm\alpha$. Selecting

$$\alpha = \pi/8 \tag{3.25}$$

enables Eq. (3.24) to compute the arctangent function for arguments in the range $\pm\tan(\pi/8)$ ($\tan(\pi/8)=\sqrt{2}-1$). To compute the arctangent for values in the interval $[\tan(\pi/8), \tan(\pi/4)]$ we follow this procedure. Using

$$\tan^{-1} x = \theta$$

with the trigonometric identity gives

$$\tan\left(\frac{\pi}{4}-\theta\right) = \frac{1-\tan\theta}{1+\tan\theta} = \frac{1-x}{1+x}$$

which implies that

$$\tan^{-1} x = \theta = \frac{\pi}{4} - \tan^{-1}\frac{1-x}{1+x}$$

For values greater than $\tan(\pi/4)$, we use the identity $\tan x = 1/\tan(\pi/2-x)$. The Chebyshev coefficients for the arctangent function are listed in Table 4. Listed also are the power series coefficients of the arctangent function for various degrees 1,3…21. On a double precision machine, the polynomial degree is limited to 21.

3.6. Arcsine and arccosine functions

We first discuss the direct way of economizing the arcsine function. From Eq. (2.1) we determine the Chebyshev expansion coefficients from the identity

$$\frac{\pi}{2} c_k = \int_{-1}^{1} \sin^{-1} x \frac{T_k(x)}{\sqrt{1-x^2}} dx \tag{3.26}$$

Substituting for $\theta = \cos^{-1} x$ in the above integral and performing the integration yields

$$c_{2k+1} = \frac{4}{\pi(2k+1)^2} \tag{3.27}$$

It is evident that these coefficients do not diminish rapidly. For k=50, $c_{2k+1}=O(10^{-4})$, which is an unacceptable error for such large sum of expansion terms.

A more economic approach is to compute the arcsine and arccosine via the arctangent function. For the arccosine we use the transformation $z=(1-x)/(1+x)$. Substituting for $x = \cos\theta$ in this transformation gives

$$z = \frac{1-x}{1+x} = \frac{1-\cos\theta}{1+\cos\theta} = \frac{2\sin^2\frac{\theta}{2}}{2\cos^2\frac{\theta}{2}} = \tan^2\frac{\theta}{2}$$

which implies that

$$\theta = \cos^{-1} x = 2\tan^{-1}\sqrt{\frac{1-x}{1+x}} \tag{3.28}$$

from which the arcsine function is computed using the trigonometric identity

$$\sin^{-1} x = \pi/2 - \cos^{-1} x = \frac{\pi}{2} - 2\tan^{-1}\sqrt{\frac{1-x}{1+x}} \tag{3.29}$$

4. Hyperbolic Functions Economization

The hyperbolic functions are of infinite range; thus they must be adapted to practically large finite range computations without sacrificing the desired accuracy. Herein we develop the economization for the hyperbolic sine, cosine, tangent and cotangent. Their scales are specified when we discuss the economized exponent function which will play a central role in the computations of these functions.

4.1. Hyperbolic sine function

The Taylor series expansion for a scaled hyperbolic sine is given by

$$\sinh(sx) = \sum_{n=0}^{\infty} \frac{1}{(2n+1)!}(sx)^{2n+1} \tag{4.1}$$

Similarity between sin and sinh functions implies, using Eqs. (3.3) and (3.4) that

$$\sinh(sx) = \sum_{k=0}^{n} c_{2k+1} T_{2k+1}, \quad -1 \le x \le 1 \tag{4.2}$$

$$c_{2k+1} = 2\sum_{n=k}^{\infty} \frac{(s/2)^{2n+1}}{(n-k)!(n+k+1)!}, \quad k = 0, 1, 2, \ldots \tag{4.3}$$

4.2. Hyperbolic cosine function

The Taylor series expansion for a scaled hyperbolic cosine is given by

$$\cosh(sx) = \sum_{n=0}^{\infty} \frac{1}{(2n)!}(sx)^{2n} \tag{4.4}$$

Similarity between the cos and cosh functions implies, using Eqs. (3.7) and (3.8) that

$$\cosh(sx) = \sum_{k=0}^{\infty} c_{2k} T_{2k}, \quad -1 \le x \le 1 \tag{4.5}$$

$$c_{2k} = 2\sum_{n=k}^{\infty} \frac{(s/2)^{2n}}{(n-k)!(n+k)!} H(k), \quad k = 0,1,2,\ldots \tag{4.6}$$

4.3. Hyperbolic Tangent Economization

The Taylor series expansion for the hyperbolic tangent is given by

$$\tanh x = \sum_{n=0}^{\infty} \frac{2^{4n+4}(1-2^{-2n-2})B_{2n+2}}{(2n+2)!} x^{2n+1} \tag{4.7}$$

Similarity between the tan and tanh functions implies, using Eqs. (3.10) and (3.11) that

$$\tanh(sx) = \sum_{k=0}^{n} c_{2k+1} T_{2k+1}, \quad -1 \le x \le 1 \tag{4.8}$$

$$c_{2k+1} = 4\sum_{n=k}^{\infty} \frac{(2s)^{2n+1}(1-2^{-2n-2})B_{2n+2}}{(n+1)(n-k)!(n+k+1)!}, \quad k = 0,1,2,\ldots \tag{4.9}$$

4.4. Hyperbolic cotangent function

The Taylor series expansion for hyperbolic cotangent $f(x) = x\coth x$, see [9], is

$$f(x) = \sum_{n=0}^{\infty} \frac{2^{2n} B_{2n}}{(2n)!} x^{2n} \tag{4.10}$$

Similarity between the cot and coth functions implies, using Eqs. (3.13) and (3.14) that

$$f(sx) = \sum_{k=0}^{n} c_{2k} T_{2k}, \quad -1 \le x \le 1 \tag{4.11}$$

$$c_{2k} = 2\sum_{n=k}^{\infty} \frac{s^{2n} B_{2n}}{(n-k)!(n+k)!} H(k), \quad k = 0,1,2,\ldots \tag{4.12}$$

4.5. Exponent function

To perform the exponent computation on a digital computer we first note that

$$\log_2 e^x = x \log_2 e = \frac{x}{\ln 2} \qquad \Rightarrow \qquad e^x = 2^{\frac{x}{\ln 2}} \tag{4.13}$$

hence, if n and z are the nearest integer and remainder of $x/\ln 2$ respectively, *i.e.*

$$n = \text{nint}\left(x / \ln 2\right) \qquad z = x / \ln 2 - n \tag{4.14}$$

then

$$e^x = 2^{x/\ln 2} = 2^n 2^z = 2^n e^{z \ln 2} = 2^n e^{x - n \ln 2} \quad -.5 \le z < .5$$

which shows that computing the exponent of any number is reduced to computing it for a corresponding value in the interval [-0.5ln2, 0.5ln2]. Thus the proper scale for computing the exponent function is s=0.5ln2. We now discuss three methods for economizing the exponent function.

4.5.1. Sum of sinh and cosh approach

Using the scales of the sinh and cosh, the exponent function can be expressed as

$$e^{sx} = \frac{1}{2}[\sinh(sx) + \cosh(sx)] \tag{4.15}$$

However, lack of symmetry in this approach will result in a large degree polynomial.

4.5.2. Hyperbolic tangent

The exponent function is transformed into a symmetric function as follows. Let

$$e^x = \frac{1+z}{1-z} \Rightarrow z = \frac{e^x - 1}{e^x + 1} = \frac{e^{x/2} - e^{-x/2}}{e^{x/2} + e^{-x/2}} = \tanh(x/2) \Rightarrow e^x = \frac{1 + \tanh(x/2)}{1 - \tanh(x/2)}.$$

Hence

$$e^{\alpha x} = \frac{1 + \tanh(\alpha x/2)}{1 - \tanh(\alpha x/2)} \tag{4.16}$$

Using α=s/2=ln2/4 in Eqs. (4.8) and (4.9) $e^{\alpha x}$ can be computed for any real number.

4.5.3. Hyperbolic cotangent approach

The above scale α=s/2=ln2/4 can be used to compute $f(\alpha x) = \alpha x \coth \alpha x$, from which we can compute the exponent function as follows: From Eq. (4.16) we get

$$e^{\alpha x} = \frac{\coth(\alpha x/2)+1}{\coth(\alpha x/2)-1} = \frac{(\alpha x/2)\coth(\alpha x/2)+\alpha x/2}{(\alpha x/2)\coth(\alpha x/2)-\alpha x/2} \tag{4.17}$$

The Chebyshev coefficients for the hyperbolic cotangent function are listed in Table 5. Listed also are the power series coefficients of the function for degrees 2, 4…10. The maximum error on the exponent function is determined below. Substitute in Eq. (4.17) with $f(x/2) = (x/2)\coth(x/2)$. The variation of the resulting expression yields

$$\delta e^x = -\frac{x\delta f}{2(f(\frac{x}{2})-\frac{x}{2})^2} = -\frac{2\delta f}{x(\coth(\frac{x}{2})-1)^2}$$

The maximum error occurs at $x/2$=s=ln2/2 which gives $\delta e^s \approx .25\delta f$. Thus any error in computing the exponent is of the same order as that of the cotangent function.

4.6. Hyperbolic arctangent function

The hyperbolic arctangent, atanh, and the logarithm functions are mutually dependent. We will exploit this dependence to determine the scales for each of them. We derive the Chebyshev expansion of atanh by utilizing its similarity with the arctangent function. By substituting $\alpha = i\beta$ in Eq. (3.24) and employing the identities

$$\tan i\alpha = i\tanh\alpha$$
$$\tan^{-1} i\beta = i\tanh^{-1}\beta$$

we get

$$\tanh^{-1}(\tanh\beta)x = 2\sum_{n=0} \frac{\tanh^{2n+1}(\beta/2)}{2n+1} T_{2n+1}(x), \quad -1 \le x \le 1 \tag{4.18}$$

The scale $\tanh\beta$ is related to the logarithm function and is determined next.

4.7. Logarithm function

The logarithm function is represented by the Taylor series expansion

$$\ln(1+y) = y - \frac{1}{2}y^2 + \frac{1}{3}y^3 - \frac{1}{4}y^4 + \cdots \tag{4.19}$$

Lack of symmetry of Eq. (4.19) will require more expansion terms to achieve a desired accuracy than those for symmetric functions. Thus the logarithm function is transformed into a symmetric function as follows: from Eq. (4.19) we get

$$\ln(1-y) = -(y + \frac{1}{2}y^2 + \frac{1}{3}y^3 + \frac{1}{4}y^4 + \cdots \tag{4.20}$$

Subtracting the above two equations

$$\ln\frac{1+y}{1-y} = 2\sum_{n=0}\frac{y^{2n+1}}{2n+1} = 2\tanh^{-1} y \tag{4.21}$$

To utilize Eq. (4.18) in the above, we substitute for $y = u\tanh\beta$ in Eq. (4.21) to get

$$\ln\frac{1+u\tanh\beta}{1-u\tanh\beta} = 2\tanh^{-1}(u\tanh\beta) = 4\sum_{n=0}\frac{\tanh^{2n+1}(\beta/2)}{2n+1}T_{2n+1}(u), \quad -1 \le u \le 1 \tag{4.22}$$

Since a real number w in a digital computer is represented by $w = x2^n$, n is an integer, then its logarithm is

$$\ln w = n\ln 2 + \ln x, \quad 1 \le x < 2 \tag{4.23}$$

which shows that the logarithm needs only be computed for $1 \le x < 2$. Now let

$$s\,x = \frac{1+u\tanh\beta}{1-u\tanh\beta} \Rightarrow u\tanh\beta = \frac{s\,x-1}{s\,x+1} \tag{4.24}$$

and substituting in Eq. (4.22) yields

$$\ln x = \ln \frac{1}{s} + 2\tanh^{-1}\left(u \tanh \beta\right), \quad -1 \le u \le 1, 1 \le x \le 2 \tag{4.25}$$

The above requires that we map the interval [1,2] of x to the interval [-1,1] of u. Therefore, from Eq. (4.24), for u=-1 and x=1 we get

$$-\tanh(\beta) = \frac{s-1}{s+1}$$

and for u=1 and x=2 Eq. (4.24) gives

$$\tanh(\beta) = \frac{2s-1}{2s+1}$$

Solving the above two equations results in,

$$\tanh(\beta) = \frac{\sqrt{2}-1}{\sqrt{2}+1} \qquad s = \frac{1}{\sqrt{2}}$$

Substituting for $\tanh(\beta)$ and s in Eq. (4.25) gives

$$\ln x = \ln \sqrt{2} + 2\tanh^{-1} \frac{x-\sqrt{2}}{x+\sqrt{2}}, \quad 1 \le x \le 2 \tag{4.26}$$

Chebyshev coefficients for the hyperbolic arctangent function are listed in Table 6. Listed also are its power series for degrees 1, 3…13.

5. Conclusions

The Chebyshev polynomials is a powerful tool for economizing transcendental functions that often results in minimal computations and uniform error distribution. Scaling these polynomials according to the given function can extend their usefulness as demonstrated for the trigonometric and hyper trigonometric functions. Nevertheless there are some functions that can not be efficiently economized as the arcsine/arccosine functions. Also there are other functions that can be efficiently economized via the use of other functions as in the exponent and the logarithm functions. The economization data for the sine, cosine, tangent, arc tangent, exponent and logarithm functions are provided in Tables 1-6.

APPENDIX

Chebyshev Recursion Equations

Rearranging the trigonometric identity

$$2\cos n\theta\cos\theta = \cos(n\theta+\theta)+\cos(n\theta-\theta) = \cos(n+1)\theta+\cos(n-1)\theta \tag{A.1}$$

yields

$$\cos(n+1)\theta = 2\cos\theta\cos n\theta - \cos(n-1)\theta \tag{A.2}$$

Substituting from Eq. (2.3) in Eq. (A.2)

$$T_{n+1}(\theta) = 2\cos\theta T_n(\theta) - T_{n-1}(\theta) \tag{A.3}$$

Using the transformation

$$x = \cos\theta \tag{A.4}$$

in Eq. (A.3) yields

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{A.5}$$

A power series as function of the Chebyshev polynomials is shown below. Let

$$u = e^{i\theta}, \quad v = e^{-i\theta} \tag{A.6}$$

The binomial theorem states

$$(u+v)^n = u^n + \binom{n}{1}u^{n-1}v + \ldots + \binom{n}{n-1}uv^{n-1} + v^n \quad \Rightarrow$$

$$(u+v)^n = (u^n+v^n) + \binom{n}{1}(u^{n-1}v+uv^{n-1}) + \binom{n}{2}(u^{n-2}v^2+u^2v^{n-2}) + \cdots$$

Substituting from Eq. (A.6) in the above gives

$$2^n \cos^n \theta = 2\cos n\theta + 2\binom{n}{1}\cos(n-2)\theta + 2\binom{n}{2}\cos(n-4)\theta + \cdots \tag{A.7}$$

Substituting from Eq. (A.4) and Eq. (2.3) in the above gives

$$2^n x^n = 2T_n(x) + 2\binom{n}{1}T_{n-2}(x) + 2\binom{n}{2}T_{n-4}(x) + \cdots \tag{A.8}$$

The last term in the RHS of Eq. (A.8) depends on whether n is odd or even, hence

$$2^n x^n = 2T_n(x) + 2\binom{n}{1}T_{n-2}(x) + \ldots + \binom{n}{\frac{n}{2}}T_0(x), \quad n \text{ is even}$$

$$2^n x^n = 2T_n(x) + 2\binom{n}{1}T_{n-2}(x) + \ldots + 2\binom{n}{\frac{n-1}{2}}T_1(x), \quad n \text{ is odd}$$

Equivalently the above two equations can be represented by the equations

$$x^{2n} = \frac{1}{2^{2n-1}}\sum_{k=0}^{n}\binom{2n}{n-k}T_{2k}H(k), \quad n = 0,1,2,\ldots \tag{A.9}$$

$$x^{2n+1} = \frac{1}{2^{2n}}\sum_{k=0}^{n}\binom{2n+1}{k}T_{2n-2k+1}(x), \quad n = 0,1,2,\cdots \tag{A.10}$$

where

$$\begin{aligned} H(k) &= 1, \quad k = 1,2,\cdots \\ &= .5, \quad k = 0 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{A.11}$$

Table 1 Chebyshev Coefficients of sin(x.π/4), -1⩽x⩽1

T_1	T_3	T_5	T_7	T_9	T_{11}	T_{13}
7.2637567669373460e-001	-1.9420029053201504e-002	1.5169292285107397e-004	-5.6058046841200102e-007	1.2053241678543556e-009	-1.6941393087095102e-012	1.6778093175966046e-015

Power Series Coefficients of sin(x), -π/4⩽x⩽π/4

n	x^1	x^3	x^5	x^7	x^9	x^{11}	x^{13}
1	9.2485023590022641e-001						
3	9.9902928275155212e-001	-1.6033938833998410e-001					
5	9.9999498989170432e-001	-1.6660157013147733e-001	8.1214933926091453e-003				
7	9.9999998616424701e-001	-1.6666636734719931e-001	8.3315839716068523e-003	-1.9462058181624460e-004			
9	9.9999999997624467e-001	-1.6666666589609330e-001	8.3333263328600488e-003	-1.9838672804571564e-004	2.7135316888187163e-006		
11	9.9999999999997213e-001	-1.6666666666540469e-001	8.3333331169595888e-003	-1.9841260735019872e-004	2.7554856370117920e-006	-2.4732065101915567e-008	
13	9.9999999999999989e-001	-1.6666666666666527e-001	8.3333333333083376e-003	-1.9841269821967074e-004	2.7557311570774417e-006	-2.5050482812758435e-008	1.5883056913369989e-010

Error bound on sin(x)

n	1	3	5	7	9	11	13
	1.9268895507187134e-002	1.5113354601436698e-004	5.5937683670697407e-007	1.2036317050184022e-009	1.6924628396502615e-012	1.6764678272216946e-015	1.2457436155176656e-018

Table 2 Chebyshev Coefficients of cos(x.π/4) , -1⩽x⩽1

T_0	T_2	T_4	T_6	T_8	T_{10}	T_{12}
8.5163191370480806e-001	-1.4643664439083687e-001	1.9214493118146466e-003	-9.9649684898292967e-006	2.7576595607187385e-008	-4.7399498081648420e-011	5.5495485414851792e-014

Power Series Coefficients of cos(x), -π/4⩽x⩽π/4

n	x^0	x^2	x^4	x^6	x^8	x^{10}	x^{12}
0	8.5163191370480806e-001						
2	9.9806855809564499e-001	-4.7478829242533427e-001					
4	9.9999000740745958e-001	-4.9970778280380668e-001	4.0397956174570697e-002				
6	9.9999997237594940e-001	-4.9999856557568295e-001	4.1655020905420913e-002	-1.3585843887443270e-003			
8	9.9999999995254496e-001	-4.9999999615145652e-001	4.1666616715874336e-002	-1.3886618604994089e-003	2.4379880313551010e-005		
10	9.9999999999994449e-001	-4.9999999999351513e-001	4.1666666543960736e-002	-1.3888880396024119e-003	2.4798929171725437e-005	-2.7173456841089967e-007	
12	1.0000000000000000e+000	-4.9999999999999267e-001	4.1666666666472389e-002	-1.3888888869983292e-003	2.4801578540009612e-005	-2.7555234093295848e-007	2.0630465643317002e-009

Error Bound on cos(x)

n	0	2	4	6	8	10	12
	1.4452513251826049e-001	1.9115118725763747e-003	9.9374392382718020e-006	2.7529251557494102e-008	4.7344049693282631e-011	5.5448388365786616e-014	4.7097049065175519e-017

Table 3 Chebyshev Coefficients of x.π/8cot(x.π/8) , -1⩽x⩽1

T_0	T_2	T_4	T_6	T_8	T_{10}	T_{12}
9.7409726717287404e-001	-2.5970025310420506e-002	-6.7541716104439387e-005	-2.5020150212295250e-007	-9.7249463554838568e-010	-3.8173451207019430e-012	-1.5017576781855927e-014

Power Series Coefficients of xcot(x) , -π/8⩽x⩽π/8

n	x^0	x^2	x^4	x^6	x^8	x^{10}	x^{12}
0	9.7409726717287404e-001						
2	1.0000672924832945e+000	-3.3680815407017939e-001					
4	9.9999975076719005e-001	-3.3330432988025988e-001	-2.2720743308627223e-002				
6	1.0000000009686922e+000	-3.3333353390088039e-001	-2.2215743042150880e-002	-2.1831349222007019e-003			
8	9.9999999999619760e-001	-3.3333333210261851e-001	-2.2222285902551153e-002	-2.1152508533200855e-003	-2.2009901470214537e-004		
10	1.0000000000000149e+000	-3.3333333334030790e-001	-2.2222221695625703e-002	-2.1164166428058939e-003	-2.1145944108258747e-004	-2.2409518727647284e-005	
12	9.9999999999999989e-001	-3.3333333333329634e-001	-2.2222222226069781e-002	-2.1164019667773465e-003	-2.1164297845290516e-004	-2.1351600229332416e-005	-2.2867104948562489e-006

Error Bound on xcot(x)

n	2	4	6	8	10	12
	2.6037818204354121e-002	6.7792893933616100e-005	2.5117782917671290e-007	9.7632705376042073e-010	3.8324182120349849e-012	1.5073091333041674e-014	5.5514551185745660e-017

Error Bound on tan(x)

	1.1376098246352341e-002	2.9619172226357848e-005	1.0974128629341012e-007	4.2656386940642206e-010	1.6744093440950972e-012	6.5855351832915799e-015	-2.4254681534169045e-017

Table 4 Chebyshev Coefficients of atan(x tan(.π/8)) , -1⩽x⩽1

T_1	T_3	T_5	T_7/T_{15}	T_9/T_{17}	T_{11}/T_{19}	T_{13}/T_{21}
3.9782473475931601e-001	-5.2467950438531983e-003	1.2455722454749680e-004	-3.5201766614312475e-006	1.0832870770174304e-007	-3.5068481349183947e-009	1.1740588439779745e-010
			-4.0259236096786315e-012	1.4055019102434098e-013	-4.9756583978892777e-015	1.7811816477512138e-016

Power Series Coefficients of atan(x) , -tanπ/8⩽x⩽tanπ/8

n	x^1	x^3	x^5	x^7/x^{15}	x^9/x^{17}	x^{11}/x^{19}	x^{13}/x^{21}
1	9.6043387010341996e-001						
3	9.9843452136500699e-001	-2.9531203542807216e-001					
5	9.9993806006897756e-001	-3.3036509848938383e-001	1.6344337884159002e-001				
7	9.9999754927664342e-001	-3.3313892658209754e-001	1.9577748863150146e-001	-1.0768971574516645e-001			
9	9.9999990303436137e-001	-3.3332184265314269e-001	1.9961549539561521e-001	-1.3751577271705970e-001	7.7261777546314608e-002		
11	9.9999999616344493e-001	-3.3333269857468339e-001	1.9996982390201859e-001	-1.4223617929487059e-001	1.0477432328425801e-001	-5.8310861036308892e-002	
13	9.9999999984820243e-001	-3.3333329991220600e-001	1.9999786271743944e-001	-1.4279648395433761e-001	1.1021714807652334e-001	-8.3689347159893165e-002	4.5512817510059506e-002
15	9.9999999993994403e-001	-3.3333333163567069e-001	1.9999985961478298e-001	-1.4285190667168293e-001	1.1100677028988537e-001	-8.9714118031158904e-002	6.8922776154370133e-002
				-3.6384863453783105e-002			
17	9.9999999999976241e-001	-3.3333333324946679e-001	1.9999999129728246e-001	-1.4285673096903709e-001	1.1110049717507468e-001	-9.0747084454347315e-002	7.5406466406530662e-002
				-5.7978986959008758e-002	2.9614064746582694e-002		
19	9.9999999999999067e-001	-3.3333333332928133e-001	1.9999999948468755e-001	-1.4285711272658866e-001	1.1111013904137161e-001	-9.0890131147222691e-002	7.6689139060161518e-002
				-6.4814154123893158e-002	4.9533201599576075e-002	-2.4441523691688023e-002	
21	9.9999999999999967e-001	-3.3333333333314108e-001	1.9999999997060344e-001	-1.4285714077811654e-001	1.1111102918781843e-001	-9.0907110559317439e-002	7.6892140631474926e-002
				-6.6346653758053542e-002	5.6626310004955688e-002	-4.2815597215545689e-002	2.0398466384482443e-002

Error Bound on atan(x)

n	1	3	5	7/15	9/17	11/19	13/21
	-5.1256530605918588e-003	1.2114198326133938e-004	-3.4152412861574160e-006	1.0493537527383149e-007	-3.3933324279115515e-009	1.1351570700684313e-010	-3.8901773909543192e-012
				1.3574621872431269e-013	-4.8039723000282888e-015	1.7168353976063915e-016	-6.4346250144822126e-018

REFERENCES

[1] C. Lanczos, Tables of Chebyshev polynomials, Applied Math. Series, U.S. Bur. Stand. 9, Washington, Government Printing Office, 1952.

