ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 7-ISSUE 1|JAN-FEB 2022

www.astesj.com
ISSN: 2415-6698

Editorial

We are pleased to present this issue featuring 11 accepted research articles that showcase recent advances across a diverse range of topics in science and engineering.

The issue opens with Okeke and Eze's paper exploring the impact of self-regulated learning strategies on the academic performance of physics students in secondary schools in Nigeria [1]. Using a quasi-experimental design, they found that students taught with self-regulated learning methods had significantly higher test scores compared to those taught using conventional teaching methods. These results highlight the potential benefits of self-regulated learning approaches to improve physics education outcomes.

Shifting focus to the realm of online education, Lin and colleagues present a case study examining student perceptions and learning outcomes using the game-based learning platform Kahoot in an online crop production course [2]. Their findings indicated that Kahoot quizzes were effective for reinforcing both theoretical knowledge and visual plant species identification, while also being enjoyable for most students. This provides insight into effective online active learning tools.

In the computer science domain, Kadappan et al. introduce a novel framework of value trace problems to support the self-study of code reading and comprehension skills in C programming [3]. Through evaluations with university students across three countries, they found that these problems allowed efficient independent practice and helped students overcome challenges in understanding complex code snippets. Such resources have high utility for novice programmers.

Turning attention to space technology, Kiryushin outlines conceptual solutions for a sodium-potassium-cesium cooled nuclear fission reactor designed for providing power on the Martian surface [4]. Through modeling and simulations, the author proposes innovative cermet fuels along with a two-circuit energy conversion system using carbon dioxide. This work lays important groundwork for enabling long-term colonization of Mars.

Shifting focus to social issues in technology, Almaiah et al. provide a comprehensive review of the numerous ethical concerns arising from increased use of social media platforms [5]. Ranging from threats to privacy and questionable content to mental health effects and cybercrimes, the authors highlight major areas society must address to develop solutions and policies that reduce adverse impacts. This will continue to be a crucial area needing further research and discussion.

In applied materials science, Garcia-Valenzuela and colleagues report on the deposition and characterization of amorphous carbon nitride thin films onto copper substrates using a DC pulsed sputtering technique [6]. Through structural and morphology analyses, they demonstrate the ability to synthesize CN films with island-like structures, elucidating the growth mechanisms. Such coatings have potential uses in mechanical and electronic devices.

Moving to thermal-electric energy systems, Eleyinmi's article provides extensive theoretical analysis alongside MATLAB simulations of thermoelectric generators (TEGs) and coolers (TECs) [7]. Determining optimal operating points to maximize efficiency, the author highlights the nonlinear relationships between currents, cooling/heating powers and other factors. The models developed offer useful tools for designing efficient thermoelectric systems.

Targeting improved reliability of electrical grids, Abdelaziz and Valles propose a novel algorithm for precisely locating faults in high-voltage transmission lines [8]. By leveraging electromagnetic transient dynamics, their method overcomes limitations of impedance-based fault localization

approaches. Simulations demonstrate 98.9% accuracy in detecting distance to faults, enabling rapid repairs and service restoration.

In another modeling contribution, da Silva Junior et al. present an approach to estimating nonhomogeneous thermal conductivity along a metallic bar using Fourier's heat transfer equation [9]. By propagating uncertainty from measurement errors, they obtain upper and lower bounds on the estimated conductivity values. This demonstrates an effective data-driven strategy for inferring material thermal properties.

Rounding out the issue, Boukhatem et al. examine integrating particle swarm optimization techniques to enhance the performance of sliding mode controllers for maximum power point tracking in photovoltaic systems [10]. Compared to classical sliding mode control, the particle swarm optimization method achieved higher precision and faster convergence in simulations. Hybridization of bio-inspired metaheuristics with conventional control methods shows strong promise.

Gritti et al. present an improved publicly verifiable proof scheme for enforcing data replication and retrievability in cloud storage systems [11]. By integrating verifiable delay functions, their method enables any third-party auditor to efficiently validate that files and replicas are correctly stored by cloud providers. This work enhances security and accountability for remote data storage services.

Overall, the excellent array of papers compiled in this issue exemplify high-quality research contributing novel techniques, frameworks and solutions across STEM fields. We hope these studies will inspire and inform future work by the research community to push forward the boundaries of knowledge and technology innovation. We thank the authors for entrusting their manuscripts to our journal and the reviewers for their diligent evaluations ensuring the highest standards.

References:

[1] T.U. Okafor, “Self-Regulated Learning and Academic Achievement of Physics Students in Selected Secondary Schools in Aguata Local Government Area of Anambra State, Nigeria,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 1–7, 2022, doi:10.25046/aj070101.

[2] M.J. Poblaciones, “Opinion and Effectiveness of Kahoot! use in Online Distance Learning in Crop Production at Higher Education Level: A Case of Study,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 8–13, 2022, doi:10.25046/aj070102.

[3] X. Lu, N. Funabiki, H.H.S. Kyaw, E.E. Htet, S.L. Aung, N.K. Dim, “Value Trace Problems for Code Reading Study in C Programming,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 14–26, 2022, doi:10.25046/aj070103.

[4] O.V. Sergeevich, “Physics behind the Concept of a Sodium-Potassium-Cesium-Cooled Martian Nuclear Reactor,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 27–46, 2022, doi:10.25046/aj070104.

[5] M.M. Qabajeh, “Ethical Implications and Challenges in using Social Media: A Comprehensive Study,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 47–52, 2022, doi:10.25046/aj070105.

[6] M.C.G. Domínguez, P.G.R. Romero, A.G. Díaz, H.M. Valencia, V.H.C. Sanchez, “Synthesis and Characterization of CN Thin Films Produced by DC-Pulsed Sputtering in an CH3CH2OH-N2 Atmosphere,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 53–59, 2022, doi:10.25046/aj070106.

[7] N.P. Bayendang, M.T. Kahn, V. Balyan, “Thermoelectric Generators (TEGs) and Thermoelectric Coolers (TECs) Modeling and Optimal Operation Points Investigation,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 60–78, 2022, doi:10.25046/aj070107.

[8] M. Ngwenyama, P. Le Roux, L. Ngoma, “A Novel Algorithm Design for Locating Fault Distances on HV Transmission Lines,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 79–89, 2022, doi:10.25046/aj070108.

[9] A. Núñez, F. Solares, A. Crisanto, “Estimation of Non-homogeneous Thermal Conductivity using Fourier Heat Equation Considering Uncertainty and Error Propagation,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 90–99, 2022, doi:10.25046/aj070109.

[10] I. Bouchriha, A. Ben Ghanem, K. Nouri, “Optimization of the Sliding Mode Control (SMC) with the Particle Swarm Optimization (PSO) Algorithm for Photovoltaic Systems Based on MPPT,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 100–106, 2022, doi:10.25046/aj070110.

[11] C. Gritti, H. Li, “Efficient Publicly Verifiable Proofs of Data Replication and Retrievability Applicable for Cloud Storage,” Advances in Science, Technology and Engineering Systems Journal, **7**(1), 107–124, 2022, doi:10.25046/aj070111.

Editor-in-chief

Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 7 Issue 1 **January-February 2022**

CONTENTS

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 01-07 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Self-Regulated Learning and Academic Achievement of Physics Students in Selected Secondary Schools in Aguata Local Government Area of Anambra State, Nigeria

Theresa Ugonwa Okafor*

Departmen ce Education, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, 432108, Nigeria

ARTICLE INFO

Article history:
Received: 07 October, 2021
Accepted: 05 December, 2021
Online: 07 January, 2022

Keywords:
Self-regulated learning
Students' Academic Success
Physics

ABSTRACT

The research investigated the consequence of self-regulated learning on students' academic performance in physics, using selected public co-education schools in Aguata Educational Zone, Anambra State as the research area. The study became necessary as a result of the observed persistent poor performance of the student in Physics in external examinations over the years. The study made use of quasi-experimental research design method. The research information was gathered from the control and experimental groups and was investigated by mean, standard deviation and ANCOVA. The tests were conducted at 0.05 significant levels. The results showed that students taught using self-regulated learning strategy had more impressive mean success scores than those taught using conventional method. The result revealed also that there was no substantial difference between the mean success scores of the male and female students. However, there was substantial difference in the mean scores of rural and urban students in favour of rural students. The interaction effects of strategy and gender on one hand and strategy and location on the other hand, were not statistically significant. The research concluded that self-regulated learning strategy has great impact on students' mean success scores in physics.

1. Introduction

Self-regulated learning is the capability to comprehend and exert influence on the learner's environment [1]. Self-regulation talents include objective setting, self-instruction, self-monitoring and self-reinforcement, according to the researchers. However, they quickly added that this should not be mistaking for intelligence or academic excellence; rather, self-regulation should be viewed as a process and collection of behaviors that are self-directed through which learners convert their mental facilities into skills. In their individual efforts, [2] and [3] noted that habit, which comes as a result of guided practice and feedback, is one of the features of self-regulation. In their self-regulated learning research in the classroom, [4] agreed wholeheartedly with the preceding position.

Self-regulated learning (SRL) is a major characteristic of self-regulation that is strongly associated with educational objectives [5]. It is, in their perspective, about metacognition (discerning about one's own reasoning), tactical approach (planning, monitoring, and evaluating personal growth against a specified benchmark), and drive to acquire knowledge. In more formal terms, self-regulated learning has three primary constituents: cognition, metacognition, and motivation [6], as [5] have already stated that self-regulated learning is contained within the domain of self-regulation. He continued by emphasizing that the cognition component covers the abilities and behaviors that are crucial when encoding, learning, and recalling information, as well as critical thinking. The metacognition component is made up of skills that allow students to comprehend and follow-up their own cognitive processes while learning. Motivation encompasses the thoughts and approaches that influence the usage and advancement in both cognitive and metacognitive skills [7]. It's no surprise that self-regulated learners tackle tough problems and put their learning activities into practice, gain a thorough comprehension of the problem at hand, and put in the work required to achieve academic achievement [5]. Self-regulated learners frequently exhibit feeling of ego to some extent as a result of these features. However, in the psychology education literature, scholars have attempted to link these qualities to achievements in and outside school

Physics, on the other hand, as one of the basic secondary school science disciplines, has proven to be a difficult hurdle for students to overcome in both internal and external examinations over the years. Nonetheless, the subject is one of the most important

*Corresponding Author: Theresa Ugonwa Okafor, teresaokafor@gmail.com

https://dx.doi.org/10.25046/aj070101

scientific subjects, as it is essential for the global growth of science and technological advances [8]. As a science or course of study that deals with natural phenomena, it aids man in understanding the nature and structure of the universe by studying the nature of matter and energy. It's no surprise that science education programs in Nigeria were created to aid students in learning problem-solving and decision-making abilities, as well as discover the interconnectedness of various science subjects, such as the relationship between physics and health, agriculture, industry, and technology, among others, in a variety of fields.

Given the importance of science education in any economy, the Nigerian government has made a concerted effort to improve science educational methods in schools. This has been accomplished through the government acquiring science equipment and hiring teachers with experience in related fields, as well as providing, to some extent, an enabling environment for students to learn without being distracted. However, students' performance in science topics, especially physics, has continued to be dismal in both internal and external examinations. Following this, concerned citizens have speculated that the method of teaching and learning physics and other science subjects may be to blame for the outcomes seen in the area. More specifically, it has been observed that most physics teachers' teaching methods do not allow students to effectively participate in class lessons, thereby limiting the rate of learning and retention in the subject. As a result, it has been suggested that introducing a learning strategy that encourages students to take charge of their learning process via a self-regulated approach may be more effective in increasing their comprehension and retention rate [8].

1.1. Statement of the Problem

Despite governments' concerted efforts directed at improving students' performance in physics results, students have continued to show dismal performance in physics external examinations such as WAEC, NECO and JAMB. This is evident from the last results of the subject at least from 2007 – 2018 where those who passed physics subject at credit level in those examinations were below 50 percent of those who enrolled in the mentioned years (WAEC, 2019). The situation has become really worrisome given the expected role and importance of physics as one of the basic sciences in the economy of Nigeria. Consequently, the research problem can be modeled as a question as follows: How can self-regulated learning mode be the answer to students' poor performance in secondary school physics?

1.2. Objectives of the Research

The main objective of the research is to survey the consequence of self-regulated learning on academic achievement of secondary school students in physics by using selected co-educational schools in Aguata Educational Zone of Anambra State, Nigeria as the research area. However, the specific objectives are to:

(i) investigate the influence of self-regulated and conventional learning on students' mean success scores in physics.
(ii) survey the consequence of gender on students' mean physics success scores when taught by self-regulated learning technique.
(iii) determine the effect of school location on students' average success scores in physics when taught by self-regulated learning method.
(iv) evaluate the interaction effects of method and gender on students' mean success scores in physics.
(v) ascertain the interaction effects of method and location on students' mean success scores in physics.

1.3. Research Questions

The following research questions were raised to guide the study:

(i) How do self-regulated and conventional learning methods affect the students' mean achievement scores in physics?
(ii) What is the impact of gender of students' on average success scores in physics when taught by self-regulated learning method?
(iii) What is the effect of school location on students' average success scores in physics when taught by self-regulated learning method?
(iv) What is the interaction consequence of self-regulated learning method and gender on students' average success scores in physics?
(v) What is the interaction consequence of self-regulated learning method and location on students' average success scores in physics?

1.4. Statement of Hypotheses

The following null hypotheses were articulated to direct the objectives of the research and fortify the analysis:

(i) There is no substantial difference in the average success scores of students taught physics by self-regulated and conventional learning methods.
(ii) There is no substantial difference between the average success scores of male and female students taught physics by self-regulated learning method.
(iii) There is no substantial difference in the average success scores of urban and rural student taught physics by self-regulated learning method.
(iv) There is no substantial interaction consequence of method and gender on students' average success scores in physics.
(v) There is no substantial interaction consequence of method and location on students' average success scores in physics.

1.5. Significance of the Research

The study is important from both a theoretical and empirical pespective. From a theoretical standpoint, the study will greatly contribute to the current body of literature in this field of research, thereby broadening the boundaries of knowledge. Empirically, both the teachers and the students will benefit greatly from the outcomes of the research. The teachers would be given insight on how to make use of self-regulated learning techniques for better imparting of knowledge in physics subjects. Similarly, the students would be enlightened on how to utilize self-regulated learning approach to accomplish great tasks in their academic endeavours

1.6. Scope of the Research

The research covers Aguata Educational Zone, Anambra State. Two co-educational secondary schools were purposefully selected namely, St Paul's Academy secondary school Ekwulobia and thus reflecting urban and rural mix respectively. In terms of variable scope, the study focuses on likely influence of self-regulated learning method on students' performance in physics subject.

2. Literature review

2.1. Conceptual Review

Regulatory education has been described as a background variable that refers to how much the teaching process encourages and outwardly helps students' self-regulated learning. It has been termed a meta-instructional variable that promotes student self-regulation and possesses hallmarks of efficient teaching [9]. But more formally, self-regulated learning has been defined as anything students engaged in order to organize and follow up their learning process, as well as to diagnose the source of any difficulties that arise while learning and design strategies to overcome it and move towards the learning objectives set initially as yardstick [2].

On the other hand, academic success as a determinate and consequence of the educational process in such a way that the degree, magnitude or level of learning aftermaths of students at various levels of education are direct effect of the teaching method [10]. It is further defined as the level to which a single student has gained information, abilities, and experiences in a certain set of academic goals [11]. However, within the context of this study, the concept is taken as the scores of individual students in physics examinations. It includes also, students' performance in tests and assignments [12].

2.2. Theoretical Framework

The study is based on Structuralist Theory, which was established by [13], and Bandura's Social Cognition Theory of Self-regulation [14]. Structuralists consider learning to be a self-regulated method in which students increase their ability to learn through the selective use of techniques. To that purpose, students monitor, control, or manage parts of their own cognition, motivation, and conduct, as well as some characteristics of their surroundings. Bandura's Social Cognition Concept of Self-Regulation posits that students have objectives and that during their learning activities, they observe, judge, and react to their views of goal processes. In his opinion, self-regulation is a self-directed method which aid the students to translate their mental aptitudes into academic capacities. The framework of the two theories above has demonstrated that if students effectively utilize motivation strategies in their studies, their learning outcomes will be enhanced substantially.

2.3. Theoretical Exposition

Self-regulated learning tactics assist students in readiness for enduring learning and acquiring crucial capabilities such as the ability to transfer skills, information, and talents from one area or context to another. This shows that it is a key method in the learning process at both the fundamental and advanced levels [15]. They observe that it is a cognitive learning approach strategy, and its application in learning has a substantial positive link with users' academic accomplishment. Students are more inspired to learn and are very likely to succeed when they apply a consolidation of academic learning skills and self-control that enables learning simpler [16]. In [10], the authors noted also that those students who are self-regulated in learning, metacognitive, motivated equally conduct themselves in a decent manner as well as offer initiative and direction for their own endeavors to learn information and skills without the assistance of teachers, parents, or others. Self-regulated learning is a recurring method that comprises of planning, execution, monitoring, and self-evaluation, which is used as feedback to achieve improved performance by students in their class works [2].

Self-regulated learners exhibit motivated beliefs as well as adaptive emotions such as academic self-efficacy [17]. The author noted further that such learners adopt learning goals and develop positive emotion towards tasks (example: joy, satisfaction and enthusiasm). Such learners, in his option, also have the capability to regulate and change them in accordance with the requirements of pre-set activities and various learning situations [18]. Self-regulated learners plan and control the time and effort spent on assignments, and they understand how to create and structure favorable situations, such as locating an appropriate study location and asking assistance from teachers and peers when challenges arise. This set of people make stronger efforts to regulate academic activities, classroom climate, and structure, which affect a number of volitional tactics aimed at avoiding external and internal distractions so that they can retain their concentration, effort, and motivation in doing tasks [17]. Therefore, self-regulated learners are those who are metacognitively, motivationally, and behaviorally involved in their own learning progressions and goal attainment [19].

2.4. Empirical Review

In [16], the authors conducted a research to determine the influence of self-regulated learning on students' mathematics success, it could be seen that self-regulated learning actually has substantial consequence on mathematics success because it makes the learning process very effective. The finding has indicated that students with strong self-regulated learning are likely to be highly motivated and successful in any topic, and vice versa. In another study carried out by [20], to verify the likely effects of self-regulated learning on the success of students in physics subject in secondary schools, using quasi-experimental and control groups, the data analysis utilizing the t-test revealed that students that practiced self-regulated learning outperformed those who did not engage in self-regulated learning.

In a research conducted by [21] to investigate the influence of self-regulated learning approach in the basic science successes of students in junior secondary school, quasi-experimental design was used. Findings from study show that students' performance in basic science improved significantly using a self-regulated learning method as against traditional approach. In a related study, [22] investigated the impact of self-regulated learning tactic on achievements of the students in basic sciences in Makurdi Council Area, Benue State, Nigeria. The research used the non-equivalent group pretest-posttest quasi-experimental method. Results of the research revealed that demonstration method improved the

achievements of students of basic science more than self-regulated approach. Secondly, no statistical substantial difference in average achievement in scores were recorded between students taught basic science by self-regulated learning technique and those taught by demonstration technique. Thirdly, average success scores of students of different gender taught basic science utilizing a self-regulated learning technique did not differ significantly. The study concludes that combination of self-regulated learning tactic and demonstration technique would produce a better student.

In another study, the researcher investigated the relationship between academic motivation and self-regulated learning in predicting academic accomplishment in school [23]. The data for the study were collected with the help of the academic motivation scale [24], and the academic self-regulated learning scale. The findings revealed that there was no correlation between average grade points and academic motivation and self-regulated learning. It means that, the combined scores of academic motivation and self-regulated learning did not determine the students' average grade points.

3. Methodology

3.1. Research Design

Quasi-experimental design was implemented in this research, which permitted the use of intact classes to circumvent the disruption of the usual class lessons. Furthermore, pre-test was also utilized to remove early variances in the two groups and to minimize selection prejudice which is a threat to core validity.

Illustrated below elucidate the design:

Group I	O_1	X_1	O_2	
Group II	O_2	X_2	O_2	

Where:

O_1 = Pre-test
X_1 = treatment for Experimental Group
X_2 = Control group
O_2 = Post test

3.2. Area of the Research

Non-private secondary schools in Aguata Council Area of Anambra State were used for this research. Specifically, St Paul's Academy Secondary School, Ekwulobia and Community Secondary School, Achina, both in Aguata L.G.A.

3.3. Research Population

The population of the research consists of the 1,356 Senior Secondary Two (SS II) Physics students in the two co-educational schools selected for the research. Since gender is one of the considerations, co-educational schools were studied. There is also issue of the location hence the rural-urban mix in the selection of schools in the study. We have also used the SS II intact classes because they are not in any external examination.

3.4. Instrument for Data Collection

Physics Achievement Test (PAT) is the tool utilized for data collection in this study. Items for PAT consists of twenty (20) multiple choice questions in tandem with the secondary school Physics curriculum and was culled from the West African Examination Council (WAEC). The PAT was used to assess the students' achievement in Physics.

3.5. Experimental Procedure

The researcher trained the teachers selected for control and experimental groups. While teachers for the experimental group were trained on guiding the students on self-regulated learning, the teachers for the control group were mentored on how to use the conventional method. Preceding the start of the experiment, the students were given a pre-test. PAT was given to the subjects for both the treatment and control groups at the end of the experiment by the teachers to answer and after that, the scripts were collated and scored and the results were used for the analysis.

3.6. Method of Data Collection

After the treatment session, posttest was conducted for the students and the results for both control and experimental groups were collated for investigations in order to answer the initial research questions and verify the claims of the null hypothesis.

3.7. Method of Data Analysis

Whereas mean ($\bar{x}$) and standard deviation (SD) were employed to answer the research questions, analysis of covariance (ANCOVA) was utilized to verify the hypotheses at 5 percent level of significance.

4. Data presentation and analysis

4.1. Answer to Research Questions

Table 1: Mean ($\bar{x}$) and Standard Deviation (SD) of Effect of Self-regulated Learning on Students Average Success Scores in Physics

Method	N	Pretest		Posttest		Gained Score
		$\bar{x}_1$	SD_1	$\bar{x}_2$	SD_1	$\bar{x}_g$
Self-regulated	71	10.58	4.30	23.85	5.76	13.27
Conventional	74	7.87	3.75	18.61	6.43	10.74

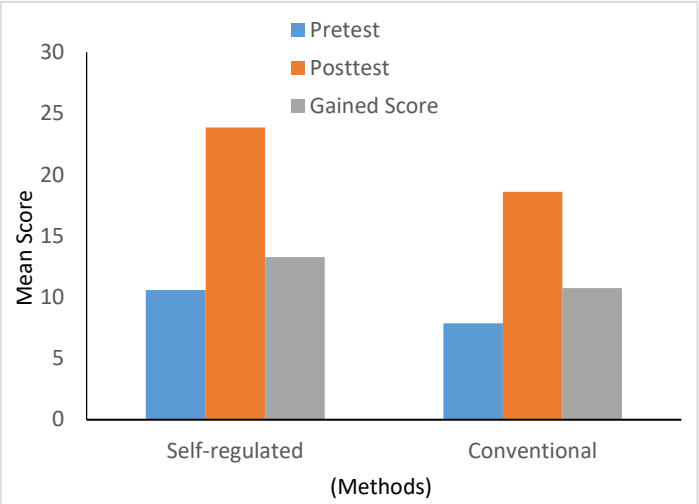

Figure 1: Mean Scores for self-regulated and conventional methods of teaching in both Pretest and Posttest

As could be seen from Table 1, the students taught using self-regulated tactic has an average score of 23.85 while those with the

conventional method is 18.61 with a gained score of 13.27 and 10.74 respectively thus showing that the students taught using SRL strategy out-performed those who were taught through conventional method.

Table 2: Mean ($\bar{x}$) and Standard Deviation (SD) of Effect of Gender on Students' Success in Physics based on SRL Strategies

Gender	N	Pretest		Posttest		Gained Score
		$\bar{x}_1$	SD_1	$\bar{x}_2$	SD_1	$\bar{x}_g$
Male	84	10.91	4.32	21.14	5.75	10.23
Female	61	9.20	3.85	20.01	6.22	10.81

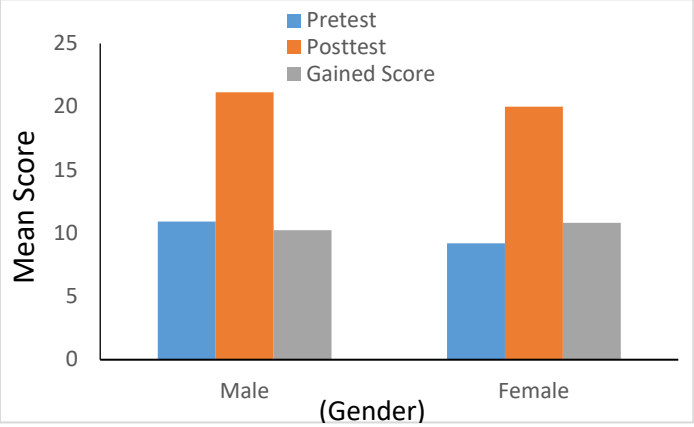

Figure 2: Mean Score for Male and Female Students in both Pretest and Posttest

The mean and standard deviation analysis presented in Table 2 reveals that the male students' average success score is 21.14 whereas that of their female counterparts is 20.01 thus showing that male students marginally did better than the female students. However, in terms of gained scores, female achieved mean was equally marginally higher than the males.

Table 3: Mean ($\bar{x}$) and Standard Deviation (SD) of Effect of Location on Students' Average Success Score in Physics when Teaching was carried out through SRL Method

Location	N	Pretest		Posttest		Gained Score
		$\bar{x}_1$	SD_1	$\bar{x}_2$	SD_1	$\bar{x}_g$
Urban	73	11.01	4.43	19.75	6.03	8.74
Rural	72	8.89	3.81	20.52	6.71	11.63

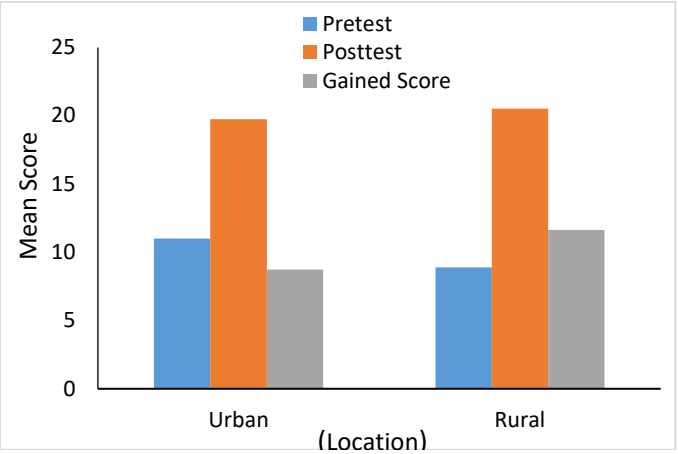

Figure 3: Mean Score for the Students in Urban and Rural Schools for both Pretest and Posttest

The results of mean and standard deviation presented in Table 3. The analysis showed that 19.75 is the average success score for urban students and 20.52 is the average success score for the rural-based students. Similarly, whereas urban students gained mean score of 8.74, the rural-based students gained mean score of 11.63 thus showing that rural-based learners did better than the urban-based learners in the test.

Table 4: Mean ($\bar{x}$) and Standard Deviation (SD) of Success Scores by Strategy and Gender

Gender/Strategy	Self-regulated Learning			Conventional		
Pretest	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Male	57	10.85	4.21	27	7.89	4.00
Female	12	8.93	3.87	49	8.05	3.90
Posttest						
Male	57	22.46	4.28	27	17.07	7.54
Female	12	23.57	6.26	49	19.45	7.38
Observed ($\bar{x}$) and (SD)		23.85	5.76		18.61	6.43

From Table 4, the average success score for the male students who were taught using self-regulated method is 22.46 while that of the female is 23.57. Similarly for the male students taught using conventional method, it is 17.07 and 19.45 for the female. However, the findings do not indicate ordinal interaction effect between tactic of teaching and gender of students. As could be seen, for different gender, the mean scores for the SRL method were higher.

Table 5: Mean ($\bar{x}$) and Standard Deviation (SD) of Success Scores by Strategy and Location

Location/Strategy	Self-regulated Learning			Conventional		
Pretest	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Urban	50	10.97	4.23	23	8.54	4.14
Rural	21	8.41	3.61	51	7.65	3.85
Posttest						
Urban	50	22.49	4.35	23	16.45	7.5
Rural	21	22.45	5.41	51	19.65	7.27
Observed ($\bar{x}$) and (SD)		23.85	5.76		18.61	6.43

The analysis in Table 5 revealed that average score for urban students taught using SRL strategy, is 22.49 while their rural counterparts is 22.45. Similarly, the average score of the urban students taught with conventional method is 16.45 while that of the rural students is 19.65. As could be seen, the results indicate ordinal interaction effect between strategy of teaching/learning and location of students.

4.2. Test of Hypotheses

All the hypotheses formulated to direct and strengthen the analysis were tested in this section through the application of Covariance technique known as analysis of covariance (ANCOVA). The result is presented in Table 6.

Table 6: Analysis of Covariance of Students Mean Achievement Scores in Physics Test

Source of Variation	Sum of Squares	df	Mean	F	Sig.

Covariate Model	1650.842	8	206.355	6.615	.000
Intercept	4387.341	1	4387.341	140.642	.000
Pretest	457.145	1	457.145	14.654	.000
Method	185.704	1	185.704	5.953	.017
Gender	68.302	1	68.302	2.189	.145
Location	215.87	1	215.87	6.920	.012
Method by Gender	6.585	1	6.585	0.211	.654
Method by Location	59.125	1	59.125	1.895	.177
Gender by Location	18.897	1	18.897	0.606	.460
Method by Gender by Location	122.524	1	122.524	3.928	.056
Error	4242.571	136	31.195	-	-
Total	62143.01	145			
Corrected Total	5788.561	144			

* *df (degree of freedom), F (F-ratio) Sig. (significance level).*

4.3. Interpretation of ANCOVA Results

In this section, the ANCOVA results were interpreted and all tests were carried out at 0.05 level of significance and f(1,144) degrees of freedom. That being the likely level at which we are willing to accept type I error.

Decision Rule I

At 0.05 level of significance and f(1, 144) = 5.953, P < .017, the null hypothesis was rejected thus showing there is substantial variance in the average success scores of the students taught physics using self-regulated learning tactic and those that used conventional method. The implication is that SRL technique is better than the conventional method of learning physics.

Decision Rule II

At 0.05 level of significance and f(1, 44) = 2.189, P > .145, the null hypothesis was accepted, and we came to the conclusion that there is no substantial difference between male and female students' average success scores in Physics.

Decision Rule III

At 0.05 level of significance and f(1, 144) = 6.920, P < .012, the null hypothesis was rejected and the alternative accepted and we conclude that there is a significant difference between the average success scores of urban and rural based students in physics.

Decision Rule IV

At 0.05, level of significance and f(1, 144) = .211, P > .654, the null hypothesis was accepted thus indicating that the interaction effect of method and gender on students average success scores in physics was not statistically significant.

Decision Rule V

At 0.05 level of significance and f(1, 144) = 1.895, P > .177, the null hypothesis was accepted thus suggesting that the interaction effect of strategy and location on students' average success scores in physics was not statistically significant.

5. Conclusion

From the answers obtained from the research questions analysed, gained score for the students who used self-regulated learning method is 13.27 while that of the conventional method is 10.74 thus showing that SRL strategy is superior to conventional method. Similarly from the mean scores of the students as presented in Table 2, it could be seen that male students performed marginally better than their female counterparts. In the same vein, Table 3 showed that rural-based students with average success score of 20.52 did better than the urban-based students with mean achievement score of 19.75. In terms of strategy and gender, at all levels of gender, the mean achievement scores for SRL strategy were higher than the conventional strategy. However, the result did not indicate ordinal interaction. Lastly, the SRL strategy has continued to maintain higher mean achievement scores even amongst the rural based students with a visible ordinal interaction effect strategy and location. The rest of the summary is specifically presented below:

1. The average success scores of students taught using the SRL technique and those taught using the traditional way differ significantly, with the SRL method having the upper hand.
2. In the physics test, there is no major difference in average success scores between male and female students.
3. In the physics test, there is a large disparity between the mean achievement scores of urban and rural based students.
4. There is no statistically significant interaction between strategy and gender on students' average success scores in the physics test.
5. There is no statistically significant interaction between strategy and location on students' average success scores on the physics test.

Furthermore, the fact that there was a marginal increase in the performance of male students over their female counterpart is an indication that both sexes have the opportunity of utilizing SRL strategy to increase performance. Therefore, it has become increasingly important that students should know how to manage their own learning as both the need and the opportunities for individuals to learn on their own outside of formal classroom settings becomes more obvious.

A major finding in this study is that self-regulated learning method is better than the conventional method in the promotion of enhanced academic success for the students. It is a finding that is coherent with previous studies such as [8,20,21, 22] that confirmed the significant importance of self-regulated learning to students' development and consequently, improved performance in academic works. This conclusion arose from the respective performance of the students after they were subjected to both pretest and posttest experiments in physics subject, where the self-regulated learning students outperformed the students taught with conventional method.

6. Recommendations

The recommendations presented herewith were based on the findings and conclusion of the study:

1. Given the significance difference in the average success scores between students taught using the SRL technique and those taught using the traditional approach in favor of the

SRL method, we recommend that secondary schools in Nigeria should start using the SRL technique because it has shown to be effective in enhancing students' performance.
2. Teachers in secondary schools should be trained to understand what SRL method is all about so that they provide proper guidance to students on SRL strategy.
3. Orientation exercise should be organized by the stakeholders in education for the students on the strategies of self-regulated learning for better academic achievement in all subjects.

7. Practical Implications of the Study

One of the challenges of the study was the limited knowledge of the teachers in self-regulated learning method. In the course of the experiment, it was found that not many teachers in the used schools used have good knowledge of self-regulated learning and it affected in no less measures, the performance of the students in the tests. Other areas of concern were inadequate number of facilities and favourable environment suitable for study. Therefore, there is the need to train the teachers on how to effectively guide the students in self-regulated learning as it has proved to be more successful in improving students' academic performance. Another policy dialogue is in the aspects of provision of necessary facilities that can aid the students to create the enabling and suitable environment for efficient utilization of self-regulated mode of learning which has proved to superior to the conventional mode which happens to be teacher-centered.

References

[1] G. Schraw, Crippen, K. & Hartley, K., "Promoting self-regulation in science education: metacognition as part of a broader perspective on learning," Research in Science Education, 36, 111-139, 2006, doi.org/10.1007/s11165-005-3917-8.

[2] B. Zimmerman, S. & Kovach, R., "Developing self-regulated learners: beyond achievement to self-efficacy," Washington D.C. American Psychological Association, 2002, https://doi.org/10.1037/10213-000.

[3] D. Butlar, "Individualizing instruction in self-regulated learning," Theory into Practice, **41,** 81-92, 2002, http://www.jstor.org/stable/1477459.

[4] S. Paris, Paris, A., "Classroom applications of research on self-regulated learning," Educational Psychology, 36, 89-101, 2001, https://doi.org/10.1207/S15326985EP3602_4

[5] J. Burman, T. Green, C.D. and Shanker, S., "On the meanings of self-regulation: digital humanities in service of conceptual clarity," Child Development, **86**(5), 1507-1521, 2015, DOI: 10.1111/cdev.12395

[6] S. Tanya, "Self-regulated learning," Teaching excellence in adult literacy (TEAL). (enter Fact Sheet No.3), 2017.

[7] E. Alvi, Iqbal, Z., Masood, F. and Batool, T., "A qualitative account of the nature and use of self-regulated learning (SRL) Strategies employed by university Students," Australian Journal of Teachers Education, 41(8), 40-59, 2016, DOI 10.14221/ajte.2016v41n8.3

[8] J. Nnaji, Okoli, J.N. "Relationship between self-regulated learning skill and achievement of secondary school students in Physics in Enugu State, Nigeria," JSTME, **3**(1), 115-127, 2019,

[9] J. Antonelli, Jolly, S., Backscheider, S. and Hawk in set J., "Understanding the self-regulated learning characteristics of first-generation college students," Journal of College Stud. Dev. **61**, 67-83, 2020, dio: 10.1353(sd.2020.0004).

[10] T. York, Gibson, C. and Rankin, S., "Defining and measuring academic success," Practical Assessment, Research and Evaluation, **20**(5), 2-12, 2015.

[11] B. Illahi, Y. and Khandai, H., "Academic achievements and study habits of college students of District Pulwama," Journal of Education and Practice, **6**(31), 1-16, 2015.

[12] E. Siahi, A. and Maiyo, J.K., "Study of the relationship between study habits and academic achievement of students: a study of Spicer Higher Secondary School, India," International Journal of Educational Administration and Policy Studies, **7**(7), 134-141, 2015.

[13] L. Steffe, Gale, J. E. (Eds.), *Constructivism in education.* Lawrence Erlbaum Associates, Inc., 1995.

[14] A. Bandura, "Social cognitive theory of self-regulation," Organizational Behavior and Human Decision Processes, 50, 248-287, 1991, https://doi.org/10.1016/0749-5978(91)90022-L

[15] G. Dettori, Donattela, M. and Persico, B., Fostering self-regulated learning through ICT, New York: Hershey, 2011.

[16] A. Fauzi, Widjajanti, D.B., "Self-regulated learning: the effect on students' mathematics achievement," Journal of Physc. Conference Series, 1097, 2018.

[17] I. Banu, "The relationship between self-regulated learning strategies and academic achievement in a Turkish EFL setting," Educational Research Review, **8**(17), 1544-1550, 2013.

[18] T. Loong, "International students' self-regulated learning and its relation to mathematics achievement in an offshore Australian Program," Academic Research International, 4(5), 507-520, 2013.

[19] B. Zimmerman, Schunk, D., Self-regulated learning and academic achievement. Theoretical perspective. Mahwah, NJ: Lawrence Eribaum Associates, 2008.

[20] N. Achufusi-Aka, Offiah, F.C., "The effect of self-regulated learning on academic achievement of secondary school physics students," African Journal of Educational Studies in Mathematics and Sciences **8**(2), 29-33, 2010.

[21] C.E. Nwafor, Obodo, A.C. and Okafor, G., "Effect of self-regulated learning approach on junior secondary school students' achievement in basic science," Journal of Education and Practice, **6**(5), 45-52, 2015.

[22] C.M. Tirgba, Bur, J.I., "Effect of self-regulated learning strategy on students' achievement in basic science in Makurdi Local Government Area, Benue State, Nigeria," African Journal of Teachers Education (AJOTE), 8. 361-379, 2019.

[23] B. Cetin, "Academic motivation and self-regulated learning in predicting academic achievement in College," Journal of International Education Research, Second Quarter, **11**(2), 95 -106, 2015.

[24] R. Vallerand, Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., and Vallieres, E. F., "The Academic Motivation Scale: a Measure of Intrinsic, Extrinsic, and Amotivation in Education," Educ. Psychol. Meas. **52**(4), 1003–1017, 1992, DOI10.1177/0013164492052004025.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 08-13 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Opinion and Effectiveness of Kahoot! use in Online Distance Learning in Crop Production at Higher Education Level: A Case of Study

Maria J. Poblaciones*

Department of Agronomy and Forest Environment Engineering, University of Extremadura. Avenida Adolfo Suárez s/n, Badajoz, 06007, Spain

ARTICLE INFO

Article history:
Received: 28 October, 2021
Accepted: 25 December, 2021
Online: 14 January, 2022

Keywords:
Game-Base learning
Higher education
Online learning

ABSTRACT

Kahoot is one of the most popular learning platforms, with over 2.5 billion players per year worldwide, used mainly for knowledge review through gamification. However, students´ opinion and its effectiveness in online distance learning as well as in another language teaching has been little studied. With these aims, this study was carried out in the second semester of 2020/21 academic year, with 73 students of the double degree between "Food Science and Technology" at University of Extremadura in Spain (UEx) and "Food Quality and Safety" at Chongquin University in China (UEC). Due to COVID-19 disease, lessons were completely online through Zoom platform. To review the theoretical content a Kahoot quiz with 15 questions was given at the end of each lesson. In the middle and at the end of the course, a partial and a final exam was held repeating two aleatory questions from each Kahoots gamed after each lesson, to determine the knowledge memorisation. For the recognition of the different plant species studied, four Kahoots were carried out with 11 questions, and a final exam repeating eight photos was carried out. Finally, students filled in the feedback available on the website itself to determine their opinion of Kahoot. The results showed the effectiveness of using the Kahoots as a tool to reinforce knowledge, not only in theoretical content, with increases of 26.0 and 31.4% in the remembering, but also in the recognition of plant species, with an increase of 17.3%. Student feedback was very positive, with more than 95% finding the tool not only fun, but that it had helped them learn and would recommend it. However, about 12% of the students had an indifferent or negative perception of its use, and this should be explored further in future research.

1. Introduction

Gamification is defined by [1] as the application of game elements in a non-game context. Its main objective is to involve students in their learning process and not only internalisation of what is exposed [2]. There are several educational software programmes based on gamification, which aim to increase enthusiasm and attention in the classroom to achieve learning outcomes. The key aim is to develop skills or to feel pleasure, all of which are intellectual characteristics of motivation in humans [3], being generally considered as an effective way to create constructivist learning environment [4-5]. However, certain aspects need to be clarified as the effectiveness in its use under online distance learning.

Since March of 2020, as a way to control the spread of the COVID-19 virus, citizens all over the world were ordered to stay at home by their governments. Therefore, all educational institutions, including universities, had to adapt their teaching methodologies to online formats in record time. So that it was necessary to implement new modes of teaching and learning compatibles with virtual classes. In this paper, an experience of using Kahoots during online teaching period due to COVID-19 pandemic is described.

The experience took place with second year students of the double degree between "Food Science and Technology" at Extremadura University (UEx) and "Food Quality and Safety" at Chongquin University (UEC) in the subject named “Bases and Technology of Plant Production”. Since the 2020/21 academic year, and due to the COVID-19 pandemy, all the formation was online and the real-time quiz Kahoot was used to consolidate knowledge. Under these special conditions, the first part of this

*Corresponding Author: Maria J. Poblaciones, majops@unex.es

paper analyse the degree of Kahoot acceptance and motivation using the existing survey on the website. Secondly, in order to study whether the use of Kahoots helps in the consolidation of knowledge, not only in theoretical questions by means of multiple-choice tests, but also in the recognition of different plant species by means of test with photos, certain questions belonging to the Kahoots used in class, were repeated in a partial exam and in the final one.

2. Literature Review

Recently, the use of digital devices with internet connection has become popular in many aspects of the life, such as social activities, business, and in the educational system [6-8]. In fact, there are several Educational Software's used in a classroom environment for teaching and learning using gamification.

In line with [9] gamification is "using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems". Games have some distinctive features which play a key role in gamification, between them [10]:

- users are all participants
- challenges/tasks that users perform and progress towards defined objectives
- points that are accumulated because of executing tasks
- levels which users pass depending on the points
- badges which serve as rewards for completing actions
- ranking of users according to their achievements

The objectives to be achieved by the implementation of games in learning can be summarised as being engaging, fun and exciting for players, thus stimulating motivation for learning, offering the opportunity for participants to explore their curiosity through new challenges [11]. Therefore, it combats one of the main frustrations of teachers, at all levels of education, namely demotivation. A problem that, as many studies have pointed out, cannot be solved by conventional methods [12-14] and gamification has been suggested as an effective tool to improve students´ motivation and engagement [15]. Current game-based learning should focus on the application of extrinsic game-like elements, such as levels, challenges, point systems and rankings, to create a user experience that engages students to complete learning tasks [16].

In recent years, tools for gamifying learning in the classroom have proliferated allowing educators new tools in giving lessons and many of them take the form of applications for mobile devices [17-19]. Kahoot is one example of a playful, game-based student response system which has received wide acceptance with more than 70 million active users across the world monthly [20] reaching that, in 2019, over 2.5 billion people from more than 200 countries have played Kahoot [21]. It is based on three basic values: play, learn and socialise, being simplicity of creating the game experience and its easy access through mobile devices are two key points.

Kahoot is a very versatile tool that allows multiple options: true-false, simple or multiple answers, allows to include images, videos, or to configure different variables, such as the available time to answer the questions [21-22]. The questions are projected in the classroom and the students, through their smartphones, answers them, adjusting to the marked time. Provides immediate feedback on the degree to which students have acquired knowledge and identifies content that has not been understood. After each question, the respective winner and points accumulated is shown to provide a final ranking, as if it were a competition. Finally, after each question, the teacher can explain each answer, which is where the pedagogical potential lies [23]. The game ranks players based on speed and accuracy, and at the end of the quiz, the names of the top five players are displayed on the leaderboard. In this way, students can compete to see who gets the most questions right in the shortest time, being this its main handicap, according to [17, 24]. The fact that higher scores are obtained for quicker answers can create anxiety and stress for many students, who risk too much to get more points, to the detriment of spending enough time thinking about the correct answer. Results can be downloaded by educators allow them to highlight problematic questions and identify students who may be struggling [21-23].

Gamification in general and Kahoot in particular, has been widely used in teaching, generally at pre-university levels and in face-to-face teaching. However, analysing the most recent reviews on the use of gamification in teaching [25-26] highlight that, of the 93 and 24 studies analysed, both at pre- and university levels, the positive properties can be summarised as an increase in motivation, in classroom dynamics and amenity, students´ and teachers´ attitudes, positive effect on learning performance and decreasing of student anxiety [26-28]. On the contrary, the main problems encountered included bad internet connection, the impossibility to change the answer after submission and the time pressure to give answers, which reduces students' reflection and some students' guessing without thinking because of the fear of losing [17, 21, 23-24]. With regards to the advantages and limitations related to game-based learning in a foreign language, the same advantages and disadvantages were found in a recent review [29].

Finally, the response of most universities worldwide to the crisis caused by the Coronavirus disease (COVID-19) was the transition to online environments. In these cases, the use of motivational and energising tools in the classroom becomes even more necessary. However, their effect on online teaching needs to be studied and clarified. To clear up these issues, with particular emphasis on the students' opinion and their effectiveness on the consolidation of knowledge, this study has been carried out.

3. Methodology

3.1. Double degree between Extremadura and Chongquin Universities

Since June 2019, the University of Extremadura in Spain (UEx) and Chongquin University in China (UEC) established the double degree between "Food Science and Technology" at UEx and "Food Quality and Safety" at UEC.

The initial planning of the double degree had a duration of four years, in which students will receive the following teaching:

- First year. UEC students will receive an intensive Spanish course.

- Second and third year. UEC students will be taught eight subjects, two per semester and year, taught in Spanish by UEX professors.
- Fourth year. UEC students will travel to UEx to be taught two subjects and to develop their Final Projects.

However, due to the outbreak of Covid-19, the governments of many countries, including China and Spain, forced their citizens to stay at home. This led to major changes not only in the lives of citizens, but also in the education system. The main consequence was the impossibility of teaching the classes in person, as the UEx teachers could not travel to the UEC. However, the UEC maintained face-to-face attendance during 2020/2021 course. Therefore, the classes were online, using the Zoom platform, and while the UEx teachers were in Spain, the UEC students and teachers were together in a classroom in China. In order to reinforce the comprehension of the lessons, and so that the language was not an impediment to the development of them, the Chinese teacher of the subject, as well as a translator Chinese-Spanish participated actively in all the sessions. Moreover, all presentations were provided to the students translated with the enough advance and the subtitles available in Power Point 365 were used.

3.2. Procedure

In the specific case of this study, the subject taught was "Bases and Technology of Plant Production", taught in the second semester of the second year. This subject is divided into 14 lessons taught in 34 sessions of 1.5 hours sessions. The number of students enrolled was 73, divided into two groups that received exactly the same contents. The evaluation of the contents´ subject was divided into two parts:

i) Evaluation of the theoretical contents by means of a multiple-choice test

ii) Evaluation of the differentiation of the main crops studied (extensive, horticultural and fruit crops) by means of their recognition by means of different photographs

In order to review the theoretical content and assess the students' level of understanding, a Kahoot quiz with 15 questions was given at the end of each lesson. To ensure that the time available to answer each question was sufficient, and taking into account the language handicap, it was extended to 1 minute. In the middle of the course, after seven topics and, therefore, seven Kahoots, a partial exam was held with 50 multiple-choice questions with four possible answers, but only one correct. To determine the knowledge memorisation or if significant learning was achieved, two aleatory questions were chosen to be repeated from each of the Kahoots gamed after each lesson, i.e. a total of 14 questions. Also, at the end of the course, there was a final exam consisting of two parts, a 50-question test, in which the same questions of the partial were repeated and other two questions aleatory chosen from the last seven lessons´ Kahoots made by the students throughout the course, thus a total of 28 questions, 14 from the partial and 14 of the last seven Kahoots (Figure 1).

The time elapsed between the midterm and the final exam, i.e. 17 weeks, will also allow us to know whether the knowledge acquired is consolidated over time.

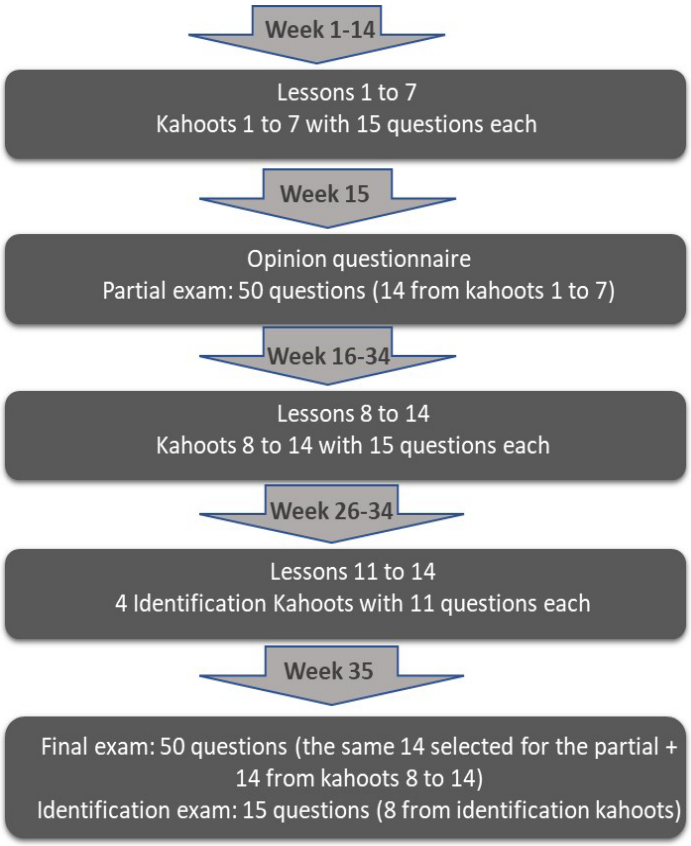

Figure 1. Phases in the study.

For the evaluation of the recognition of the different plant species studied, along the last month of the course, four Kahoots were carried out with 11 questions asking which species could be seen in the photo (Figure 2). In this case, the final exam was a test with 15 questions, in which eight photos were repeated, two from each of the Kahoots carried out in class (Figure 1).

Figure 2. Example of identification question on the Kahoot website.

3.3. Opinion on the use of Kahoot as a learning tool

After the seventh questionnaire, students were asked to fill in the feedback available in the website itself to determine the students' perception of the use of the tool based on the student response system with different scales (Table 1).

Table 1. Opinion questions and their scales.

Question	Scale
1. How fun was it?	1 (no fun) to 5 (very fun)
2. Did the players learn anything from this Kahoot?	No-Yes
3. Could players recommend this Kahoot?	No-Yes
4. How did the players feel about this Kahoot?	Negative – Neutral - Positive

3.4. Data analysis

The obtained data after the evaluation of the questionnaire items was quantitative and based on a descriptive analysis in SPSS 23.

4. Result and Discussion

4.1. Degree of theory knowledge consolidation.

Figure 3 shows how Kahoot improved students´ memorisation of knowledge in the theoretical questions, as they obtained better results when the questions were repeated, not only in the midterm, with much less time elapsed, but also in the final exam. Specifically, the average score obtained in the 14 chosen questions from Kahoots 1 to 7 increased in about 31.4% in the midterm, and 29.6% in the final exam. This is a very good result because the time elapsed from the first to seven Kahoots to the repetition of those questions was between 32 to 17 weeks. Regarding the repeated questions from Kahoots 8 to 14, the increases were about 26.0%.

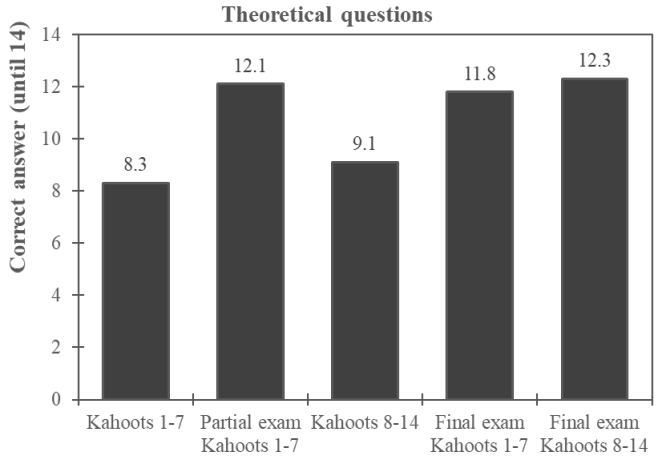

Figure 3. Average of the obtained scores in the 14 chosen questions corresponding to Kahoots 1 to 7 and 8 to 14, in their repetition in the midterm and in the final exam.

The creator of Kahoot stated in his study at the Norwegian University of Science and Technology (NTNU) that the use of Kahoot increased learning by 22% over students who used other game dynamics [30]. A previous study [24]], also found lightly lower increases in a similar study, but in presential learning, and [31-32] found a good degree on knowledge consolidation, in contrast to [33-34] who did not find any significant effect on the knowledge consolidation, considering that repetition and frequency of use are directly related to its efficiency.

4.2. Degree of identification knowledge consolidation.

The first rounds of each Kahoot had an average of 6.2 correct answers from 8 possible, a high percentage, indicating that the contents are easily interpreted and memorised. The increase between the questions in the first Kahoot and their repetition in the final exam was of 17.3%, quite lower in comparison with the theory questions. But, considering the rate achieved is out of 8 points, the results are good, being even better in their repetition. In this case, two facts are worth noting: firstly, the time that has passed is much shorter, from 2 to 8 weeks; and secondly, the fact that it is a question of recognising photos, with the photographic memory being much more actively involved.

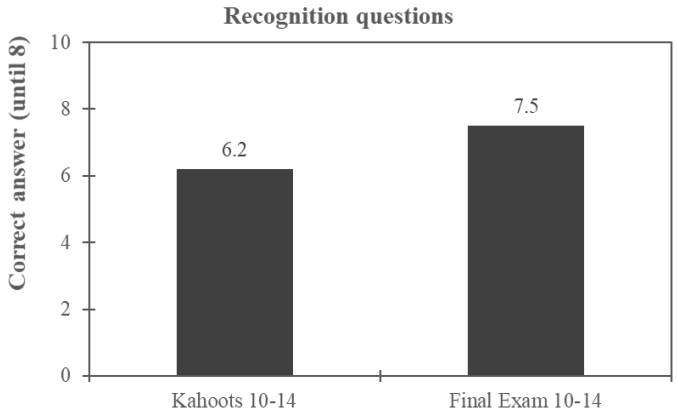

Figure 4. Average of the scores obtained in the 8 chosen questions corresponding to Kahoots 10 to 14 and in their repetition in the final exam.

So, our finding would be according with that found in the review of [25], which included 36 studies investigating the learning outcome of only using Kahoot. Of them, 70% of the studies showed that Kahoot significantly improves the final grade or test results compared to other teaching approaches.

4.3. Opinion on the use of Kahoot as a learning tool

Table 2 confirms that the students´ acceptation was really good, with percentages of the positive answers for all the items higher than 90%.

For the question 1 relative to the fun of playing the game, only 3 of the surveyed students showed a neutral or negative opinion regarding the fun of the game. As is showed in question 2, more than 97% of students feel they have learned something and more than 95% would recommend using Kahoot as a review tool (question 3). Only when they were asked about how it makes them feel (question 4), the percentage of students with either an indifferent (9.59%) or negative (2.74%) opinion was substantially higher, although a large majority (87.67%) had a positive perception (Table 2).

Table 2: Questionnaire data analysis (N=73).

	1 (%) No Negative	2 (%)	3 (%) Neutral	4 (%)	5 (%) Yes Positive	Average (1 to 5)	Median
1. Fun	0	1.37	2.74	23.29	72.60	4.74	5
2. Learn	2.74	-	-	-	97.26	-	Yes
3. Recom.	4.11	-	-	-	95.89	-	Yes
4. Feel	2.74	-	9.59	-	89.67	-	Positive

The results found in this study are notable for the fact that the teaching was online and not face-to-face. Numerous studies [27-28, 31-32, 35] agree in conclude that Kahoot is an effective instrument for carrying out presential activities in the classroom in a fun way, breaking the monotony and improving student participation, fostering a positive relationship between the group of students. In fact, authors such as [33] indicate a higher rate of class attendance. Most students perceive Kahoot as a game and not as an assessment system [36]. All these high levels of acceptance such as those found by [24], with very similar values in classroom teaching of this subject, with more than 95% of students with very positive evaluations of the use of Kahoot, or [31] who obtained lower values, with 87% acceptance. Rejection is often related to the inability to change the answer after submission, stressful time pressure to give answers, lack of time to answer, fear of losing, and difficulty in catching up if an incorrect answer has been given. [25, 33].

Moreover, the fact that the teaching is in another language does not seem to influence the students' opinion. However, the reinforcement of the Chinese-Spanish translator probably reduced the complication of understanding another language. Finally, in terms of perception in language classes, [37] found a greater willingness to actively participate in Arabic classes, with [38] demonstrating that students learned vocabulary more easily using Kahoot.

5. Conclusion

Although some of the limitations of this study are that it was developed in a single semester, that the number of students is quite limited and that, all students belong to the same group instead of dividing them into two groups, one control and one using Kahoot, more than 90% of students found it fun, that helps to learn and it's recommended as an engagement and students´ motivation tool. However, a small proportion felt indifferent or unpleasant when using this tool. This is where future research should investigate further to see what factors lead to lower acceptance, with particular emphasis on the stress caused by limited response time, the impossibility of changing the response, and how to remedy them.

The results demonstrate its effectiveness in consolidating knowledge. The number of answers that students remembered between questions asked in class using Kahoot and then repeated between 17 and 34 weeks was between 26 and 31%, regardless of the time elapsed between the two. Despite of the obtained results are consistent with previous research; further research must be done in order to corroborate the effectiveness of these questionnaires. This would require more experiments with more students and of longer duration, and more attention to the anxiety caused by the short time to answer.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

Author gratefully thank the students of "Bases and Technology of Plant Production" at Chongquin University, for their participation.

References

[1] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, D. Dixon, Gamification: Using game design elements in non-gaming contexts, 2011, doi:10.1145/1979742.1979575.

[2] R. Al-azawi, F. Al-faliti, M. Al-blushi, "Educational gamification vs. game-based learning: Comparative study," in International Journal of Innovation, Management and Technology, **7**(4), 132–136, 2016, doi: 10.18178/ijimt.2016.7.4.659.

[3] R.E. White, "The power of play: a research summary on play and learning," Smart Play, 15–25, 2013.

[4] M.J. Poblaciones, T. Garcia-White, C. Marin, "Students´ Perception of Real-Time Quiz Kahoot! As a Review Tool in Higher Education," International Journal of Engineering Pedagogy, **11**(4), 165-175, 2021. doi:10.3991/ijep.v11i4.21359

[5] S. Papadakis, "The use of computer games in classroom environment," International Journal of Teaching, **9**(1), 1-25, 2018. doi:10.1504/IJTCS.2018.10011113.

[6] B. Bature, "The role of information and communication technology as a tool for effective teaching and learning of mathematics," Journal of Applied of Computational Mathematics, **5**(6), 1–3, 2016, doi:10.4172/2168-9679.1000333

[7] A. Colbert, N. Yee, G. George, "The digital workforce and the workplace of the future," Academy of Management Journal, **59**(3), 731–739, 2016. doi: 10.5465/amj.2016.4003.

[8] K. Bedi, "Tablet PC and smartphone uses in education (TabletTours)," In 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 940-945, 2014, doi: 10.1109/MIPRO.2014.6859703.

[9] K.M. Kapp, "The gamification of learning and instruction: Game-based methods and strategies for training and education," Pfeiffer, San Francisco, CA, 2012.

[10] G. Kiryakova, N. Angelova, L. Yordanova, "Gamification in Education," 9th International Balkan Education and Science Conference, 2014.

[11] F. Rozi, Y. Rosmansyah, B. Dabarsyah, "A Systematic Literature Review on Adaptive Gamification: Components, Methods, and Frameworks," in 2019 International Conference on Electrical Engineering and Informatics (ICEEI), 187–190, 2019, doi:10.1109/ICEEI47359.2019.8988857.

[12] G. Barata, S. Gama, J. Jorge, D. Gonçalves, "Gamification for smarter learning: Tales from the trenches," Smart Learning Environments, **2**(1), 1–23, 2015, doi:10.1186/s40561-015-0017-8

[13] K.R. Bell, "Online 3.0-The rise of the gamer educator the potential role of gamification in online education," Dissertation available from ProQuest. AAI3635727. https://repository.upenn.edu/dissertations/AAI3635727, 2014.

[14] P. Buckley, E. Doyle, "Gamification and student motivation," Interactive Learning Environments, **24**(6), 1162–1175, 2016, doi: 10.1080/10494820.2014.964263.

[15] D.O. Göksün, G. Gürsoy, "Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz," Computers &Education, **135**, 15-19, 2019, doi:10.1016/j.compedu.2019.02.015

[16] F.L. Khaleel, N.S. Ashaari, T.S.M.T. Wook, A. Ismail, "Gamification Elements for Learning Applications," International Journal on Advanced Science, Engineering and Information Technology, **6**(6), 868, 2016. doi:10.18517/ijaseit.6.6.1379.

[17] M. Esteves, A. Pereira, N. Veiga, R. Vasco, A. Veiga, "The use of new learning technologies in Higher education classroom: case of study," International Journal of Engineering Pedagogy, **8**, 115-127, 2018, doi:10.3991/ijep.v8i2.8146

[18] T. Hailikari, T. Tuononen, A. Parpala, "Students' experiences of the factors affecting their study progress: differences in study profiles," Journal of Further and Higher Education, **42**(1), 1–12, 2018. doi: 10.1080/0309877X.2016.1188898

[19] J. Hamari, J. Koivisto, H. Sarsa, "Does Gamification Work? - A Literature Review of Empirical Studies on Gamification," Proceedings of the 47th Hawaii International Conference System Science, 2014, doi: 10.1109/HICSS.2014.377.

[20] E. Harrell, "Kahoot! Reached 70 million unique users on its platform. Kahoot!," [Accessed 7 November 2021], 2019, Available at: https://kahoot.com/blog/2018/01/18/70-million-unique-users-kahoot/

[21] I. Vick, "Training professionals from three countries share their Kahoot!'," **19**, 2019, https://kahoot.com/blog/2019/09/10/top-training-tips-kahoot-around-world/: Kahoot!.

[22] Kahoot! (2015). KAHOOT! is one of the worlds faster growing learning brands. [Accessed 14 November 2021], 2015, available at http://support.kahoot.com/hc/en-us/categories/115000091647-About-Kahoot.

[23] A. Alawadhi, E.A.S. Abu-Ayyash, "Students' perceptions of Kahoot! An exploratory mixed-method study in EFL undergraduate classrooms in the UAE," Education and Information Technologies, 1-30, 2021, doi:10.1007/s10639-020-10425-8.

[24] M.J. Poblaciones, T. Garcia-White, C. Marin, "Students´ Perception of Real-Time Quiz Kahoot! As a Review Tool in Higher Education: A Case of Study," International Journal of Engineering and Pedagogy, **11**(4)**,** 2021, doi: 10.3991/ijep.v11i4.21359.

[25] A. Wang, R. Thair, "The effect of using Kahoot! For learning. A literature reviews," Computers & Education, **149**, 103818, 2020, doi:10.1016/j.compedu.2020.103818.

[26] M. Kalogiannakis, S. Papadakis, A.I. Zourmpakis, A.I. "Gamification in Science Education," A systematic review of the Literature, **11**, 22. 2021, doi:10.3390/educsci110.10022.

[27] D. López Carrillo, A. Calonge García, T. Rodríguez Laguna, G. Ros Magán, A.J. Lebron Moreno, "Using Gamification in a teaching innovation Project at the University of Alcalá," A New approach. Electronic Journal e-Learn, **17**, 93-106, 2021, doi:10.34190/jel.17.2.03.

[28] D.R. Asa, C. Gunn, "Improving problems solving skills in introductory physics using Kahoot!," Physics Education, **53**, 053001, 2018, doi:10.1088/1361-6552/aacade.

[29] B. Klimova, J. Kacetl, "Computer game-based foreign language learning: Its benefits and limitations". In International Conference on Technology in Education, 26–34, 2018, Springer.

[30] A. Inge, "Results from using various quiz-approaches in class," Kahoot Journal. [Accessed 14 November 2021], 2013, available at: https://goo.gl/Wqu8D1

[31] R. Cutri, L. Marim, J. Cordeiro, H. Gil, C. Guerald, "Kahoot, A New and Cheap Way to Get Classroom-Response Instead of Using Clickers," ASEE Annual Conference & Exposition Proceedings, 2016, doi:10.18260/p.25512.

[32] R. Raju, S. Bhat, S. Bath, R. D´Souza, A.B. Singh, "Effective usage of gamification techniques to boost student´s engagement," Journal of Engineering Transformations, **34**, 713-717, 2021, doi:10.16920/jeet/2021/v34i0/157171.

[33] A. Alawadhi, E.A.S. Abu-Ayyash "Students' perceptions of Kahoot! An exploratory mixed-method study in EFL undergraduate classrooms in the UAE," Education and Information Technologies, 1-30. 2021, doi:10.1007/s10639-020-10425-8.

[34] L. Rodríguez-Fernández, "Smartphones and learning: use of Kahoot in the university classroom," Mediterranean Journal of Communication, **8**(1), 181-190, 2017. doi: 10.14198/MEDCOM2017.8.1.13

[35] G. Del Cerro, "Aprender jugando, resolviendo: diseñando experiencias positivas de aprendizaje," XII Jornadas Internacionales de Innovación Universitaria Educar para transformar: Aprendizaje experiencial, 2015, Available in: https://goo.gl/x6Z70t.

[36] A. Fernández, J. Olmos, J. Alegre, "Valor pedagógico del repositorio común de conocimientos para cursos de dirección de empresas," Revista d´innovació Educativa, **16,** 2016, doi: 10.7203/attic.16.8044

[37] P. Kaur, R. Naderajan, "Kahoot! In the English language classroom," South East Asia Journal of Contemporary Business, Economics and Law., **20**(6), 49–54, 2019.

[38] M. Sahrir, G. Yusri, "Online vocabulary games for teaching and learning Arabic," Journal of Language Studies, **12**(3), 961–977, 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 1, 14-26 (2022)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Value Trace Problems for Code Reading Study in C Programming

Xiqin Lu[1], Nobuo Funabiki[*,1], Htoo Htoo Sandi Kyaw[1], Ei Ei Htet[1], Shune Lae Aung[2], Nem Khan Dim[2]

[1]*Department of Electrical and Engineering, Okayama University, Okayama, 700-8530, Japan*

[2]*Department of Computer Studies, University of Yangon, Yangon, 11041, Myanmar*

ARTICLE INFO

Article history:
Received: 05 July, 2021
Accepted: 02 January, 2022
Online: 13 January, 2022

Keywords:
C programming
Value trace problem
Code reading
Self-study
Grammar concept
Algorithm
Pointer

ABSTRACT

C programming is taught in a lot of universities across the world as the first computer programming language. Then, for novice students, it is important to read many simple C source codes and understand their behaviors to be familiar to the programming paradigm. Unfortunately, effective tools to support independent code reading study at home have not been well designed. Heretofore, we have proposed the value trace problem (VTP) for Java programming. A VTP instance consists of one source code, several questions, and the correct answers to them. Each question asks the value of a critical variable or output message in the source code. The correctness of any student answer is checked instantly by string matching at the answer interface for self-study. In this paper, we present the value trace problem (VTP) for code reading self-study of C programming. 42 VTP instances are generated using simple C source codes on basic grammar concepts and fundamental data structures & algorithms in textbooks and websites. In addition, for hard instances on pointer and algorithms, the devices of hints, multiple choice questions, and references are provided to improve their solution performances. For evaluations, we requested 49 *undergraduate students in Japan, China, and Myanmar to independently solve them at home. Their average correct answer rate reached* 94.29%*, where our devices for hard instances improved it by* 33.26%*. Thus, the effectiveness of our proposal is confirmed in motivating self-study of C programming to novice students.*

1 Introduction

Presently, *C programming* is widely taught in many universities across the world as the first computer programming language. In addition to *information technology (IT)* departments, many departments including science, agriculture and mechanical/electrical engineering are teaching *C programming* courses. Actually, the study of *C programming* can offer basic knowledge to study more advanced and practical languages such as *Java*, *JavaScript*, and *Python*. Besides, students in IT departments should study *C programming* in parallel with computer architecture, because they can learn accesses to memories or registers through it. Furthermore, *C* is still the third most popular programming language, despite the long time passed from the appearance. Then, for such novice students, it is important to read and understand a lot of simple source codes to be familiar to the *C programming* structure. Unfortunately, effective tools to support independent *code reading self-study* at home have not been well designed within our knowledge.