[2] J.C. Mason and D.C. Handscomb, Chebyshev Polynomials, Chapman &Hill/CRC, Boca Raton, FA, 2003.

[3] J. Rivlin, Chebyshev Polynomials, John Wiley &Sons Inc., New York, NY, 1990.
[4] B.W. Arden and K.N. Astill, Numerical Algorithms: Origins and Applications, Addison-Wesley, Reading, MA, 1970.
[5] M.A. Snyder, Chebyshev Methods in Numerical Approximation, Prentice Hall, Englewood Cliffs, NJ, 1966.
[6] L. Fox and I.B. Parker, Chebyshev Polynomials in Numerical Analysis, Oxford University Press, London, 1968.
[7] C.W. Clenshaw, "A Note on the Summation of Chebyshev Series" http://www.ams.org/journals/mcom/1955-09-051/S0025-5718-1955-0071856-0/S0025-5718-1955-0071856-0.pdf
[8] Ying Guang Shi, Power Orthogonal Polynomials, Nova Science Publishers, New York, NY, 2006.
[9] M. Abramowitz and I.A. Stegun (Ed.), Handbook of Mathematical Functions, Dover Publications, Inc., New York, N.Y., 1970.
[10] G.P. Michon, Numericana.com, http://home.att.net/~numericana/.

Current Address: Esmat BEKIR, 5805 Serrania Avenue, Woodland Hills, California, USA, 91367
E-mail: ebekir@hotmail.com
Orcid ID: https://orcid.org/0000-0001-8500-5131

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 57-67 (2019)
DOI: 10.33769/aupse.493048
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

QUALITY OF SOFTWARE PROJECTS – A CASE STUDY

Alper KIRAL and Tülin ERÇELEBİ AYYILDIZ

ABSTRACT. In software projects, many points that are overlooked such as time constraints and human factors are causing great problems in the future. By measuring the quality of software projects, problems that may arise in important parameters such as maintenance-repair, functionality and reliability can be eliminated. In this study, metrics that can be used for measuring maintainability quality attribute within the scope of ISO 9126 Quality Standard are examined. In order to perform the study, 40 open source object-oriented software was selected and code complexity analysis was performed. Values of metric sets such as Chidamber and Kemerer (CK), Lorenz and Kidd (LK) and McCabe's complex Suite were determined by the Understand Code Analysis tool. It was determined whether the obtained values exceeded the threshold values indicated in the literature. Frequencies of metrics passing threshold values were determined for 40 open source object-oriented software projects, and the consistency among the metrics was evaluated using WEKA Machine Learning Software and EXCEL Data Analysis Tool. When the results were evaluated, it was observed that in addition to CK metrics such as WMC, CBO, and RFC, which measure the maintainability quality attribute, NOC (CK), NIM (LK), and the ratio of comment/code metrics have been observed to yield significant measurement results

1. INTRODUCTION

As the technology sector becomes a big part of daily life, the software used is constantly expanding as code and manpower. The growth in the software project leads to a significant increase in maintenance costs, project costs and software development time. If these and similar factors cannot be correctly predicted and carried out from the beginning, it is inevitable that problems that cannot be corrected afterwards are encountered. Economic loss of software projects that are rejected by customers, unavailable to use efficiently, canceled due to increased costs, require

Received by the editors: December 06, 2018; Accepted: May 14, 2019.
Key word and phrases: software quality metrics, software quality measurement, maintainability, CK, JK, McCabe, ISO 9126, Understand, WEKA.

high maintenance and repair costs may be far higher than those predicted. As an example,

according to 2002 data, the annual loss of failed software projects to the American economy is around $ 59 billion [1]. In addition, according to Tricentis 2017 research, the loss of failing software in the global economy is around $ 1.7 trillion [2]. By measuring the software quality; early decision can be made for factors that need to be calculated and implemented at early stages such as the ratio of customer needs met by software, the clarity of the software for the developers, the structural quality of the software and the cost and price balance of the software.

Since it is not possible to perform individual code analysis in large projects, there are tools and add-ons that can perform these analyzes in a short time. In this study, 40 open source object-oriented software projects written in Java programming language belonging to Space and Aviation domain were examined with static code analysis tool called Understand. In the scope of ISO 9126 quality standard, the object-oriented metric values which are recommended in the literature are calculated for maintainability quality characteristic and the metric probabilities which are possible to be used in addition to the literature are investigated with the help of WEKA machine learning software.

This study is based on our previous study which investigates the maintainability perspective of software quality metrics [3].

The rest of paper is organized as follows: Section II presents an overview of the related work. In Section III, the metric values analysis and metric threshold exceeding frequencies of 40 open source object-oriented software projects are evaluated. Section IV discusses additional metrics that can be used under the Maintainability quality characteristic. Section V concludes with research results..

2. Related works

A. ISO 9126 QUALITY STANDARD

For software quality measurement, various metric clusters are presented by people working in this field. With the help of these metrics, quality requirements are measured. Quality requirements for the ISO 9126 quality standard are shown in Table I [4].

TABLE I. ISO 9126 QUALITY STANDARET AND CHARACTERISTICS[4]

Characteristics	Sub-characteristics
Functionality	Suitability, accuracy, interoperability, security
Reliability	Maturity, fault tolerance, recoverability
Usability	Understandability, operability, attractiveness etc.
Efficiency	Time behavior, resource utilization
Maintainability	Analyzability, changeability, stability, testability
Portability	Adaptability, installaebility, replaceability e.g.

B. SUGGESTED METRICS FOR MAINTAINABILITY QUALITY CHARACTERISTICS MEASUREMENT

One of the metric sets proposed in the literature due to the sub-characteristics of the maintainability attribute includes all Chidamber and Kemerer metrics and additionally the TCC, LCC metrics. [5]. Characteristics, sub-characteristics and metrics are shown in Fig. 1.

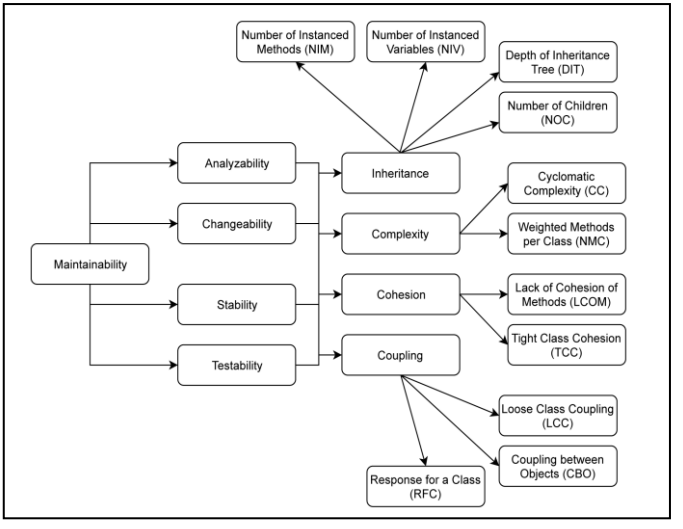


Fig. 1. Maintainability Characteristic and object-oriented metrics

Another suggestion includes WMC, CBO metrics from the CK metric set for the maintainability characteristics measurement, and Table II shows the CK metrics that are suitable for the design phase quality attributes [6].

TABLE II. RELATIONSHIP BETWEEN DESIGN PHASE QUALITY ATTRIBUTES AND CK METRICS

Quality Attribute	CK Metrics
Maintainability	WMC, CBO
Reusability	WMC, CBO, DIT, NOC
Testability	RFC, CBO, NOC
Understandability	RFC, CBO, DIT
Development Effort	WMC, LCOM

In this study, by considering WMC, CBO metrics and the importance of machine learning and other metrics, ISO 9126 Standard Maintainability Characteristics will be investigated in the light of Table II and Fig 1. As a result, it is expected that metrics such as DIT, NOC, RFC for reusability and testability attributes will be some of the recommended metrics for measuring the maintainability attribute.

3. PROJECT MEASUREMENTS AND THRESHOLD EVALUATION

Within the scope of the study, 40 open source object-oriented software project written in Java was downloaded from GitHUB and NASA Open Source Software Library. Medium scale projects were chosen for the study, the related projects and total KLOCs are shown in Table III.

TABLE III. PROJECTS AND LINE OF CODES

#	KLOC	#	KLOC	#	KLOC	#	KLOC	#	KLOC
1	13K	9	52K	17	11K	25	114K	33	25K
2	21K	10	29K	18	44K	26	199K	34	78K
3	11K	11	17K	19	10K	27	124K	35	166K
4	36K	12	49K	20	75K	28	50K	36	20K
5	22K	13	49K	21	15K	29	119K	37	27K
6	54K	14	19K	22	15K	30	68K	38	42K
7	12K	15	14K	23	42K	31	115K	39	10K
8	16K	16	18K	24	63K	32	134K	40	32K

A. DETERMINING THRESHOLD VALUES

Various threshold values were calculated for metrics and used in software quality measurement. Some recommended threshold values for CK metrics are shown in Table IV.

TABLE IV. THRESHOLDS FOR CK METRIC SUITE

	CK Metrics					
Related Works	**LCOM**	**DIT**	**CBO**	**NOC**	**RFC**	**WMC**
[7]	3	6	9	3	6	30
[8]	1	6	8	6	35	15
[9]	Low	4	8	6	35	11
[10]	20	2	14	2	44	20
[11]	Low	4	94	5	10	108
[12]	Low (%85.6)	6	10	14	62	25

Since Reference [12] derives CK metric thresholds over projects in the same domain, this study was continued by using these metric thresholds.

From the McCabe's Complexity Suite metric set, the Cyclomatic Complexity (CC) metric threshold is set to 10, and the Essential Complexity (EC) metric threshold is set to 4 [13-16]. Also recommended Ratio Code/Comment (Ratio C/C) metric threshold value is set to 0.16 [16].

NIV and NIM metrics, which are class metrics from the Lorenz and Kidd metric set, have been selected [14]. A value of 0.8 for the NIM metric threshold and 9 for the NIV metric threshold is suggested [15].

B. CLASS FREQUENCIES EXCEEDING THE THRESHOLD VALUES IN PROJECTS

For the threshold values specified in the previous section, the ratio of the class numbers exceeding the threshold values to the total class numbers in the projects is shown in Table V. Reference [12] showed that 10% of the frequencies did not exceed much for this study domain. Therefore, projects with CBO and WMC frequency less

than 10% for WEKA machine learning operations are classified as 0 (successful) and those above as 1 (unsuccessful).

TABLE V. FREQUENCIES FOR CLASSES EXCEED THRESHOLD VALUES

Project #	CBO	NOC	RFC	DIT	LCOM	WMC	CC	EC	Ratio C/C	NIM	NIV	Classification
1	0.05	0.00	0.00	0.00	0.06	0.05	0.04	0.01	0.86	0.15	0.21	1
2	0.02	0.00	0.00	0.00	0.06	0.08	0.02	0.02	0.32	0.14	0.13	1
3	0.08	0.04	0.32	0.00	0.14	0.13	0.02	0.02	0.83	0.01	0.11	0
4	0.06	0.00	0.04	0.00	0.13	0.15	0.01	0.02	0.81	0.13	0.14	0
5	0.03	0.00	0.05	0.00	0.13	0.08	0.02	0.02	0.59	0.05	0.10	1
6	0.04	0.00	0.02	0.00	0.06	0.06	0.00	0.02	0.76	0.16	0.05	1
7	0.02	0.00	0.00	0.00	0.02	0.03	0.02	0.02	0.66	0.31	0.06	1
8	0.05	0.00	0.00	0.00	0.11	0.05	0.01	0.03	0.84	0.19	0.13	1
9	0.12	0.00	0.01	0.00	0.06	0.08	0.00	0.02	0.81	0.13	0.05	0
10	0.03	0.00	0.00	0.00	0.07	0.08	0.01	0.01	0.21	0.13	0.07	1
11	0.02	0.00	0.00	0.00	0.07	0.02	0.03	0.02	0.88	0.24	0.12	1
12	0.09	0.00	0.00	0.00	0.04	0.07	0.04	0.07	0.54	0.14	0.10	1
13	0.02	0.00	0.01	0.00	0.09	0.06	0.01	0.03	0.69	0.04	0.10	1
14	0.01	0.00	0.01	0.00	0.08	0.10	0.02	0.02	0.71	0.08	0.12	0
15	0.01	0.00	0.00	0.00	0.08	0.05	0.02	0.03	0.88	0.30	0.05	1
16	0.05	0.00	0.00	0.00	0.07	0.05	0.01	0.02	0.94	0.14	0.07	1
17	0.01	0.00	0.00	0.00	0.09	0.07	0.04	0.04	0.77	0.23	0.20	1
18	0.06	0.01	0.00	0.00	0.05	0.05	0.01	0.02	0.84	0.13	0.05	1
19	0.03	0.00	0.00	0.00	0.09	0.06	0.02	0.03	0.87	0.20	0.09	1
20	0.12	0.01	0.01	0.00	0.11	0.07	0.04	0.04	0.09	0.04	0.15	0
21	0.01	0.00	0.00	0.00	0.12	0.08	0.01	0.01	0.45	0.38	0.10	1
22	0.10	0.00	0.15	0.00	0.07	0.10	0.05	0.01	0.37	0.05	0.18	0
23	0.08	0.01	0.01	0.00	0.07	0.08	0.03	0.05	0.48	0.06	0.18	1
24	0.04	0.00	0.01	0.00	0.08	0.10	0.04	0.06	0.31	0.12	0.15	1
25	0.12	0.00	0.01	0.00	0.04	0.04	0.00	0.02	0.71	0.10	0.04	0
26	0.09	0.01	0.25	0.00	0.06	0.07	0.01	0.04	0.79	0.08	0.05	1
27	0.11	0.01	0.24	0.01	0.09	0.07	0.02	0.05	0.79	0.21	0.04	0
28	0.08	0.00	0.03	0.00	0.12	0.06	0.01	0.02	0.62	0.07	0.07	1
29	0.10	0.00	0.03	0.00	0.14	0.11	0.06	0.06	0.55	0.12	0.12	0
30	0.03	0.00	0.01	0.00	0.10	0.09	0.05	0.05	0.70	0.18	0.23	1
31	0.03	0.01	0.00	0.00	0.08	0.05	0.03	0.03	0.78	0.14	0.08	1
32	0.02	0.00	0.01	0.00	0.11	0.08	0.02	0.02	0.76	0.16	0.15	1
33	0.03	0.00	0.01	0.00	0.07	0.07	0.01	0.02	0.88	0.10	0.10	1
34	0.02	0.00	0.00	0.00	0.08	0.06	0.03	0.01	0.94	0.11	0.14	1
35	0.20	0.01	0.01	0.00	0.10	0.06	0.01	0.01	0.33	0.07	0.12	0
36	0.15	0.00	0.01	0.00	0.07	0.06	0.01	0.01	0.41	0.04	0.08	0
37	0.02	0.00	0.02	0.00	0.09	0.10	0.02	0.02	0.59	0.16	0.18	0
38	0.00	0.00	0.00	0.00	0.23	0.02	0.00	0.00	0.96	0.02	0.01	1
39	0.07	0.01	0.09	0.00	0.14	0.11	0.03	0.06	0.30	0.10	0.22	0
40	0.01	0.00	0.00	0.00	0.02	0.01	0.00	0.01	0.10	0.02	0.26	1

The ANOVA table of multi regression test which belongs to Table V is shown in Table VI. Classifications were used as criterion variables and metric frequency values were used as predictors for the regression test. Sig. F value shows that classification with WMC and CBO metrics is significant for this research.

TABLE VI. ANOVA TABLE

ANOVA					
	df	**SS**	**MS**	**F**	**Sig. F**
Regression	11	5.81	0.53	4.98	0.00029
Residual	28	2.97	0.11		
Total	39	8.78			

4. ESTIMATING NEW MAINTAINABILITY METRICS

40 open source object-oriented software projects are classified according to the amount in which they exceed WMC and CBO metrics as in Table V. These classifiers are included in WEKA machine learning software.

A. ATTRIBUTE SELECTION PHASE

By using classified projects, the correlation and significance between the metrics were investigated by correlation attribute selection evaluation option in WEKA.

TABLE VII. SELECTED METRICS BY WEKA CORRELATION RANKING RESULT

Attribute Evaluator (supervised, Class (nominal): 12 maintainability): Correlation Ranking Filter	
Ranked attributes:	
Score:	**Metric:**
0.637	CBO
0.503	WMC
0.37	RFC
0.306	NOC
0.303	NIM
0.237	Ratio C/C
Selected attributes	1, 6, 3, 2, 10, 9 : 6

6 metrics are chosen to be used in the next phase by ranker selection method as a result of machine learning process. Table VII shows these results. It is already

foreseen that the first two orders should be CBO and WMC. Remaining four recommendations will be examined in following sections.

B. TRAINING AND TESTING PHASE

In the WEKA Classification section, data belonging to the previous WMC and CBO based classification were used as training data and machine learning was performed. By using CBO and WMC metrics a Naïve Bayes classification was made in WEKA and the results are shown in Table VIII. Projects are correctly classified by 80% accuracy.

TABLE VIII. WEKA NAIVE BAYES CLASSIFICATION RESULTS FOR WMC AND CBO METRICS

Correctly Classified Instances			32		80%			
Incorrectly Classified Instances			8		20%			
Kappa Statistics			0.5616					
Mean Absolute Error			0.2053					
Root Mean Squared Error			0.3967					
Relative Absolute Error			46.1509%					
Root Relative Squared Error			84.1191%					
Total Number of Instances			40					
Detailed Accuracy By Class								
TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
0.769	0.185	0.667	0.769	0.714	0.565	0.865	0.709	0
0.815	0.231	0.880	0.815	0.846	0.565	0.863	0.935	1
0.800	0.216	0.811	0.800	0.803	0.565	0.864	0.862	**Weighted Average**
Confusion Matrix								
a	**b**	← Classified As						
10	**3**	**a=0**						
5	**22**	**b=1**						

Afterwards, RFC, NOC, NIM, Ratio C / C metrics were added to CBO and WMC and a new classification was made by Naïve Bayes method too. Results for this classification are shown in Table IX. Projects are correctly classified by 90% accuracy by WEKA. In terms of maintainability, two projects that were previously

marked as successful were considered unsuccessful and two projects that were previously unsuccessful were classified as successful.

TABLE IX. WEKA NAIVE BAYES CLASSIFICATION RESULTS FOR WMC, CBO, RFC, NOC, NIM AND RATIO C/C METRICS

Correctly Classified Instances			36		90%			
Incorrectly Classified Instances			4		10%			
Kappa Statistics			0.7721					
Mean Absolute Error			0.1352					
Root Mean Squared Error			0.288					
Relative Absolute Error			30.6223%					
Root Relative Squared Error			61.4892%					
Total Number of Instances			40					
Detailed Accuracy By Class								
TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
0.846	0.074	0.846	0.846	0.846	0.772	0.943	0.872	0
0.926	0.154	0.926	0.926	0.926	0.772	0.943	0.975	1
0.900	0.128	0.900	0.900	0.900	0.772	0.943	0.942	**Weighted Average**
Confusion Matrix								
a	**b**	← Classified As						
11	**2**	**a=0**						
2	**25**	**b=1**						

5. CONCLUSIONS

The key contribution of this study is the investigation of the software metrics that are required to measure maintainability characteristic. After researching scientific literature for the recommended metrics to measure the maintainability characteristic, 40 object-oriented open source projects examined with the aid of machine learning tools.

It is seen that WEKA machine learning software has the correct classification with 90% accuracy with new six metrics which is better than the classification with WMC and CBO metrics that have 80% accuracy. Having two projects that are successful and misclassified is considered an acceptable result as false positives. Having small number of true negatives does not hinder our judgement about the project. If general

accuracy is high, common procedure of reviewing the negatives would only result in diminishing risks rather than the waste of time.

On the other hand, false positives might affect the projects at hand by misguiding the review process thus, increasing risk. This misclassification should remain minimal with increasing number of projects.

As a result, with 90% accuracy score of classification, it is observed that the usage of CBO, WMC, RFC, NOC, NIM and Ratio C/C metrics in the measurement of maintainability characteristic of ISO 9126 Quality Standard gives consistent results. As a future work, we are planning to increase the number of projects and a more comprehensive machine learning study can be carried out and it can be discussed whether or not to add or remove existing metrics.

References

[1] National Institue of Standards and Technology, 'The Economic impacts of inadequate infrastructure for software testing', 2002. [Online]. Available: https://www.nist.gov/sites/default/files/documents/director/planning/report02-3.pdf. [Accessed: 10- Oct- 2018].

[2] Tricentis.com, 'software fail watch 5th edition', 2017. [Online]. Available: https://www.tricentis.com/wp-content/uploads/2018/01/20180119_Software-Fails-Watch_Small_Web.pdf. [Accessed: 10- Oct- 2018].

[3] Kıral, A., Ayyıldız, T.E., Software Metrics Required to Measure Maintainability, *III. International Conference on Theoretical and Applied Computer Science and Engineering* (ICTACSE 2018), Ankara, TURKEY, November 23-24, 2018.

[4] Jung H., Kim S., and Chung C., "Measuring software product quality: *a survey of ISO/IEC* 9126", September/October 2004, IEEE Software, 88-92.

[5] Y. Fitrisia and B. Hendradjaya, "Implementation of ISO 9126-1 quality model for asset inventory information system by utilizing object oriented metrics," 2014 *International Conference on Electrical Engineering and Computer Science* (ICEECS), Kuta, (2014), 229-234.

[6] U. L. Kulkarni, Y. R. Kalshetty and V. G. Arde, "Validation of CK metrics for object oriented design measurement," *3rd International Conference on Emerging Trends in Engineering and Technology*, Goa, (2010), 646-651.

[7] Antony P., "Predicting reliability of software using thresholds of CK metrics", International Journal of Advanced Networking, (2013) 1778–1785.

[8] Yuming Zhou and Hareton Leung, "Empirical analysis of object-oriented design metrics for predicting high and low severity faults," *IEEE Transactions on Software Engineering,* vol. 32, no. 10, , Oct. 2006, 771-789.

[9] Mago J., Kaur P., "Analysis of quality of the design of the object oriented software using fuzzy logic." *International Conference on Recent Advances and Future Trends in Information Technology* (iRAFIT2012) *Proceedings Published in International Journal of Computer Applications*® (IJCA), (2012), 21–25.
[10] A.D. Bakar, A. Sultan, H. Zulzalil and J. Din, "Predicting maintainability of object-oriented software using metric threshold." *Information Technology Journal*, 13 (2014) 1540-1547.
[11] S. Singh, M. Kaur, "Deriving and validating software metrics threshold values for design errors" *International Journal of Engineering Technology and Scientific Research* Volume 2 Issue 2, (2016).
[12] Kıral, A., Ayyıldız, T.E., "Comparison of software quality metrics: a case study." 2018, *12th Turkish National Software Engineering Symposium* (UYMS'18), Istanbul, TURKEY, 10-12 September, in press.
[13] K. Yamashite et al., "Thresholds for size and complexity metrics: a case study from the perspective of defect density," 2016 *IEEE International Conference on Software Quality, Reliability and Security* (QRS), Vienna, (2016), 191-201.
[14] G. Marcela et al. "A survey of metrics for UML class diagrams." *Journal of Object Technology* 4 (2005), 59-92.
[15] S. Demeyer, S. Ducasse and M. Lanza, "A hybrid reverse engineering approach combining metrics and program visualisation," *Sixth Working Conference on Reverse Engineering* (Cat. No.PR00303), Atlanta, GA, USA, (1999), 175-186.
[16] Mccabe.com, "All metrics thresholds in McCabe iq", [Online]. Available: http://www.mccabe.com/pdf/McCabe%20IQ%20Metrics.pdf. [Accesed: 7- Oct-2018].

Current Address: Alper KIRAL: Başkent University, Department of Computer Engineering, Ankara, 06790, Türkiye
E-mail: alperkiral@gmail.com
Orcid ID: https://orcid.org/ 0000-0002-4018-0419
Current Address: Tülin ERÇELEBİ AYYILDIZ: Başkent University, Department of Computer Engineering, Ankara, 06790, Türkiye
E-mail: ercelebi@baskent.edu.tr
Orcid ID: https://orcid.org/ 0000-0002-7372-0223

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 68-75 (2019)
DOI: 10.33769/aupse.525368
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

ESTIMATING CO2 EMISSIONS BY USING ENERGY INTENSITY DATA OF OECD COUNTRIES

Recep ERYİĞİT and Semra GÜNDÜÇ

ABSTRACT. It is discussed that economic development has an essential effect on the country's CO2 emission which plays an important role in global warming. In this research well-known machine learning algorithm Extreme Learning Machine, ELM, is used to investigate the relationship between CO2 emission and energy intensity for countries in OECD. The results indicate a strong correlation and the method perform well for estimation.

1. INTRODUCTION

The rising amount of Greenhouse gases in the atmosphere contribute one of the most critical problems that we have today, global warming. Even though there are many sources of greenhouse gases, CO_2 is heading with a contribution of more than 60% in total [1]. For this reason, to concern global warming it is essential to reduce CO_2 emission. Since pre-industrial revolution period, it is clear that CO_2 emission has increased intensively [2], and correspondingly result in an increase in the global temperature approximately 1.3 degrees [3].

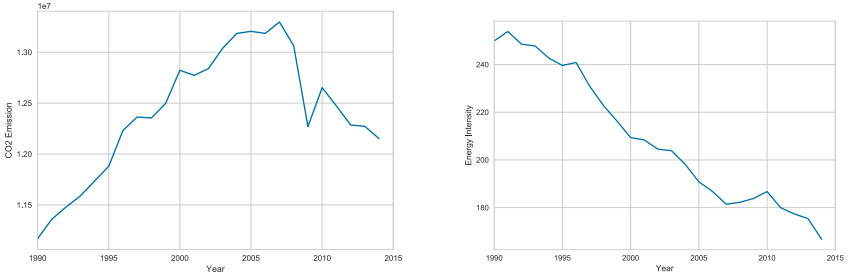
The population of the world is also increasing. As a result, energy demand (mainly produced from fossil sources), pollution, usage of natural sources and deforestation (essential for the absorption of CO_2), is increasing. It is clearly discussed in [4] that human activities have many effects on the environment.

Also, the development in the economy for many countries is rising. Besides, Gross Domestic Product, (GDP), mostly used as a metric to measure this development, is increasing. For this reason the studies which aim the relation of many factors, mostly economic, with CO_2 emission has become popular in the last decades. In [6] it is supposed that economic output is the main driving factor for CO_2 emission. The

Received by the editors: February 11, 2018; Accepted: May 24, 2019.
Key word and phrases: Extreme Learning Machine (ELM), Energy Intensity, CO2 emission.

work [7] says that about half of the increase in CO_2 emission is due to economic growth.

The total amount of CO_2 emission of the countries in Organisation for Economic Co-operation and Development (OECD) in years from 1990 to 2015 is introduced in Figure 1a. The data is taken from the World Data Bank [8]. The effect of strategies which are taken by policymakers to decrease the CO2 emission is seen in the figure since 2005.



(a) (b)

Figure 1. (a)Total CO2 emission values (b) Total Energy Intensity values, of OECD Countries in the years between 1990 and 2015.

Both increasing energy usage/demand and economic development are considered as main factors affecting the CO2 emission in the literature. Besides, in many countries, there is an enormous effort to decrease the damage to the natural environment. For this reason, to use energy as efficiently as possible a new metric can be considered to measure the amount of unit energy per unit of GDP which defined as energy intensity. High energy intensity value means the high cost for converting this energy to GDP. Countries in (OECD) have been decreased energy intensity considerably in the years between 1990 and 2015 which is an indication of a great effort to reduce the CO_2 emission. Figure 1b shows the total energy intensity of OECD countries as a function of years.

All in all such efforts are needed. The cause of emission mostly depend on national sources of countries and therefore shows a great variety in different geographical regions. Also, some costly and long-term investment is required. For this reason,

OECD, EU, even governments are supposed to construct radical policies related to the reduction of CO_2 emission.

For many reasons, such as advising to policymakers, for future planning, etc., in the literature, there is an increasing amount of work which aims to find the effects and also the relation between some parameters and CO_2 emissions, which are mainly nonlinear. However, the lack of driving mechanisms of such a process forces to use soft programming techniques to reach a significant result. Also increasing power and advances in technology allow to generate and collect data at an incredible rate. This situation results in having an extensive data set in size and dimension. This truth has an important effect to build new and more efficient computational methods than ever.

To that end in this work to investigate the CO_2 emission by considering the energy intensity of OECD countries, machine learning algorithm ELM is used. In the next section, the details of the method and the results obtained by using this method are introduced.

2. Model

In this paper, it is aimed to find out the relationship between energy intensity and the amount of CO_2 emission. In the literature conventional neural network algorithm is frequently used in many research for estimation and building a correlation between parameters of the systems [9]. Because of updating all parameters at every iteration in the neural network algorithm, the time used for calculation is very high. Also in many cases, there is no analytical approach to the driving mechanism of the process. For this reason, soft computing techniques introduce a unique way to get the solution. Therefore to find out the relation between the energy intensity and the amount of CO2 emission, Extreme Learning Machine (ELM) [8] algorithm, based on Single-Layer Neural Networks, is used in this work.

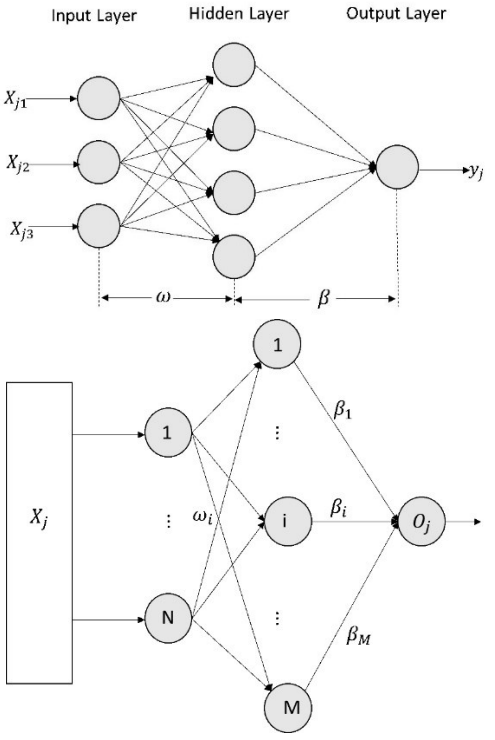


Figure 2. A schematics representation of ELM.

ELM algorithm was first introduced in [8] for training Single-Layer Feed forward Neural Networks (SLFN) and since used in many researches [10,11] because of its success and fast learning speed. A schematic representation is given in Figure 2.

For N arbitrary distinct samples (x_i ,y_i) the output of an SLFN which have M hidden neurons can be mathematically modelled as,

$$f_M(x) = \sum_{i=1}^{M} \beta_i \, g_i(x_j) = \sum_{i=1}^{M} \beta_i \, g(w_i \, . \, x_j + b_i) = O_j, \qquad j = 1, \dots, N$$

Here b_i is the threshold of the ith hidden nodes, $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the weight vector between ith hidden node and input nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{in}]^T$ is the weight vector between output and hidden nodes.

In its original article [8], these equations are written in an compact form as

$$\boldsymbol{H}\beta = \boldsymbol{T}$$

where

$$\boldsymbol{H} = \begin{bmatrix} g(w_1\boldsymbol{x_1} + b_1) & \cdots & g(w_M\boldsymbol{x_1} + b_M) \\ \vdots & \cdots & \vdots \\ g(w_1\boldsymbol{x_N} + b_1) & \cdots & g(w_M\boldsymbol{x_N} + b_M) \end{bmatrix}_{N*M}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M*m} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_M^T \end{bmatrix}_{N*m}$$

$$\beta = \mathbf{H}^{\dagger}\boldsymbol{T}$$

H is called the hidden layer output matrix of the neural network.

In the calculations to understand how well future samples are likely to be predicted by the model, the coefficient of determination is used which provides a measure. Having n different samples, for the ith sample, if the true value of the variable is y_i and its predicted value is, the coefficient of determination is defined as,

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}$$

here $\overline{y}$ is

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The value of the coefficient of determination between 0 and 1 means that the independent variable can be predicted, 0 means cannot be predicted and 1 means can be predicted without error from the dependent variable.
For each estimation, we also calculated the mean absolute error, MAE, over n sample, is defined in Equation below, which is a measure of the average of the absolute errors

$$MAE(y, \widehat{y}) = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

The results are introduced in the next section.

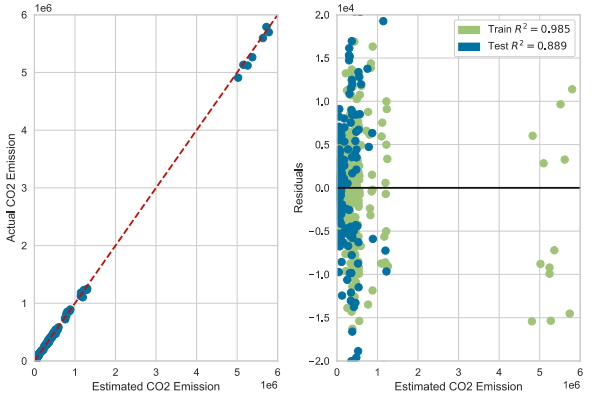
3. Results and discussion

In this work, ELM algorithm is used to estimate the amount of CO2 emission from energy intensity values for OECD countries. The energy intensity value is used as an input parameter and amount of CO2 emission is estimated as output. The results including the coefficient of determinations and MAE are introduced in Table 1.

Table 1. Coefficient of determination and Mean Absolute Error for ELM.

Method	R^2 Test	R^2 Train	Mean Absolute Error
ELM	0.8891	0.9847	8947

As it is seen, the value of coefficient of determination indicates that energy intensity, input value, is a suitable variable to estimate the CO_2 emission, output value, and ELM algorithm perform well for this estimation. Fig.3 shows the graphical representation of prediction of CO2 emission by considering the energy intensity under ELM algorithm. In the graphic, the left figure shows the prediction and the right one shows the residuals. The residuals, difference between the observed value and the estimated value, are also considered in this work to have a sense of how accurate our model is by relating the amount of CO2 emission with energy intensity. The countries in OECD have two main behaviors about CO2 emission which is seen clearly in Figure 3a that some have low-level emission, some have high but no middle. For this reason, residual plot is affected this having lack of some data region but still give good results.



(a) (b)

Figure 3. (a) Estimation of CO_2 emission by using ELM and (b) Residuals of estimation.

References

[1] I. Ozturk and A. Caravci, CO2 emissions, energy consumption and economic growth in Turkey. Renew Sustain Energy Rev 2010;14:3220–5.

[2] BP. BP statistical review of world energy June 2016; 2016. http://www.bp.com/statisticalreview.

[3] SJ. Davis, K. Caldeira and HD. Matthews, Future CO2 emissions and climate change from existing energy infrastructure. Science 2010;329:1330–3.

[4] P.R. Ehrlich and J.P. Holdren, Impact of population growth. Science 1971, 3977, 1212–1217.

[5] A. Shi, The impact of population pressure on global carbon dioxide emissions, 1975–1996: Evidence from pooled cross-country data. Ecol. Econ. 2003, 1, 29–42.

[6] M. Wang and C. Feng, Decomposition of energy-related CO2 emissions in China: an empirical analysis based on provincial panel data of three sectors. Appl Energy 2017;190:772–87.

[7] B. Lin and H. Liu, CO2 emissions of China's commercial and residential buildings: Evidence and reduction policy. Build Environ 2015;92:418–31.

[8] www.worlddatabank.com

[9] J. Long, L. Shuai, H. Bin and L. Mei, A survey on projection neural networks and their applications Applied Soft Computing, Volume 76, 2019, Pages 533-544

[10] G.B. Huang, Q.Y. Zhu and C.K. Siew, Extreme learning machine:Theory and applications, Neurocomputing 70 (2006a) 489501

[11] S. Gang and D. Qun, A novel double deep ELMs ensemble system for time series forecasting , Knowledge-Based Systems, Volume 134, 2017, Pages 31-49.

[12] K. Marius, Y. Yang, L. Caihong, C. Yanhua and L. Lian Mixed kernel based extreme learning machine for electric load forecasting Neurocomputing, Volume 312, 2018, Pages 90-106.

Current Address: RECEP ERYİĞİT: Ankara University, Computer Engineering Department, Gölbaşı Ankara Turkey
E-mail: eryigit@eng.ankara.edu.tr
Orcid ID: https://orcid.org/0000-0002-4282-6340
Current Address: SEMRA GÜNDÜÇ: Ankara University, Computer Engineering Department, Gölbaşı Ankara Turkey
E-mail: gunduc@.ankara.edu.tr
Orcid ID: https://orcid.org/0000-0002-3811-9547

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 76-101 (2019)
DOI: 10.33769/aupse.547407
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

FOCUSING, IMAGING AND FOURIER TRANSFORMING WITH A LARGE NUMERICAL APERTURE, DIELECTRIC FLAT PHOTONIC CRYSTAL LENS IN METAMATERIAL REGIME

EVRIM COLAK, ATILLA OZGUR CAKMAK

Abstract. Traditional Gradient Index Photonic Crystals (GRIN PCs) are designed under the assumption of the paraxial wave approximation and the slowly changing refractive index. In contrast, in this study, an untraditional Gradient Index Photonic Crystal (GRIN PC) is employed within the long wavelength regime to exhibit focusing, imaging and Fourier Transforming. A very large numerical aperture is aimed by breaking both the paraxial wave approximation and the slowly changing refractive index assumption. This untraditional GRIN PC is shown to exhibit very similar lensing characteristics as its analytically solvable traditional counterparts demonstrate. The results suggest that Fourier Transforming together with Fractional Fourier Transforms can be obtained from the present design. The performance of the GRIN PC lens is very much dependent on the excitation source as it is expected from GRIN optics and adiffractional propagating beams over large distances can be obtained.

1. INTRODUCTION

Metamaterials (MTM) have emerged with the promise of building superlenses as early as in 2000s. Such lenses would exceptionally recapture the high spatial frequencies to reconstruct the subwavelength features [1]. There have been several demonstrations of subwavelength focusing [2- 4] and cloaking [5,6] where simultaneous negative dielectric permittivity (ε_r) and magnetic permeability (μ_r) could be observed. The difficulty of obtaining simultaneous negative electromagnetic constitutive parameters with moderate optical losses compelled the researchers to shift to hyperbolic MTMs which are also volumetric, chiral structures and plasmonic counterparts [7-11]. Anisotropic nature of the MTM mimicked the negative refraction for hyperbolic MTMs whereas the bianisotropy helped to relieve the strict requirement (simultaneous negative ε_r and μ_r) on the negative refraction for resonating chiral structures. Plasmonic lenses were found to be the solution to

Received by the editors: April 01, 2018; Accepted: May 27, 2019.
Key word and phrases: Graded Index Photonic Crystal (GRIN) lens, fractional Fourier transform, Fourier optics, adiffractional beam propagation.

reach below 100nm resolution in lithographic methods to be utilized as an active element of the exposure systems [12]. In the following years, metasurfaces specifically metalenses were commonly employed to end up with superior imaging performances that could battle chromatic aberrations, as well [13,14]. Hence, volumetric MTM structures seemingly left their place to wavefront modification based lower dimensional devices.

On the other hand, while still being volumetric, dielectric based Graded Index Photonic Crystals (GRIN PCs) have been proposed in order to build lenses that would be immune to non-idealities stemming from parasitic losses and resonator configurations [15,16]. Since then the GRIN PCs have been efficiently shown to guide the light carefully to serve as waveguide input couplers [17], to slow down the light [18] and to achieve self-collimation, super-bending and mirage effects [19-22]. The main ideas have been borrowed from the already well-established GRIN optics as the lattice configurations have been carefully adjusted to mimic GRIN media [23].

In this study, a PC with a quadratic index profile will be analyzed in depth to illustrate the focusing properties of a full dielectric configuration for a most general case, i.e., outside the constraints of a paraxial approximation and the assumption of a slowly changing index profile. The proposed GRIN PC's imaging performance will be studied while staying inside the metamaterial regime. Metamaterial regime is attributed to be the operating wavelength range where the lattice spacings of the designed PC is much smaller than the incident light's wavelength. The present work will start with the analytical investigation of a perfect lens. The underlying evanescent wave amplification mechanism will be studied with the help of Transfer Matrix Method (TMM). The historical difficulties arising from the creation of such a perfect lens will be presented to the reader and the GRIN PC configuration, which is based on the homogenization methods will be examined. To the best of our knowledge, a prior analysis of the homogenization procedure for PC based GRIN lenses for the most general case without the paraxial wave approximation or the gradual refractive index change have not been carried out before even though several self-focusing (SELFOC) GRIN medium have been thoroughly examined. The present paper serves as a bridge between the GRIN optics and GRIN PCs while adapting the Maxwell-Garnett theory to study the focusing and imaging phenomenon in depth. The paper will attempt to explain ray optics, imaging, Fourier optics, bandwidth management and source dependency of the current GRIN PC by forming links with the well-established GRIN optics.

2. METAMATERIAL BASED DESIGNS

Subwavelength focusing has been accepted as a pioneering accomplishment for possible future applications, especially in the near-field science. The appearance of the negative-index materials (NIM) has escalated the endeavors towards the construction of improved lenses. MTMs and PCs have been the fundamental building blocks of such lensing activities. The main physic entails the negative refraction and evanescent wave amplification.

An evanescent field ($E_{1,2,3}$) propagating along z can be written in three sections (see Fig. 1) separately with the given field coefficients (A, B, C, D and F) based on the standard TMM analysis. $E_1 = Ae^{-k_{1z}z} + Be^{k_{1z}z}$, $E_2 = Ce^{-k_{2z}z} + De^{k_{2z}z}$, $E_3 = Fe^{-k_{3z}z}$, where the wave propagation constants along z-axis are $k_{1z,2z,3z} = \sqrt{k_{1x,2x,3x}^2 - \omega\varepsilon_{1,2,3}\mu_{1,2,3}/c}$ due to the evanescent nature of the field. ω and c are angular frequency and the speed of light, respectively. The detailed discussions and analysis of TMM formalism is given in Chs. 9 and 10 of ref. 24. Hence, applying the boundary conditions dictating that the tangential electric and magnetic fields are to be continuous on both sides at $z = -a$ and $z = a$, after some algebra C and D can be linked to the coefficient F as:

$$2\begin{bmatrix} 1+\dfrac{k_{2z}}{k_{3z}}\dfrac{\mu_3}{\mu_2} & 1-\dfrac{k_{2z}}{k_{3z}}\dfrac{\mu_3}{\mu_2} \\ 1-\dfrac{k_{2z}}{k_{3z}}\dfrac{\mu_3}{\mu_2} & 1+\dfrac{k_{2z}}{k_{3z}}\dfrac{\mu_3}{\mu_2} \end{bmatrix}^{-1} \begin{bmatrix} Fe^{-k_{3z}a} \\ 0 \end{bmatrix} = \begin{bmatrix} Ce^{-k_{2z}a} \\ De^{k_{2z}a} \end{bmatrix} \tag{2-1}$$

The remaining parameters (A and B) can also be linked to F and after a renormalization procedure to satisfy the boundary conditions, Fig. 1(a) can be obtained. As it can be clearly seen, the evanescent wave reaching the boundary z=-a is amplified within the normalized wave impedance matched ($Z = \sqrt{\mu_r/\varepsilon_r} = 1$) region. The main difficulty in recreating the subwavelength images stems from the fact that the high spatial frequency components that carry very critical information about the subwavelength image cannot propagate to the far-field. Such components are normally expected to die out as it can be noticed in medium 1 in Fig. 1(a). By employing a 100nm artificial superlens, these high spatial frequency components can have a chance to be recovered as they are amplified to the output plane during their propagation.

In reality, achieving a perfect lens has never been an easy task as the medium also needs to have impedance matched to the air on both sides. As soon as the impedance matching condition is relaxed, the reflections start to appear as it is demonstrated in

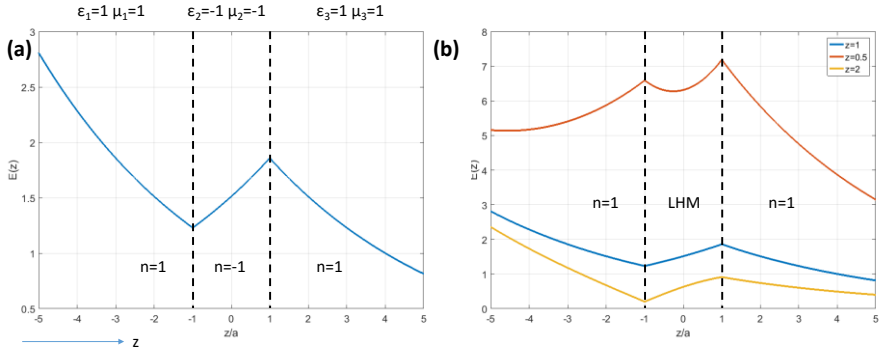


Figure 1. A 100nm thickness left-handed medium (LHM) is sandwiched between two dielectric media along *z*-axis. The evanescent field in medium 1 is launched from the left hand side. The electric fields ($E(z)$) are plotted at 500 THz for (a) wave impedance (Z, normalized by $Z_0 = 120\pi\ \Omega$) matched case; LHM has Z=1 and (b) Z-matched and unmatched cases together. LHM impedance cases: Z=1 (blue), Z=0.5 (red), Z=2 (orange).

Fig. 1(b) for two different wave impedance cases. The evanescent amplification is still observed. Figure 2 illustrates a more realistic exemplary negative index lens that consists of a dispersive homogenized medium in order to mimic the characteristics of a MTM. The constitutive parameters of the medium, and thereby the permittivity and permeability of the MTM are assumed to be of the Drude and Lorentz form, respectively. A similar model had earlier been studied at a different frequency in ref. 25 and the Drude/Lorentz description of a double negative MTM is a common approach. The simulations have been carried out in CST Microwave Studio, a commercial software which is based on Finite Integration Technique. Meanwhile, the scalability of the Maxwell's equations allows us to make analogies between the microwave and optical frequency ranges. The absolute values of the electric field distribution maps have been plotted at 17.7 GHz under the illumination of a modeled horn antenna. The launched beam possesses wavefronts with spherical features. The phase advancement is along the propagation direction in free space (at P_1 and P_3), whereas the left handed behavior dictates a negative phase velocity with respect to the group velocity inside the MTM (at P_2). The conservation of the momentum at the air-MTM boundary compels the wave fronts to collapse and the focusing is attained. The beam quickly starts to diverge once it leaves the MTM.