Heretofore, we have proposed and implemented the *Java programming learning assistant system (JPLAS)* for *Java programming* study [1]. As the objected-oriented programming language, *Java* is now widely used in IT societies. For assisting *Java programming* studies at different learning levels, JPLAS provides several types of *exercise problems* with different difficulties. For any problem type, the student answer to a question is marked automatically at the system to support programming self-studies.

Among them, the *value trace problem (VTP)* is presented for novice students to study basic concepts of programming grammar and skills through *code reading study* [2]. A VTP instance consists of one source code, several questions, and the correct answer strings to them. Each question asks the value of a critical variable or an important output in the given code. The correctness of any student answer is marked through *string matching* with the correct answer string at the answer interface using the web browser instantly.

The manual collection of C source codes and the manual selection of the variables or output messages from each source code for questions can be the limitation of *VTP*. When a teacher uses *VTP* in the programming course, he/she needs to carefully collect the

[*]Corresponding Author: Nobuo Funabiki, 3-1-1 Tsushimanaka, Okayama University, funabiki@okayama-u.ac.jp

source codes and select the variables or messages with their timing whose actual values or contents should be traced in the questions, so that they can be synchronized with the course progress. One way to generate VTP instances by a teacher efficiently is the use of sample source codes that will be provided with the textbook in the course. Usually, the teacher can download them from the website. Then, he/she can select the variables or output messages that are closely related to the teaching topics in each class of the course. The impossibility of practicing source code writing by a student will be another limitation of *VTP*. It should be offered by other problem types such as the *code writing problem* [3].

In this paper, we present the *value trace problem (VTP)* for *code reading self-study* of *C programming* by extending our previous works for Java programming. To assist the study of novice students, 42 simple C source codes on *basic grammar concepts* and *fundamental data structures & algorithms* are collected from textbooks and websites. Here, *fundamental data structures & algorithms* are often taught using *C* programs in IT departments. Then, by analyzing critical variables/messages and adding necessary standard output statements in them, the corresponding VTP instances are generated manually.

The memory management using *pointer* in *C programming* can be a hard topic for novice students to study, although it is the remarkable feature in implementing fast and efficient programs. Besides, some fundamental algorithms can be difficult, because novice students are often not familiar to them. Therefore, we additionally provide the devices of hints, multiple choice questions, and references to improve the solving performances by the students for the hard VTP instances on *pointer* and *algorithms*.

For evaluations of the proposal, we asked 49 students in Japan, China, and Myanmar to independently solve the generated 42 VTP instances at home. Then, the results showed that the average correct answer rate among them reached 94.29% that is sufficiently high, where the devices for hard instances improved it by 33.26%. Thus, the effectiveness of our proposal is confirmed in motivating self-study of *C programming* by novice students [4]–[10].

VTP can be applied for studying other programming languages such as *JavaScript* and *Python* in addition to *Java* in a straightforward way. Besides, *VTP* can be used for studying a natural language such as English or Japanese as the potential application. For natural language study, the instance may give a sentence or a paragraph and ask the appearing keywords in the questions. Here, the variable for programming study becomes the key concept in the sentence or paragraph for natural language study, and the value becomes the keyword.

The remaining part of this paper is organized as follows: Section 2 discusses related works in literature. Section 3 presents the *value trace problem (VTP)* for *C programming*. Section 4 presents devices to improve the solving performances for hard instances on *pointer* and *algorithms*. Section 5 shows their evaluation results. Finally, Section 6 concludes this paper with future works.

2 Related Works in Literature

In this section, we introduce related works in literature. Some papers presented programming study tools, and some discussed the importance of code reading in programming study.

2.1 Programming Study Tools

In [11], the author reviewed some tools to support teaching and learning of programming, and categorized them into four groups: 1) tools that include simple reduced development environments, 2) example-based environments, 3) tools that are based on visualizations and animations, and 4) simulation environments. Our proposal can be categorized to 2).

In [12], the author presented a tool called *QuizPACK*, which is similar to the *value trace problem (VTP)*, and showed that this tool significantly improved the knowledge of semantics. A question in *QuizPACK* asks the value of a particular expression (usually a variable) in a fragment of a program. A teacher needs to prepare a set of problems using source code fragments and expressions to be questioned manually. If compared with *VTP* in this paper, the advantage of *QuizPACK* is that the constants related to the questioned variables in the code and their correct answers can be generated dynamically so that different students can solve the same question with the different correct answers. On the other hand, *VTP* fixes the constants and the correct answers. The disadvantage of *QuizPACK* is that it can only ask the ending values of the variables in the code fragment. It cannot ask the values at the different timing of the program. Thus, *QuizPACK* is not suitable for studying *algorithms*, where different algorithms can give the same variable values.

In [13], the author presented a problem-solving environment named *LECGO (Learning Environment for programming using C using Geometrical Objects)* for learning *C programming* by beginners. *LECGO* emphasized: (a) multiple external representations in student learning, (b) motivation through performing problem-solving activities from the familiar and meaningful context, (c) the active participation of students by using hands-on experience, (d) appropriate feedback to aid self-corrections, and (e) holistic, activity-based, multi-media, multi-representational and multi-layered content for learning basic concepts.

In [14], the author presented *Gidget*, which is a game such that the eponymous robot protagonist is cast as a fallible character that blames itself for not being able to correctly write code to complete its missions. Players of *Gidget* can learn programming by debugging the problematic code.

In [15], the author presented a game-based learning support environment for novice students to learn programming. This environment exploits game composition tasks to make the elementary programming more intuitive to be learnt, and comprises concept visualization techniques, to allow the students to learn key concepts in programming through game object manipulations.

In [16], the author proposed the *web-based programming assisted system for cooperation (WPASC)* for facilitating cooperative programming learning, and investigated cooperative programming learning behaviors of students and its relationships with learning performances.

In [17], the author presented systematic literature review results on assessment tools for programming assignments, to help instructors make their selections in programming courses. They identified three specialties in assessment tools, namely, *contests*, *quizzes*, and *software testing*. In *contests*, a tool compiles the source code of a

student, runs it with a set of test cases for the input, and informs whether the code is accepted or not. In *quizzes*, it gives a set of questions, where for each question, a student enters a fragment of a code in the interface of the tool. In *software testing*, it checks the correctness of the source code of a student by comparing its output with the output of the model code, or by running the test code against the source code. VTP in this paper is different from any of them.

In [18], the author presented a system that combines the dual objectives of automated grading and program repairing for introductory programming courses called *GradeIT*. It grades the submitted source code with the number of passed test cases, the inverse of time spent to solve, and the fraction of successful compilations. For the last one, the compilation score is defined by 1 – (number of compilation errors in the code) / (maximum number of compilation errors among students). For repairing, it uses simple re-writing rules to fix simple but frequent compile time errors.

In [19], the author proposed a plug-in system to *Moodle* called *LAPLE (Learning Analytics in Programming Language Education)* to provide a learning dashboard to capture the behaviors of the students in the classroom and identify the difficulties faced by different students looking at different knowledge. *LAPLE* asks the students to write full source codes, and collects and analyzes the compiling logs of the students every five minutes to yield real-time visualizations for feedbacks during the class. The authors analyzed the distribution of the classified 36 error types, and encouraged the students to pay attentions to the frequent error types. *LAPLE* is designed for use at programming classes, while *VTP* is for use at homes for self-studies.

2.2 Code Reading in Programming Study

In [20], the author concluded that through literature reviews and interviews, *code reading* is connected to comprehending programs and algorithms, or algorithmic ideas, as well as details, and is needed in many aspects of learning programming, but at the same time, there is not much knowledge about the reading and comprehension process of learners. They claimed that a possible means to foster programming learning is to teach code reading directly, including reading strategies. The value trace problem in this paper can be one way to achieve it for *C programming* study, which supports the novelty of this paper.

In [21], the author concentrated on finding out whether reading programs before writing programs is more efficient for students to learn programming. He asked software developers and found that computer science graduates should understand existing codes, since most developers need to maintain and extend existing systems. He concluded that if students get through the syntax of programming more quickly by conducting code reading exercises, class time could be spent on issues such as the efficiency and readability of software. Code reading problems are important for beginners to learn a new programming language.

In [22], the author concluded that programming learning should focus on algorithmic thinking, rather than programming language features and skills. Programming learning could help improve problem solving skills through algorithmic thinking. The generated VTP instances in this paper cover important algorithmic topics such as stack, queue, and sorting algorithms, which will be helpful for developing algorithmic thinking.

In [23], the author indicated that source code reading behaviors are important in teaching programming skills as well as designing IDEs and programming languages. A source code is a different type of text than a natural-language text. It is highly formal and structured, and has a very limited vocabulary of keywords, operators, and separators, yet, tremendous combination possibilities for literals and identifiers. They recorded programmers' eye movements when reading short source codes. The results show that most attention is oriented towards understanding of identifiers, operators, keywords, and literals, relatively little reading time is spent on separators.

3 Value Trace Problem for C programming

In this section, we present the *value trace problem (VTP)* for *C programming*.

3.1 Definition of Value Trace Problem

The mission of *VTP* is to let students trace the value of a critical variable or an important output in the given source code by carefully reading and understanding its behaviors. To generate a VTP instance, one *source code*, a set of *questions* with the *answer forms*, and the *correct answer strings* to them need to be prepared. A question requests the current value of the important variable or message in the source code that is specified by the teacher. Therefore, the source code should contain the *standard output* statements that correspond to them, so that the students can easily find which values are asked in the questions.

3.2 Design Goals of Value Trace Problem

VTP for *C programming* have the following design goals:

1) A variety of C source codes are given to beginners for *code reading study*.

2) A student can solve the questions in a VTP instance by reading the given source code and understanding its behavior carefully.

3) Any answer from a student will be checked through *string matching* at the interface automatically, and the result will be returned to the student instantly.

4) A teacher can choose proper source codes in terms of contents and difficulty levels for the VTP instance that will be assigned at each class in the programming course.

3.3 VTP Instance Generation Procedure

A new VTP instance can be generated by the following procedure:

1. A proper *source code* for studying a *basic grammar concept* or a *fundamental data structure/algorithm* is selected.

2. The important *variables* or *output messages* to be traced in the source code are selected.

3. The corresponding *standard output statements* are added in the code.

4. The questions of asking the values/messages at the standard outputs and their *correct answers* are prepared.

5. The source code, the corresponding questions, and the correct answers are put together into one *input text file*.

6. The program in [4] is executed with the *input text file* to generate the *CSS/HTML/JavaScript files* for the *answer interface* on the web browser.

7. The generated VTP instance is registered as the assignment to students.

3.4 Answer Interface

The *answer interface* using the web browser in [4] is adopted for the students to solve the generated VTP instances in this paper. The necessary functions such as *string matching* for answer marking and the answer data storage for submissions are implemented in *JavaScript* so that this interface can be used without the Internet connection.

To use this answer interface, the correct answers to the questions need to be distributed to the students, so that their answers can be marked instantly at offline. Then, to prevent the students from looking at the correct answers before solving the questions, the hash values of them are taken using *SHA256* before the distributions. Actually, each correct answer is concatenated with the assignment and problem IDs before hashed, so that the same correct answers for the different questions are converted to the different hash values. It can avoid knowing the correct answer from the same hash value for a different question. At the marking, any answer from a student is concatenated and hashed using *SHA256* for *string matching* with the hashed correct answer.

3.5 Example VTP Instances

Two VTP instances generated in this paper are illustrated here as the typical examples for the *variable trace study* and the *pointer trace study* in *C programming*.

3.5.1 VTP Instance for Variable Trace Study

Figure 1 shows the *input text file* for the *if in for-loop* instance at *ID=12* in Table 1. This file contains the source code, the questions with four answer forms, and their correct answers, which must be done manually. Each correct answer is separated by ",". This instance asks to trace the value of *num* and the logic implemented by two *if*. A student needs to read and understand the source code, and then fill in the forms with correct answers.

```
int main(void){
   int max = 9;
   int n;
   for (n = 2; n <= max; n++){
      if (n == max){
         printf("%d is prime number \n", max);
      }
      else if(max % n == 0){
         printf("%d is not prime \n", max);
         break;
      }
   }
   return 0;
}

What is the value of max? _1_
What is the output? _2_ is _3_ _4_

9,9,not,prime
```

Figure 1: Input text file for VTP instance at *ID=12*.

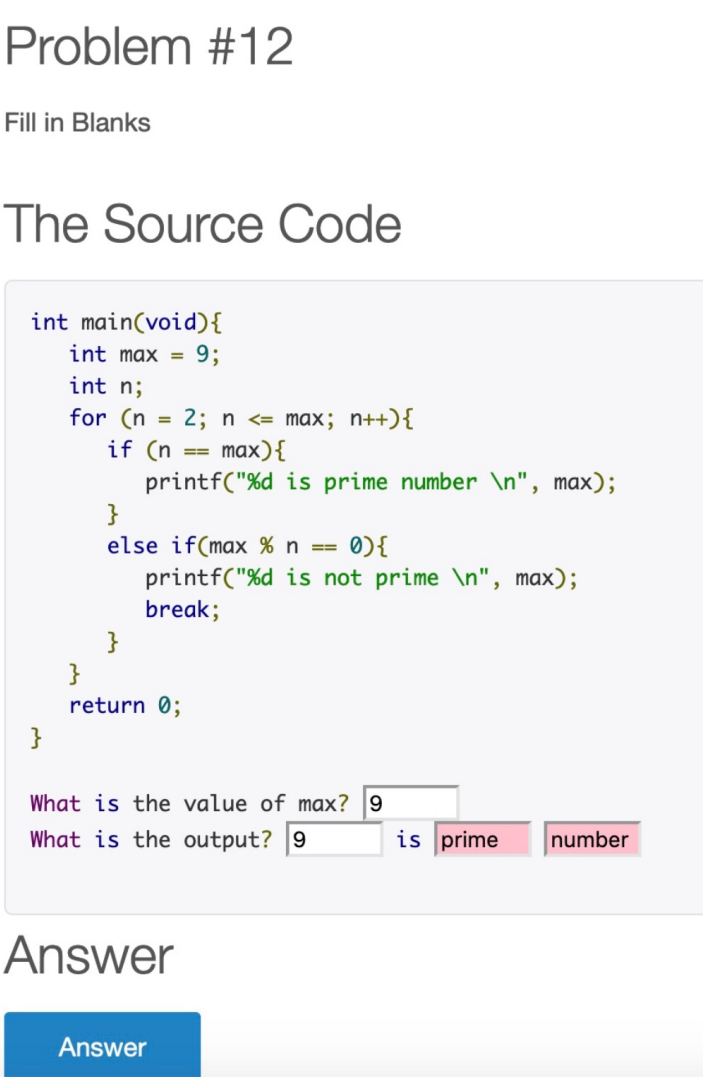

Figure 2: Answer interface for VTP instance at *ID=12*.

```
int main(void)
{
   int test[5] = {80,60,55,22,75};

   printf("test[0] is %d \n", test[0]);
   printf("test[1] is %d \n", test[1]);
   printf("The address of test[0] is %p \n", &
       test[0]);
   printf("test is %p \n", test);
   printf("The address of test[1] is %p \n", &
       test[1]);
   printf("That is *test is %d \n", *test);
   return 0;
}

hint:The address of test[0] is 0028FF20
     1 word 4 bit in this system

What is the output of this program?
test[0] is  _1_
test[1] is  _2_
The address of test[0] is  _3_
test is  _4_
The address of test[1] is  _5_
That is *test  _6_

80,60,0028FF20,0028FF20,0028FF24,80
```

Figure 3: Input text file for VTP instance at *ID=9*.

Figure 2 illustrates the answer interface on a web browser for this VTP instance. After a student fills in the answer forms, he/she should click the blue button "Answer". If the student answer is incorrect, the form color will be *red*. For the correct one, it becomes *white*. Using this interface, a student can fix the mistakes and submit the answers instantly, until all the answers become correct.

3.5.2 *VTP Instance for Pointer Trace Study*

Figure 3 shows the *input text file* for the *memory address with pointer for integer array* instance at *ID=9* in Table 1. This instance asks to trace the memory address or the value given by *pointer* including the one for the array *test*. Figure 4 depicts the answer interface for this sample VTP instance.

3.6 *Generations of Value Trace Problem Instances*

Now, we present the generated 42 *value trace problem (VTP)* instances with 586 answer forms for studying *basic grammar concepts* and *fundamental data structures & algorithms* in *C programming*.

Table 1 shows the topic, the number of lines (#of lines) in the source code, the number of questions, and the number of answer forms in each VTP instance. Here, to encourage novice students solving all of them without giving up halfway, we took the following stances at their generations:

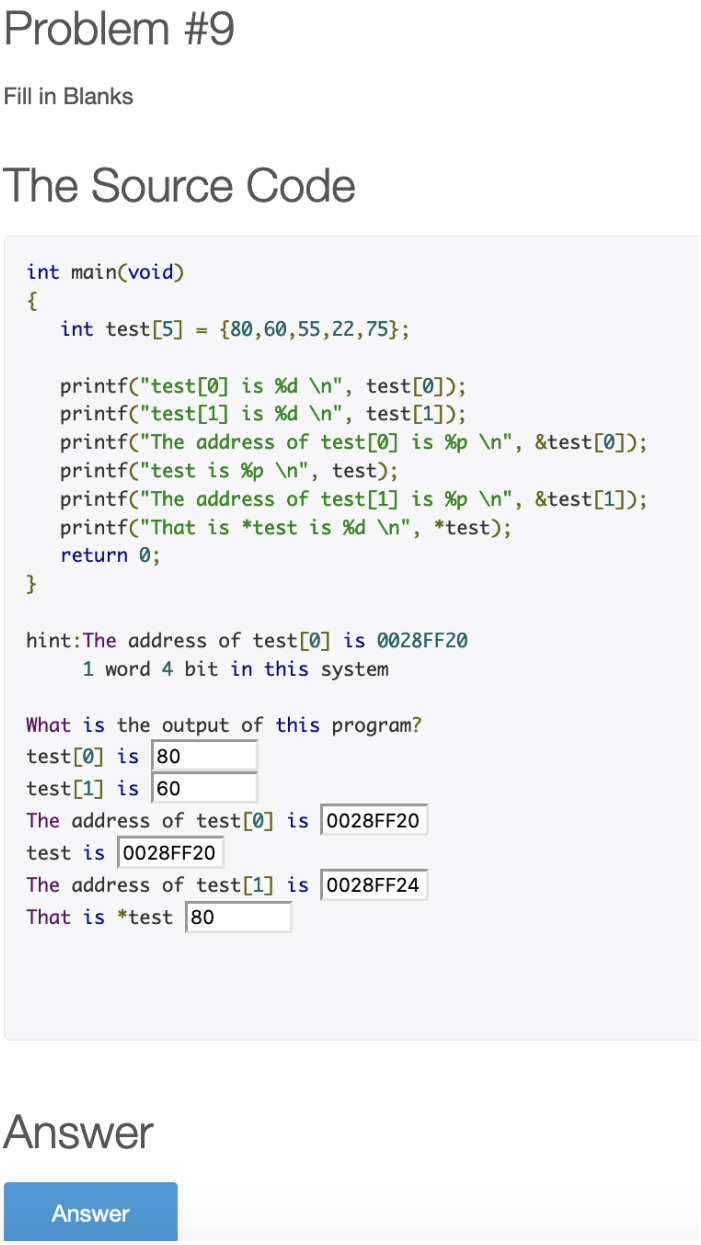

Figure 4: Answer interface for VTP instance at *ID=9*.

1. A short source code with less than 30 lines is basically selected for each instance. Actually, the average number of lines among the source codes becomes 24.5. However, source codes for *fundamental data structures & algorithms* are longer due to their natures.

2. The number of questions is limited to at most four for each instance. Actually, the average number of questions becomes 1.26.

3. The difficulty level of each instance is gradually increased from the first to the last roughly, by following *C programming* textbooks.

3.6.1 VTP Instances for Basic Grammar Concepts

For studying *basic grammar concepts*, 26 VTP instances were generated using the C source codes in the websites [5]-[7]. The selected grammar concepts are essential in fundamental *C programming*.

3.6.2 VTP Instances for Fundamental Data Structures & Algorithms

To study *fundamental data structures & algorithms* of C programming, 16 VTP instances were generated using the source codes in the textbook [8]. They can be hard subjects for novice students in IT departments. To let them read the implemented algorithm procedure in each source code step-by-step and understand the behaviors, the hallway actual values of the important variables in the code are asked in the questions. For example, the actual values of the array variables to store the sorted data and the variable for *pivot* at each iteration are asked in the VTP instance for *quick sort*. These questions can also avoid answering them without correctly understanding the algorithm implementation in the source code.

4 Devices for Hard VTP Instances

In this section, we present our devices to improve the solution performances for hard VTP instances on *pointer* and *algorithms*.

4.1 Hard VTP Instances

In our preliminary application of our generated VTP instances to limited students, the instances at *ID=22*, *ID=28*, *ID=31*, *ID=38*, and *ID=41* in Table 1 showed low correct answer rates compared with the others. These results suggest that not a few students do not understand the topics of *memory address*, *linked list*, *reverse polish notation*, *quick sort*, and *data size of pointer* well. They are related to memory managements by *pointer* or *algorithms*.

To improve the comprehensions on these topics and the solution performances of students, the following devices are presented in this paper:

1) We add *tips* in questions on pointers.
2) We give *hints* in the answer interface on memory addresses.
3) We prepare *multiple choice questions* on linked lists.
4) We add *links to reference websites* in the answer interface on algorithms.

4.2 Tips in Questions

On *tips* for pointers, Figures 5 and 6 show the source code and the questions for the VTP instance at *ID=41*. In the questions, the size of one word in this system is described by *(64-bit size system)*, and the targeting data in each question is described by *Data size of normal variables*, *Data size of pointer*, and *Data size of pointer reference*.

Table 1: Generated 42 VTP instances.

ID	topic	#of lines	#of questions	#of forms
1	two-dimensional array	18	1	12
2	max function	7	1	5
3	arithmetic	6	1	6
4	data type	12	1	16
5	for-loop	10	1	1
6	for-loop with character	11	1	9
7	function call	12	1	10
8	function for swap	21	1	4
9	memory address with pointer for integer array	11	1	56
10	car structure	12	1	2
11	output in for-loop	9	1	8
12	if in for-loop	15	2	4
13	book structure	14	1	3
14	pointer for one-D array	8	1	2
15	if in if with integer	24	2	2
16	arithmetic with if	11	2	2
17	if in if with character	27	2	4
18	while loop	19	1	8
19	structure data setting	18	1	7
20	output format specifier	11	1	4
21	memory address with pointer for character array	10	4	11
22	memory address with pointer for various array	24	1	13
23	four arithmetic operations	11	1	4
24	greatest common divisor (GCD)	21	1	4
25	inner product operation	16	1	1
26	linear search	23	1	5
27	binary search	29	1	17
28	linked list	67	1	7
29	tree traversal algorithms	81	3	63
30	stack data structure	47	1	8
31	reverse polish notation	67	2	12
32	queue data structure	55	1	4
33	insertion sort algorithm	25	1	30
34	selection sort algorithm	28	1	18
35	bubble sort algorithm	27	1	37
36	shell sort algorithm	24	1	23
37	merge sort algorithm	34	1	16
38	quick sort algorithm	38	1	56
39	heap sort algorithm	44	1	42
40	pointer	12	1	6
41	data size of pointer	20	1	9
42	list structure	47	1	12
maximum		81	4	63
minimum		6	1	1
average		24.50	1.26	13.95
total		1029	53	586

```
#include <stdio.h>
main()
{
    char vch, *pch;
    int vin, *pin;
    double vdo, *pdo;
    printf("sizeof(vch) = %d\n", sizeof(vch));
    printf("sizeof(vin) = %d\n", sizeof(vin));
    printf("sizeof(vdo) = %d\n", sizeof(vdo));

    printf("sizeof(pch) = %d\n", sizeof(pch));
    printf("sizeof(pin) = %d\n", sizeof(pin));
    printf("sizeof(pdo) = %d\n", sizeof(pdo));

    printf("sizeof(*pch) = %d\n", sizeof(*pch));
    printf("sizeof(*pin) = %d\n", sizeof(*pin));
    printf("sizeof(*pdo) = %d\n", sizeof(*pdo));
}
```

Figure 5: Source code for VTP instance at *ID=41*.

```
(64-bit size system)
What is the output of this program?
Data size of normal variables
sizeof(vch) = _1_
sizeof(vin) = _2_
sizeof(vdo) = _3_

Data size of pointer
sizeof(pch) = _4_
sizeof(pin) = _5_
sizeof(pdo) = _6_

Data size of pointer reference
sizeof(*pch) = _7_
sizeof(*pin) = _8_
sizeof(*pdo) = _9_
```

Figure 6: Questions for VTP instance at *ID=41*.

4.3 Hints in Answer Interface

On *hints* for memory addresses, Figures 7-9 show the source code, the hints, and the questions for the VTP instance at *ID=22*. The hints describe the byte size of each data type with the word size that is necessary to calculate the increase of the memory address, and the memory addresses of the first two elements in the array variables, so that students can answer the question on them.

```
#include <stdio.h>
int main(void){
    int vch, vsh, vin, vfl, vdo;
    vch = sizeof(char);
    vsh = sizeof(short);
    vin = sizeof(int);
    vfl = sizeof(float);
    vdo = sizeof(double);

    printf("%d %d %d %d %d \n", vch, vsh, vin,
        vfl, vdo);//hintA

    char arr1[] = {'o','k','a','y','a','m','a'};
    int arr2[5] = {1,2,3,4,5};
    float arr3[5] = {1.1,2.2,3.3,4.4,5.5};
    double arr4[5] = {1.0,2.2,3.5,22.4,10.0};

    printf("arr1[4] is %c \n", arr1[4]);
    printf("arr2[3] is %d \n", arr2[3]);
    printf("arr3[2] is %f \n", arr3[2]);
    printf("arr4[1] is %f \n", arr4[1]);

    printf("The address of arr1[2] is %p \n", &
        arr1[2]);//hintB
    printf("The address of arr2[2] is %p \n", &
        arr2[2]);//hintC
    printf("The address of arr3[3] is %p \n", &
        arr3[3]);//hintD
    printf("The address of arr4[2] is %p \n", &
        arr4[2]);//hintE
    return 0;
}
```

Figure 7: Source code for VTP instance at *ID=22*.

```
hintA: Data types and sizes(32-bit system)
char: 1 byte
short: 2 bytes
int 4: bytes
float: 4 bytes
double: 8 bytes

hintB: The address of arr1[0] is 0028FE30
       The address of arr1[1] is 0028FE31

hintC: The address of arr2[0] is 0028FE20
       The address of arr2[1] is 0028FE24

hintD: The address of arr3[1] is 0028FEF4
       The address of arr3[2] is 0028FEF8
hexadecimal: the symbols "0"~"9" to represent
    values zero to nine, and "A"~"F" to represent
     values ten to fifteen.

hintE: The address of arr4[0] is 0028FE10
       The address of arr4[1] is 0028FE18
```

Figure 8: Hints for VTP instance at *ID=22*.

```
What is the output of this program?
_1_ _2_ _3_ _4_ _5_
arr1[4] is _6_
arr2[3] is _7_
arr3[2] is _8_
arr4[1] is _9_
The address of arr1[2] is _10_
The address of arr2[2] is _11_
The address of arr3[3] is _12_
The address of arr4[2] is _13_
```

Figure 9: Questions for VTP instance at *ID=22*.

4.4 Multiple Choice Questions

On *multiple choice questions* for linked lists, Figures 10 and 11 show the source code and the questions for the VTP instance at *ID=28*. The multiple choice questions of the correct answers are used at Q1 and Q3 as the hints to understand the pointer and structure in the linked list.

```
#include <stdio.h>
#include <errno.h>
#include <string.h>
#include <time.h>
#include <stdlib.h>
struct cell{
    struct cell *next;
    int data;
};
typedef struct cell cell_t;
cell_t *list_alloc(int data){
    cell_t *new = NULL;
    new = (cell_t *)malloc(sizeof(cell_t));
    if(new == NULL){
        fprintf(stderr, "ERROR: list_alloc(): %s\
           n", strerror(errno));
        return(NULL);
    }
    new->next = NULL;
    new->data = data;
    return(new);
}
int list_add(cell_t *header, int data){
    cell_t *next = NULL;
    cell_t *prev = header;
    next = list_alloc(data);
    if(next == NULL) return(-1);
    while(prev->next != NULL){
        prev = prev->next;
    }
    prev->next = next;
    return 0;
}
void list_free(cell_t *header){
    cell_t *temp = header;
    cell_t *swap = NULL;
    while(temp != NULL){
        swap = temp->next;
        free(temp);
        temp = swap;
    }
}
static void list_print(cell_t *header){
    cell_t *p = header;
    printf("list{ ");
    while(p != NULL){
        printf("%d ", p->data);
        p = p->next;
    }
    printf("}\n");
}
int main(void){
    int cnt = 0;
    cell_t *header = list_alloc(0);
    if(header == NULL) return -1;
    for(cnt = 1; cnt < 5; cnt++){
        list_add(header, cnt*7%10);
    }
    list_print(header);
    list_free(header);
    return 0;
}
```

Figure 10: Source code for VTP instance at *ID=28*.

```
Q1: Select one word for each blank from a list of
     choices.
What is data type of *header? _1_
What is data type of header? _2_
(A) pointer
(B) data
(C) cell
(D) cell_t

Q2: What is "data" of the cell pointed by "header
    " in the code?  _3_
What is "next" of the cell pointed by "header" in
     the code?  _4_

Q3: Select one word for each blank from the list
    of choices given below from list_free method.
What is the name of cell that will be freed this
    time? _5_
What is the name of cell that will be freed next
    time? _6_
(A) swap
(B) temp

Q4: In list_add method, what is the name of new
    cell? _7_

Q5: What is the output of this programming?
list { _8_ _9_ _10_ _11_ _12_ }
```

Figure 11: Questions for VTP instance at *ID=28*.

4.5 Links to Websites in Answer Interface

On *links to reference websites* for algorithms, the links to the websites that explain *reverse polish notation* in [9] and *quick sort algorithm* in [10] are included in the answer interface. By clicking these links, the students can easily access to the necessary information to solve them.

5 Evaluations

In this section, we evaluate the generated 42 VTP instances with 586 answer forms by applying them to 49 undergraduate students in Myanmar (21 students), Japan (18 students), and China (10 students) who are currently studying or have studied *C programming* and *fundamental data structures & algorithms*.

5.1 Solution Results for Individual VTP Instances

First, the solution results for the 42 individual VTP instances by the 49 students are analyzed. Figure 12 shows the total number of answer submission times and the average correct answer rate (%) for each VTP instance submitted by the students. As the summary, Table 2 shows the maximum, minimum, average, and standard deviation (SD) of the total number of submissions and the average correct answer rate among the 42 VTP instances. The average correct rate 94.29% and the average number of submissions 139.17 by all the students (2.84 by one student) for each instance suggest that the generated VTP instances are of moderate difficulty for novices to study *C programming*.

Table 2: Summary of solution results for individual VTP instances.

\	total# of submissions	average correct rate
maximum	628	100 %
minimum	55	82.32 %
average	139.17	94.29%
SD	106.14	4.84%

5.1.1 VTP Instances with Low Results

The VTP instances for *pointer* or *fundamental data structures & algorithms* are generally difficult for the students. For the eight instances at *ID=21*, *24*, *28*, *29*, *30*, *32*, *36*, and *37*, which are on *pointer*, *stack*, *linked list*, and *sorting*, the average correct answer rate is under 90%.

For the VTP instance *ID=41 linked list*, the correct answer rate 82.32% is lowest and the number of answer submissions 628 is highest among the 42 instances. This topic is still hard for the students, regardless of the presented devices in Section 4. However, it is observed that the students tried to solve this instance by referencing to the devices.

For the VTP instance for *merge sort* at *ID=37*, both the correct rate 85.87% and the number of submissions 101 are relatively low. Many students gave up solving it correctly at early time, which suggests their insufficiency in understanding this algorithm. Like for *quick sort*, the device of adding the link to the reference website in the answer interface will be necessary for this instance.

5.1.2 Distribution of Correct Answer Rates

Figure 13 shows the distribution of the correct answer rates among the 42 VTP instances. 34 instances achieved more than the 90% correct rate and seven did 100%, which indicates that these instances are moderate for novice students. However, the rates of eight instances were less than 90%, which need to be improved.

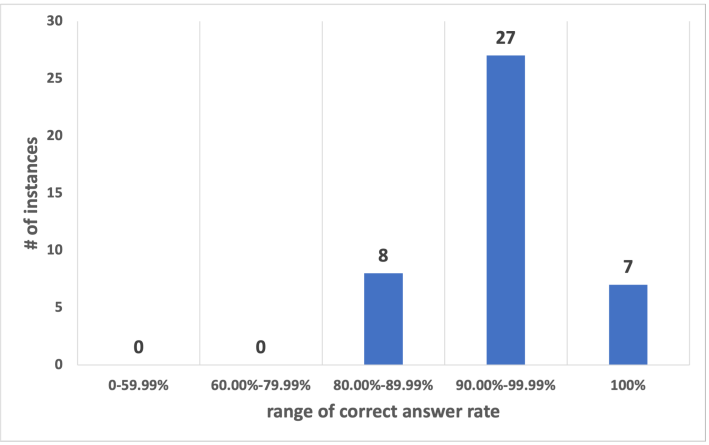

Figure 13: Distribution of correct answer rates among instances.

5.1.3 Distribution of Submission Times

Figure 14 shows the distribution of the number of answer submission times among the VTP instances. 21 instances were solved within 100 submissions. 30 instances were within 150 submissions by all the students or 3.57 submissions by one student. However, for the three instances at *ID=22*, *28*, and *29*, the answers were submitted more than 300 times, which will need to be improved.

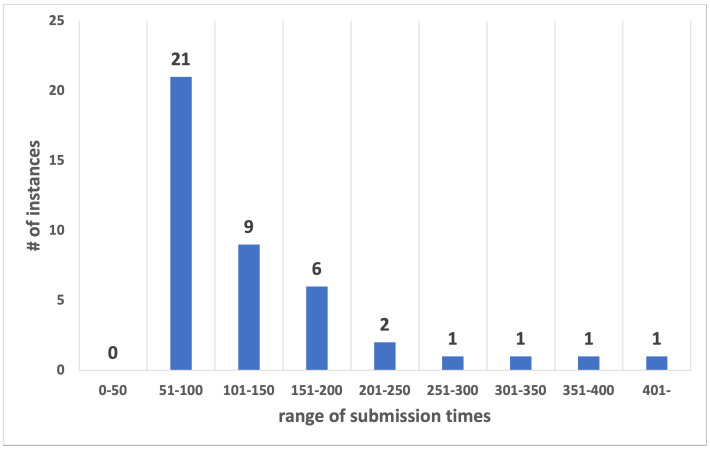

Figure 14: Distribution of submission times among instances.

5.2 Solution Results for Individual Students

Next, the solution results for the individual students to the 42 VTP instances are analyzed. Figure 15 shows the total number of answer

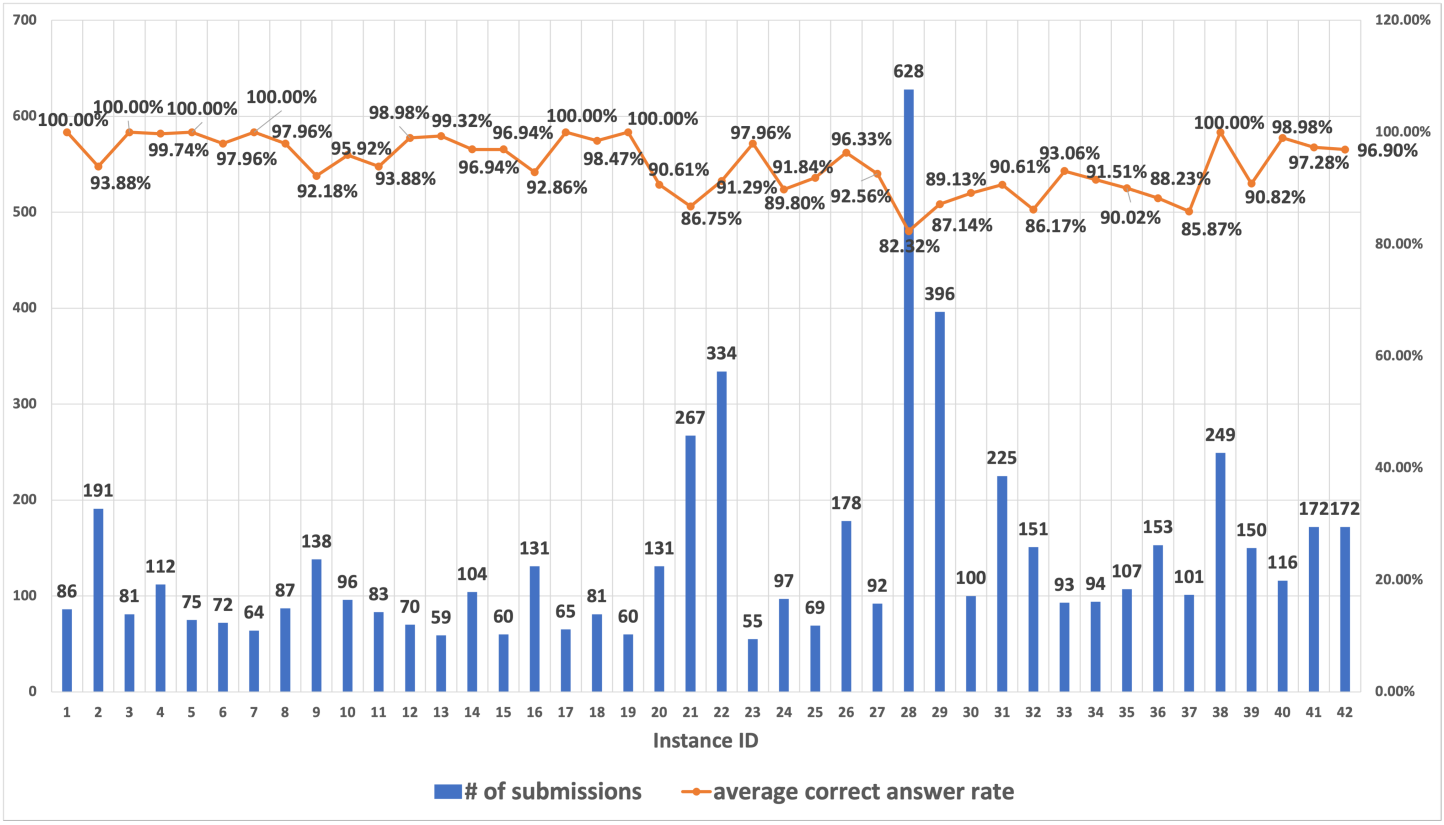

Figure 12: Solution results for individual VTP instances.

submissions and the average of the correct answer rates among all the instances for each student. Table 3 shows the summary results for each student. The results suggest that most of the students well solved the VTP instances.

Table 3: Summary of solution results for individual students.

	total # of submissions	correct answer rate
maximum	394	100%
minimum	38	58.75%
average	119.29	94.29%
SD	89.61	8.25%

5.2.1 *Students with Low Results*

For 11 students at *ID=8*, *10*, *11*, *12*, *13*, *15*, *18*, *19*, *33*, *46*, and *49*, the average correct answer rate among the 42 VTP instances is smaller than 90%.

For the student at *ID=18*, both the correct answer rate 58.75% and the number of answer submissions 38 are lowest among the students. Even, this student did not try to solve some instances. The teacher should find the reason and help this student studying *C programming*. The five students at *ID=10*, *12*, *33*, *46*, and *49* also have the similar proclivities.

For the two students at *ID=8* and *15*, the number of submissions is relatively high. Thus, they made sufficient efforts to solve the VTP instances. It can be expected that the proper guidance by the teacher can quickly improve their performances.

5.2.2 *Distribution of Correct Answer Rates*

Figure 16 shows the distribution of the correct answer rates among the 49 students. 38 students achieved over 90% correct answer rates. Eight students at *ID=22*, *24*, *26*, *27*, *28*, *32*, *39*, and *44* achieved 100%.

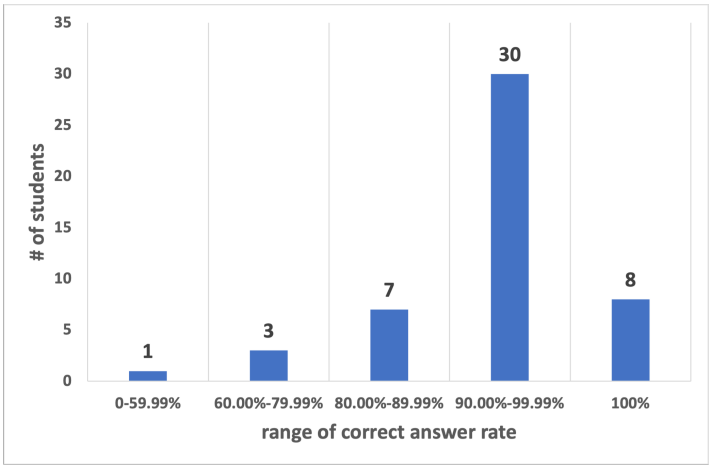

Figure 16: Correct answer rate distribution of students.

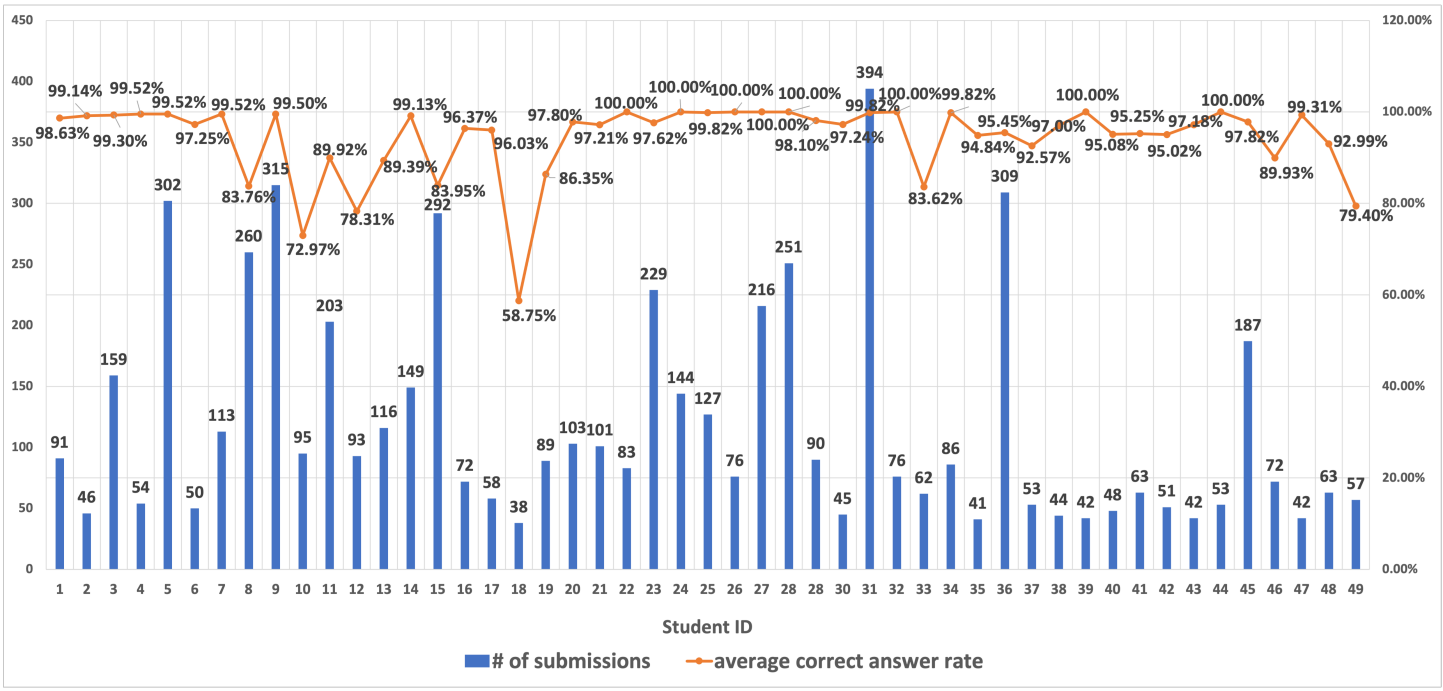

Figure 15: Results for each student.

5.2.3 *Distribution of Submission Times*

Figure 17 shows the distribution of the number of answer submission times among the students. For 30 students, the total number of submissions was smaller than 100, where the excellent student at *ID=39* correctly solved any instance except one instance by submitting the answers only one time. For the four students at *ID=5*, *9*, *31*, and *36*, the total number of submissions exceeded 300, and the answer rate is larger than 90%. It was found that these students did not give up solving them.

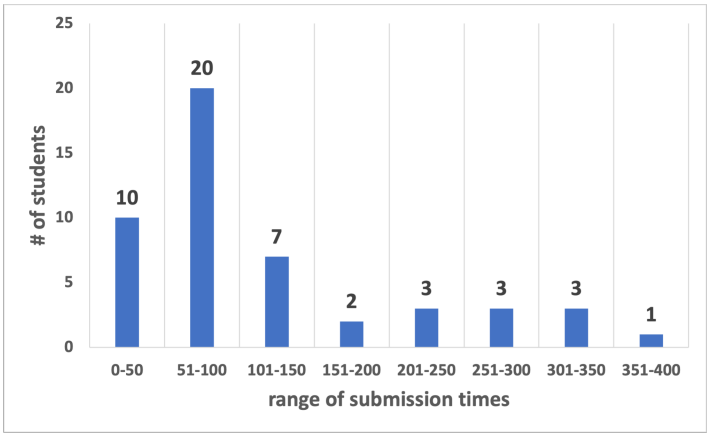

Figure 17: Submission times distribution among students.

5.3 *Evaluations of Devices for Hard Instances*

To evaluate the effectiveness of the presented devices in Section 4, we compare the solution results *with* and *without* using them for the four hard VTP instances at *ID=22*, *28*, *31*, and *38*.

5.3.1 *Comparison Results*

In our preliminary application, we applied the four instances to 21 students among 49 *without* using the proposed devices. Then, to avoid the solution improvements by solving the same instances twice by the same students, the results of the remaining 28 students are used as the results *with* them.

Table 4 compares the average number of answer submission times and the average correct answer rate for the four instances by the students between the *with* group and the *without* group. In the *without* group, the average correct answer rate for any instance was smaller than 70%, which suggests that many students suffered from solving them. Then, in the *with* group, it was improved to higher than 80%, which can be considered as the minimum rate that every student should exceed. The average number of submissions was increased except for *ID=31*, because the students were motivated to continue solving them due to the devices.

Table 4: Comparisons of solution results for hard instances.

ID	topic	average # of submissions		average correct answer rate	
		with	*without*	*with*	*without*
22	memory address with pointer for various array	7.95	7.17	91.29%	54.40%
28	linked list	14.95	1.92	82.32%	56.00%
31	reverse polish notation	5.36	6.88	90.61%	52.00%
38	quick sort algorithm	5.93	1.71	100.00%	68.79%
average		8.55	4.42	91.06%	57.80%

5.3.2 t-test *Analysis*

t-test is applied to the solution results to check the significant difference between the average correct answer rates of the two groups. The *p-value* is $1.179E-16$ and is smaller than 0.05, which means that there is the significant difference between them. Thus, it is confirmed that the proposed devices for the hard VTP instances are effective in improving the solution performances of novice students.

5.4 *Discussions*

At solving the VTP instances, some students may reach the correct answers without understanding them well. They may randomly submit possible answers or copy the answers of other students. Fortunately, the answer interface can record any submitted answer of a student. By analyzing the records, it may be possible to find the undesirable behaviors of students, which will be in our next study.

6 Conclusion

This paper presented the *value trace problem (VTP)* for independent *code reading study* of *C programming*. 42 VTP instances were generated using simple C source codes for *basic grammar concepts* and *fundamental data structures & algorithms* in textbooks or websites. Besides, the devices of tips, hints, multiple choice questions, and links to reference websites were presented for hard instances on *pointer* and *algorithms* to improve the solution performances.

For evaluations, the generated 42 VTP instances were assigned to 49 undergraduate students in Japan, China, and Myanmar to be solved at home using the answer interface that can run offline. The average correct answer rate among all the VTP instances reached 94.29%, where by our proposed devices, the average rate for the four hard instances was improved by 33.26% from 57.80% before improvements. Thus, the effectiveness of our proposal was confirmed in motivating *self-study* of *C programming* to novice students at home.

A limitation of *VTP* is the manual selection of variables or output messages in the source code for questions, in addition to the manual collection of C source codes. An algorithm should be investigated to automatically generate a new VTP instance from a given source code. Another limitation is the impossibility of practicing source code writing by students.

As future works, we will study the automatic VTP generation algorithm, make new VTP instances using C source codes for other grammar concepts and algorithms, such as file I/O, error handing, recursions, and graph algorithms, and apply them to students in *C programming* courses. Then, more comprehensively and formally, we will study the methodologies of organizing and conducting pedagogical experiments to verify the effectiveness of the proposal in increasing motivations of novice students to study *C programming*, and the methods of processing the results of pedagogical experiments.

References

[1] Sio-Iong Ao, IAENG Transactions on Engineering Sciences - Special Issue for the International Association of Engineers Conferences 2016 Volume II, 2018.

[2] N. Funabiki, K.K. Zaw, W.-C. Kao, "A proposal of value trace problem for algorithm code reading in Java programming learning assistant system," Information Engineering Express, 1(3), 2015, doi:10.52731/iee.v1.i3.39.

[3] K.K. Zaw, W. Zaw, N. Funabiki, W.C. Kao, "An informative test code approach in code writing problem for three object-oriented programming concepts in java programming learning assistant system," IAENG International Journal of Computer Science, **46**(3), 2019.

[4] N. Funabiki, H. Masaoka, N. Ishihara, I.W. Lai, W.C. Kao, "Offline answering function for fill-in-blank problems in Java Programming Learning Assistant System," in 2016 IEEE International Conference on Consumer Electronics-Taiwan, ICCE-TW 2016, 2016, doi:10.1109/ICCE-TW.2016.7521045.

[5] Pointer and gcc, https://ubuntuforums.org/archive/index.php/t-858030.html.

[6] FCP 2 If else.ppt, https://dokumen.tips/documents/fcp2ifelseppt.html.

[7] Loop and break, https://www.loopandbreak.com.

[8] T. Hikita, Algorithms by C, Science Pub., 1995.

[9] Reverse polish notation, http://www.cs.man.ac.uk/ pjj/cs212/fix.html.

[10] Quick Sort, https://www.programiz.com/dsa/quick-sort.

[11] M. G`ómez-Albarr`án, The teaching and learning of programming: A survey of supporting software tools, Computer Journal, **48**(2), 2005, doi:10.1093/comjnl/bxh080.

[12] P. Brusilovsky, S. Sosnovsky, "Individualized Exercises for Self-Assessment of Programming Knowledge: An Evaluation of QuizPACK," ACM Journal on Educational Resources in Computing, **5**(3), 6, 2005, doi:10.1145/1163405.1163411.

[13] M. Kordaki, "A drawing and multi-representational computer environment for beginners' learning of programming using C: Design and pilot formative evaluation," Computers and Education, **54**(1), 69–87, 2010, doi:10.1016/j.compedu.2009.07.012.

[14] M.J. Lee, A.J. Ko, "Personifying programming tool feedback improves novice programmers' learning," in ICER'11 - Proceedings of the ACM SIGCSE 2011 International Computing Education Research Workshop, 2011, doi:10.1145/2016911.2016934.

[15] F.W.B. Li, C. Watson, "Game-based concept visualization for learning programming," in MM'11 - Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops - ACM International Workshop on Multimedia Technologies for Distance Learning, MTDL'11, 2011, doi:10.1145/2072598.2072607.

[16] W.Y. Hwang, R. Shadiev, C.Y. Wang, Z.H. Huang, "A pilot study of cooperative programming learning behavior and its relationship with students' learning performance," Computers and Education, **58**(4), 2012, doi:10.1016/j.compedu.2011.12.009.

[17] D.M. Souza, K.R. Felizardo, E.F. Barbosa, "A systematic literature review of assessment tools for programming assignments," in Proceedings - 2016 IEEE 29th Conference on Software Engineering Education and Training, CSEEandT 2016, Institute of Electrical and Electronics Engineers Inc.: 147–156, 2016, doi:10.1109/CSEET.2016.48.

[18] S. Parihar, R. Das, Z. Dadachanji, A. Karkare, P.K. Singh, A. Bhattacharya, "Automatic grading and feedback using program repair for introductory programming courses," in Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE, Association for Computing Machinery: 92–97, 2017, doi:10.1145/3059009.3059026.

[19] X. Fu, A. Shimada, H. Ogata, Y. Taniguchi, D. Suehiro, "Real-time learning analytics for C programming language courses," in ACM International Conference Proceeding Series, Association for Computing Machinery: 280–288, 2017, doi:10.1145/3027385.3027407.

[20] T. Busjahn, C. Schulte, "The use of code reading in teaching programming," in ACM International Conference Proceeding Series, 3–11, 2013, doi:10.1145/2526968.2526969.

[21] T. Vandegrift, "Reading before writing: can students read and understand code and documentation?," SIGCSE '05 Cojoined Meeting, 2005.

[22] D. Kwon, I. Yoon, and W. Lee, "D.Y. Kwon, I.K. Yoon, W.G. Lee, "Design of programming learning process using hybrid programming environment for computing education," KSII Transactions on Internet and Information Systems, **5**(10), 1799–1813, 2011, doi:10.3837/tiis.2011.10.007.

[23] T. Busjahn, R. Bednarik, C. Schulte, "What influences dwell time during source code reading? Analysis of element type and frequency as factors," in Eye Tracking Research and Applications Symposium (ETRA), 2014, doi:10.1145/2578153.2578211.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 27-46 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Physics behind the Concept of a Sodium-Potassium-Cesium-Cooled Martian Nuclear Reactor

Okunev Viacheslav Sergeevich*

Bauman Moscow State Technical University, Department of Physics, The Faculty of Fundamental Science, Moscow, 2-ya Baumanskaya ul. 5, 105005, Russia

ARTICLE INFO

Article history:
Received: 22 November, 2021
Accepted: 25 December, 2021
Online: 22 January, 2022

Keywords:
Martian Nuclear Power Plant
Fast reactor
Eutectic alloy of sodium, potassium and cesium
Safety
Cermets fuel

ABSTRACT

The main goal of the work is to determine the basic conceptual solutions of a nuclear reactor operating on the surface of another planet. The problem was solved using the example of the Martian nuclear power plant. The article uses calculation and optimization research methods and corresponding program codes (well-known and author's). The results of the study made it possible to formulate the basic requirements for the Martian nuclear power plant and select the type of reactor. This is a new type of reactor: pressurized liquid metal fast reactor. It is proposed to use an innovative cermets nuclear fuel based on mixed mononitride and uranium metal nanopowder, which was previously considered by the author for new generation BN and BREST ground-based reactors. It is proposed to use a eutectic (or near-eutectic) NaKCs alloy as a coolant. Optimization of the alloy composition has been carried out. The fuel and coolant of the reactor contains long-lived radioactive waste to be transmuted. NaKCs alloy is less reactive than pure alkali metals including Na, K and Cs. With an electric power of 600 MW, it is possible to ensure the internal self-protection of the reactor. All emergency modes of the ATWS type (anticipated transient without scram) are not hazardous. This means that with a decrease in power to values characteristic of the initial stages of the colonization of Mars, the safety of the reactor is easily ensured. The relatively low chemical activity of the coolant makes it possible to use a two-circuit energy conversion scheme. The second circuit can use water or carbon dioxide. Carbon dioxide is preferred because of its presence in the atmosphere of Mars (95% CO_2). The significance of the research lies in the possibility of constructing a Martian nuclear power plant within the framework of existing technologies.

1. Introduction

The exploration of outer space, the planets closest and distant from us, requires the creation of compact, economical, reliable, low-maintenance (practically autonomous) and safe energy sources with a high energy density. Technologically mastered energy sources using the fission reaction of heavy atomic nuclei are attractive.

There are two directions of using space nuclear reactors. The first is related to the need to travel long distances. This is a nuclear electric propulsion system that combines the concepts of "nuclear space tug", "interorbital tug", "and transport and power module". The second direction is associated with the supply of energy to the spacecraft, space or alien station and crew. It can be a low-power alien nuclear power plant (NPP) that provides power to the station, or a space NPP that provides power to a spacecraft. Its functions are not related to the movement of the spacecraft. This work is devoted to solving the second problem and is a continuation of the research published by the author earlier [1]. At present, the development of nuclear electric propulsion systems is being actively pursued, capable of operating in the mode of obtaining thrust (engine) and in the mode of generating electrical energy (NPP) [2, 3]. If it is necessary to supply power to stations located on other planets, it will be necessary to build a NPP (still of low power, that is, electric power up to 100 ... 300 MW). The present work is devoted to this problem.

All reactors that have ever experienced severe accidents were considered safe. They met all the rules and regulations that were in force at the time of the accident on the territory of the country where these reactors were operated. Serious accidents occurred at the plutonium production reactor of the Windscale nuclear complex (Great Britain, 1957); Unit 2 of the Three Mile Island NPP (USA,

*Corresponding Author: Okunev Viacheslav Sergeevich, okunevvs@bmstu.ru

1979); 4th block of the Chernobyl NPP (USSR, 1986); four units of the Fukushima-1 NPP (Japan, 2011).

After a series of accidents at NPPs, the philosophy of ensuring the safety of nuclear power facilities was radically rethought. The first conceptual designs of nuclear reactors appeared, in which it is possible to exclude all accidents leading to unacceptable releases of radioactive substances outside the NPP.

A comprehensive solution to the following tasks was proposed:

- Provision of energy on the required scale;
- Economic efficiency;
- Safety;
- Provision (self-sufficiency) with fuel.

Safety is understood not only to exclude severe accidents at nuclear power facilities, but also to safely handle waste and create physical barriers to the theft of nuclear materials.

Nuclear power in the 20th century was not limited to terrestrial reactors. Nuclear power has entered low earth orbit. The first developments of NPPs were started simultaneously in the USSR and the USA in 1954 [4]. Space nuclear power was ahead of its time. Modern interplanetary spacecraft use radioisotope energy sources (less than 1 kW) [2]. Russia has experience in the development and use of nuclear electric propulsion systems and NPPs with electric power up to 6 kW in outer space. But such installations have not yet been in demand [2]. Until now, energy sources with a capacity of about 1 MW and more have not been in demand. The provision of spacecraft and orbital stations with energy is achieved through the use of solar batteries (the power of which is estimated at several kilowatts, but can reach ~ 10^5 W), chemical sources of electric current (galvanic cells, batteries, fuel cells), radioisotope sources, nuclear reactors, etc. [5].

Despite the relatively low efficiency (up to 13% [5]), the preference is still given to solar panels. Their specific weight is 4.6 kg / kW; in the future it can be reduced by almost 2 times [5]. Flights to Mars, the creation of stations on the Moon and Mars will require much more energy (from 1 MW and more). Space nuclear power can easily solve this problem.

2. Background

2.1. Specific Requirements for a Space Reactor

Unlike ground-based nuclear power reactors for space reactors, it is necessary to provide a high density of energy release in the core with much lower safety requirements, and the possibility of long-term autonomous operation. For reactors and NPPs built on the surface of the planet, minimization of mass and size is not so urgent.

Relatively low safety requirements are associated with the fact that space reactors are designed to operate with an extremely low population density (crew of a ship or an orbital station; personnel of a station located on the planet under study). When placing a NPP on the surface of a planet (the Moon or Mars), it is advisable to bury the reactor under the soil layer. This is necessary to protect against meteorites in a highly rarefied atmosphere or in the absence of an atmosphere. With this arrangement of the reactor, a severe accident would not destroy or damage the entire space station. In addition, long-term dust storms are characteristic of Mars. Martian dust is iron oxide particles about 1.5 microns in size [6].

The situation is aggravated if the reactor is located on a spacecraft or orbital station. A severe accident of a space reactor will lead to huge economic losses. According to various estimates, the cost of the International Space Station is $ 150 billion. The cost of the Mir station is $ 4.2 billion (in 2001 prices). The commercial value of the Soyuz-2.1 disposable spacecraft (without additional modules) is about $ 35 million ($ 20-30 thousand per 1 kg of payload) [7]. The commercial value of the heavy rocket Proton (Russia) and Falcon 9 (USA, SpaceX) is $ 60-65 and $ 62 million, respectively [7]. The Soyuz-2.1 spacecraft is capable of putting up to 8.25 tons of cargo into low orbit, the Proton spacecraft - about 22 tons, the Falcon 9 - 22.8 tons [7]. NASA estimates the cost of a manned flight to Mars at between $ 400 billion and $ 1 trillion [8]. For comparison, the cost of building a land-based NPP with four VVER-1200 reactors (Turkish nuclear power plant "Akkuyu") is $ 22 billion [9]. The cost of building a fast reactor BN-800 (4th power unit of the Beloyarsk NPP, Zarechny, Russia, electric capacity 880 MW) is estimated at $ 2 billion [10]. The cost of building BREST-OD-300 (Seversk, Russia) is estimated at $ 1.4 billion. [11] The high cost of the project is explained by the rise in prices for materials [11].

When designing space reactors, compactness is important: minimum weight and dimensions. This is achievable for liquid metal cooled reactors (LMR) operating on intermediate neutrons (in the resonance spectrum of neurons). One of the advantages of intermediate neutron LMRs is its relatively small size and lighter weight compared to light water reactors (LWRs). This is attractive for the use of LMRs as transport reactors (on ships of surface and submarine fleets, spaceships and space stations).

2.2. Existing Devices and Developments

The first space reactor on intermediate neutrons (SNAP-10A, thermal power 40 kW, electric power 500 ... 650 W, thermoelectric power converter) was developed by Boeing by order of the Air Force and the US Atomic Energy Commission [12]. Highly enriched fuel was used in the core, the moderator was zirconium hydride, and the coolant was a eutectic alloy of sodium and potassium [12]. In the period from 1970 to 1988, 31 spacecraft with the fast reactor "Buk" (thermoelectric energy conversion, electric power 3 kW, service life up to 4400 hours) were launched into low-earth orbits in the USSR. Two experimental Soviet reactors on intermediate neutrons "TOPAZ" (thermionic energy conversion, electric power 5 kW) were launched into space in 1987 and 1988.

Modern concepts of space reactors, performing the functions of a tug, being developed from Russia and the United States, are guided by an electrical power of ~ 1 MW. In Russia, since 2010, research has been carried out within the framework of the project of the state corporations "Roscosmos" and "Rosatom" ("Creation of a transport and energy module based on a megawatt-class NPP") [13]. The installations are designed for flights into deep space, to the Moon and Mars. For the same purposes, the USA is developing a nuclear electric motor plant "Kilopower" with an electric power of 1 kW (in the future, with an increase in power up to 2 MW) [14]. The service life of such installations is designed for 15 years [14].

At present, the development of nuclear electric propulsion systems is carried out in three directions: three types of installations are being developed [2, 3]:

- The first type is based on the concept of a reactor with direct conversion of thermal energy into electrical energy (thermal emission conversion) and an electric rocket motor;
- The installation of the second type is a reactor with machine energy conversion and an electric rocket engine;
- The installation of the third type should operate in two modes: obtaining traction in engine mode and generating electrical energy in a closed loop.

Several conceptual designs of new generation space NPPs have been developed in Russia [4, 15].

- Four projects with electrical capacity from 0.1 to 150 MW for operation as part of a transport an energy module for the study of distant planets of the solar system;
- Four projects of the Martian NPP with electric power from 25 to 500 kW;
- Four projects of the lunar NPP with electric power from 25 to 500 kW.

It is proposed to use a high-temperature fast reactor in all projects.

NASA and the US Department of Energy have selected three nuclear thermal propulsion concepts. The reactors are intended for deep space exploration (see [16] with reference to "World Nuclear News"). The concepts were developed by the following companies.

- "BWX Technologies" и "Lockheed Martin";
- "General Atomics Electromagnetic Systems", "X-energy" and "Aerojet Rocketdyne";
- "Ultra Safe Nuclear Technologies", "Ultra Safe Nuclear Corporation", "GE Hitachi Nuclear Energy", "GE Research", "Framatome" и "Materion".

Companies are encouraged to develop "different design strategies for specific performance requirements that could help deep space exploration" [16].

The plans of the Russian state corporation "Roscosmos" (within the framework of the "Nucleon" project) include the construction of a NPP on Mars to supply the future Russian Martian base with electrical energy [17]. The NPP is planned to be delivered to Mars orbit using the "Zeus" nuclear tug (nuclear rocket engine, 1 MW) [17]. The NPP will descend to the surface of the planet by parachute. Immediately after landing, the NPP will be ready for operation [17]. The developer of the project is the "Arsenal" design bureau (St. Petersburg). The "Arsenal" design bureau includes a pilot production plant. The plant is capable of performing a full cycle of work from the development of the corresponding (design, technological) documentation to the manufacture of finished products. The United States is planning to build a small NPP (with an electrical power of 10 kW, based on the "Kilopower" concept) on the Moon and Mars.

2.3. Operating Conditions

In the absence of a dense layer of the atmosphere in order to minimize the risk of damage to an alien NPP by meteorites, it should be buried in the soil of the planet [1]. In addition, it is necessary to take into account the daily and seasonal temperature drops on the planet's surface [1]. According to [18, 19], daily temperature drops on the lunar surface average 300 ºC: from –173 ºC (at night) to +127 ºC (during the day). During eclipses (lasting about one and a half hours) the temperature of the lunar surface during the day decreases by 250 ... 300 ºC. In lunar craters near the pole, a temperature of –249 ºC was recorded. On the other hand, at a depth of 1 m, the temperature of the lunar regolith is stable and amounts to –35 ºC [18, 19]. This is attractive for burying NPPs into the ground.

The seasonal change in the temperature of the atmosphere near the surface of Mars is 105 ºC: from –80… 125 ºC to + 20… 25 ºC. The maximum temperature spread is 178 ºC: from –143 ºC (in winter at the pole) to +35 ºC [18, 19]. In mid-latitudes, temperatures range from -50 °C in winter nights to 0 °C in summer days [18]. The average atmospheric pressure on the average radius of Mars is 636 Pa (data from NASA, 2004) [18]. Depending on the season, the pressure varies from 400 to 870 Pa [18]. Due to the large difference in altitude on Mars, the pressure at the surface varies greatly. In deep depressions (Hellas plain), atmospheric pressure is 1.24 kPa [20]. Atmospheric pressure on the lunar surface is about 10 nPa [21]. The lower layer of the Mars atmosphere contains carbon dioxide (95.32%), nitrogen (2.7%), argon (1.6%), other gases and traces of organic compounds [22]. The atmosphere of the Moon is composed mainly of hydrogen, helium, neon and argon [21].

One of the features of space NPPs operating in space far from the planet (relatively large gravitational mass) is associated with the absence of natural circulation of the coolant when the pumps fail. The high level of natural circulation in land-based reactors plays an important role (with rare exceptions) in ensuring their safety. An exception to this pattern is a fast reactor with a heavy coolant (lead or lead-bismuth alloy). For such reactors (for example, BREST-OD-300 [23]), a two-circuit energy conversion scheme is adopted, and water is used as a working fluid (in the second circuit). In emergency modes with depressurization of the steam generator tubes, a high level of natural circulation promotes the entrainment of water vapor bubbles into the core, which leads to an increase in reactivity and can lead to an accident.

In reactors located on Mars and the Moon, the level of natural circulation is noticeably lower than in terrestrial reactors. The reason lies in the lower value of the acceleration g of gravity. If on the surface of the Earth the average value of g is 9.81 m / s^2, then on the surface of Mars $g \approx 3.86$ m / s^2, on the surface of the Moon $g \approx 1.61$ m / s^2. The flow rate G of the coolant through the core in the natural circulation mode is determined by the formula [1]:

$$G = [Hg\Delta\gamma/\Delta p]^{1/1,75}, \quad (1)$$

where H is the height of the natural circulation contour (thrust section), equal to the difference in heights between the average levels of the core and the intermediate heat exchanger for

emergency cooling; $\Delta\gamma$ is the difference in the density of the coolant at the inlet and outlet to the core; Δp is hydraulic pressure loss in the primary circuit.

In fig. 1 shows the qualitative dependence of the coolant flow rate in the natural circulation mode on the acceleration of the combined fall, all other things being equal (i.e., at $H\Delta\gamma / \Delta p = \text{const} = B$).

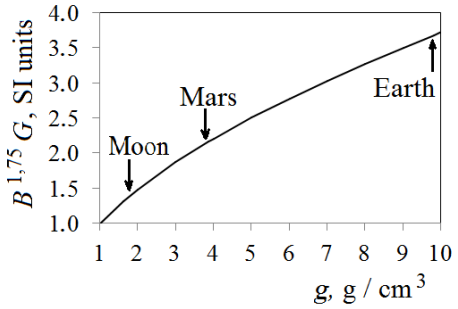

Figure 1: Dependence of the coolant flow rate in the natural circulation mode on the acceleration of the consolidated fall

As noted in [1], "to compensate for the decrease in G, all other things being equal (in comparison with a ground-based reactor), it is necessary to increase H and, consequently, the total height of the NPPt". For reactors located on Mars or the Moon, high values of the height of the natural circulation loop should be chosen. This is possible due to the deepening of the reactor into the soil of the planet. Table 1 shows the characteristic values of H for the lunar and Martian NPPs in relation to H_0 of a terrestrial-based reactor (according to [1]).

Based on the data in table 1, it can be concluded that for a space reactor located at a shallow depth under the surface of Mars or the Moon, acceptable values of the level of natural circulation are achievable to ensure safety in emergency modes with violation of forced circulation of the coolant. Circulation pumps can be eliminated in the design of low-power reactors located on the surface of planets with relatively high gravity.

Table 1: Characteristic values of the height of the natural circulation contour of terrestrial, Martian and lunar-based reactors

The planet on which the reactor is located	H/H_0	Typical values of H, m
Earth	1	5…10
Mars	2.54	up to 10
Moon	6.09	24…25

2.4. Possible Borrowing of Technologies of Ground-Based New Generation of Fast Power Reactors

The design of NPPs (installations capable of operating only in the mode of generating electrical energy) on other planets can be based on borrowing technologies from new generation of ground-based power reactors, taking into account the specifics of the operating conditions of the reactor.

Liquid metal cooled fast reactors (fast reactors or LMFRs) have a higher core energy density than light water reactors (LWRs) and have good potential for improving internal security. Currently, two directions of high-power, ground-based fast power reactors are being developed. The first direction (traditional) involves the use of a liquid sodium coolant. The second direction (innovative) involves the use of a liquid lead coolant. In 2021, Russia began construction of a pilot demonstration reactor BREST-OD-300 with a lead coolant and mixed mononitride fuel (UN-PuN) based on waste uranium and plutonium recovered from spent LWR fuel. Another advantage of fast reactors is the high value of the average number of neutrons ν_f produced during the fission of a heavy nucleus (Fig. 2). This is due to the relatively high kinetic energy of neutrons in the fast reactor.

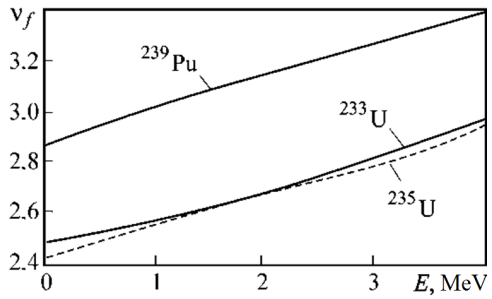

Figure 2: Dependence of the average number of neutrons produced during nuclear fission on the kinetic energy of the neutron that caused the fission [24]

In space reactors, fuel based on highly enriched uranium (without plutonium in the launch load) is used. This is due to some of the properties of plutonium. When using plutonium-239 as the main fissile nuclide, the following advantages are achieved (in comparison with the fissile isotopes of uranium ^{233}U and ^{235}U):

- Maximum energy release;
- Maximum value of parameter ν_f;
- when using mixed fuel (based on ^{238}U and ^{239}Pu), the reactor can operate for a long time (up to 20 years) in a self-fueling mode, i.e., the number of fissile nuclides in the core is constant and practically does not change over time (the BRC - breeding ratio in the core is approximately equal to 1).

These advantages are attractive for the use of mixed fuel (based on ^{238}U and ^{239}Pu) in terrestrial power engineering.

There are two significant disadvantages of plutonium fuel.

- A small fraction of β delayed neutrons produced during the fission of heavy nuclei (during the fission of uranium-235 β = 0.0065, plutonium-239 β = 0.0021 [25]) complicates reactor control and leads to a potential accident hazard (reactor is prompt critical) [24];
- Unpredictable behavior of the crystal lattice (when using metallic plutonium) and a spontaneous change in its shape during operation of the reactor leads to a noticeable change in the volume of fuel [26].

These disadvantages impede the use of plutonium-239-based fuel in transport reactors (space, ship, etc.).

One of the topical and fundamental tasks is the development of space NPPs, combining the advantages of new generation fast power reactors and existing or under development space reactors.

2.5. Tasks solved by a space NPP (based on data [2, 3])

In the near future, space NPPs with an electric power of up to 250 kW are designed to solve the problems of a global communication system, remote sensing of the Earth, environmental monitoring, warning of natural disasters, etc .; comprehensive study of the Earth and solar-terrestrial relations; detailed studies of planets and their satellites; explore the possibilities of industrial production in space; the task of cleaning from space debris; warnings about the approach of dangerous space objects. In the future, space nuclear power (the electric power of the reactor is from 5 to 40 MW) is designed to provide manned flights to the Moon and Mars; sending probes outside the solar system; delivery to Earth and processing of asteroids; combating natural disasters.

3. Materials and Methods

In addition to the general requirements for the choice of materials for space reactors, specific requirements must be taken into account, taking into account the extreme operating conditions. Extreme conditions include high temperatures (usually oriented at 3000 K), intense radioactive irradiation (neutron flux density 10^{14} s^{-1} cm^{-2}), chemically active media (hydrogen in a nuclear rocket engine) [27, 28]. Most of the "traditional" materials for ground-based nuclear technology do not withstand these operating conditions.

In the manufacture of structural materials, preference is given to refractory metals (tungsten, molybdenum, chromium) [29]. Preferred fuels are carbides and nitrides of transition metals, ceramics [29]. To expand the temperature range of operation of fuel and structural materials, introduction phases are used [29-31]: ordered structures that are formed during the interaction of some metals with carbon, nitrogen, hydrogen and oxygen atoms. The disadvantage of interstitial phases is brittle fracture at relatively low temperatures (up to 1000 K) [29-31]. This drawback is eliminated by a special design technology. Materials must be in a compressive stress field. Tensile stresses should be minimized [29, 31].

3.1. "Traditional" promising fuel

Uranium dioxide is rightfully considered the most widespread nuclear fuel. The melting point of UO_2 is 2750 °C [32]. Such fuel is characterized by low density and thermal conductivity. Uranium mononitride melts congruently at 2850 °C if the nitrogen pressure in the system exceeds 250 kPa [32]. Low nitrogen pressure leads to the decomposition of UN to form a liquid uranium phase. The melting point of uranium monocarbide is 2275 °C [32]. Mononitride and monocarbide are characterized by high density and thermal conductivity. This is attractive for the use of such fuels in reactors with a high bulk heat density.

A mixture of uranium and zirconium hydride was used as fuel for the SNAP-10A space reactor. An alloy of uranium with molybdenum was used as fuel for the "Buk" reactor, a uranium dioxide was used for the "TOPAZ" reactor.

Fuel materials (introduction phases) are produced by interaction of heavy metals (Th, U, Pu) with non-metals (oxygen, carbon and nitrogen). The simplest ceramic compounds are UO_2, UC and UN. The advantages and disadvantages of these fuels are well understood. They were studied at the dawn of the development of fast reactors. In the BOR-60 reactor, fuel based on UC, UPuC and UNC was studied [33]. The fuel (UPu) $C_{1-x}N_x$ (where $x < 0.8$), PuO_2-MgO, PuN-CrN, (UZr) N [18, 33], UZrCN (uranium carbonitride) [34] was studied. It is proposed to use highly enriched uranium (up to 90% ^{235}U) [34]. For a nuclear rocket engine, high-temperature fuel based on solid solutions of carbides is considered: UC-ZrC, UC-NbC and ZrC-NbC-UC [35]. The density of ZrC-NbC-UC is relatively low: 7.6 g / cm^3, melting point 3520 K [35]. The carbonitride fuel UC-ZrC-ZrN is being investigated [35]. Fuel based on uranium-zirconium carbonitrides has the highest density among high-temperature fuels.

A space NPP is being developed in the United States. The NPP was developed on the basis of the "Kilopower" project. Uranium metal, UMo alloy, or uranium dioxide is considered as fuel [4]. The installation is designed to generate electricity (up to 10 kW) on the surface of Mars. The service life is 10 years.

So, in Russian space reactor designs, UZrCN carbonitride fuel is considered the most promising. The multi component composition opens up possibilities for optimizing the composition and properties of such a fuel. In the USSR, high-temperature and high-density carbonitride fuel based on highly enriched (96%) and low-enriched (less than 20%) uranium was developed for various types of reactors [36].

Scientists from Belarus, Russia and the United States are conducting joint computational and experimental studies of 0.9U-0.1Zr-0.5C-0.5N fuel (with 19.75% enrichment in ^{235}U) with a density of 12 ... 12.5 g / cm^3 and thermal conductivity

$$\lambda = 11,4 \text{ W m}^{-1} \text{ K}^{-1}$$

for the Russian high-flux research reactor SM-3 [36, 37].

3.2. New Promising Fuel

When designing Martian and lunar NPPs (in fact, space nuclear power plants), borrowing technical solutions suggests that the cermets fuel previously considered by the author for ground-based BN and BREST power reactors can be used in space NPPs [1]. This fuel is sintered pellets of mixed mononitride fuel micro grains (UN-PuN with additives of ^{237}Np, ^{241}Am, ^{243}Am long-lived waste) and uranium metal nanopowder. The composition of UN uses natural (^{235}U content about 0.72%) or depleted (^{235}U content about 0.4%) uranium. PuN uses plutonium recovered from LWR spent fuel and purified from the ^{238}Pu isotope. Plutonium isotopic composition: 0.6 ^{239}Pu - 0.25 ^{240}Pu - 0.109 ^{241}Pu - 0.041 ^{242}Pu. The Pu / (Pu + U) ratio does not exceed 0.2. The mass content of radioactive waste (^{237}Np, ^{241}Am and ^{243}Am) in fresh fuel is about 5%.The total content of fissile nuclides (^{235}U, ^{239}Pu and ^{241}Pu) in the fuel is about 16%.

The nanopowder occupies the voids between the micro grains of the ceramic. As a result, the porosity of the fuel is reduced to 5%. "The ratio of the volumes of micro grains of ceramics and metal nanopowder 20/75 corresponds to a mass ratio of 0.39. This is the maximum permissible mass content of metallic nanopowder in the fuel, associated with the limitation of the maximum permissible temperature" [1]. When using enriched uranium

nanopowder, the reactor power at fixed dimensions can be increased by 40%.

The average density of such a fuel is about 16 g / cm^3, the thermal conductivity is about 27.8 W m^{-1} K^{-1}, which significantly exceeds the corresponding values for mononitride and carbonitride. This contributes to the safe end of emergency situations, including those accompanied by a failure of the emergency protection (ATWS - anticipated transient without scram).

When the fuel temperature rises in emergency modes, melting of uranium nanopowder is possible. However, it will not go beyond the fuel element cladding. At the end of the emergency mode, the nanopowder will again go into the solid phase.

Uranium metal nanopowder acts as a getter of free nitrogen released from micro grains during reactor operation. Nitrogen migrates to the fuel element cladding. In the presence of free nitrogen, the corrosion rate of the inner surface of the shell increases. Getter binds free nitrogen, helping to minimize the rate of corrosion of the inner surface of the shell. After a long-term operation of the reactor, which "survived" emergency situations, uranium metal gradually loses its functions of an effective getter. Its functions in ensuring the safety of the reactor (due to its high thermal conductivity and density) are preserved.

Optimization of fuel properties is possible by changing the ratio of micro grains of mononitride and uranium nanopowder, as well as by changing the enrichment of uranium nanopowder.

From the point of view of ensuring safety in emergency situations (including the most dangerous of them - ATWS), it is sufficient to ensure the mass content of uranium nanopowder in the fuel at the level of 15 ... 20%. From the point of view of maximizing the reactor power, it is advisable to increase the nanopowder content to the limiting value (39%). This value corresponds to the filling of the pores between the micro grains of the ceramic. When melting a nanopowder in emergency situations, liquid droplets should not completely surround the ceramic micro grains.

Particular attention is paid to solving the problem of non-proliferation of materials that can be used in the creation of weapons of mass destruction. As noted, when using a mixed mononitride fuel, a self-fueling regime is achieved, i.e., the breeding ratio in the core is approximately equal to 1. On the one hand, at all stages of the fuel cycle, the need to separate isotopes of heavy nuclei is completely eliminated. The burnout of the fissile isotopes of uranium (a small amount of ^{235}U) and plutonium (^{239}Pu, ^{241}Pu) is compensated by the production of ^{239}Pu from ^{238}U. On the other hand, NPP fuel contains material for direct use. This is attractive to the terrorist kidnappers. But terrorists are unlikely to be able to reach the Moon or Mars. If they can, then these will only mean that they are representatives of states with high technologies and (most likely) possessing nuclear weapons. If necessary, they will be able to produce the required amount of materials for direct use on Earth. It is extremely difficult to steal irradiated fuel characterized by a high background radiation, especially in space.

3.3. Traditional Coolants

The eutectic alloy NaK was used as a coolant for the first space reactors SNAP-10A, "Buk" and "TOPAZ".

In fast space reactors of the second generation, it was proposed to use the eutectic alloy NaK (electric power 70 ... 150 kW) and lithium (two reactors with electric power 80 ... 150 and 160 ... 400 kW) as a coolant [38]. The eutectic NaK and ^7Li (not used in space) can be considered traditional coolants for space reactors with liquid metal cooling [39].

Alkali metals are characterized by the highest chemical activity. The activity towards oxygen increases from lithium to cesium. In the series of alkali metals from lithium to cesium (Li, Na, K, Rb, and Cs), the melting point decreases from Li to Cs, the chemical activity increases, and the hardness weakens. Of all alkali metals, only lithium interacts with water without exploding and has the highest melting point. There is practically no oxygen in the lunar and Martian atmosphere. Carbon dioxide (the main constituent of the Mars atmosphere [22]) oxidizes all alkali metals.

Table 2 shows the melting and boiling points of some liquid metal coolants with a relatively low fast neutron absorption cross section (according to [40–42]).

Table 2: Operating temperatures of some liquid metal coolants

Coolant	Melting point, ºC	Boiling point, ºC
Lithium	180.54	1339.85
Sodium	97.86	883.15
Potassium	63.7	758.85
Mercury	-36	357
Bismuth	271	1490
Lead	327.4	1740
77,2 % K - 22,8 % Na	−12.6	785
55,5 % Bi - 44,4 % Pb	123.5	1670

A NaK alloys are chemically active, low-melting, and flammable in air. With a mass fraction of potassium of 40 ... 90%, the alloy remains liquid under normal conditions [41]. An alloy of 56% K and 44% Na is also a eutectic.

3.4. Alternative Coolants. Sodium, Potassium and Cesium Alloys

Is there an alternative to lithium-7 and NaK alloy used for cooling the core of space NPPs?

In the reactors of nuclear submarines, a coolant based on a eutectic alloy of blue bream with bismuth was used [43]. (Eutectic is a liquid solution that crystallizes at the lowest temperature for the alloys of this system.) The first such reactor (BM-40A) of the nuclear submarine "Lira" was lighter than traditional reactors of the LWR type by 300 tons at the same power (155 MW) [43]. An eutectic alloy of lead and bismuth provides a high level of inherent safety in relation to accidents [44]. The operating temperature range of such a coolant is wide: from 123.5 °C to 1670 °C [40, 41]. The technology of using such an alloy as a coolant for a nuclear reactor has been developed [43]. The PbBi eutectic alloy cooled reactor is intrinsically safe. In such a reactor, severe accidents can be excluded. The use of the PbBi coolant will require the reliable operation of the heating system of the idle reactor circuit.

Gallium has a maximum operating temperature range. The freezing point is 30 °C; the boiling point is 2230 °C [42]. Gallium is classified as a rare element. It is highly dispersed in the earth's

crust (on average, its mass content in the earth's crust is about 19 ppm) [42]. Gallium is extracted from zinc and aluminum ores. Gallium is attractive as a high-temperature heat carrier. But gallium is a very expensive coolant. In addition, gallium is highly corrosive to traditional structural materials for fast reactors (AISI 316 SS, Russian stainless steel "12X18H9") [44]. Gallium has never been used as a coolant in nuclear reactors [44]. Gallium is characterized by a relatively high fast neutron absorption cross section. When using mononitride fuel or cermets based on micro grains of mononitride and uranium metal nanopowder, it is possible to ensure a good balance of neutrons in the reactor core. The fuel self-sufficiency mode is easily achievable, i.e., the fuel breeding ratio in the core is slightly more than 1. However, the void reactivity effect is positive (when bubbles are drawn into the central part of the core) even in low-power reactors (electric power of 150 MW and more).

Among alkali metals, NaKCs alloys with different concentrations of components are attracted. Sodium, potassium and cesium are highly reactive. The NaKCs alloy forms a eutectic with a sodium content of 12% (at.), Potassium 47%, and cesium 41%. The properties of such an alloy are well studied [44–46]. The melting temperature of such an alloy ("Soviet alloy") is minimal for all metals and is –78 ºC [47]. This eliminates the need for heating the coolant in the idle reactor. It is possible to exclude freezing of the coolant in emergency modes such as OVC WS (overcooling accident without scram). Alloy NaKCs with a certain content of components are characterized by a much lower chemical activity than their individual components. It was found experimentally that NaKCs alloys with a certain content of components "are capable of A NaK alloys are chemically active, low-melting, and flammable in air. With a mass fraction of potassium of 40 ... 90%, the alloy remains liquid under normal conditions [41]. An alloy of 56% K and 44% Na is also a eutectic.

"NaKCs alloys with a certain content of components are capable of oxidizing in air at 300-1000 K without intensive release of aerosols and a noticeable increase in temperature" [48]. Eutectic alloy NaKCs does not ignite in the temperature range 293 ... 973 K [45]. Thus, by changing the concentration of Na, K and Cs in the alloy, it is possible to control the properties of this alloy. The manufacturing technology of NaKCs alloys is presented in [45].

The melting point, density, electrical conductivity, electric potential, reactivity, plasticity, and flammability of the NaKCs alloy depend on the concentration of the components [45]. These differences are a consequence of differences in the interactions between the components of the alloy during its formation [45]. In addition to the eutectic, alloys 20% (at.) Na - 30% K - 50% Cs and 40% (at.) Na - 30% K - 30% Cs have been well studied. Melting point of 20% Na - 30% K - 50% Cs alloy is equal minus 37.90 ºC [45]. Melting point of 40% (at.) Na - 30% K - 30% Cs alloy is equal 0.00 ºC [45].

Ignition of liquid metal occurs due to a sharp violation of thermal equilibrium. This leads to a self-accelerating temperature rise [45]. The ability of a metal to ignite is largely determined by the characteristics of the oxide film formed on the surface [45].

When analyzing three-component systems (liquid alloys), appropriately delineated triangles are used (they are called Gibbs-Rosebohm triangles or Rosebohm triangles). On such a triangle for the NaKCs alloy, it is possible to distinguish the concentration range corresponding to the non-flammability of the alloy [48]. The boundary of the region is determined by the equation of an ellipse with the ratio of the semi axes 2.4 m center at the point corresponding to an alloy close to eutectic [48].

The center of the ellipse is determined by the content of sodium 12% (at.), potassium 44%, cesium 44% (or the mass content of 3.52% Na - 21.88% K - 74.60% Cs) [48]. The boundaries of the non-flammable range (% at.): 2 ... 24 Na - 19 ... 69 K - 19 ... 69 Cs or (% wt.) 0.5 ... 7.8 Na - 7.6 ... 43.7 K - 46.5 ... 89.2 Cs [48]. The region of non-flammability is determined by the inequality [48]:

$$(6C_{Na} - 1)^2 - 8C_K C_{Cs} + 1 \le 0, \qquad (2)$$

where C_{Na}, C_K and C_{Cs} are the atomic concentrations of sodium, potassium and cesium, respectively.

Sodium and potassium are considered the most common in the earth's crust (the mass content of sodium in the earth's crust is 2.3%; potassium is 2.1% [49]). Cesium is a rare element (the mass content in the earth's crust is $3 \cdot 10^{-4}$% [42]). The volume of world production of cesium is small: 20 tons / year [49]. Increase in cesium production is possible. It can be extracted from spent nuclear fuel (~ 10 t / year).

In [50], with reference to [51], the total yield of relatively long-lived cesium isotopes during the fission of uranium and plutonium nuclei is given (Table 3). This yield is quite large: from 6.2 to 7.5%.

Table 3: Half-life and total (cumulative) yield of some isotopes of cesium during fission of heavy nuclei

Parameter	Cesium isotope		
	^{133}Cs	^{135}Cs	^{137}Cs
Half-life, years	Stable	$2.3 \cdot 10^6$	30.0
The yield upon fission of ^{235}U nuclei by a thermal neutron, %	6.6149	6.5803	6.2444
The yield upon fission of ^{235}U nuclei by a fast neutron, %	6.6100	6.3696	5.9962
The yield upon fission of ^{239}Pu nuclei by a thermal neutron, %	6.9034	7.2411	6.5090
The yield upon fission of ^{239}Pu nuclei by a fast neutron, %	7.0466	7.5384	6.3534

The distribution of the yield w of fission fragments over the mass numbers A is determined by the presence of about magic, magic, and doubly magic nuclei. According to the shell structure (model) of the atomic nucleus, asymmetric fission is a consequence of the predominant formation of fission fragments with filled neutron and proton shells.

The w (A) curve corresponding to the fission of uranium or plutonium contains two maxima. This reflects the well-known fact that fission into two fragments with a mass ratio of 1.6 is most

likely. The first maximum corresponds to the values $A \approx \approx 90$ ... 100, the second corresponds to the values $A \approx 130$ ... 140 (Fig. 3). The yield of cesium isotopes illustrates the second (right) maximum.

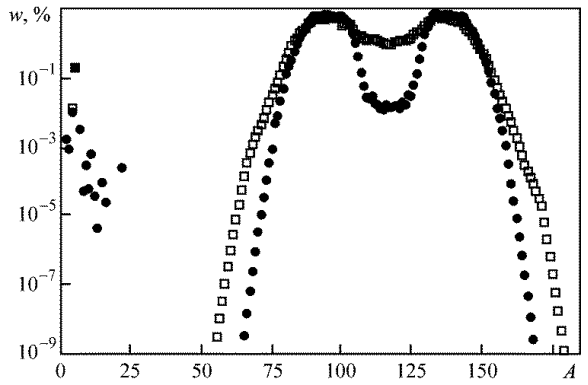

Figure 3: Distribution of the yield of fission fragments of uranium ^{235}U by atomic masses A [24]: □ - fission of nuclei by neutrons with an energy of 14.06 MeV; ● - fission of nuclei by neutrons with an energy of 0.0253 eV

Cesium is strongly activated in the core of a nuclear reactor [40, 44]). This is a serious problem that prevents the use of such alloys as coolants for ground-based power reactors. The use of cesium as a component of the NaKCs alloy, if it can be justified, is only by the non-flammability of such a coolant. However, the specific operating conditions of the reactor on the surface of Mars or the Moon (extremely low population density - the personnel of the alien station) make it possible to use radioactive waste from terrestrial nuclear power engineering as reactor materials (fuel and coolant).

Due to the wide range of operating temperatures at an extremely low freezing point and a relatively high boiling point (721 ºC [47]), the coolant based on the NaKCs eutectic is of great interest for space reactors. You can eliminate the need for heating the coolant in the idle reactor. Freezing of the coolant in OVC WS is excluded.

The use of such a coolant will make it possible to exclude the combustion of the coolant (as a result of which the void reactivity effect is realized), and the boiling problem can be solved due to the optimal choice of the reactor layout parameters. For medium to high power LMFRs, there is a problem with the positive void reactivity effect. However, in low-power reactors, this effect is negative. It can be expected that the use of such a coolant will make it possible to create a safe space reactor.

3.5. The Research Methods

The author uses calculation and optimization methods and codes to substantiate the possibility of creating a space reactor for an alien NPP. The author's codes [1] and well-known codes [52, 53] are used:

- FRISS-2D for simulation of emergency modes (including ATWS) [1];
- Multiple modernized calculation and optimization code Dragon-M [1];
- MCU code for precision neutron-physical calculation [52];
- WIMS code for neutron-physical calculation of a fuel element cell [53].

The FRISS-2D mathematical model is based on the solution of the following equations:

- Systems of equations of point neutron kinetics with a six-group description of delayed neutrons (for the relationship between power and reactivity);
- Non-stationary heat conduction equation, which is the heat balance equation, and the coolant energy equation (to determine the relationship between the reactor power and changes in the fuel and coolant temperatures);
- balance equation of reactivity (linking temperature changes with reactivity).

One of the versions of the FRISS-2D code can work as part of the Dragon-M computational and optimization code.

Dragon-M code is intended for solving mathematical programming problems in a deterministic setting and under conditions of uncertainty in the initial data. To solve optimization problems, the method of sequential linearization is used [54]. The code allows you to solve problems with a given criterion of optimality. The restrictions for the functionals characterizing the economic efficiency of the NPP, reliability and safety are considered. The emergency modes of the ATWS type of reactor operation are taken into account. Comprehensive optimization is carried out: the code contains modules for neutron-physical calculation (in the multi-group diffusion approximation), calculation of changes in the nuclide composition of fuel, thermal-hydraulic calculation, strength calculation, and assessment of the economic characteristics of NPPs. A homogeneous cylindrical reactor consisting of radial and axial zones differing in composition is considered. The control parameters include the geometric characteristics of the fuel element array, the fuel composition, the coolant flow rate, the dimensions of the reactor zones, etc.

The MCU code allows obtaining the solution of the stationary gas-kinetic equation of neutron transport (Boltzmann equation) by the Monte Carlo method.

In the calculations using the WIMS code (WIMS-D5B), the S4-approximation of the discrete ordinate method was chosen to solve the integral neutron transport equation (Peierls equation).

The author has developed the FRISS-2D, Dracon-M codes and a number of auxiliary codes. The author for the first time included in the optimization problem the constraints for functionals that simulate the safe termination of emergency processes (including ATWS). Limitations are considered for extreme temperatures of fuel, coolant, fuel-element cladding, power, pressure in the cavity for collecting gaseous fission products, etc.

As an optimality criterion, it is customary to consider the functional characterizing the economic efficiency of a NPP. The rest of the functionals can be taken into account among the limitations of the mathematical programming problem. However, when safety functionals are taken into account, the area of safe layouts in the space of control parameters is sometimes so small

that the choice of the economic criterion of optimality does not matter. The criterion related to the safety of the reactor can be selected as the target functional (for example, the void reactivity effect). The control vector includes the characteristics of the core.

The FRISS-2D and Dracon-M codes assume that the reactor consists of several radial and axial zones of homogenized composition. The Dracon-M code has no analogues yet.

With the help of the developed software, the author for the first time obtained the results of optimizing the safe layouts of ground-based and alien-based LMFRs, taking into account safe operation. In such reactors, severe accidents can be ruled out.

The software developed by the author allows solving computational and optimization problems in conditions of uncertainty, including the uncertainty of emergency scenarios. This is important at the initial stages of nuclear reactor design (when its characteristics are not finally determined).

4. Results

4.1. Initial Provisions

The relatively low chemical activity of the NaKCs alloy in comparison with its components makes it possible to consider a two-circuit energy conversion scheme. The first circuit is filled with liquid NaKCs alloy, the second - with water. The transition to a two-circuit scheme will greatly simplify the design and reduce the cost of NPPs.

The most dangerous are emergency modes, accompanied by the failure of emergency protection (ATWS). As you know, all emergency modes of LMFR operation can be triggered by the following events.

- Draining the core or part of it (including entrainment of bubbles into the core) - modes of the LOCA WS (loss of coolant without scram) type;
- Input of limited value of positive reactivity - modes of TOP WS (transient overpower without scram) type;
- Violation of the forced circulation of the coolant in the primary circuit - modes of the LOF WS (loss of flow without scram) type;
- Violation of heat sinks from the first circuit to the second - modes of the LOHS WS type (loss of heat sink without scram);
- Overcooling of the primary coolant - modes of the OVC WS type.

The most dangerous are the combinations of these modes.

It is assumed that the reactor of the Martian NPP has an electrical capacity of 600 MW. (This is an abnormally high power for such a NPP.)

The traditional layout of the core is considered. It is assumed that there are two zones of profiling along the radius of the reactor. The initial loading of the reactor assumes the presence of enriched uranium and does not contain plutonium isotopes. Plutonium builds up as the reactor runs.

Both zones contain uranium mononitride fuel of different enrichment (11 and 15% for the central and peripheral zones, respectively) with the addition of long-lived transuranic elements (5% by weight ^{237}Np, ^{241}Am and ^{243}Am). A nanopowder of metallic uranium (39% by weight) with an enrichment of 16% in ^{235}U is placed in the fuel pellet between the micro grains of the mononitride.

In order to increase the flow rate of the coolant in the natural circulation mode (when the main circulation pumps are de-energized), it is necessary to increase the effective height of the natural circulation circuit in comparison with ground-based NPPs. To ensure the required level of natural circulation up to 14 ... 15% (which is required for the trouble-free completion of the LOF WS mode), it is necessary that the height difference between the averages levels of the core and the emergency cool-down heat exchanger be 23 ... 25 m. This can be easily achieved when the reactor is buried under soil of the planet.

Ground-based fast reactors, cooled with sodium or NaK alloy, operate and operated at a pressure in the core close to the Earth's atmospheric pressure (0.1 MPa). The average atmospheric pressure at the surface of Mars is about 636 Pa [18], i.e., almost 160 times less than at the surface of the Earth. The boiling point of the coolant decreases with decreasing operating pressure. For this reason, it is necessary to provide pressure in the core at the level of the earth's atmosphere, i.e., 0.1 MPa. The result is a new type of reactor called the LMFR under pressure (PLMFR).

Reactors of the LMFR type with increased pressure in the core are being developed in Russia. An example is the lead-cooled BREST-OD-300 reactor under construction [23]. Under a thick and heavy layer of lead, the pressure in the core is about 1 MPa or 10 atm. This does not preclude the safety of the reactor.

4.2. Optimal Layout of the Martian Reactor

The calculations were performed using the Dragon-M code, which includes the FRISS-2D modules. The void reactivity effect was chosen as an optimality criterion. When the entire reactor is drained, it is negative. The WIMS code is used to calculate some of the coefficients of reactivity. The MCU code is used to refine the meaning of the void reactivity effect. (The diffusion approximation used in the Dragon-M code can lead to large errors in estimating this effect.)

It is assumed that the reactor core operates in a self-fueling mode. The breeding ratio in the core is slightly more than 1 (which is necessary to compensate for the production of "slag").

Several layouts of the Martian reactor core were obtained. Two types of fuels based on cermets were considered. The first is a mixture of micro grains MN and nanopowder U, the second is a mixture of micro grains MOX and nanopowder U. The nanopowder fills the pores between micro grains of the ceramic. In both cases, the condition BRC ≈ 1 is satisfied. This limitation takes into account the minimum reactivity margin for fuel burnup.

The void reactivity effect was considered as an optimality criterion, realized when draining the reactor.

The core contains two radial zones surrounded by a reflector. The zones differ in the value of the fuel enrichment. The initial charge is formed from enriched uranium and does not contain

plutonium. The fuel rods are located at the nodes of the triangular lattice. Perforated covers for fuel assemblies are used.

The reflector consists of two zones, assembled from assemblies containing steel tubes with long-lived radioactive waste (^{14}C, ^{99}Tc). The assemblies are washed with a coolant. The first rows of assemblies (closest to the core) contain ^{99}Tc, while the far rows contain a mixture of ^{14}C and ^{99}Tc powders. Carbon-14 slows down neutrons, which contributes to the creation of favorable conditions for the transmutation of ^{14}C and ^{99}Tc. In addition, the presence of such nuclides in the reflector allows the thickness of the reflector to be reduced. The end reflectors are technetium-99 tablets. The height of the upper and lower reflectors is 20 cm each. The lower reflector may be absent.

Two optimal versions (A and B) of the MN-U-fuel core have been proposed.

Option A corresponds to the highest inlet temperature value. The total radius of the core and side shield is limited to 2.6 m. The thicknesses of the side and end reflectors are specified and did not change during the optimization process.

For both variants the diameter of the fuel pellet is 5 mm. The fuel rods are located at the nodes of the triangular lattice.

The grating spacing is 5.75 and 6.2 mm for two zones, respectively (Option A). The height of the fuel column is 1 m. The outer radius of the core is 2.15 m. The temperature at the entrance to the core is 198 ºC. Fuel enrichment with respect to ^{235}U is 10.9 and 14.5% for the central and peripheral zones, respectively. The radius of the core is 129 cm. The radial dimensions of the two zones (with fuel of different enrichment) are 86.7 and 42.3 cm, respectively. When the reactor is operating at rated power, the maximum fuel temperature does not exceed 641 ºC, the coolant temperature does not exceed 516 ºC, and the fuel element cladding does not exceed 524 ºC. The presence of long-lived radioactive cesium-135 in the coolant (decaying into stable barium-135) and short-lived cesium-137 (decaying into metastable barium-135 and with the subsequent emission of γ-quanta) require taking into account the possibility of contamination of the coolant with barium. The melting point of barium is 727 ºC. There is no data on the formation of intermetallic compounds in the Ba-Na, Ba-K, Ba-Cs systems.

Thus, the presence of barium in the coolant can cause increased erosion of structural materials (in particular, fuel-element cladding). This is a payment for the use of "free" radioactive cesium-137 as part of the coolant. For this reason, the speed of the coolant in the core should be limited. In the considered options A and B, the average speed of the coolant does not exceed 3.2 and 3.5 m / s, respectively. (In sodium reactors, the average speed does not exceed 5 m / s.)

There are three ways to solve the problem:

- limiting the speed of the coolant in comparison with a sodium reactor;
- refusal to use cesium-137 as part of the coolant;
- use of cladding of fuel elements with tungsten sputtering (the technology is based on low-temperature plasma spraying of tungsten powder).

Options A and B are taught using the first path. In the options under consideration, the average speed of the coolant does not exceed 3.2 and 3.5 m / s, respectively. (In sodium reactors, the average speed does not exceed 5 m / s.)

Option B was obtained with the same lattice of fuel elements in two radial zones. This is a less compact core. The fuel element lattice spacing is 6.3 mm. Fuel enrichment in 235U in zones is 9.9 and 13.5%. The radius of the core is 144.5 cm. The radial dimensions of the two zones (with fuel of different enrichment) are 111.0 and 33.5 cm, respectively. The height of the fuel column is 0.92 m. The temperature at the inlet to the core is 198 ºC. When the reactor is operating at rated power, the maximum fuel temperature does not exceed 684 ºC, the coolant temperature does not exceed 502 ºC, and the fuel cladding does not exceed 527 ºC.

4.3. On the role of Doppler void and density coefficients and reactivity effects in ATWS

Doppler broadening of resonances in the energy dependence of the neutron-matter interaction cross sections plays an important role in ensuring the self-protection of reactors. This broadening is characterized by the Doppler effect and the coefficient of reactivity. If resonances are present in the energy dependence of the cross sections, then with increasing temperature they broaden (the width increases and the amplitude decreases). With a significant increase in temperature, narrow isolated resonances can overlap. If the kinetic energy of a neutron after collision (scattering) with a nucleus becomes comparable to the energy of some resonance, a nuclear reaction, in the energy dependence of which this resonance is present, is realized with a high probability (a neutron is captured by a nucleus, a compound nucleus is formed, decaying along some channel). If the kinetic energy of the neutron corresponds to the energy between resonances, where the cross section is close to zero, the neutron continues to slow down until its energy coincides with the energy of another resonance (up to the width of the level) and the neutron enters into a nuclear reaction.

If a neutron is captured by the uranium-238 nucleus, then one neutron decreases in the reactor. If a neutron is captured by a uranium-235 nucleus, then the fission of this nucleus with the production of new ν_f (as a rule, $\nu_f > 1$) neutrons is possible, and the number of neutrons in the system increases. (In the resonance region, ^{238}U nuclei practically do not fission.) Thus, broadening of resonances in the energy dependence of the cross sections for ^{238}U leads to a decrease in the number of neutrons in the system, i.e., to a decrease in the effective neutron multiplication factor and reactivity. Broadening of resonances in the energy dependence of the fission cross section for fissile fuel nuclei (^{235}U, ^{239}Pu, etc.) promotes an increase in the number of fissions and, consequently, the number of neutrons, and hence, an increase in the effective neutron multiplication factor and reactivity. The first case corresponds to a negative Doppler reactivity effect, the second to a positive one. Thus, the sign and significance of the Doppler reactivity effect are mainly determined by the nuclei on which the broadening of resonances predominates.

If the reactor uses fuel with a relatively low enrichment in the fissile nuclide, then with an increase in the fuel temperature, the effect of broadening of resonances on the raw material nuclides (^{238}U, etc.) dominates. This effect is negative. It has a beneficial effect on the safety of the reactor. In reactors with high enrichment

(some research, transport), the dominant effect is associated with the broadening of resonances on the dependence of the fission cross section for fissile nuclides, which leads to an increase in the number of neutrons and a positive Doppler reactivity effect. For this reason, it is proposed to use fuel with a relatively low enrichment for LMFR (about 16%) in our space reactor.

It should be noted that in LMFRs, the negative Doppler reactivity effect can compensate for the positive void reactivity effect, since when the core is drained, the heat removal from the fuel elements deteriorates and the fuel temperature increases. Depending on the design of the LMFR and the type of emergency, the role of the Doppler coefficient of reactivity is different. The generally desirable negative Doppler coefficient of reactivity can in some cases exacerbate an accident. This remark applies primarily to LMFRs with medium to high power oxide fuels. This is well known. In such reactors, an increase in the absolute value of the negative Doppler coefficient leads to a favorable termination of emergencies caused by an unauthorized increase in the reactor power (for example, due to the introduction of limited positive reactivity or overcooling of the coolant), and the aggravation of emergency processes initiated by a decrease in the coolant flow rate (for example, as a result of blackout main circulation pumps).

It is known that the negative Doppler coefficient of reactivity helps to minimize the deviation of power and maximum fuel temperature from the nominal values. Fig. 4 illustrates the role of the Doppler coefficient of reactivity with decreasing and increasing maximum power W and maximum fuel temperature T_f. (These parameters are considered among the functionals of the optimization problem.) The dashed line corresponds to the values of these functionals in the nominal operating mode of the reactor. The arrows indicate the trend in the change in the functionals (ΔW or ΔT_f) with an increase in modulus of the negative Doppler coefficient of reactivity. It can be seen that the negative Doppler coefficient of reactivity, which is large in absolute value, plays a favorable role. For this reason, a negative Doppler factor contributes to the safe termination of the TOP WS emergency mode (initiated by injection of limited positive reactivity). Moreover, the higher this coefficient in absolute value, the smaller the deviation of the temperature and power of the reactor in emergency mode.

The nature of the change in the maximum fuel temperature over time during the LOF WS process depends on the type of fuel used (the relative amount of light and heavy atoms). As a consequence, the nature of the change in the maximum fuel temperature depends on the temperature difference between the fuel and the coolant in the nominal mode (before the start of the emergency). If this difference is large (for example, when using oxide fuel), the following scenario is realized. As the flow rate decreases, the coolant temperature rises and the temperature difference between the fuel and the coolant decreases. Consequently, the heat flux from the fuel to the coolant decreases. Under the influence of negative feedback on reactivity due to the thermal expansion of the core, the thermal power of the reactor decreases. The power decreases with time faster than the heat flux from the fuel to the coolant, i.e. more heat is removed from the fuel than is generated in it. As a result, the maximum fuel temperature decreases over time, which leads to a positive contribution to reactivity from the Doppler reactivity effect. To neutralize the latter, a higher temperature of the coolant and structural materials of the core is required. A negative Doppler coefficient of reactivity always prevents the power from deviating from the nominal value (in this case, it prevents it from decreasing) and leads to a stronger heating of the coolant. Thus, in order to increase the self-protection of a fast reactor with oxide fuel from accidents of the LOF WS type, it is necessary to decrease the negative Doppler coefficient of reactivity in absolute value. This fact was known at the dawn of the development of LMFR technologies and was described by many authors. However, the ATWS modes were not considered.

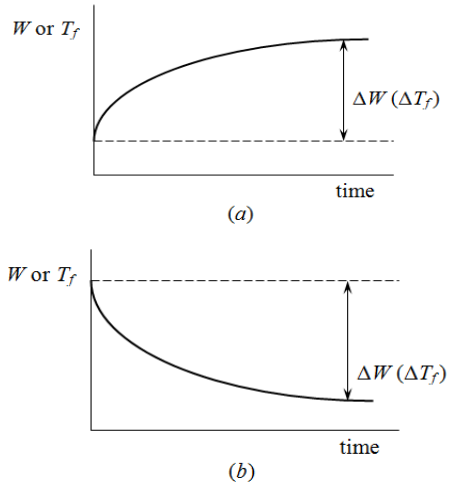

Figure 4: Influence of the Doppler coefficient of reactivity with decreasing (*a*) and increasing (*b*) functional W or T_f

When using metallic fuel, the temperature difference between the fuel and the coolant in the nominal mode is small and the temperatures of the fuel and the coolant in the LOF WS process increase. In this case, negative Doppler reactivity coefficient and the effect of thermal expansion of the core prevent an increase in temperatures, and in order to increase self-protection from accidents, it is necessary to increase the negative Doppler coefficient in absolute value, i.e., the Doppler reactivity effect plays a favorable role. This is also known.

In reactors with nitride or carbide fuel, one of these scenarios can be realized or an intermediate case, when the fuel temperature increases, reaches a maximum, and then decreases to a value below the nominal value. Different scenarios of changing the maximum fuel temperature can be realized in different grading zones.

When using the fuel considered in this work (UN-PuN-U or UN-U) with a relatively high density and thermal conductivity, the role of the Doppler reactivity coefficient in the emergency modes LOF WS and TOP WS is the same. An increase in the absolute value of the negative Doppler coefficient of reactivity contributes to an increase in the self-protection of the reactor from such accidents.

The results of the analysis of the Doppler, void and density coefficients and reactivity effects are presented in [1]. In fig. 5 shows the dependence of the Doppler reactivity effect on the deviation of the fuel temperature from the nominal. Fig. 6 illustrates the dependence of the spectral component of the density

effect on the density of the coolant. Zero density in Fig. 6 corresponds to the void reactivity effect.

Dependence *1* in Fig. 5 and 6 correspond to traditional mixed mononitride fuel (porosity 25%). Dependences *2*, *3* and *4* correspond to fuel (5% porosity) based on a mixture of micro grains UN-PuN and metallic uranium nanopowder (dependence *2* corresponds to waste uranium in the composition of a nanopowder; *3* - enriched to 20% in ^{235}U; *4* - enriched to 80% in ^{235}U). Obviously, the point of intersection of all straight lines corresponds to the reactor operating mode at rated power.

4.4. ATWS Analysis

The LOF WS emergency mode is initiated by de-energizing all main circulation pumps. For variant A, with a pump run-on time of 30 s, the maximum temperature of the heating medium reaches 694 ºC, which corresponds to a 27-degree boiling margin. For variant B it is 6 ºC higher than the boiling point. With an increase in the run-on time of the pumps to 40 s, the maximum temperature of the heat carrier is 709 ºC. In the steady-state mode of natural circulation, the maximum temperature of the heat carrier decreases by 68 and 58 ºC for options A and B, respectively. The maximum fuel temperature in the LOF WS transition mode (with a pump run-on time of 30 s) reaches 714 and 762 ºC for options A and B, respectively. The reactor power is reduced in LOF WS mode.

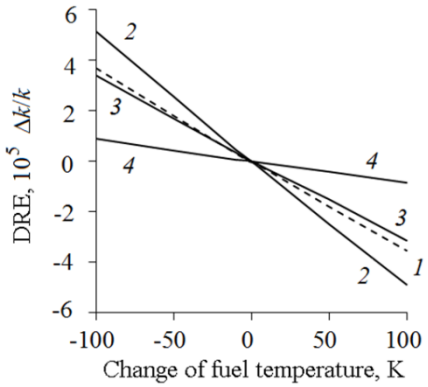

Figure 5: Dependence of the Doppler reactivity effect (DRE) on the deviation of the fuel temperature from the nominal [1]

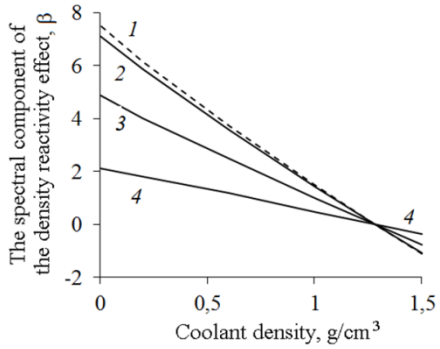

Figure 6: Dependence of the spectral component of the density and void reactivity effects on the density of the coolant [1]

The TOP WS emergency mode is initiated by the input of reactivity 0.9 β (where β is the effective fraction of delayed neutrons) for 10 s according to the linear law. This corresponds to the erroneous (unauthorized) removal from the core of both absorbing rods - reactivity compensators. The maximum fuel temperature increases in the transient mode to 777 and 789 ºC for options A and B, respectively. The maximum coolant temperature is 577 and 562 ºC for options A and B, respectively. The steady state power (under the action of reactivity feedbacks) is approximately 24% higher than the nominal power.

In OVC WS mode, the heating medium temperature decreases over time. A significant temperature margin is maintained until the coolant freezes. When the main circulation pumps are converted to increased performance (with an increase in flow rate by 20%), the maximum fuel temperature reaches 733 and 767 ºC for options A and B, respectively. The power increases by 58 and 24% (respectively) of the nominal. This is a hypothetical emergency that practically cannot be realized. When a “cold” backup loop is connected (if it is provided), the maximum fuel temperature is 746 and 787 ºC for options A and B, respectively.

LOHS WS emergency mode is initiated by de-energization (or failure) of all secondary circulation pumps. Even if a coolant with a temperature equal to the outlet temperature is supplied to the entrance to the core, the emergency mode is not dangerous. Fuel temperature decreases over time. In the LOHS WS emergency mode, the maximum temperature of the coolant is 535 ºC. The maximum fuel temperature decreases over time.

When the LOF WS and OVC WS, LOHS WS and OVC WS modes are superimposed, their mutual neutralization is observed. The most dangerous is the overlap of the LOF WS, LOHS WS and TOP WS modes and the non-simultaneous overlap of LOF WS, TOP WS and OVC WS processes (with OVC lag). In this case, boiling of the coolant is possible.

Optimal configurations of the reactor with fuel based on oxide micro grains and uranium nanopowder have been obtained. The problem statement (optimality criterion, control parameters, constraints) is the same. The same scenarios for the development of emergency modes are postulated. It is a more heat-resistant fuel. Let us consider one of such layouts with the most compact core. Fuel enrichment in two zones is 12.5 and 19.8%, respectively. The diameter of the fuel pellet in both zones is the same and equal to 5.21 mm. The spacing of the fuel element lattice in the zones is 5.84 and 6.40 mm, respectively. The radial dimensions of the zones are 86.7 and 42.3 cm, respectively. The radius of the core is 129 cm. The height of the core is 1 m. The temperature of the coolant at the entrance to the core is 198 ºC. The maximum temperatures of fuel, coolants and fuel element cladding in the nominal mode do not exceed 938, 523 and 527 ºC, respectively. The void effect of reactivity is positive and dangerous. In the most dangerous combination of emergency modes, the coolant boils. The behavior of the reactor in ATWS is similar to that of using MN fuel without uranium nanopowder.

Thus, MN-U fuels are of the greatest safety interest.

4.5. Optimization of the Composition of the NaKCs Alloy

Multi component alloys of liquid metals have the greatest potential for regulating the properties of the LMFR coolant. If the problem of optimizing the ratio of the concentrations of the components of a binary alloy (for example, NaK used in space

reactors) from the point of view of one or another quality criterion is quite trivial, then when switching to coolants based on three-component systems (for example, NaKCs), the solution of such a problem is laborious and can be associated with serious problems.

The complexity of solving such problems increases many times, if individual components of the alloy are capable of forming a eutectic, and if, at certain ratios of the concentrations of the constituent elements, a qualitative change in the chemical properties of the alloy is possible. First of all, these remarks refer to three-component alloys of alkali metals based on sodium, potassium and cesium, which, like individual components, including those forming binary eutectic alloys, can be considered as potential coolants for space reactors. Among the components of the NaKCs alloy, the eutectic is formed by the binary alloys NaK, NaCs, KCs [45, 55]. The melting point of the eutectic 21% (at.) Na - 79% Cs is equal to minus 31.70 ºC; the melting temperature of the eutectic 30% (at.) Na - 70% Cs is equal to minus 7.85 ºC [45]. The melting point of the eutectic 50% (at.) K - 50% Cs is equal to minus 37.90 ºC [45].

Optimization of the composition of the NaKCs alloy in terms of minimizing the cost of the coolant, increasing the internal self-protection against accidents LOHS WS, OVC WS and LOCA WS is of a conflict nature.

An increase in the concentration of cesium in the alloy is hindered by its high activability in the reactor core and its high cost. The problem is not relevant when using radioactive waste in the coolant of an alien reactor. An increase in the concentration of cesium in the NaKCs alloy leads to a decrease in the void reactivity effect (hence, to an increase in self-protection against a LOCA WS accident) and a deterioration in the self-protection against accidents of LOF WS and LOHS WS. Heavier elements (in comparison with sodium) slow down neutrons worse due to elastic scattering. The deceleration due to inelastic scattering is characterized by a high energy threshold. As a result, when the coolant is lost, the neutron spectrum in the reactor does not change significantly. As a result, the void reactivity effect is relatively small, and the danger of a LOCA WS emergency is less. The boiling points of the NaKCs alloy are lower than that of pure sodium, potassium (but higher than that of cesium).

The content of sodium and potassium in the alloy is higher than the content of cesium. As a result, the danger of accidents LOF WS and LOHS WS is higher. At the same time, the freezing point of the NaKCs eutectic is so low that this advantage can only be used in space. (For a Martian NPP buried in the ground, it is enough to ensure the coolant is non-freezing.) A slight change in the concentration of cesium in the NaKCs alloy will lead to an insignificant increase in the freezing point of the alloy (in comparison with the eutectic) from the point of view of ensuring self-protection from accidents with coolant cooling.

Thus, it makes sense to solve the problem of minimizing the concentration of cesium in the NaKCs alloy, provided that the main advantages of using such a coolant are retained: non-flammability and the elimination of freezing in the reactor. The first condition is determined by expression (1), the second - by the non-freezing of the alloy. The condition of non-flammability is more important when choosing a heat carrier. According to inequality (1), the alloy non-freezing is provided for a fairly wide range of cesium concentration in the alloy: from 19 to 69% (at.), Or from 46.6 to 89.2% (wt.). Non-freezing of the alloy is ensured at and near the eutectic point, i.e., with a slight deviation of the cesium concentration in the alloy from 73.8%, as well as with a small (ideally zero) cesium content in the NaKCs alloy. As noted earlier, in the absence of cesium, sodium and potassium can form eutectic alloys.

The optimization problem can be formulated as a continuous multi criteria problem with constraints for several functionals. These functionals include the maximum and minimum coolant temperatures, the maximum temperatures of the fuel and cladding of fuel elements, and the maximum power of the reactor in emergency modes of the ATWS type. The control parameters of the problem are the concentration of alloy components, the parameters of the reactor and core layout, the parameters of the lattice of the fuel elements, the flow rate of the coolant, the enrichment of the fuel, etc.

The solution to a multi criteria problem is a set of informal procedures. For example, the formation of a system of quality criteria (at the stage of setting the problem) and the choice of the most preferred solution (at the final stage of the solution) are not formalized and are subjective. To solve the problem, multiple correlation coefficients are introduced. This coefficient is equal to zero when the void reactivity effect and the cost of the coolant are selected as criteria. For the criteria related to the cost of the coolant and internal self-protection against accidents LOF WS and LOHS WS, the correlation coefficient is maximal and is equal to 1.

Let us analyze the preferred solutions, the number of which can be large. Let us choose nine solutions U_1, U_2, U_9, each of which is obtained under the condition that one or another quality criterion is preferable.

- The U_1 solution is obtained if the cost of the heat carrier is selected as the most preferable criterion: U_1 = {0% Cs, 0% K, 100% Na} (hereinafter, the mass content is indicated as a percentage);
- The U_2 solution meets the criteria of "cost of the coolant" and "internal self-protection against accidents OVC WS": U_2 = {0% Cs, 78% K, 22% Na}, i.e. eutectic Na-K;
- The U_3 solution corresponds to the minimum void reactivity effect: U_3 = {100% Cs, 0% K, 0% Na};
- The U_4 solution corresponds to internal self-protection against accidents LOF WS and LOHS WS: U_4 = {0% Cs, 0% K, 100% Na};
- The U_5 solution corresponds to ensuring the non-flammability of the coolant: a set of optimal concentrations that satisfy condition (1), i.e., the optimality area, centered at the point 3.5% Na - 21.9% K - 74.6% Cs;
- The U_6 solution corresponds to the internal self-protection against accidents OVC WS: U_6 = {73.8% Cs, 22.0% K, 4.2% Na}, i.e. eutectic;
- The U_7 solution corresponds to internal self-protection against accidents OVC WS and non-flammability: U_7 = {73.8% Cs, 22.0% K, 4.2% Na}, i.e. eutectic;

- The U_8 solution corresponds to the non-flammability of the coolant and the minimum void reactivity effect: determined by condition (1) at the maximum cesium content - 89.2% Cs;
- The U_9 solution corresponds to the low cost of the coolant, the non-flammability of the coolant and the minimum void reactivity effect: determined by condition (1) with the minimum cesium content U_9 = {46.6% Cs, 46.3% K, 7.1% Na}, i.e. eutectic.

The next task is to determine the most preferable solution from options U_1, U_2, U_9.

Optimum corresponding to the low cost of the coolant; negative void reactivity effect; the best self-protection against accidents OVC WS, LOF WS, LOHS WS, LOCA WS (loss of coolant as a result of its combustion); and the minimum cost of the coolant can be determined unambiguously only when taking into account the weight coefficients reflecting the significance of all the listed criteria. (This is not a feature of the Martian reactors. This is a feature of the materials used in the core and the secondary circuit (mainly a non-flammable coolant with a relatively high boiling point and the lowest freezing point, with a small absorption and deceleration cross section for fast neurons) and the optimal layout of the core.) In other words, the determination of optimal control requires a preliminary ranking of the above criteria and emergency situations according to the degree of significance and danger. Obviously, this optimum belongs to the range of cesium concentration at which the NaKCs alloy does not ignite, i.e. 46.6 ... 89.2% Cs. Internal self-protection against accidents such as ATWS and the negative void reactivity effect can be achieved by the optimal choice of the parameters of the layout, the fuel element array, and other design solutions. For this reason, the criterion of non-flammability, other things approximately equal, should be given preference when choosing a coolant for an alien reactor.

Using the well-known ideas and methods for solving discrete multi criteria problems, let us rank the options U_1, U_2, ..., U_9 according to the degree of significance. The set of analyzed objects of the problem (U_1, U_2, ..., U_9) is finite. Of these nine objects of the discrete optimization problem, a set of the most preferable solutions (objects) can be distinguished. This set includes options U_5, U_6, U_7, U_8 and U_9.

We will give preference to economic efficiency (in the problem under consideration, the goal is to minimize the cost of the coolant) and safety. In single power units (in the absence of massive construction of reactors on Mars), improving the economic characteristics is not so urgent. Giving preference to three criteria of the problem: non-flammability, the smallest void reactivity effect and the best self-protection against accidents LOF WS and LOHS WS, we can single out effective objects: U_5, U_6, U_7, U_8 and U_9. They are preferred. Thus, in the problem of finding the most preferred object, options U_1, U_2, U_3 and U_4 can be ignored.

Let F_1, F_2, and F_3 be dimensionless optimality criteria that determine non-flammability, the smallest PER and the best self-protection against accidents with a violation of heat removal (LOF WS and LOHS WS), respectively. Let's define them. The area of effective objects is limited by three planes A, B and C (Fig. 7), corresponding to $F_i \geq L_i$ (where i = 1, 2, 3), where L_i is the minimum allowable value of the corresponding functional.

The lines of intersection of planes A and B; B and C; A and C, are parallel to the axes [0 F_1), [0 F_3) and [0 F_2), respectively. Point 0_2 meets the conditions: $F_1 = L_1$, $F_2 = L_2$ and $F_3 = L_3$. The goal of the problem is to maximize F_i under the condition $F_i \geq L_i$ (in the case under consideration, i = 1, 2, 3).

Let's sort options U_5, U_6, U_7, U_8 and U_9 according to preference. Note that the solutions U_6 and U_7 are coincide (correspond to the eutectic). One of them (for example, U_6) can be excluded from consideration. Let's write down the remaining objects in order of preference.

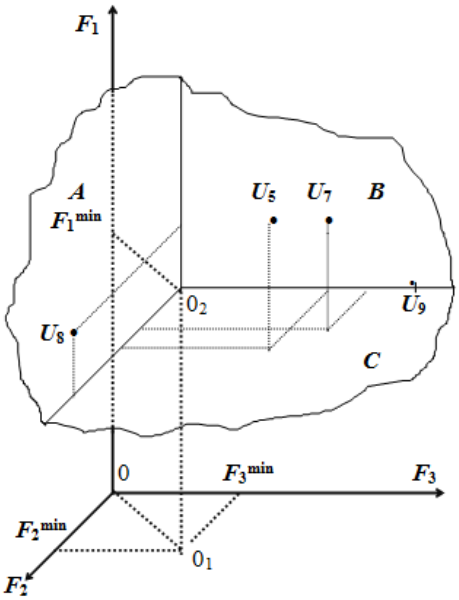

Figure 7: General view of the area of effective objects for choosing the preferred composition of the NaKCs alloy

- From the point of view of maximizing the functional F_1: $U_5 > U_6 > U_7 > U_8 > U_9$ (here the symbol ">" means "more preferable");
- From the point of view of maximizing the functional F_2: $U_8 > U_5 > U_7 > U_9$;
- From the point of view of maximizing the functional F_3: $U_9 > U_7 > U_5 > U_8$.

Object U_8 belongs to plane A (Fig. 7), i.e., for it $F_3 = L_3$. Object U_9 belongs to the F_3 axis, that is, to the planes B and C. For it, $F_1 = L_1$ and $F_2 = L_2$. Thus, object U_8 can be considered more preferable than U_9. Variants U_5 and U_7 lie inside the area of dominant objects at some distance from the border of this area. They satisfy the strict inequalities $F_i > L_i$ (i = 1, 2, 3); moreover, U_5 dominates twice, i.e., according to two criteria: when maximizing F_1, and when maximizing F_2, and the U_7 variant dominates once: when the functional F_3 is maximized.

Thus, variant U_5 can be considered the most preferable object. If we consider all alloys based on sodium, potassium, and cesium with combinations of concentrations of components from region (1) "equally non-flammable", then approaching the center of the

non-flammable region inside (1) makes no sense, in other words, it makes no sense to maximize the functional F_1 inside the region (1). In this case, it is enough to satisfy the condition $F_i \geq L_i$ and the most preferable solution should be considered U_7 (U_7 is equivalent to the object U_6), corresponding to the eutectic. Strictly speaking, the choice of the best solution from U_5 and U_7 should be carried out taking into account the coefficients of significance of the criteria F_1, F_2 and F_3.

Solution U_9 belongs to the line of intersection of planes B and C, that is, it is determined by the condition $F_3 = L_3$. A further increase in the economic efficiency of the reactor is possible with a comprehensive optimization of its parameters, which presupposes the search for the optimal control u, the components of which are not only the concentrations of sodium, potassium and cesium in the alloy, but also the parameters of the layout, the lattice of the fuel elements, the flow rate of the coolant, the enrichment of the fuel, etc.

So, the optimal composition of the NaKCs alloy can be obtained by solving the problem of multi criteria optimization. Moreover, the most preferable solutions may be different, and for alloys forming a eutectic, the optimum does not always coincide with the point of the eutectic.

5. Discussion

5.1. Further Prospects in Fuel Use

A two-circuit scheme of energy conversion with water in the second circuit was considered. Due to the relatively low chemical activity of the primary coolant (eutectic alloy NaKCs) during depressurization of the heat exchanger tubes, the interaction of NaKCs with water is not explosive. Water is the simplest and cheapest heat carrier. The technology of using water for cooling is well established. This applies to land-based reactors. The construction of a NPP on Mars will require large volumes of water. Water will have to be transported from Earth. This will increase the cost of the Martian NPP.

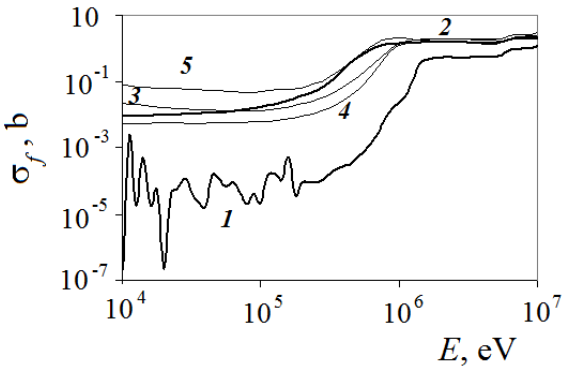

Figure 8: Dependence of the fission cross section on the kinetic energy of a neutron for ^{238}U (curve 1) and some minor actinides: ^{237}Np (*2*), ^{241}Am (*3*), ^{243}Am (*4*), ^{244}Cm (*5*)

Among the possible innovations, it is possible to propose an increase in the proportion of long-lived radioactive waste (isotopes of neptunium and americium with an even number of neutrons) in the fuel composition by reducing the proportion of waste uranium. The fission cross section for ^{237}Np, ^{241}Am and ^{243}Am in the neutron energy range of more than 1.4 MeV is 2 - 3 times higher than that of ^{238}U [51].

In fig. 8, according to [51], the dependences of the microscopic fission cross section σ_f on the kinetic energy E of the neutron for uranium-238 and some minor actinides are given.

When water is used in the second loop, the ingress of a small portion of the moderator into the core will lead to a decrease in reactivity and shutdown of the reactor. (Initiating event: emergency situation with depressurization of the tubes of the heat exchanger "liquid metal - water".)

5.2. NPP with a Closed Gas Turbine Energy Conversion Cycle

What can be considered as an alternative to water (in the second circuit)? The atmosphere of Mars is 95.32% carbon dioxide [22]. Carbon dioxide was used as a coolant for the British nuclear reactors "Magnox", AGR and the French reactor UNGG [56]. It is possible to offer a NPP with a closed gas turbine energy conversion cycle. In this case, carbon dioxide or unpurified air of Mars can serve as a coolant. A closed gas turbine cycle is proposed to be used in high-temperature gas-cooled reactors [57].

It is often proposed to use helium as a working medium for reactors of NPPs with a gas turbine energy conversion cycle. As is known, helium is the second most abundant in the Universe (after hydrogen) and makes up about 23% of the mass of the Universe. Natural helium consists of two stable isotopes with mass numbers 3 and 4. The fraction of ^{4}He in the natural mixture of isotopes is 99.99986%) and can vary considerably within wide limits [58].

Helium (consisting mainly of the doubly magical isotope ^{4}He) is not activated in the reactor core. Its practical use is hindered by its high cost and high fluidity [59]. For this reason, instead of helium, it is advisable to use expensive helium – xenon mixture [59]. Heat exchange equipment using a He-Xe mixture is the most compact [59].

In three-circuit NPPs with sodium reactors, the void reactivity effect can be realized in the following cases.

- Boiling of the coolant;
- Combustion of the coolant;
- Coolant leak (loss of coolant);
- Entrainment of gas from the argon bed into the core;
- Depressurization of fuel element cladding and release of gaseous fission products into the coolant.

In two-circuit NPPs with a gas turbine energy conversion cycle, another additional scenario for the implementation of the void reactivity effect appears: gas entrainment into the core when the gas heater tubes (liquid metal - gas heat exchanger) are depressurized.

Potential coolants of fast gas-cooled reactors (superheated water vapor, helium, carbon dioxide, dissociating gas, for example, $N_2O_4 \leftrightarrow 2\ NO_2$, helium - xenon mixtures, etc), as well as working fluids of gas turbine plants, can be considered as the working fluid (coolant of the secondary circuit).

One problem with medium to high power LMFRs is the positive void reactivity effect. The most dangerous is the draining of the central part of the core (for example, when the working fluid is involved in the event of depressurization of the heat exchanger tubes). For the considered reactor, the void reactivity effect in the most dangerous scenario of its implementation reaches 6.1 β (slightly less than when using a sodium coolant).

When the steam generator pipes are depressurized, steam bubbles (the working fluid of the second circuit) emerging from them can be entrained by the coolant flow into the core. From the point of view of possible depressurization of the steam generator pipes, the most dangerous mode is LOF WS (or its overlap with TOP WS and LOHS WS), accompanied by the greatest increase in the temperature of the primary coolant.

When the main circulation pumps of the primary circuit are de-energized, the mode of natural circulation of the coolant is established. The higher the flow rates of the coolant in this mode, the greater the likelihood of the entrainment of bubbles in the core. Thus, the normally desired high level of natural circulation can worsen a steam generator pipe leakage accident.

The mechanism of the effect of bubbles (in the core) on reactivity is as follows. If water is used in the second loop, then, entering the core, water vapor (hydrogen) effectively slows down neutrons. Hydrogen also absorbs neutrons, but this effect is relatively small at LMFR energies. Deceleration of neutrons leads to a decrease in the average neutron energy in the core. In this case, the microscopic fission cross sections for all fissile nuclides (^{235}U, ^{239}Pu, etc.) increase significantly. As a result, the effective neutron multiplication factor and reactivity increase. This could lead to an accident.

If, as a working fluid (in the second circuit), a substance is used that strongly absorbs fast neutrons and weakly slows them down, then the result is exactly the opposite. Additional absorption of neutrons in the core leads to the shutdown of the reactor: the void effect is negative.

Let us assume that a substance that weakly interacts with neutrons is used as a working fluid. In this case, the void effect of reactivity is realized in “pure form”. Any coolant one way or another absorbs and slows down neutrons. The NaKCs alloy is no exception. When entering the core in the form of bubbles, the efficiency of absorption and slowing down of neutrons deteriorates. As a result of the worst absorption in the core, additional neutrons appear. This leads to increased reactivity and the possibility of an accident. As a result of the worst moderation of neutrons in the core, the neutron energy increases. This leads to intensification of the fission of raw material nuclides (^{238}U) by neutrons with energies above 1.4 MeV (the fission threshold for most raw material nuclides). As a result, the reactivity and the risk of an accident increase.