The conventional MTM configuration suffers from possible losses which originate from the imaginary parts of the constitutional parameters. The impedance mismatches at the air-MTM interface causes standing waves between the source and

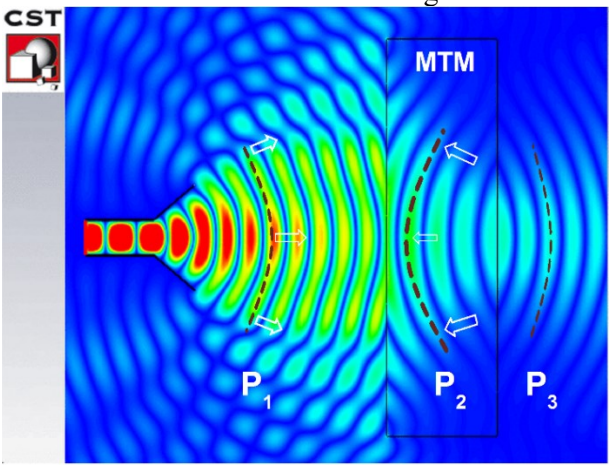


Figure 2. The absolute values of the electric field distributions for a MTM based lens at f=17.7 GHz. P_1, P_2 and P_3 signify three different positions along the direction of propagation. The dashed wine-colored curves exhibit the wave fronts roughly at P_1, P_2 and P_3. The excitation source is modeled with a horn antenna on the left hand side of the figure. The white arrows indicate the direction of the wave-vectors on the same wave front. The effective permittivity function is $\varepsilon_{eff}(\omega) = \varepsilon_\infty - \omega_p/[\omega(\omega - i\nu_c)]$ and the effective permeability function is $\mu_{eff}(\omega) = \mu_\infty + (\mu_s - \mu_\infty){\omega_0}^2/({\omega_0}^2 + i\omega\delta - \omega^2)$, $\varepsilon_\infty = 1.62$, $\omega_p = 2\pi \times 31$ GHz, $\nu_c = 30.69$ MHz, $\mu_\infty = 1.12$, $\mu_s = 1.26$, $\omega_o = 2\pi \times 17$ GHz, $\delta = 500$ MHz.

the structure. Besides, the planar negative index lenses do not have a definite focal length. The MTM lens could focus the incident beams launched from a source that is 4λ away in Fig. 2, but an object at infinity would not be resolved by the planar MTM lens. Instead, planoconcave NIM lenses are built with PCs [26] and MTMs [27], which can accept wave fronts with larger radius of curvatures. Consequently, a planar GRIN formalism presents itself as a favorable choice in view of the fact that it may avoid the cost and manufacturing complexities that a planoconcave lens brings. Hence, an increasing number of publications have rapidly adapted the GRIN concept for the MTM lenses as well [28-32]. Essentially, the main disadvantage of the MTM based lens is the frequency dependency of the negative refraction mechanism since the double negative behavior can be sustained for a limited operational band. Liu *et al*. had suggested working with non-resonant metamaterials in ref. 33, which radically increased the frequency bandwidth. The broadband MTM based designs have been utilized by relying on the small variations of the design parameters of the unit lattice along the propagation direction, which results in a gradient change of the refractive index to control the light's trajectory.

In contrast, the GRIN PC already works over a relatively broader frequency region. Given that the lensing effect is based on the phase retardation mechanism owing to the direct lattice spacing modulation, an intuitive examination even at this stage suggests that the GRIN PCs are less sensitive to the frequency variations. Yet, this point is going to be elaborately covered in the succeeding parts.

3. GRIN PC IN METAMATERIAL REGIME

3.1. Design of the GRIN PC and Ray Trajectories:

GRIN PCs have been exploited in the long wavelength regime, for which the effective medium theory holds. Actually, the earliest designs that utilized the effective medium theory within the context of the 1-D GRIN PCs go back to the year 2005. The proposed designs were envisioned to be used in imaging, guiding, mode matching and coupling [34,35]. The GRIN PC has been shown to simultaneously possess the multi-functionality as beam deflectors and beam aperture modifiers [36]. Consequently, waveguide bends [37] and a cylindrical optical black hole with an omnidirectional absorbance [38] have been made available in the following years. Even a free-space carpet cloak has been proposed by making use of the GRIN PCs in metamaterial regime [39]. An extensive study of the possible GRIN PC lenses in metamaterial regime has been carried out in ref. 40 and similar lens design techniques have rapidly been addressed in the literature [41-43].

Accordingly, we have adapted a GRIN PC lens design that is based on the homogenization procedure. The implementation of the homogenization procedure is derived from Maxwell-Garnett theory. The results are presented in Fig. 3. GRIN PC comprises alumina rods ($\varepsilon_{rod} = n_{rod}^2$=9.61) with varying radii. The dispersion graph in Fig. 3(a) indicates the phase difference ($\Delta\varphi$) within the 1st band along the propagation distance ($\Delta\varphi \propto \Delta k$). The phase difference is expected to be accumulated between two sets of rods with different radii (r_1 and r_2). The radius of the rods changes in the transverse direction (x-axis). Then, the refractive index at any point along x-axis is calculated with the formula that is given in the inset of Fig. 3(a). Thus, a quadratic refractive index variation can be sustained for n(x), where f(x)=π r^2⁄a^2 is the filling factor for each unit cell and a is the constant lattice spacing in the transverse direction. As a result of the quadratic variation, a focusing effect is observed at the end of the structure.

The presented lensing will occur due to the wavefront shaping and thereby phase retardation mechanisms. Hence, the phase velocities (v_p) for each lattice point corresponding to the rods with different filling factors need to be calculated to

estimate $v_p = \omega/k$. The isofrequency contours (IFCs) are presented in Fig. 3(b), which turn out to be centered around Γ point in the lattice symmetry as already hinted in Fig. 3(a). The IFCs are plotted for the smallest radius value (r_2) to the

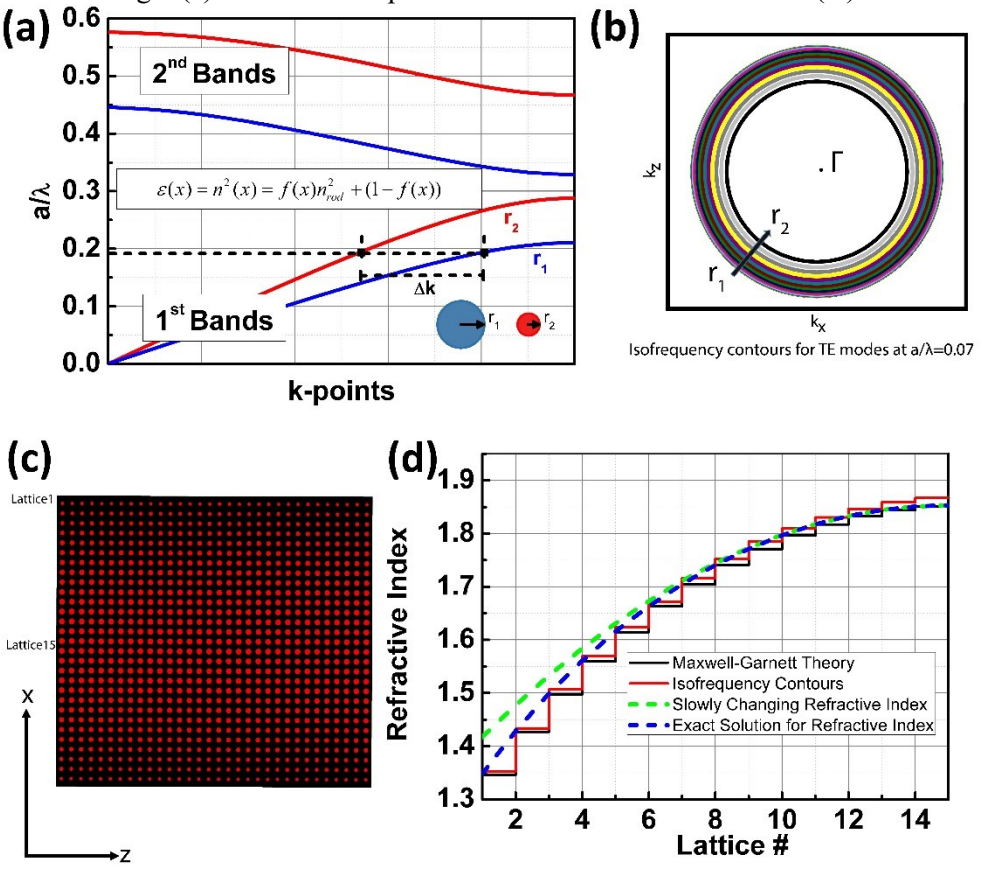


Figure 3. (a) Dispersion graph for two radius values: $r_2 = 0.3a$ (blue line) and $r_1 = 0.173a$ (red line). (b) Isofrequency contours (IFCs) calculated for each rod with different filling factors. The circular IFCs belong to the 1st band of the rod configurations again at $f = 0.07\ a/\lambda$. The smallest filling factor constitute the smallest circle in the IFC whereas the largest is plotted as the largest circle. (c) The realized GRIN PC by modifying the rod radii. (d) The refractive index evaluated at different points of the GRIN PC structure as a function of the rod number where lattice 1 corresponds to the outermost rod and vice versa. Maxwell-Garnett theory predicts the black step-wise plot at distinct lattice points for the corresponding rods, IFC calculations predict red step-wise plot, a slowly changing refractive index treatment gives the continuous homogenized GRIN medium shown in dashed green and the exact solution of the polynomial GRIN medium is shown with dashed blue curve.

largest element (r_1). It becomes hard to distinguish the larger IFCs since they get very close to each other. The centered isotropic IFCs mean that we can estimate the regarding phase velocity directly and correspondingly an effective refractive index for each lattice element. Hence, a secondary homogenized refractive index can be estimated on top of the values offered by Maxwell-Garnett theory. These values are plotted together in Fig. 3(d) as step plots for each homogenized lattice element. The consistency between these plots is evident. The small mismatches are mainly caused by the difficulty of precisely estimating the phase velocity for larger radius values.

In the end, the structure as shown in Fig. 3(c) is constructed with 29 rods along x-axis. The wave propagation will be along z-axis. The homogenized refractive index values retrieved from both Maxwell-Garnett theory and IFCs are compared with the polynomial continuous change that has the form of $n^2(x) = n_0^2(1 - \alpha^2 x^2)$. This quadratic GRIN medium is very well known in traditional GRIN optics. Most of the time, this function is expanded so that a direct estimation of the refractive index rather than the dielectric permittivity can be obtained. Then, the refractive index becomes $n(x) = n_0(1 - \alpha^2 x^2/2)$. Yet, this approximation is obviously only valid when α is very small, which is also attributed to be the slowly changing GRIN medium. In our case, the half width of the GRIN PC lens is $14a$ and the refractive index is not slowly changing. The two refractive index functions are plotted together in Fig. 3(d) for the retrieved n_0 and α values, which also clearly show that the Taylor expansion for small α cannot be applied for the current study. On the other hand, the continuous GRIN optics $n(x)$ almost perfectly agrees with the Maxwell-Garnett theory and closely follows IFC based GRIN PC homogenization efforts. Accordingly, Fermat's Eikonal equation for the shortest path of the rays in GRIN optics for the exact solution needs to be obtained. Such rays would describe how light would propagate inside the GRIN PC and alter its wavefront.

While the ray propagates from point A to B, Fermat's principle dictates that the optical path calculated needs to be at an extremum:

$$\int_A^B n(r)ds = extremum \tag{3-1}$$

where $n(r)$ is the defined refractive index along the path ds and $ds = \sqrt{dx^2 + dz^2}$ assuming that the rods are infinitely extending and the refractive index is invariant along y-axis. A very quick solution can be obtained by strictly working with paraxial waves, which states that the slope of the ray with the main optical/propagation axis is very small and the ray is almost parallel to the z-axis, in other words $ds \approx dz$ [44].

Yet, for our GRIN PC configuration we cannot assume the validity of the paraxial waves, either. The variation principle applied to Eq. (3-1) will bring $\delta \int_A^B n(r)ds = 0$. The solutions to the variation principle will be obtained with optical Euler-Lagrangian equations which give the relationship between the length covered by the ray (s) and the refractive index function (n) as shown in Eq. (3-2) [45]:

$$\frac{d}{ds}\left(n\frac{dr}{ds}\right) = \nabla n \Rightarrow \frac{d}{ds}\left(n\frac{dx}{ds}\right) = \frac{dn}{dx}, \frac{d}{ds}\left(n\frac{dz}{ds}\right) = \frac{dn}{dz} \tag{3-2}$$

Even though the refractive index in our GRIN PC is changing along z-axis in reality as it is seen in Fig. 3(c), the homogenization efforts simply dictate that $dn/dz = 0$ as the same type lattices are assembled in this direction, which would imply the following relationship as it is given in Eq. (3-3):

$$n(x)\frac{dz}{ds} = \frac{n(x)}{\sqrt{dx^2 + dz^2}}dz = \frac{n(x)}{\sqrt{\frac{dx^2}{dz^2} + 1}} = C_1 \tag{3-3}$$

where n is only a function of x and C_1 is a constant quantity which will give $dn/dz = 0$. A rearrangement of Eq. (3-3) can be done in order to obtain z after plugging in the exact quadratic function for $n(x)$ where a secondary constant $C_2 = \sqrt{(n_0^2 - C_1^2)/n_0^2\alpha^2}$ is introduced as the ratio of the other constants in our problem. Hence, Eq. (3-4) can predict the position of the ray along z-axis once the ray's initial position X_i and the final position X_f are known. Yet, most of the time we are interested in the position of the ray along x-axis as the propagation continues along z-axis:

$$\begin{aligned} z &= \int_{X_i}^{X_f} \frac{C_1}{\sqrt{n(x)^2 - C_1^2}}dx = \int_{X_i}^{X_f} \frac{\frac{C_1}{n_0\alpha}}{\sqrt{\frac{n_0^2 - C_1^2}{n_0^2\alpha^2} - x^2}}dx \\ &= \frac{\sqrt{1 - C_2^2\alpha^2}}{\alpha}\int_{X_i}^{X_f} \frac{1}{\sqrt{C_2^2 - x^2}}dx \end{aligned} \tag{3-4}$$

Once we solve the integral given in Eq. (3-4) and try to express X_f as a function of X_i and z, we end up with Eq. (3-5):

$$X_f = C_2 sin\left(\frac{\alpha z}{\sqrt{1 - \alpha^2 C_2^2}} + sin^{-1}\left(\frac{X_i}{C_2}\right)\right) \tag{3-5}$$

The second important quantity will be the slope of the ray which is defined as $tan(\theta) = dx/dz$ in our coordinate axis arrangement. In order to understand the ray propagation in our GRIN PC, we need to solve for the boundary conditions. At the input side we have waves coming at normal incidence with respect to the *x*-axis just before they penetrate into the GRIN PC, ie. θ_i=0 when z=0. Eq. (3-5) can be solved at the input surface to find out the constant C_2:

$$X_i = \pm C_2 \Rightarrow sin^{-1}\left(\frac{X_i}{C_2}\right) = (2n+1)\frac{\pi}{2} \tag{3-6}$$

where n is any integer number. The boundary condition can be applied for every different case of the impinging fields. Then, similar to the metamaterial counterpart, TMM (also called ABCD matrix for the ray optics treatment) of the GRIN PC structure can be obtained for the known input slope and position for the most general case. Eq. (3-5) can be expanded in the following form and the slope can also be retrieved:

$$X_f = C_2\left[sin\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right)cos\left(sin^{-1}\left(\frac{X_i}{C_2}\right)\right) + cos\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right)\frac{X_i}{C_2}\right]$$

$$tan(\theta_f) = \frac{C_2\alpha}{\sqrt{1-\alpha^2 C_2^2}}\left[cos\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right)cos\left(sin^{-1}\left(\frac{X_i}{C_2}\right)\right) - sin\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right)\frac{X_i}{C_2}\right] \tag{3-7}$$

However, since the boundary condition for z=0 will yield $tan(\theta_i)\sqrt{1-\alpha^2 C_2^2}/(\alpha C_2) = cos\left(sin^{-1}(X_i/C_2)\right)$, we can plug it into Eq. (3-7) to end up with the TMM:

$$\begin{bmatrix} X_f \\ tan(\theta_f) \end{bmatrix} = \begin{bmatrix} cos\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right) & sin\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right)\frac{\sqrt{1-\alpha^2 C_2^2}}{\alpha} \\ -\frac{\alpha}{\sqrt{1-\alpha^2 C_2^2}}sin\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right) & cos\left(\frac{\alpha z}{\sqrt{1-\alpha^2 C_2^2}}\right) \end{bmatrix}\begin{bmatrix} X_i \\ tan(\theta_i) \end{bmatrix} \tag{3-8}$$

$$\begin{bmatrix} X_f \\ tan(\theta_f) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} X_i \\ tan(\theta_i) \end{bmatrix}$$

One should be cautious that the solutions highlighted in Eq. (3-3) do not impose the SELFOC gradient medium that is satisfied in Fig. 4(b) under the assumption of paraxial waves and a slowly changing refractive index. On the other hand, the ray

trajectories are plotted in Fig. 4(a) for the general case. It is immediately seen that a very thick GRIN medium is required when SELFOC type lenses are built. The focusing takes place at a position drastically further away while the rays are propagating inside GRIN medium in Fig. 4(b) compared to Fig. 4(a) at the same wavelength. Another important drawback comes from the strict numerical aperture of the lens. The lens in Fig. 4(a) has a much larger width. In contrast, our GRIN PC will be prone to aberrations as the rays have different pitch lengths depending on the exact position of the entrance into the GRIN PC. A relatively blurred focusing will be achieved around $z=2\lambda$ as it can be depicted from Fig. 4(a).

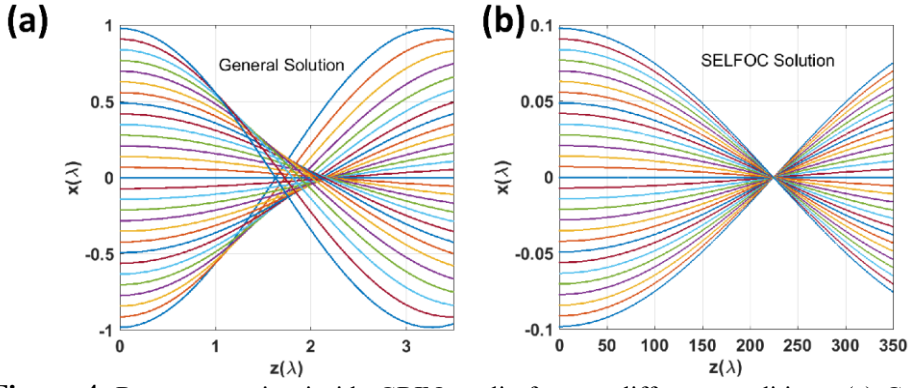


Figure 4. Ray propagation inside GRIN media for two different conditions. (a) General solution as highlighted in Eq. (3-8). (b) SELFOC solution under the assumption of paraxial rays and slowly changing refractive index.

As the focal length of the TMM given in Eq. (3-8) depends on the constant C_2, it would be a critical choice to find the most suitable thickness of the GRIN PC for focusing. In the case of a SELFOC lens, the period of the cosine function in the TMM matrix yields simply:

$$cos(\frac{\alpha L}{\sqrt{1-\alpha^2 C_2^2}}) \approx \cos(\alpha L) = 0 \Rightarrow L = \frac{\pi}{2\alpha} \tag{3-9}$$

which is the well-known periodic pitch length while $\alpha^2 C_2^2 \ll 1$. Two different thicknesses have been chosen in Fig. 5 to show the focusing phenomena solely based on ray optics. As the ray reaches the end of GRIN PC which is depicted for $\Delta z=1.75\lambda$ and 2.24λ, the refraction will determine the outgoing ray trajectory. Then, the outgoing ray angle can be written in the following way. Recognizing $n^2(X_f) = n_0^2(1-\alpha^2 X_f^2)$, $C_2 = X_i$ and $tan(\theta_i) = 0$, we obtain:

$$tan^{-1}\left(-\frac{\alpha C_2}{\sqrt{1-\alpha^2 C_2^2}}sin(\frac{\alpha\Delta z}{\sqrt{1-\alpha^2 C_2^2}})\right) = \theta_f$$
$$\theta_{out} = sin^{-1}\left(n_0\sqrt{1-\alpha^2 C_2^2 cos^2\left(\frac{\alpha\Delta z}{\sqrt{1-\alpha^2 C_2^2}}\right)}sin(\theta_f)\right) \quad (3\text{-}10)$$

Furthermore, the outgoing angle would determine the focal point (Δf) for that particular ray as follows:

$$tan(\theta_{out}) = \frac{X_f}{\Delta f} \Rightarrow \Delta f = \frac{X_f}{tan(\theta_{out})} \quad (3\text{-}11)$$

Combining the results of Eqs. (3-10) and (3-11) would give an idea about Δf. The change in the trajectory can also be seen in Fig. 5. When $\Delta z = 2.24\lambda$, the ray is already very close to the optical axis. Nonetheless, it is very difficult to determine a single Δf for this configuration as we critically witness the aberration in Fig. 4(a), as well. As a result, the determination of the most suitable Δf value requires implementation of wave optics and Finite Difference Time Domain (FDTD) based results to be utilized.

3.2. Wave Optics: Focusing and Source Dependent Propagation under Gaussian beam illumination

Ray optics relies on the fact that $\lambda \rightarrow 0$ whereas Maxwell-Garnett effective medium theory is valid in the long wavelength, namely the metamaterial regime $\lambda \gg a$. These two requirements will bring in differences between the short-wavelength ray optics and long-wavelength metamaterial assumptions. Ray optics still offer an insight regarding the preliminary design of the GRIN PCs but wave optics based solutions are needed to overcome the limits of the ray optics.

Hence, GRIN PCs that are $\Delta z = 1.75\lambda$ and 2.24λ long have been chosen with $\Delta x = 2\lambda$ width. A Gaussian beam source illuminates the input side with almost normal incidence. FDTD based solutions are given in Fig. 6. The real part of the transverse electric field ($\mathrm{Re}\{E_y\}$) has been plotted in Fig. 6(a) and (b) in order to explicitly show the phase retardation mechanism that creates the focusing effect. The intensity plots clearly show that the focusing occurs at the exit side of the GRIN PC. Interestingly, the thicker GRIN PC on the right has a longer Δf compared to the shorter GRIN PC. This is consistent with previous results reported in ref. 46 where

a secant hyperbolic GRIN PC was investigated. Yet, it would not be predictable solely based on Eq. (3-11). Furthermore, the thicker GRIN PC has a better impedance matching to the incoming beam such that it can channel more energy to the output side.

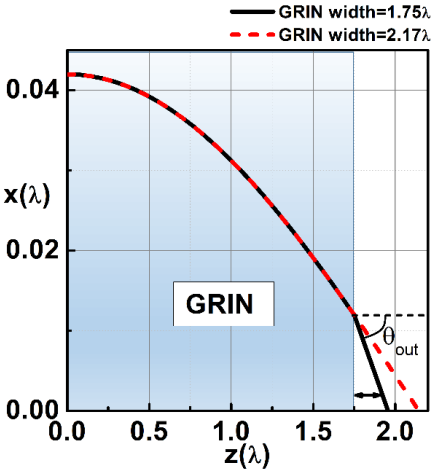


Figure 5. Ray trajectories when GRIN has a thickness of Δz=1.75λ (black solid curve) and 2.24λ (dashed red curve).

The illumination by a Gaussian beam brings in differences when compared to ray optics and also plane waves. Such effects are covered in depth for GRIN optics. One of the first trivial mismatch occurs in the Δf. This shift between different illumination sources is analytically shown in ref. 23 Ch. 4 under paraxial and slowly changing refractive index assumptions for a quadratic GRIN medium. Figure 7 shows two cases with uniform (plane wave) and Gaussian-type illuminations. Periodic boundary conditions have been assumed for the plane wave excitation. FDTD solutions also demonstrate the imaging and focusing differences, which depend on the illumination sources. As expected, the realistic Gaussian beam type illumination creates a waist at the output while the working distance (Δf) has now more depth. The Gaussian beam can be focused over a slightly longer depth along z-axis, which is expressed as the quality of the depth of focus (DOF) for lenses. In other words, the depth of focus of the lens is improved by sacrificing from the resolution under Gaussian beam type illumination. We can conclude that such phenomena that were observed in GRIN optics are also valid for our GRIN PC structure.

Another overlooked property of the GRIN PC is the source size for the Gaussian type illuminations. A Gaussian beam does not automatically yield lensing in GRIN PC structures. The beam interaction depends heavily on the width (w) of the

Gaussian beam. Figure 8 depicts 4 different cases with varying w. Ref. 23 explains that the GRIN optics have Hermite-Gaussian eigenmodes and the incoming wave would be coupled into such modes. The half-width of the fundamental mode is defined as w_{fm}. In case $w_{fm}=w/2$ at the GRIN medium's entrance, the mode would be an adiffractional Gaussian beam and would propagate as the fundamental mode. We clearly see a very similar phenomenon happening in Fig. 8. The beam diverges for Fig. 8(a) and we can label this case as $w<w_{fm}$. On the other hand, Figs. 8(c) and (d)

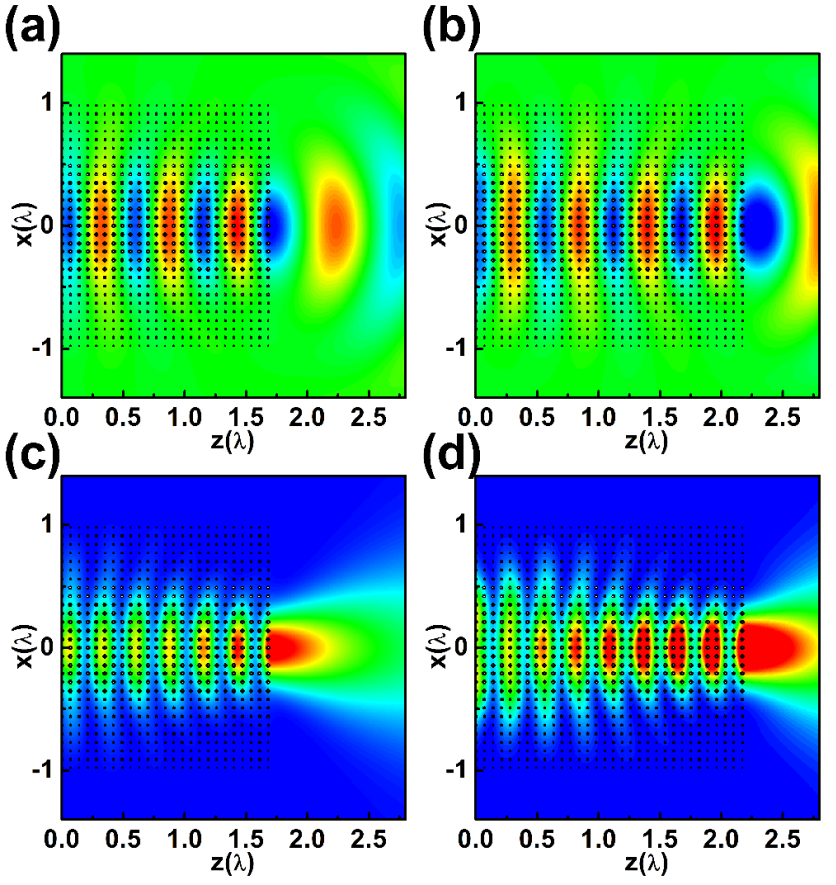


Figure 6. Electric field (E_y) spatial distributions from FDTD results at $f = 0.07\ a/\lambda$. The GRIN PC has been highlighted with black circles for each case. (a) Re$\{E_y\}$ and (c) $|E_y|^2$ when $\Delta z=1.75\lambda$, (b) Re$\{E_y\}$ and (d) $|E_y|^2$ when $\Delta z=2.24\lambda$. The source is a Gaussian beam with a width of 1.14λ. The plots are normalized with respect to the unit intensity.

show the lensing characteristics when $w_{fm}<w$. The adiffractional propagation almost occurs in Fig. 8(b). We would conclude that a very similar assessment of the Gaussian beam propagation is also valid for our GRIN PC.

3.3. Imaging and Fourier Transform with GRIN PC

In order to understand the imaging characteristics of the GRIN PC lens, a current source infinitely extending along y-axis is placed at x=-0.56λ at the input side of the

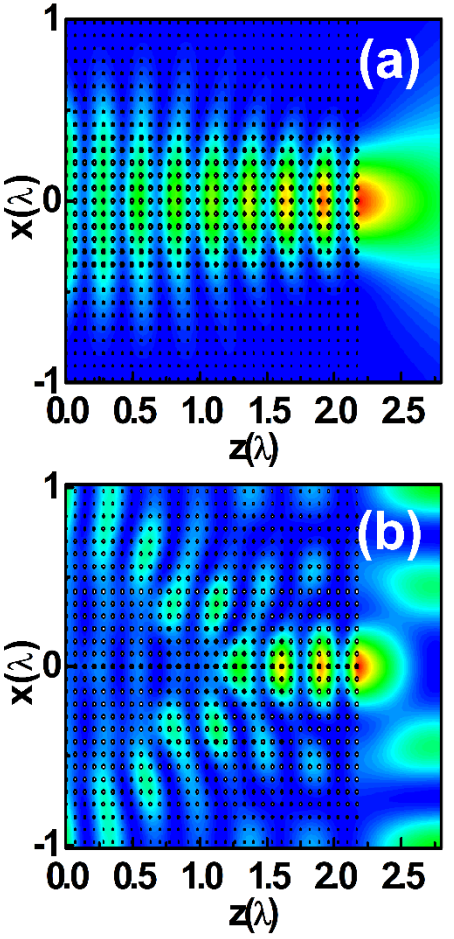


Figure 7. Spatial distributions of the intensity ($|E_y|^2$) when Δz=2.24λ from FDTD results at $f = 0.07\ a/\lambda$. The GRIN PC has been highlighted with black circles for each case. (a) Gaussian beam type illumination with a width of 1.14λ and (b) Plane wave (uniform) illumination. The plots are normalized with respect to their maximum intensity values. The same intensity distribution in Fig. 6(d) is replotted in (a) with a different normalization in order to study the impact of the illumination sources.

lens. The current source sits 0.14λ away from the lens. The created image of any lens at the output side is expected to take a double Fourier Transform in the form of:

$$\mathcal{F}\{\mathcal{F}\{f(x)\}\} = f(-x) \tag{3-12}$$

where $\mathcal{F}$ stands for Fourier Transform and $f(x)$ is the image at the input side. Hence, the image would be inverted. This is also seen in Fig. 9 as the image is created on the $+x$ side. The image is created 0.23λ away from the lens at the output side. The aberration problems caused by the astigmatism of the GRIN PC is evident during the image formation. The ray optics treatment in Fig. 4(a) already gave clues about this aberration problem for this particular GRIN PC, which is not considered to be a traditional SELFOC medium. The peak of the intensity at the exit side is formed at $x=0.6\lambda$, which is very close to the location of the current source at the input side.

The GRIN optics elements are characterized with kernel (point spread) functions. The kernel function ($K(x,x_0;z)$) would be used to form the images at the output side. For the SELFOC media, a canonical integral transformation is used:

$$\Psi(x;z) = \int \Psi(x_0; z=0) K(x_0, x; z) dx_0 \tag{3-13}$$

Eq. (3-13) images the field at the input transversal plane $\psi(x_0;z=0)$ to the output transversal plane $\psi(x;z)$. Since imaging systems are linear systems, the kernel function would define the impulse response of the system. For the slowly changing refractive index condition, the kernel function is derived to be in the following format in ref. 23:

$$K(x_0, x; z) = \frac{k_0 n_0 \exp(ik_0 n_0 z)}{i2\pi B(x,z)} \times exp\left\{\frac{-k_0 n_0}{i2B(x,z)}\left[A(x,z)x_0^2 + D(x,z)x - 2xx_0\right]\right\} \tag{3-14}$$

where A, B and D are the TMM 2x2 matrix parameters illustrated in Eq. (3-8) and k_0 and n_0 are the free space wave vector and the refractive index at the center of the GRIN medium, respectively. For the SELFOC medium ($\alpha^2 C_2^2 \ll 1$) at $z = L = \pi/2\alpha$, both A and D would become zero while $B = \sin(\alpha L)/\alpha = 2L/\pi$ as Eq. (3-9) also implies. Under these conditions, Eq. (3-13) can be rewritten as:

$$\Psi(x;z) = \frac{k_0 n_0 \exp(ik_0 n_0 z)}{i4L} \int \Psi(x_0; z=0) exp\left\{\frac{-i2\pi x x_0 k_0 n_0}{4L}\right\} dx_0 \tag{3-15}$$

$$= \mathcal{F}\{\Psi(x_0; z = 0)\}$$

which is the Fourier Transform of the input wave. This condition is very much similar to the propagation through the standard lens. Now, we need to assign the spatial frequency for this lensing scenario. The spatial frequency is $v_x = \pi x n_0 / 2\lambda L$ as opposed to $v_x = x/\lambda L$ in standard lensing.

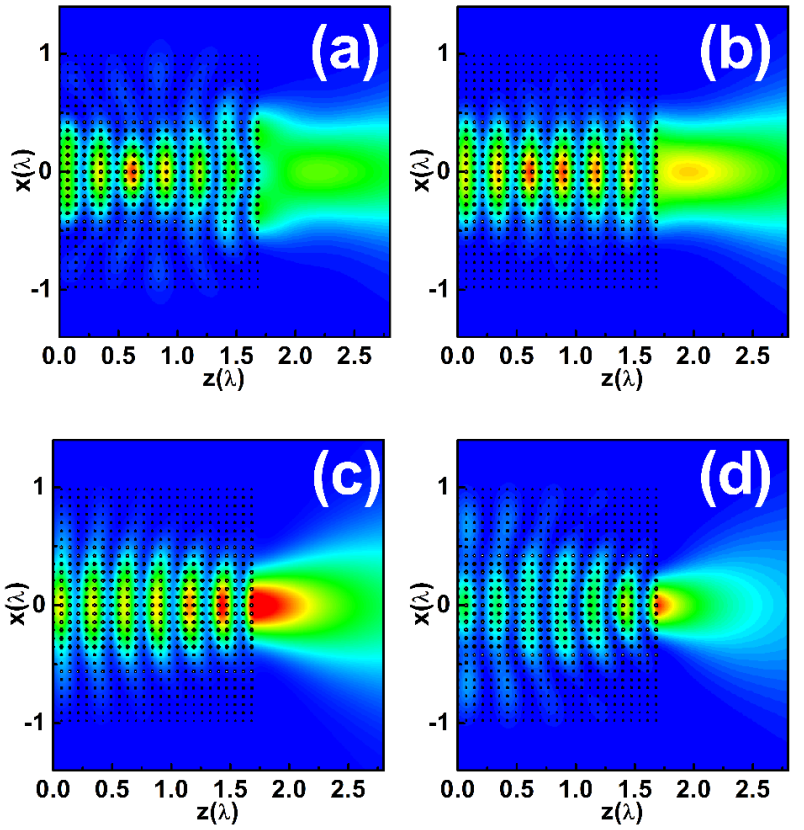


Figure 8. Spatial distributions of the intensity ($|E_y|^2$) when Δz=1.75λ from FDTD results at $f = 0.07\ a/\lambda$ under Gaussian beam type illumination for different beam widths. The GRIN PC has been highlighted with black circles for each case. Gaussian beam source has a width of (a) w=0.38λ, (b) w=0.76λ, (c) w=1.14λ and (d) w=1.52λ. The plots are normalized with respect to their maximum intensity values. The same intensity distribution in Fig. 6(c) is replotted in (c) with a different normalization in order to study the impact of the width of the Gaussian beam type sources.

The validity of Eq. (3-15) for our GRIN PC which is not a SELFOC medium will be tested. If we stick with the same kernel function, both A and D become very small when GRIN PC is Δz=2.24λ and B is almost 1/α. However, we also have the finite aperture of the lens that we need to take into consideration. This aperture (w_{ap}) function will determine the resolution of the lens since the GRIN medium is not infinite. The aperture can be modelled with a $rect(x/w_{ap})$ function as follows:

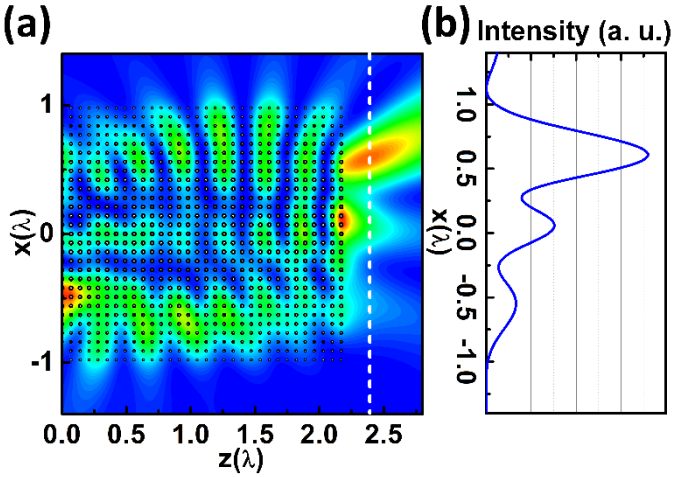


Figure 9. (a) Spatial distribution of the intensity ($|E_y|^2$) when Δz=2.24λ from FDTD results at $f = 0.07a/\lambda$. A dipole source is located at the input side (x=-0.56λ, z=-0.14λ) of the GRIN PC and image is formed at the exit side (x=+0.6λ, z=2.4λ) to understand the imaging characteristics. (b) The intensity distribution at z=2.4λ is plotted.

$$rect\left(\frac{x}{w_{ap}}\right) = \begin{cases} 1, |x| \le w_{ap}/2 \\ 0, |x| > w_{ap}/2 \end{cases} \tag{3-16}$$

Then, simply if an input function $\psi_{test}(x_0,z=0)$ is launched towards the GRIN PC, the wavefunction at the input side of the transversal plane of the GRIN PC will be the multiplication of the test function with the aperture function. Hence, Eq. (3-15) will become:

$$\Psi_{out} = \frac{k_0 n_0}{i4L}\exp(ik_0 n_0 z)\int \Psi_{test}(x_0) rect(\frac{x_0}{w_{ap}}) exp\left\{\frac{-i2\pi x x_0 k_0 n_0}{4L}\right\} dx_0 \tag{3-17}$$

$$\Psi_{out} = \mathcal{F}\{\Psi_{test}(x_0)\} \otimes \mathcal{F}\left\{rect(\frac{x_0}{w_{ap}})\right\}$$

as convolution theorem suggests once we drop the constant phase term ($\exp(ik_0n_0z)$) in front. Figure 10 shows the case when we feed the GRIN PC with a Gaussian test function of the form $\Psi_{test}(x_0) = \exp(-4x_0^2/w_g^2)$ where $w_g \approx 1.52\lambda$ is the full waist of the Gaussian and $w_{ap} \approx 2.03\lambda$. Hence, Eq. (3-17) will turn into:

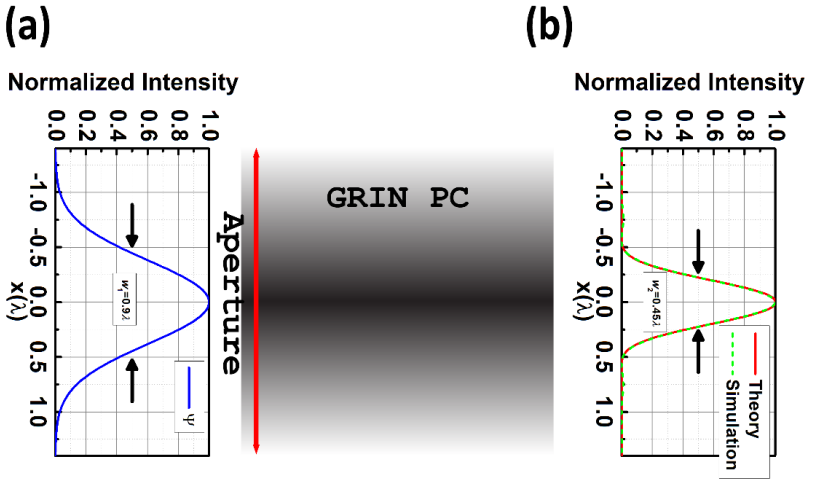


Figure 10. Fourier Transform capability of the GRIN PC with the applied test function. (a) The normalized intensity distribution of a Gaussian test function at the input of the GRIN PC. The width of the test Gaussian function is 0.9λ. (b) The normalized intensity distribution at the output of the GRIN PC (dashed green curve) as the FDTD simulation results offer at $f = 0.07a/\lambda$ and Δz=2.24λ. The result of the normalized intensity distribution (solid red curve) at the output based on Eq. (3-17). The width of the output wavefunction decreases from w_1=0.9λ to w_2=0.45λ.

$$\mathcal{F}\left\{rect(\frac{x_0}{w_{ap}})\right\} = \frac{\sin(\pi w_{ap} v_x)}{\pi v_x}\sqrt{\frac{k_0 n_0}{i4L}} \tag{3-18}$$

$$\mathcal{F}\{\Psi_{test}(x_0)\} = \frac{\sqrt{\pi}}{2} w_g exp\left(\frac{-\pi^2 w_g^2 v_x^2}{4}\right)\sqrt{\frac{k_0 n_0}{i4L}}$$

The resulting wavefunction at the exit side has been plotted together with the FDTD based simulation results in Fig. 10(b). The results suggest that even though our GRIN PC is not a SELFOC medium, the kernel and canonical integral transformation do yield very similar results to the SELFOC versions. The GRIN PC narrows down the width of the outgoing beam from 0.9λ to 0.45λ. The fact that Fourier transforming capability of the lens is valid has already been already backed

up by the previously discussed focusing cases related to FDTD based results in Figs. 6-8, even without the presence of a SELFOC medium.

Finally, the Fractional Fourier Transform capability of the GRIN PC is tested in a similar fashion. GRIN optics dictate that the modes inside the GRIN medium propagate in such a way that would naturally realize Fractional Fourier Transform. GRIN medium provides the Fractional Fourier Transform in its purest simplest form [47]. Then, the kernel function would be modified such that $\alpha L < \pi/2$ for the SELFOC medium as the self-focusing distance has not been reached yet. Accordingly, a fractional scaling factor can be defined for the SELFOC and adapted for our GRIN PC:

$$\beta = \frac{\alpha L}{\pi/2} \Rightarrow \frac{\frac{\alpha \Delta z}{\sqrt{1-\alpha^2 C_2^2}}}{\pi/2} \tag{3-19}$$

This modification brings in the fact that depending on the C_2 value, similar but different β values could be proposed. As it can be seen in Fig. 11(a), a GRIN PC that is much shorter, Δz=1.05λ is selected and the fields are collected at the output of the GRIN PC. Unlike the treatment of a homogenous GRIN medium, the near-field interactions make this process quite tricky together with the impedance mismatch at the air-GRIN PC interface. The field should be obtained not from the air but as close as possible to the GRIN PC as we are interested in the natural Fractional Fourier Transform of the GRIN PC. The simulation result is plotted in Fig. 11(b) together with the results based on Fractional Fourier Transform for two possible β values. The preliminary results suggest that Fractional Fourier Transform capability can also be addressed with GRIN PC that is not a SELFOC medium, itself. When Δz=1.05λ, β value ranges between 0.4 and 0.43 based on Eq. (3-19) for different ray trajectories depicted in Fig. 4(a).

A frequency sweep shows that the dielectric lenses are much more suitable for larger bandwidth applications, as depicted in Fig. 12. Two higher frequencies were selected as $a/\lambda = 0.1167$ and $a/\lambda = 0.2$. These two frequencies still reside in the first band of the dielectric rods, which is shown in Fig. 3(a). Maxwell-Garnett theory based homogenization is frequency independent. Nevertheless, it presents very consistent results for $a/\lambda = 0.07$. As we move up close to $a/\lambda = 0.2$, Maxwell-Garnett homogenization method results will deviate from what IFCs present. IFCs will still be circular centered around Γ point, resembling Fig. 3(b). Yet, the actual refractive index of each lattice will be different from Maxwell-Garnett's predictions. The refractive index gradual change can still be expressed with a quadratic relationship (with some tolerable deviation). However, α value will be different than what has

been assumed throughout the paper for $a/\lambda = 0.07$. A quick analysis shows that α will need to increase as the refractive index change becomes more pronounced over each lattice in the transverse direction while frequency of operation is increasing. On the other hand, the ray optics assumptions will be more suitable for these higher frequencies. Eq. (3-9) will dictate that a shorter lens is required for focusing. Figure 12 also illustrates that focusing condition has been already met before the waves reach the full width. Nevertheless, Fig. 12(a) still offer focusing at the exit side of

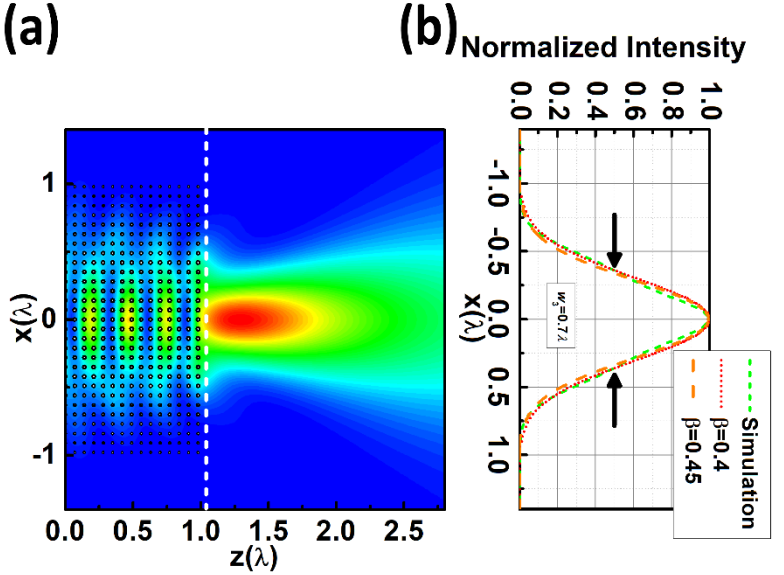


Figure 11. Fractional Fourier Transform capability of the GRIN PC with the applied test function. (a) Spatial distribution of the intensity ($|E_y|^2$) when Δz=1.05λ from FDTD results at $f = 0.07\ a/\lambda$. The same Gaussian test function also with the same width as in Fig. 10 is applied at the input of the GRIN. (b) The normalized intensity distribution of the output of the GRIN PC (dashed green curve) as the FDTD simulation results offer at the output of the GRIN PC (at the dashed white line). The result of the normalized intensity distribution at the output based on Fractional Fourier Transform for β=0.4 (dotted red curve) and β=0.45 (dashed orange curve). The width of the output wavefunction decreases from w_1=0.9λ to w_2=0.7λ.

the GRIN PC, which can also be claimed for Fig. 12(b). There is a competing phenomenon coming into play for Fig. 12(b), which is the traditional Bragg interactions of the wave with the periodic crystal. As the wavelength is decreased, the wave starts to actually "see" each scatterer instead of a homogenous medium and the peculiar wavefront formations are observed as in the case of Fig. 12(b). Shortly, since the change in α is not very large compared to the originally studied case in $a/\lambda = 0.07$, the same GRIN PC length could be used for both cases in Fig. 12. Once the focusing is achieved and the modified Gaussian beam reaches its shortest

waist inside the GRIN PC, it starts diverging in Fig. 12(b) more significantly. The bandwidth of the lens is limited by the Bragg interactions rather than by the bandgap. The GRIN PC still offers much better performance than the metallic, resonance analogs that are shown in Fig. 2. Lastly, the same physical size of the Gaussian beams that have been used in the original case has been launched in Fig. 12. Since, $w_{fm}<w$ at the largest wavelength ($a/\lambda = 0.07$), the shorter wavelength cases would already force the beam to focus.