Thus, it is necessary either to exclude the ingress of bubbles into the core, or to carefully select the materials of the reactor. It is preferable to use substances that efficiently absorb fast neutrons as a working fluid (secondary circuit). In this case, the void reactivity effect is negative.

Suppose, as a result of depressurization of the heat exchanger tubes, the material of the second circuit (working fluid) is drawn into the central part of the core (the most dangerous scenario). In fig. 9 shows the dependences of the neutron infinite multiplication factor k on the volumetric content ε of the working fluid of the second circuit in the coolant of the core. (Calculations were performed using the WIMS code [53]). As a working fluid, the following were considered: water vapor (fig. 9, curve *1*), ^{4}He (*2*), ^{3}He (*3*), CO_2 (*4*), N_2 (*5*), N_2O_4 (*6*), air of the earth's atmosphere (*7*) and Xe. Curve *8* corresponds to vacuum and is close to xenon. The dependence of k on the volume fraction f of gas when steam, He or CO_2 enters the core is non-monotonic, and it is possible to increase and decrease k, therefore, the scenario of the development of an emergency is less predictable.

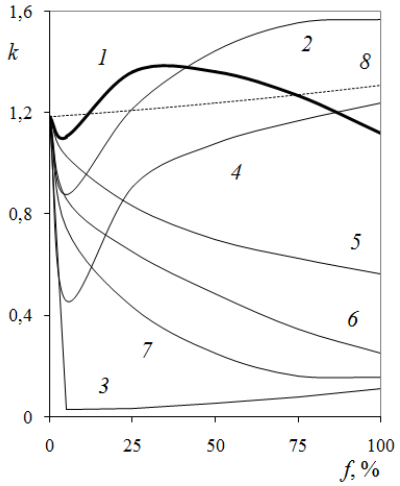

Figure 9: Dependence of k on the volume fraction f of gas in the primary coolant during depressurization of the heat exchanger tubes and the entrainment of gas bubbles into the core

As follows from the data in Fig. 9, with a small fraction ε of water vapor in the coolant, deceleration beyond the fission threshold of ^{238}U, ^{237}Np, ^{241}Am and ^{243}Am prevails (k decreases). With an increase in the vapor fraction, a slowdown towards resonances is observed in the fission cross sections of ^{235}U, ^{239}Pu and ^{241}Pu (k increases). The key role in these processes is played by elastic deceleration on hydrogen nuclei in the composition of water. With a further increase in ε, an intense absorption of neutrons slowed down to resonances in the radiation capture cross section (on these nuclei and on hydrogen) is observed. When ^{3}He (exotic), CO_2, N_2, N_2O_4 and air are used as a working medium, the deterioration of the conditions for heat removal from fuel elements can be compensated for by negative reactivity from the ingress of gas into the core.

The ingress of carbon dioxide (Martian atmospheric air) into the central part of the core and the replacement of the liquid metal coolant with gas by 7 ... 8% by volume, leads to a sharp decrease in the criticality of the reactor (k decreases by 2.4 times), i.e., to the shutdown of the reactor.

The disadvantage of a NPP with a gas turbine energy conversion cycle is its large size. However, if a reactor of relatively small (according to earthly estimates) electrical power (500 ... 600 MW) is built on the surface of another planet, this is not a serious problem.

5.3. Plans

The Russian design bureau "Arsenal" plans to implement the "Nucleon" project to build a NPP on Mars. It is assumed that the NPP will be fully built in terrestrial conditions, and not on Mars. The nuclear tug "Zeus" will deliver it to Mars [17]. The NPPs will start operating immediately after landing on Mars. Certain problems can arise here. The atmosphere of Mars is so rarefied that NPPs will be attacked by meteorites. Without special protection (containment), this can lead to damage or complete destruction of the NPP and reactor. (NASA's MRO robotic interplanetary station, which orbited 60,000 times around Mars from 2006 to 2019, "received images of an unusual trail of recent meteorite impacts on the surface of Mars, similar to the traces of shots from a giant space shotgun" [60].)

It is advisable to use the Martian NPP "Nucleon" with an electrical capacity of about 1 MW at the initial stage of Mars exploration. It can be used, for example, to generate energy for the construction of NPPs of greater capacity, buried under the ground. Larger NPPs can be assembled from blocks built on Earth and delivered to Mars by a nuclear tug. If the concept of a NPP with a gas turbine energy conversion cycle is chosen as the basis, then carbon dioxide for the second circuit can be pumped from the Martian atmosphere.

There are at least three reasons for burying the Martian NPP under the ground.

- Burying the NPP under the ground is necessary to protect against meteorites and dust storms;
- It is necessary to ensure the required height of the natural circulation circuit (to ensure the natural circulation of the primary coolant in emergency LOF WS modes);
- It is also necessary to maintain a stable ambient temperature during NPP operation.

An erosion hazard for NPPs can be posed by Martian dust storms and tornadoes (particles consisting of iron oxide) [6]. Their speed reaches 100 m / s, duration from 50 to 100 days [6].

When building a NPP on Mars, the following factors should be taken into account.

- The soil of Mars is a rocky rock (basalt) with an admixture of iron oxide (that is, it does not differ much from the earth's soil), covered with a layer of dust (consisting mainly of iron oxide);
- Mars is characterized by large daily and seasonal temperature drops;
- Mars is characterized by dust storms, meteorite attacks;
- In some regions of Mars (covering about 25% of the planet's surface) volcanic and tectonic activity is pronounced [61].

It is possible to single out specific tasks that are not typical for ground-based NPPs. One of them is the production of oxygen from the atmosphere of Mars (from carbon dioxide) [62, 63]. Oxygen is needed to support the life of researchers working on Mars and as an oxidizer for traditional rocket (non-nuclear) fuel, if used. Another challenge is to provide energy to greenhouses for growing food crops. At the University of Dartmouth, by order of NASA, a project for a Martian greenhouse was developed [64]. The greenhouse will provide plant food for four astronauts daily for 600 days on Mars. [64] Another challenge is to provide energy for vehicles and machines for laying mines and tunnels.

Economic assessments of the construction of the considered rector (electric power of 600 MW) on Mars have not yet been carried out. It is obvious that the proposed concept of the core (and the reactor as a whole) in terms of construction costs significantly exceeds a NPP with a low-power reactor (~ 10 ... 1000 kW). The diameter of the BN-600 reactor vessel is 12.8 m, the height is 12.6 m. (For comparison, the Hubble telescope is 13.3 m long and 4.3 m in diameter.) It is not yet possible to transport such a "product" to Mars. But at first glance, unsolvable problems have always stimulated and intensified scientific research in related fields. The coolant should be transported separately by elements in solid form (in the form of small briquettes). It is proposed to use long-lived cesium isotopes. This requires the use of special containers for the transport of cesium or full automation of loading and unloading a space tug. By the time a NPP of such a large capacity is in demand, a certain infrastructure and sources of electricity will already exist on Mars (for example, a low-power NPP). Robots will play an important role. The process of building a Martian NPP, filling it with coolant and other procedures can be fully robotized. It is advisable to make the reactor without a vessel. Instead of a body, a concrete shaft lined with a thin layer of steel (5 ... 6 mm) can be used. The technology of creating such a mine (for the BREST-OD-300 reactor) exists in terrestrial conditions [23]. The lining of a concrete shaft can be made with steel or with a composite material. There is a technology for obtaining such a material. It is lightweight, durable, heat-resistant, radiation-resistant material. The material is reinforced with nanofibers of another polymer composite material with a high capture cross section for neutron and γ-radiation [65].

Concrete is required for the construction of the mine. It can be produced from Martian soil. There are many types of cement. As in terrestrial conditions, the technological process for the production of Martian cement consists of three main stages: extraction and processing of raw materials (CaO, Si_2O_3, Al_2O_3, Fe_2O_3); roasting of crushed and mixed raw materials to obtain clinker; *and* grinding of clinker, gypsum and additives. Gypsum can exist as minerals on Mars. Gypsum or its analog can be made artificially from Martian rock. Cement can be produced from Martian soil. It is not necessary to use earthy materials as additives. Mars has significant and accessible deposits of water ice [66]. In 2008, the Phoenix space probe (NASA) obtained water from soil samples collected from the surface of Mars [66].

6. Conclusion

The proposed innovations associated with the use of new nuclear fuel based on UN-PuN micro grains and non-powder of metallic uranium and a coolant based on the NaKCs alloy will make it possible to create an alien NPP within the framework of existing technologies. Computational and optimization studies demonstrate the possibility of creating a safe space reactor with an electric power of up to 600 MW. A reactor of such a relatively large capacity may be in demand only in the distant future. It is important to note that if it is possible to ensure the internal safety

of a 600 MW reactor, then there will be no problems in ensuring the safety of a smaller reactor.

The following physical foundations of the concept of a Martian NPP can be noted:

- Fast reactor with electric power up to 600 MW (at the initial stage it is possible to restrict oneself to a much lower power, for example, 50 ... 100 MW);
- Low-maintenance or maintenance-free (fully autonomous) NPP, operating without fuel replacement for 10 ... 20 years;
- Coolant based on eutectic (or near eutectic) NaKCs alloy;
- Double-circuit power conversion scheme;
- As an option, a NPP with a gas turbine cycle (CO_2) energy conversion;
- Burying the NPP under the ground;
- Land-based factory fabrication of NPP modules and their transportation to Mars;
- Use of well-mastered technologies (it makes sense for the construction of a ground-based prototype of the Martian NPP).

LMFR type reactors are characterized by high safety potential. They are compact. The neutron spectrum (the absence of a neutron moderator and the fuel used) allows such a reactor to operate in the mode of self-supply of the core with fuel (the breeding ratio in the core is about 1), i.e., for a long time. Essentially, the Martian NPP will use a new type of nuclear reactor: a pressurized fast reactor (PLMFR).

The proposed new cermets fuel will allow not only to ensure accident-free completion of ATWS modes, but also to increase the content of fissile materials in the fuel elements of the core (hence, to increase the reactor power).

The power of the NPP will be determined by the range of tasks to be solved, including those unconventional for the terrestrial energy sector. For example, with the help of electricity generated by a NPP, the problem of producing oxygen from carbon dioxide will be solved. As Mars exploration progresses, the relevance of charging batteries for an increasing number of electric vehicles capable of moving across the surface of Mars will increase. There will be a need to provide power to electric machines for walking tunnels.

One should not be afraid of plutonium theft (material of direct use in weapons of mass destruction) from "fresh" fuel before the launch of the reactor. States whose representatives will visit Mars on their spacecraft (especially with a nuclear rocket engine) already possess nuclear weapons. In addition, the fuel cycle is protected from proliferation.

The coolant based on the NaKCs alloy is characterized by a much lower chemical activity compared to pure alkali metals. This allows us to consider a two-circuit energy conversion scheme (even with water in the second circuit). For Mars, a NPP with a gas turbine cycle (based on the use of carbon dioxide in the second loop) energy conversion with a low reactor power is more preferable.

The double-circuit scheme helps to simplify and reduce the cost of the NPP design. With such an energy conversion scheme, an additional scenario for the implementation of the void reactivity effect is possible. It is associated with the possible involvement of the working fluid in the reactor core. When carbon dioxide is used in the second circulation loop, the involvement of the working fluid into the core helps to reduce the reactivity and shutdown the reactor.

It is known that it is much easier to ensure the safety of a low-power reactor than a high-power reactor. The electric capacity of the 600 MW power units is unlikely to be in demand for the exploration of Mars in the near future. However, if it is possible to ensure the safety of such a reactor, then the safety of a reactor with a power of 10 times less does not pose a problem. The results are focused on the distant future.

Deepening under the ground, ground-based factory production of all modules and equipment, the use of well-mastered technologies are the general tasks of creating infrastructure facilities for the future Martian station.

It is obvious that the maximum use of the natural resources of the developed planet can significantly reduce research costs.

Obviously, for the implementation of the construction of a relatively large Martian NPP, international cooperation will be required. Before the construction of such a NPP on Mars, it is advisable to build a terrestrial prototype for testing technologies. Of course, the construction of a NPP on Mars with a reactor of electrical power of 600 MW may remain science fiction. But work on such an idea stimulates the development of new technologies that may be in demand in the future when exploring other planets.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The author is grateful to the head of the Department of Physics at the Bauman Moscow State Technical University, Corresponding Member of the Russian Academy of Sciences A.N. Morozov for comprehensive support.

References

[1] V.S. Okunev, "Analysis of Reactivity Effects and Coefficients of a Space Nuclear Reactor Cooled by a Sodium-Potassium-Cesium Alloy", AIP Conference Proceedings, **2318**, 040007, 2021, https://doi.org/10.1063/5.0036182

[2] A.S. Koroteev, "Analysis of Advanced Space Problems and the Place of Nuclear Power Propulsion Plants in their solution ", in 2005 International Conference Nuclear Power Engineering in Space-2005 (NPS-2005), Moscow, Podolsk, 2005, Plenary report 01.

[3] A.S. Koroteev, "Actual problems of modern rocket and space technology", Izvestiya RAN: Energetika, **5**, 9-18, 2004.

[4] Yu.G. Dragunov, B.A. Gabaraev, E.L. Romadova, "Space nuclear power: past, present, future", in 2018 5th Int. Scientific and Technical Conf. Innovative Design and Technologies of Nuclear Power (ISTC NIKIET-2018), Moscow, 689-705.

[5] V.N. Gushchin, "Fundamentals of Spacecraft Construction", Moscow: Mashinostroenie, 2003.

[6] M.T. Lemmon et. al., "Atmospheric Imaging Results from the Mars Exploration Rovers: Spirit and Opportunity", Science, 2004, **306**(5702), 1753-1756, DOI:10.1126/science.1104474.

[7] "The cost of the commercial launch of Soyuz with the Fregat block has become known", RIA Novosti: Science (Moscow), 02 October 2018, https://ria.ru/20181002/1529826936.html

[8] S. Kalashnikov, "The cost of a flight to Mars", Ferra.ru (Moscow), May 10, 2017, https://www.ferra.ru/news/techlife/mars-cost-10-05-2017.htm

[9] "Akkuyu NPP (Turkey)", RIA Novosti (Moscow), December 10, 2017, https://ria.ru/20171210/1510584872.html

[10] "The cost of building a fast breeder reactor BN-800 is estimated at 145.6 billion rubles (Russian)", Moscow: TASS, January 19, 2018, https://tass.ru/ural-news/2597666

[11] "The cost of building a fast reactor BREST is estimated at 100 billion rubles", Moscow: TASS, June 8, 2021, https://tass.ru/ekonomika/11594071

[12] R.F. Wilson, "SNAP 10A - A Status Report. / Book: Space Power Systems Engineering", Progress in Astronautics and Rocketry, 1966, **16**, 581-593, https://doi.org/10.1016/B978-1-4832-3056-6.50029-3

[13] N. Yachmennikova. "Roskosmos will continue to create a space tug with a nuclear engine", Rossiyskaya Gazeta RGRU, 04/29/2020, https://rg.ru/2020/04/29/roskosmos-prodolzhit-sozdanie-kosmicheskogo-buksira-s-iadernym-dvigatelem.html

[14] "Nuclear Reactor for Mars Outpost Could Be Ready to Fly by 2022", SPACE.com, 12 August 2019, https://www.youtube.com/watch?v=vQLStYhIR7Q

[15] Yu.G. Dragunov, E.L. Romadova, L.A. Sleptsov, V.V. Kudinov, "Space Reactor Installation of a New Generation",in 2012 Int. Scientific and Technical Conf. Innovative Design and Technologies of Nuclear Power (ISTC NIKIET-2012), Moscow.

[16] "NASA and US Department of Energy Selected Three Reactor Concepts for Thermal Nuclear Engines", Atominfo.Ru, 15 July 2021, http://www.atominfo.ru/newsz03/a0869.htm

[17] A. Demidov, "Roscosmos proposed to build a nuclear power plant on Mars", Gazeta.ru (Moscow), July 10, 2021, https://www.gazeta.ru/science/news/2021/07/10/n_16224692.shtml

[18] R.D. Williams, "Mars Fact Sheet", National Space Science Data Center, NASA, Sept. 1, 2004, https://nssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html.

[19] "Extreme Planet Takes Its Toll. Mars Exploration Rover Mission: Spotlight", NASA, California Institute of Technology, Jet Propulsion Lab., June 12, 2007, https://mars.nasa.gov/mer/spotlight/20070612.html.

[20] "Making a Splash on Mars", NASA Science, June 29, 2000, https://science.nasa.gov/science-news/science-at-nasa/2000/ast29jun_1m/

[21] S.A. Stern, "The Lunar atmosphere: history, status, current problems, and context", Rev. Geophys, 1999, **37** (4), 453-491, DOI: 10.1029/1999RG900005.

[22] "Mars Pathfinder - Science Results", Atmospheric and Meteorological Properties, https://mars.nasa.gov/MPF/science/atmospheric.html (дата обращения 7 ноября 2021)

[23] V.V. Lemekhov, A.V. Moiseev, M.K. Sarkulov, V.S. Smirnov, O.A. Yarmolenko, Yu.V. Lemekhov, Yu.S. Cherepnin, V.P. Vasyukhno, D.A. Afremov, "Present-Day Status and Development Prospects of Fast-Neutron Lead-Cooled Reactors", in. 2018 5th Int. Scientific and Technical Conf. Innovative Design and Technologies of Nuclear Power (ISTC NIKIET-2018), Moscow, 35-37.

[24] V.S. Okunev, "Fundamentals of Applied Nuclear Physics and an Introduction to the Physics of Nuclear Reactors", Moscow: Publishing house of Bauman Moscow State Technical University, 2015.

[25] DOE Fundamental Handbook: "Nuclear physics and Reactor Theory", DOE-HDBK-1019/1-93, U.S. Department of Energy, Washington, D.C. 20585, FSC-6910, **1**, 32, 1993.

[26] O.J. Wick (Ed), "Plutonium Handbook, A Guide to the Technology", American Nuclear Society, 1980, **2**.

[27] N.M. Vlasov, I.I. Fedik, "Fuel elements of nuclear rocket engines", Moscow: TsNIIatominform, 2001.

[28] L.H. Caveny (ed.), "Orbit-Raising and Maneuvering Propulsion: Research Status and Needs", Progress in Astronautics and Aeronautics. **89**, New York: American Institute of Aeronautics and Astronautics, 1984, https://doi.org/10.2514/4.865633

[29] I.I. Fedik, N.M. Vlasov, "New Materials in Space Nuclear Power Engineering", Perspective Materials, 2001, **6**, 24-30.

[30] L.E. Von Toth, "Transition Metal Carbides and Nitrides", New York – London: Academic Press, 1971.

[31] M.L. Taubin, "Some Radiation Effects in the Interstitial Phases of ZrC, NbC, ZrN", Atomic Energy 1990, **69** (3), 176-177.

[32] A.E. Waltar, A.B. Reynolds, "Fast Breeder Reactors", New York, Oxford, Toronto, Sydney, Paris, Frankfurt: PERGAMON PRESS, 1981.

[33] IAEA-TECDOC-1083, "Status of liquid metal cooled fast reactor technology", Vienna: IAEA, 1999.

[34] N.N. Ponomarev-Stepnoy, V.S. Rachuk, V.P. Smetannikov, I.I. Fedik, "Space nuclear power and propulsion systems based on a reactor with external heat conversion of a solid-phase core", in 2005 International Conference Nuclear Power Engineering in Space-2005 (NPS-2005), Moscow, Podolsk, 2005, Plenary report 07.

[35] A.G. Lanin and I.I. Fedik, "Selecting and using materials for a nuclear rocket engine reactor", Phys.-Usp, 2011, **54** 305, DOI: 10.3367/UFNr.0181.201103f.0319

[36] S.N.Sikorin, A.V.Kuzmin, S.G. Mandzik, S.A. Polozov, T.K. Grigorovich, Sh.T. Tukhvatulin, I.E. Galev, A.N. Bakhin, A.L. Izhutov, V.E. Alekseev, D. Kaiser, I. Bolshinsky, "Low Enriched Nuclear Fuel on the Basis Uranium Zirconium Carbonitride: Preparation to Reactor Test and Research on the Critical Assemblies", in 2018 5th Int. Scientific and Technical Conf. Innovative Design and Technologies of Nuclear Power (ISTC NIKIET-2018), Moscow, 522-532.

[37] A.N. Bakhin, I.E. Galev, D.S. Kiselev, D.M. Soldatkin, A.A. Urusov, D.A. Chesnokov, A.F. Ginevsky, Ya.S. Volgin, V.S. Volgin, "Study of Thermomechanical Strength of Uranium-Zirconium Carbonitride Fuel", in 2018 5th Int. Scientific and Technical Conf. Innovative Design and Technologies of Nuclear Power (ISTC NIKIET-2018), Moscow, 533-543.

[38] V.S. Vasilkovsky, P.V. Andreev, G.A. Zaritsky, N.N. Ponomarev-Stepnoy, G.V. Kompaniets, V.A. Usov, V.V. Vasilenko, A.Yu. Danilyuk, V.N. Zubrev, K.A. Pavlov, L.A. Rachev, V.P. Chupakhin, G.A. Shevtsov, V.V. Viter, M.V. Arakin, E.A. Ksenofontov, L.N. Tararin, B.I. Poletaev, E.G. Lyannoy, A.Yu. Pavlov, A.V. Romanov, "Problems of space energy and the role of nuclear power plants in their solution", in 2005 International Conference Nuclear Power Engineering in Space-2005 (NPS-2005), Moscow, Podolsk, 2005, Plenary report 05.

[39] A.A. Gafarov, A.B. Prishletsov, N.M. Rozhdestvensky, "Comparative analysis of transport and energy modules based on nuclear power plants with direct and dynamic energy conversion systems", in 2005 International Conference Nuclear Power Engineering in Space-2005 (NPS-2005), Moscow, Podolsk, 2005, Section 1, report 01.

[40] V.M. Borishansky, S.S. Kutateladze, I. I. Novikov, S. S. Fedynsky, "Liquid metal coolants", Moscow: Atomizdat, 1976.

[41] N.S. Akhmetov, "General and Inorganic Chemistry", Moscow: Higher School, 2001.

[42] L.M. Sulimenko, "Gallium. Popular library of chemical elements", T. 1. M .: Nauka, 1983.

[43] Yu. V. Apalkov, "Submarines of the Soviet fleet 1945-1991", Moscow: Morkniga, 2012, **3**, 229-233.

[44] V.S. Okunev, "Some Issues of Using Liquid Metals and Their Alloys for Cooling Fast Reactors", Moscow: Bauman Moscow State Technical University - Research Institute of Power Engineering, 2004.

[45] E.V. Sulim, N.G. Bogdanovich, O.V. Starkov, E.A. Kochetkova, V.E. Levchenko, "Investigation of the properties of the ternary system of alkali metals sodium-potassium-cesium in the temperature range 293-973 K" , Obninsk (Russia): Institute of Physics and Power Engineering, 2007, FEI-3087.

[46] A.G. Mozgovoy, V.V. Roshchupkin, S.N. Skovorodko, M.A. Pokrasin, A.I. Chernov, "Density and pressure of saturated vapors of liquid sodium-potassium-cesium eutectic", High Temperature, **41(**4), 560-562, 2004. http://mi.mathnet.ru/tvt1699

[47] P.I. Bystrov, D.N. Kagan, G.A. Krechetova, E.E. Spielrain, "Liquid metal coolants for heat pipes and power plants", Moscow: Nauka, 1988.

[48] O.D. Kazachkovsky, A.V. Starkov, E.A. Kochetkova, N.G. Bogdanovich, E.V. Sulim, V.E. Levchenko, N.F. Lyubchenko, V.I. Zharinov, "Some features of the alloys of the sodium-potassium-cesium system", Atomic Energy, 1992, **73** (6), 500-502.

[49] J. Emsley, "The Elements", Oxford: Clarendon Press, 1991.

[50] V.S. Okunev, **"**Nuclear Power in Space: a Look into the Future**",** AIP Conference Proceedings, **2171**, 040001, 2019.

[51] A. Koning, R. Forrest, M. Kellett, R. Mills, H. Henriksson, Y. Rugama (Ed.), "The JEFF Nuclear Data Library", OECD 2006 NEA No. 6190, Nuclear energy agency organisation for economic co-operation and development, 2011, https://inis.iaea.org/collection/NCLCollectionStore/_Public/45/026/45026295.pdf

[52] E.A. Gomin, "MCU-4 status", VANT: Physics of nuclear reactors, 2006, **1**, 6-32.

[53] "WIMSD-IAEA Library", IAEA, Nuclear Data Services, 2014, https://www.iaea.org/resources/databases/wimsd-iaea-library

[54] Geraskin N.I., Kuzmin A.M., Kashutin A.A, A.V. Kobiak, D.V. Morin, A.E. Novikov, V.S. Okunev, M.O. Shvedov, V.V. Khromov, "Complex Optimization of the Fast Reactors", in 1990 proc. Of International Conference on the Physics of Reactors (PHYSOR-90), Marseille, Fr., **4,** 1990.
[55] I.S. Grigorieva, E.Z. Meilikhova (Ed.), "Physical quantities: Handbook", Moscow: Energoatomizdat, 1991.
[56] IAEA-TECDOC-1521, "Characterization, Treatment and Conditioning of Radioactive Graphite from Decommissioning of Nuclear Reactors", Vienna: IAEA, 2006.
[57] IAEA-TECDOC-1198, "Current Status and Future Developments of Modular High Temperature Gas C"ooled Reactor Technology", Vienna: IAEA, 1999.
[58] V.G. Fastovsky, A.E. Rovinsky, Yu.V. Petrovsky, "Inert gases", Moscow: Atomizdat, 1972.
[59] E.A. Manushin, "Gas turbines problems and prospects", Moscow: Energoatomizdat, 1986.
[60] A.Bina, E. Harrington, E. Pilles, L. Tornabelle, "A recent Cluster of Impacts", 2017 (A HiRISE. Lunar & Planetary Laboratory. The University of Arizona), https://hirise.lpl.arizona.edu/ESP_047768_1995
[61] S.C. Solomon, J.W. Head, "Evolution of the Tharsis Province of Mars: The Importance of Heterogeneous Lithospheric Thickness and Volcanic Construction", J. Geophys. Res., 1982, **87**(B12), 9755-9774, DOI:10.1029/JB087iB12p09755.
[62] "A plant that extracts oxygen from carbon dioxide was created in the USA, Pasadena, USA", Federal News Agency Regnum (Russia), May 30, 2019, https://regnum.ru/news/innovatio/2638707.html
[63] A. Deryabin, "Decomposed CO2: Nizhny Novgorod residents figured out how to effectively decompose carbon dioxide", Russian newspaper RGRU - Week - Volga region, 160, 22 July 2020, https://rg.ru/2020/07/22/reg-pfo/nizhegorodcy-pridumali-kak-effektivno-razlozhit-uglekislyj-gaz.html
[64] K. Damadeo (Ed.), "NASA's 2019 BIG Idea Challenge Winner Designs Best Planetary Greenhouse", NASA, Space Tech, Apr 24, 2019, https://www.nasa.gov/feature/langley/nasa-s-2019-big-idea-challenge-winner-designs-best-planetary-greenhouse
[65] P.V. Matyukhin. Heat-resistant polymeric composites for neutron and gamma protection. International research journal. **9**(28), 39-40. DOI: 10.18454/IRJ.2227-6017.
[66] Phoenix managed to get water from Martian soil. Lenta.ru. August 1, 2008. https://lenta.ru/news/2008/08/01/water/

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 47-52 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Ethical Implications and Challenges in using Social Media: A Comprehensive Study

Mohammad M. Qabajeh*

Department of Computer Science, Faculty of Applied Science, Palestine Technical University Kadoorie, 1524200, Palestine

ARTICLE INFO

Article history:
Received: 05 November, 2021
Accepted: 12 January, 2022
Online: 25 January, 2022

Keywords:
Social media
Ethics
Privacy
Society
Digitization

ABSTRACT

The technological revolution penetrates every aspect of our dailylives; it changed our lives in different fields such as communication, decision-making, information access and work environment. These changes have benefits, but they also have costs. These costs include many ethical and social problems that need more investigation. The human element is the most affected part of this technological revolution; it affects our sense of privacy, our concepts of ownership, our consumption patterns and human relationships. A comprehensive review of various ethical concerns of using social media and how they affect our lives has been outlined in this paper. This study gives some recommendations and advices that help the society and reduce the heinous crimes that take place over social media.

1. Introduction

Recently, the emerging technology breaks through in different fields including Internet of Things (IoT), robotics, artificial intelligence, block chain, big data, cloud computing , autonomous vehicles, 3-D printing, nanotechnology, biotechnology and quantum computing. Mobile technology has grown rapidly over the past few years and the capacities of cellular networks are developed from 2G to 3G, 4G and 5G. This increase in network capacity is engaged with unprecedented processing power and huge storage capacity of the mobile devices. This leads to high increase on number of mobile phone users to reach 5.22 billion users in October 2020 [1], which is around 70 percent of the world population.

The Internet service gains increasing popularity nowadays by all generations as a forum for communication and self-expression in the society. The number of people using the Internet around the world has grown to 4.66 billion in October 2020 [2]. The use of Internet varies from browsing the web, downloading music and watching movies, buying and selling goods and products and playing games. The Internet is also a source for social networking, which is recently becoming increasingly popular [3].

Social media platforms such as Facebook and LinkedIn use websites and applications to communicate informally with others, find people, connecting with others through groups, developing social and professional contacts, promote business and share similar interests. Moreover, some platforms such as Twitter and Tumblr are used to for posting very short entries or updates on a social networking site, share and create hashtags and send direct messages. Other platforms allow users to publish and share photos with others either publicly or privately such as Instagram, Snapchat, Pinterest, and Flickr [4].

Social media is easily accessible by people from all ages. In a recent study [1], the number of people around the world that use social media each month is more than 4 billion, and every day an average of nearly 2 million new users are joining these platforms. People around the world are spending more time on social media too; the typical users roughly spending 15 percent of their life using social media platforms, exploring the social networking websites and online chat rooms, downloading photos, and playing games [1], [5].

The lockdowns during COVID-19 pandemic affects the habits and behaviors of most of people in different ways. It is clear that the time spent in browsing the Internet has been significantly increased. Also, the time spent on watching movies and streaming service such as Netflix and Disney+ increased compared with the time spent on watching broadcast and cable channels, since the Netflix gets 16 million new sign-ups during COVID-19 pandemic [1], [6]. In addition, YouTube videos reach four billion video views a day and it is considered as the second largest search engine after Google. At the same time, the use of social media increased to interact with family and friends and helped in boosting moods and sense of wellbeing during the lockdowns.

*Corresponding Author: Mohammad M. Qabajeh, mohammad.qabajeh@ptuk.edu.ps

It is a truth that social media become one of the megatrends affecting our everyday life; it gives us both opportunities and challenges. However, there are also significant risks that come with it, especially the ethical issues. Organizations, governments and individuals gather information through several means including search engines, invisible data gathering mechanisms, and marketing platforms. The information that is collected from several sources can be used to build complete profile for each user. Since available information on social media are globally accessible, some individuals and organizations can use this information for several purposes such as political issues, election and marketing. Hence, this available data can be increasingly utilized by most companies in the future [7, 8].

Many researches have examined the ethical issues of social media including [1, 6, 7]. These studies were not covering all the ethical issues and challenges arised recently. In addition to the recent effects of COVID-19 pandemic on using social media, which are not covered in the recent research articles.

Most people have their own ethics and morals and they have their own beliefs as to what is right and what is wrong. However, the ethical concerns of using social media is increased [9]. In this study, section 2 investigates the unethical practices on social media platforms and how they affects individual physical and mental wellbeing. We also discussed the challenges that face the users in many ways including personal, social, behavioral and ethical challenges. Section 3 provides some discussions and recommendations. In section 4 some concluding remarks are given.

2. Literature Review

Many published research papers dealing with the ethical issues in digitization and information technology, but they did not cover all the issues raised when the individuals use new applications and platforms. Also, they do not give recommendations and advices for the users of such platforms specially for users with poor background in information technology. This paper highlights a wide range of the unethical issues that rise with the increased usage of the social networks and digital technology. In addition, we provide technical recommendations to prevent or reduce the victims of this unethical activities in the future. Here, we will give an overview about the similar efforts from the research community that have relevance to this paper.

In [10], the social and ethical issues that arise when using technology is discussed in different fields including Internet of Things, robotics, biometrics, persuasive technology, virtual & augmented reality, and digital platforms. The paper analized different ethical concerns including privacy, security, autonomy, justice and human dignity.

Authors in [11] discussed the ethical implications of social media. This review shows the potential risks associated with using social media by psychotherapists. They also highlight that the absence of guidance and better understanding of using social media by psychotherapists can lead to inadvertent self-disclosures to clients and can place both the client and clinician at risk.

Authors in [12] showed the influence of social media on business and the effects on the sales process. They discussed how the social media has directly affect sales and marketing and highlights the challenges of using such technology.

Different types of unethical practices on social media are studied and discussed in [3]. This paper identifies the physical and mental effects of using such platforms. These unethical activates include depression, suicidal thoughts, stress, family reputation and disagreement. The authors highlighted the role of the community, government, and family in reducing the overcoming unethical activities resulted when using social media.

In [13], a discussion about the spread of using social media among people with mental disease is presented. They consider the impact of social media on mental wellbeing, they also study the risks, possible harms, and needed safety precautions with using social media for mental health. Many of the people with mental diseaseS use social media to share their experience with mental illness, requesting support from others, and to search for information about treatment recommendations, accessing mental health services and coping with the mental illness symptoms

3. Ethical Issues Of Social Media Platforms

Social media is now becoming more and more popular since it facilitates the sharing of ideas, information, business interests and other forms of expression and communication via virtual communities and networks. The backbone of social media is the activities of the users through sharing contents, posting of text and comments and posting digital photos or videos. This data may be sold to organizations for marketing purposes. Using social media in unethical way resulted in a violation of individual privacy and affected both information and physical security. In this section, we will discuss some of the challenging issue that may result from the unethical use of social media.

3.1. Privacy and Confidentiality

The rapid development of technology changes the way that we live and work. Such technologies include global positioning system (GPS) devices, smartphones, digital cameras, databases, bigdata handling and greater Internet access. This technology can document every aspect of our lives and we do not know what people knows about us and how they can use that information. In October 2020, the number of users of Internet reach 4.66 billion and the users of social media reach 4.14 billion [1]. Unfortunately, not all of them adhere by the rules of the community and behave in ethical way.

Privacy is defined as to be free from intrusion, to be free from secret surveillance, or unwanted disclosure of personal data or information by government, organizations, or individual [9]. To protect our privacy, we have to understand first what type of information can be collected and what the risks is and problems may rise when using this data. For professional IT users, they understand the problems and they can plan to protect their privacy. While, the ordinary users do not know how to protect their privacy. The privacy threats come in several categories:

- The intentional use of our personal information from the government or from private sector (in the government sector primarily for law enforcement and tax collection, and in the private sector primarily for marketing and decision making).

- Unauthorized use or release by "insiders," the people who maintain the information.
- Theft of information.
- Inadvertent leakage of information through negligence or carelessness.
- Our own actions (sometimes intentional trade-offs and sometimes when we are unaware of the risks).

To protect our privacy from unethical and crazy social media and Internet users, we have to protect things that we believe that they should be prevented from being accessed online and not being shared with others. Some private information includes our financial data, our photos and personal information, addressing information, family issues, religious thoughts and beliefs, our political views, and our travel plans.

3.2. Cyberbullying

Cyberbullying means using technology to harass, threatening, torment or humiliation by a person or group upon another person (the victim) via using Internet or cell phone [10]. The accessibility and the amount of time that people spend on the Internet and social media has increased significantly. This led to new opportunities for online aggression due to several reasons including problematic use, talk to strangers, competition for status and esteem, dealing with fake social media accounts, sexting by sending sexually suggestive or explicit text or photos to boyfriends or girlfriends. Based on [3, 11, 12], Facebook and Twitter are the two highest platforms for cyberbullying.

There are several forms of cyberbullying, including and not limited to the following:

- Impersonate the victim by sending messages to threaten him and sending inappropriate messages
- Stealing the password of the victim and updating his profile to include inappropriate material such as sexual, racist, homophobic, or posting inappropriate material that attracts the attention of undesirable people.
- Using the social media pages of the victims to post personal or false information related to the victim.
- Creating social networking account or building a web site to threaten or humiliate the victim.
- Taking personal photos or inappropriate videos of the victim and posting them online or sending them via cell phone.
- Using interactive games to send unethical messages while playing and pass these messages between the players.

There are different forms of cyberbullying, which makes it difficult to identify. This leads to a situation that children's, youngsters and females be the most targeted victims of cyberbullying. Unfortunately, few of them tell their parents and relatives when they became victims of cyberbullying. All the victims should be educated and be aware of the potential effects of cyberbullying in order to prevent their privacy violation and stop cyberbullying from happening and worsening. In addition, they need to know the consequences of their actions during using social media .

Cyberbullying has direct harmful consequences on health such as emotional distress, mental and behavioral health issues, embarrassment, isolation from other community members, anxiety, and depression and in some cases it can lead to an increased risk of suicide.

3.3. Anonymity

Nowadays, social networking supports exchange of news, ideas, opinions, discussions, rumors, and other information with large number of followers and users. However, users of such networks must often take ethical decisions about the efficient use of this amazing freedom and power.

Anonymity is defined as the expression of opinions and ideas by persons who do not show their identity [3]. In any democratic society, the freedom of speech and expressing your opinion without fear of punishment and reprisal is essential right. However, communication without showing the identity may be exploit to commit a crime or illegal and unethical activities. Hence, it is unethical for someone not to present himself or represent himself with wrong affiliations, credentials or expertise, it is unethical to become anonymous ad showing himself to be someone different way .

In cyberspace, users use anonymity and confidentiality for harassment, extortion and fraud, to distribute pornography for children's, to threaten victims, to steal business documents or other personal information, and to infringe copyrights. Anonymity can seriously damage personal relations, damage business and incite violence. In addition, anonymity makes it difficult to track wrongdoers.

3.4. Defamation

Recently, the rapid increase of social networks and content gathering websites increased the risk of slanderous and reaching of false statements to broad audience. Defamation involves writing or saying something about someone, group of people, or small company that injures or damages their reputation. Defamation comes in the form of libel, or written defamation, and slander, or verbal defamation. Our Facebook status update or tweet may receive dozens of 'likes', comments and views by tens of thousands of users. So, it's important that the potential victims, users and sharers fully understand the law of defamation and impact of online defamation [13].

It's very easy for users to re-publish posts by sharing them and re-tweeting materials in a short period of time and re-broadcast it over social media. The users may fall in the trap when using the social media, they feel that they are free to express anything they want and after that find themselves liable for defamatory publications. Therefore, it is easy that a user may become a defendant and find himself faced with claims of defamation because of this posts or comments that have different meanings in different cultures for different people. Therefore, when users creates, reposts, shares, or forwards defamatory content, they have to know that they will be strictly liable for these publications. Therefore, they need be careful and review the content they are publishing since it may contain defamatory imputations concerning the aggrieved [14].

3.5. Cyberstalking

Nowadays, many social media users freely share their feelings and desires, post personal information, share location information and publish family photos. However, if they are not aware of the danger of social networking, this can lead to numerous undesirable consequences such as cyberstalking. Cyberstalking is defined as harassment or stalking that takes place over social media or the Internet in general. The victims might be individuals, groups, or even organizations and can take different forms including threats, frighten, defamation and slander. The motivation behind this action may be to control or threaten the victim or to gather information that may be used in other crimes such as identity theft or offline stalking. This kind of crimes are naturally planned and continued over a period of time [15].

When you feel that you are a victim of cyberstalking, it is important to gather as much evidence as possible and immediately make a report and document the harassment as it goes on. Also, you should block any person who you wish to stop hearing from and inform the involved platform about their abusive behavior to technically protect you from cyberstalking. Additionally, you should contact the local authorities if you feel that you need protection.

3.6. Violence and Extortion

Violence on the social media platforms is a widespread and continuous issue due to the large number of users of it and the divergence of the user's ideas and orientations because there is a wrong understanding of the freedom that exists in the Internet space. So, we find some users resorting to pseudonyms and concealing their personalities, and it is no longer surprising that we read insulting and insulting expressions in the dirtiest terms, whether by users' peers or celebrities, direct threats, or publishing offensive images using photo and video editing programs.

Electronic violence is defined as violence against the other, whether through electronic extortion, photography, verbal threat, or writing, and it targets all groups, especially women, whether through photography or recording conversations, messages, or offensive clips. The forms of electronic violence specially against women comes in several forms. In cases of extortion, the Snapchat application is considered a window for women's freedom and a door to violence after the emergence of some applications that provide downloading video clips and images from this application, which made it easier for the extortionist to extort women because of the clips they publish and a video of her without knowing who the perpetrator is when he threatens her or publishes her in case he does not agree to his requests.

3.7. Sexual Harassment

Sexual Harassment is the use of social media sites to send sexual material that harassment the recipient, this material may be messages to identify the recipient for sexual purposes, messages that contain sexual expressions, sharing of lewd jokes, sending pictures or video scenes, or sending threats and blackmail using pictures, or using pictures without the consent of the owner or without his knowledge, and sharing it through the means of communication .

There are various factors and motives behind the phenomenon of electronic harassment. These factors include the reducing of family control, looking at a woman as a body, ease of anonymity, desire for revenge with an inability to cope, sexual deprivation, incorrect childhood and lack of awareness, culture shock and constant irritation across the media.

Many of the victims of sexual harassment are cut off from their surroundings and escape from the society. The major complication of the phenomenon is the psychological effects that may extend for years. In addition, the victims may become more likely than others to suffer from anxiety, depression, phobias and panic.

4. Summary and Discussion

All technologies have important benefits and face several challenges. These challenges make using them unpleasant especially using social media. Several challenges seems to be trivial but they may have a serious effect on the mental and physical capabilities of the users. From the discussions presented in the previous sections, many points are concluded. These points are summarized as follows:

- The social media platforms including Twitter, Facebook, YouTube, Google, Instagram and Pinterest are not free tools and treat their users as products. They are designed to collect and utilize our activities an sell to others. Therefore, it is important to remember that if you do not pay for a product, you will be the product.
- Educating users specially users with limited background about information technology and the risks, dangers and threats of posting of sensitive information on social networks, and teaching them how to protect their privacy and that they have not to share their sensitive information such as contact number or detailed address, because this data can be dangerous and can harm them.
- With the increased storage media devices and intelligent databases, evert details of our lives are recorded and can last forever. Moreover, the websites and search engines record everything we visit and search. They can sell it to anyone who pay for them, and so users must protect their data especially adult users.
- Relatively few people do not carry CellPhones. Most people use CellPhones rather than laptops and PCs and they are mostly connected to the Internet. Billons of applications are downloaded to these devices without previous knowledge about the purpose and identity of the developers of such applications. These Apps can access more information than we can imagine.
- The variety of anticipated applications with the availability of tools such as sensors, call followers, motion detection, tiny cameras, data detectors and invisible microphones they can access and generate valuable information about everything

about us. This data can craved by politicians, advertisers, companies, organizations and by individuals.

- Companies such as Apple and Google have tools called "kill switches" which can remotely enable and disable any application on our CellPhones and can delete all our data from our devices. Can you imagine what will happen if any outsider can access to this "kill switches" and operate them on our devices?
- Using social media for long time will reduce the time spent on doing physical activities (sport, reading books), change our habits and culture (respect for rules/ beliefs), and change our behavioral and mental activities (life management, interaction/ communication). Based on recent reports, the time per day spent using Internet exceeds 9 hours in some countries (Philippines, South Africa, Brazil, Colombia, Thailand), which is a troubling number. While in Japan they only connect for an average of 4 hours and 22 minutes per day.
- During COVID-19 pandemic and lockdowns, more and more people join the Internet and spent more time on social media. This will generate a new wave of digital innovations and resulting in new devices, platforms and services.
- The increased number of platforms such as Facebook, TikTok, Twitter, Snapchat, Youtupe, Instagram, WhatsApp and a lot other platforms make it difficult to secure the accounts and keep our information safe. The platforms is changing faster than we know how to understand them.

5. Recommendations

To minimize the challenges related to using social media some advices and recommendations can be taken.

- Using the technological approaches for protection user accounts such as encryption and filters to protect them from exploring sites that have objectionable material or pornography.
- Relationships with others on the Internet may be developed based on mutual interests without caring or knowing information about age, gender, nationality or race. Some of those others might not be a human at all. There are Artificial Intelligence programs that can simulate a human being in social media.
- Information on social media might be inaccurate and misleading, so do not take your decisions on what you read or hear on social media.
- The governments must put strict laws, heavy penalties and fines on people who misuse the social media and cyberspace. This action will reduce the heinous crimes that take place over the social media.
- The civil society organizations and governments must develop and build systematic culture to save the youth and the upcoming generations from the disorders and maladies of social media platforms by promoting cultural education.
- We have to form a balance view of the value and impact of using technology that can be difficult. Some of the values are obvious. Some of the impacts might be dramatic and frightening on people lives.
- Educating users about the ethical and legal issues related to using social media. Grasping these issues may minimize or prevent the negative incidents associated to social media. Users need to know that they are responsible for everything they are linked to and they have to think twice before they act on social media.
- Encourage users to use the social media in ethical way, treat others as you would like to be treated.

6. Conclusion

Widespread of using social media make it essential part of human's life everywhere. Younger generations and older generations are quickly becoming members of this technology. It enable its users to obtain and share information all over the world in a very short time. Unfortunately, the consequences of unethical use of social media are dangerous and harmful. There are many ethical issues in using social media that must be taken into deep consideration. The main objective of this paper is to determine what is the ethical issues arise in using social media platforms and how can we minimize the effect of the unethical use social media.

Most population of social media are unaware and not educated of the procedures that are needed to minimize the level of vulnerability of their sensitive data. Therefore, they should be educated about the risk of participating in social media and to use them in a responsible way. Also, the governments should have ethical and moral principles that can be applied for all users and lead to pleasant relationships regardless of nationality, race, culture, religious and political persuasion and social status.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The author wish to thank Palestine Technical University Kadoorie, Palestine for their cooperation and support to publish this research.

References

[1] S. KEMP, "Global Digital Overview, Digital October," Singapore 2020, Available: https://datareportal.com/reports/digital-2020-october-global-statshot.

[2] W. E. Forum, "Fourth Industrial Revolution , Strategic Intelligence," World Economic Forum, https://intelligence.weforum.org/topics/a1Gb0000001RIhBEAW?tab=publications.

[3] K. Chandra, S. Gounder, R Verma, S. S. Mudliar, "The Unethical Practices on Social Media," IOSR Journal Of Humanities And Social Science (IOSR-JHSS), **22**(7), 46-54, 2017, doi: 10.9790/0837-2207064654.

[4] C. Véliz, "Privacy and digital ethics after the pandemic," Nature Electronics, **4**(1), 10-11, 2021, doi: 10.1038/s41928-020-00536-y.

[5] J. Nicholas, S. Onie, M. E. Larsen, "Ethics and Privacy in Social Media Research for Mental Health," Current Psychiatry Reports, **22**(12). 84-112, 2020, doi:10.1007/s11920-020-01205-9.

[6] Z. Thomas, "Netflix gets 16 million new sign-ups thanks to lockdown," in BBC News, 2020, https://www.bbc.com/news/business-52376022

[7] V. Kumar, P. Nanda, "Social Media to Social Media Analytics: Ethical Challenges," International Journal of Technoethics (IJT), **10**(2), 57-70, 2019, doi:10.4018/IJT.2019070104.

[8] L. Royakkers, J. Timmer, L. Kool, R. van Est, "Societal and ethical issues of digitization," **20**(2), 127-142, 2018, doi: 10.1145/3437075.3437093.
[9] N. Barrett-Maitland, J. Lynch "Social Media, Ethics and the Privacy Paradox," in Security and Privacy From a Legal, Ethical, and Technical Perspective, **70**(6), 2-12, 2020, doi: 10.5772/intechopen.90906.
[10] S. Baase, A Gift of Fire, New Jersey, Prentice Hall, 2012, doi: 10.1145/572260.572271.
[11] J. M. Kizza, Ethical and Social Issues in the Information Age, 6 ed. Springer International Publishing, 2017, doi: 10.1007/978-1-4757-2950-4.
[12] M. Khalid;, S. Singh;, K. Singh;, J. Jeevitha, G. P. Anand, "Medicus: A Doctor Appointment Booking System,", International Journal of Computing and Technology, **5**(4), 48-52, 2018.
[13] V. R. Bhargava, M. Velasquez, "Ethics of the attention economy: The problem of social media addiction," **31**(3), 321-359, 2021, doi: 10.1017/beq.2020.32.
[14] C. Lewis, "Social Media - Cyber trap door to defamation," Masaryk University Journal of Law and Technology, **9**(1), 65-94 2015, doi: 10.5817/MUJLT2015-1-5.
[15] M. Pietkiewicz, M. Treder, "Cyberstalking in social media – Polish view," Journal of Modern Science, **38**(3) 29-40, 2018, doi: 10.13166/jms/99217.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 53-59 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Synthesis and Characterization of CN Thin Films Produced by DC-Pulsed Sputtering in an CH_3CH_2OH-N_2 Atmosphere

Marcos Crescencio González Domínguez[*,1], Pedro Guillermo Reyes Romero[1], Aarón Gómez Díaz[1], Horacio Martínez Valencia[2], Víctor Hugo Castrejón Sanchez[3]

[1]*Universidad Autónoma del Estado de México, Laboratorio de Física Avanzada, Facultad de Ciencias, Toluca, 50000, México*

[2]*Universidad Nacional Autónoma de México, Laboratorio de Espectroscopía, Instituto de Ciencias Fisicas, Cuernavaca, 48, México*

[3]*Tecnologico de Estudios Superiores de Jocotitlán, Jocotitlán, México.*

ARTICLE INFO

Article history:
Received: 30 November, 2021
Accepted: 31 January, 2022
Online: 10 February, 2022

Keywords:
CN_x
Sputtering
Raman
SEM
AFM
OES

ABSTRACT

This paper reports on the successful deposition of amorphous carbon nitride thin films (a-CN_x). Samples were deposited using a DC-pulsed sputtering technique in CH_3CH_2OH-N_2 atmosphere at constant pressure (1 Torr). Raman spectra were collected to investigate the bonding structure of the deposed thin films, while scanning electron microscopy (SEM) and atomic force microscopy (AFM) were used to study thin-film surface morphology. Plasma characterization was performed during the deposition process using optical emission spectroscopy (OES), and the influence of the deposition process parameters on the chemical fragmentation of species present in the plasma was determined. Raman results were typical for a-CN_x films, and SEM analysis showed that carbon clusters deposited onto the Cu substrate form a non-homogeneous surface. The morphology observed by AFM indicated that the thin films grow as islands, which corroborate the generation of amorphous structures grown on the Cu surface. OES spectra verified the existence of CN radicals within the CH_3CH_2OH/N_2 plasma during thin-film deposition.

1. Introduction

The miniaturization demands in the electronic industry required development of new materials that allow the creation of these miniature devices. A good candidate material is carbon, which has a wide variety of allotropic forms with interesting physical and chemical properties. Although the properties of diamond and graphite have been extensively investigated, there are other allotropic forms (fullerenes) that are of great interest today. Since its discovery, carbon nitride (CN) has aroused great interest in theoretical [1] and experimental studies, given its physical and chemical properties. Carbon nitrides have attracted attention owing to their potential applications in anti-corrosion photo-electrochemical technology [2], photo-degradation [3], and photo-catalysis [4, 5], which can be used in optoelectronics [6] and the manufacture of biosensors [7, 8]. From an experimental point of view, the most important material successes have been in the synthesis of carbon nitrides in the form of thin films by chemical and physical deposition of vapor. Among the physical deposition methods, the use of plasmas [9], especially the sputtering method [10], has been successfully implemented for deposition of carbon nitrides. Recently, a hybrid coating process that combines RF and DC sputtering systems has been successfully used: DC magnetron or DC-pulsed discharge [11, 12]. The advantages of these hybrid coating processes are that the various deposition parameters (e.g., working pressure, sputtering power, current, substrate bias voltage, and target material) can be adjusted separately to produce films with excellent mechanical and tribological properties.

The physical properties of carbon-based films depend upon different forms of hybrid bonds of the carbon atom: sp^1, sp^2, and sp^3. The incorporation of nitrogen into the system enables the formation of different kinds of carbon configurations, including sp^3 (tetrahedral bond), sp^2 (triple bond) and C=N and N=C-type chains. These configurations improve the mechanical and physical properties of CN_x compounds, increasing the application range of the films [13]. To the best of our knowledge, there are no publications in the literature reporting attempts to synthesize C-N compounds by CH_3CH_2OH-N_2 DC-pulsed sputtering. Even more

[*]Corresponding Author: Marcos Crescencio González Domínguez, kenny.sxe4@gmail.com

significantly, there still exist various unanswered queries about how the growth process impacts the resulting structure in amorphous CN (a-CN_x) films. The purpose of this article is to investigate the morphological and microstructural properties of a-CN_x films deposited by DC-pulsed sputtering in a CH_3CH_2OH-N_2 plasma mixture. Characterization of the bond structure of the thin films was performed by Raman spectroscopy, and their morphology and microstructure were investigated using scanning electron microscopy (SEM) and atomic force microscopy (AFM). Plasma characterization during thin-film deposition was also carried out using optical emission spectroscopy (OES), and these findings are correlated to variations in microstructure and morphology of the thin films.

2. Materials and Method

Reactive sputtering is a suitable coating technique for the preparation of thin films. The pulsed DC sputtering deposition system used in this work (Figure 1) consists of a cylindrical stainless-steel chamber (33-cm long, 20-cm radius, and a volume of 1.04×10^3 cm^3) and two mobile Cu electrodes (5-cm in diameters) placed in the center of the vacuum chamber with a separation space of 20 mm. The power supply consisted of Pinnacle Plus+ generators, which delivered DC power in a pulsing configuration to enable reactive sputtering. Pirani pressure sensors (MKS, model 103170027SH) were used to determine the total pressure in the chamber, and a flowmeter (Matheson Tri. Gas FM-1050) was used to regulate the gas flow within the system. An ocean optics spectrometer monitored the optical emission spectra generated within the discharge and the chamber vacuum was formed using a mechanical pump (Varian DS302). The background and working pressure of the vacuum chamber were 2×10^{-2} and 1 Torr, respectively.

High-purity CH_3CH_2OH (99.9%) was used as the carbon precursor, and high-purity nitrogen (99.9%) was used both as the sputtering ion source, which bombards the target, and as the nitrogen species, which is introduced into the thin films with carbon ions. The relation of the mixture was 20%-CH_3CH_2OH to 80%-N_2. To generate plasma, the vacuum chamber is first evacuated with a mechanical pump until a base pressure of 2×10^{-2} Torr is reached; next, it is filled with CH_3CH_2OH-N_2 mixture gas and discharge is generated by applying a pulsed current (300 mA at 450 V) between two electrodes. Then, the CH_3CH_2OH-N_2 mixture gas is ionized and accelerated by the electrical field produced by the electrodes, forming an energetic ion beam that provides both sputtering ions and reacting nitrogen and carbon precursors. A Cu disk was used as a substrate for the growth of CNx films. Pulsed discharge was performed for 1 hour. The intention of growing CNx films on the surface of Cu

The bond structure and microstructure of the thin films was investigated using Raman spectroscopy (micro-Raman, LabRam HR 800 system). Raman measurements were recorded using a He–Ne laser, which was focused using a 50X lens. During the measurements, 60 recoded data were collected every 60 s. SEM (Jeol IT-100 coupled to a Bruker X-Ray microscope) and AFM (EasyScan 2 Flex AFM) were used to determine the thin-film morphology. The SEM microscope was operated in high voltage mode with an accelerating voltage of 20 kV. SEM studies were carried out on the electrolytic Cu (99.99%) substrate to observe the formation of attached carbon structures and thus corroborate their existence.

To characterize the plasma during film deposition, the species generated within the plasma discharge was observed by optical emission spectroscopy (OES). An ocean optics HR 4000CG-UV-NIR spectrometer was used, and spectra were observed in a wavelength range of 200 to 850 nm, with a resolution of 0.75-nm FWHM.

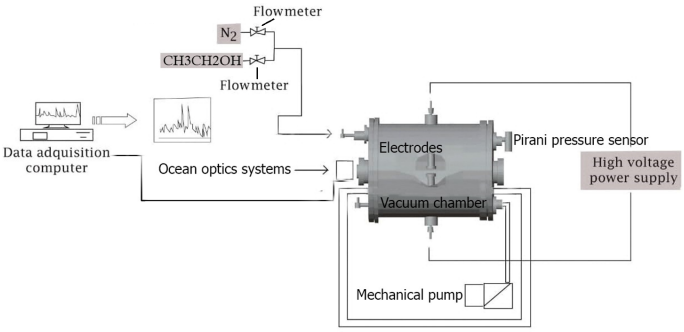

Figure 1: Experimental apparatus

3. Results and Discussion

3.1. Raman spectroscopy

Raman spectroscopy is commonly used to analyze the bond structure of carbon-related materials. Visible Raman spectroscopy is 50 to 230 times more sensitive to sp^2 sites because π electrons are preferentially excited with visible photons [12]. Thus, the Raman spectra obtained for the deposited CN_x films provided double-link information, in addition to indication of changes in film microstructure.

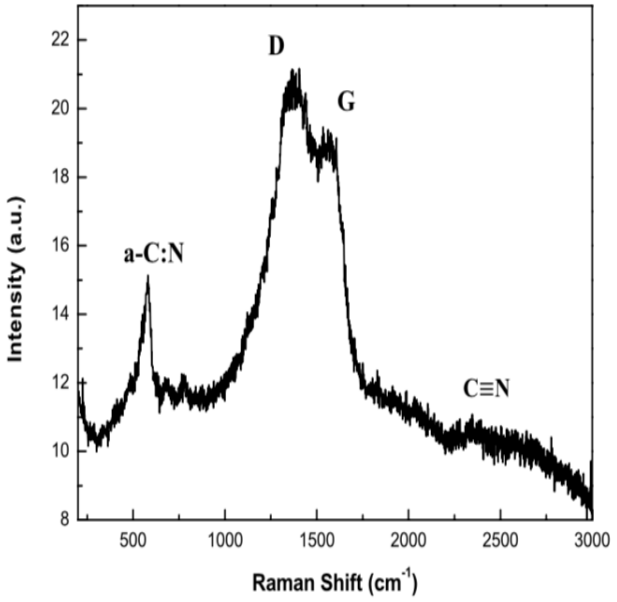

Figure 2: Raman spectra of thin CN_x deposited by pulsed discharge at a pressure of 1 Torr.

Figure 2 shows Raman spectra for CN_x films deposited onto Cu at 300 mA, 550 V and 120 W at a working pressure of 1 Torr. Raman bands appear at 692, 1,362 (D peak), 1,588 (G peak), and 2,250 cm^{-1} [14]. The 692 cm^{-1} band is characteristic of CN films

deposited under the influence of a reactive gas. This band may also appear owing to structural disorder in the film. The G peak arises from stretching vibrations at all sp^2 sites, both in olefinic chains and in aromatic rings, while the existence of the D peak is associated with breathing vibrations in the aromatic rings. These peaks, therefore, indicate the presence of distorted-ring structure, or the lack of coherence between adjacent planes of graphite-like structures [15, 16]. This is due to the disorder in the link angles, despite the graphite-like micro-domains induced by the binding of sp^3 carbon atoms or the finite size of the sp^2 micro-domains.

The D peak is only related to the presence of aromatic rings. The D peak arises in materials that exhibit formations of small clouds or clusters organized into such rings (graphite clusters) [17]. The band centered at about 2250 cm^{-1} is due to the stretching vibration of a C-N triple bond, suggesting that the C and N atoms are chemically bonded in the film. The position of the D and G peaks, together with their widths and the relationship between their inetnsities (I_D/I_G) provide valuable information concerning the deposited films.

The position of the G peak is shifted toward higher energies as a result of the current used within our system (300 mA). This current value guarantees an increase in the system temperature. The position of the G peak, independent of the composition of the film or the growth method, is related to the vibrational energy of the present bonds. Therefore, the shift toward high frequencies is due to a decrease in sp^3 bonds and an increase in sp^2 links [18]. This indicates that a conversion of sp^3 to sp^2 bonds occurred in the a-CN_x films due to the increase in temperature. This conversion process could form graphite domains with larger size and/or a number of sp^2 clusters, eventually leading to graphitization.
The Raman shift of the D peak at high frequencies suggests a densification of sp^2-ringed structures in the atomic network, which also leads to increased sp^2-link content [15]. On the other hand, $I_D/I_G = 1.11$, which suggests that C sp^2 are more likely to form C=C and/or C=N bonds in rings and less likely to form C=C and/or C=N bonds in chains. Additionally, the I_D/I_G value indicates the growth of graphite domains or an increase in their number [19–21], i.e., graphitization by heating with an increase in the disorder of the bond angles of C atoms and N sp^2 in the films [20, 22]. The width of the D peak also reflects the crystallinity of the structure: a narrower band suggests the structure is more crystalline, but if it is wider, the structure is more amorphous. For the obtained spectrum, it can be seen that peak D is wider, which suggests derangement of the structure due to the creation of defects, possibly through the introduction of C or N atoms between the graphite-like layers. This may indicate that the structures formed in the thin-film are more amorphous.

3.2. SEM analysis

The SEM images (Figure 3) acquired at different magnifications reveal the morphology of the deposited thin-film. In Figure 3(a), the carbon clusters deposited onto the Cu surface form a non-homogeneous surface. It is likely that this lack of uniformity is due to the distance between the electrodes (20 mm) because the deposition rate is proportional to this distance [23]. The agglomerations observed in the SEM image cause the formation of large islands with small grains, in turn generating an uneven surface [24]. Images of the thin-film at different magnifications, shown in Figures 3(b) and 3(c), were collected to better observe the surface. Thus, it is inferred that internal stresses decrease within the material, allowing the diffusion of particles within the film surface. This uniformity on the surface can be explained by two factors: the energy with which the plasma interacts with the target and the high internal strain generated within the samples as a result of the formation of amorphous structures

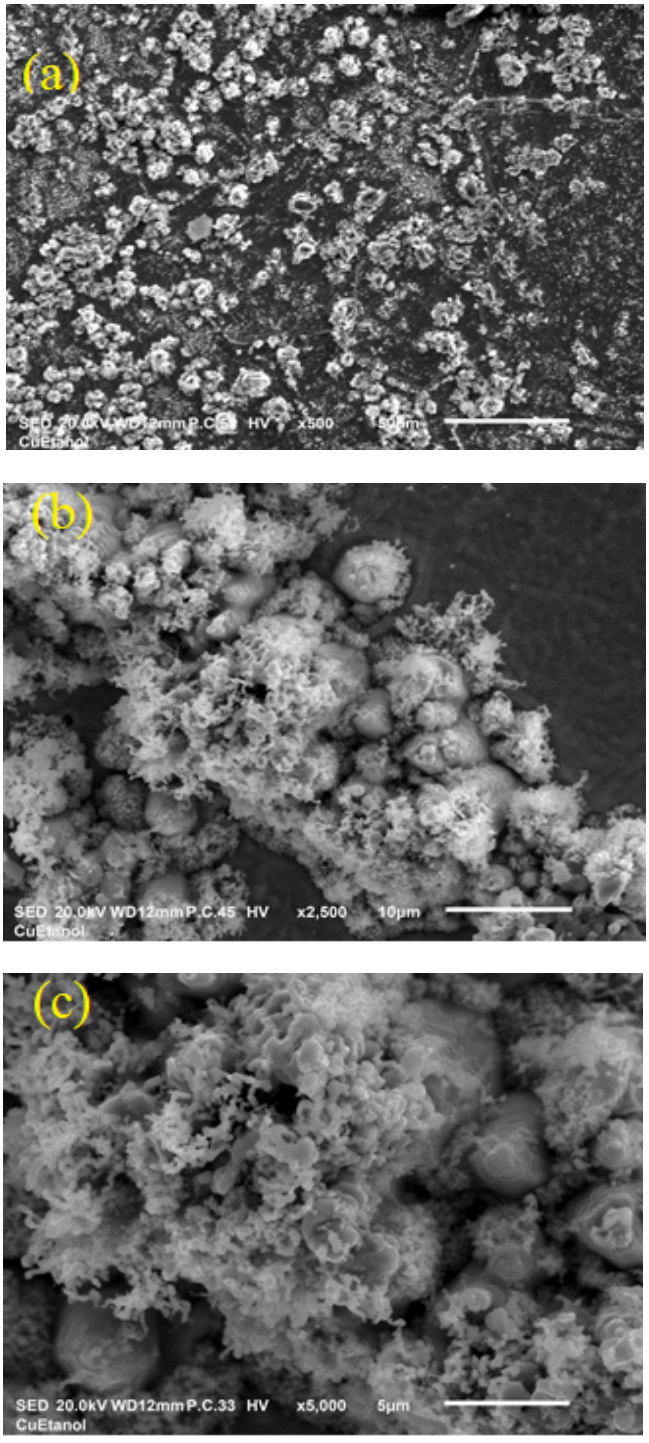

Figure 3: SEM micrographs of thin CN_x deposited by pulsed discharge at a pressure of 1 Torr, taken at (a) 10X, (b) 50X, and (c) 1000X magnification..

3.3. AFM analysis

Figure 4 shows the three-dimensional (3D) AFM images of the deposited thin-film using a current of 300 mA. Figures 4(a), 4(b), and 4(c) show results obtained by AFM at different scales in contact mode that is, at swept lines of 50, 25, and 10 µm, respectively. The morphology that occurs in the deposited films is determined by the non-uniformity in the surface morphology, which can be explained by the presence of large graphitic domains and a polymer component of CN.

Higher polymer components lead to a smoother film, whereas higher graphite-cluster contents and larger cluster sizes lead to more agglomerations [23]. A large polymeric component results in low roughness, whereas a greater number of graphitic clusters induce the formation of more or larger agglomerations. Although no defined grains were observed on the surfaces of the films, we have established that the CN_x films had linear and graphite-like polymer components, which could lead to the surface non-uniformity. These results agree with those presented by [25] regarding the morphology of CN_x films.

The microstructure, in addition to the morphology formed during thin-film formation, are important issues in determining the mechanical (hardness and elastic recovery) and tribological properties (friction coefficient and wear rate) of CN_x films [26]. Therefore, such thin films could be important materials for application in many tribological uses, because of their mechanical properties, and in solar cells, due to their optical and electrical properties [27]. Another interesting application of the present work is that the observed morphology determined the non-uniformity in the surface that effect could be reproduce using phase-field model for planer film surface [28], by the corresponding different nanoscale effects [29], or by thin planar interfacial models [30, 31].

3.4. Plasma analysis

Figure 5 display the OES spectrum for CH_3CH_2-N_2 in the spectral range of 200–900 nm. In the mixture, decomposition of CH_3CH_2OH takes place by breaking the C-H chemical bonds in the discharge, producing atomic and molecular impacts and excitation reactions between radicals and N atoms. The species observed are as follows: Swan bands ($d^3\Pi_g$ - $a^3\Pi_u$), which are characteristic of the C_2 molecule at 473.71 nm ($\Delta\nu = +1$), 516.52 nm ($\Delta\nu = 0$), and 563.55 nm ($\Delta\nu = -1$) [16], and CN violet bands ($B^2\Sigma^+$ - $X^2\Sigma^+$), for which the band heads appear at 359.04 nm ($\Delta\nu = +1$), 388.34 nm ($\Delta\nu = 0$), and 421.60 nm ($\Delta\nu = -1$). Here $\Delta\nu = \nu' - \nu''$ is the difference in vibrational quantum numbers between the upper (ν') and lower (ν'') transition states. Also seen are CH ($B^2\Sigma$-$X^2\Pi$) at 390 nm and CH ($A^2\Delta$-$X^2\Pi$) at 426.76 nm.