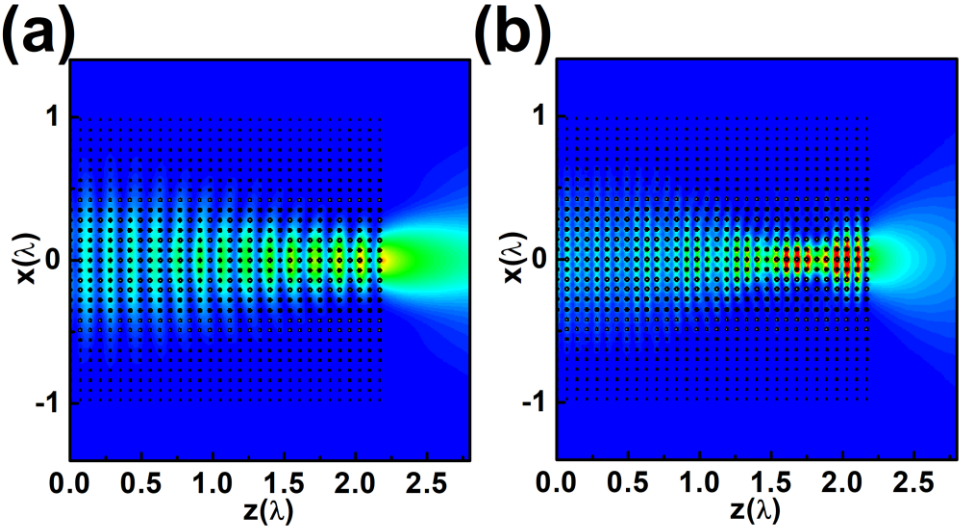


Figure 12. Spatial distribution of the intensity ($|E_y|^2$) when $\Delta z=2.24\lambda$ from FDTD results at (a) $f = 0.1167a/\lambda$ and (b) $f = 0.2a/\lambda$. The source is a Gaussian beam with a fixed width used in Fig. 6(d). The normalization has been carried out with respect to the highest intensity values for each frequency.

4. CONCLUSION

In conclusion, a GRIN PC medium which has not been traditionally investigated because of its abrupt refractive index distribution is studied in details. The studied GRIN PC is shown to behave in very close correlation to its slowly perturbed refractive index possessing GRIN counterparts at its core physics. Concepts from GRIN optics have been borrowed to explain the ray trajectories, wave optics phenomena together with imaging and Fourier optics in such untraditional GRIN PCs to be used as lenses with wide numerical apertures and relatively much shorter lengths over a considerably large bandwidth.

The analysis has started by discussing and comparing GRIN PCs with volumetric MTMs. The resonance nature and the impedance matching difficulties of the MTMs have been highlighted as the main factors which directed the researchers to switch to GRIN optics. The ray trajectory of the studied GRIN PC has been derived, which also pointed out the ABCD matrix for this untraditional GRIN PC design. The wave nature of the current GRIN PC design has been investigated under the light of the derived ABCD matrix. The simulations have shown that the imaging capabilities of a point source can easily be demonstrated for the current GRIN PC. Since the current GRIN PC can perform imaging by taking the double Fourier Transform of a point source, a kernel function adapted from the GRIN optics has been applied to understand the single Fourier Transform capability. The presented results have illustrated that the impulse response which is valid for the SELFOC GRIN medium can be utilized and the present GRIN PC can take the Fourier Transform of a test Gaussian function. Moreover, the Fractional Fourier Transform of the test Gaussian function is also verified at the output of the GRIN PC.

The bandwidth is shown to be limited by the emergence of the Bragg interactions of the wave with the periodic scatterers. A seemingly ambiguous GRIN PC, such as the one currently studied that does not possess an easy analytical solution without the paraxial approximation and a large refractive index variance, has not been studied extensively in the literature. The present work shows that even such a GRIN PC can exhibit not only very similar Fourier Transforming capabilities but also Fractional Fourier Transform properties as one would expect to see in GRIN optics. The adiffractional beam propagation is shown with the careful selection of the excitation sources. The adifractional beam propagation is very critical at the end of fiber optic waveguides where the beam tends to quickly diverge and a small portion of the wave information can be transmitted to the next optical device. The present design with its high numerical aperture can be used to collimate the light as an output coupler in a real life application. The characteristics of such an output coupler will be studied in a future work. Furthermore, the current design is a member of the broad research family of transformation optics. The gradient filling factor has been employed to show how light can be focused in this study but similar transformational optics designs could also be analyzed from the perspective of wave impedance matching and a cloaking scheme can be implemented. The refractive index can be gradually remodulated to facilitate the cloaking by limiting the interaction of the incident light with the emerging beam. Moreover, the imaging capability of the GRIN PC can be utilized for advanced optical lithography. Exemplary application fields can be listed for the current design aside from being merely used as an imaging component. Another very useful implementation would be to utilize the current design as a fast Fractional Fourier Transform calculator. The Fractional Fourier

Transform could be calculated at the speed of light, which would be a promising improvement on top of the present methods relying on the computational power of electronics.

REFERENCES

[1] J.B. Pendry, Negative Refraction Makes a Perfect Lens, *Phys. Rev. Lett.*, 85/18 (2000) 3966–3969.

[2] D.R. Smith, J.B. Pendry and M.C.K. Wiltshire, Metamaterials and Negative Refractive Index, *Science*, 305/5685 (2004) 788-792.

[3] A. Grbic and G.V. Eleftheriades, Subwavelength focusing using a negative-refractive-index transmission line lens, *IEEE Antennas Wirel. Propag. Lett.*, 2 (2003) 186-189.

[4] K. Aydin, I. Bulu and E. Ozbay, Subwavelength resolution with a negative-index metamaterial superlens, *Appl. Phys. Lett.*, 90 (2007) 254102.

[5] D. Schurig, J.J. Mock, B.J. Justice, S.A. Cummer, J.B. Pendry, A.F. Starr and D.R. Smith, Metamaterial Electromagnetic Cloak at Microwave Frequencies, *Science*, 314/5801 (2006) 977-980.

[6] R. Liu, C. Ji, J.J. Mock, J.Y. Chin, T.J. Cui, D.R. Smith, Broadband Ground-Plane Cloak, *Science*, 323/5912 (2009) 366-369.

[7] D. Sell, J. Yang, E.W. Wang, T. Phan, S. Doshay and J.A. Fan, Ultra-High-Efficiency Anomalous Refraction with Dielectric Metasurfaces, *ACS Photonics*, 5/6 (2018) 2402-2407.

[8] C. Argyropoulos, N.M. Estakhri, F. Monticone and A. Alù, Negative refraction, gain and nonlinear effects in hyperbolic metamaterials, *Opt. Express*, 21 (2013) 15037-15047.

[9] W. Liang, Z. Li, Y. Wang, W. Chen and Z. Li, All-angle optical switch based on the zero reflection effect of graphene–dielectric hyperbolic metamaterials, *Photon. Res.*, 7 (2019) 318-324.

[10] J. Noonan and T.G. Mackay, On electromagnetic surface waves supported by an isotropic chiral material, *Opt. Commun.*, 434 (2019) 224-229.

[11] T. Gric and O. Hess, Disorder in Metamaterials Chapter 12, Advanced Thermoelectric Materials (editted by C.R. Park), (John Wiley & Sons, Inc., 2019)

[12]M. Takeda, A. Tsuchiyama, M. Okada, S. Matsui, T. Inoue and K. Aizawa, Improvement of focusing characteristics of a surface plasmonic lens for UV wavelength, *Jpn. J. Appl. Phys.*, 56 (2017) 09NC02.

[13] W.T. Chen, A.Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, F. Capasso, A broadband achromatic metalens for focusing and imaging in the visible, *Nat. Nanotechnol*., 13 (2018) 220−226.

[14] N. Kundtz and D.R. Smith, Extreme-angle broadband metamaterial lens, *Nat. Mater.*, 9 (2010) 129–132.

[15] H. Kurt and D.S. Citrin, Graded index photonic crystals, *Opt. Express*, 15 (2007) 1240-1253.

[16] D. Schurig, J.B. Pendry and D.R. Smith, Calculation of material properties and ray tracing in transformation media, *Opt. Express*, 14 (2006) 9794-9804.
[17] A.O. Cakmak, E. Colak, H. Caglayan, H. Kurt and E. Ozbay, High efficiency of graded index photonic crystal as an input coupler, *J. Appl. Phys.*, 105/10 (2009) 103708.
[18] A.E. Serebryannikov, A.O. Cakmak, E. Colak, H. Caglayan, H. Kurt and E. Ozbay, Multiple slow waves and relevant transverse transmission and confinement in chirped photonic crystals, *Opt. Express*, 22 (2014) 21806-21819.
[19] M. Gumus, I.H. Giden and H. Kurt, Broadband self-collimation in C2 symmetric photonic crystals, *Opt. Lett.*, 43/11 (2018) 2555-2558.
[20] D. Luo, G. Alagappan, X.W. Sun, Z. Raszewski and J.P. Ning, Superbending effect in two-dimensional graded photonic crystals*, Opt. Commun.*, 282/2 (2009) 329–332.
[21]E. Centeno, D. Cassagne and J.-P. Albert, Mirage and superbending effect in two dimensional graded photonic crystals, *Phys. Rev. B*, 73/23 (2006) 235119–235123.
[22] E. Akmansoy, E. Centeno, K. Vynck, D. Cassagne and J. M. Lourtioz, Graded photonic crystals curve the flow of light: An experimental demonstration by the mirage effect, Appl. Phys. Lett. 92(13), 133501 (2008).
[23] C. Gomez-Reino, M.V. Perez and C. Bao, Gradient-Index Optics: Fundamentals and Applications (Springer, 2002).
[24] P. Markos and C. M. Soukoulis, Wave Propagation From Electrons to Photonic Crystals and Left-Handed Materials, (New Jersey, Princeton University Press, 2008).
[25] G. Lubkowski, R. Schuhmann and T. Weiland, Extraction of Effective Metamaterial Parameters by Parameter Fitting of Dispersive Models, *Microw. Opt. Technol. Lett.*, 49 (2007) 285-288.
[26] B.D.F. Casse, W.T. Lu, Y.J. Huang and S. Sridhar, Nano-optical microlens with ultrashort focal length using negative refraction, *Appl. Phys. Lett.*, 93 (2008) 053111.
[27] C. G. Parazzoli, R.B. Greegor, J.A. Nielsen, M.A. Thompson, K. Li, A.M. Vetter, M.H. Tanielian and D.C. Vier, Performance of a negative index of refraction lens, *Appl. Phys. Lett.*, 84 (2004) 3232-3234.
[28] T. Driscoll, D.N. Basov, A.F. Starr, P.M. Rye, S. Nemat-Nasser, D. Schurig and D.R. Smith, Freespace microwave focusing by a negative-index gradient lens, *Appl. Phys. Lett.*, 88 (2006) 081101.
[29] R.B. Greegor, C.G. Parazzoli, J.A. Nielsen, M.A. Thompson, M.H. Tanielian and D.R. Smith, Simulation and testing of a graded negative index of refraction lens, *Appl. Phys. Lett.*, 87 (2005) 091114.
[30] D. R. Smith, J. J. Mock, A. F. Starr and D. Schurig, Gradient index metamaterials, *Phys. Rev. E*, 71 (2005) 36609.
[31] O. Paul, B. Reinhard, B. Krolla, R. Beigang and M. Rahm, Gradient index metamaterial based on slot elements, *Appl. Phys. Lett.*, 96 (2010) 241110.
[32] R. Liu, X. M. Yang, J. G. Gollub, J. J. Mock, T. J. Cui and D. R. Smith, Gradient index circuit by waveguided metamaterials, *Appl. Phys. Lett.*, 94 (2009) 073506.
[33] R. Liu, Q. Cheng, J. Y. Chin, J. J. Mock, T. J. Cui and D. R. Smith, Broadband gradient index microwave quasi-optical elements based on non-resonant metamaterials, *Opt. Express*, 17 (2009) 21030-21041.
[34] U. Levy, M. Nezhad, H.-C. Kim, C.-H. Tsai, L. Pang and Y. Fainmann, Implementation of a graded-index medium by use of subwavelength structures with graded fill factor, *J. Opt. Soc. Am. A*, 22 (2005) 724-733 (2005).

[35] U. Levy, M. Abashin, K. Ikeda, A. Krishnamoorthy, J. Cunningham and Y. Fainman, Inhomogenous Dielectric Metamaterials with Space-Variant Polarizability, *Phys. Rev. Lett.*, 98 (2007) 243901.
[36] M. Lu, B.K. Juluri, S.-C.S. Lin, B. Kirally, T. Gao and T.J. Huang, Beam aperture modifier and beam deflector using graded-index photonic crystals, *J. Appl. Phys.*, 108 (2010) 103505.
[37] H.-W. Wang and L.-W. Chen, High transmission efficiency of arbitrary waveguide bends formed by graded index photonic crystals, *J. Opt. Soc. Am. B*, 28 (2011) 2098-2104.
[38] H.-W. Wang and L.-W. Chen, A cylindrical optical black hole using graded index photonic crystals, *J. Appl. Phys.*, 109 (2011) 103104.
[39] M. Yin, X.Y. Tian, H.X. Han and D.C. Li, Free-space carpet-cloak based on gradient index photonic crystals in metamaterial regime, *Appl. Phys. Lett.*, 100 (2012) 124101.
[40] B. Vasic, G. Isic, R. Gajic and K. Hingerl, Controlling electromagnetic fields with graded photonic crystals in metamaterial regime, *Opt. Express*, 18 (2010) 20321-20333.
[41] Z.L. Mei, J. Bai and T.J. Cui, Gradient index metamaterials realized by drilling hole arrays, *J. Phys. D: Appl. Phys.*, 43 (2010) 055404.
[42] B. Vasic, R. Gajic and K. Hingerl, Graded photonic crystals for implementation of gradient refractive index media, *J. Nanophoton.*, 5 (2011) 051806.
[43] B. Vasic and R. Gajic, Self-focusing media using graded photonic crystals: Focusing, Fourier transforming and imaging, directive emission and directional cloaking, *J. Appl. Phys.*, 110 (2011) 053103.
[44] B. Saleh and M. Teich, Fundamentals of photonics (New York, Wiley, 1991).
[45] R. Weinstock, Calculus of Variations with Applications to Physics and Engineering (New York: Dover Publications, 1974).
[46] M. Turduev, I.H. Giden and H. Kurt, Design of flat lens-like graded index medium by photonic crystals: Exploring both low and high frequency regimes, *Opt. Comm.*, 339 (2015) 22-33.
[47] H.M. Ozaktas and D. Mendlovic, Fractional Fourier optics, *J. Opt. Soc. Am. A*, 12 (1995) 743-751.

Current Address: Evrim COLAK: Electrical and Electronics Engineering Department Ankara University, TURKEY
E-mail : evrim.colak@ankara.edu.tr
Orcid ID: https://orcid.org/ 0000-0002-4961-5060

Current Address: Atilla Ozgur CAKMAK: Engineering Science and Mechanics Department The Pennsylvania State University, USA
E-mail: aozgur@psu.edu
Orcid ID: https://orcid.org/ 0000-0002-3933-3203

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 102-110 (2019)
DOI: 10.33769/aupse.506181
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

DESIGNING A SECURE BLOCKCHAIN-BASED TRADING PLATFORM FOR INTERNET OF THINGS

Aydın ELBUZ, Murat OSMANOGLU and Omer Ozgur TANRIOVER

ABSTRACT. Blockchain, a distributed database technology attracts intensive attention recently. It was first described in 1991 by Stuart Haber and W. Scott Stornetta, in order to prevent tampering for document timestamps [1]. After that, Satoshi Nakamoto conceptualized this technology with his work, a peer-to-peer electronic cash system named as Bitcoin. Due to its authenticity, user anonymity and data immutability; this technology has also been applied to different areas ranging from finance, supply-chain managements to social services. In this paper, we propose a blockchain based application for trading data, collected from IoT devices. Our application enables the sellers to commercialize their data in exchange of the currency produced by the application. The application also utilizes smart contracts to establish trust between both sides of the exchange. Furthermore, we discuss the security and reliability of the system components.
.

1. INTRODUCTION

Blockchain is an efficient technology that realizes distributed ledger, which is a decentralized and publicly available data structure for storing transactions. These transactions are verified through a majority consensus [2,3]. So once the transactions are written in the ledger, they can't be changed and erased unless someone can take control majority of the system at the same time [4]. Blockchain is a read-only database system that makes it impossible to create and add a fraudulent or incorrect transaction to the chain. This feature presents a safe and secure system that eliminates third parties along with data immutability and user anonymity.

Bitcoin is the most popular application of blockchain and it took the world by storm when it showed up. After that, mostly in finance sector, many other similar implementations appeared that uses blockchain technology. Due to its success in finance, blockchain's popularity has surpassed these levels and spread in many

Received by the editors: January 01, 2019; Accepted: May 14, 2019.
Key word and phrases: Blockchain, internet of things, distributed ledger.

areas.

Internet of things (IoT) network was also affected from blockchain technology. IBM states that blockchain allows IoT to accelerate transactions and reduce the risk of collusion and tampering, which also reduces costs and complexity of operations [5]. Nowadays, this technology used for controlling and configuring IoT devices [6] and shared economy applications [7]. Besides, blockchain has also been used in smart buildings and food and pharmaceutical industries [8].

Huh, Cho and Kim [6] proposed a blockchain based solution for designing an IoT system. They aimed to control and configure IoT devices using the blockchain. The system uses RSA public key cryptography for managing keys. Private keys saved on individual IoT devices. Beside Ethereum platform is used for smart contracts and storing public keys. They simulated IoT system using Raspberry Pi devices and ran a Turing-complete code for smart contracts. However, they observed that the transaction time of Ethereum blockchain is slow. Thus, they needed a proxy or a large storage for saving entire blockchain because of Ethereum's light client protocol supporting issue.

Dorri et al. [9] proposed a tiered lightweight scalable blockchain, the optimized blockchain system, for IoT security and privacy requirements. They used a smart home setting for IoT applications. The system achieves decentralization by forming an overlay network where high resource devices jointly manage a public blockchain. The overlay designed as distinct clusters to reduce overheads. The cluster heads are responsible for managing the public blockchain. In conclusion, they observed that the lightweight scalable blockchain increases blockchain scalability compared to relevant baselines. It also resilient to some security attacks and it decreases packet overhead and delay.

In this paper, we introduce an application of blockchain to IoT data. Our application can be viewed as an extension of the construction introduced by Elbuz et al. [10]. It can be considered as a platform in which sellers can commercialize their data collected from an IoT device, and customers can contact with sellers, examine the data and buy it. This platform gives confidence to both seller and customer side. Our system uses blockchain technology together with smart contracts for reliability of data and secure money transfer.

The rest of the paper is organized as follows. Section II provides an overview of the used technologies. Section III describes the design of the proposed system and its scenario. Section IV analyzes the system and components' security. Finally, Section V concludes the paper and explains future works.

2. DEFINITION

In this section, we give the definitions of the core components that will be used to build our platform.

Cryptographic hash function is a one-way mathematical function or an algorithm which produces a unique fixed length string of text from a data or a file of an arbitrary length. Because the same data always results in the same hash, a small change to a data changes the whole hash value. And it is computationally infeasible to generate the data from its hash value due to the pre-image resistance feature. Also, cryptographic hash functions have collision resistance, i.e. it is very hard to find the same hash value from any two different inputs with any length.

Digital signature is a mathematical technique for validating the authenticity of digital document or message. This process is generated as a result of a sequence of operations: key generation, signing and verification algorithm. In the key generation, public and private keys are computed for the user. In the signing algorithm, sender signs the document with his private key. After signing, he attaches his signature to the document and sends it to the recipient. In the verification algorithm, recipient verifies the validity of the signature using the signature and the message together with sender's public key. A digital signature algorithm is called ‘unforgeable’ if any message/signature pair (m, s) not exists where s was not produced by the legitimate signer.

Merkle tree is a hash tree that contains hash values of data blocks and data blocks’ hashes hierarchically. In Merkle tree, every leaf node contains the hash value of a data block and every non-leaf node contains the hash value of child nodes. Trees are generated by hashing pairs of nodes until there is only one hash left, and this hash is called the root hash. It is the efficient and secure way to verify the contents of large data. Due to this feature, it is preferred in designing the blockchain applications.

Smart contract is a self-executing computer code running on top of a blockchain. It contains a set of rules to be executed and helps to exchange money, data or

anything of value. These contracts execute only when certain conditions are met. They reduce transaction costs by eliminating the third parties. Note that a smart contract can never be changed and no one can break the contract. Also it is distributed; the outcome of contract is verified by every node on the network.

Distributed ledger is a decentralized database system for recording transactions. Each node of the network can access the recordings and own an identical copy. Every changes or additions are reflected to each node's ledger. Periodically, nodes construct the transactions and update their copy through a consensus algorithm. Once the correct copy has been determined, all nodes update themselves with the correct copy of the ledger. Cryptographic hash functions and digital signatures are used to ensure the security.

Proof of work is a type of consensus algorithm that is used for confirming transactions and producing new blocks for the chain. When a new transaction is made, it must be validated and written to the block. To do this, users on the network try to solve a mathematical problem at a certain difficulty level. The user, who solves this problem, earns the right to write the next block to the chain. Note that the miners are rewarded with the certain amount of currency in order to be encouraged for mining.

3. Model

In this section, we explain our model and its core components that will be used to build our platform.

Our platform focuses on trading data collected from IoT devices in a safe environment. We utilize a number of primitives to realize our objective: blockchain for immutable and decentralized data storage; smart contracts for exchanging data and money authentically; digital signatures for providing authentication of users; and cryptographic hash functions for providing data integrity.

We design our platform as an open blockchain so that anyone can register the platform to commercialize his IoT data or to buy the IoT data sold in the platform. There are three different roles defined in our system, i.e. sellers, customers, and miners. A seller can advertise and sell his data through the platform by putting a brief information that describes the data, and a short proof that will be used to

prove the authenticity of the data. A customer checks the data advertised in the platform, and buys the data via a smart contract. A miner collects a certain number of available transactions in a certain period of time, and creates a block of these transactions through proof of work mechanism. Note that during the registration, the system generates a secret key-public key pair that will be used in the signature scheme deployed by the platform.

Our model can be viewed as two-phase protocol: uploading the data and executing the trade.

Uploading the data. In our model, the sellers do not keep their data in the blockchain. Instead, they use blockchain to market his data and to prove the integrity of the data. When a seller wants to commercialize his data through the platform, he prepares a small proof associated to the data using the technique proposed in [11].

Briefly, he divides the data into n small pieces $S_1, \dots, S_n$ of same length. Then he creates the Merkle tree of the data as shown in Figure 1. He then uploads the root R of the tree together with a brief information that describes the data. Note that the root of the tree R will be used to convince the customer that the integrity of the data will have been preserved during the marketing.

Executing the trade. Our platform enables the customers to check the quality of the data through a smart contract before buying it. When a customer wants buy some data advertised in platform, he first picks a random integer x from [1, n] as a challenge, and initiates a smart contract with the seller of the data.

The smart contract works as follows: for the challenge x, the seller sends the corresponding piece S_x to the customer. Note that, in our platform, it is assumed that any small piece of the data reflects the general structure of the data, and the quality and the usefulness of the data can easily be checked through this small piece. The customer then examines the sample. If the data satisfies the customer, he sends a confirmation (that can be just the integer 1 and 0 otherwise) to the seller. The seller prepares a proof π for the challenge x that consists of the siblings of the nodes on the path from S_x to the root in the Merkle tree of the data. Then it is checked through the smart contract whether the proof π is compatible with the root R uploaded in the platform.

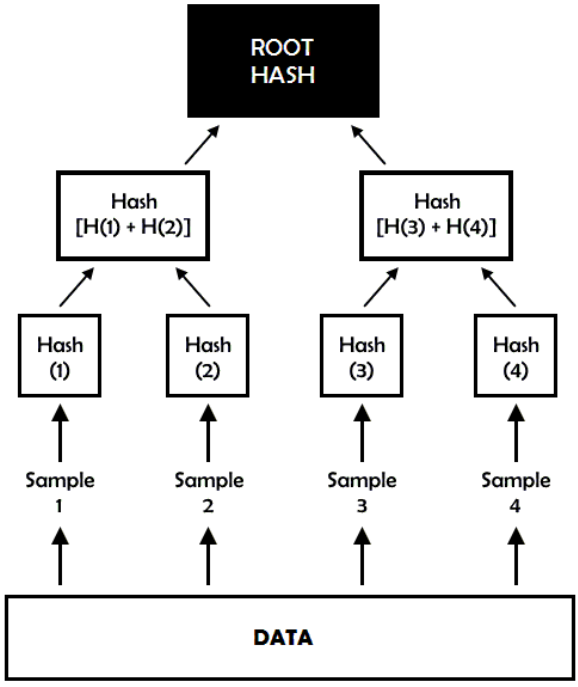


FIGURE 1. Generating Merkle tree from the data

For example, assume $n = 4$, $x = 2$, and the seller has created the Merkle tree of the data as in Figure 1. Then the proof π for the challenge the seller prepares will be $\pi = \left[H(S_1), H\left(H(S_3), H(S_3)\right)\right]$. It is checked whether

$$H\left(H\left(H(S_1), H(S_2)\right), H\left(H(S_3), H(S_4)\right)\right) \stackrel{?}{=} R. \tag{3.1}$$

If the proof is compatible with the root, the customer's money and seller's data are exchanged through the smart contract. Smart contracts allow us to protect both customers and sellers to incur losses in presence of any problems.