Figure 6 shows CN, N_2, and H_β emissions observed in the plasma discharge. The N_2 ($B^3\Pi_g$-$C^3\Pi_u$) band, which corresponds to the second positive system (SPS, 337.1 nm), may be mainly due to the presence of two consecutive reactions: (1) the electron impact excitation of the N_2 molecular energy ground state and (2) radiative decay to a lower level with the emission of a characteristic photon. These reactions are respectively expressed as

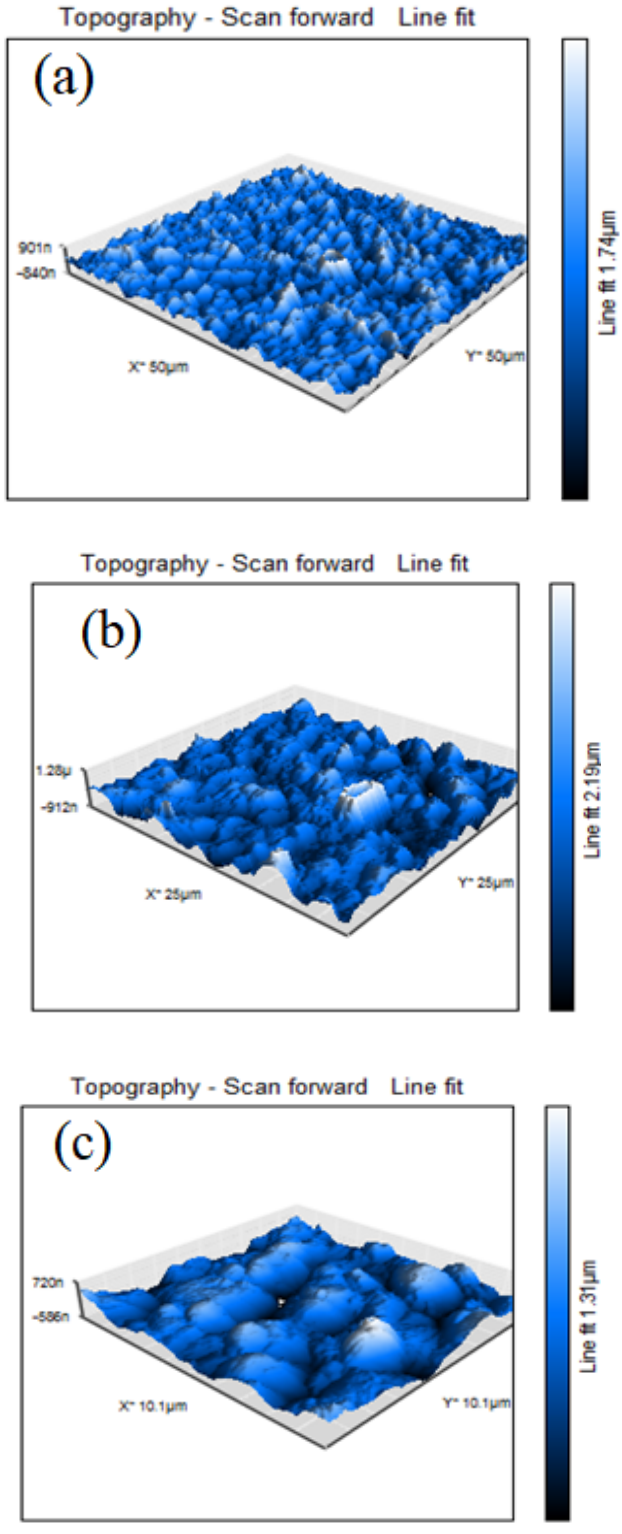

Figure 4: AFM images of thin CN_x deposited by pulsed discharge at a pressure of 1 Torr, taken at different scales. (a) 50 µm; (b) 25 µm; (c) 10 µm

$$e + N_2(X_1\,\Sigma^+_g) \rightarrow N_2(C^3\,\Pi_u) + e \quad (1)$$

and

$$N_2(C^3\,\Pi_u) \rightarrow N_2(B^3\Pi_g) + h\nu\ (337.1\ nm) \quad (2)$$

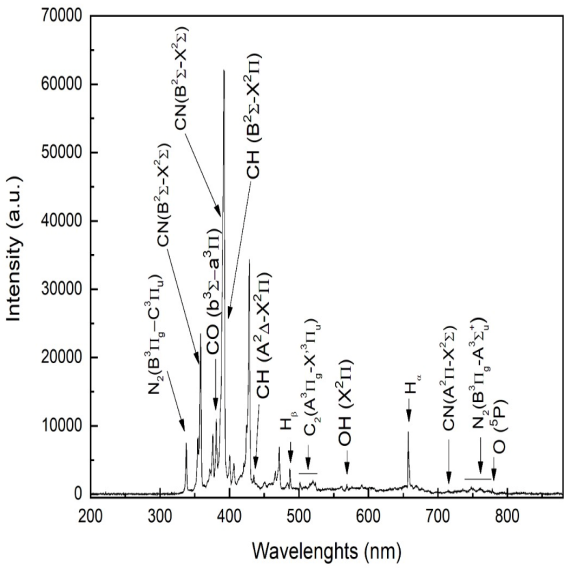

Figure 5: OES CH_3CH_2OH/N_2 plasma.

The interactions of N_2 species with C_2H_5OH lead to the formation of CN species [32]. The violet and red spectral systems of CN are shown in Figure 6. The violet spectral system of CN may be produced by the following reaction:

$$CN(B^2\Sigma^+) \rightarrow CN(X^2\Sigma^+) + h\nu \quad (3)$$

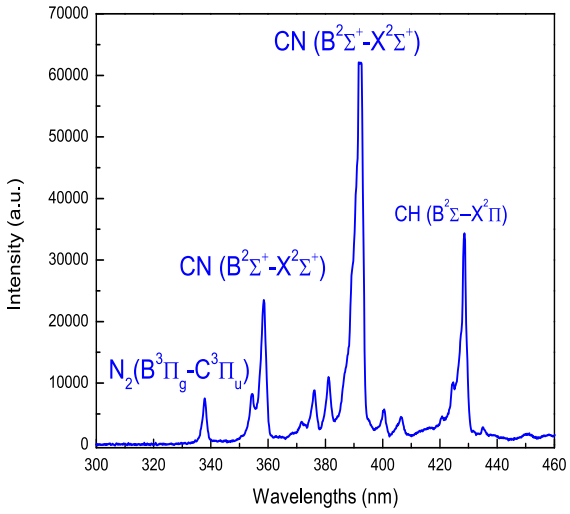

Figure 6: OES spectrum CN, N_2 and CH species.

which corresponds to $\Delta\nu = 0$ and a radiative lifetime τ = 60 ns. The resulting vibrational distribution of CN ($X^2\Sigma^+$) can be modified by relaxation processes that include vibrational collision relaxation within each electronic state, inter-electronic transfer by collision between vibrational levels of the states $X^2\Sigma^+$ and $A^2\Pi$, and the radiative deexcitation of the vibrational levels of $A^2\Pi$ toward vibrational levels lower than the $X^2\Sigma^+$ state. The latter of these relaxation processes result in the appearance of the red spectral system, as shown in Figure 7, and is expressed as

$$CN(A^2\Pi) \rightarrow CN(X^2\Sigma^+) + h\nu \quad (4)$$

with $\Delta\nu = 3$ and a radiative lifetime τ = 6000 ns.

Conversely, from Figure 5 it can be seen that the CN species at 388.3 nm and the N_2^+ species at 390 nm overlap, implying that the production of CN radicals in that interval is mainly linked to the excited state of N_2 ($C^3\Pi_u$). This phenomenon has two causes: (*i*) resonant excitation in collisions between highly excited neutral N_2 molecules (3.26 eV) and CN radicals (3.19 eV) or (*ii*) the high energy of the excited N_2 (11.1 eV) favoring the generation of atomic N, which in turn favors CN formation. The reaction of N atoms with C or C-containing species is exothermal, which facilitates these reactions [33–35]. N atoms within the discharge could not be detected because their lifetimes are shortened by bonding with N atoms and C-containing species.

Its peaks are also covered up by CN compounds with higher intensities. The presence of CN radicals within the discharge is mainly related to the excited state of N_2 [32]. C_2 Swan band emissions arise from transitions between electronic states ($d^3\Pi_g$-$a^3\ \Pi_u$) as a probable result of the dissociation of CH_3CH_2OH in the following way:

$$CH_3CH_2OH \rightarrow e$$
$$CH_3CH_2OH \xrightarrow[e\ impact]{} CH_2OH, CH_3 \xrightarrow[e,rad\ impact]{} CO^*, C_2^* \xrightarrow[c\ condensation]{} C(solid, CO) \quad (5)$$

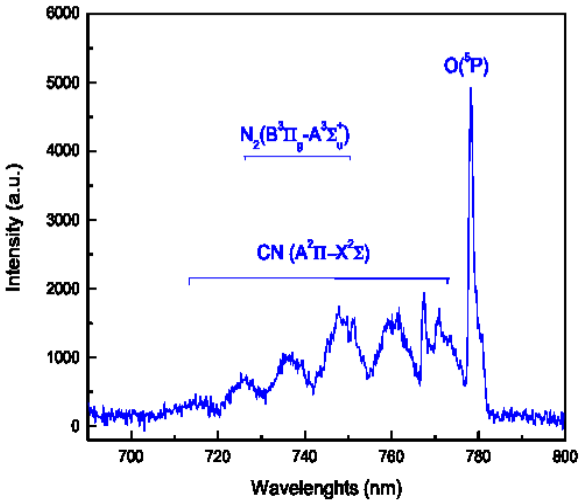

Figure 7: OES spectrum CN (red spectral system), N_2 and O atom species.

As can be seen in (5), atomic hydrogen is removed from CH_3CH_2OH [32, 36] and the C-C bond finally dissociates by direct electron impact, which leads to the formation of solid carbon. The presence of C_2 can be taken as an indication of solid carbon formation, as it is known that the C_2 species initially nucleates in solid carbon clusters [37]. This would result in the presence of C_2 Swan bands in the range of 480–530 nm.

Swan band heads correspond to (0,0), (1,1), (2,2), (3,3), and (4,4) with a $\Delta\nu = 0$. Carbon molecules react with N atoms in the following way:

$$N + C_2 \rightarrow CN + C \quad (6)$$

We assume that generation of C_2 molecules was caused by collisions in the plasma, which promote recombination of the dissociated C of the CH_5 species. The C_2 species and the CN radicals generated within the mixture contribute to the growth of the CN thin films. The availability of C_2 species provides a considerable fraction of sp^2 hybridization, as opposed to sp^3 hybridization, which is an important factor in the formation of graphite-like or fullerene-like structures with sp^2 hybridization in deposited CN films.

The emission of CH ($B^2\Sigma^+$-$X^2\Pi$) at 388.9 nm becomes embedded in the CN signal, as seen in Figure 5. The presence of CH in the CH_3CH_2OH discharge can be considered an indication of solid carbon and higher order hydrocarbon formation. This is because CH species are typically found as intermediate radicals in the formation of these products [36]. CH species react with nitrogen atoms in the following process:

$$N + CH \rightarrow CN + H \quad (7)$$

However, CH radicals can arise as intermediate reaction products to generate CH_4, C_2H_2, C_2H_4, C_2H_6, or C_3H_8 [38, 39] when hydrogen atoms (H_α and H_β) react with solid carbon fragments [40–42]. The presence of H_α and H_β in the discharge is a result of (7), which forms the hydrogen species. In Figure 5, these species are seen as the most intense lines of the Balmer series.

$$e + OH \rightarrow O + H + e \quad (8)$$

The presence of H_α and H_β is related to dissociation of the CH and OH within the discharge, as seen in (7) and (8), which could also be a new source for hydrogen generation.

4. Conclusions

CN_x thin films were deposited on Cu substrates using a pulsed discharge method with CH_3CH_2OH as the source of carbon and N_2 as the sputtering ion and nitrogen source.

Through Raman analysis, it was demonstrated that the structure of the deposited films was amorphous. The characteristic Raman D and G peaks for carbon were identified. The position of the G peak is related to the vibrational energy of the present bonds; therefore, there is a shift toward higher frequencies due to the decrease of sp^3 bonds and hence the increase in sp^2 bonds. This indicates that a conversion of sp^3 bonds to sp^2 bonds occurred in the CN_x films owing to the increase in temperature. Such a conversion process could form large graphite domains with large size and/or an increased number of sp^2 clusters, eventually leading to graphitization. The observed high-frequency D peak shift, on the other hand, suggested densification of ringed sp^2 in the atomic network, also increasing the sp^2-bond content. In addition, the Raman analysis showed an additional band centered at approximately at 2250 cm^{-1}, which appears because of the stretching vibration of a triple bond C-N. This suggests that the C and N atoms are chemically bonded in the film. An I_D/I_G value of 1.32 suggested that C sp^2 is more likely to have formed in the film, considering that C=C and/or C=N bonds can be formed in rings and there are fewer opportunities to form C=C and/or C=N links in chains. In addition, this I_D/I_G value indicated the growth of the graphitic domains or the increase of their number, i.e., graphitization by heating with an increase in disorder of the angles of the bonds of the C and N atoms sp^2 in the films.

Through SEM analysis, we were able to observe the thin-film morphology, wherein carbon clusters deposited on the surface of the Cu were observed. The lack of uniformity of the non-homogeneous surface can be attributed to the distance between the electrodes during the deposition process. Since the deposition rate is considered proportional to the distance, these observed agglomerations cause the formation of large islands with small grains, in turn generating an irregular surface. These results were later corroborated with AFM results, from which we determined the morphology of the CN_x film. The surface was found to lack uniformity, which can be explained by the presence of large graphite domains and a polymeric component of CN. This reveals that the greater the content of graphite groups or the larger their size, the greater the number of agglomerations. Finally, we conclude that the films grow via island growth, very similar to the results presented by [25].

OES measurements confirmed the decomposition of the CH_3CH_2OH molecule as a result of electron impact dissociation in the plasma discharge. Products obtained from the dissociation collide with the N species to generate CN radicals and form violet and red emission spectra characteristic of CN. CN radicals are important because of their variety of industrial applications, mainly in materials science. In addition, the C_2 Swan band, a species that also plays an important role in the appearance of CN radicals, was identified. The presence of CH indicated the formation of solid carbon. Hydrogen species were also detected in the CH_3CH_2OH-N_2 plasma mixture. Therefore, the mixture plasma is not only useful for producing CN but also for generating hydrogen.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We are grateful to the Advanced Physics Laboratory (FC/UAEM), Spectroscopy laboratory (ICF/UNAM) and TESJo for their infrastructure provided to carry out the research. This research was supported by CONACyT scholarship for doctoral study, DGAPA[IN102222], CONACyT [225991 and 268644] and UAEM [4307/2017/CI].

Reference

[1] M. Uz A. Y. Liu, M. L. Cohen, (1992), Theorical study of a hypothetical metallic phase of carbon, Phys. Rev. B, **45**, 4579-4581. https://doi.org/10.1103/PhysRevB.45.

[2] Y. Bu, Z. Chen, J. Yu & W. Li, (2013), A novel application of g-C_3N_4 thin film in photo-electrochemical anticorrosion, Electrochim. Acta., **88**, 294-300. https://doi.org/10.1016/j.electacta.2012.10.049

[3] S. Yan, C., Li, Z. S. & Zou, Z. G., (2009), Photo-degradation Performance of g-C_3N_4 Fabricated by Directly Heating Melamine, Langmuir, **25**, 10397-10401. https://doi.org/10.1021/la900923z

[4] F. K. Kessler; Y. Zheng; D. Schwarz; C. Merschjann; W. Schnick; X. Wang; M. J. Bojdys, (2017), Functional carbon nitride materials- design strategies for electrochemical devices, Nature Reviews Materials, **17030**. https://doi.org/10.1038/natrevmats.2017.30

[5] K. Schwinghammer, M. B. Mesch, V. Duppel, C. Ziegler, J. Senker and B. V. Lotsch, (2014), Crystalline Carbon Nitride Nanosheets for Improved

Visible-Light Hydrogen Evolution, J. Am.Chem.Soc.,**136**, 5, 1730-1733. https://doi.org/10.1021/ja411321s

[6] J. Xu, M. Antonietti, and M. Shalom, (2016), Moving Graphitic Carbon Nitride from Electro-catalysis and Photo-catalysis to a Potential Electrode Material for Photoelectric Devices, Chem. Asian J., **11**, 2499-2512. https://doi.org/10.1002/asia.201600857

[7] R. Kour, Sandeep Arya, Sheng-Joue Young, Vinay Gupta, Pankaj Bandhoria, and Ajit Khosla, (2020), Review-Recent Advances in Carbon Nanomaterials as Electrochemical Biosensors, Journal of The Electrochemical Society, **167** 037555. https://doi.orfg/10.1149/1945-7111/ab6bc4

[8] H. Yu, P. Li, J. Robertson, (2011), Fabrication and bio-functionalization of tetrahedral amorphous carbon thin films for bio sensor applications, Diamond & Related Materials, **20**, 1020-1025. https://dx.doi.org/10.1016/j.diamond.2011.06.005

[9] Z. A. Umar, R. S. Rawat, R. Ahmad, Z. Chen, Z. Zhang, J. Siddiqui, A. Hussnain, T. Hussainc and M. A. Baiga, (2017), Structural, compositional and hardness properties of hydrogenated amorphous carbon nitride thin films synthesized by dense plasma focus device, Surface and Interface Analysis, **49**, 548-553. https://doi.org/10.1002/sia.6192

[10] P. Wang, T. Takeno, K. Adachi, H. Miki & T. Takagi, (2012), Preparation and tribological characterization of amorphous carbon nitride coatings in a RF PECCVD-DC PVD hybrid coating process, Applied Surface Science, **258**, 6576-6582. https://doi.org/10.1016/j.apsusc.2012.03.082

[11] D, Nicolas et al., Thierry Fouquet, Patrick Choquet, (2015), Atmospheric Pressure Plasma Initiated Chemical Vapor Deposition Using Ultra-Short Square Pulse Dielectric Barrier Discharge, Plasma Process. Polym., **12**, 66-74. https://doi.org/10.1002/ppap.201400094

[12] N. Nedfors, Oleksiy Vozniy, and Johanna Rosen, (2018), Effect of synchronized bias in the deposition of TiB_2 thin films using high power impulse magnetron sputtering, J. Vac. Sci. Technol. A, **36**, 3. https://doi.org/10.116/1.5003194

[13] A. Aijaz Kostas Sarakinos, Mohsin Raza, Jens Jensen and Ulf Helmersson, Principles for designing sputtering-based strategies for high-rate synthesis of dense and hard hydrogenated amorphous carbon thin films, Diamond and Related Materials. Vac., **44**, 117-122, 2014. https://doi.org/10.1016/j.diamond.2014.02.014

[14] Y. Kang, Yongqiang Yang, Li-Chang Yin, Xiangdong Kang, Gang Liu, and Hui-Ming Cheng, (2015), An Amorphous Carbon Nitride Photo-catalyst with Greatly Extended Visible-Light-Responsive Range for Photo-catalytic Hydrogen Generation, Adv. Mater., **27**, 4572-4577. https://doi.org/10.1002/adma.201501939

[15] R O Dillon, J Woollam, V Katkanant, (1984), Use of Raman scattering to investigate disorder and crystallite formation in as-deposited and annealed carbon films, Phys. Rev. B, **29**, 3482. https://doi.org/10.1103/PhysRevB.29.3482

[16] Y Taki, T Kitagawa, O Takai, (1997), Shielded arc ion plating and structural characterization of amorphous carbon nitride thin films, Thin Solid Films, **304**, 183-190. https://doi.org/10.1016/S0040-6090(97)00193-4

[17] A C Ferrari, S E Rodil, J Robertson, (2003), Interpretation of infrared and Raman spectra of amorphous carbon nitride, Phys. Rev. B, **67**, 155306, 155. https://doi.org/10.1103/PhysRevB.67.155306

[18] H Jung, H Ho Park, (2000), Studies on the structure and bonding state of nitric amorphous carbon (a-CN_x) films by reactive rf magnetron sputtering, Thin Solid Films, **377-378**, 320-325. https://doi.org/10.1016/S0040-6090(00)01363-8

[19] J. Wang, N. Huang, C. J. Pan, S.C.H. Kwok, P. Yang, Y.X. Leng, J. Y. Chen, H. Sun, G. J. Wan, Z. Y. Liu and P. K. Chu, (2004), Bacterial repellence from polyethylene terephthalate surface modified by acetylene plasma immersion ion implantation-deposition, Surface and Coatings Technology, **186 (1-2)**, 299-304. https://doi.org/10.1016/j.surfcoat.2004.02.046

[20] E. Cappelli, C. Scilletta, S. Orlando, V. Valentini and M. Servidori., (2009), Laser annealing of amorphous carbon films, Applied Surface Science, **255**, 5620-5625. https://doi.org/10.1016/j.apsusc.2008.10.062

[21] L. Marcinauskas, A. Grigonis, P. Valatkevicius and A. Medvid, (2012), Irradiation of the graphite-like carbon films by ns-laser pulse, Applied Surface Science, **261**, 488-492. https://doi.org/10.1016/j.apsusc.2012.08.042

[22] X. Wang, Z. Li, P. Wu, E. Jiang and H. Bai, (2006), Annealing effects on the microstructure of amorphous carbon nitride films, Applied Surface Science, **253 (4)**, 2087-2092. https://doi.org/10.1016/j.apsusc.2006.04.003

[23] R. Maeda & K. Kikuch, (1997), Deposition of thin films by uv light laser ablation, Surf. Eng., **13**,71-74. https://doi.org/10.1179/sur.1997.13.1.71

[24] Ming Gao, Danni Liu, Huanhuan Yang, Hao Huang, Qian Luo, Yifan Huang, Xue-Feng Yu and Paul K.Chu, (2019), Modification of Layered Graphitic Carbon Nitride by Nitrogen Plasma for Improved Electrocatalytic Hydrogen Evolution, Nanomaterials, **9**, 568. https://doi.org/10.3390/nano9040568

[25] A. Stanishevsky, (1998), On the surface morphology of C: N films deposited by pulsed cahodic arc discharge method, Mater. Lett., **37**, 162-167. https://doi.org/10.1016/S0167-577X(98)00085-8

[26] E. Broitman et al., Mechanical and tribological properties of CN_x films deposited by reactive magnetron sputtering, Wear, **248,** 55-64. https://doi.org/10.1016/S00431648(00)00519-6

[27] Z. B. Zhoua, R. Q. Cui, Q. J. Pang, G. M. Hadi, Z. M. Ding, W. Y. Li, (2002), Schottky solar cell with amorphous carbon nitride thin films prepared by ion sputtering technique, Solar Energy Materials & Solar Cell, **70,** 487-493. https://doi.org/10.1016/S0927-0248(01)00086-1

[28] A. M. Roy, (2021), Barrierless melt nucleation at solid-solid interface in energetic nitramine octahydro-1, 3, 5, 7-tetrazocine, Materialia, **15**, 101000. https://doi.org/10.1016/j.mtla.2021.101000

[29] A. M. Roy, (2021), Influence of Nanoscale Parameters on Solid-Solid Phase Transformation in Octagen Crystal: Multiple Solution and Temperature Effects, JETP Letters, **113** (4), 265-272. doi.org/10.1134/S0021364021040032

[30] A. M. Roy, (2021), Formation and stability of nanosized, undercooled propagating intermediate melt during β→δ phase transformation in HMX nanocrystal, Europhysics Letters, **133**, 56001. doi: 10.1209/0295-5075/133/56001

[31] A. M. Roy, (2021), Energetics and kinematics of undercooled nonequilibrium interfacial molten layer in cyclotetramethylene-tetranitramine crystal, Physica B: Condensed Matter, **615**, 412986. https://doi.org/10.1016/j.physb.2021.412986

[32] E. Tatarova, N. Bundaleska, F. M Dias, D. Tsyganov, R. Saavedra and C. M. Ferreira, (2013), Hydrogen production from alcohol reforming in a microwave 'tornado'-type plasma, Plasma Sources Sci. Technol., **22**. 065001. https://doi.org/10.1088/0963-0252/22/6/065001

[33] K. J. Clay, S. P. Speakman, G. A. J. Amaratunga, and S. R. P. Silva, (1996), Characterization of a-C:H:N deposition from CH_4/N_2 rf plasmas using optical emission spectroscopy, J. Appl. Phys., **79** (9),7227-7233. https://doi.org/10.1063/1.361439

[34] S. Veprek, J. Weidmann and F. Glatz, (1995), Plasma chemical vapor deposition and properties of hard C_3N_4 thin films, J. Vac. Sci. Thechnol., **13**, 2914. https://doi.org/10.1116/1.579613

[35] Wei Hu, Jun-Yi Tang, Jia-da Wu, Jian Sun, Yi-qun Shen, Xiao-feng Xu, and Ning Xu, (2008), Characterization of carbon nitride deposition from CH_4/N_2 glow discharge plasma beams using optical emission spectroscopy, Physics of plasma, **15**, 073502. https://doi.org/10.1063/1.2953521

[36] A. Yanguas-Gil, J. L. Hueso, J. Cotrino, A. Caballero, and A. R. González-Elipe, (2004), Reforming of ethanol in a microwave surface-wave plasma discharge, Appl. Phys. Lett., **85**,4004-4006. https://doi.org/10.1063/1.1808875

[37] A. Kumar, P. A., A. Xue, B. Hao , Y. K. Yap and R. M. Sankaran, (2013), Formation of nanodiamonds at near-ambient conditions via micro-plasma dissociation of ethanol vapour, Nature Communication, **4,** 2618. https://doi.org/10.1038/ncomms3618

[38] J. Serrano, J. Moros and J. J. Laserna, (2015), Exploring the formation routes of diatomic hydrogenated radicals using femtosecond laser-induced breakdown spectroscopy of deuterated molecular solids, J. Anal. At. Spectrom., **30,** 2343-2352. https://doi.org/10.1039/C5JA00192G

[39] G. M. Harris and A. W. Tickner, (1947), Reaction of Hydrogen Atoms with Solid Carbon, Nature, **160** ,871. https://doi.org/10.1038/160871a0

[40] A. Okita, Y. Suda, A. Ozeki, H. Sugawara, Y. Sakai, A. Oda, and J. Nakamura, (2006), Predicting the amount of carbon in carbon nanotubes grown by CH_4 RF plasmas, J. Appl. Phys., **99**, 014302. https://doi.org/10.1063/1.2150599

[41] A. Okita, Y. Suda, A. Oda, J. Nakamura, A. Ozeki, K. Bhattacharyya, H. Sugawara, and Y. Sakai, (2007), Effects of hydrogen on carbon nanotube formation in CH_4/H_2 plasma, Carbon, **45**, 7, 1518-1526. https://doi.org/10.1016/j.carbon.2007.03.022

[42] A. Oda, Y. Suda, and A. Okita, (2008), Numerical analysis of pressure dependence of carbon nanotube growth in CH_4/H_2, Thin Solid Films, **516**,19, 6570-6574. https://doi.org/10.1016/j.tsf.2007.11.088

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 60-78 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Thermoelectric Generators (TEGs) and Thermoelectric Coolers (TECs) Modeling and Optimal Operation Points Investigation

Nganyang Paul Bayendang*, Mohamed Tariq Khan, Vipin Balyan

Department of Electrical, Electronic and Computer Engineering (DEECE), Cape Peninsula University of Technology (CPUT), Bellville Campus, Cape Town, Western Cape, 7535, South Africa

A R T I C L E I N F O

Article history:
Received: 11 September, 2021
Accepted: 05 December, 2021
Online: 10 February, 2022

Keywords:
Alternative Energy
Energy Efficiency
Energy Harvesting
Thermoelectric Coolers (TECs)
Thermoelectric Generators (TEGs)
TEGs and TECs Modeling
TEGs/TECs Optimal Operation
Thermoelectricity

A B S T R A C T

Sustainable energy is gradually becoming the norm today due to greenhouse warming effects; as a result, the quests for different renewable energy sources such as photovoltaic cells as well as energy efficient electrical appliances are becoming popular. Therefore, this article explores the alternative energy case for thermoelectricity with focus on the steady-state mathematics, mixed modelings and simulations of multiple TEGs and TECs modules to study their performance dynamics and to establish their optimal operation points using Matlab and Simulink. The research substantiates that the output current from TEGs or input current to TECs, initially respectively increases the output power of TEGs and the cooling power of TECs, until the current reaches a certain maximum optimal point, after which any further increase in the current, decreases the TEGs' and or TECs' respective output and cooling powers as well as efficiencies, due to Ohmic heating and or entropy change caused by the increasing current. The research main contributions are elaborate easy to understand TEGs/TECs theoretical formulations as well as static and dynamic simulated models in Matlab/Simulink, that can be used initially to dynamically investigate an infinite quantity of TEG and TEC modules connections, be it in series and or in parallel. This is to assist system designers grasp TEGs and TECs theoretical operations better and their limits, when designing energy efficient waste heat recovery (using TEGs)/cooling (using TECs) systems for industrial, residential, commercial and vehicular applications.

1. Introduction

According to [1], energy security and green economy are becoming paramount today; as a result, the demands for renewable and alternative energy sources such as solar, wind, hydro energy, bio-fuels and fuel cells, as well as energy efficient loads, are on the rise in an effort to ensure energy sustainability and carbon free environment. In this regards, we investigated thermoelectricity as a potential alternative energy for sustainable energy source and loads − that is, as a clean DC power source for low energy lighting/applications and as well to provide clean cooling/heating in various human habitats. Thermoelectricity as reviewed in [2], practically focuses on the Seebeck and Peltier effects. Seebeck effect is basically converting heat to DC electricity and the device that does this is a thermoelectric generator (TEG). The reversed phenomenon is a Peltier effect − which is basically the production of cold from DC electricity and if the direction of current flow changes (swap voltage polarity), heat is also produced and the device that does this is called a thermoelectric cooler (TEC). Therefore, by efficiently applying thermoelectricity prudently, a clean alternative energy source for DC low power applications using TEGs and or energy efficient loads in the forms of heat pumps, air conditioners, refrigerators etc using TECs; can be passably implemented to help sustain some human habitats basic energy consumption such as lighting, cooling and heating; as well as reduce environmental pollution.

As already examined in [2], thermoelectricity lends itself to various applications with focus on how TEGs and TECs can be used respectively as a power source and as a load. Furthermore, studied in [3], is a re-configurable TEG DC-DC converter for maximum TEG energy harvesting in a battery-powered wireless sensors network (WSN). Described in [4], is the analysis and design of a thermoelectric energy harvester (TEH) prototype for

*Corresponding Author: NP Bayendang, CPUT DEECE, +27765404896, bayendangn@cput.ac.za

powering up outdoor sensors and devices. Solar energy was harvested using different TEG arrays in [5] and a theoretical analysis of implementing a re-configurable TEG was researched in [6]. Electronic cooling was investigated in [7] and the findings revealed that the TEC cooling capacity could be increased by increasing its cold side junction temperature and decreasing its temperature difference. A multi-stage TEC module in cascade was examined in [8]; whereas in [9], an extensive mathematical analyses were articulated for TEG and TEC design and materials. A TEG model was developed in [10] for maximum power point tracking but lacks the detailed underlining maths and the parallel TEG combinations was limited to just 2. A comprehensive TEG and TEC models with the detailed maths supporting the TEG and TEC models, were presented respectively in [11] and [12]. In [13], a modeling of TEG using Modelica is asserted but deficient in the comprehensive maths, especially considering modeling infinite multiple TEGs and as well TECs modules —which were not articulated. A parametric ANSYS study of TEG and TEC was presented in [14]; however, the detailed maths and especially for the case for infinite TEG and TEC modules use/ connection, was inadequate. In addition, for large scale TEGs and TECs applications, the following studies were examined. In [15], 600 TEGs with a temperature difference of ~120 ℃, were applied to harvest and generate up to 1 kW of DC power from geothermal heat. It was further indicated a 2 kW power could be achieved with a higher temperature difference and also the TEG cost is much lower to generate equivalent amount of power than using photovoltaic. However, the study lacks the theoretical details to substantiate it. TEG harvesting of waste thermal energy from household heat sources such as a generator exhaust pipe and a kerosene stove, were performed in [16] and various parameters measurements were made but without detailing the maths to calculate these parameters. Light and heat from the Sun are the most common forms of energy abundant on Earth; as a result, [17] reviewed the possibility of integrating photovoltaic and TEG in a hybrid photovoltaic-TEG system and further examined the efficiency improvement. A 128 TEGs system was assembled in [18] to generate ~684 W of power from radiation heat transfer at a temperature difference of ~125 K and with a corresponding power density of 845 W/m^2. Their results further justified that with a greater practical temperature difference of 200 K, the respective generated power and power density of their TEGs system could attain 1.23 kW and 1.51 kW/m^2. Their TEGs system open circuit voltage, its output power, its power density and its conversion efficiency were investigated in details at different temperature differences; however, the underlining maths was not elaborated. A grid-tied 20 W TEG experimental model using 24 modules in series with the heat harvested from a waste incinerator, was experimented in a lab and the preliminary and analytical models of the electric output power as a function of specific temperatures, were investigated in [19]. A micro combined cold, heat and power system for a small household with a TEC as the cooler and achieving a cooling power of 26.8 W, was presented in [20]. In [21], a 3D printable TEG device architecture with a high thermocouple density of 190 per cm² by using a thin substrate as an electrical insulation between the thermoelectric elements, resulted in a high-power output of 47.8 μW/cm^2 from a 30 K temperature difference. A stove-powered TEG (SPTEG) was used in [22] to generate power from waste heat released during cooking. They researched series and parallel TEGs connections and the effect of pressure to address low power output due to irregular temperature. Finally, an experimental and a numerical investigations on TECs for comparing air-to-air and air-to-water refrigeration were investigated in [23], with the findings revealing that air-to-water achieves 30-50% efficiency, compared to air-to-air cooling.

These are just a few noted studies; however, lacking in the TEGs/TECs literature are comprehensive details on their maths, modeling and operations when connected in series and also in parallel to increase the output power (in the case of TEG) and the cooling power (in the case of TEC). This article therefore, expands on i) developing and expressing further the theoretical maths covering TEGs and TECs various parameters/modules with focus on the total internal resistance, ii) the modeling of multiple TEGs and TECs modules focusing on their electrical parameters and finally iii) their static and dynamic simulations with focus on the optimal operation points investigation as well as the interpretations thereof. The results are then validated with established published studies and concluding remarks are drawn.

2. TEGs and TECs Mathematical Analyses and Modeling

In [9], [11] and [12], the standard static mathematics defining various TEG and TEC parameters as well as their modeling are demonstrated. We developed further and present in the following sections: i) TEGs and TECs maths and ii) the implemented models (based on their maths) using Matlab/Simulink and the simulations of TEG/TEC modules, be it in series and or in parallel connections.

2.1. TEGs and TECs Steady-state Mathematical Analyses

The derivations thus far of the TEG and TEC parameters have been based-on the p-n junction thermoelement resistance at the thermocouple level and by extension at the module level as indicated in [9], [11] and [12]. However, in practice, more than one TEG and TEC modules will be needed for more power production and this will take the form of series and or parallel connections; as a result, the electrical resistance will often change. This section redefines the change in R to R_t and is articulated next.

(I) *TEGs Steady-state Mathematical Analysis*

The following TEG parameters mathematics are developed and presented step-wise for multiple TEGs case as follows:

- Thermoelectric (TE) device p-n junction thermocouple resistance (r)

The TE device p-n thermocouple resistance r in ohm is:

$$r = \rho L/A \quad (\Omega) \tag{1}$$

with ρ being the TEG/TEC electrical resistivity in Ω.m, L is the length in (m) of the TEG/TEC p-n thermocouple and the TEG/TEC p-n thermocouple area is A in metre squared (m^2).

- TE device (TEG and TEC) module resistance (R)

The resistance in (Ω) of a TEG/TEC module is computed as:

$$R = nr \quad (\Omega) \tag{2}$$

where n (which differs, could be 100, 127, 199, 255 etc) is a TEG/TEC manufacturer p-n thermocouples amount used in a TEG/TEC. The more the n, the more powerful is the TEG/TEC.

- TEG/TEC module(s) total resistance (R_t)

The total resistance R_t in (Ω) of a TEG/TEC module(s) is simply calculated as:

$$R_t = n\frac{T_s}{T_p}r = R\frac{T_s}{T_p} \quad (\Omega) \tag{3}$$

with T_p being the TEGs/TECs (TEG/TEC modules) amount connected in parallel and T_s the TEGs/TECs (TEG/TEC modules) amount connected in series. NB: all the TEGs/TECs used in (3), have to be identical model to make sure the R of each TEG/TEC is not vastly different; if not, (3) would be inaccurate.

- TEG(s) output voltage (V_o)

The TEG(s) voltage generated in volt, can be derived as:

$$V_o = nS\Delta T - IR_t \quad (V) \tag{4}$$

with S being the TE device Seebeck coefficient in V/K, $\Delta T = T_h - T_c$ the TEG(s) temperature difference in kelvin or °C and the output current of the TEG(s) is I in ampere.

- TEG(s) output current (I)

The TEG(s) generated current I in ampere is deduced as:

$$I = \frac{nS\Delta T}{R_L + R_t} \quad (A) \tag{5}$$

with R_L being the resistance of the electrical load connected to the TEG(s) output. NB: more I causes the TEG(s) more Joule heating, which negatively affects the TEGs efficiency.

- TEG(s) hot-side heat absorbed (Q_h)

TEG(s) produce DC power when their hot-side is at a high temperature T_h, during which the TEG(s) becomes hotter and the absorbed heat in watt is Q_h, given as:

$$Q_h = n[(SIT_h) + (K\Delta T)] - 0.5I^2R_t \quad (W) \tag{6}$$

with K being the TEG(s) thermal conductance in W/K.

- TEG(s) cold-side heat emitted (Q_c)

TEG(s) produce DC power when the cold-side of the TEG(s) is at a low temperature T_c releasing the heat Q_c in watt.

$$Q_c = n[(SIT_c) + (K\Delta T)] + 0.5I^2R_t \quad (W) \tag{7}$$

- TEG(s) output power (P_o)

The TEG(s) modules generated power P_o in watt, is found variously as follows:

$$P_o = Q_h - Q_c \quad (W) \tag{8}$$
$$P_o = IV_o = n\,[(SI\Delta T)] - I^2R_t \quad (W) \tag{9}$$

- TEG(s) electrical/conversion/thermal efficiency (η)

η is the TEG(s) power output P_o divided by the TEG(s) hot-side heat absorbed Q_h. η being a performance parameter is:

$$\eta = P_o/Q_h \tag{10}$$

The conversion efficiency details is presented later.

- TEG/TEC Carnot's efficiency (η_c)

Carnot efficiency is the efficiency determined based-on the temperatures T_h and T_c.

$$\eta_c = \frac{\Delta T}{T_h} = \frac{T_h - T_c}{T_h} = 1 - \frac{T_c}{T_h} \tag{11}$$

- TEG(s) conversion efficiency expression (η_e)

Simply, η_e is the raw expression of η. That is, when equations of Q_h and P_o (respectively (6) and (8) or (9)) are both substituted in (10).

$$\eta_e = \eta_c \frac{(nR_L/R_t)}{[(1+nR_L/R_t) - 0.5\eta_c + ((1/(2Z\bar{T}))(1+nR_L/R_t)^2(1+T_c/T_h))]} \tag{12}$$

with $Z\bar{T}$ being the TE device average dimensionless merit figure. NB: Z is the TE device merit figure in per K (K^{-1}) and $\bar{T} = (T_h + T_c) / 2$, is the TE device average temperature in K.

- TEG(s) maximum conversion efficiency (η_m)

η_m is the efficiency of the TEG(s) at $R_t / R_L = \sqrt{1 + Z\bar{T}}$. The η_m expression as a function of TEG temperatures and Z is:

$$\eta_m = \eta_c\left(\frac{(\sqrt{1+Z\bar{T}}) - 1)}{(\sqrt{1+Z\bar{T}} + (T_c/T_h))}\right) \tag{13}$$

- TEGs maximum power conversion efficiency (η_{mp})

As a function of temperatures and Z, η_{mp} is the efficiency of the TEG at its maximum output power P_o − that is, at $R_t = R_L$.

$$\eta_{mp} = \eta_c/[2 - 0.5\eta_c + (2/Z\bar{T})(1+T_c/T_h)] \tag{14}$$

- TEG(s) maximum power output (Po_{max})

The TEG(s) maximum transfer of power theoretically happens at $R_t = R_L$. NB: in practice, $R_t = R_L$ is hardly ever the case.

$$Po_{max} = (nS\Delta T)^2(R_L/R_t)/R(1+(R_L/R_t))^2 \quad (W) \tag{15}$$

- TEG(s) maximum voltage output (Vo_{max})

TEG(s) Vo_{max} happens at open circuit, that is when R_L is not connected or R_L is infinity (extremely large), $I = 0$A.

$$Vo_{max} = nS(T_h - T_c) = nS\Delta T \quad (V) \tag{16}$$

- TEG(s) maximum current output (I_{max})

TEG(s) I_{Max} happens at short circuit – meaning, when the load R_L is 0Ω. NB: R_t will therefore ideally be the sole resistance.

$$I_{Max} = nS\Delta T/R_t = nS(T_h - T_c)/R_t \quad (A) \tag{17}$$

- TEG(s) generated current normalized (I_n)

I_n is the normalized current of the TEG(s) in the range $0 \leq I_n \leq 1$. At the TEG(s) maximum transfer of power ($R_t = R_L$), I_n =

0.5. Simply, I_n is the TEG(s) generated current divided by the TEG(s) maximum current output. It is calculated as:

$$I_n = \frac{I}{I_{Max}} = \frac{R_t}{R_t + R_L} \quad (18)$$

- TEG(s) generated voltage normalized (V_n)

V_n is the normalized voltage of the TEG(s) ranging from 0 ≤ V_n ≤ 1. At the TEG(s) maximum transfer of power (i.e. R_L=R_t), V_n = 1/2. V_n is the TEG(s) voltage generated divided by the TEG(s) maximum (ideal) voltage generated. It is given as:

$$V_n = \frac{V_o}{Vo_{max}} = \frac{R_L}{R_L + R_t} \quad (19)$$

- TEG(s) output power normalized (P_n)

P_n is the normalised TEG(s) power bounded between 0 ≤ P_n ≤ 1. P_n = 1 at the TEG(s) maximum transfer of power (R_L=R_t). P_n is the TEG(s) power generated divided by the TEG(s) maximum output power. It is expressed as:

$$P_n = \frac{P_o}{Po_{max}} = \frac{4(R_L/R_t)}{[(R_L/R_t)+1]^2} \quad (20)$$

- TEG(s) conversion efficiency normalized (η_n)

η_n is the conversion efficiency of the TEG(s) in the region 0≤ η_n ≤1. η_n depends on R_t/R_L, T_c/T_h and $\overline{ZT}$. η_n is the conversion efficiency of the TEG(s) divided by the maximum conversion efficiency of the TEG(s), deduced as:

$$\eta_n = \eta/\eta_m \quad (21)$$

- TEG(s) effective Seebeck coefficient (S_e)

S_e measured in volt/kelvin, is expressed as:

$$S_e = 4Po_{max}/(nI_{max}\Delta T) \quad \text{(V/K)} \quad (22)$$

- TEG(s) effective electrical resistivity (ρ_e)

ρ_e measured in ohm metre, is found using:

$$\rho_e = 4[(A/L)Po_{max}]/nI_{max}^2 \quad (\Omega.\text{m}) \quad (23)$$

- TEG(s) effective figure of merit (Z_e)

Z_e measured in per kelvin, is computed as:

$$Z_e = [(2/\overline{T})(1+(T_c/T_h))]/[\eta_c((1/\eta_{mp})+0.5)-2] \quad (K^{-1}) \quad (24)$$

- TEG(s)/TEC(s) effective thermal conductivity (k_e)

k_e measured in watt per metre kelvin, is expressed as:

$$k_e = S_e^2/(\rho_e Z_e) \quad \text{(W/mK)} \quad (25)$$

TEGs/TECs effective parameters enable researchers to factor in TEGs/TECs system losses using maximum parameters to bridge the theoretical and measured specifications differences [9].

- TEG(s) Heat Flux Density (HFD)

HFD is the amount of heat absorbed per TEGs hot-side surface area (*TEGsa*) in watt per centimetre square.

$$HFD = Q_h/TEG_{sa} \quad (\text{W/cm}^2) \quad (26)$$

This concludes the TEG(s) modules static mathematical analysis.

(II) *TECs Steady-State Mathematical Analysis*

The following TEC parameters mathematics are examined and developed step-wise for multiple TECs case as follows:

- TEC(s) voltage input (V_{in})

The TEC(s) applied voltage in volt, is expressed as:

$$V_{in} = n[S(T_h - T_c)] + I_{in}R_t \quad \text{(V)} \quad (27)$$

where I_{in} is the TECs input current from the power supply.

- TEC(s) input current (I_{in})

The TECs input current in ampere is derived as:

$$I_{in} = \frac{nS\Delta T}{R_s - R_t} \quad \text{(A)} \quad (28)$$

where R_s is the internal source electrical resistance of the power supply connected to the TECs.

- TEC(s) cold-side heat absorbed (Q_c)

TECs create cold when their cold-side is at a low temperature T_c to absorb heat and supply a steady cooling power Q_c in W.

$$Q_c = n[(SI_{in}T_c) - (K\Delta T)] - 0.5I_{in}^2R_t \quad \text{(W)} \quad (29)$$

- TEC(s) hot-side heat emitted (Q_h)

TECs produce cold when their hot-side is at a high temperature T_h emitting the heat Q_h in watt.

$$Q_h = n[(SI_{in}T_h) - (K\Delta T)] + 0.5I_{in}^2R_t \quad \text{(W)} \quad (30)$$

- TEC(s) power input (P_{in})

The applied power P_{in} in watt required to power the TECs, is calculated variously as follows:

$$P_{in} = Q_h - Q_c = n[(SI_{in}\Delta T)] + I_{in}^2R_t \quad \text{(W)} \quad (31)$$

$$P_{in} = I_{in}V_{in} \quad \text{(W)} \quad (32)$$

- TEC(s) coefficient of performance (*CoP*)

This is TECs cooling power Q_c divided by its input power P_{in}.

$$CoP = Q_c/P_{in} \quad (33)$$

- TEC(s) CoP Expression (CoP_e)

CoP_e is the raw expression of CoP when the equations of Q_c and P_{in} (respectively (29) and (31) or (32)) are put in (33).

$$CoP_e = \frac{[(SI_{in}T_c) - (K\Delta T) - (0.5I_{in}^2R_t/n)]}{[(SI_{in}\Delta T) + (I_{in}^2R_t/n)]} \quad (34)$$

- TEC(s) current to yield CoP (I_{cop})

I_{cop} is the TECs input current in (A) needed to attain *CoP*.

$$I_{cop} = \frac{nS\Delta T}{R_t[(\sqrt{1+Z\bar{T}})-1]} \quad (A) \qquad (35)$$

- TECs maximum CoP (CoP_{max})

CoP_{max} is the TECs maximum CoP that can be achieved.

$$CoP_{max} = \frac{[T_c/\Delta T]\left((\sqrt{1+Z\bar{T}}) - \frac{T_h}{T_c}\right)}{((\sqrt{1+Z\bar{T}})+1)} \qquad (36)$$

- TEC(s) maximum cooling power current (Icp_{max})

Icp_{max} is TECs current in ampere needed to realise max Q_c.

$$Icp_{max} = nST_c/R_t \quad (A) \qquad (37)$$

- TEC(s) I_{cop} maximum cooling power (Qcp_{max})

Qcp_{max} in (W), is TECs maximum Q_c attained based-on I_{cop}.

$$Qcp_{max} = n[(SI_{cop}T_c) - (K\Delta T)] - 0.5I_{cop}^2R_t \quad (W) \qquad (38)$$

- TEC(s) maximum temperature difference (ΔT_{max})

TEC(s) ΔT_{max} in (K), occurs at maximum I_{in} and at $Q_c = 0$W.

$$\Delta T_{max} = \left(T_h + \frac{1}{Z}\right) - \sqrt{(T_h + \frac{1}{Z})^2 - {T_h}^2} \quad (K) \qquad (39)$$

- TEC(s) maximum input current (I_{max})

I_{max} is TEC(s) maximum input current in (A) at $Q_c = 0$W.

$$I_{max} = nS(T_h - \Delta T_{max})/R_t \quad (A) \qquad (40)$$

- TEC(s) maximum input voltage (Vin_{max})

Vin_{max} is the maximum V_{in} in volt, that produces maximum ΔT_{max} when $I_{in} = I_{max}$, R_t=0, $T_c = 0$, $Q_c = 0$ and T_h is maximum.

$$Vin_{max} = nST_h \quad (V) \qquad (41)$$

- TEC(s) maximum cooling power (Qc_{max})

Qc_{max} is the maximum absorbable heat or cooling power in watt, at $I_{in} = I_{max}$ and $\Delta T = 0°C$.

$$Qc_{max} = (nS)^2(T_h^2 - \Delta T_{max}^2)/2R_t \quad (W) \qquad (42)$$

- TEC(s) input current normalized (Iin_n)

TEC(s) Iin_n is I_{cop} divided by I_{max}.

$$Iin_n = I_{cop}/I_{max} \qquad (43)$$

- TEC(s) input voltage normalized (Vin_n)

TEC(s) Vin_n is V_{in} divided by Vin_{max}.

$$Vin_n = V_{in}/Vin_{max} \qquad (44)$$

- TEC(s) cooling power normalized (Qc_n)

TEC(s) Qc_n is Q_c divided by Qc_{max}.

$$Qc_n = Q_c/Qc_{max} \qquad (45)$$

- TEC(s) CoP normalized (CoP_n)

TEC(s) CoP_n is CoP divided by CoP_{max}.

$$CoP_n = CoP/CoP_{max} \qquad (46)$$

- TEC(s) normalized temperature difference (ΔT_n)

TECs ΔT_n, is ΔT divided by ΔT_{max} and it is expressed as:

$$\Delta T_n = \Delta T/\Delta T_{max} \qquad (47)$$

Normalized parameters give dimensionless parameters.

- TEC(s) effective Seebeck coefficient (S_e)

TECs S_e measured in VK^{-1}, is defined as:

$$S_e = 2Qc_{max}/[nI_{max}(T_h + \Delta T_{max})] \quad (V/K) \qquad (48)$$

- TEC(s) effective electrical resistivity (ρ_e)

TECs ρ_e measured in ohm metre, is written as:

$$\rho_e = AS_e(T_h - \Delta T_{max})/LI_{max} \quad (\Omega.m) \qquad (49)$$

- TEC(s) effective figure of merit (Z_e)

TECs Z_e measured in per kelvin, is given as:

$$Z_e = 2\Delta T_{max}/(T_h - \Delta T_{max})^2 \quad (K^{-1}) \qquad (50)$$

- TEC(s) midpoint current (I_{mid})

I_{mid} measured in ampere, is the mean of Icp_{max} and I_{cop}.

$$I_{mid} = 0.5(Icp_{max} + I_{cop}) \quad (A) \qquad (51)$$

- TEC(s) midpoint cooling power (Qc_{mid})

Qc_{mid} measured in watt, is expressed as:

$$Qc_{mid} = n[(SI_{mid}T_c) - (K\Delta T)] - 0.5I_{mid}^2R_t \quad (W) \qquad (52)$$

- TEC(s) midpoint input power (Pin_{mid})

Pin_{mid} measured in watt, is deduced as:

$$Pin_{mid} = n[(SI_{mid}\Delta T)] + I_{mid}^2R_t \quad (W) \qquad (53)$$

- TEC(s) midpoint CoP (CoP_{mid})

CoP_{mid} is computed as:

$$CoP_{mid} = Qc_{mid}/Pin_{mid} \qquad (54)$$

Midpoint parameters ascertain safer optimal TECs design.

- TEC(s) cold flux density (CFD)

CFD is the cold amount produced (heat absorbed) per TECs cold-side surface area ($TECsa$) in W/cm^2. It is computed as:

$$CFD = Q_c/TECsa \quad (W/cm^2) \qquad (55)$$

2.2. TEGs and TECs Modelling and Simulations

Covered in Section 2.1., are the TEGs and TECs parameters of interests —which were extensively expressed mathematically with emphasis/basis on the total internal resistance R_t—which was derived and the regular TEG/TEC equations re-expressed based-on R_t to now cover TEG(s)/TEC(s). The above equations are herein further modeled in Matlab and Simulink, to institute the TEGs and TECs models that can now be utilized to simulate and investigate many connected TEGs and TECs optimal performance.

Exemplified in Figures 1a and 1b, are the TEGs static and dynamic simulated model GUIs, from which the TEGs parameters expressed in Section 2.1.I, can all be statically and dynamically configured for an infinite amount of TEGs connections and then simulated to obtain the TEG(s) optimum operation points. Figures 1c and 1d zoom-in on the TEGs internal modeling. Figure 1e expands on the TEGs R_t modeling − this must be matched to the load resistance R_L − which can be changed before or while the simulation is running to match the TEGs R_t for maximum power transfer simulation. Figure 2 exemplifies the TECs simulator user interface. Also, multiple TECs combinations in T_s and T_p and the various parameters presented in Section 2.1.II, can be optimally simulated. Likewise, maximum power will be transferred also from the DC power supply to the TECs by matching its R_t to R_s.

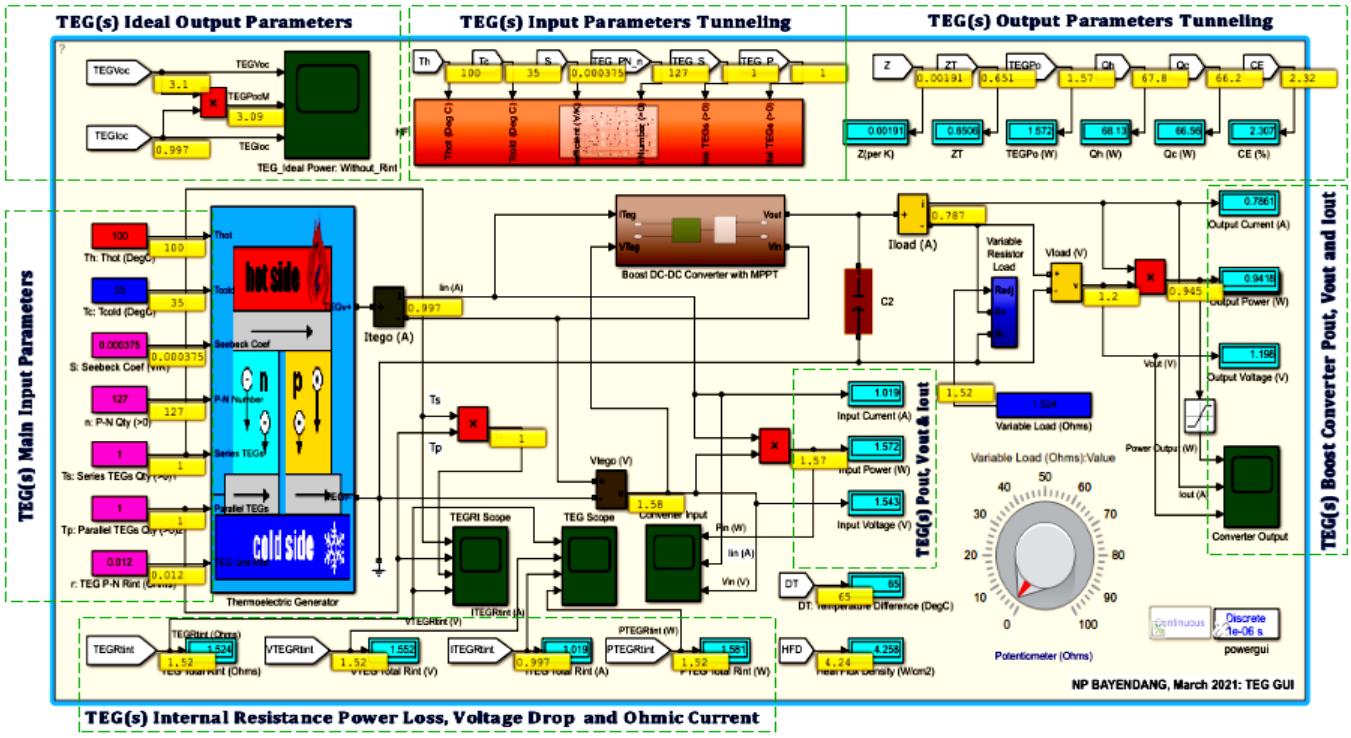

Figure 1a: TEG(s) static simulator user's interface − shows the steady-state simulation with all the input parameters fixed (though can change) over-time

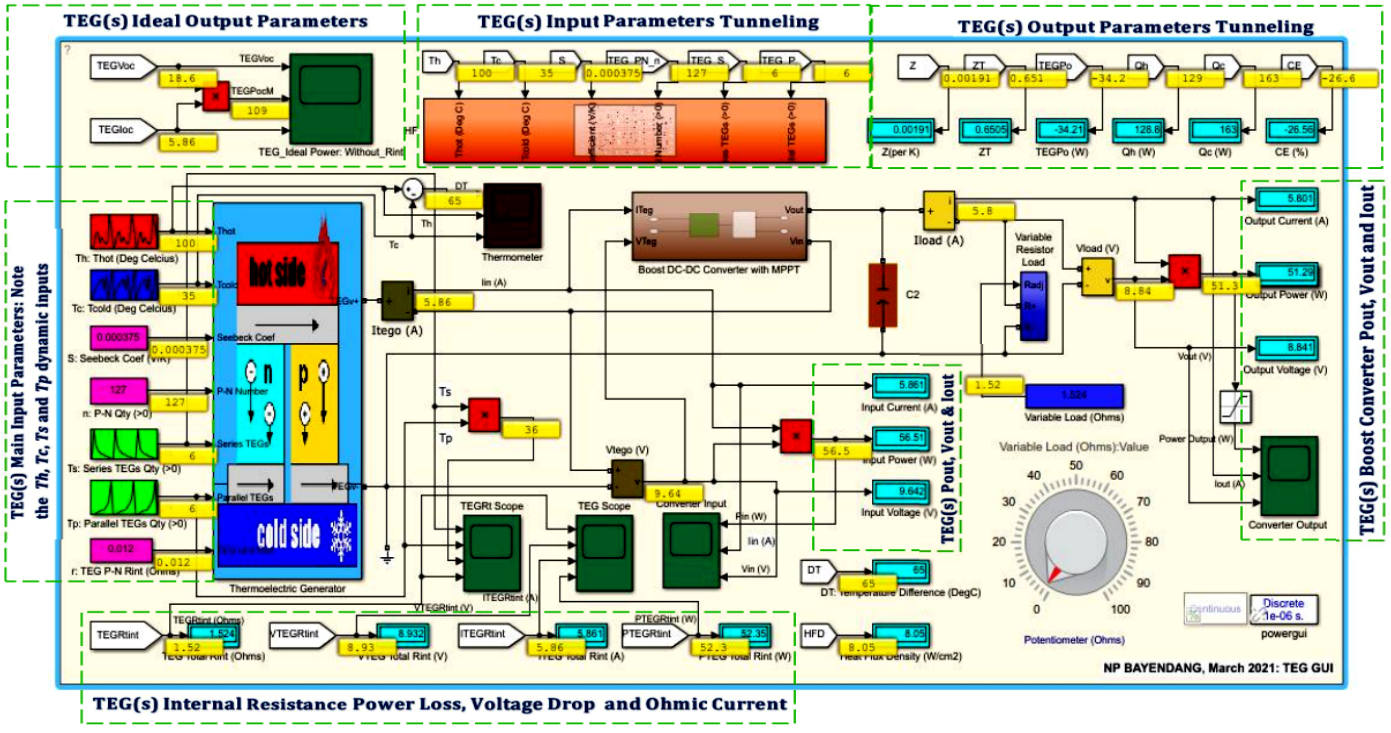

Figure 1b: TEG(s) dynamic simulator user's interface − shows the transient simulation with the T_h, T_c, T_s *and* T_p input parameters auto changing with time

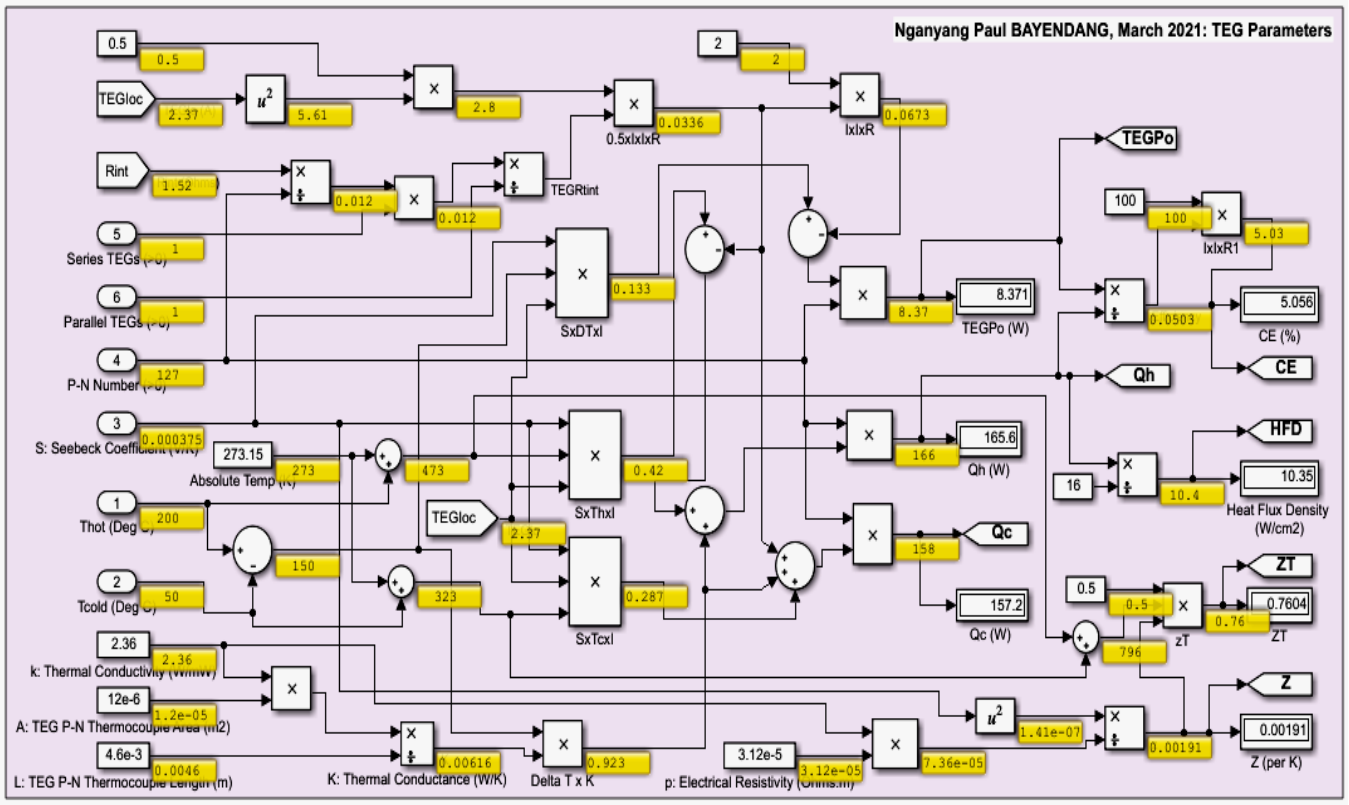

Figure 1c: TEG(s) modeling and simulation − TEG(s) parameters

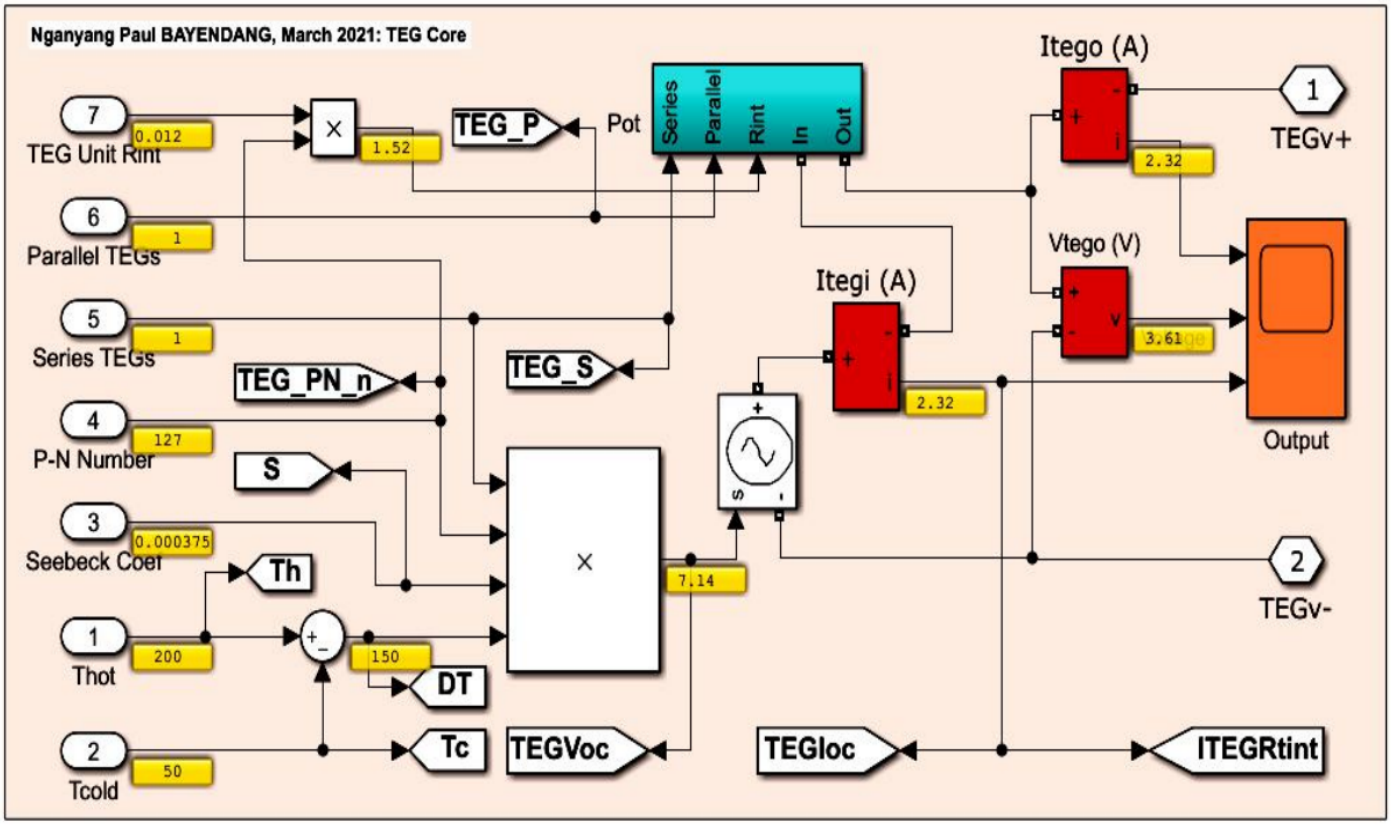

Figure 1d: TEG(s) modeling and simulation − TEG(s) engine

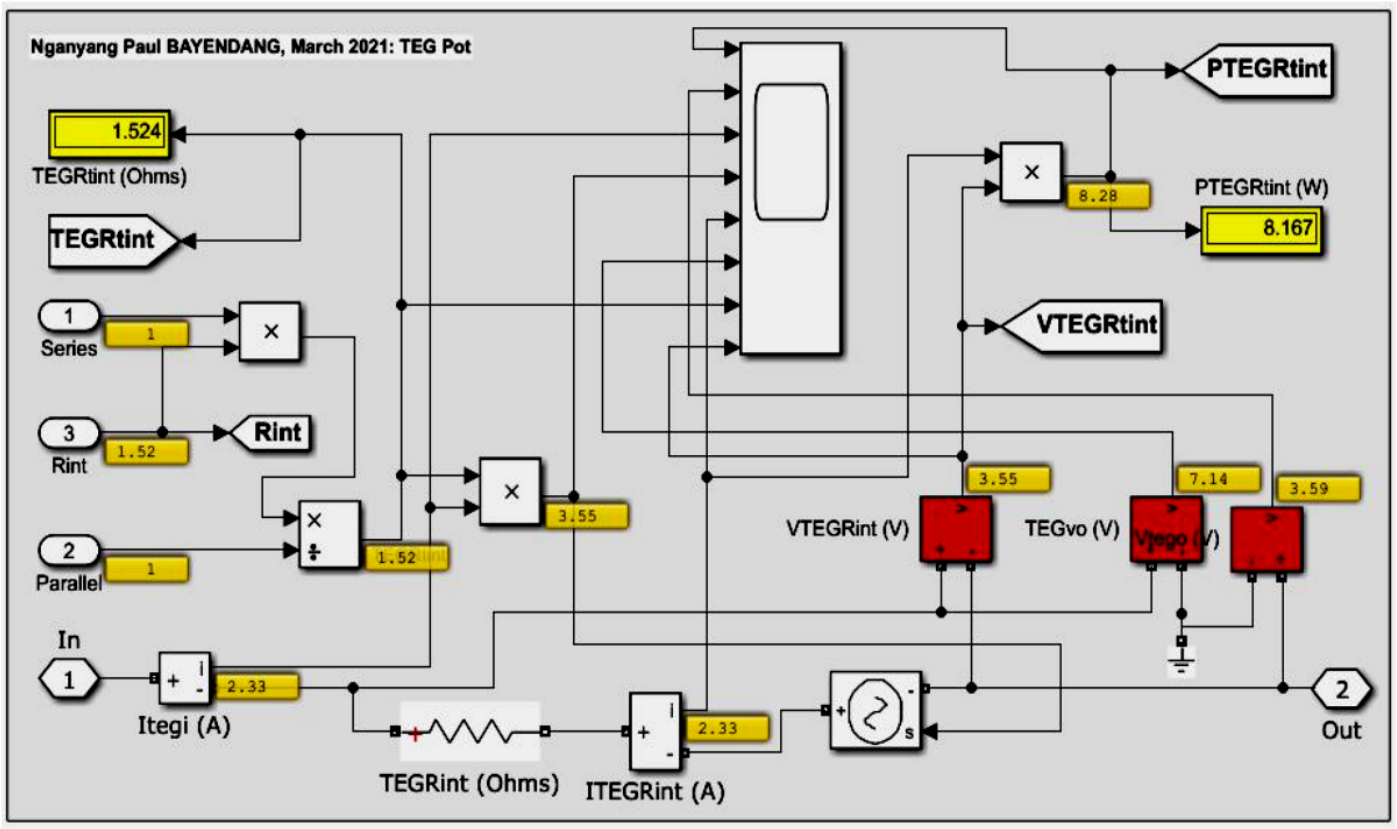

Figure 1e: TEG(s) modeling and simulation − TEG(s) automatic internal source total electrical resistance R_t

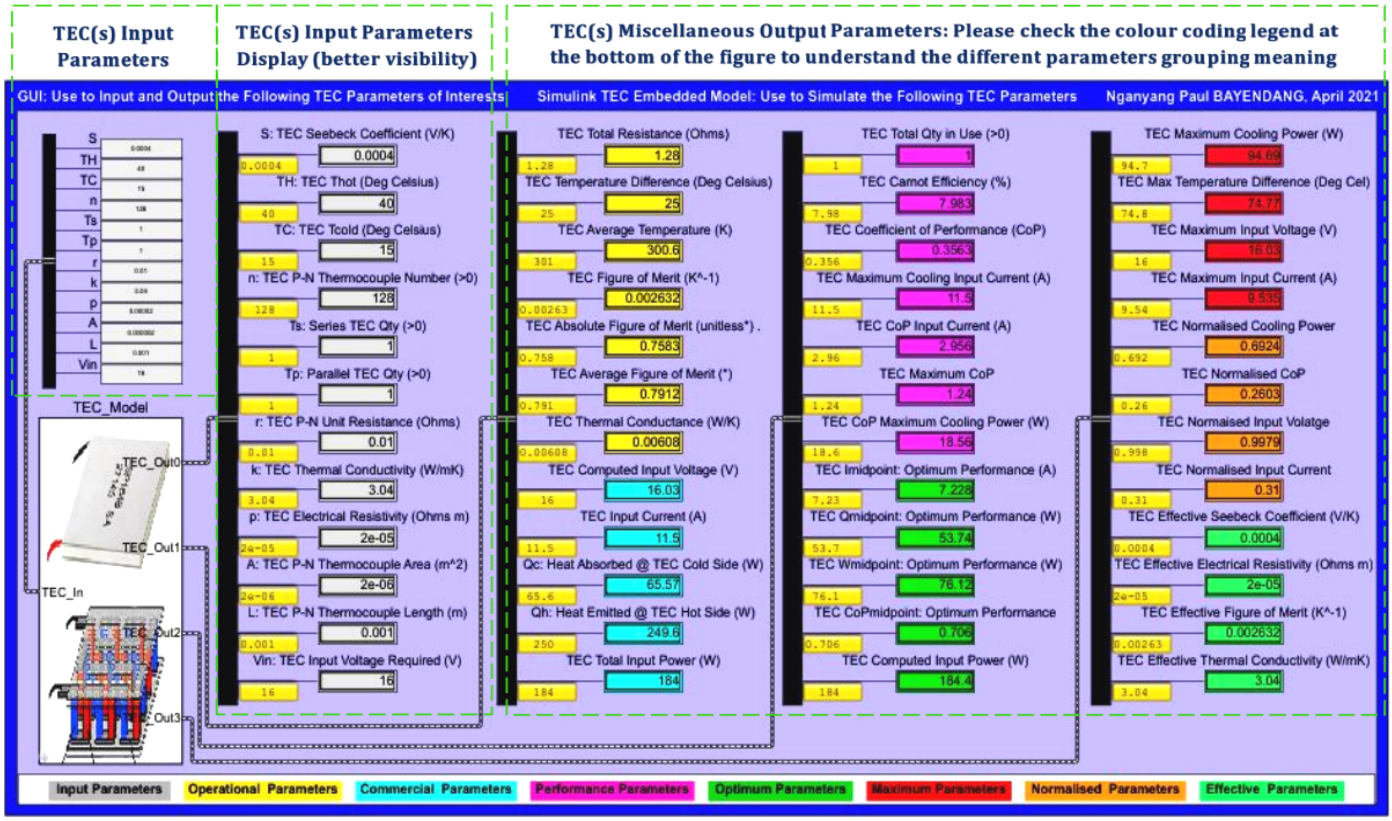

Figure 2: TEC(s) simulator − simulates TECs various parameters by inputting a TEC specific data sheet parameters and calculates its theoretical outputs

3. TEGs and TECs Simulations Results

The TEGs and TECs simulations results are presented in three parts as follows, the i) TEGs parameters static simulation results ii) TECs parameters static simulation results and iii) TEGs parameters dynamic simulation results. Understanding these parameters operation is very paramount; otherwise, doing the physical design would just be a matter of taking chances and hoping for the best − which is sometimes the case, as most designers have reported very bad design results, likely from not understanding TE devices dynamic operations and limitations. The results from investigating the TEG(s) and TEC(s) parameters optimal operation points are discussed in details in Section 4.