Figure 2 shows the basic operations in the smart contract for trading data and money, between sellers and customers. In a nutshell, the proof π of the challenge provided by the seller is compared with the root of the data already uploaded in the blockchain. Besides, it is also checked whether the customer has enough money or not. If one of them fails, the smart contract is terminated unsuccessfully and trade

operation is cancelled. Otherwise, the smart contract is terminated successfully and trade operation is completed.

All the transactions related to the trade are written to the blocks by the miners. Miners use the proof of work concept to establish a consensus for the current state of the ledger among all the entities in the network, i.e. miners collect the available valid transactions at the moment, calculate a proof of work for the new block consisting of the collected transactions, and add it to the valid chain. Miners get rewards corresponding to the effort they made by the system. Besides, a certain amount of fee for each transaction is paid to the miners in order to incentivize them for mining.

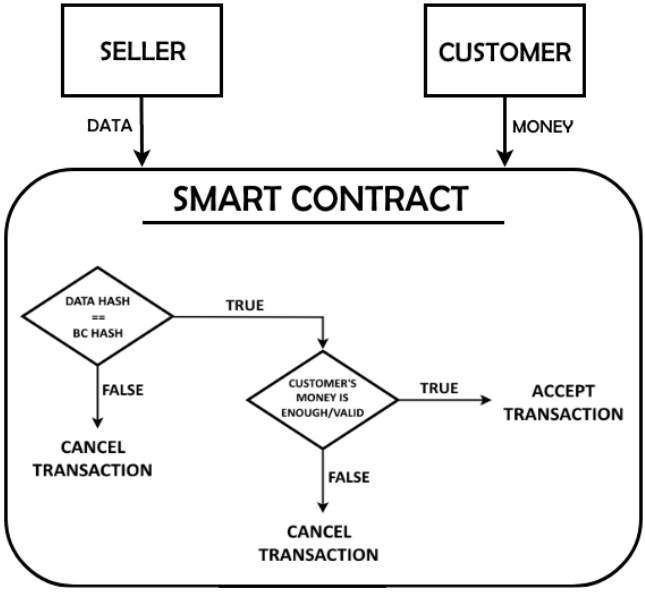


FIGURE. 2. Smart contract for trade operation.

4. Security analysis

This section provides a discussion on security and performance of our blockchain system and its structures.

As we stated before, when the customer requests sample from data, there is a possibility that the seller may send the wrong or useless sample. On the other hand,

the sample may belong to a useful data, but the data to be sent at the end may be an unusable data that is not relevant to the sample. Our platform solves this issue by having the sellers to put the root of the Merkle tree created on the data. Thus, the seller cannot make any change on the data after uploading this proof to the platform.

The users, who will trade among themselves, attach their signature to the transactions they create. The unforgeability of the digital signature scheme ensures that a user cannot impersonate any other users in the system. Also, no one in the system can alter the content of the transaction after it is deployed to the system.

Besides, any malevolent third party node can try to break or alter the transaction and message sequence in the chain. Due to the avalanche effect of the cryptographic hash function, block sequences and integrities can be checked fast and correctly. Even the slightest change in the block causes the hash value to change completely. Therefore, the consensus algorithm will not accept the manipulated block.

5. Conclusion

In this paper, we proposed an application for blockchain based IoT. It aims to market IoT data for money authentically. The system provides an immutable distributed database using blockchain. Besides, it provides confidence between the seller and the customer through smart contracts.

In our future work, we plan to implement this proposed system. We will also explore which file extension the IoT data should have.

References

[1] S. Haber, W.S. Stornetta, “How to time-stamp a digital document”, in Journal of Cryptology, vol. 3, no. 2, pp. 99-111, 1991.

[2] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, J. Wangasdad, “Untangling Blockchain: A Data Processing View of Blockchain Systems” in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 7, pp. 1366-1385, 2018.

[3] P. Ghuli, U. P. Kumar, R. Shettar, “A Review on Blockchain Application for Decentralized Decision of Ownership of IoT Devices” in Advances in Computational Sciences and Technology, vol. 10, no. 8, pp. 2449-2456, 2017.

[4] I. C. Lin, T. C. Liao, "A Survey of Blockchain Security Issues and Challenges" in International Journal of Network Security, vol. 19, no. 5, pp. 653-659, 2017.
[5] "Watson IoT and Blockchain: Disruptor and game changer," https://public.dhe.ibm.com/common/ssi/ecm/ww/en/ww912350usen/watson-iot-cognitive-solutions-ww-infographic-general-ww912350usen-20180306.pdf
[6] S. Huh, S. Cho, S. Kim, "Managing IoT Devices using Blockchain Platform" in 19th International Conference on Advanced Communication Technology, 2017.
[7] S. Huckle, R. Bhattacharya, M. White, N. Beloff , "Internet of Things, Blockchain and Shared Economy Applications" in Procedia Computer Science, vol. 98, pp. 461-466, 2016.
[8] "Ambrosus - Enabling Sensors to Talk to Blockchain," https://ambrosus.com/.
[9] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "LSB: A Lightweight Scalable BlockChain for IoT Security and Privacy", arXiv preprint arXiv: 1712.02969, 2017.
[10] Elbuz, M. Osmanoglu, O. Tanriover, "Application of Blockchain Technology on Internet of Things" in 3rd International Conference on Theoretical and Applied Computer Science and Engineering (ICTACSE), Ankara, Turkey, 2018.
[11] S. Dziembowski, S. Faust, V. Kolmogorov, and K. Pietrzak. "Proof of Space", in Crypto 2015, 2015.

Current Address: Aydın ELBUZ: Ankara University, Computer Engineering Department, Ankara, Turkey
E-mail: aelbuz37@gmail.com,
Orcid ID: https://orcid.org/0000-0001-5932-8754
Current Address: Murat OSMANOGLU: Ankara University, Computer Engineering Department, Ankara, Turkey
E-mail: murat.osmanoglu@ankara.edu.tr,
Orcid ID: https://orcid.org/0000-0001-8693-141X
Current Address: Omer Ozgur TANRIOVER: Ankara University, Computer Engineering Department, Ankara, Turkey
E-mail: tanriover@ankara.edu.tr
Orcid ID: https://orcid.org/0000-0003-0833-3494

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 1, Pages 111-128 (2019)
DOI: 10.33769/aupse.560812
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

INVESTIGATING DOUBLY CHARGED LEPTONS AT FUTURE ENERGY FRONTIER MUON-PROTON COLLIDERS

Aysuhan OZANSOY

ABSTRACT. We study the doubly charged leptons considered in extended isospin models at muon-proton colliders. We respect the lepton flavor conservation and take into consideration the single production of the doubly charged leptons related to second generation. We give the effective Lagrangian describing the doubly charged lepton gauge interactions. We calculate the signal and corresponding background cross sections and analyze the kinematical distributions to obtain the suitable cuts for the discovery. We choose the W-boson hadronic decay channel to get the accessible mass limits and couplings of doubly charged leptons for various muon-proton colliders.

1. INTRODUCTION

The fundamental particle spectrum of the Standard Model (SM) has a pattern with three generations of quarks and leptons. These generations are replication of each other except from their mass. A natural explanation for the SM fermionic family replication would be that there could be one more layer of matter constituents. In composite models, known fermions are bound states of more fundamental constituents called preons [1-3]. Compositeness is one of the beyond the SM (BSM) theories that give satisfying explanations for the fermionic family replication, quark-lepton symmetry and fundamental particle inflation in the SM. Observation of excited fermionic states would be a direct evidence of compositeness [4-8]. Spin and isospin-1/2 excited fermions are considered as lowest radial and orbital excited

Received by the editors: May 06, 2019; Accepted: May 31, 2019.
Key word and phrases: Doubly charged leptons, muon-proton colliders, weak isospin.

states. Excited leptons with weak isospin states IW=0 (singlet) and IW=1/2 (doublet) are studied widely in the literature [9- 25].

Here, we take into consideration a different point of view of the compositeness: the weak isospin (IW) invariance. In extended isospin models the weak isospin values are extended to IW =1 (triplet) and IW =3/2 (quadruplet) multiplets. Excited leptonic states with exotic charges with $Q = -2e$ (doubly charged leptons) take part in these exotic multiplets [26]. Any signal for the doubly charged leptons at future colliders would give considerable explanation for the SM flavor structure and hints for the new physics.

There are great effort to build more powerful particle colliders for the Large Hadron Collider (LHC) era and post LHC era. These projects can be classified in three groups concerning the colliding particle types:

1) *Lepton colliders:* International Linear Collider (ILC) with $\sqrt{s} = 0.5 - 1$ TeV [27] and Compact Linear Collider (CLIC) with up to $\sqrt{s} = 3$ TeV [28] as e^+e^- colliders, and muon colliders with center-of-mass energies from Higgs resonance up to $\sqrt{s} = 6$ TeV [29, 30]. Recently, a 100 TeV muon collider was proposed in [31].
2) *Hadron colliders:* Future Circular Collider (FCC) is proposed as *pp* collider with $\sqrt{s} = 100$ TeV [32] at CERN. (Besides the *pp* option, FCC has also an e^+e^- option known as TLEP or FCC-*ee* [33] and several *ep* collider options [34]). Super proton-proton Collider (SppC) is Chinese analog of the FCC with center-of-mass energy about 70 TeV [35].
3) *Lepton-hadron colliders:* Large Hadron-electron Collider (LHeC) with $\sqrt{s} = 1.3$ TeV and possibly upgraded to $\sqrt{s} = 1.96$ TeV [36, 37] is proposed as *ep* collider at CERN. Other designments are FCC-based lepton-hadron (*ep* and μp colliders) [34], SppC-based lepton-hadron colliders [38], and LHC-based muon-proton colliders [39].

Aside from compositeness, one may encounter doubly charged leptonic states in Type II seesaw mechanisms [40-42], in some extensions of supersymmetric models [43-47], in flavor models in warped extra dimensions and in more general models [48-49], and also in string inspired models [50]. In the literature, phenomenological searches for doubly charged leptons are investigated so far at the LHC [51-60], at the future linear colliders [61-64], at the LHeC [65], and at the FCC-based ep colliders [66].

In this work, we give the effective Lagrangians responsible for the gauge interactions of doubly charged leptons and calculate the decay widths in Section 2. In Section 3, we introduce the various muon-proton colliders. We give our analysis in Section 4 and then we conclude.

2. Effective lagrangians and decay widths

In the beginning of hadron physics, strong isospin invariance enlightened so much to find out some patterns of baryon and meson resonances despite the existence of quarks and gluons were not understood yet. Identically, similar situations may be occured in the electroweak sector. By using the weak isospin symmetry arguments, the possible fermionic resonances can be placed into electroweak isospin multiplets. Accordingly, the quantum numbers of excited fermionic states could be obtained without needing the dynamics of the preons explicitly. Since light fermions have $I_W = 0$ or $1/2$ (singlets or doublets), and gauge bosons have $I_W = 0$ or 1; excited fermionic states with $I_W \leq 3/2$ can be allowed. Basics of extended isospin model is discussed in [26]. In the extended isospin models exotic doubly charged leptons appear in triplets with $I_W=1$ and in quadruplets with $I_W=3/2$. The form of these multiplets are listed as

$$L_1 = \begin{pmatrix} L^0 \\ L^- \\ L^{--} \end{pmatrix}, \qquad L_{3/2} = \begin{pmatrix} L^+ \\ L^0 \\ L^- \\ L^{--} \end{pmatrix} \tag{1}$$

and similar for the antiparticles. To calculate the decay widths and cross sections, it is necessary to specify the doubly charged lepton couplings to SM leptons and gauge fields. For the reason that all the gauge fields have $Y=0$ hypercharge, a given exotic multiplet couples through the gauge fields to a SM multiplet only with the same Y. To satisfy the gauge invariance, the couplings have to be of anomalous magnetic moment type. Therefore, doubly charged leptons can couple to SM fermions only via W-boson. The interaction Lagrangians describing the gauge interaction of doubly charged leptons with ordinary leptons are

$$\mathcal{L}_{GM}^{(1)} = i\frac{gf_1}{\Lambda}\left(\bar{L}\sigma_{\mu\nu}\partial^{\nu}W^{\mu}\frac{1+\gamma_5}{2}l\right)+h.c. \qquad (2)$$

$$\mathcal{L}_{GM}^{(3/2)} = i\frac{gf_3}{\Lambda}\left(\bar{L}\sigma_{\mu\nu}\partial^{\nu}W^{\mu}\frac{1-\gamma_5}{2}l\right)+h.c. \qquad (3)$$

where g is the $SU(2)$ coupling constant which is equals to $g = g_e/\sin\theta_W$ where $g_e = \sqrt{4\pi\alpha}$. f_1 and f_3 are dimensionless new coupling constants related to effective interactions of exotic multiplets I_W=1 and I_W=3/2. Their exact values can only be obtained from the underlying model for compositeness. L and l stand for doubly charged lepton and SM lepton, respectively. Λ is the compositeness scale and $\sigma_{\mu\nu} = i\left(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu\right)/2$ where γ_μ being the Dirac matrices.

Since doubly charged leptons can interact with the SM leptons only via the W-boson, doubly charged leptons decay to SM leptons only with $L^{--} \to l^{-}W^{-}$ process (L^{--} is the lightest one among the new exotic leptons). The same values of decay width for the doubly charged leptons are obtained whether one can use the interaction Lagrangian given in Eq. (3) or Eq. (4). The decay width of the doubly charged lepton with respect to its mass (M_L) for f_1=1 (or f_3=1) is given in Figure 1 for two different values of compositeness scale.

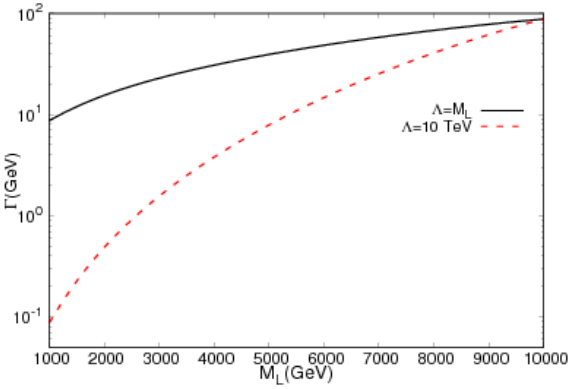


FIGURE 1. Total decay width of doubly charged lepton as a function of its mass for $\Lambda=M_L$ and Λ=10 TeV.

3. MUON-PROTON COLLIDERS

Muon-proton colliders were proposed in the original work [67] by making the suggestion of construction of an additional proton ring in $\sqrt{s}$=4 TeV muon collider tunnel. Afterwards, construction of additional muon ring with 200 GeV energy in the Tevatron tunnel in order to handle $\sqrt{s}$=0.9 TeV μp collider was considered in [68], and ultimate μp collider with 50 TeV proton ring was suggested in [69]. Recently, future pp colliders FCC and SppC based energy frontier muon-proton colliders were proposed in [34] and [38], respectively. Also LHC-based μp colliders were considered in [39]. We list briefly the main parameters of proposed μp colliders in Tables 1 – 3.

A lot of BSM topics such as excited muon, excited muon neutrino, second family leptoquarks, supersymmetry (SUSY), doubly charged leptons related to second family, color octet muon, contact interactions etc. can be investigated at μp colliders.

Table 1. Main parameters of LHC-based μp colliders. (*These values are obtained from* [39]).

E_μ (TeV)	E_p (TeV)	$\sqrt{s}$ (TeV)	$L_{\mu p}$ ($cm^{-2}s^{-1}$)
0.75	7	4.58	1.4×10^{33}
1.5	7	6.48	2.3×10^{33}
3	7	9.16	0.9×10^{33}

Table 2. Main parameters of SppC-based μp colliders. (*These values are obtained from* [38]).

E_μ (TeV)	E_p (TeV)	$\sqrt{s}$ (TeV)	$L_{\mu p}$ ($cm^{-2}s^{-1}$)
0.75	35.6	10.33	5.5×10^{32}
0.75	68	14.28	12.5×10^{32}
1.5	35.6	14.61	4.9×10^{32}
1.5	68	20.2	42.8×10^{32}

Table 3. Main parameters of FCC-based μp colliders. (*These values are obtained from* [34]).

E_μ (TeV)	E_p (TeV)	$\sqrt{s}$ (TeV)	$L_{\mu p}$ ($cm^{-2}s^{-1}$)
0.063	50	3.50	0.2×10^{31}
0.75	50	12.2	4.9×10^{32}
1.5	50	17.3	4.3×10^{32}

4. MUON-PROTON COLLIDERS

Respecting the lepton family number conservation, we searched for the doubly charged leptons which carry muonic lepton number. They can be produced singly through the process $\mu^- p \to L^{--} X$ at μp colliders. The Feynman diagrams representing the subprocess $\mu^- q(\bar{q}') \to L^{--} q'(\bar{q})$ are given in Figure 2.

FIGURE 2. Feynman diagrams for the subprocess $\mu^- q(\bar{q}') \to L^{--} q'(\bar{q})$.

In our analysis, we considered only I_W=1 multiplet. We embedded the doubly charged lepton interaction vertices in high-energy simulation programme CALCHEP [70-72] and used it for our calculations. To make a comparision between the μp colliders listed in Table 1-3; we chose the muon beam energy as 0.75 TeV and proton beam energies 7, 35.6, and 50 TeV for the LHC, SppC, and FCC, respectively. We summarized our selection in Table 4.

Table 4. Selected μp colliders to make a comparision in our calculations.

μp Collider Name	E_p (TeV)	E_μ (TeV)	$\sqrt{s}$ (TeV)
LHC-μp	7	0.75	4.58
SppC-μp	35.6	0.75	10.33
FCC-μp	50	0.75	12.2

Total production cross section for the single production of doubly charged leptons
($\mu^- p \to L^{--} X$) at various μp colliders for Λ=M_L and f=1 is shown in Figure

3. (Since we take into consideration only I_W=1 multiplet, f is refer to f_1 henceforth). In numerical calculations we used the CTEQ6L parton distribution function.

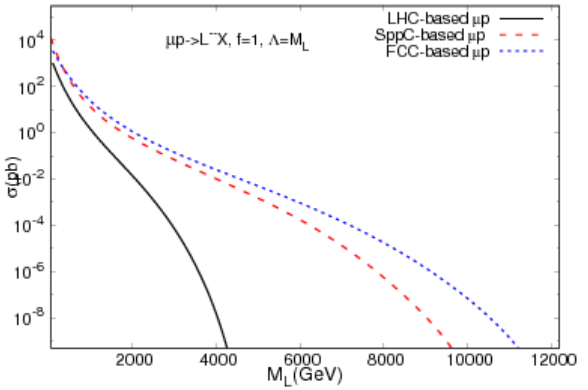


FIGURE 3. Total cross section for the single production of doubly charged leptons at various μp colliders.

After the decay of L^{--}, we deal with the $\mu^- q(\bar{q}') \to W^- \mu^- q'(\bar{q})$ process. We impose the generic cuts for the final state muon and jets as

$$\begin{aligned} p_T^{\mu} &> 20\ GeV \\ p_T^{j} &> 30\ GeV. \end{aligned} \tag{4}$$

By applying the generic cuts we get the transverse momentum (p_T) and normalized pseudorapidity (η) distributions of final state muon. We show these kinematical distributions in Figures 4-9.

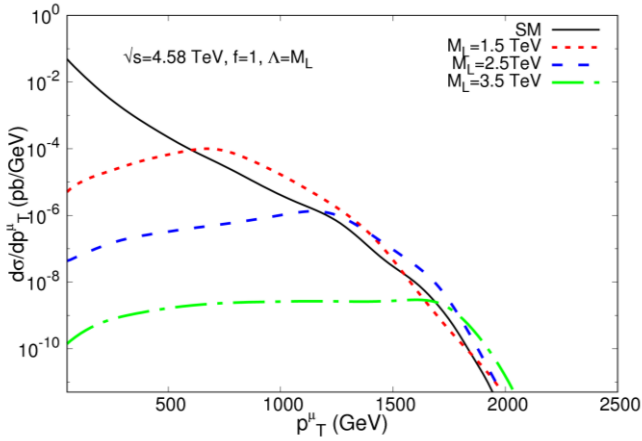


FIGURE 4. Muon p_T distribution for the LHC-μp.

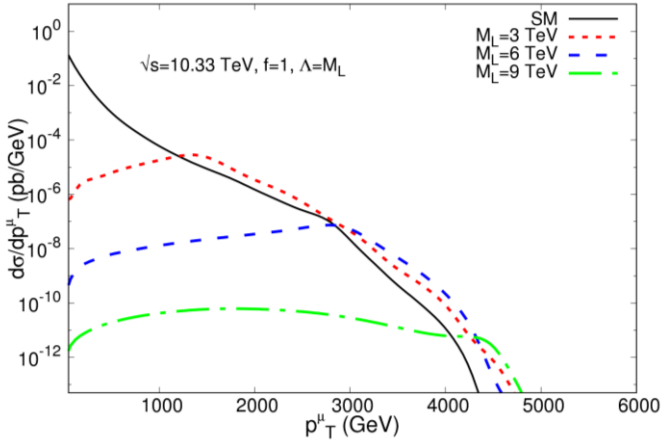


FIGURE 5. Muon p_T distribution for the SppC-μp.

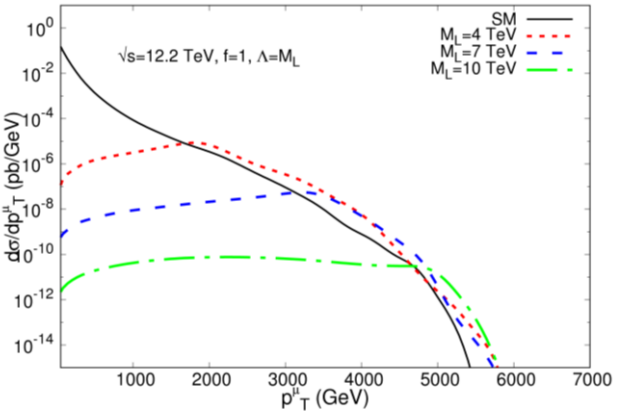


FIGURE 6. Muon p_T distribution for the FCC-μp.

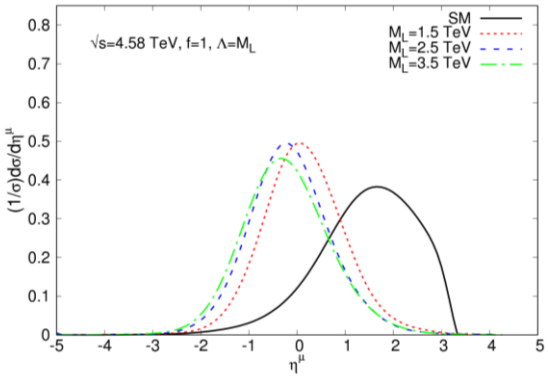


FIGURE 7. Muon normalized η distribution for the LHC-μp.

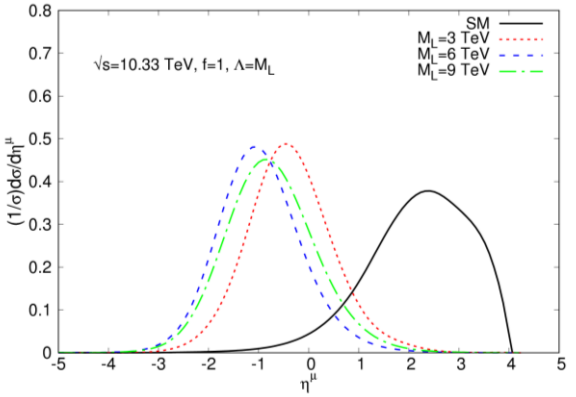


FIGURE 8. Muon normalized η distribution for the SppC-μp.