3.1. TEGs Parameters Static Simulation Results

Figures 3 − 6 expound the TEGs parameters simulated to determine their optimal operation points − marked in green.

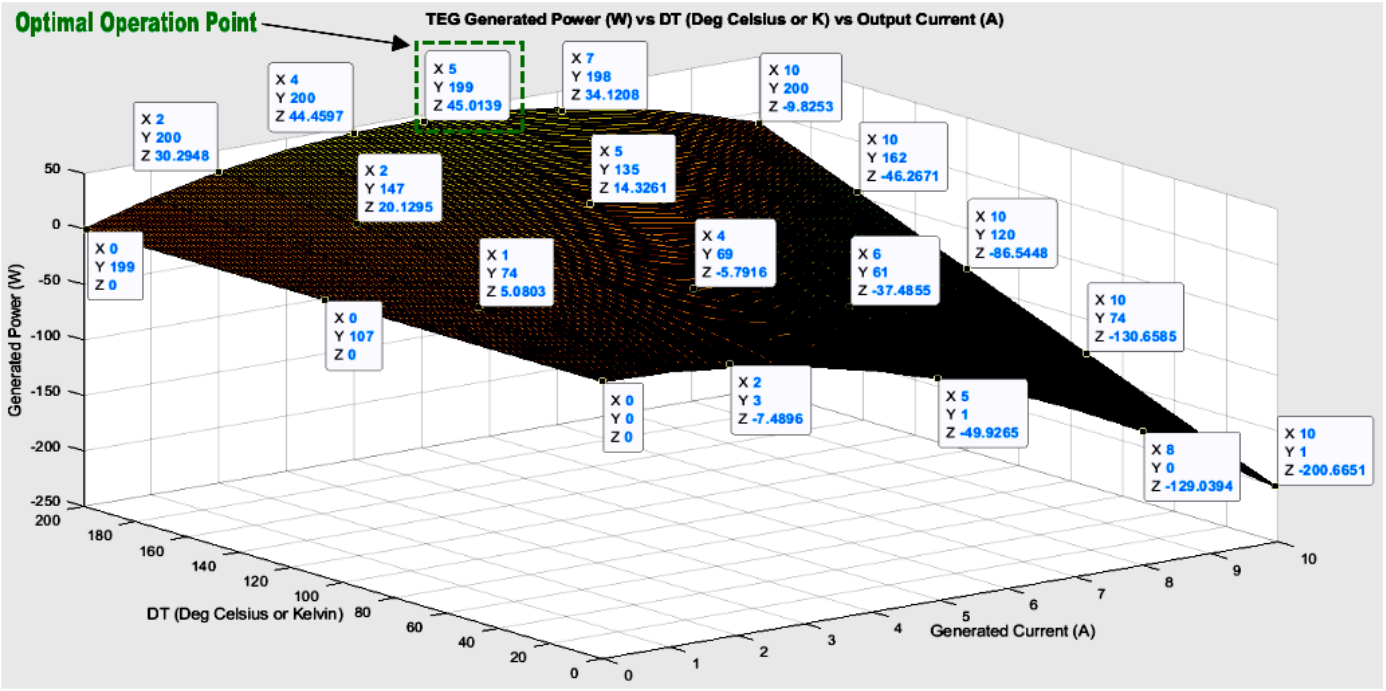

Figure 3: TEG power output P_o (W) vs temperature difference ΔT (°C) vs current output I (A)

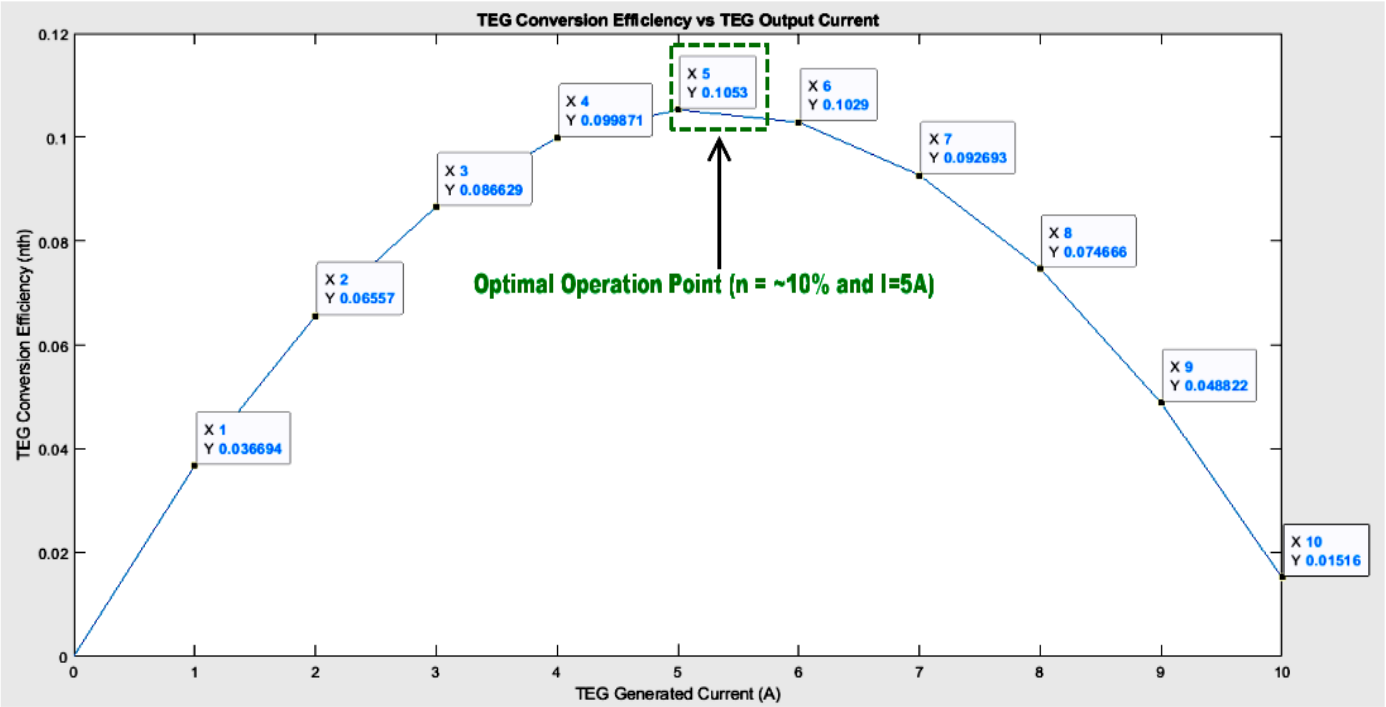

Figure 4: TEG conversion efficiency η vs current output I (A)

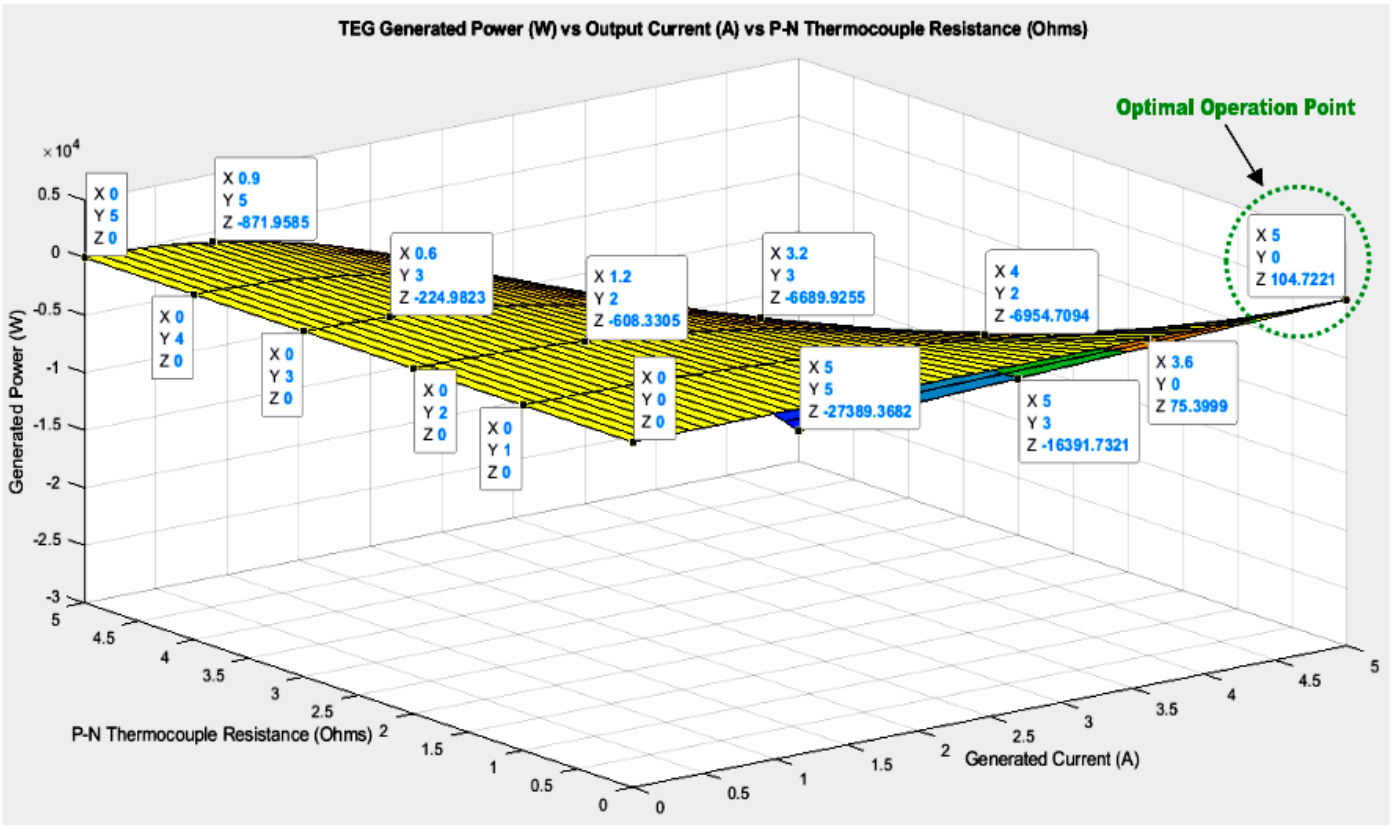

Figure 5: TEG power output P_o (W) vs r or R or R_t (Ω) vs current output I (A)

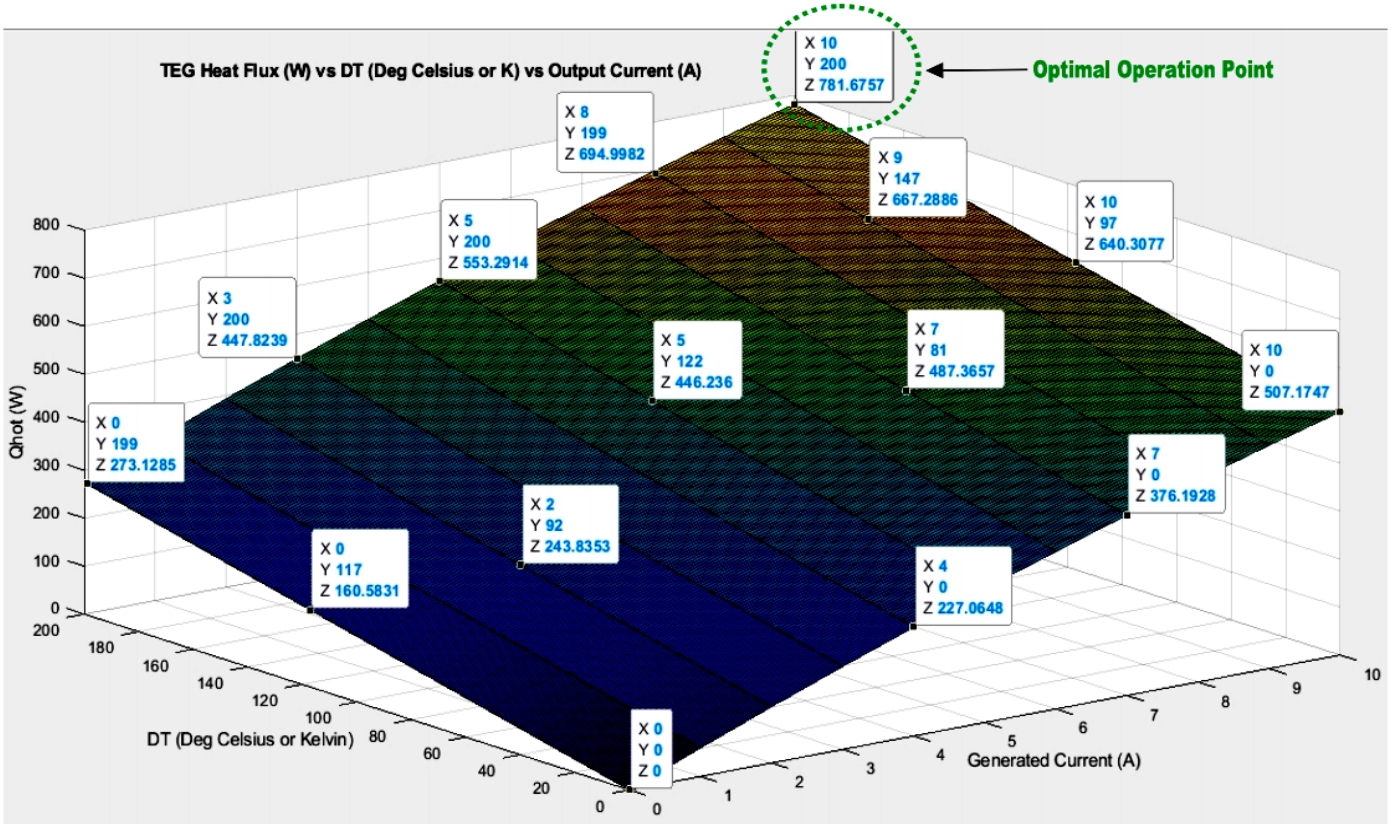

Figure 6: TEG absorbed heat Q_h (W) vs temperature difference ΔT (°C) vs output current I (A)

3.2. TECs Parameters Static Simulation Results

TECs parameters are simulated in Figures 7 - 10 to determine their possible optimal operation points − shown highlighted in red.

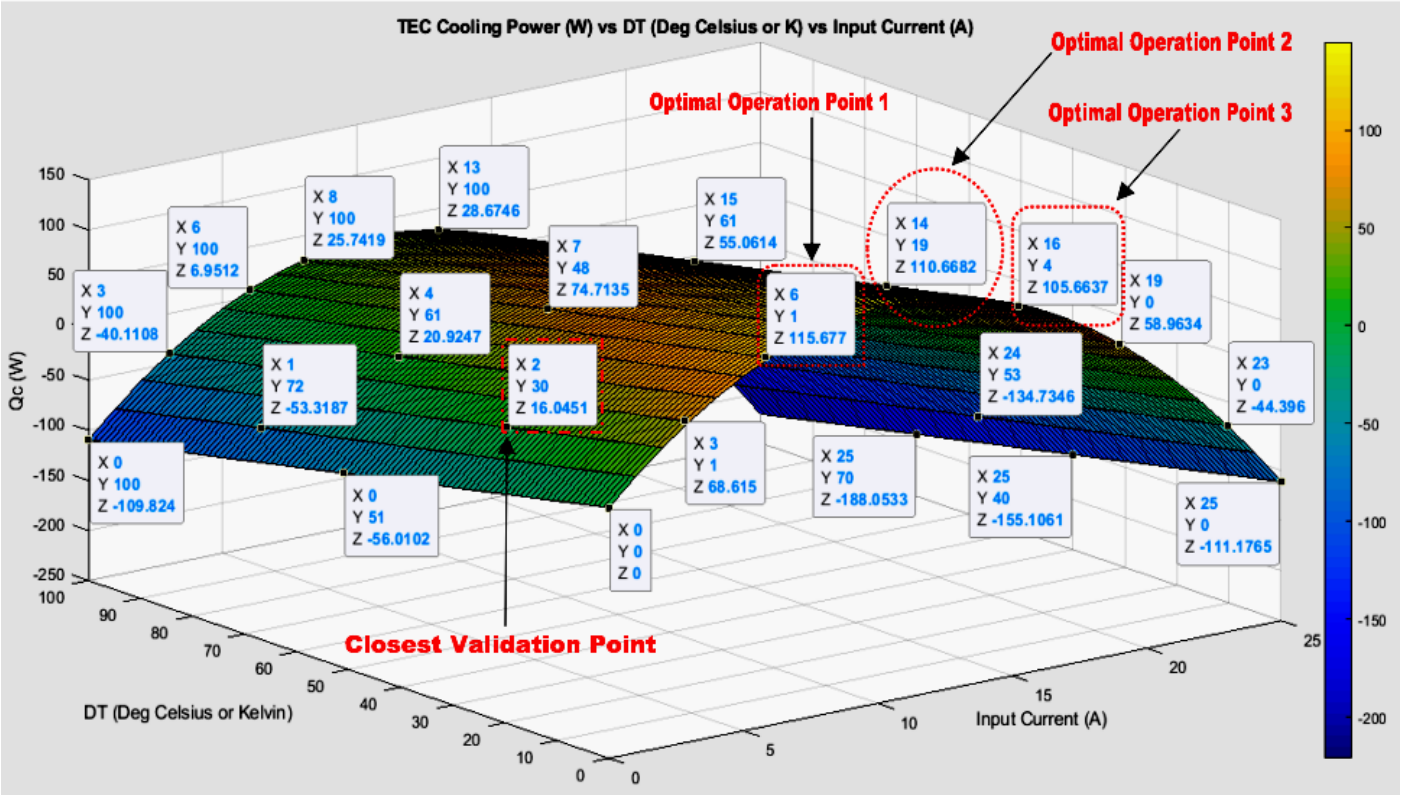

Figure 7: TEC cooling power or heat absorbed Q_c (W) vs temperature difference ΔT (°C) vs input current I_{in} (A)

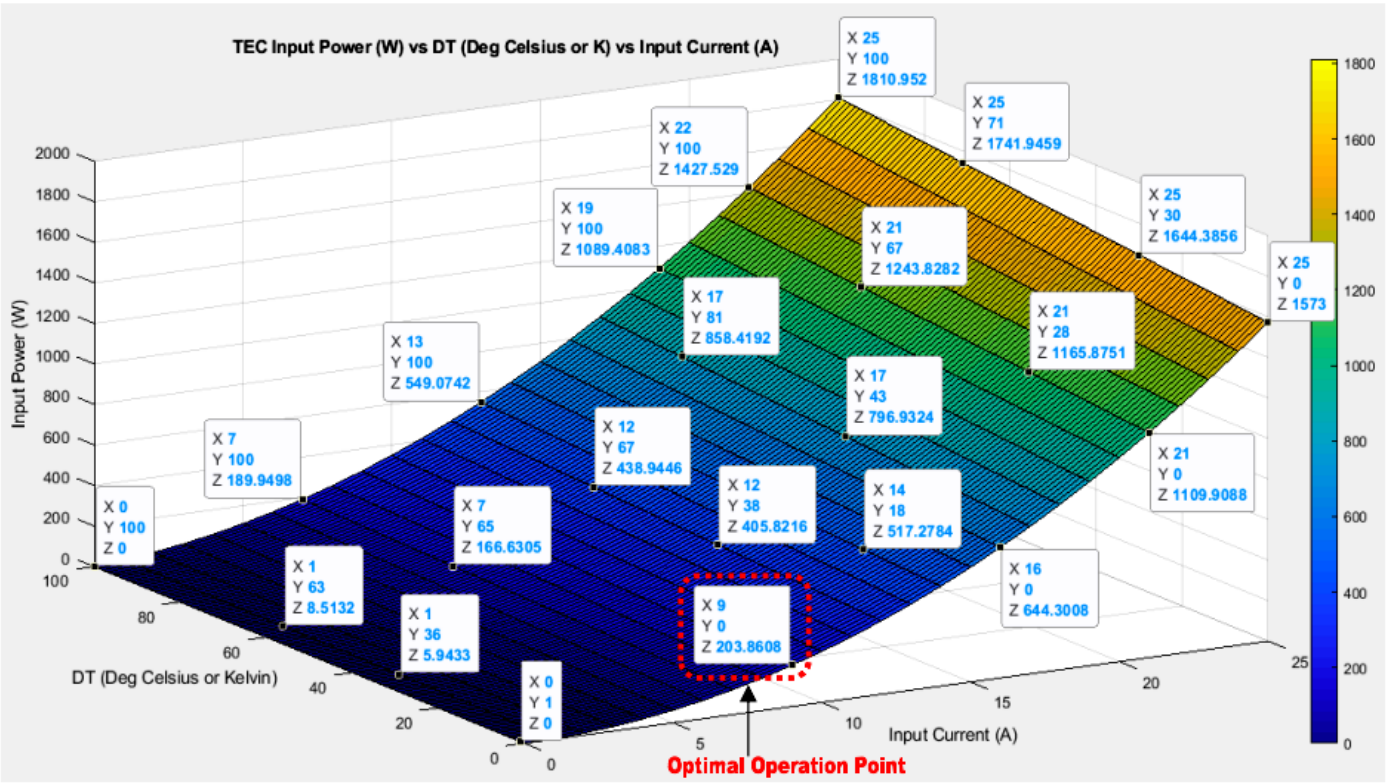

Figure 8: TEC input power P_{in} (W) vs temperature difference ΔT (°C) vs input current I_{in} (A)

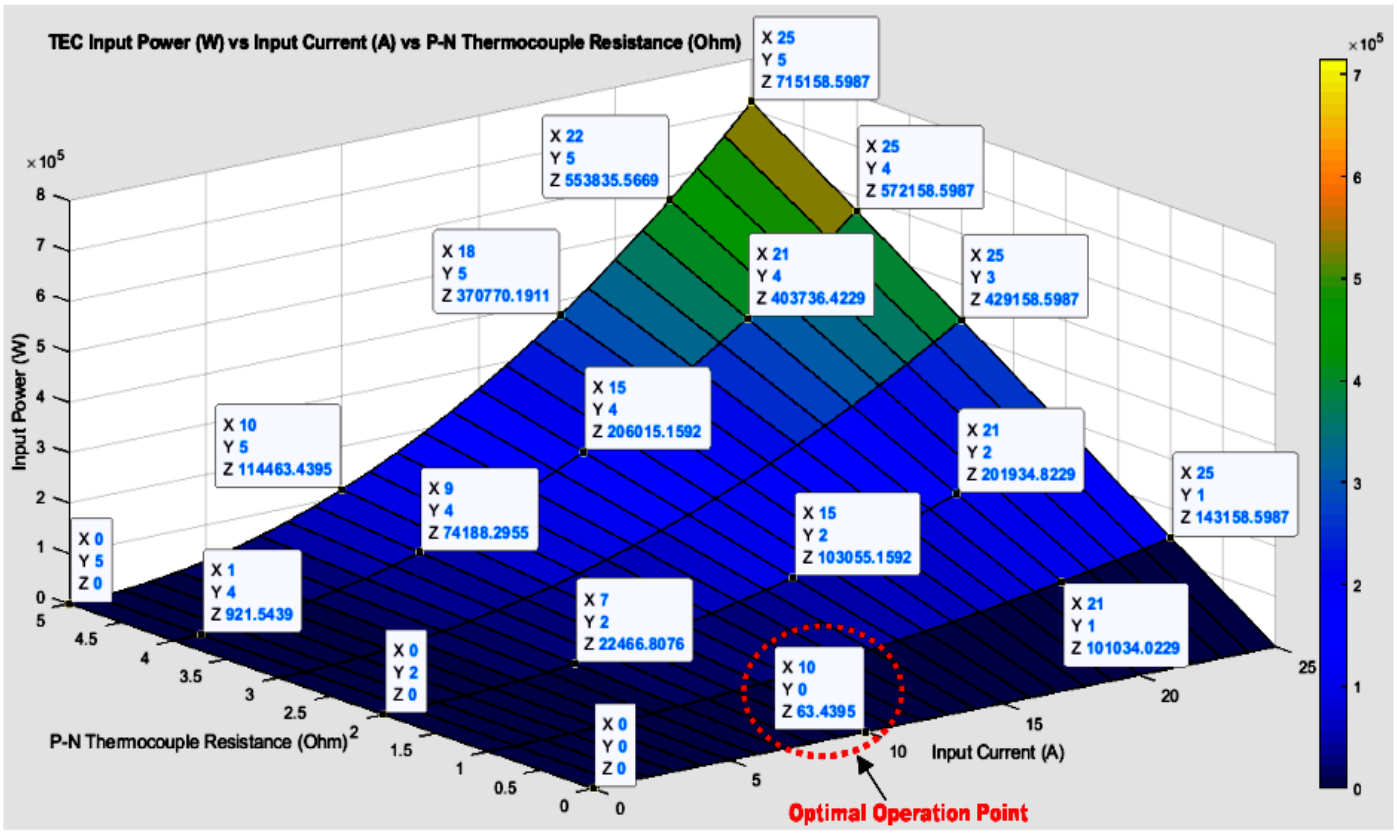

Figure 9: TEC input power P_{in} (W) vs internal resistance r or R or R_t (Ω) vs input current I_{in} (A)

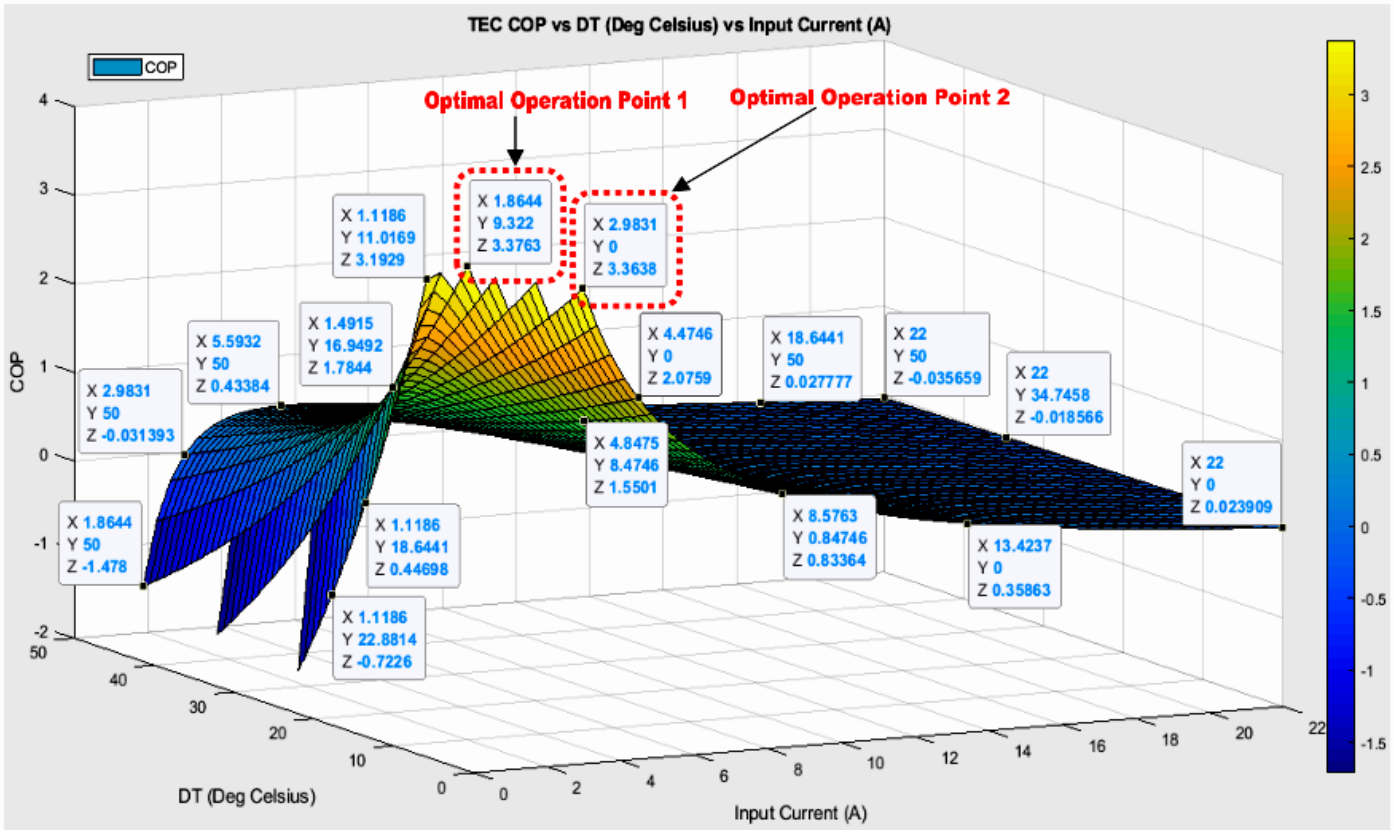

Figure 10: TEC coefficient of performance CoP (%) vs temperature difference ΔT (°C) vs input current I_{in} (A)

3.3. TEGs Parameters Dynamic Simulation Results

TEG modules temperatures, its series and parallel connections dynamic simulation results are depicted in Figures 11a - 11f.

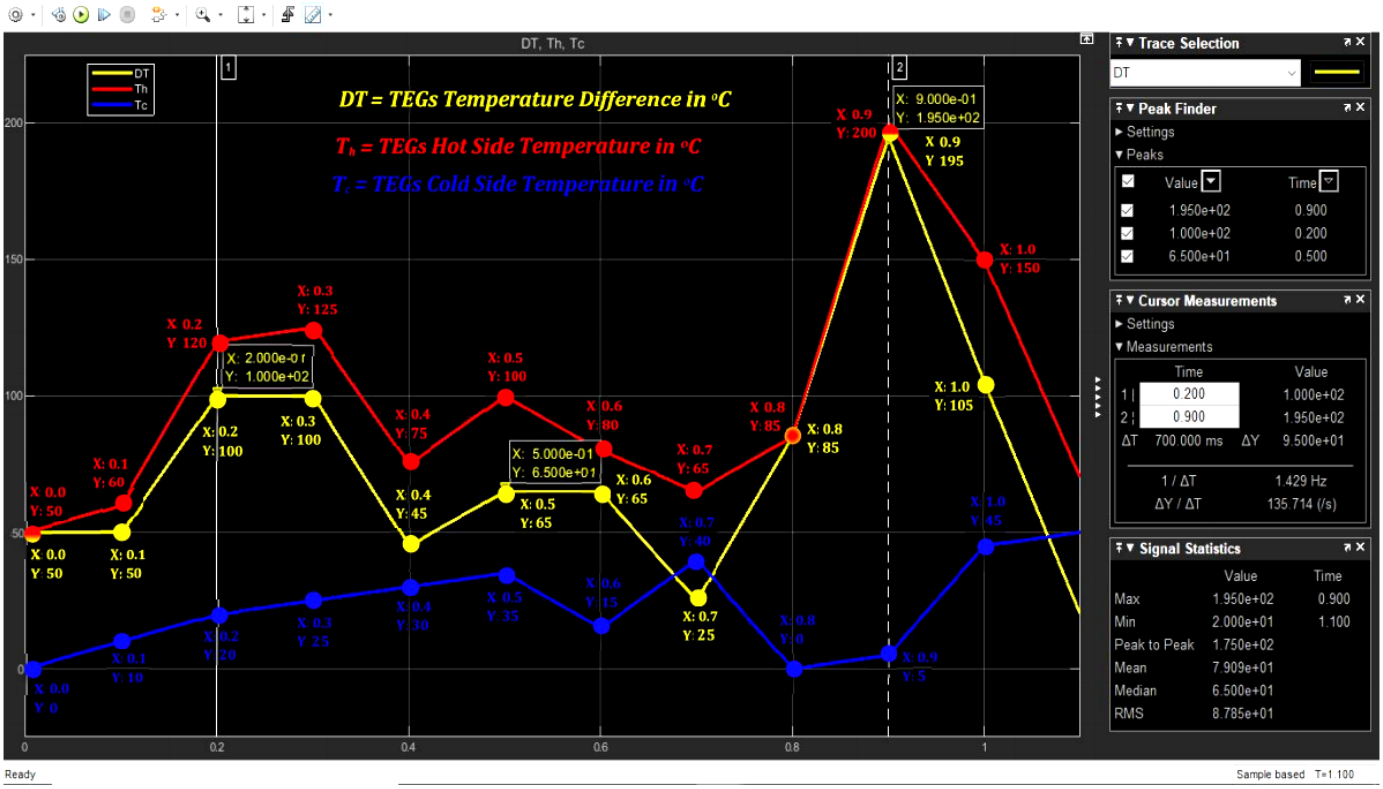

Figure 11a: 36 TEGs hot (T_h) and cold (T_c) temperatures as well as temperature difference (DT) dynamics − temperature changes as simulation progresses

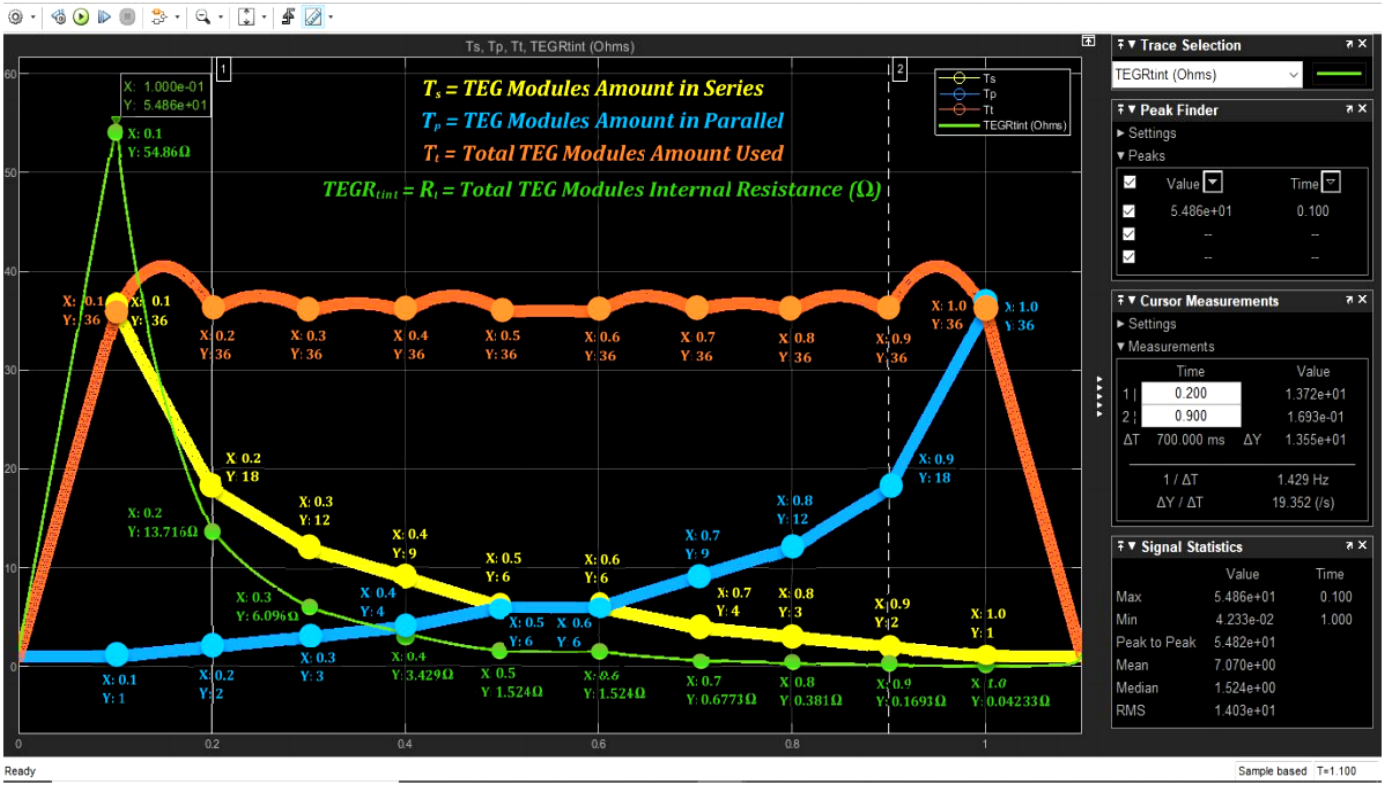

Figure 11b: TEGs in series (T_s), parallel (T_p) and total internal resistance (R_t) dynamics − 36 TEG modules simulated in 10 different auto reconfiguration

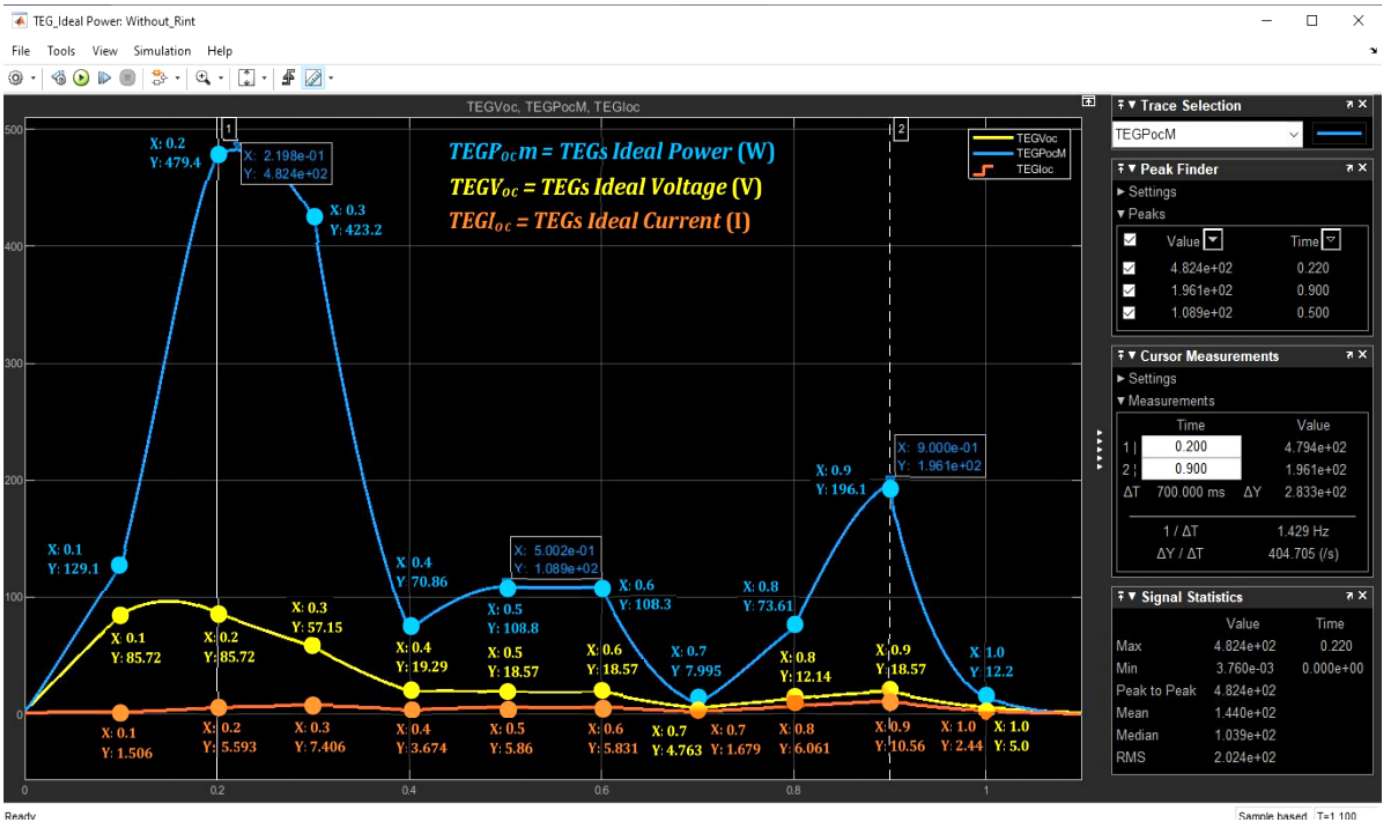

Figure 11c: 36 TEGs ideal output power, voltage and current dynamics; as TEGs temperatures and its 10 configurations change as simulation progresses

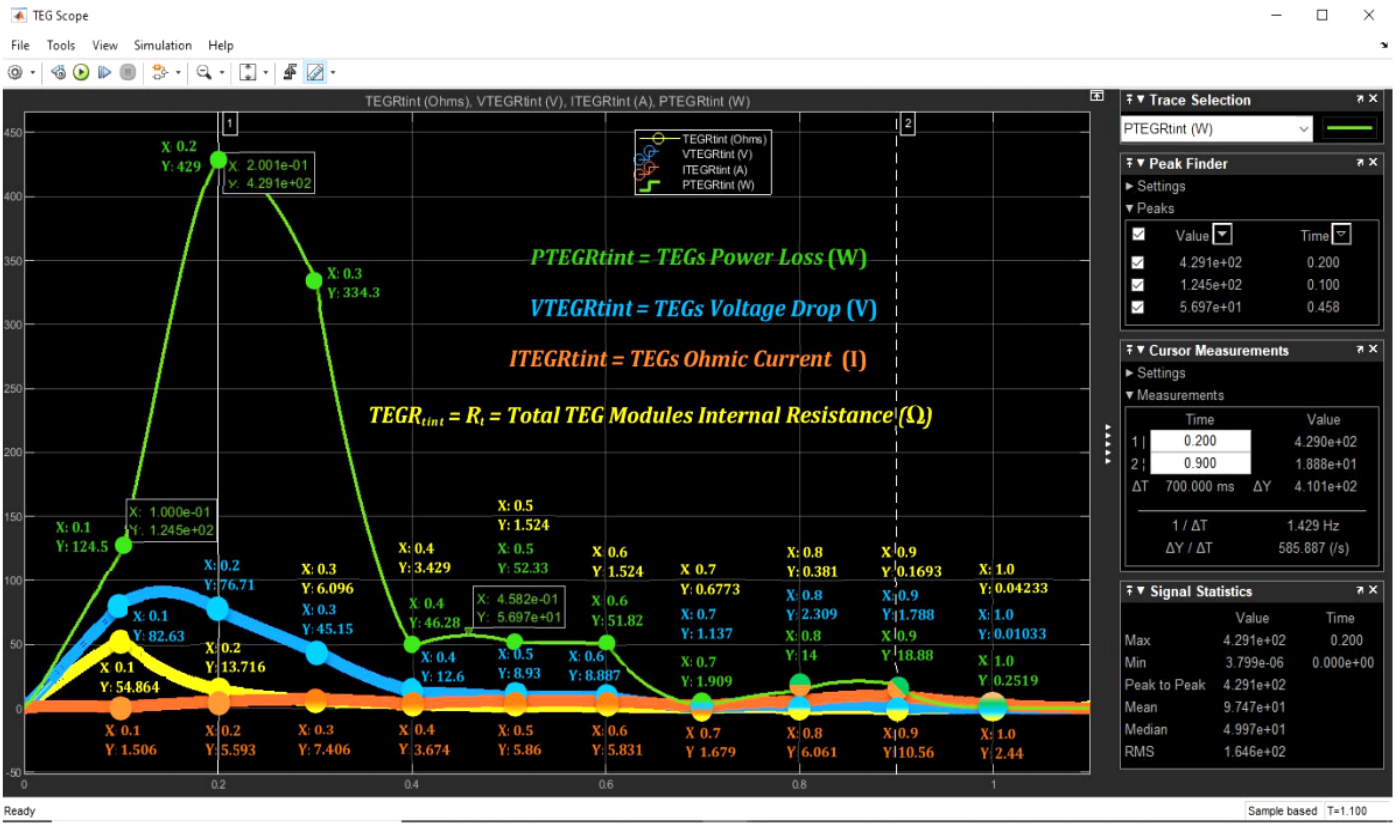

Figure 11d: 36 TEGs total internal resistance current, voltage and power losses dynamics; as the TEGs 10 configurations and temperatures auto change

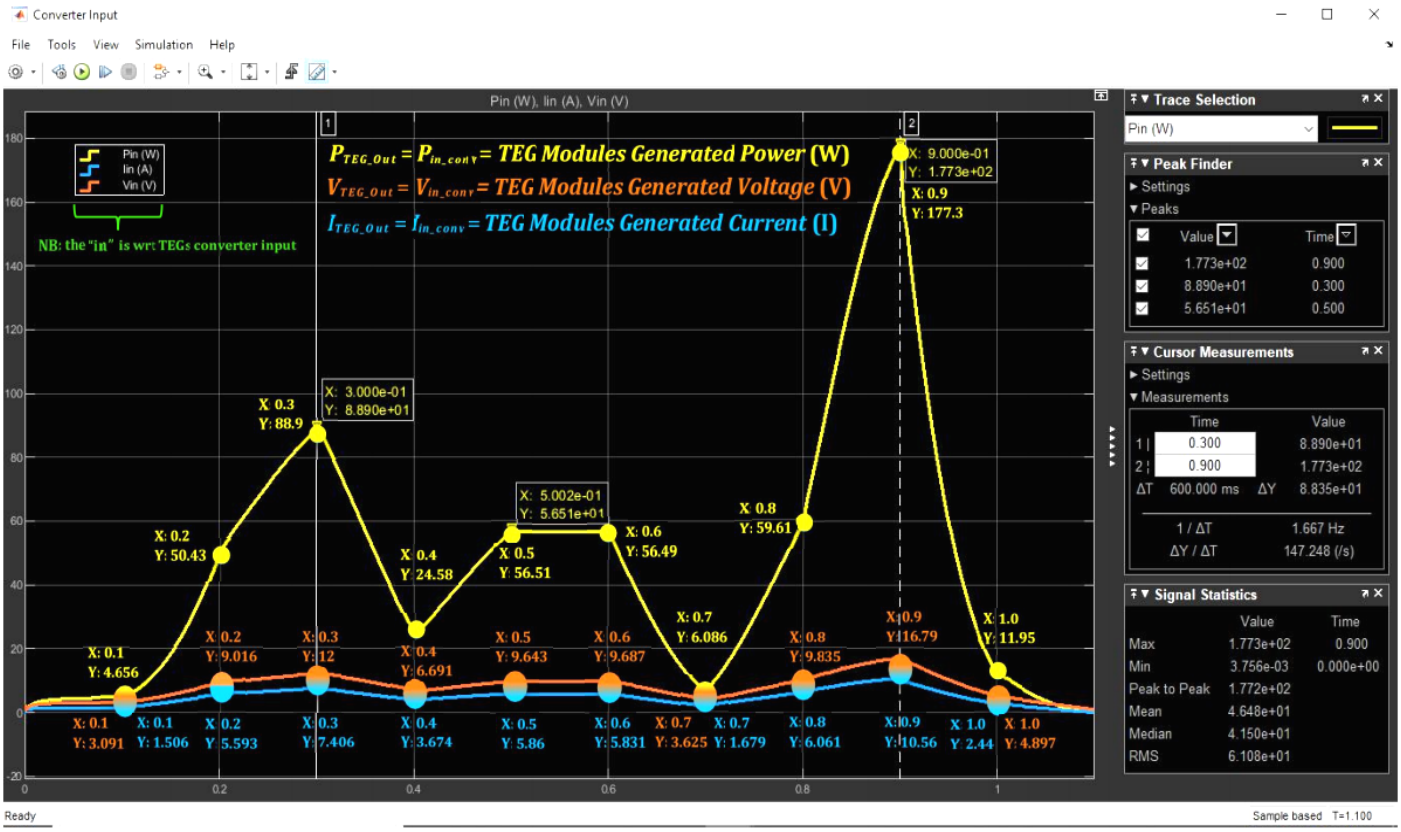

Figure 11e: 36 TEGs output power, voltage and current dynamics as the TEGs temperatures and 10 configurations auto change as simulation progresses

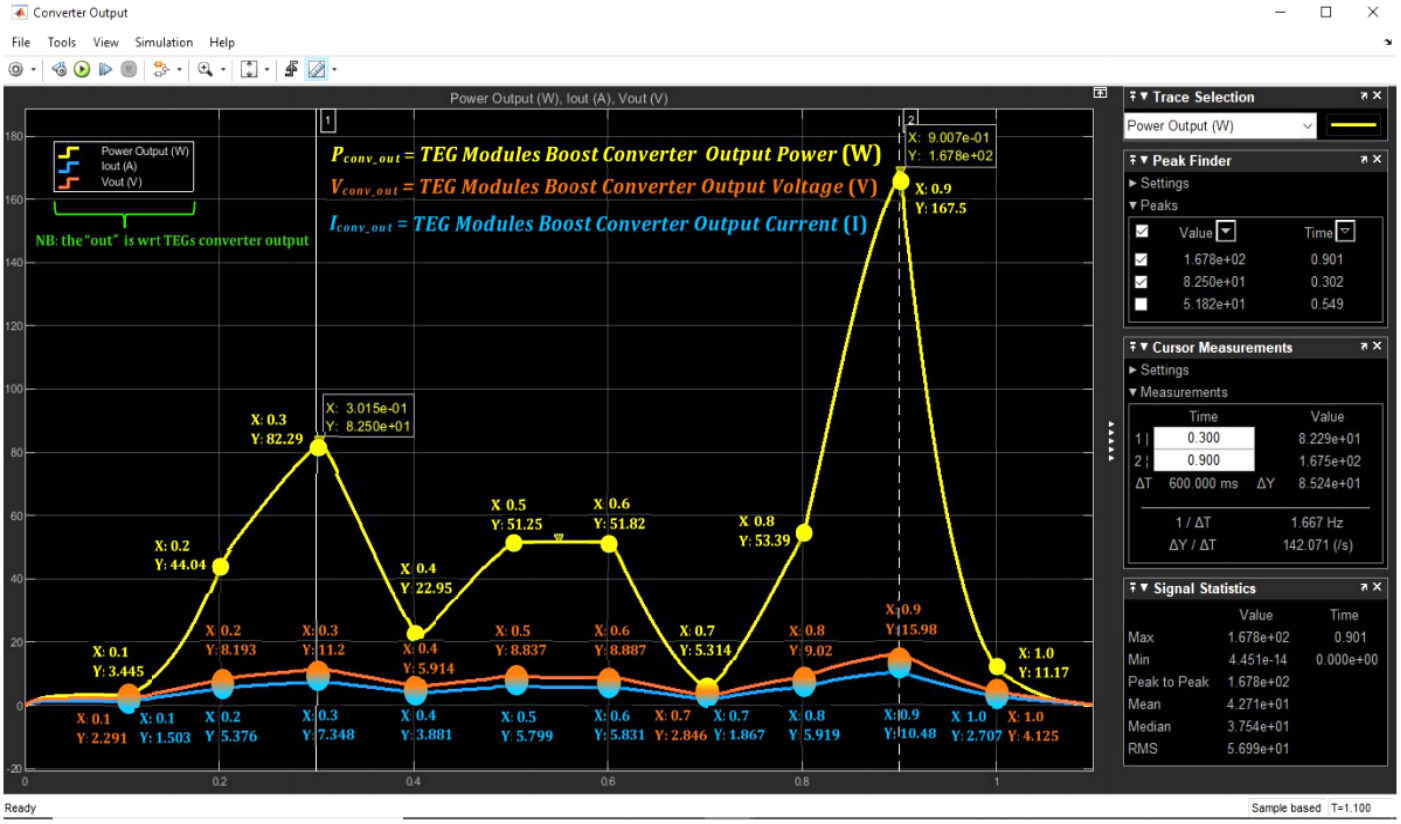

Figure 11f: 36 TEGs boost converter output power, voltage and current dynamics as the TEGs temperatures and the TEGs 10 configurations auto change

4. TEGs and TECs Simulations Results Discussions

The TEGs/TECs simulations results demonstrated in Section 3, are engaged below in their following respective sub-sections.

4.1. TEGs Parameters Static Simulation Results Discussion

Some of the crucial TEGs parameters simulated in Section 3.1. and the significance of the results are herein asserted. As exemplified in Figure 3, a TEGs generated power P_o is proportional to its temperature difference ΔT and output current I; however, I above 5A (in this case) will decrease P_o − which is because of the TEGs internal Ohmic heating as a result of the increasing output current I. The ΔT, P_o and I optimum operation points are emphasized in green in Figure 3. In Figure 4, a TEGs conversion efficiency η is directly proportional to current output I up to ~5A max (in this case) and decreases later as highlighted in green. It should be noted that η is as well directly proportional to P_o. However, a TEG P_o is reciprocally proportional to its p-n thermocouple junction resistance r and as well to its total internal resistance R_t (more than one connected TEG modules), though pro rata to I up to ~5A (in this case) as portrayed in Figure 5. At this optimal point; R_t or R is 0Ω, I is ~5A maximum and P_o is ~105W as highlighted in green. In Figure 6, the TEGs current output I is proportional directly to the TEGs absorbed heat Q_h (at temperature T_h on the TEG hot-side) which in turn is directly dependent on the TEG ΔT. Figure 6 pictured the optimum point stressed-out in green. It should be noted that these results are not specific to a particular TEGs' connections – the results just fundamentally give a holistic theoretical understanding on what TEGs physical parameters must be taken into considerations, how they are interrelated, their associated dynamics and technical limitations and how they can be practically traded-off or optimized for optimal performance when designing TEGs power supply systems.

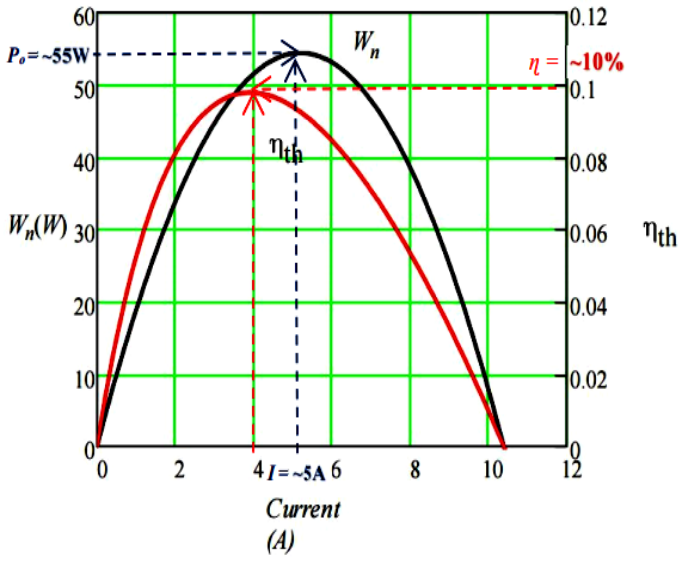

Figure 12: Validating our model with [9] − TEG (i) output power P_O = ~55W vs output current I = ~5A validating our Figure 3 result and (ii) conversion efficiency η = ~10% vs output current I = ~ 4A validating our Figure 4 result.

Depicted in Figure 12, is a result of a typical TEG model simulated with Mathcad using TEG standard specifications from typical manufacturers data-sheet as presented in [9]. This was used as a benchmark to validate our TEG model simulation accuracy − which is very close, besides a few discrepancies due to minor simulation settings differences. In light of this, our implemented TEG model can be used and developed further to simulate TEGs, including infinite series and parallel connections, which are central to our research and in large scale TEGs uses.

4.2. TECs Parameters Static Simulation Results Discussion

Some of the critical TECs parameters simulated in Section 3.2. and the importance of the results are herein articulated. Figure 7 illustrates that TECs Q_c on TECs cold-side T_c, is reciprocally proportional to ΔT but proportional directly to I_{in} up to a maximum point, after which Q_c starts dropping. The reasons are due to i) Joule heating (the more I_{in}, the more the internal heating effect) and also ii) the second law of thermodynamics − simply put, heat flows from a hotter to a colder body; in this regards, the heating caused by the increasing I_{in}, increases the TECs internal temperature up to a temperature greater than that surrounding the TECs hot-side T_h; consequently, heat now starts to flow from the TECs hot-side to its cold-side, thus making the cooling process (heat pumping) on the TECs cold-side inefficient. In Figure 7 and highlighted in red, the Q_c, ΔT and I_{in}; display three optimal operation points depending on the TECs design constraints/ priorities. In option 1, Q_c is 115.677W with a ΔT of 1°C and I_{in} of 6A. In option 2, Q_c is 110.668W with a ΔT of 19°C and I_{in} of 14A. In option 3, Q_c is 105.664W with a ΔT of 4°C and I_{in} of 16A. As evident, either ΔT and or I_{in} depending on the design constraints, can be optimized by either minimizing the TECs ΔT and or maximizing TECs I_{in} to increase Q_c within max operational limits. In Figure 8, P_{in} and I_{in} are directly proportionally, which will initially increase Q_c until a certain maximum limit, after which increasing P_{in} and I_{in} drop Q_c − contrary to ΔT which is inversely proportional to Q_c. The optimal operation point is highlighted in red. Figure 9 shows a TECs P_{in} vs I_{in} vs R. Normally R is set fixed when designed by the manufacturer but now, with R_t introduced, R can be fairly altered and if it is matched to R_s, maximum power will be transferred to the TEC(s); thereby, optimizing P_{in} and maximizing Q_c as highlighted in red. Figure 10 demonstrates how *CoP* akin to Q_c; increases with decreasing ΔT and initially with increasing I up to a maximum value and then starts decreasing, as current I increases as shown variously in Figure 10. Depending on the design constraints, two optimal *CoP* operation points are evident as highlighted in red − in optimal operation point 1, a *CoP* of 3.3763 is achievable by minimizing I_{in} to 1.8644A and maximizing ΔT to 9.322°C; whereas in optimal operation point 2, a *CoP* of 3.3638 is attainable by maximizing I_{in} to 2.9831A and minimizing ΔT to 0°C. Finally, our TECs model is reasonably validated by comparing a specific Q_c of Figure 7 with that of Figure 13, as shown. The discrepancy is due to different TECs parameters setting. In sum, understanding the theory of TECs parameters and taking the various operational dynamics involved into considerations are very crucial in TEC(s) design/performance.

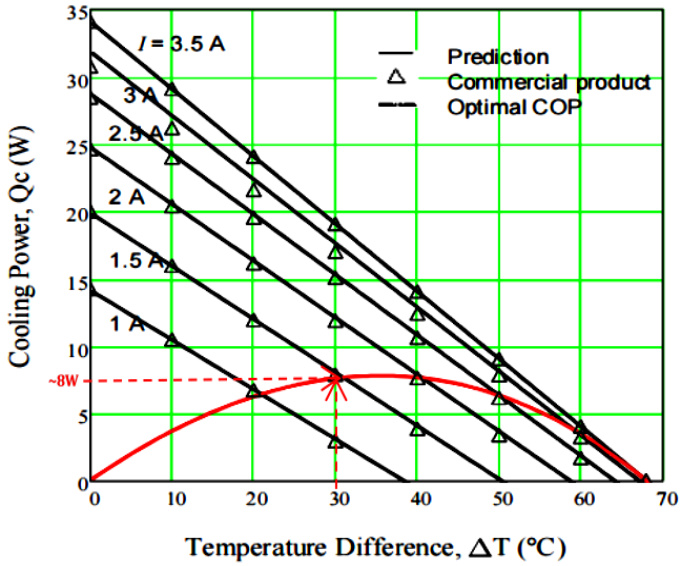

Figure 13: Validating our model with Lee, 2016 [9] − using TEC cooling power Q_c = ~8W vs input current I = ~1.5A vs ΔT= ~30°C to validate our TECs Q_c in Figure 7 result with cooling power Q_c = ~16W vs input current I = ~2A vs ΔT= ~30°C.

4.3. TEGs Dynamic Simulation Results Discussion

Some of the critical TEGs dynamic simulated in Section 3.3. and the importance of the results are herein discussed. The TEGs temperatures and modules electrical connections (series, parallel, series/parallel) dynamics were simulated. In which beginning with the TEGs temperature dynamics, various arbitrary temperatures on the TEGs hot and cold sides as demonstrated in Figure 1b and Figure 11a, as well as summarized in Table 1, were simply dynamically simulated using time-series inputs. As expected, the TEGs dynamically generated power, voltage and current; increased with increasing T_h and DT but with decreasing T_c.

Table 1: TEGs time-series inputs dynamic simulations results summary

Parameters	Matlab / Simulink Simulation Time									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Figure 11a	**TEG modules T_h, T_c and DT dynamic temperature inputs in ℃**									
TEGs T_h	60	120	125	75	100	80	65	85	200	150
TEGs T_c	10	20	25	30	35	15	40	0	5	45
TEGs DT	50	100	100	45	65	65	25	85	195	105
Figure 11b	**36 TEG modules in 10 dynamic T_s, T_p, T_t and R_t auto configuration**									
T_s	36	18	12	9	6	6	4	3	2	1
T_p	1	2	3	4	6	6	9	12	18	36
T_t	36	36	36	36	36	36	36	36	36	36
R_t (Ω)	54.86	13.72	6.096	3.429	1.524	1.524	0.677	0.381	0.169	0.0423
Figure 11c	**36 TEG modules ideal (if TEG_{Rtint} = 0) power, voltage and current**									
TEG_{PocM} (W)	129.1	479.4	423.2	70.86	108.8	108.3	7.995	73.61	196.1	12.2
TEG_{Voc} (V)	85.72	85.72	57.15	19.29	18.57	18.57	4.763	12.14	18.57	5
TEG_{Ioc} (A)	1.506	5.593	7.406	3.674	5.86	5.831	1.679	6.061	10.56	2.44
Figure 11d	**36 TEG modules internal resistance power, voltage and current**									
TEG_{Rtint} (Ω)	54.86	13.72	6.096	3.429	1.524	1.524	0.677	0.381	0.169	0.0423
$PTEG_{Rtint}$ (W)	124.5	429	334.3	46.28	52.33	51.82	1.909	14	18.88	0.2519
$VTEG_{Rtint}$ (V)	82.63	76.71	45.15	12.6	8.93	8.887	1.137	2.309	1.788	0.0103
$ITEG_{Rtint}$ (A)	1.506	5.593	7.406	3.674	5.86	5.831	1.679	6.061	10.56	2.44
Figure 11e	**36 TEG modules generated (terminal) power, voltage and current**									
P_{TEG_Out} (W)	4.656	50.43	88.9	24.58	56.51	56.49	6.086	59.61	177.3	11.95
V_{TEG_Out} (V)	3.091	9.016	12	6.691	9.643	9.687	3.625	9.835	16.79	4.897
I_{TEG_Out} (A)	1.506	5.593	7.406	3.674	5.86	5.831	1.679	6.061	10.56	2.44
Figure 11f	**36 TEG modules boost converter output power, voltage and current**									
P_{conv_out} (W)	3.445	44.04	82.29	22.95	51.25	51.82	5.314	53.39	167.5	11.17
V_{comv_out} (V)	2.291	8.193	11.2	5.914	8.837	8.887	2.846	9.02	15.98	4.125
I_{conv_out} (A)	1.503	5.376	7.348	3.881	5.799	5.831	1.867	5.919	10.48	2.707

The TEG modules quantity used and most vitally in series, parallel and mixed connection were simulated, whereby as shown in Figure 1b and Figure 11b, as well as summarized in Table 1; 36 TEGs were arbitrary chosen and then arranged in 10 different combinations to study the effects of the various arrangements and when matched to a 1.524Ω electrical load. Each arrangement gives a different R_t, consequently giving different generated powers, voltages and currents. Figure 11c depicts the TEGs ideal power, voltage and current generated − assuming the TEGs R_t or TEG_{Rtint} is trivial. Figure 11d shows the power loss, voltage drop and Ohmic current due to the presence of TEG_{Rtint}. Finally, Figures 11e and 11f, show the resultant output power, voltage and current supplied to the DC-DC boost converter and from it. As apparent, more TEG modules increased the output values; however, what is more insightful is how TEGs opt to be connected and matched to a R_L − to ensure maximum power is transferred between R_t and a R_L.

5. Conclusions

Sustainable energy is becoming popular to supplement the traditional grid and for private use, as well as for green economy. In view of this, we proffer thermoelectricity as an alternative energy source (TEGs) as well as an energy efficient load (TECs) for assorted applications that require low DC power, cooling and heating. However, TEG and TEC require multiple units connected in series and or in parallel to provide decent output and cooling powers respectively. Usually, the uninformed perception would be trying to utilize more TEGs and TECs with the hope to get more output and cooling powers respectively. However, our findings asserted this is not really the case, since i) TEG and TEC are not entirely linear devices, especially with increasing current, ii) TEG and TEC temperature difference ΔT and current parameters have performance dynamics which must be operated within very strict optimal operation limits to guarantee efficiency and iii) TEGs and TECs total electrical resistance R_t changes − increases when connected in series and decreases when connected in parallel. Thus, the overall power and efficiency will be affected, especially if the source and load resistances are not matched to transfer maximum power. In essence, our research major contributions include formulas developed for various TEGs/TECs parameters with focus on the TEG and TEC modules total resistance R_t variations − when more than one TEG and or TEC modules are connected in infinite series and or in parallel combinations. Further contributions include detailed TEGs and TECs theoretical simulated models using Matlab/Simulink, whereby the TEGs and TECs models were used to easily simulate and investigate some thermoelectricity profound parameters performance dynamics, R_t losses and to validate some of their operation points with industry standard models. Assorted large scale practical applications of TEGs and TECs were examined and in light of their results, our future work will include embarking on an actual lab design, testing our implemented models using them and refining ours accordingly while taking the physical dynamics into account. Thereafter, a practical pilot 1kW implementation shall be devised for a low energy combined cooling, heating and power (CCHP) system − as an alternative energy green option for private use.

Nomenclature/Symbols

A	TEG/TEC p-n junction thermocouple area in m^2
$CCHP$	Combined cooling, heating and power
CFD	TEC(s) cold flux density in W/m
CoP	TEC(s) coefficient of performance
CoP_e	TEC(s) CoP expression
CoP_{max}	TECs maximum CoP
CoP_{mid}	TEC(s) midpoint CoP
CoP_n	TEC(s) normalized CoP TEC(s)
ΔT	TEG(s) temperature difference ($T_h - T_c$) in °C or K
ΔT_{max}	TEC(s) maximum temperature difference in °C
ΔT_n	TEC(s) normalized temperature difference
HFD	TEG(s) heat flux density in W/m^2
I	TEGs output current in ampere through the TEG(s)
I_{conv_out}	TEGs booster converter output current
I_{cop}	TEC(s) current in ampere to yield CoP
Icp_{max}	TEC(s) maximum cooling power current in ampere
I_{in}	TEC module(s) input current in ampere
Iin_n	TEC(s) normalized input current is the ratio of I_{cop} and I_{max}
I_{Max}	TEG(s) maximum output current in ampere
I_{max}	TEC(s) maximum input current in ampere when $Q_c = 0$
I_{mid}	TEC(s) midpoint current in ampere
I_n	TEG(s) normalized output current
$ITEG_{Rtint}$	TEG ohmic current − results to TEG Ohmic or Joule heating
I_{TEG_Out}	TEGs generated current (input current to the boost converter)
K	TEC/TEG(s) thermal conductance in (W/K)
k_e	TEG(s)/TEC(s) effective thermal conductivity in W/mK
L	TEG/TEC p-n junction thermocouple length in meter
n	P-N thermocouples amount used in a TEG/TEC
η	TEG(s) thermal/electrical/conversion efficiency
η_c	Carnot efficiency
η_e	TEG(s) conversion efficiency expression
η_n	TEG(s) conversion efficiency normalized
η_m	TEG(s) maximum conversion efficiency
η_{mp}	TEGs max power conversion efficiency at the TEGs maximum P_o
ρ	TEG/TEC electrical resistivity in Ω.m
ρ_e	TEG(s)/TEC(s) effective electrical resistivity in Ω.m
P_{conv_out}	TEGs booster converter output power
P_{in}	TEC module(s) input power in watt
Pin_{mid}	TEC(s) midpoint input power in watt
Po	TEG(s) output power in watt − which is $Q_h - Q_c$
Po_{max}	TEG(s) maximum output power in watt
P_n	TEG(s) normalized output power
$PTEG_{Rtint}$	TEG generated power loss − due to TEG internal resistance
P_{TEG_Out}	TEGs generated power (input power to the boost converter)
Q_c	TEC module(s) cooling power on its cold-side in (W)
Q_c	TEG module(s) heat emitted on its cold-side in watt
Q_h	TEC module(s) heat emitted on its hot-side in watt
Q_h	TEG module(s) heat absorbed on its hot-side in watt
Qcp_{max}	TEC(s) Icop maximum cooling power in watt
Qc_{max}	TECs maximum absorbable heat in watt, when $\Delta T = 0$°C
Qc_{mid}	TEC(s) midpoint cooling power in watt
Qc_n	TEC(s) normalized cooling power is the ratio of Q_c and Qc_{max}
r	TE device p-n thermocouples unit resistance in ohm
R	TE device (TEG and TEC) module unit resistance in ohm
R_L	TEGs electrical load resistance in Ω connected to the TEG(s)
R_s	Power source resistance in ohm connected to the TECs
R_t	TEG/TEC module(s) total resistance in ohms
S	TE device Seebeck coefficient in V/K
Se	TEG(s)/TEC(s) effective Seebeck coefficient in V/K
$\bar{T}$	TE device average temperature ($T_h + T_c$)/2 in K or °C
T_c	Temperature on TEG/TEC cold-side in °C
T_h	Temperature on TEG/TEC hot-side in °C
TE	Thermoelectric
TEC	Thermoelectric cooler
$TECsa$	TEC cold-side surface area
TEG	Thermoelectric generator
TEG_{Ioc}	TEG ideal generated current − assuming there is no TEG_{Rtint}
TEG_{PocM}	TEG ideal generated power − assuming there is no TEG_{Rtint}
TEG_{Rtint}	TEG internal resistance (R_t) − responsible for the power loss
$TEGsa$	TEG hot-side surface area
$TEGs\ T_c$	TEGs cold side temperature
$TEGs\ DT$	TEGs temperature difference
$TEGs\ T_h$	TEGs hot side temperature
TEG_{Voc}	TEG ideal generated voltage − assuming there is no TEG_{Rtint}
TEH	Thermoelectric Energy Harvester
T_p	TEGs/TECs module quantity connected in parallel
T_s	TEGs/TECs module quantity connected in series
T_t	TEG/TEC modules total quantity connected
V_{comv_out}	TEGs booster converter output voltage
Vin	TEC module(s) input voltage in volt
Vin_{max}	TEC's max V_{in} in (V) that produces max ΔT_{max} when $I_{in}=I_{max}$
Vin_n	TEC(s) normalized input voltage is the ratio of V_{in} and Vin_{max}
Vo	TEG module(s) output voltage in volt
Vo_{max}	TEG(s) maximum output voltage in volt
Vn	TEG(s) normalized output voltage
V_{TEG_Out}	TEGs generated voltage (input voltage to the boost converter)
$VTEG_{Rtint}$	TEG generated voltage drop − due to TEG internal resistance
WSN	Wireless Sensors Network
Z	TE device figure of merit in per K
Ze	TEG(s)/TEC(s) effective figure of merit in per K

$Z\bar{T}$ TE device average dimensionless figure of merit

Acknowledgment

This work was supported in parts by the Cape Peninsula University of Technology CPGS and the University of the Western Cape HySA Systems.

Data availability

Research in progress − data available upon completion.

Conflict of Interest

Authors declare no conflict of interest.

References

[1] M.L. van der Walt, J. van den Berg, M. Cameron, "State of Renewable Energy in South Africa", South Africa Department of Energy, Pretoria, 2017, http://www.energy.gov.za/files/media/Pub/2017-State-of-Reewable-Energy-in-South-Africa.pdf.

[2] N.P. Bayendang, M.T. Kahn, V. Balyan, "A structural review of thermoelectricity for fuel cells CCHP applications," Hindawi Journal of Energy 2020, 1-23, 2020, https://doi.org/10.1155/2020/2760140/

[3] Y.-S. Noh, J.-I. Seo, W.-J. Choi, J.-H. Kim, H.V. Phuoc, H.-S. Kim, S.-G. Lee, "17.6 A Re-configurable DC-DC Converter for Maximum TEG Energy Harvesting in a Battery-Powered Wireless Sensor Node," 2021 IEEE International Solid-State Circuits Conference (ISSCC),San Francisco, CA, USA, 266-268, 2021, doi: 10.1109/ISSCC42613.2021.9365811.

[4] D. Charris, D. Gómez, M. Pardo, "A Portable Thermoelectric Energy Harvesting Unit to Power Up Outdoor Sensors and Devices," 2019 IEEE Sensors Applications Symposium (SAS), Sophia Antipolis, France, 1-6, 2019, doi: 10.1109/SAS.2019.8705985.

[5] J. Singh, P. Kuchroo, H. Bhatia, E. Sidhu, "Floating TEG based solar energy harvesting system," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 763-766, 2016, doi: 10.1109/ICACDOT.2016.7877689.

[6] Q. Wan, Y. Teh, P.K.T. Mok, "Analysis of a Re-configurable TEG Array for High Efficiency Thermoelectric Energy Harvesting," 2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Jeju, Korea (South), 662-665, 2016, doi: 10.1109/APCCAS.2016.7804084.

[7] R. Chein, G. Huang, "Thermoelectric cooler application in electronic cooling", Applied Thermal Engineering, **24**(14–15), 2207-2217, 2004, https://doi.org/10.1016/j.applthermaleng.2004.03.001.

[8] I.R. Belovski, A.T. Aleksandrov, "Examination of the Characteristics of a Thermoelectric Cooler in Cascade," 2019 X National Conference with International Participation (ELECTRONICA), 1-4, 2019, doi: 10.1109/ELECTRONICA.2019.8825631.

[9] H. Lee, Thermoelectrics: design and materials, John Wiley & Sons, Inc., Wiley, New Jersey, USA, 2016.

[10] H. Mamur, Y. Çoban, "Detailed Modeling of a Thermoelectric Generator for Maximum Power Point Tracking", Turkish Journal of Electrical Engineering & Computer Sciences, **28**, 124–139, 2020, https://doi.org/10.3906/elk-1907-166.

[11] N.P. Bayendang, M.T. Kahn, V. Balyan, I. Draganov, S. Pasupathi, "A Comprehensive Thermoelectric Generator (TEG) Modelling", AIUE Congress 2020: Energy and Human Habitat Conference, Cape Town, South Africa, 1-7, 2020; Publisher Zenodo: Geneva, Switzerland, Available online 2020, http://doi.org/10.5281/zenodo.4289574.

[12] N.P. Bayendang, M.T. Kahn, V. Balyan, I. Draganov, S. Pasupathi, "A Comprehensive Thermoelectric Cooler (TEC) Modelling", AIUE Congress 2020: International Conference on Use of Energy, Cape Town, South Africa, 1-7, 2020; Publisher SSRN: Rochester, NY, USA, Available online 2021, https://ssrn.com/abstract=3735378 or http://dx.doi.org/10.2139/ssrn.3735378.

[13] F. Felgner, L. Exel, M. Nesarajah, G. Frey, "Component-oriented modeling of thermoelectric devices for energy system design," in IEEE Transactions on Industrial Electronics, **61**(3), 1301-1310, 2014, doi: 10.1109/TIE.2013.2261037.

[14] C. Luo, R. Wang, W. Yu, W. Zhou, "Parametric study of a thermoelectric module used for both power generation and cooling," Renewable Energy, **154**, 542-552, 2020, https://doi.org/10.1016/j.renene.2020.03.045.

[15] C. Liu, P. Chen, K. Li. "A 1kW Thermoelectric Generator for Low-temperature Geothermal Resources," PROCEEDINGS, 39th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA, 24-26, 2014, SGP-TR-202. https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2014/Li.pdf [Date accessed: August 3, 2021].

[16] S.O. Giwa, C.N. Nwaokocha, A.T. Layeni, O.O. Olaluwoye, "Energy harvesting from household heat sources using a thermoelectric generator module," Nigerian Journal of Technological Development, **16**(3), 2019, http://dx.doi.org/10.4314/njtd.v16i3.6.

[17] M.W. Aljibory, H.T. Hashim, W.N. Abbas, "A Review of Solar Energy Harvesting Utilizing a Photovoltaic–Thermoelectric Integrated Hybrid System," IOP Conference Series: Materials Science and Engineering, 4th International Conference on Engineering Sciences (ICES 2020), 1067, 2021, Kerbala, Iraq, doi:10.1088/1757-899X/1067/1/012115.

[18] X. Hu, C. Jiang, X. Fan, B. Feng, P. Liu, Y. Zhang, R. Li, Z. He, G. Li, Y. Li, "Investigation on waste heat recovery of a nearly kilowatt class thermoelectric generation system mainly based on radiation heat transfer," Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2020, https://doi.org/10.1080/15567036.2020.1829190.

[19] M.Y. Fauzan, S.M. Muyeen, S. Islam, "Experimental modelling of grid-tied thermoelectric generator from incinerator waste heat," International Journal of Smart Grid and Clean Energy, 2020, doi: 10.12720/sgce.9.2.304-313.

[20] M. Ebrahimi, E. Derakhshan, "Design and evaluation of a micro combined cooling heating and power system based on polymer exchange membrane fuel cell and thermoelectric cooler", Energy Conversion and Management, 171, 507-517, 2018, https://doi.org/10.1016/j.enconman.2018.06.007.

[21] A.G. Rösch, A. Gall, S. Aslan, M. Hecht, L. Franke, M.M. Mallick, L. Penth, D. Bahro, D. Friderich, U. Lemmer, "Fully printed origami thermoelectric generators for energy-harvesting," NPJ Flex Electrons, **5**(1), 2021, https://doi.org/10.1038/s41528-020-00098-1.

[22] V.B. Abhijith, K. Narayanaswamy, C.M. Pooja, S.V. Prasad, R. Sambhu, "Household Application of Thermoelectric Generator in the Field of Propagating Power from Waste Heat," AIP Conference Proceedings 2297, 020010, 2020, https://doi.org/10.1063/5.0031699.

[23] F. Afshari, "Experimental and numerical investigation on thermoelectric coolers for comparing air-to-water to air-to-air refrigerators," Journal of Thermal Analysis and Calorimetry, **144**, 855–868, 2021, https://doi.org/10.1007/s10973-020-09500-6.

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 79-89 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Novel Algorithm Design for Locating Fault Distances on HV Transmission Lines

MK Ngwenyama[1,*], PF Le Roux[1], LJ Ngoma[2]

[1]Electrical Engineering Department, Tshwane University of Technology, Witbank, 1034, South Africa

[2]Electrical Engineering Department, Tshwane University of Technology, Pretoria, 0183, South Africa

ARTICLE INFO

Article history:
Received: 25 November, 2021
Accepted: 26 January, 2022
Online: 21 February, 2022

Keywords:
Conventional method
Electrical energy
Electrical fault
Impedance-based technique
MATLAB/SIMULINK
Transmission line
Transmission network

ABSTRACT

The transmission network has been considered among the globe's prevalent complex systems, comprised of hundreds of electrical transmission lines and other equipment used to transmit electrical energy from one location to another. Over a decade, power engineers have worked tirelessly to ensure that the transmission network operates reliably, transmitting electrical energy from the power station to the consumers without interruption. With growing generation capacity and the recent introduction of renewable energy systems (RES) such as wind turbines and solar energy, the transmission lines are increasingly being forced to run near their design limitations and greater unpredictability on the network operational configuration. As a result, the transmission network faces greater challenges than previously. As a worst-case scenario, large-scale electrical network power outages caused by electrical faults can disrupt electricity availability for several hours, impacting millions of customers and inflicting massive economic damage. These electrical faults must be repaired before electricity is restored to consumers. This necessitates a thorough grasp of the challenge and potential remedies to assure improved power efficiency. In the present work, an expansion of preceding work, a novel algorithm for estimating faults on transmission lines is presented. Impedance-based techniques are susceptible to producing errors or incorrect predictions. The presence of faults induced from high impedance sources produces an extra impedance to the ground, which negates the impedance calculation and produces errors in the distance to the fault. This results in inaccuracies that can affect a distance-to-fault estimation by 1-15 % of the overall line length. In this work, a design of a fault detection-location element (FDLE) algorithm is proposed. This algorithm relies on the dynamics of current and voltage signals on the transmission line while deserting impedance. Comparison research is undertaken against the impedance-based techniques to validate the proposed algorithm. Finally, the proposed algorithm findings are compared to fault location estimations using an impedance-based technique. Extensive trials on a simulated transmission line prove that the proposed algorithm is responsive to faults with an error as low as 1%, reaching a precision of 98.9%.

1. Introduction

This paper is an extension of work originally presented at the *2021 6th Asia Conference on Power and Electrical Engineering (ACPEE)* [1]. The transmission network has been considered among the globe's prevalent complex systems, comprised of hundreds of electrical transmissions such as the addition of wind turbines and solar energy. Therefore, the transmission network faces greater challenges than previously. Over a decade, power designers have worked diligently to guarantee that the transmission network runs adequately, transmitting electrical energy from the power station to customers without disruption. The demand for transmission lines is increasing as the penetration of renewable energy systems (RESs), electric vehicles (EVs), and energy storage systems (ESSs) grows in modern transmission networks [2]. To achieve dependable, efficient, and reliable RESs, significant improvement in the conventional transmission line protection method is necessary. The rapid increase of current during a fault doubles the potential risk of transmission network degradation and thus presents the requirement that faults

*Corresponding Author: MK Ngwenyama, South Africa, NgwenyamaMK@yahoo.com

in RESs be separated by a protection method with the shortest operational time. Due to the increased dependability and stability of RESs, the ring model is the ideal design for RESs because of the bidirectional flow of current in these configurations; the standard protection system would not perform adequately. As a worst-case scenario, large-scale electrical network power outages caused by electrical faults can disrupt electricity availability for several hours, impacting millions of customers and inflicting massive economic damage. These electrical faults must be repaired before electricity is restored to consumers. The introduction of the proposed algorithm may aid in the improvement of these systems.

Many electric utilities have opted to invest minimal, or no effort in electrical transmission fault detection devices in the past, alleging that faults are usually temporary thus do not necessitate location. Moreover, [3] suggested that the data produced by protection relays were not accurate enough to support deploying personnel to check temporary fault locations. As a result, fault location operations were commenced immediately after a fault had become permanent and a solution was required. With the restructuring and privatisation of electric power production enterprises across the globe, such sentiments are beginning to change [4, 5]. To attain the greater plant efficiency levels and increased service reliability regulations that a dynamic industry needs, electricity utilities are increasingly taking a more "pro-active" strategy to most business operations, notably in the field of transmission line fault location. The capability of impedance-based techniques to address the errors and constraints of protection relays was discovered in the mid-20th century when various trial designs were introduced; several of them were converted into industrial uses [6]. Although such designs are more accurate than other methods, they were discontinued due to dependability and maintenance issues, leading to a lack of involvement and credibility. Impedance-based fault location techniques are used repeatedly on transmission lines [7]. Many causes drive the resurgence of impedance-based techniques, the foremost of which is an industry need for quick and precise fault location on economically vital, extremely lengthy high-voltage transmission lines. Precise fault location lowers operational expenses by eliminating costly and time-consuming inspections [8]. Precise fault location speeds up line maintenance and recovery, decreasing revenue shortfall due to outages. Comparison research is undertaken against other techniques to validate the proposed algorithm.

In [9], the authors proposed an artificial intelligence-based approach for detecting faults in photovoltaic (PV) plants and transmission lines. The method is known as recurrent neural networks (RNN). The authors stress that faults in PV plants are usually foreseeable, and while conventional methods are precise, they are not cost-effective for mass application, allowing most PV plants to operate unmonitored. They developed an algorithm for precise fault diagnosis of PV plants that uses satellite weather data and low-frequency inverter sensors. The algorithm enables machine learning-based fault detection even for PV plants with the absence of detectors, and it utilises a recurrent neural network to detect all forms of defects based on the last 24 hours of observations, rather than just the latest detected fault. The authors also demonstrated that the proposed algorithm determines the generated power loss induced by the fault, for example, the fault level, whereas the traditionally utilised techniques are restricted to classifying the fault type. The results show that the algorithm is responsive to as low as 5 % intensity defects, with a precision of 96.9% utilising accurate climate data and 86.4% utilising satellite weather data. The results reveal that the algorithm could also detect unspecified defects, that is, defects that appeared to be not included in the training data.

The authors in [10] presented a closed-loop sinusoidal pulse width modulation (SPWM) monitoring device for the operations of converters utilised in wind turbine plants for utility grid purposes. Wind energy is a major form of clean energy; therefore, windmills are commonly deployed in electrical distribution systems and thus are coupled physically to electrical transmission systems. Wind energy significantly influences the performance of current configuration systems as network saturation improves, resulting in the probability of faults on the transmission lines transporting power to the main grid. The presented technique is utilised to investigate wind turbine efficiency. The authors state that voltage and current inverters produce discrete output waveform signals, harmonic pollution, extra power losses, and higher frequency noises. As a result, to achieve the required current waveforms, large inductors must be connected in line with the related load. The simulated results showed that the mentioned technique has a higher harmonic elimination strength. As a result, the transmission network will require fewer fault detecting and locating devices.

In [11], the authors presented a technique for performing serial fault repair in a wind turbine plant with synchronous actuator and sensor defects. The technique is known as a robust sliding mode observer (SMO). The authors used a diagonal transform matrix and a post-filtering system to design a novel improved model. The novel model then transforms the sensor fault into an actuator fault to locate the fault. The novel model allows the SMO technique to detect and pinpoint faults on transmission lines. The simulated tests indicate how the presented technique can properly restore actuator and sensor defects; thus, the active fault controller can obtain optimum wind power extraction. Although this technique works optimally for a microgrid that operates in an islanded mode, there is no concise evidence of how this technique functions when employed to a transmission line of a microgrid configuration integrated into the main grid.

The theories of the single and double-ended impedance-based fault location techniques are also discussed in this paper and explain the concepts related to fault location and present different impedance-based fault location algorithms. The work aims to investigate numerous network faults and assess the efficiency of fault locators in light of potential causes of inaccuracy [12].

In the present work, an expansion of preceding work [1], a novel algorithm for estimating faults on transmission networks is presented. Impedance-based techniques are susceptible to producing errors or incorrect predictions. The presence of faults induced from high impedance sources produces an extra impedance to the ground, which negates the impedance calculation and produces errors in the distance to the fault. This results in inaccuracies that can affect a distance-to-fault estimation by 1 -15 % of the overall line length. This will have a detrimental impact on the transmission network and renewable

energy systems. In this work, a design of a fault detection-location element (FDLE) algorithm is proposed. The model utilises MATLAB/SIMULINK software to study the electrical faults on a transmission line. The proposed algorithm relies on the dynamics of current and voltage signals on the transmission line while deserting impedance. The authors provide simulation results to demonstrate the efficiency of the proposed algorithm. The FDLE algorithm findings are compared to fault location estimations using an impedance-based approach. Extensive trials on a simulated transmission line prove that the proposed algorithm is responsive to faults with an error as low as 1%, reaching a precision of 98.9%.

2. Techniques and Specifications for Fault Diagnosis

Various techniques of predicting fault position are currently in practice in the industry [13]:

- Conventional methods:
  - Impedance-based methods [14]:
    - Single-ended method, and
    - Double-ended method (Synchronized, Unsynchronized, and Unsynchronized Current-only).
  - Travelling wave methods [15, 16]:
    - Single-ended method, and
    - Double-ended method.
- Artificial Intelligence methods:
  - Artificial Neural Network (ANN) [17],
  - Support Vector Machine (SVM) [18],
  - Fuzzy Logic [19], and
  - Matching Approach [20].

In this work, comparison research is undertaken against the impedance-based techniques to validate the proposed algorithm and presents the outcomes using real-world simulated faults.

3. Theoretical Foundations of Impedance-based Fault Location Techniques

Various impedance-based fault location techniques have been designed for transmission lines purposes. Single-ended techniques are fault-locating techniques that use information recorded by a fault location detection element at one end-point of the transmission line to locate faults [21]. Double-ended techniques use information from fault location detection elements at both end-points of the connection. The impedance between both the fault locator and the position of the fault is estimated using voltage and current signals collected by fault location detection elements during a fault [22]. The distance to the fault may be precisely determined if the transmission line impedance in ohms is known. When determining the proximity to a fault, each technique has different basic data needs and requires different estimates. These estimates may or may not be valid in a specific fault position case [23]. Selecting the appropriate technique for pinpointing faults within quite a diverse set of impedance-based fault location techniques is thus a challenging effort that demands a diligent grasp of the essential principles of an individual fault-locating technique. Constructed on the preceding knowledge, this paper discusses the fundamental concept of single-ended and double-ended impedance-based fault location techniques. The aims are to properly articulate each fault-finding technique's input data requirements, explain each technique's applicability in pinpointing actual transmission line faults, and offer advice on selecting the optimal fault-locating technique.

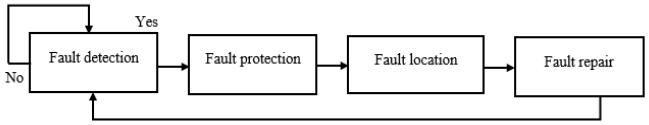

Figure 1: Flow Chart Procedure of a Conventional Technique.

4. Techniques and Prerequisites for Impedance-based Fault Location

The underlying analysis is required for impedance-based techniques:

- Voltage and current amplitudes should be recorded,
- Extract the essential elements,
- Identify the fault category and faulted phasor/s, and
- Employ the impedance-based technique.

Single-ended impedance techniques employ a simplified approach. Hence, links between communicating and off-site data are typically unnecessary [24]. Whereas double-ended techniques are extremely precise. However, they require data from both end-points of the transmission line. For this technique to be executed [25].

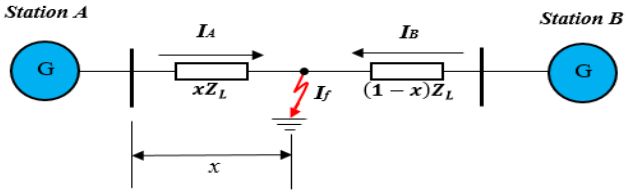

Figure 2: A Single-ended Impedance-based Algorithm.

Table 1: Different Formulae for Fault Types [1].

Fault Type	Positive-Sequence Impedance Equation *mZ1L is equal to:*
A-earth	$V_a/(I_a + (k \times 3 \times I_O))$
B-earth	$V_b/(I_b + (k \times 3 \times I_O))$
C-earth	$V_c/(I_c + (k \times 3 \times I_O))$
A-B / A-B-earth	V_{ab}/I_{ab}
B-C / B-C-earth	V_{bc}/I_{bc}
C-A / C-A-earth	V_{ca}/I_{ca}
A-B-C	*Any of the following:* V_{ab}/I_{ab}, V_{ab}/I_{ab}, V_{ca}/I_{ca}

The single-ended impedance-based fault location technique determines the fault position using the detected impedance through examining the transmission line from a single end-point. Every line's line-to-line and line-to-ground output signal must be recorded to pinpoint all fault forms [12]. The impedance equations shown in Table 1 may be utilised to approximate the fault position assuming zero fault resistance.

Where V_A, V_B, and V_C are the phase voltages,

I_A, I_B, and I_C are the phase currents,

Z_O the zero-sequence impedance,

Z_1 the positive sequence impedance, and

K_O the residual compensation factor and may be expressed as:

$$K_O = (Zo - Z_1)\, K * Z_1 \quad (1)$$

K may be 1 or 3, depending on the fault location detection element design. According to Kirchhoff's law, the zero-sequence Io in Equation (2) will be zero in a balanced system.

$$Io = (I_A + I_B + I_C)/3 \quad (2)$$

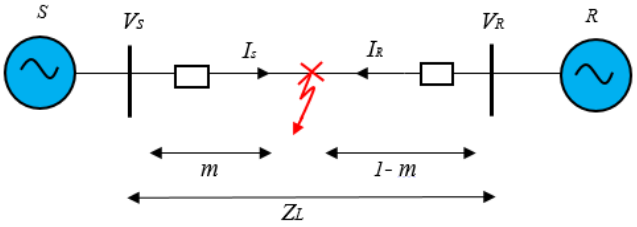

Figure 3: A Double-ended Impedance-based Algorithm.

Where m in Figure 3 denotes the distances to a fault per unit [26], V_S the sending voltage, V_R the receiving voltage, I_S the sending current, I_R the receiving current, and Z_L the impedance of the line.

The difficulties in accurately locating faults using single-ended techniques are recognised and documented in numerous publications [27].

To summarise, the preceding circumstances may result in single-ended impedance-based fault location techniques producing errors:

- Imprecise fault-type (faulted phase/s) detection.
- Incorrect line parameters that do not correspond to actual parameters.
- Lack of accuracy of the line model.
- Current and voltage transformer errors.

5. Single-ended Impedance-based Method

5.1. Simple Reactance Technique

The potential difference loss at the sending (S) end of the transmission line, illustrated in Figure 4 is:

$$V_S = (m \times Z_{1L} \times I_S) + (R_f \times I_f) \quad (3)$$

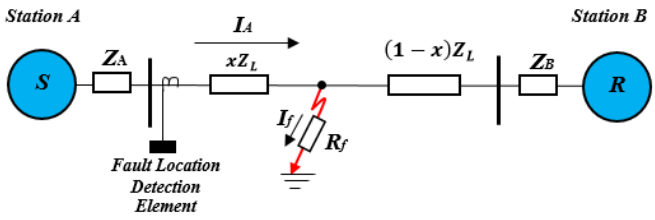

Figure 4: Faulted Transmission Network with Fault Location Detection Element.

For an A phase to earth fault, it must have the following characteristics,

$$V_S = V_{a-g} \text{ and } I_S = I_a + (k \times 3 \times I_O) \quad (4)$$

The objective is to reduce the impact of the $R_f \times I_f$ variable. The basic reactance algorithm divides all variables by I_S (I obtained at the fault location) and excludes the variable $R_f \times (I_f/I_S)$.

To achieve this, preserve the imaginary component and solve m.

$$I_m(V_s/I_s) = I_m(m \times Z_{1L}) = m \times X_{1L} \quad (5)$$

$$m = \frac{I_m\left(\frac{V_S}{I_S}\right)}{X_{1L}} \quad (6)$$

Error is 0 if $\angle I_s = \angle I_f$ or $R_f = 0$

5.2. Takagi Technique

The Takagi technique needs the collection of pre-fault and fault data. It mostly enhances the simple reactance technique by minimising the influence of fault resistance and lowering the impact of load flow [28].

$$V_s = (m \times Z_{1L} \times I_f) + (R_f \times I_f) \quad (7)$$

Superposition current (I_{sup}) may be used to identify a variable that is in conjunction with I_f:

$$I_{sup} = I_f - I_{pre} \quad (8)$$

$$I_f = Fault\ Current \quad (9)$$

$$I_{pre} = Pre - Fault\ Current \quad (10)$$

Potential difference loss produced by the sending bus:

$$V_s = (m \times Z_{1L} \times I_s) + (R_f \times I_f) \quad (11)$$

Multiply both sides of (7) with the complex conjugate of I_{sup} (I^*_{sup}) and keep the imaginary component. Secondly, solve m:

$$I_m(V_s \times I_{sup^*}) = [(m \times I_m)(Z_{1f} \times I_s \times I_{sup^*})] + [(R_f I_m)\ (I_f \times I_{sup^*})] \quad (12)$$

$$m = \frac{I_m(V_s \times I_{sup^*})}{I_m(Z_{1L} \times I_S \times I_{sup^*})} \quad (13)$$

The fact that I_s and I_f angles are the equivalents that are important to the Takagi technique's efficiency. These angles are the same in a perfect homogenous network. The inaccuracy in the fault position estimation improves as the angle between I_s and I_f improves [29].

5.3. Modified Takagi Technique

For earth faults, the modified Takagi utilises zero-sequence current from the sending end $(3 \times I_{os})$ rather than the superposition current. As a result, no pre-fault data is required for this technique. Angle rectification is similarly possible with the Modified Takagi technique. Provided that the user understands the network source impedances, he/she can change the zero-sequence current using angle T to enhance the fault position estimation for a specific transmission line [30].

$$m = \frac{I_m(V_S \times (3 \times I_{OS})^* \times e^{-jT})}{I_m(Z_{1L} \times I_S \times (3 \times I_{OS})^* \times e^{-jT})} \quad (14)$$

The defined angle T is only applicable for a single fault point within the transmission line. Figure 5 demonstrates methods to compute T.

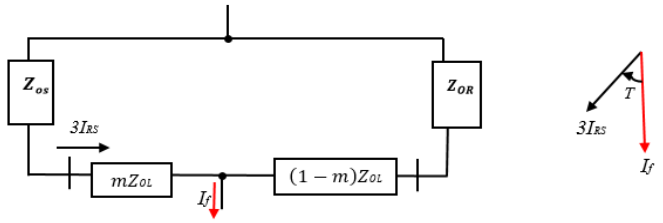

Figure 5: Correction of a Zero Sequence Angle.

$$\frac{I_f}{3 \times I_{RS}} = \frac{Z_{OS}+Z_{OL}+Z_{OR}}{(1-m) \times (Z_{OL}+Z_{OR})} = A \angle T \quad (15)$$

Even though the Modified Takagi technique outperforms the Takagi technique, the precision of position estimations is dependent on having the source impedance characteristics precisely.

5.4. Eriksson Technique

This technique employs the generated impedance characteristics to avoid any reactance inaccuracy induced from fault resistance, loads, or non-homogeneity network to determine the distances to the fault. Furthermore, such a technique evaluates the magnitude of fault impedance, which is important in detecting the main reason for a fault and verifying a transmission network short-circuit design [31].

Equation (16) summarises and updates the variables in Equation (14).

$$m^2 - (k_1 m) + k_2 - (k_3 R_F) = 0 \quad (16)$$

Wherein k_1, k_2, and k_3 represent complex algebraic expressions of a transmission line voltage, current, impedance, and supply impedances; thus, can be presented as follows:

$$k_1 = a + jb = 1 + \left(\frac{Z_{H1}}{Z_{L1}}\right) + \left(\frac{V_{G1}}{Z_{L1} \times I_G}\right)$$

$$k_2 = c + jd = \left(\frac{V_{G1}}{Z_{L1} \times I_G}\right)\left(1 + \frac{Z_{H1}}{Z_{L1}}\right)$$

$$k_3 = e + jf = \left(\frac{\Delta I_{G1}}{Z_{L1} \times I_G}\right)\left(1 + \frac{Z_{H1}+Z_{G1}}{Z_{L1}}\right)$$

By dividing (16) into actual and unreal components, total distance to fault m may be calculated using the quadratic formula (17).

$$m = \frac{\left(a - \frac{eb}{f}\right) \pm \sqrt{\left(a - \frac{eb}{f}\right)^2 - 4\left(c - \frac{ed}{f}\right)}}{2} \quad (17)$$

Should m be known, any other unknown variable may be obtained using Equation (17). Given that the fault position prediction has to be smaller than the entire transmission line distance, an m value of between 0 and 1 per unit must be utilised for the distance measurement.

Therefore, Equation (18) may be used to determine fault tolerance:

$$R_f = \frac{d - mb}{f} \quad (18)$$

Assuming that the supply impedance Z_{G1} is not known, and the load impedance Z_{H1} is precisely calculated, the impedance Z_{G1} can be estimated using fault occurrence records as:

$$Z_{G1} = -\frac{V_{G1} - V_{G1pre}}{I_{G1} - I_{G1pre}} \quad (19)$$

5.5. Novosel et al. Technique

The Novosel et al. technique is a revised edition of the Eriksson technique for pinpointing faults on a compact, radial transmission line. At downstream of the transmission line, all the loads supplied or connected to the line are combined. Given that there is a fixed load impedance design, the initial action is to determine the load impedance using the pre-fault current and voltage as inputs [14]; this is expressed as:

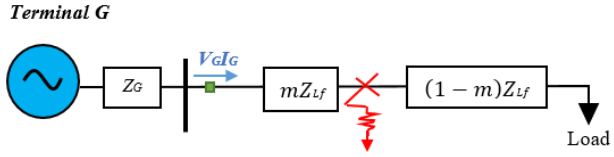

Figure 6: Novosel et al. Technique Applied to a Transmission Network with a Constant Impedance Load

$$Z_{Load} = R + jX = \frac{V_{G1pre}}{I_{G1pre}} - Z_{L1} \quad (20)$$

Resolving the quadratic Equation in (17) results in the distance to the fault per unit, whereby the variables are specified as:

$$k_1 = a + jb = 1 + \left(\frac{Z_{Load}}{Z_{L1}}\right) + \left(\frac{V_{G1}}{Z_{L1} \times I_G}\right)$$

$$k_2 = c + jd = \left(\frac{V_{G1}}{Z_{L1} \times I_G}\right) + \left(1 + \frac{Z_{Load}}{Z_{L1}}\right)$$

$$k_3 = e + jf = \left(\frac{\Delta I_{G1}}{Z_{L1} \times I_G}\right) + \left(1 + \frac{Z_{Load}+Z_{G1}}{Z_{L1}}\right)$$

Similar to the Eriksson technique, *m* can include any of these variables on the Novosel et al. technique. Given that the fault position prediction must be less than the total transmission line distance, several *m* between 0 and 1 per unit must be used for distance measurement.

6. Double-ended Impedance-based Method

6.1. Synchronised Two-ended Technique

This technique is built on the hypothesis that data from both end-points of a transmission line are coordinated to a similar period response using a global positioning system (GPS). During fault position analysis, any three symmetric variables can be employed. However, utilising negative-sequence elements is highly beneficial because it is unaffected by load current or zero-sequence connected loads feeding to the system. To demonstrate the fault-finding concept, a negative-sequence system amid an imbalanced fault must be considered [32].

Figure 7 illustrates how to compute V_{F2}, the negative-sequence potential difference at the fault location *F*, using Terminals *G* and *H*. The Equations are represented as:

$$\textit{Terminal G: } V_{F2} = V_{G2} - (mZ_{L2}I_{G2}) \tag{21}$$

$$\textit{Terminal H: } V_{F2} = V_{H2} - (1-m)(Z_{L2}I_{H2}) \tag{22}$$

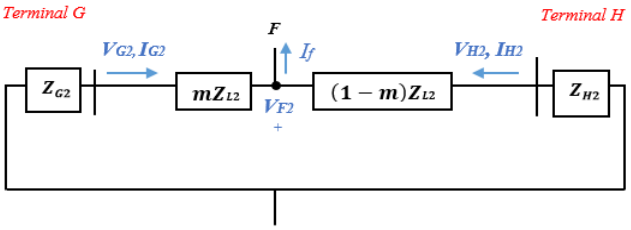

Figure 7: Negative Sequence System during Asymmetrical Fault.

The potential difference V_{F2} determined on either bus of the transmission line is equivalent. As a result of matching Equations (21) and (22), the distance measurement to a fault is calculated as:

$$m = \frac{V_{G2} - V_{H2} + (Z_{L2}I_{H2})}{(I_{G2} + I_{H2})Z_{L2}} \tag{23}$$

A symmetrical fault may be located using Equation (23). On the other hand, negative-sequence components do not occur when a symmetrical fault is present. In this scenario, the identical fault pinpointing approach is used for the positive sequence system, and the fault distance is calculated as:

$$m = \frac{V_{G1} - V_{H1} + (Z_{L1}I_{H1})}{(I_{G1} + I_{H1})Z_{L1}} \tag{24}$$

6.2. Unsynchronised Two-ended Technique

In this technique, current and voltage signals recorded by fault location detection elements at opposite ends of a transmission line could be out of synchronisation. The GPS sensor could be damaged or barely working properly. Fault location detection elements can also have various sample speeds or pinpoint the fault at marginally varying time intervals [33].

A phase shift can be caused by the communication route that transmits data from one fault location detection element to the other. As a result, $e^{j\delta}$ must be employed as a synchronising controller to synchronise the voltages and currents measured at Terminals G and H. This process is as follows:

$$\textit{Terminal G: } V_{Fi} = V_{Gie^{j\delta}} - \left(mZ_{Li}I_{Gie^{j\delta}}\right) \tag{25}$$

$$\textit{Terminal H: } V_{Fi} = V_{Hi} - (1-m)(Z_{Li}I_{Hi}) \tag{26}$$

Asymmetrical faults are calculated using the negative-sequence element, whereas symmetrical faults are calculated using the positive-sequence element. The sync-controllers, as indicated in Equation (27), are formed by matching Equations (25) and (26).

$$\left|e^{j\delta}\right| = 1 = \left|\frac{V_{Hi} - (1-m)(Z_{Li}I_{Hi})}{V_{Gi} - (mZ_{Li}I_{Gi})}\right| \tag{27}$$

The distance measurement to fault *m* is an algebraic expression obtained after simplification and arranging the variables.

$$m = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \tag{28}$$

Where the parameters are specified as follows:

$$A = |Z_{Li}I_{Gi}|^2 - |Z_{Li}I_{Hi}|^2$$

$$B = -2 \times Re[V_{Gi}(Z_{Li}I_{Gi})^* + (V_{Hi} - Z_{Li}I_{Hi})(Z_{Li}I_{Hi})^*]$$

$$C = |V_{Gi}|^2 - |V_{Hi} - Z_{Li}L_{Hi}|^2$$

When the algebraic expression in (28) is solved, two parameters of *m* are obtained. The fault position estimation must have a value of 0 to 1 per unit.

6.3. Unsynchronised Current-only Two-ended Technique

This technique is used when there are data accessibility constraints, assuming that merely current signals at buses *G* and *H* are provided for fault analysis and the voltage signals V_{G2} and V_{H2} are either absent or unavailable. V_{F2} is determined from the two buses using just the current as well as the supply impedance characteristics as:

$$\textit{Terminal G: } V_{F2} = -(Z_{G2} + mZ_{L2})I_{G2} \tag{29}$$

$$\textit{Terminal H: } V_{F2} = -(Z_{H2} + (1-m)Z_{L2})I_{H2} \tag{30}$$

V_{F2} is removed by dividing Equations (29) by (30). This is to prevent difficulties of synchronising data collections from all buses of the transmission line and simply use real numbers:

$$|I_{H2}| = \left|\frac{(Z_{G2} + mZ_{L2})}{(Z_{H2} + (1-m)Z_{L2})} \times I_{G2}\right| \tag{31}$$

The fault distance *m* is then calculated using the algebraic Equation (28), where the variables are specified:

$$a + jb = I_{G2}Z_G$$

$$c + jd = Z_{L2}I_{G2}$$

$$e + jf = Z_{H2} + Z_{L2}$$

$$g + jh = Z_{L2}$$

$$A = |I_{H2}|^2 \times (g^2 + h^2) - (c^2 + d^2)$$

$$B = -2 \times |I_{H2}|^2 \times (eg + fh) - 2(ac + bd)$$

$$C = |I_{H2}|^2 \times (e^2 + f^2) - (a^2 + b^2)$$

This technique is only convenient for pinpointing imbalanced faults. Given that current data is available, the position predictions' precision depends on having the supply impedance characteristics precisely [34].

7. Proposed Model

The proposed model is shown in Figure 8. The model is applied to a simulated transmission line design. The sequencing plan of the algorithm utilised to pinpoint the transmission line faults is indicated in Figure 9.

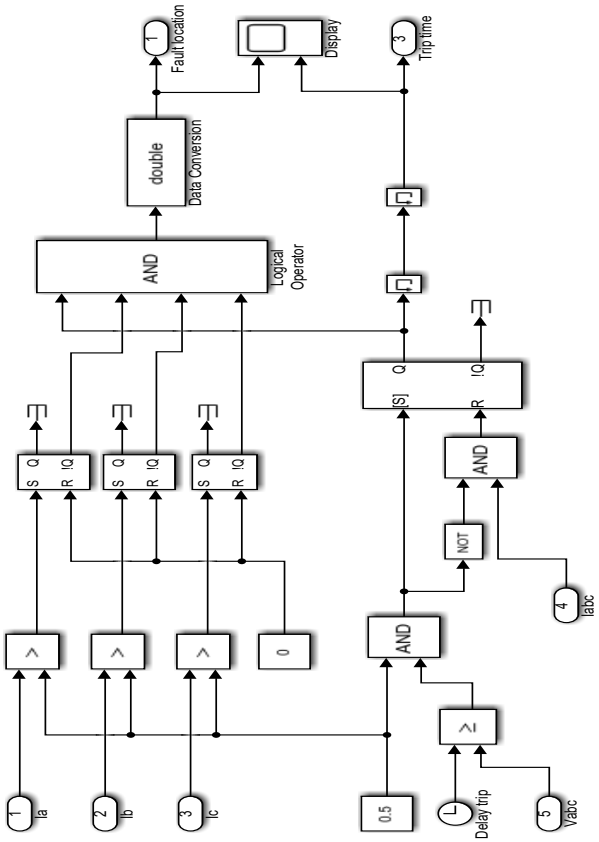

Figure 8: Proposed Algorithm Model.

Measurements will be obtained, parameters will be computed to construct an appropriate transmission line model utilising MATLAB/SIMULINK software, and several adjustments will be required to achieve a robust and efficient network.

The second process is implementing multiple faults on a selected line at various intervals from a designated end terminal. The pre-fault, fault, and post-fault voltage and current waveforms will be studied. These data can then be utilised to determine the location of the fault. Assuming that all the phases are completed successfully, the precise current and voltage signals are gathered and used to pinpoint the exact position of the fault, as illustrated in Figure 9.

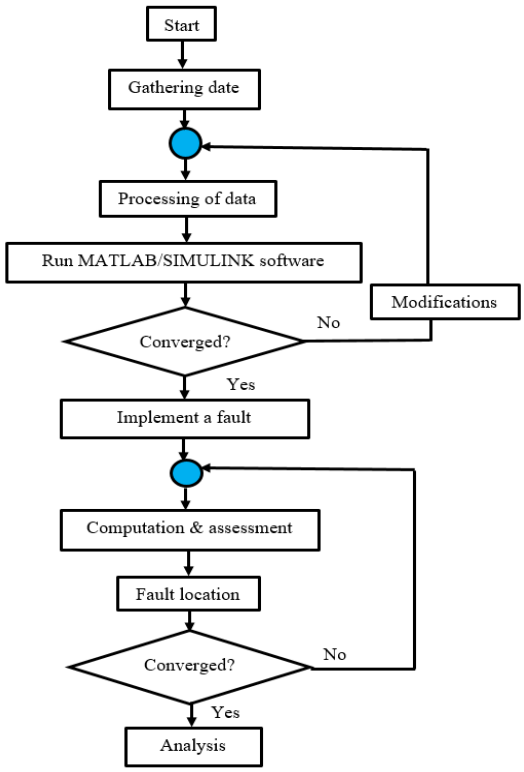

Figure 9: Proposed Algorithm Flow Chart.

The MATLAB/SIMULINK Model is presented in figure 10.

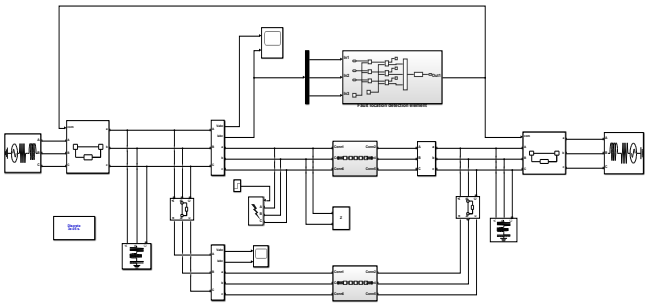

Figure 10: Transmission Line Design with a Single-ended Impedance-based Fault Location Technique.

8. Simulation Results

The Simulation results are as fellows

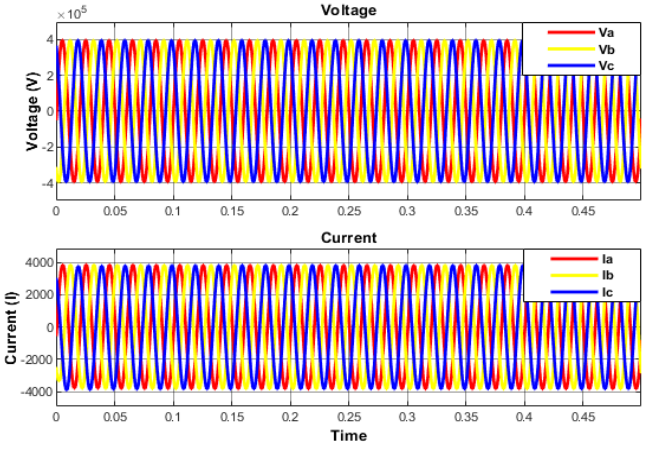

Figure 11: Balanced Voltage & Current Waveforms.

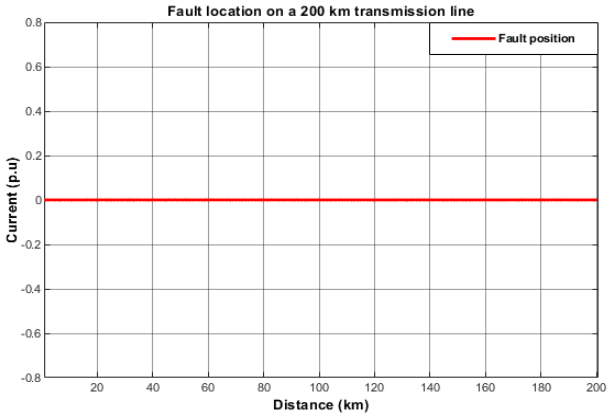

Figure 12: No Located Fault/s.

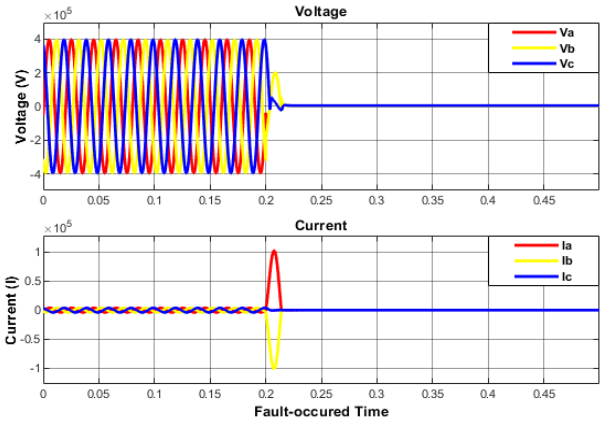

Figure 13: Distorted Voltage & Current Oscillations Induced by L-L Fault.

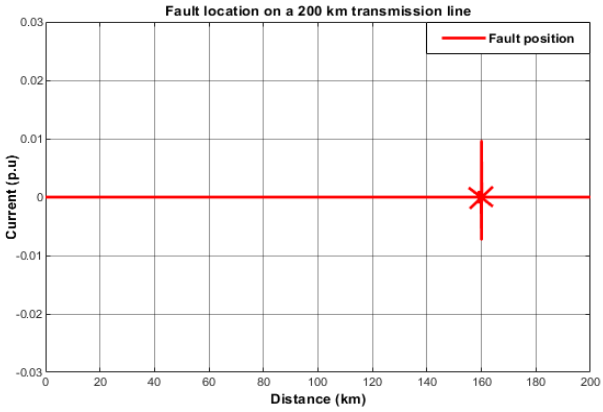

Figure 14: L-L Fault Pin-pointed at 160.5 km.

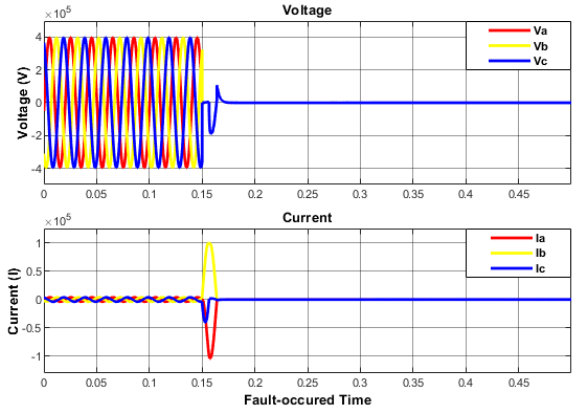

Figure 15: Distorted Voltage & Current Oscillations Induced by L-L-L Fault.

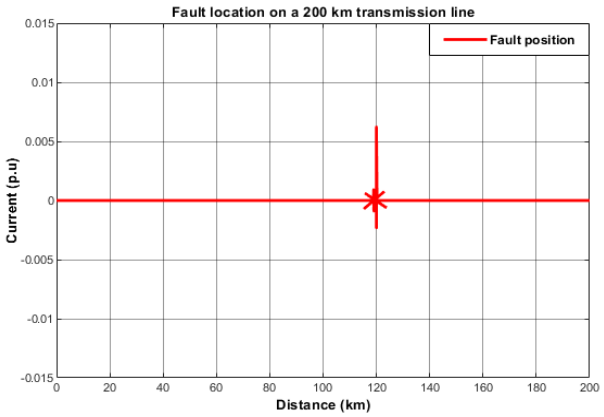

Figure 16: L-L-L Fault Pin-pointed at 119.7 km.

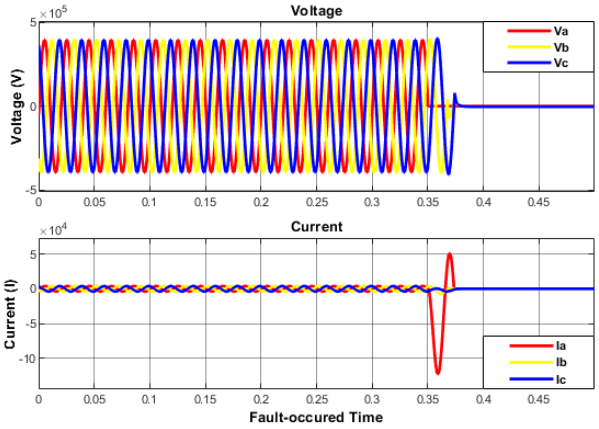

Figure 17: Distorted Voltage & Current Oscillations Induced by L-G Fault.

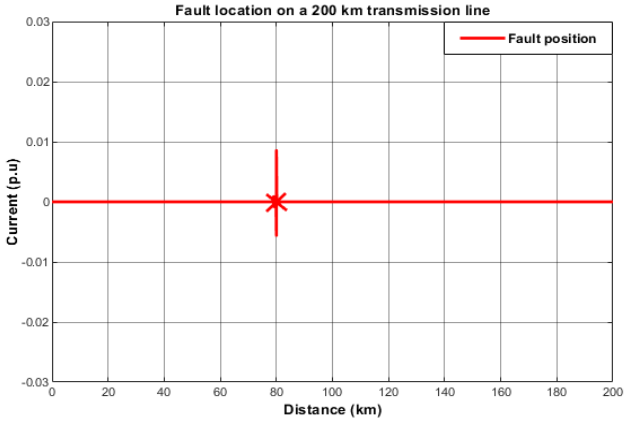

Figure 18: L-G Fault Pin-pointed at 79.9 km.

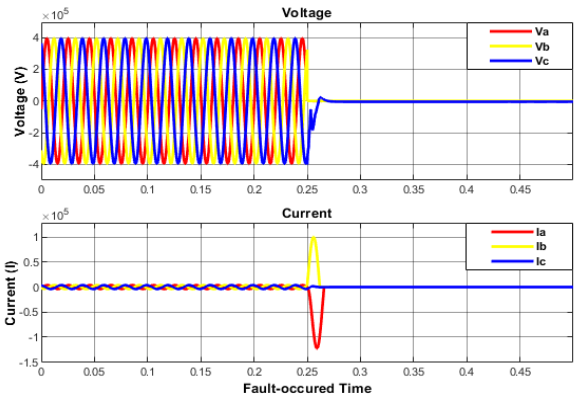

Figure 19: Distorted Voltage & Current Oscillations Induced by L-L-G Fault.

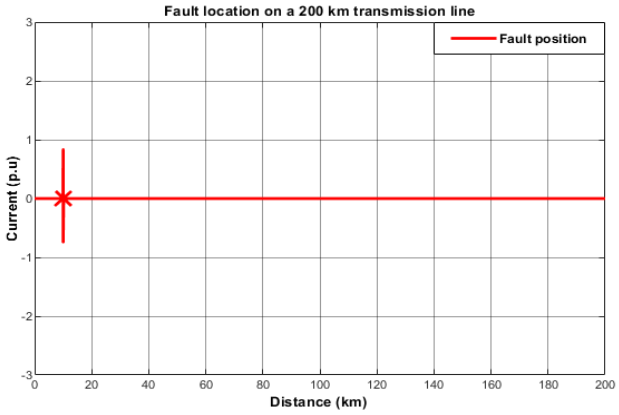

Figure 20: L-L-G Fault Pin-pointed at 10.6 km.

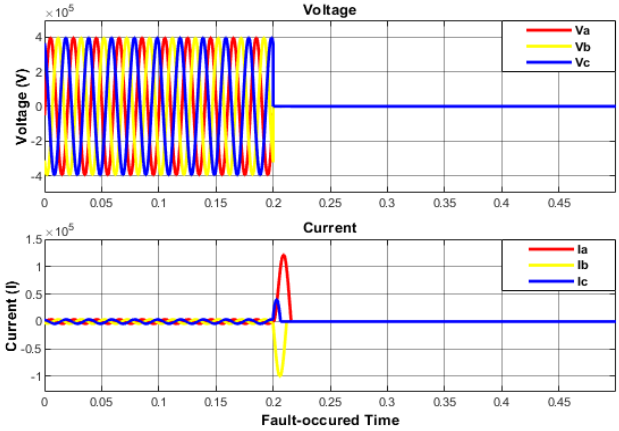

Figure 21: Distorted Voltage & Current Oscillations Induced by L-L-L-G Fault.

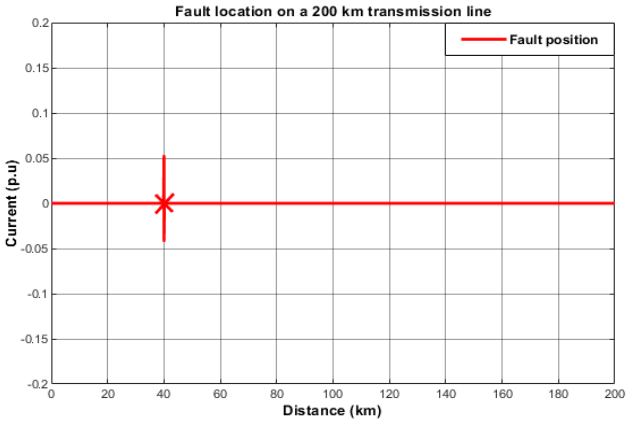

Figure 22: L-L-L-G Fault Pin-pointed at 41 km.

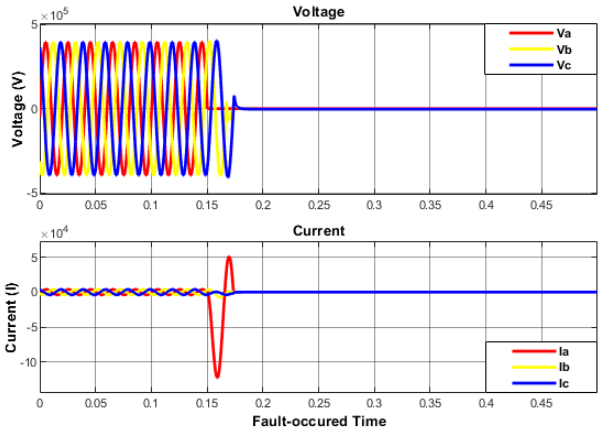

Figure 23: Distorted Voltage & Current Oscillations Induced by L-G Fault.

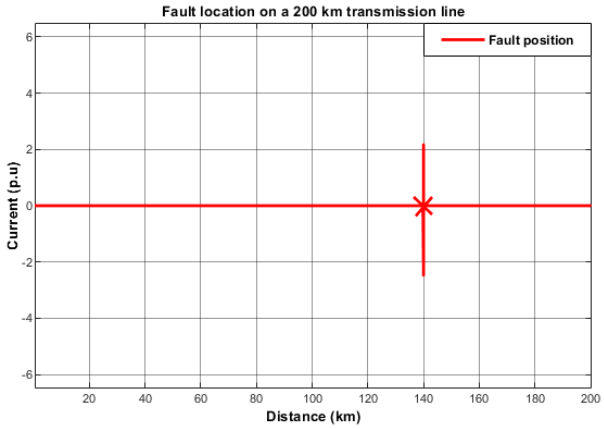

Figure 24: L-G Fault Pin-pointed at 140.9 km.

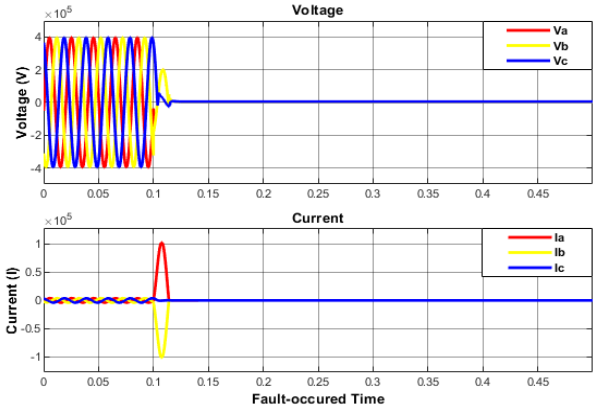

Figure 25: Distorted Voltage & Current Oscillations Induced by L-L Fault.

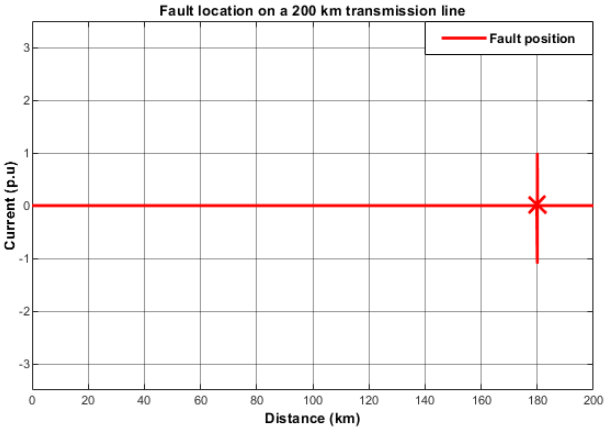

Figure 26: L-L Fault Pin-pointed at 179.2 km.

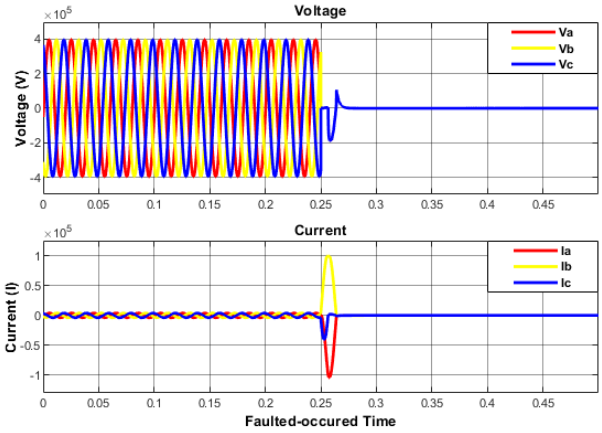

Figure 27: Distorted Voltage & Current Oscillations Induced by L-L-L Fault.

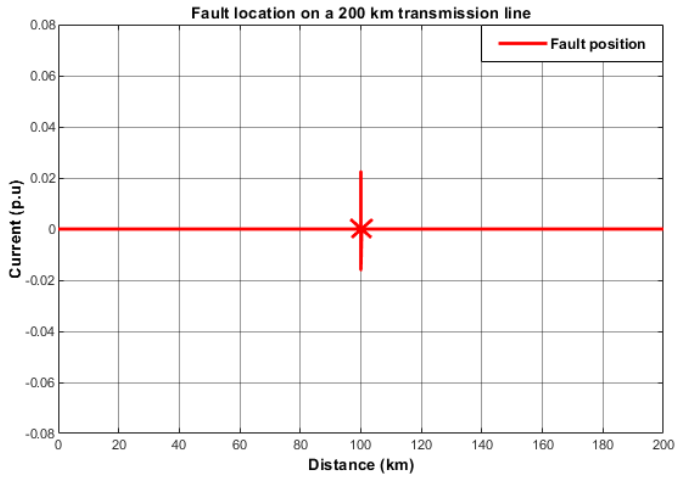

Figure 28: L-L-L Fault Pin-pointed at 98.9 km.

9. Practical Results

In this section, the calculated fault distances using the impedance-based technique are presented in Table 2. The results were obtained from test fault cases involving different faults on a 200 km transmission line. The accuracy of the fault distances was calculated using Equation (32).

$$\%error = \frac{|Actual\ distance - Calculated\ distance|}{Total\ length\ of\ the\ line} \times 100 \qquad (32)$$

Table 2: Error of Estimation: Impedance-based Technique.

Length of the Line (km)	Fault Type	Distance of the Fault (km)	Calculated Fault Distance (km)	Error (%)
200	L-G	80	78.7	1.65
200	L-L	160	146.8	6.6
200	L-L-G	10	6.4	1.8
200	L-L-L	120	115.2	2.4
200	L-L-L-G	40	32.1	3.95
200	L-L	180	176.5	1.75
200	L-G	140	136.7	1.65
200	L-L-L	100	106.4	-3.2

It is observed that the impedance-based technique achieves accurate estimations that are within 1-15% of the designed transmission line.

10. Results Comparison

This section presents the percentage error of the simulated results utilising the proposed algorithm against the calculated results of the impedance-based technique for the designed transmission line. Table 3 illustrates that the proposed algorithm provides a more accurate fault location than the impedance-based technique in all scenarios.

It is shown in Table 3 that the proposed algorithm gives precise predictions with an approximation of less than 1% from the simulated transmission line. As an outcome, it is evident that the proposed algorithm provides more precision.

Table 3: Error of Estimation: Impedance-based Technique against the Proposed Algorithm.

Length of the Line (km)	Fault Type	Distance of the Fault (km)	Calculated Fault Distance (km)	Error (%)
200	L-G	80	1.65%	0.05%
200	L-L	160	6.6%	-0.25%
200	L-L-G	10	1.8%	-0.3%
200	L-L-L	120	2.4%	0.15%
200	L-L-L-G	40	3.95%	-0.5%
200	L-L	180	1.75%	0.4%
200	L-G	140	1.65%	-0.45%
200	L-L-L	100	-3.2%	0.55%

11. Conclusion

This work presents the results of an impedance-based technique to the results of the proposed algorithm on a 200-km transmission line. It was observed that the impedance-based techniques are susceptible to producing errors or incorrect predictions. The presence of faults induced from high impedance sources produces an extra impedance to the ground, which negates the impedance calculation and produces errors in the distance to the fault. This results in inaccuracies that can affect a distance-to-fault estimation by 1-15 % of the overall length. Comparison research was undertaken against the impedance-based techniques to validate the proposed algorithm. The simple reactance technique is considered to be the most basic technique. However, the precision of such a technique suffers in a non-homogenous network due to fault resistance, load current, and distant infeed. The Takagi technique, for instance, is load resistant; however, insensitive to distant infeed. For most scenarios, double-ended impedance-based fault location techniques achieve better outcomes. Studying the fault location usage situation can help determine what extra devices are required to improve the performance of fault-location techniques.