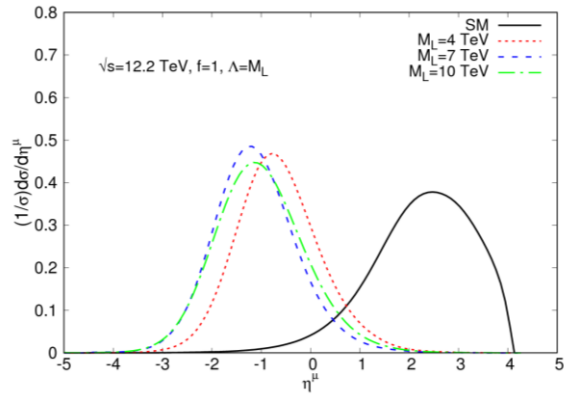


FIGURE 9. Muon normalized η distribution for the FCC-μp.

As it can be seen from Figures 4-6, doubly charged leptons carry high transverse momentum having a peak around M_L / 2, and from Figures 7-9, they are produced mainly in the backward direction. We define the cuts best suited for the discovery by investigating these kinematical distributions. To find out the discovery cuts we seek for the appropriate regions where we remove the most of the background but at the same time do not influence the signal so much. These cuts are presented in Table 5.

Table 5. Discovery cuts fot the doubly charged lepton production at μp colliders.

μp **Collider Name**	p_T^μ ***cut***	η^μ **cut**
LHC-μp	$p_T^\mu > 600$ GeV	$-2 < \eta^\mu < 1$
SppC-μp	$p_T^\mu > 1200$ GeV	$-2.5 < \eta^\mu < 0.9$
FCC-μp	$p_T^\mu > 1700$ GeV	$-2.5 < \eta^\mu < 0.6$

We choose hadronic decay mode of W-boson as $W \to 2j$ and no further cut is made on these jets. We define the statistical significance as

$$SS = \frac{\sigma_S}{\sqrt{\sigma_B}} \sqrt{L_{\text{int}}} \tag{5}$$

where σ_S and σ_B are the signal and background cross sections, respectively; L_{int} is the integrated luminosity of the collider.

We show the SS - M_L plot for the LHC-based μp collider in Figure 10, for the SppC-based μp collider in Figure 11, and for the FCC-based μp collider in Figure 12, specifying the 2σ (exclusion), 3σ (observation), and 5σ (discovery) regions, respectively. In Table 6, we determine

the doubly charged lepton mass limits at different μp colliders for f=1 and Λ=M_L, taking into account the criteria $SS \geq 2$, $SS \geq 3$ and $SS \geq 5$ which denote the 2σ, 3σ and 5σ limits, respectively.

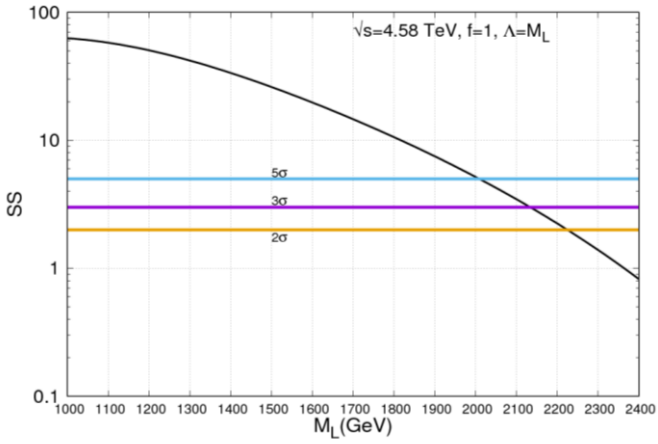


FIGURE 10. *SS* as a function of M_L for the LHC-μp.

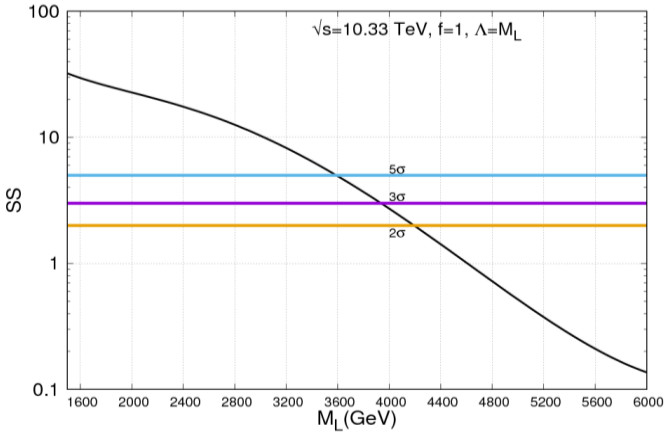


FIGURE 11. *SS* as a function of M_L for the SppC-μp.

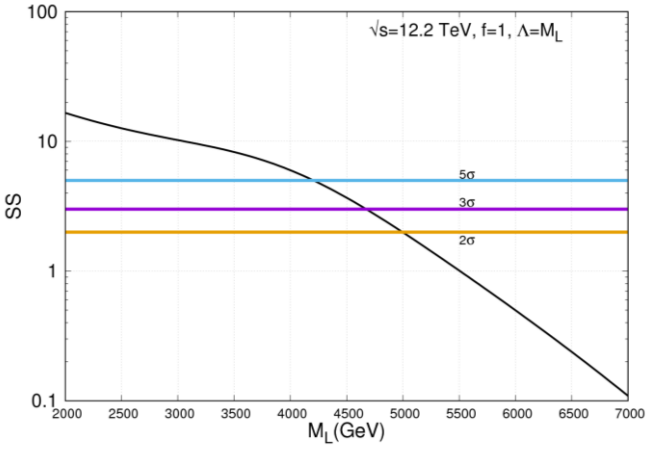


FIGURE 12. *SS* as a function of M_L for the FCC-μp.

Table 6. Mass limits for doubly charged lepton at different μp colliders.

Collider name	L_{int} (pb^{-1})	2σ (Exclusion)(GeV)	3σ (Observation)(GeV)	5σ (Discovery) (GeV)
LHC-μp (√s = 4.58 TeV)	1.4×10^4	2220	2140	2000
SppC-μp (√s = 10.33 TeV)	5.5×10^3	4200	3900	3600
FCC-μp (√s = 12.2 TeV)	4.9×10^3	5000	4700	4200

5. CONCLUSION

The weak isospin invariance is a particular point of view of the compositeness. We can use the weak isospin invariance to extend the weak isospin values to $I_W = 1$ (triplet) and $I_W = 3/2$ (quadruplet) multiplets. These multiplets contain exotic lepton states of charge $Q = -2e$, namely doubly charged leptons. μp colliders will give the oppurtunity to search for the doubly charged leptons related to second family. In our

study, we make a comparision for the search potential of doubly charged leptons at future energy frontier muon-proton colliders. Taking into account the lepton flavor conservation, we showed doubly charged leptons carrying muonic lepton number can be observed up to 2.1, 3.9, and 4.7 TeV at LHC-μp, SppC-μp and FCC-μp, respectively.

References

[1] Terazawa H., Chikashige Y., and Akama K., Unified Model of the Nambu-Jona-Lasino Type for all Elementary Particle Forces, Phys. Rev. D 15-2 (1977) 480-487.

[2] Neeman Y., Primitive Particle Model, Phys. Lett. 82B-1 (1979) 69-70.

[3] Terazawa H., Yasue M., Akama K., and Hayashi M., Observable Effects of the Possible Sub-structure of Leptons and Quarks, Phys. Lett. 112B 4-5 (1982) 387-392.

[4] Renard F. M., Excited Quarks and New Hadronic States, Il Nuovo Cimento A 77-1 (1983) 1-20.

[5] Eichten E. J., Lane K. D., and Peskin M. E., New Tests for Quark and Lepton Substructure, Phys. Rev. Lett. 50-11 (1983) 811-814.

[6] de Rujula A., Maiani L., and Petronzio R., Search for Excited Quarks, Phys. Lett. 140B 3-4 (1984) 253-258.

[7] Cabibbo N., Maiani L., and Srivastava Y. N., Anomalous Z Decays: Excited Leptons?, Phys. Lett. 139B 5-6 (1984) 459-463.

[8] Kühn J. and Zerwas P. M., Excited Quarks and Leptons, Phys. Lett. 147B 1-3 (1984) 189-196.

[9] Hagiwara K., Komamiya S., and Zeppenfeld D., Excited Lepton Production at LEP and HERA, Z. für Physik C 29-1 (1985) 115-122.

[10] Boudjema F., Djouadi A., and Kneur J. L., Excited Fermions at e^+e^- and eP Colliders, Z. für Physik C 57-3 (1993) 425-449.

[11] Cakir O., Yilmaz A., and Sultansoy S., Single Production of Excited Electrons at Future e^+e^-, ep and pp Colliders, Phys. Rev. D 70-7 (2004) 075011.

[12] Cakir O., Turk Cakir I., and Kirca Z., Single Production of Excited Neutrinos at Future e^+e^-, ep and pp Colliders Phys. Rev. D 70-7 (2004) 075017.

[13] Caliskan A., Kara S. O., and Ozansoy A., Excited Muon Searches at the FCC-Based Muon-Hadron Colliders, Adv. High Energy Phys. 2017 (2017) 1540243.

[14] Caliskan A., Excited Neutrino Search Potential of the FCC-Based Electron-Hadron Colliders, Adv. High Energy Phys. 2017 (2017) 4726050.

[15] Caliskan A. and Kara S. O., Single Production of the Excited Electrons at the Future FCC-Based Lepton-Hadron Colliders, Int. J. Mod. Phys. A 33-24 (2018) 1850141.

[16] Ginzburg I. F. and Ivanov D. Yu., Excited Leptons and Quarks at $\gamma\gamma$ / γe Colliders, Phys. Lett. B276 1-2 (1992) 214-218.

[17] Koksal M., Analysis of Excited Neutrinos at the CLIC, Int. J. Mod. Phys. A 29-24 (2014) 1450138.

[18] Ozansoy A. and Billur A. A., Search for Excited Electrons Through γγ Scattering, Phys. Rev. D 86-5 (2012) 055008.
[19] Kirca Z., Cakir O., and Aydin Z. Z., Production of Excited Electrons at TESLA and CLIC Based e gamma Colliders, Acta Phys. Polon. B 34-8 (2003) 4079.
[20] Eboli O. J., Lietti S. M., and Mathews P., Excited Leptons at the CERN Large Hadron Collider, Phys. Rev. D 65-7 (2002) 075003.
[21] Inan S. C., Exclusive Excited Leptons Search in Two Lepton Final States at the CERN LHC, Phys. Rev. D81-11 (2010) 115002.
[22] Cakir O., Leroy C., Mehdiyev R., and Belyaev A., Production and Decay of Excited Electrons at the LHC, Eur. Phys. J. C 32-2 (2004) s1-s17.
[23] Belyaev A., Leroy C., and Mehdiyev R., Production of Excited Neutrinos at the LHC, Eur. Phys. J. C 41-2 (2005) 1-10.
[24] Boos E., Volodgin A., Toback D., and Gaspard J., Prospects of Searching for Excited Leptons During Run II of the Fermilab Tevatron, Phys. Rev. D 66-1 (2002) 013011.
[25] Baur U., Spira M., and Zerwas P. M., Excited Quark and Lepton Production at Hadron Colliders, Phys. Rev. D 42-3 (1990) 815-824.
[26] Pancheri G. and Srivastava Y. N., Weak Isospin Spectroscopy of Excited Quarks and Leptons, Phys. Lett. 146B (1984) 87-94.
[27] Carwardine J. (Editorial Board Chair), The International Linear Collider Technical Design Report – Volume 3. II: Accelerator Baseline Design, (2013) https://arxiv.org/abs/1306.6328.
[28] Aicheler M., Burrows P., Draper M., Garvey T., Lebrun P., Peach K., Phinney N., Schmickler H., Schulte D., and Toge N. (editors), A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report, CERN-2012-007, KEK Report JAI-2012-001, 2012-1, PSI-12-01, SLAC-R-985, CERN, Geneva, Switzerland, (2012)
[29] https://edms.cern.ch/ui/file/1234244/7/CERN-2012-007.pdf.
[30] Delahaye J-P., Ankenbrandt C., Bogacz A., Brice S., Bross A., Denisov D., Eichten E., Huber P., Kaplan D. M., Kirk H., Lipton R., Neuffer D., Palmer M. A., Palmer R., Ryne R., and Snopok P. (editors), Enabling Intensity and Energy Frontier Science with a Muon Accelerator Facility in the U.S. : A White Paper Submitted to the 2013 U.S. Community Summer Study of the Division of Particles and Fields of the American Society, FERMILAB-CONF-13-307-APC (2013), https://arxiv/org/abs/1308.0494
[31] Palmer R. B., Muon Colliders, Rev. Accel. Science and Technology 7 (2014) 137-159.
[32] Zimmermann F., LHC/FCC-Based Muon Colliders, 9th Int. Particle Accel. Conference, IPAC2018, doi:10.18429/JACow-IPAC2018-MOPMF065 (2018) 273-276.
[33] Future Circular Collider Study (FCC), https://fcc.web.cern.ch
[34] Bicer M. et al. (The TLEP Design Study Working Group), First Look at the Physics Case of TLEP, Journal of High Energy Physics 2014-1 (2014) 164.
[35] Acar Y. C., Akay A. N., Beser S., Canbay A. C., Karadeniz H., Kaya U., and Oner B.B., Future Circular Collider Based Lepton-Hadron and Photon-Hadron Colliders: Luminosity and Physics, Nuclear Inst. and Methods in Phys. Research A 871 (2017) 47-53.

[36] Su F., Gao J., Xiao M., Wang D., Wang Y-W., Bai S., Bian T-J., Method Study of Parameter Choice for a Circular Proton-Proton Collider Chin. Phys. C 40-1 (2016) 017001.
[37] Abelleira Fernandez J. L. et al. (LHeC Study Group), A Large Hadron Electron Collider at CERN, Journal of Phys. G 39 (2012) 075001.
[38] Bruening O. and Klein M., The Large Hadron Electron Collider, Mod. Phys. Lett. A 28 - 16 (2013) 1330011.
[39] Canbay A. C., Kaya U., Ketenoglu B., Oner B. B., and Sultansoy S., Adv. in High Energy Phys. 2017 (2017) 4021493.
[40] Kaya U., Ketenoglu B., and Sultansoy S., The LHC Based μp Colliders, (2018) https://arxiv.org/abs/1807.09867.
[41] Chua C. K. and Law S. S. C., Phenomenological Constraints on Minimally Coupled Exotic Lepton Triplets, Phys. Rev D 83 (2011) 055010.
[42] Foot R., Lew H., He X. G., and Joshi G. C., Seesaw Neutrino Masses Induced by a Triplet of Leptons, Zür. Phys. C 44 (1989) 441-444.
[43] Kumericki K., Picek I., and Radovcic B., Exotic Seesaw-Motivated Heavy Leptons at the LHC, Phys. Rev. D 84 (2011) 093002.
[44] Demir D. A., Frank M., Huitu K., Kumar Rai S., and Turan I., Signals of Doubly-Charged Higgsinos at the CERN Large Hadron Collider, Phys. Rev. D 78 (2008) 035013.
[45] Franceschini M. and Mohapatra R., Radiatively Induced Type II Seesaw Models and Vectorlike 5/3 Charge Quarks, Phys. Rev. D 89-5 (2014) 055013.
[46] Dutta B., Mohapatra R. N., and Muller D. J., The Signature at the Tevatron for the Light Doubly Charged Higgsino of the Supersymmetric Left-Right Model, Phys. Rev. D 60 (1999) 095005.
[47] Chacko Z. and Mohapatra R. N., Supersymmetric Left-Right Model and Light Doubly Charged Higgs Bosons and Higgsinos, Phys. Rev. D 58 (1998) 015003.
[48] Frank M., Doubly Charged Higgsino Mediated Lepton Flavor Violating Decays, Phys. Rev. D 62 (2000) 053004.
[49] Cirelli M., Fornengo N., and Strumia A., Minimal Dark Matter, Nucl. Phys. B 753 (2006) 178-194.
[50] del Aguila F., de Blas J., and Perez-Victoria M., Effects of New Leptons in Electroweak Precision Data, Phys. Rev. D 78 (2008) 013010.
[51] Cvetic M., Halverson J., and Langacker P., Implications of String Constraints for Exotic Matter and Z' s Beyond the Standart Model, Journal of High Energy Phys. 1111 (2011) 058.
[52] Delgado A., Garcia Cely C., Han T., and Wang Z., Phenomenology of Lepton Triplet, Phys. Rev. D 84 (2011) 073007.
[53] Biondini S., Panella O., Pancheri G., Srivastava Y. N., and Fano L., Phenomenology of Excited Doubly Charged Heavy Leptons at the LHC, Phys. Rev. D 85 (2012) 095018.
[54] Alloul A., Frank M., Fuks B., and de Traubenberg M. R., Doubly-Charged Particles at the Large Hadron Collider, Phys. Rev. D 88 (2013) 075004.
[55] Leonardi R., Panella O., and Fano L., Doubly Charged Heavy Leptons at LHC via Contact Interactions, Phys. Rev. D 90-3 (2014) 035001.

[56] You Y., Chong-Xing Y., and Yun X., Pair Production of the Doubly Charged Leptons via Electroweak Vector Boson Fusion at the Large Hadron Collider, Chin. Phys. Lett. 31 (2014) 021201.
[57] Ma T., Zhang B., and Cacciapaglia G., Triplet with a Doubly-Charged Lepton at the LHC, Phys. Rev. D 89-1 015020 (2014); Doubly Charged Lepton from an Exotic Doublet at the LHC, Phys. Rev. D 89-9 (2014) 093022.
[58] Yu Y., Yue C-X., and Yang S., Signatures of the Quintuplet Leptons at the LHC, Phys. Rev. D 91-9 (2015) 093003.
[59] Okada H. and Yagyu K., Three-loop Neutrino Mass Model with Doubly Charged Particles from Isodoublets, Phys. Rev. D 93-1 (2016) 013004.
[60] Chen C-H. and Nomura T., Bounds on LFV Higgs Decays in a Vector-like Lepton Model and Searching for Doubly Charged Leptons at the LHC, Eur. Phys. J. C 76 (2016) 353.
[61] Leonardi R., Alunni L., Romeo F., Fano L., and Panella O., Hunting for Heavy Composite Majorana Neutrinos at the LHC, Eur. Phys. J. C 76 -11 (2016) 593.
[62] Zeng Q-G., Ji L., and Yang S., Pair Production of the Doubly Charged Leptons Associated with a Gauge Boson γ or Z in e^+e^- and $\gamma\gamma$ Collisions at Future Linear Colliders, Commun. Theor. Phys. 63-3 (2015) 331-339.
[63] Biondini S. and Panella O., Exotic Leptons at Future Linear Colliders, Phys. Rev. D 92-1 (2012) 015023.
[64] Guo Y-C., Yue C-X., Lui Z-C, The Signatures of Doubly Charged Leptons in Future Linear Colliders, J. Phys. G. Nucl. Part. Phys. 44 (2017) 085004.
[65] Zeng Q-G., Production of the Doubly Charged Leptons at the ILC, Eur. Phys. Lett. 111 (2105) 21003.
[66] Yu Y., Guo Y-C., Lui Z-C., Fan W-J., Mei Y., and Zhang J., The Signatures of Doubly Charged Leptons at LHeC, J. Phys. G. Nucl. Part. Phys. 45 (2018) 125003.
[67] Ozansoy A., Doubly Charged Lepton Search Potential of the FCC-Based Energy Frontier Electron-Proton Colliders, (2019) in preparation .
[68] Ginzburg I. F., Physics at Future ep, γp (linac-ring) and μp Colliders, Turk. J. Phys. 22-7, (1998) 607-610.
[69] Shiltsev V. D., An Asymmetric Muon-Proton Collider: Luminosity Consideration, Proceedings of the Part. Accelerator Conf., Vancouver Canada, Conf. Proc. C970512 (1997) 420-421.
[70] Sultansoy S., The Post-HERA Era: Brief Review of Future Lepton-Hadron and Photon-Hadron Colliders, (1999) https://arxiv.org/abs/hep-ph/9911417.
[71] Belyaev A., Christensen N. D., and Pukhov A., CalcHEP 3.4 for Collider Physics Within and Beyond the Standard Model, Computer Phys. Commun. 184-7 (2013) 1729-1769.
[72] Pukhov A., CalcHEP 2.3: MSSM, Structure Functions, Event Generation, Batchs, and Generation of Matrix Elements for other Packages, (2004) https://arxiv.org/abs/hep-ph/0412191.

[73] Pukhov A., Boos E., Dubinin M., Edneral V., Ilyin V., Kovalenko D., Kryukov A., Savrin V., Shichanin S., and Semenov A., CompHEP- a Package for Evaluation of Feynman Diagrams and Integration Over Multi-particle Phase Space. User's Manual for Version 33, (1999) https://arxiv.org/abs/hep-ph/9908288.

Current Address: Aysuhan OZANSOY: Ankara University, Faculty of Sciences, Department of Physics, Ankara, TURKEY
E-mail: aozansoy@science.ankara.edu.tr
Orcid ID: https://orcid.org/0000-0002-9042-1372

INSTRUCTIONS TO CONTRIBUTORS

Communications Faculty of Sciences University of Ankara Series A2-A3: Physical Sciences and Engineering is a peer reviewed journal which has been published since 1948 by Ankara University, accepts original research articles written in English in the fields of Physics, Engineering Physics, Electronics/Computer Engineering, Astronomy and Geophysics. Review articles written by eminent scientists can also be invited by the Editor.
Manuscripts should be submitted as a single PDF file attached to an e-mail with a covering letter. In the covering letter, authors should nominate three potential reviewers and e-mailed the file to the most appropriate Area Editor of the research. The editorial office may not use these nominations, but this may help to speed up the selection of appropriate reviewers.
Manuscripts should be typeset using the LATEX typesetting system. Authors should prepare the article using the COMMUNICATIONS style before submission by e-mail. Manuscripts written in DOC form are also acceptable. A template of manuscript can be reviewed in **https://dergipark.org.tr/communc.** After the acceptance of manuscripts for publication, we will ask you to submit the **TeX** form of the manuscript prepared in accordance with the style of the Journal. Authors are required to submit their Open Researcher and Contributor ID (**ORCID**) 's which can be obtained from http://orcid.org as their URL address in the format http://orcid.org/xxxx-xxxx-xxxx-xxxx. Acknowledgements should be given as short as possible at the end of the text. Formulas should be numbered consecutively in parentheses ( ). Footnotes should be avoided if possible, but when necessary, should be short and never contain any important part of the work and should be numbered consecutively by superscripts. All illustrations not including tables (photographs and other films, drawings, graphs, etc) must be labeled as "Figure". The proper position of each table and figure must be clearly indicated in the paper.

All tables and figures must have a number (Table 1, Figure 1) and a caption or legend. References including comments must be numbered consecutively in order of first appearance in the text. The reference number should be put in brackets [] where referred to in the text. References should be listed at the end of the manuscript in the numbered order in which they appear in the text as follows:

[1] Bairamov, E, Ozalp N., Uniform convergence and numerical computation of the Hubbell radiation rectangular source integral, *Radiation Physics and Chemistry*, 80 (2011) 1312–1315.
[2] Kelley, J. L., General Topology, Van Nostrand, 1970.

It is a fundamental condition that articles submitted to COMMUNICATIONS have not been previously published and will not be simultaneously submitted or published elsewhere. After the manuscript has been accepted for publication, the author will not be permitted to make any new additions to the manuscript.

Before publication, the galley proof is sent to the author for correction. Thus, it is solely the author's responsibility for any typographical mistakes occur in their article as it appears in the Journal. The contents of the manuscript published in the COMMUNICATIONS are the sole responsibility of the authors.

The PDF copies of accepted papers are free of charges, but hard copies of the paper, if required, are due to be charged for the amount of which is determined by the administration each year.

Editor in Chief
Commun. Fac. Sci. Univ. Ank. Ser. A2-A3.
Ankara University, Faculty of Sciences
06100, Besevler - ANKARA TURKEY

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 61 **Number : 1** **Year :2019**