In the present work, an expansion of preceding work [1] introduced a novel algorithm for estimating faults on transmission lines. Extensive trials on a simulated transmission line led to the conclusion of this work. It was observed that the impedance-based technique achieves accurate estimations within 1-15% of the designed transmission line, and the proposed algorithm gives precise predictions with an approximation of less than 1% from the simulated transmission line, reaching a precision of 98.9%. As an outcome, it was evident that the proposed algorithm provides more precision.

In future work, the authors will test the proposed algorithm on a distributed system due to the availability of multiple incomers and feeders connected to the system. Furthermore, the authors will perform a study by simultaneously integrating the proposed algorithm and an artificial neural network (ANN) technique on a complex transmission line and comparing the two algorithms' accuracy.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The author wishes to convey his gratitude to the university for offering him the possibility to work with such a stellar supervisor as Dr. PF Le Roux and offer impactful commentary.

References

[1] M.K Ngwenyama, P.F Le Roux, L.J. Ngoma, "Conventional Method for Electrical Transmission System Fault Location Detection," in 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), 69-76, 2021, doi: 10.1109/ACPEE51499.2021.9437010.

[2] S.M. Hakimi, A. Hasankhani, M. Shafie-khah, M. Lotfi, J.P. Catalão, "Optimal sizing of renewable energy systems in a Microgrid considering electricity market interaction and reliability analysis," Electric Power Systems Research, **10**, 7678, 2022, doi: 10.1016/j.epsr.2021.107678.

[3] P. Bunnoon, "Fault detection approaches to power system: state-of-the-art article reviews for searching a new approach in the future" International Journal of Electrical and Computer Engineering, **3**(4), 553, 2013.

[4] S.A. Aleem, N. Shahid, I.H. Naqvi, "Methodologies in power systems fault detection and diagnosis," Energy Systems, **6**(1), 85-108, 2015, doi: 10.1007/s12667-014-0129-1.

[5] S.S. Gururajapathy, H. Mokhlis, H.A. Illias, "Fault location and detection techniques in power distribution systems with distributed generation: A review," Renewable and sustainable energy reviews, **74**, 949-958, 2017.

[6] A. Keshavarz, R. Dashti, M. Deljoo, H.R. Shaker, "Fault location in distribution networks based on SVM and impedance-based method using online databank generation," Neural Computing and Applications, **34**(3), 1-17, 2021.

[7] J. Wang, Y. Zhang, T. Li, "Equivalent characteristic impedance based hybrid-HVDC transmission line fault location," Electric Power Systems Research, **10**, 7055, 2021, doi: 10.1016/j.epsr.2021.107055.

[8] X. Tong, H. Wen, "A novel transmission line fault detection algorithm based on pilot impedance," Electric Power Systems Research, **10**, 6062, 2020, doi: 10.1016/j.epsr.2019.106062.

[9] J. Van Gompel, D. Spina, C. Develder, "Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks," Applied Energy, **11**, 7874, 2022, doi: 10.1016/j.apenergy.2021.117874.

[10] A.H. Hassanabad, D. Nazeipur, "Design and Simulation of a Control System for Investors in Wind Turbines.", **6**, 6.

[11] X. Wang, Y. Shen, "Fault-tolerant control strategy of a wind energy conversion system considering multiple fault reconstruction," Applied Sciences, **7**, 94, 2018.

[12] M. Nemati, M. Bigdeli, A. Ghorbani, "Impedance-based fault location algorithm for double-circuit transmission lines using single-end data," Journal of Control, Automation and Electrical Systems, **31**(5), 1267-1277, 2020, doi: 10.1007/s40313-020-00620-w.

[13] J. Doria-Garcia, C. Orozco-Henao, L. Iurinic, J.D. Pulgarín-Rivera, "High impedance fault location: Generalized extension for ground faults," International Journal of Electrical Power & Energy Systems, 105387, 2020, doi: 10.1016/j.ijepes.2019.105387.

[14] L. De Andrade, T.P. de Leão, "Impedance-based fault location analysis for transmission lines," in PES T&D 2012, 1-6, 2012, doi: 10.1109/TDC.2012.6281527.

[15] D.W. Thomas, C. Christopoulos, R.J.d.O. Carvalho, E.T. Pereira, "Single and double ended travelling-wave fault location on a MV system," 2004, doi: 10.1049/cp_20040098.

[16] M. Ngwenyama, P. Le Roux, L. Ngoma, "Traveling Wave fault location detection technique for high voltage transmission lines," in 2021 2nd International Conference for Emerging Technology (INCET), 1-7, 2021, doi: 10.1109/INCET51464.2021.9456334.

[17] M.T. Hagh, K. Razi, H. Taghizadeh, "Fault classification and location of power transmission lines using artificial neural network," in 2007 International Power Engineering Conference (IPEC 2007), 1109-1114, 2007.

[18] S. Ekici, "Support Vector Machines for classification and locating faults on transmission lines," Applied soft computing, **12**(6), 1650-1658, 2012, doi: 10.1016/j.asoc.2012.02.011.

[19] A. Prasad, J.B. Edward, C.S. Roy, G. Divyansh, A. Kumar, "Classification of faults in power transmission lines using fuzzy-logic technique," Indian Journal of Science and Technology, 1-6, 2015, doi: 10.17485/ijst/2015/v8i30/77065.

[20] L. Wei, W. Guo, F. Wen, G. Ledwich, Z. Liao, J. Xin, "Waveform matching approach for fault diagnosis of a high-voltage transmission line employing harmony search algorithm," IET generation, transmission & distribution, **4**(7), 801-809, 2010, doi: 10.1049/iet-gtd.2010.0104.

[21] L. Ji, X. Tao, Y. Fu, Y. Fu, Y. Mi, Z. Li, "A new single ended fault location method for transmission line based on positive sequence superimposed network during auto-reclosing," IEEE Transactions on Power Delivery, **34**(3), 1019-1029, 2019, doi: 10.1109/TPWRD.2019.2901835.

[22] A. Di Tomasso, G. Invernizzi, G. Vielmini, "Accurate single-end and double-end fault location by traveling waves: a review with some real applications," in 2019 AEIT International Annual Conference (AEIT), 1-6, 2019.

[23] F. Aboshady, D. Thomas, M. Sumner, "A new single end wideband impedance based fault location scheme for distribution systems," Electric Power Systems Research, 263-270, 2019, doi: 10.1016/j.epsr.2019.04.034.

[24] J. Barati, A. Doroudi, "Novel modified impedance-based methods for fault location in the presence of a fault current limiter," Turkish Journal of Electrical Engineering & Computer Sciences, 1881-1893, 2018, doi: 10.3906/elk-1711-127.

[25] F. Wang, X. Feng, L. Zhang, Y. Du, J. Su, "Impedance-based analysis of grid harmonic interactions between aggregated flyback micro-inverters and the grid," IET Power Electronics, **11**(3), 453-459, 2018, doi: 10.1049/iet-pel.2017.0356.

[26] A. Abu-Siada, M.I. Mosaad, S. Mir, "Voltage–current technique to identify fault location within long transmission lines," IET Generation, Transmission & Distribution, **14**(23), 5588-5596, 2020, doi: 10.1049/iet-gtd.2020.1012.

[27] K. Morgan, W. Gamal, K. Samuel, S.D. Morley, P.C. Hayes, P. Bagnaninchi, J.N. Plevris, "Application of impedance-based techniques in hepatology research," Journal of clinical medicine, **50**, 2020, doi: 10.3390/jcm9010050.

[28] D. Guillen, C. Salas, L. Fernando Sanchez-Gomez, L.M. Castro, "Enhancement of dynamic phasor estimation-based fault location algorithms for AC transmission lines," IET Generation, Transmission & Distribution, **14**(6),1091-1103, 2020, doi: 10.1049/iet-gtd.2019.0051.

[29] C. Zhang, Y. Yu, Y. Wang, M. Zhou, "Takagi–Sugeno fuzzy neural network hysteresis modeling for magnetic shape memory alloy actuator based on modified bacteria foraging algorithm," International Journal of Fuzzy Systems, **22**(4), 1314-1329, 2020, doi: 10.1007/s40815-020-00826-9.

[30] A. Macioł, P. Macioł, B. Mrzygłód, "Prediction of forging dies wear with the modified Takagi–Sugeno fuzzy identification method," Materials and Manufacturing Processes, **35**(6), 700-713, 2020, doi: 10.1080/10426914.2020.1747627.

[31] S. Roostaee, M.S. Thomas, S. Mehfuz, "Experimental studies on impedance based fault location for long transmission lines," Protection and Control of Modern Power Systems, 1-9, 2017.

[32] M.N. Hashim, M.K. Osman, M.N. Ibrahim, A.F. Abidin, "Investigation of features extraction condition for impedance-based fault location in transmission lines," in 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 325-330, 2017, doi: 10.1109/ICCSCE.2017.8284428

[33] A. Dalcastagne, S. Zimath, "A study about the sources of error of impedance-based fault location methods," in 2008 IEEE/PES transmission and distribution conference and exposition: Latin America, 1-6, 2008, doi: 10.1109/TDC-LA.2008.4641697

[34] K. Ramar, H. Low, E.E. Ngu, "One-end impedance based fault location in double-circuit transmission lines with different configurations," International Journal of Electrical Power & Energy Systems, **64**, 1159-1165, 2015, doi: 10.1016/j.ijepes.2014.09.006.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 1, 90-99 (2022)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Estimation of Non-homogeneous Thermal Conductivity using Fourier Heat Equation Considering Uncertainty and Error Propagation

Alexander Núñez[*,1], Fernando Solares[2], Alejandro Crisanto[3]

[1]*Application engineer/Associated Researcher, Measurement Systems, CIATEQ, AC, Querétaro, México*

[2]*Project Leader, Measurement Systems, CIATEQ, AC, Querétaro, México*

[3]*Measurement Systems Manager, Measurement Systems, CIATEQ, AC, Querétaro, México*

ARTICLE INFO

Article history:
Received: 9 September, 2021
Accepted: 25 January, 2022
Online: 21 February, 2022

Keywords:
State-space model
Thermal conductivity
Uncertainty
Fourier hear equation
Partial differential equations

ABSTRACT

The present work develops an estimator for thermal conductivity using a simple experiment implemented in a simulated solid metallic bar. The bar is sectioned in a finite number of segments, lately called nodes, and a discretization of the Fourier heat equation is applied in each node to generate a timed-spaced model of the temperature behavior along the bar. Considering only one-dimensional heat flow, an algorithm based on the temperature measured in each node generates the calculus of the estimated thermal conductivity for every segment of the bar. The calculations of thermal conductivity depends on previous values, such as temperature measurements and adjacent segments thermal conductivity, leading to an error propagation. The analysis of uncertainty related to this values is used to establish a range of values for thermal conductivity estimation. Using the proposed technique allows to calculate thermal conductivity in real time and add to the results a uncertainty estimation for thermal conductivity, providing a more complete information about the measurement procedure. Knowing the uncertainty allows to indicate, in statistical terms, the dispersion of the actual values for thermal conductivity, since the values calculated may vary from real, a higher uncertainty implies a lest reliable calculation according to statistics.

1 Introduction

Modeling physical phenomenons is an increasing topic in science and engineering projects. A very accurate model is not only a validation of theory, but a significantly affordable option for test performing without put in jeopardy any actual equipment or material. Hence, the necessity for accurate models provides a research line leading to expand the state of art of current techniques. Alongside, in order to incorporate realistic behavior and obtain valid simulation results, software algorithms are more powerful and faster. However, the more complicated is the model, the more computational work is required. The need of accurate model must be compared with the processing power available for the project and, in most of the cases, the researches has limited resources. The mathematical model used to describe the behavior of thermal conduction is the Fourier heat equation, a second order-partial-differential equation that includes the time and space behavior of the temperature in a solid body. The solution of this equation has been widely studied in different papers, and it has become in a handy useful tool for introduction in partial-differential equations theory. Tough the solution is known, interpretation and programing in computational simulator may not be as simple, due to the solution is described in literature as a Fourier series [1] and different approaches to deal with the solution [2], [3]. In order to obtain a simplified version of this model, a discrete Fourier equation is developed in [4]. The space discrete model becomes then in a linear system of ordinary differential equations, with the temperature in each segment as the dependent variables and time as only independent variable. This new model fits to results given by the Fourier series solution, with an error estimation related to the physical parameter involved in Fourier equation, density, specific heat and thermal conductivity. These three parameters are non-constant, even two objects of the same material, but they are easily obtained with experimental data. Density is the simplest of them all, due to density is the ratio of mass and volume of the object. Specific heat is more sophisticated, but with a simple temperature test, the value can be calculated as shown in [5], [6] and [7] . A similar experiment may be applied for thermal conductivity, obtaining the average thermal conductivity of the object. Using the space-discrete Fourier equation and a experiment shown in [8] to calculate the thermal conductivity in every

[*]Corresponding Author:Alexander Núñez, Av. Manantiales 23, 52 4421961500 & carlos.nunez@ciateq.mx

segment of the object, the resulting values provides a description of the thermal conductivity along the object. Furthermore, using multiple measure points provides multiple conductivity values, and using a Gauss-Markov estimator, the thermal conductivity in any space position of the bar is calculated. Moreover, the calculation are supposed correct, but there is an uncertainty related to each calculation. In order to describe the error propagation in the calculation of thermal conductivity, an uncertainty analysis is conduced, leading to a range of values in which the correct thermal conductivities are bounded. The work is presented in the following order. Section 2 provides a fundamentals review including the mathematical involved, such as the Fourier equation, its space-continuous and space discrete solution and comparison, and the concept of thermal parameters. Section 3 presents the thermal conductivity estimation using simulated data and the methodology of the experiment, including varying punctual thermal conductivity. In Section 4 it is presented an uncertainty analysis, and the estimation of propagation error. Finally, in section 5, the conclusions and future work are presented.

2 Fundamentals

In this section, the fundamentals and theory applied in the present work are introduced. The general concepts and descriptions are focused referring to the present article. For further information the reader is referred to the selected bibliography included at the document closure.

2.1 Fourier heat equation

The thermal transference in a solid rigid body is represented by the Fourier heat equation, a partial-differential equation developed by Fourier at lately 19th century. The equation balances the thermal energy provided by the temperature evolution in time and the temperature evolution in space. This balance is represented in equation 1. Fourier heat equation is aim of several studies in differential equations solutions. Some of this works provide an analytical solution using Fourier series, but there are numerical solutions also. The solution of Fourier equation may be obtained in analytical form using Fourier series [1], applying Fourier analysis [2] or using semi-analytical approaches [3].

$$\rho C \frac{\partial T}{\partial t} = k \nabla^2 T \tag{1}$$

With

- ρ the body density in $\frac{Kg}{m^3}$
- C the body specific heat in $\frac{J}{Kg^o C}$
- T the body temperature as a function $T(t, x, y, z)$
- k the body thermal conductivity in $\frac{W}{m^o C}$
- ∇ the operator defined as $\nabla = (\frac{\partial}{\partial x} \frac{\partial}{\partial y} \frac{\partial}{\partial z})$

The right side of equation 1 contains the Laplacian of temperature. Solving for $\frac{\partial T}{\partial t}$, the right side becomes in the expression below.

$$\frac{\partial T}{\partial t} = \frac{k}{\rho C} \nabla^2 T \tag{2}$$

The thermal parameters of the body then are combined into a general thermal parameter known as thermal diffusivity α.

$$\frac{k}{\rho C} = \alpha \tag{3}$$

Merging equations 1 and 3 and the definition of ∇ operator, the balance equation may be described as follows.

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) \tag{4}$$

2.2 Continuous solution of 1 space-dimensional Fourier heat equation

The solution presented is based on traditional analytic approach performed in reference [1]. In order to simplify the mathematical involved in this work, the solution of the Fourier heat equation is shown in the 1 dimensional case, thus, the temperature function has only two independent variables, time t and dimension x. Therefore, the Fourier equation in this case is represented en equation 5. For this solution method, the thermal parameter α is considered constant for both of the independent variables.

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2}\right) \tag{5}$$

Applying the method of separable variables, as shown in the refernce [1], the solution is presented as a Fourier series.

$$T(x,t) = \sum_{n=1}^{\infty} B_n sin(n\pi x) e^{-n^2\pi^2 t} \tag{6}$$

$$B_n = 2 \int_0^1 sin(n\pi x) T(x,0) dx \tag{7}$$

Find a numerical solution is an accurate path to avoid the mathematical complexity shown in the continuous solution above. Furthermore, the solution in 3 dimension is more complex and requires larger calculations. For the Fourier equation, one of the most applied numerical methods is the differential discretization proposed by Euler. Using a discrete-space approximation leads to result similar to finite volumes technique, as seen in reference [9].

To achieve a space discretization, the finite differences method is applied [10].

Definition 1 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous and differentiable function. The differential f' at the point $x = a$ is approximated by the following expression*

$$f'(a) = \frac{f(a) - f(b)}{a - b} \tag{8}$$

with b a value sufficiently close to a.

Previous definition is applied for non-even partitions. For the discrete method, even partitions are required, leading to following definition.

Definition 2 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous and differentiable function. Assume the domain of the function is partitioned in sufficiently small intervals with length Δx. The differential $f'(x)$ is given by*

$$f'(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{9}$$

Using definition 2, the first order differential of Fourier heat equation becomes discrete. For the second order differentials, the same algorithm may be used for the first order differential.

Definition 3 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous and at least twice differentiable function. Assume the domain of the function is partitioned in sufficiently small intervals with lenght Δx. The second order differential $f''(x)$ is given by*

$$f''(x) = \frac{f(x + \Delta x) - 2f(x) - f(x - \Delta x)}{(\Delta x)^2} \tag{10}$$

However, the second order differentials in Fourier equation are partial. Moreover, the function T is not defined from real numbers to real numbers. A more accurate interpretation of the second order partial differentials is given in definition 4.

Definition 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous and at least twice differentiable function. Assume the domain of every independent variable x_i of the function is partitioned in sufficiently small intervals with lenght Δx_i. The second order partial differential $\frac{\partial^2 f}{\partial x_i^2}$ of the variable x_i is given by equation 11.*

Using definitions 2 and 4, both sides of the Fourier equation are expressed in discrete form in equations 12-14 and 15 respectively. Remembering that temperature $T(t, x, y, z)$ is function defined from $R^4 \to R$.

$$\frac{\partial^2 T}{\partial x^2} = \frac{T(t, x + \Delta x, y, z) - 2T(t, x, y, z) + T(t, x - \Delta x, y, z)}{(\Delta x)^2} \tag{12}$$

$$\frac{\partial^2 T}{\partial y^2} = \frac{T(t, x, y + \Delta y, z) - 2T(t, x, y, z) + T(t, x, y - \Delta y, z)}{(\Delta y)^2} \tag{13}$$

$$\frac{\partial^2 T}{\partial z^2} = \frac{T(t, x, y, z + \Delta z) - 2T(t, x, y, z) + T(t, x, y, z - \Delta z)}{(\Delta z)^2} \tag{14}$$

$$\frac{\partial T}{\partial t} = \frac{T(t + \Delta t, x, y, z) - T(t, x, y, z)}{\Delta t} \tag{15}$$

In order to compute the differentials appearing in equations 12-15, the rigid body is then meshed in a set of stationary points, such in time as in space.

Definition 5 *A node in a rigid body is a space position with coordinates (x, y, z). The temperature of each node in a time instant t is given by $T(t, x, y, z)$. The distance between two adjacent nodes in space variable x is ΔX. Same description for distance in space variable y and z. The value Δt is known as the sampling period.*

To ignore the Δ values, the nodes are numerated and indexed for each of the four variables.

Definition 6 *The temperature in the node (l, i, j, k) is the result of value the function*
$T(l * \Delta t, i * \Delta x, j * \Delta y, k * \Delta z)$,
and it is identified by the indexing $T_{(i,j,k,l)}$

Rewriting the equations 12-15 using definition 6, the following expressions are obtained.

$$\frac{\partial^2 T_{(l,i,j,k)}}{\partial x^2} = \frac{T_{(l,i+1,j,k)} - 2T_{(l,i,j,k)} + T_{(l,i-1,j,k)}}{(\Delta x)^2} \tag{16}$$

$$\frac{\partial^2 T_{(l,i,j,k)}}{\partial y^2} = \frac{T_{(l,i,j+1,k)} - 2T_{(l,i,j,k)} + T_{(l,i,j-1,k)}}{(\Delta y)^2} \tag{17}$$

$$\frac{\partial^2 T_{(l,i,j,k)}}{\partial z^2} = \frac{T_{(l,i,j,k+1)} - 2T_{(l,j,i,k)} + T_{(l,i,j,k-1)}}{(\Delta z)^2} \tag{18}$$

$$\frac{\partial T}{\partial t}_{l,i,j,k} = \frac{T_{(l+1,i,j,k)} - T_{(l,i,j,x)}}{\Delta t} \tag{19}$$

Equations 16-19 are expressions depending only of node index and the constant Δ values. Using this expressions, the Fourier heat partial-differential equation becomes in an ordinary linear equation, as shown in equation 24

2.3 Sampling considerations

In order to defined the parameters $\Delta t, \Delta x, \Delta y$ and Δz, the sampling theorem postulated by Nyquist and Shanon is applied.

Theorem 7 (Sampling theorem [11]) *Let $s(t)$ a continuous time function and $F(s(t))$ its Fourier transform. Assume that F is bandwidth limited by the frequency f_c. The function $s(t)$ can be reconstructed if and only if the sampling frequency f_s satisfy*

$$f_s > 2f_c \tag{20}$$

For a system response, the frequency f_c is associated to the slowest pole. The value of the slowest pole of the system matches the value of f_c in rad/s. In a thermal system, the value of poles may be extremely near to 0, due to the slow temperature variation. Sampling period T_s is defined as the inverse of sampling frequency f_s. Therefore, sampling periods in thermal system can be significantly extended in comparison to other dynamical systems, such as electrical. Time sampling periods for thermal systems are defined in minutes.

For the space sampling, meaning, the distance between the nodes, a slight variation of sampling theorem is used.

Lemma 8 *Let $R(x)$ be a one dimensional space function with domiain $[0, l]$ and $F(R(x))$ its space Fourier transform. F domain is upper bounded by $\frac{1}{l}$. Signal R can be reconstructed at 99.9% of accuracy if the sample nodes are placed with a distance d_s*

$$d_s = \frac{l}{10} \tag{21}$$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{f(x_1, \cdots, x_i + \Delta x_i, \cdots, x_n) - 2f(x_1, \cdots, x_i, \cdots, x_n) + f(x_1, \cdots, x_i - \Delta x_i, \cdots, x_n)}{(\Delta x_i)^2} \tag{11}$$

$$\frac{T_{(l+1,i,j,k)} - T_{(l,i,j,x)}}{\Delta t} = \\ = \alpha\left(\frac{T_{(l,i+1,j,k)} - 2T_{(l,i,j,k)} + T_{(l,i-1,j,k)}}{(\Delta x)^2} + \frac{T_{(l,i,j+1,k)} - 2T_{(l,i,j,k)} + T_{(l,i,j-1,k)}}{(\Delta y)^2} + \frac{T_{(l,i,j,k+1)} - 2T_{(l,j,i,k)} + T_{(l,i,j,k-1)}}{(\Delta z)^2}\right) \tag{23}$$

Using these approach, the minimal distance between nodes along x axis can be calculated as

$$\Delta x = \frac{l_x}{10} \tag{22}$$

with l_x the length of the body along x axis. Hence, the minimal required nodes per axis is eleven.

Using equation 24 and solving for $T_{(l+1,i,j,k)}$, it is possible to calculate the temperature for a fixed node, using information of adjacent nodes in the previous sample.

2.4 Space-state model for temperature distribution

As shown in previous subsection, the discrete approach of Fourier equation leads to a numerical solution, and it is widely used in discrete estimations. However, in order to propose a continuous control algorithm as well as a continuous state observer a continuous space-state model is required. For the continuous model the following considerations are imperative

- Every state of the model represents the temperature of a particular node in the rigid body.
- The nodes on the body surface experiment conduction and convection at the same time, so the equations for these nodes are different from the rest.
- Fourier equation defines the system natural response, no input is applied.
- The number of nodes is previously calculated and exposed in subsection 2.3.
- Only space discretization is being considered, time remains continuous.

Following the previous considerations, the time differentials of equation 24 remains continuous. The l index is ignored, hence the equations is rewritten as seen on equation **??**

The symbol $\dot{T}_{(i,j,k)}$ is used as differential representation. Using equation **??**, the continuous differential of each node temperature is calculated as a linear combination of the rest of node temperatures, in particular only the adjacent nodes contribute. Therefore, the temperatures are presented in a space-state representation.

2.5 One-dimensional state-space representation

In order to explain the model in more detail, the first approach is made for only one space dimension heat flow, considering n nodes evenly distributed.

Excluding the border nodes, the remaining $n-2$ nodes are represented by equation **??**, but ignoring the j and k indexes and the y and z respective differentials. Hence

$$\dot{T}_{(i)} = \alpha\frac{T_{(i+1)} - 2T_{(i)} + T_{(i-1)}}{(\Delta x)^2}, 1 \geq i \geq n-1 \tag{25}$$

$$\dot{T}_{(i)} = \frac{\alpha}{(\Delta x)^2}T_{(i+1)} - 2\frac{\alpha}{(\Delta x)^2}T_{(i)} + \frac{\alpha}{(\Delta x)^2}T_{(i-1)} \tag{26}$$

In matrix condensed form, the space state is represented as seen in equation 31.

For the border nodes, thermal energy interaction is defined by two independent processes, conduction and convection. In convection process, Newton law is applied.

$$\rho c\frac{dT}{dt} = hl(T - Ta) \tag{27}$$

Solving for $\frac{dT}{dt}$

$$\dot{T}_{(i)} = \frac{dT_{(i)}}{dt} = \frac{hl}{\rho c}(T_{(i)} - Ta) \tag{28}$$

Moreover, a border node has only one adjacent node in the one-dimensional approach. Hence, the fourier equation is reduced.

$$\begin{aligned}\dot{T}(0) &= -\frac{\alpha}{(\Delta x)^2}T_{(0)} + \frac{\alpha}{(\Delta x)}T_{(1)} \\ \dot{T}(n) &= -\frac{\alpha}{(\Delta x)^2}T_{(n)} + \frac{\alpha}{(\Delta x)}T_{(n-1)}\end{aligned} \tag{29}$$

Merging equations 28 and 29, the following expression is obtained.

$$\begin{aligned}\dot{T}(0) &= \frac{hl}{\rho c}(T_{(0)} - Ta) - \frac{\alpha}{(\Delta x)^2}T_{(0)} + \frac{\alpha}{(\Delta x)}T_{(1)} \\ \dot{T}(n) &= \frac{hl}{\rho c}(T_{(n)} - Ta) - \frac{\alpha}{(\Delta x)^2}T_{(n)} + \frac{\alpha}{(\Delta x)}T_{(n-1)}\end{aligned} \tag{30}$$

The complete model obtained from Fourier and Newton equation is presented in equation 32. This expression represents a continuous space-state model for the temperature of the nodes in the body.

$$\dot{T}_{(i,j,k)} = \frac{\partial T_{(i,j,k)}}{\partial t} = \\ = \alpha\left(\frac{T_{(i+1,j,k)} - 2T_{(i,j,k)} + T_{(i-1,j,k)}}{(\Delta x)^2} + \frac{T_{(i,j+1,k)} - 2T_{(i,j,k)} + T_{(i,j-1,k)}}{(\Delta y)^2} + \frac{T_{(l,i,j,k+1)} - 2T_{(j,i,k)} + T_{(i,j,k-1)}}{(\Delta z)^2}\right) \tag{24}$$

$$\begin{bmatrix} \dot{T}_{(1)} \\ \dot{T}_{(2)} \\ \dot{T}_{(3)} \\ \vdots \\ \dot{T}_{(i)} \\ \vdots \\ \dot{T}_{(n-2)} \\ \dot{T}_{(n-1)} \end{bmatrix} = \frac{\alpha}{(\Delta x)^2} \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} T_{(0)} \\ T_{(1)} \\ T_{(2)} \\ T_{(3)} \\ \vdots \\ T_{(i)} \\ \vdots \\ T_{(n-2)} \\ T_{(n-1)} \\ T_n \end{bmatrix} \tag{31}$$

$$\begin{bmatrix} \dot{T}_{(0)} \\ \dot{T}_{(1)} \\ \dot{T}_{(2)} \\ \vdots \\ \dot{T}_{(i)} \\ \vdots \\ \dot{T}_{(n-2)} \\ \dot{T}_{(n-1)} \\ \dot{T}_{(n)} \end{bmatrix} = \begin{bmatrix} \frac{hl}{\rho c} - 2\frac{\alpha}{(\Delta x)} & \frac{\alpha}{(\Delta x)} & \cdots & 0 & \cdots & 0 & 0 \\ \frac{\alpha}{(\Delta x)^2} & -2\frac{\alpha}{(\Delta x)^2} & \cdots & 0 & \cdots & 0 & 0 \\ 0 & \frac{\alpha}{(\Delta x)^2} & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & -2\frac{\alpha}{(\Delta x)^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \frac{\alpha}{(\Delta x)^2} & 0 \\ 0 & 0 & \cdots & 0 & \cdots & -2\frac{\alpha}{(\Delta x)^2} & \frac{\alpha}{(\Delta x)^2} \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \frac{hl}{\rho c} - 2\frac{\alpha}{(\Delta x)} \end{bmatrix} \begin{bmatrix} T_{(0)} \\ T_{(1)} \\ T_{(2)} \\ \vdots \\ T_{(i)} \\ \vdots \\ T_{(n-2)} \\ T_{(n-1)} \\ T_n \end{bmatrix} + \begin{bmatrix} \frac{hl}{\rho c} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ \frac{hl}{\rho c} \end{bmatrix} T_a \tag{32}$$

Though the model is no linear, due to the appearance of environmental temperature, the linear system decomposition matches with the $n \times n$ matrix in equation 32. Using these matrix, the stability of the system can be analyzed obtaining the eigenvalues. Moreover, the model can be extended, including external energy inputs and output temperature sensors, leading to a controllability and observability analysis.

2.6 Thermal conductivity

Thermal conductivity is one of the three thermodynamical parameters involved in Fourier heat equation (eq. 1), and there are previous analysis remarking the heat coefficients, as in [12]. Thermal conductivity, definition and formulas are introduced in this subsection.

Definition 9 *Thermal conductivity k is the ratio of heat per length unit and temperature variation.*

$$k = \frac{QL}{A\Delta T} \tag{33}$$

With

- k thermal conductivity in $\frac{W}{m^2}$
- Q the heat flow
- L the body length
- A the traversal section area
- ΔT the temperature variation

A comparative method for calculating thermal conductivity is shown in [8]. The comparison requires a material with known thermal conductivity k_2 coupled to an unknown thermal conductivity material k_1. Considering no heat loss, heat flow remains constant in steady state. Thermal conductivity k_1 can be calculated using equation 34.

$$k_1 = k_2 \frac{A_2 \Delta T_2 L_1}{A_1 \Delta T_1 L_2} \tag{34}$$

2.7 Non-homogeneous thermal conductivity

Furthermore in this section, lets consider the thermal conductivity of the bar modeled by equation 32 is not constant. Hence, the case a non-homogeneous thermal conductivity is analyzed. Thermal conductivity, as explained in section 2.6, represents the ratio of heat flow along a determinate length and the temperature variation. Then, thermal conductivity is defined in length intervals. Considering the nodes defined in definition 5, the segment between every couple of adjacent nodes may have a different thermal conductivity. This distribution is illustrated in figure 1. Considering non-homogeneous thermal conductivity, the expression 3 can not be used as a simplification for thermal parameters. In order to obtain a model that considers non-homogeneous thermal conductivity, other assumption is applied. Using equation 25, the epxression is rewritten to appreciate the different thermal conductivities, in equation 35 consider K_m as the total thermal conductivity of the two segments. Then, in equation 36, the nodes are isolated from every segment, hence thermal conductivity for every segment is used. All the terminus are regrouped to obtain equation 37.

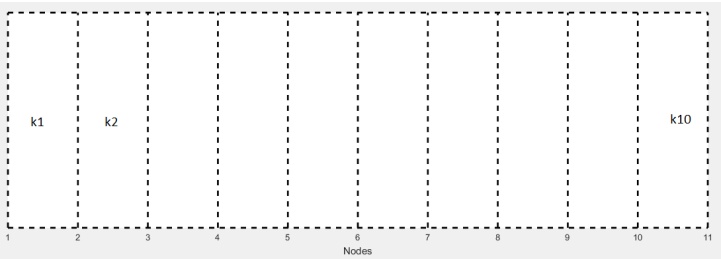

Figure 1: Rigid body with non-homogeneous thermal conductivity

$$\dot{T}_{(i)} = \frac{K_m}{\rho c(\Delta x)^2}(T_{(i+1)} - 2T_{(i)} + T_{(i-1)}) \tag{35}$$

$$\dot{T}_{(i)} = \frac{k_i}{\rho c(\Delta x)^2}(T_{(i+1)} - T_{(i)}) + \frac{k_{i-1}}{\rho c(\Delta x)^2}(-T_{(i)} + T_{(i-1)}) \tag{36}$$

$$\dot{T}_{(i)} = \frac{1}{\rho c(\Delta x)^2}(k_i T_{(i+1)} - (k_i + k_{i-1})T_{(i)} + k_{i-1}T_{(i-1)}) \tag{37}$$

For the border nodes a similar analysis is applied. Considering that in border nodes the conduction is only present in one side, the expression for border nodes is presented in equation 38.

$$\begin{aligned} \dot{T}(0) &= \frac{hl}{\rho c}(T_{(0)} - Ta) + \frac{k_0}{\rho c(\Delta x)^2}(T_{(1)} - T_{(0)}) \\ \dot{T}(n) &= \frac{hl}{\rho c}(T_{(n)} - Ta) + \frac{k_0}{\rho c(\Delta x)^2}(T_{(n-1)} - T_{(n)}) \end{aligned} \tag{38}$$

3 Thermal conductivity estimation

This section presents a method for thermal conductivity characterization in a solid rigid body. The experiment is based in the technique of concentric cut bars shown in [8] for homogeneous thermal conductivity. Merging the described experiment with the node representation of a solid body introduced in section 2.7, a method to determinate thermal conductivity per node in real time. The experiment presented in [8] consist of two different materials aligned in concentrically form. A heat source is placed in an edge of material A. This material has a known thermal conductivity. As the heat is transferred along the materials by conduction, the temperature at the edge of the materials is measured. Three points are measured, the edge of material A (nearest from heat source), the contact point of two materials and edge of material B (farthest to heat source). Only one directional heat flow is allowed, so the materials are isolated. Once thermal equilibrium point is achieved, the values of temperature are used in equation 40 to calculate thermal conductivity of material B.

Thermal conductivity can be calculated using the equation 39, as seen on section 2.6

$$k = \frac{QL}{A\Delta T} \tag{39}$$

Lets consider one dimensional heat flow, an isolated rigid body composed of two different materials, with respectively thermal conductivities k_a and k_b and a heat source located in one of the body's edges. Using equation 34 with the steady state temperature measurements and considering the cross area is equal, thermal conductivity k_b of an unkown material is calculated with equation 40, with ΔT_a the temperature difference between the borders of material a, L_a the length of material a and ΔT_b the temperature difference between the borders of material b, L_b the length of material b.

$$k_b = k_a \frac{\Delta T_a L_b}{\Delta T_b L_a} \tag{40}$$

Knowing the nominal thermal conductivity in material b, and using the approach given by the node representation defined in 5 and equation 36, in steady state the heat flow in each of the nodes is equal to 0. This fact is corroborated for the model presented in equation 32, due to equilibrium point of the space state system is calculated when the differentials of each node temperature is zero. $\dot{T}_i = 0$. Hence, using Fourier heat equation represented in the form of equation 25

$$\alpha \frac{T_{(i+1)} - 2T_{(i)} + T_{(i-1)}}{(\Delta x)^2} = 0 \tag{41}$$

Merging this expression with equation 36 leads to

$$\frac{k_i}{\rho c(\Delta x)^2}(T_{(i+1)} - T_{(i)}) + \frac{k_{i-1}}{\rho c(\Delta x)^2}(-T_{(i)} + T_{(i-1)}) = 0 \tag{42}$$

$$k_i(T_{(i+1)} - T_{(i)}) + k_{i-1}(-T_{(i)} + T_{(i-1)}) = 0 \tag{43}$$

With K_i the thermal conductivity of the segment between the nodes i and $i+1$. From previous equation, thermal conductivity of segment i can be calculated using thermal conductivity of segment i and temperature measurements in nodes $i-1$, i and $i+1$, as seen on equation 44

$$k_i = k_{i-1} \frac{(T_i - T_{i-1})}{(T_{i+1} - T_i)} \tag{44}$$

Using the calculated conductivity of the material k_b provided by equation 34, the first iteration k_0 becomes k_b. Then, for the following nodes, the resulting equations are obtained

$$\begin{aligned} k_0 &= k_b \\ k_1 &= k_b \frac{(T_1 - T_0)}{(T_2 - T_1)} \\ k_2 &= k_1 \frac{(T_2 - T_1)}{(T_3 - T_2)} \\ &\vdots \end{aligned} \tag{45}$$

3.1 Simulated thermal conductivity estimation

The procedure above is used to estimate thermal conductivity for simulated data. Using temperature derived from the model described in section 2.7. For the simulation is performed in the complement Simulink of Matlab. For the simulation the following values are considered.

- k_1= 220 (Aluminum)
- L=0.5 m
- Δx = 0.05m
- ρ = 2675 $\frac{kg}{m^3}$
- c= 897 $\frac{J}{kg^oC}$

The aluminum bar is represented in nodes and the space state equation given by 32 is programed in simulink using a value of $\Delta x = 0.05m$, and the temperature values for each node are obtained. The parameter of thermal conductivity in the segment between nodes six and seven is set to a value of 187, the rest of the nodes thermal conductivity are set to 220. Using the temperature values for each node, thermal conductivity of each segment is the calculate using equation 44, considering the nominal thermal conductivity of aluminum for the initial thermal conductivity. The results for thermal conductivities in each segment are presented in table 1.

Table 1: Thermal conductivity of each segment in the simulated aluminum bar

Segment	Left temperature ($\circ C$)	Right temperature ($\circ C$)	k
1	45.1478	44.1361	220
2	44.1361	43.1244	220
3	43.1244	42.1126	220
4	42.1126	41.1009	220
5	41.1009	40.0892	220
6	40.0892	38.8991	187.03
7	38.8991	37.8874	220
8	37.8874	36.8756	220
9	36.8756	35.8639	220
10	35.8639	34.8522	220

The temperature in the nodes are presented in figure 3. Results of table 1 in graphic form are observed in figure 2, the variation of temperature along the bar is small, hence the slope of the graphic may not vary even the thermal conductivity is different in the red segment. The method not only provides an accurate value of thermal conductivity, it also isolates the only thermal conductivity variation as if it was related to another material incrusted in segment 5. Figure 4 shows the evolution of segment 5 thermal conductivity along time.

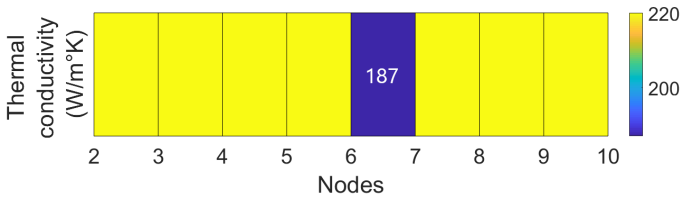

Figure 2: Thermal conductivity along the bar in steady state

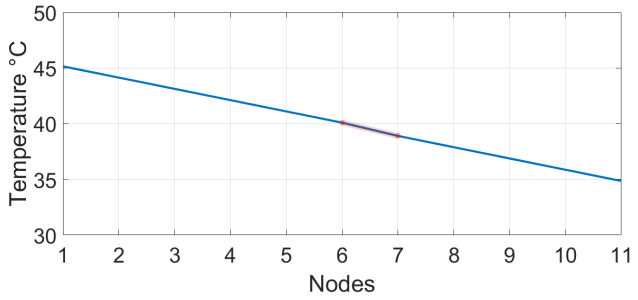

Figure 3: Temperature along the bar in steady state

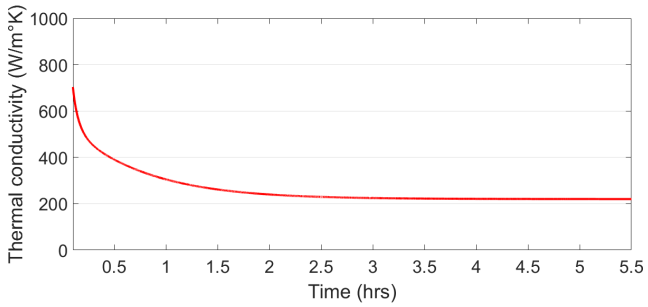

Figure 4: Thermal conductivity of segment 5

4 Uncertainty analysis

The results presented in section 3 are consistent with the calculation of thermal conductivity for the anomaly detected in the simulated bar. Since this is a simulation, no external consideration has been made. However, the data obtained for the temperature measurements are limited by the uncertainty of the sensor used for this purpose. Moreover, with each calculation a new thermal conductivity is obtained, and this new value becomes an uncertainty source for the new calculation, leading to a error propagation. In this section, an analysis of uncertainty and error propagation is presented in order to obtain a range of values to bound the thermal conductivity values obtained using the method presented in section 3.

Definition 10 (Uncertainty) *Uncertainty is a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand [13]*

Let y be the final measurement of a procedure and variables $x_1, x_2, \cdots, x_n$ the uncertainty sources. Assume all the variable x_i are not correlated. Then, y can be expressed as a function of variables x_i as seen in equation 46.

$$y = f(x_1, x_2, \cdots, x_n) \tag{46}$$

The combined uncertainty $U_c(y)$ for the measurement y is calculed by equation 47.

$$U_c^2(y) = \sum_{i=1}^{n} (\frac{\partial f}{\partial x_i})^2 u^2(x_i) \tag{47}$$

With $u(x_i)$ the corresponding uncertainty of each x_i variable.

The term $\frac{\partial f}{\partial x_i}$ is known as the sensibility coefficient for variable x_i.

In section 3, equation 44 is used to calculate a thermal conductivity using temperature measurement points. Matching this with equation 46, leads to a function to calculate thermal conductivity k_{i+1} using as inputs: thermal conductivity k_i and temperatures T_{i-1}, T_i, T_{i+1}, as seen on equation 48

$$y = k_{i+1} = f(k_i, T_{i-1}, T_i, T_{i+1}) = k_i \frac{(T_i - T_{i-1})}{(T_{i+1} - T_i)} \tag{48}$$

The sensibility coefficients for this measurement are

$$\frac{\partial f}{\partial k_i} = \frac{T_i - T_{i-1}}{T_{i+1} - T_i} \tag{49}$$

$$\frac{\partial f}{\partial T_{i-1}} = \frac{k_i}{T_i - T_{i+1}} \tag{50}$$

$$\frac{\partial f}{\partial T_i} = k_i \frac{T_{i+1} - T_{i-1}}{(T_{i+1} - T_i)^2} \tag{51}$$

$$\frac{\partial f}{\partial T_{i+1}} = k_i \frac{T_{i-1} - T_i}{(T_{i+1} - T_i)^2} \tag{52}$$

Considering the measurements are made with an extremely accurate temperature meter, the uncertainty related to temperature measurement in each node is set as $u(T_k) = 0.05^o C$. Hence, for the first segment the thermal conductivity is taken as nominal, then $k_0 = 220$. Using equations 53-56, and information in table 1, the sensibility coefficients for output k_1 are calculated next.

$$\frac{\partial f}{\partial k_1} = \frac{T_1 - T_0}{T_2 - T_1} = 1 \tag{53}$$

$$\frac{\partial f}{\partial T_0} = \frac{k_0}{T_1 - T_2} = 217 \tag{54}$$

$$\frac{\partial f}{\partial T_1} = k_0 \frac{T_2 - T_0}{(T_2 - T_1)^2} = 434 \tag{55}$$

$$\frac{\partial f}{\partial T_2} = k_0 \frac{T_0 - T_1}{(T_2 - T_1)^2} = 217 \tag{56}$$

Using equation 47, the uncertainty related to output k_1 is then calculated.

$$U_c^2(k_1) = \\ [(1)(0)]^2 + [(217)(0.05)]^2 + [(434)(0.05)]^2 + [(217)(0.05)]^2 \tag{57}$$

$$U_c(k_1) = 26.57 \tag{58}$$

Thus, the interval of values for thermal conductivity are [193.43 246.57]. Even if the range is quite near the expected value, the iterative method to calculate the rest of thermal conductivities in the bar leads to an error propagation, increasing the uncertainty value

for every segment of the bar. In table 2 the values of uncertainty for each thermal conductivity are shown.

Table 2: Uncertainty for each thermal conductivity

Segment	Sensibility coefficients $(k_{i-1}, T_{i-1}, T_i T_{i+1})$	$U_c(k_{i-1})$ All $u(T) = 0.05$	$U_c(k_i)$
k_1	(1, 217, 434, 217)	0	26.57
k_2	(0.99,217 ,434 ,217)	26.57	37.66
k_3	(1,217 ,434 ,217)	37.66	46.13
k_4	(1,217 ,434 ,217)	46.13	53.26
k_5	(0.85,217 ,434 ,217)	53.26	52.53
k_6	(1.17,157 ,369 ,184)	52.53	65.63
k_7	(1,217 ,434 ,217)	65.63	70.82
k_8	(1,217 ,434 ,217)	70.82	75.67
k_9	(1,217 ,434 ,217)	75.67	80.22
k_{10}	(1,217 ,434 ,217)	80.22	84.52

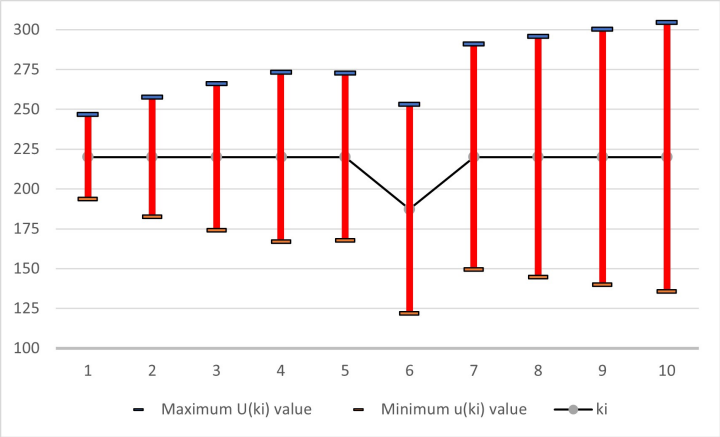

Figure 5: Thermal conductivity uncertainty along the bar in steady state

Table 2 shows that the uncertainty in the segment 10 of the bar is 84, leading to a range between 136 and 304 W/m^2. It is clear that every recursion increases the value of uncertainty associated.

5 Conclusions and further work

The model proposed fits the results for the numerical solution of the Fourier Heat equation, using a correct spacing it is possible to reconstruct the temperature distribution along the aluminum bar. The discrete approximation of the partial differential is a well discussed method, however, maintaining the time variable as continuous provides a useful model in state-space representation with applications in control theory such as control design and implementation, state observers, mathematical simulation, linear algebra tools, differential geometry theory, among others, besides, the information of the space-discrete model is preserved, at least the 99 percent of the general partial differential equation of Fourier Heat. Moreover, the thermodynamic parameters associated to the model are normally non-homogeneous in a rigid body, particularly the thermal conductivity. The reinterpreted experiment to estimate the thermal conductivity distribution along the bar allows to calculate the thermal conductivity variations along the bar, using simulated data in Matlab. The model provides an accurate estimation of the temperature distribution, and the estimation method for thermal conductivity is proven using simulated data, calculating then the thermal conductivity on-line. Using this technique allows to detect and isolate thermal conductivity changes in a physic thermodynamic system. The thermal conductivities obtained from this procedure matches the originally proposed in the simulated model, with an error below 0.05 units. Even that a segment of the bar has a different thermal conductivity, the procedure is capable of recognize this variation and calculate an estimate value. The implications of this feature are 2: First, the model now can be completed considering non-homogeneous materials, and second, the procedure can predict the location for thermal loads along the bar, this could be generated by other materials addition in an specific location. The results in simulation are quite near to perfect, since the estimation error is low, but this could change for a real time implementation. For instance, the measurement instruments used for the temperature may not be calibrated. The uncertainty of measurements could have a direct impact in the estimation of thermal conductivity, due to a recursion equation using in the procedure. When a temperature instrument is used for the node temperature measurement, the instrument has associated a measurement uncertainty. Considering a extremely accurate temperature measurement instrument, and obtaining the uncertainty related to every segment, it is possible to estimate the combined uncertainty for each segment thermal conductivity. Thermal conductivity uncertainty increases with every new calculation, leading to an error of 84 units in the worst case scenario. This error propagation impacts directly in the state-space model and the accuracy of thermal conductivity estimation is reduced. Thus, the error propagation needs to be corrected in future works to avoid this problem and provide an optimal model for the rigid body temperature distribution. Knowing the measurement uncertainty is a useful information in order to validate the procedures and maintain conformance with normative standards, but it also provides a reference for the reliability of the calculation, in statistical terms, the calculation of thermal conductivity may vary at most 84 units, leading to different values in each repeated experiment.

The analysis presented in this article has a limited approach for many physical characteristics, in order to maintain the models and calculations simple to make a comparative with the simulation data. Hence, the assumptions conceptualized before the article development had been confirmed, e.g. the model using finite difference and state space model can be used to determine thermal conductivity in different segments of a bar, the next step in this research is to study more complex situations. For instance, the geometry of the element in the case study must be different than a rectangular bar. More complex geometries lead to a different node and segment definition. The heat flux is considered uni dimensional in this approach, this is made for matching the requirements of the documented bibliography for the experiment to determinate thermal conductivity. The analysis of multidimensional heat flow is closer to real world conditions. The uncertainty is calculated using the minimum necessary number of nodes, but using software tools on Finite Element, the number of nodes is not a problem for the calculation. A follow-up article may

explore the possibility of merging Finite Element techniques with uncertainty propagation and analyze the error derived for each node calculation.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The author acknowledges CIATEQ,AC for funding this research paper.

References

[1] R. Guenther, J. Lee, Partial differential equation of mathematical physics and integral equations, Dover Publications, 1996.

[2] R. Danchin, P. B. Mucha, New Maximal Regularity Results for the Heat Equation in Exterior Domains, and Applications, 101–128, Springer New York, New York, NY, 2013, doi:10.1007/978-1-4614-6348-1_6.

[3] M. Oane, F. Scarlat, I. N. Mihailescu, "The semi-analytical solution of the Fourier heat equation in beam â€“ 3D inhomogeneous media interaction," Infrared Physics and Technology, **51**(4), 344–347, 2008, doi:https://doi.org/10.1016/j.infrared.2007.11.002.

[4] J. H. P. Vázquez, C. A. N. Martín, O. R. U. Gosebruch, H. J. Z. Osorio, C. E. C. González, "Analysis of the Dynamic Behavior of the Temperature Distribution Inside a Domestic Refrigerator," in ICONS 2018, The Thirteenth International Conference on Systems, 63–66, IARIA, 2018.

[5] C. H. Huang, M. N. özisik, "Direct Integration Approach for Simultaneously Estimating Temperature dependent Thermal Conductivity and Heat Capacity," Numerical Heat Transfer, Part A: Applications, **20**(1), 95–110, 1991, doi:10.1080/10407789108944811.

[6] C. Coskun, Z. Oktay, N. Ilten, "A new approach for simplifying the calculation of flue gas specific heat and specific exergy value depending on fuel composition," Energy, **34**(11), 1898–1902, 2009, doi:https://doi.org/10.1016/j.energy.2009.07.040.

[7] X. Qian, J. Xue, Y. Yang, S. W. Lee, "Thermal Properties and Combustion-Related Problems Prediction of Agricultural Crop Residues," Energies, **14**(15), 2021, doi:10.3390/en14154619.

[8] N. W. Pech-May, Á. Cifuentes, A. Mendioroz, A. Oleaga, A. Salazar, "Simultaneous measurement of thermal diffusivity and effusivity of solids using the flash technique in the front-face configuration," Measurement Science and Technology, **26**(8), 085017, 2015, doi:10.1088/0957-0233/26/8/085017.

[9] S. Han, "Finite volume solution of two-step hyperbolic conduction in casting sand," International Journal of Heat and Mass Transfer, **93**, 1116–1123, 2016, doi:https://doi.org/10.1016/j.ijheatmasstransfer.2015.10.061.

[10] N. Perrone, R. Kao, "A general finite difference method for arbitrary meshes," Computers and Structures, **5**(1), 45–57, 1975, doi:https://doi.org/10.1016/0045-7949(75)90018-8.

[11] Yang-Ming Zhu, "Generalized sampling theorem," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, **39**(8), 587–588, 1992, doi:10.1109/82.168954.

[12] X. Qian, S. W. Lee, Y. Yang, "Heat Transfer Coefficient Estimation and Performance Evaluation of Shell and Tube Heat Exchanger Using Flue Gas," Processes, **9**(6), 2021, doi:10.3390/pr9060939.

[13] W. G. . of the Joint Committee for Guides in Metrology (JCGM/WG 1), Evaluation of measurement data â€" Guide to the expression of uncertainty in measurement, 2008.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 1, 100-106 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Optimization of the Sliding Mode Control (SMC) with the Particle Swarm Optimization (PSO) Algorithm for Photovoltaic Systems Based on MPPT

Ilhem Bouchriha*, Ali Ben Ghanem, Khaled Nouri

The Laboratory of Advanced Systems LSA, Polytechnique School, University of Carthage, Tunis, 3200, Tunisia

ARTICLE INFO

Article history:
Received: 20 September, 2021
Accepted: 25 January, 2022
Online: 21 February, 2022

Keywords:
Boost converter
Maximum Power Point Tracking
Sliding Mode Control
Particle Swarm Optimization

ABSTRACT

Photovoltaic systems are classified as non-linear systems. On the other hand, the characteristics of photovoltaic cells are non-linear. For this reason, researchers use several methods of maximum power monitoring (MPPT) to improve the performance of photovoltaic system. It is therefore necessary to use an adaptation stage between the photovoltaic generator (GPV) and the load to take out at any time the maximum power available of the GPV and to transfer it to the load. This stage acts as an interface between the two elements by ensuring the transfer of the maximum power supplied by the generator using a control system used for this purpose. This work described a study on two Maximum Power Point Tracking (MPPT) techniques, classic sliding mode control (SMC) and particle swarm optimization (PSO) for controlled boost chopper. The decisive simulation results that by applying the PSO and the conventional SMC to a well-built model, the PSO achieves high precision results compared to the sliding mode control. In fact, we have suggested a new MPPT method based on the optimization of the command parameters in sliding mode by the PSO optimization algorithm.

1. Introduction

The integration of renewable energies into power generation systems is becoming important, especially with the increase in the world's population and technological advancement. Solar energy is considered a renewable energy source with great potential and widely distributed geographically. Its drawback remains the high initial cost of implementing the system [1] [2]. To make a photovoltaic system more profitable, an method for finding the maximum power point tracking MPPT is released to command a DC-DC converter placed between the photovoltaic panels and the load, what is called the adaptation step [3] [4]. The objective of this algorithm is to maximize the energy extracted of the solar panels, and thus growth the overall performance of the system.

Besides, and because of the strongly nonlinear electrical characteristics, many researchers have proposed control solutions whatever the climatic conditions [5].

Several researchers have worked on the MPPT method as in [6–8] they have shown that the technique founded on fuzzy logic control (FLC) can quickly go along with the point of M. In [9], we studied the sliding mode control (SMC) and compared it with the P&O algorithm we concluded that this method is more robust and faster. We also found that they worked on the backstepping method but they improved it in [10] we combine it with the artificial neural network (ANN) method [11]. This gave us a good performance in maximum point tracking. Indeed, the first work on the evolution of Sliding Mode Control for DC-DC converters introduced in 1983 [6] and 1985 [9]. This work shows how SMC can be used in different topologies of second order DC-DC converters. The concept of fetching the equivalent command to the duty cycle in order to obtain a sliding mode controller controlled by PWM can be found in [6]. A general SMC synthesis method applicable to most DC-DC converter topologies is proposed by [5] which shows that SMC allows greater robustness to variations in load and circuit parameters.

This method appears more adequate and gives a good transient response. On the other hand, many works propose the control of the duty cycle rather than directly controlling the state of the transistor ensuring the commutation and this without degrading the properties of the controller. In this case, the discontinuous command is replaced by the equivalent command [12]. The latter can be assimilated to the duty cycle since we are working at high frequency. The advantages of this method are that it does not require additional circuits and that the transient response remains good.

This work is an extension of the work initially presented in International Conference on Advances in Science, Engineering and Technology 2021(ICASET) [2]. In this article, the MPPT

*Corresponding Author: Ilhem Bouchriha, ilhem.bouchriha@gmail.com

control of the photovoltaic system is processed under varying climatic irradiation.

The photovoltaic system is composed a photovoltaic panel, boost chopper, MPPT controller and resistive load.

This paper describes a study of the two MPPT techniques, this is firstly the SMC (sliding mode control) method and the second is the PSO (particle swarm optimization) algorithm. Then we optimize the parameters of the SMC using the PSO algorithm. The two proposed techniques are implemented on the Boost converter with MATLAB.

In the following, we will first present a model of a photovoltaic panel. Then modeling a DC-DC step-up converter as a linear switching system. Next, we will describe the two MPPT commands SMC and PSO. And finally our contribution is the combination between the two methods to improve the performance of the sliding mode method. The results of the simulation are presented in the last part.

2. Modeling of the proposed system

A photovoltaic PV system is a set of elements of electricity production. Our system is shown in Figure 1. These components are essentially the PV, a boost chopper and the command MPPT.

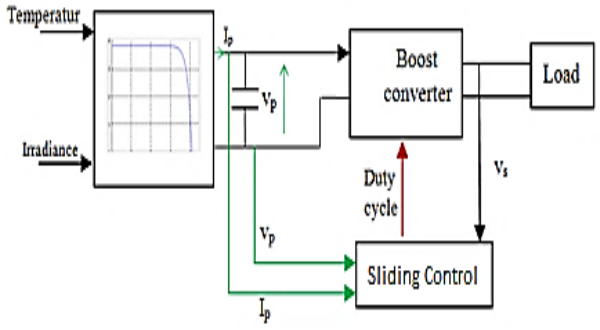

Figure1: PV System

where:
Ip: the output currents of the photovoltaic panel (PV).
Vp: the output voltage of the PV.
Vs: the output Voltage of the DC/DC converter.

3. PV Array

The electric model of a solar cell consists of one or two diodes in parallel and a series resistor Rs supplied by an ideal current source, the first diode D1 describes the semiconductor properties of the cell and the second D2 models the phenomenon. Recombination of charge carriers. We have, this time, two diodes for the phenomena of polarization of the PN junction and two resistors. When the surface of the PV cell is exposed to light, it absorbs light energy and turns directly into electric current. The voltage generated can vary between 0.3V and 0.7V depending on the aging of the cell, the temperature of the material used and the sunshine. The operation of a solar cell can be described by considering the equivalent electrical schema below figure 2 [13–15].

$$I = I_{ph} - I_{s1}\left[e^{\frac{q(V+IRs)}{KTn1}} - 1\right] - I_{s2}\left[e^{\frac{q(V+IRs)}{kTn2}} - 1\right] - \frac{V+IRs}{Rp} \quad (1)$$

Knowing that: I and V are the current and output voltage of the photovoltaic cell. I_{d1} and I_{d2} are the saturation currents of the diodes. d1 and d2 are the diode factors. Iph is the photo generated current, it is often given by.

$$I_{ph}(T) = I_{phmax} * S.10^{-3}[1 + (T - 298.K).5.10^{-4}] \quad (2)$$

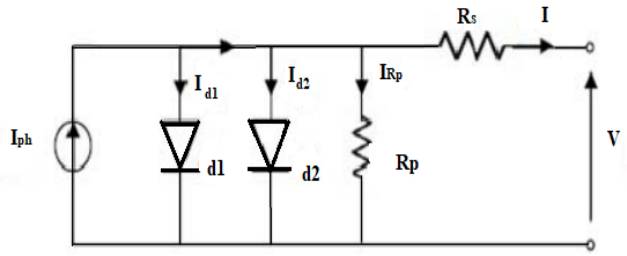

Figure 2: Diagram of the model equivalent to two diodes of a PV cell.

With : S is a Sunstroke. T is the temperature of the cell in Kelvin. From these equations, it is evident that the characteristic I-V strongly depends on insolation and of the temperature. Temperature dependence is further amplified by saturation currents inverse of the diodes which are given by:

$$I_{s1} = K_1 T^3 e^{\frac{Eg}{KT}} \quad (3)$$

$$I_{s2} = K_2 T^{\frac{5}{2}} e^{\frac{Eg}{KT}} \quad (4)$$

where Eg is the gap energy of the semiconductor and k_1, k_2 having the following values.

K_1=1,2A/cm^2.K^3
K_2=2,9.105A/cm^2.K$^{5/2}$

To confront our model a little more with reality, it is necessary to study how these parameters will influence the characteristics of the cell. The figure 3 and 4 presents the curves of I and P in function of V, for different levels of sunshine at constant temperature, and for different temperatures with constant sunshine.

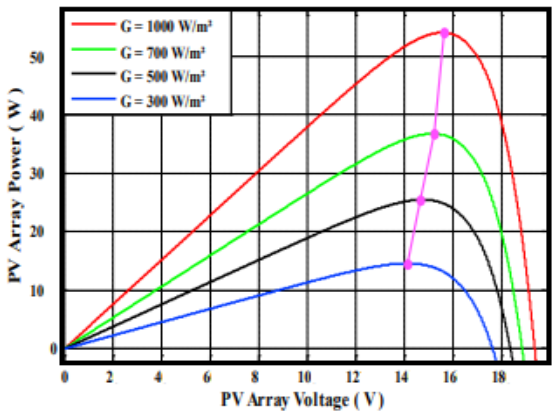

Figure 3: Power curves of PV module at various temperatures simulated with the MATLAB model Boost converter

4. Boost converter

The DC / DC converter makes it possible to connect the PV field to a DC bus, to which the various energy machines of the system are connected to the load. A chopper can be realized using an electronic switch which can be controlled on opening and

closing, such as GTO thyristors or bipolar or insulated gate field effect transistors operating in switching mode. The principle of a chopper consists in establishing and then periodically interrupting the source-load connection using the electronic switch. This must be able to be closed or opened at will in order to have an adjustable continuous output voltage. Our Boost chopper is shown in the figure 5.

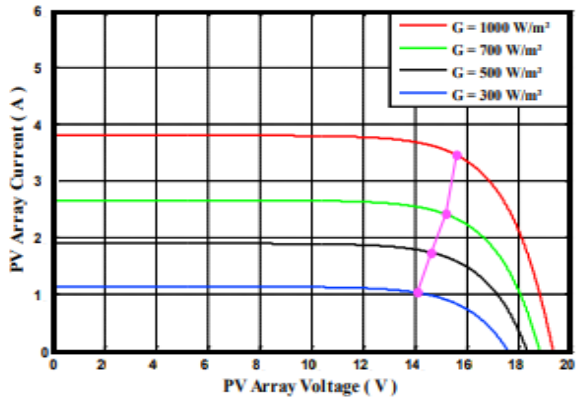

Figure 4: Curve of the current as a function of the voltage of a PV panel at various solar radiations simulation with MATLABC

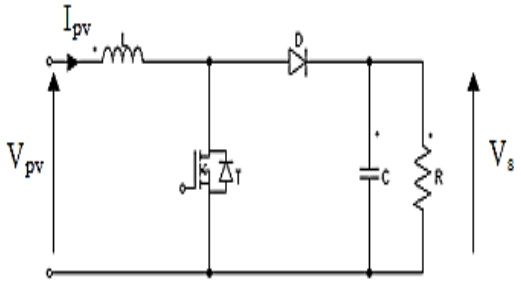

Figure 5: BOOST converter

Table 1 shows the nominal parameters of the circuit. Throughout the duration dTs of the switching period Ts, the switch is closed and the diode is off. The current increases until it reaches the peak value in the inductor. When the thyristor is blocked, the diode will be conductive, the current will decrease until it reaches its minimum value and the output voltage equals the input voltage [16].

Table 1: The Value of the Boost Chopper Coefficient

Symbol	Description	Value
V_{pv}	Input voltage	220v
F	Switching Frequency	50Hz
C	Capacitance of capacitor	2000.10-6
L	Inductance	0.3µH
R	Load resistance	50Ω

The energy accumulator in the coil is dissipate into the capacitor and the load. In the interval $t_0 \leq t \leq t_0 + dT_s$, the switch T is shut and the diode is off. We obtain :

$$L\frac{di_L}{dt} = V_e \tag{5}$$

And

$$C\frac{dV_s}{dt} = -\frac{V_s}{R} \tag{6}$$

The first configuration is represented by the linear model:

$$\begin{bmatrix} \frac{di_L}{dt} \\ \frac{dV_s}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i_L \\ V_s \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_e \tag{7}$$

In the interval $t_0 + \mathrm{dT_s} \leq \mathrm{t} \leq \mathrm{t_0} + \mathrm{T_s}$, the diode D is passant and T is open . We will then have:

$$L\frac{di_L}{dt} = V_e - V_c \tag{8}$$

And

$$C\frac{dV_s}{dt} = i_L - \frac{V_s}{R} \tag{9}$$

The second configuration is described by the following linear model:

$$\begin{bmatrix} \frac{di_L}{dt} \\ \frac{dV_s}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i_L \\ V_s \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_e \tag{10}$$

The general equation which describes the functioning of the Boost converter is:

$$\begin{bmatrix} \frac{di_L}{dt} \\ \frac{dV_s}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{(1-d)}{L} \\ \frac{(1-d)}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i_L \\ V_e \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_e \tag{11}$$

Based on equations (11) and (12), the equivalent model representation of the state space is described by the following model (12).

$$\begin{cases} \dot{x_1} = -\frac{1-d}{L} x_2 + \frac{1}{L} V_p \\ \dot{x_2} = \frac{1-d}{C} x_1 - \frac{1}{RC} x_2 \end{cases} \tag{12}$$

where : (x_1, x_2)= (iL,vs)

And $y=\begin{bmatrix} 0 \\ 1 \end{bmatrix}$(iL,vs)

5. MPPT control strategy

PV systems are classified as nonlinear systems. Indeed, researchers have proven that simple commands and PID correctors are effective for linear systems. But for nonlinear systems sometimes the PID correctors are not sufficient. This is why we are always looking for other methods to have good precision and the system will be more stable and robust [17], [18].

5.1 Sliding mode control (SMC)

The variable structure controls by the sliding mode, which appeared in the 1960s, due to the research of the mathematician

A.F. Philipov, is a more nonlinear method fonder on the use of discontinuous expressions. After the work developed by the team of Professor Emelyanov in the Soviet Union and due to problems with chatter and implementation, the variable structure control is obtained in the late 1970s. It has appeared well in the fields of electronics and computing. Indeed, this command is based on a high frequency switching for a better sliding speed [19], [20,21].

The principle of operation of the SMC method is to bring the system of states to arrive at the sliding surface and switch it by a suitable logic switching round it until the point of equilibrium, hence the sliding phenomenon. [23] The form of a variable structure command can be given by:

$$\dot{S} < 0 \; if \; S > \mathrm{x_{2d}} \qquad (13)$$
$$\dot{S} > 0 \; if \; S > \mathrm{x_{2d}}$$

where: S is the output voltage.

Equation 14 defines the path of the first order system. And so, the speed of convergence will be controlled.

$$\dot{S} = -\lambda(S - \mathrm{x_{2d}}) \qquad (14)$$

When we do the combination of equation 14 with the two equations 12 and 10, we get the duty cycle equation d which allows us to tune the output of our system. d will be the duty cycle of our boost chopper.

$$d(t) = 1 - \frac{v_{PV} + \sqrt{{v_{pv}}^2 + aS_2(S_2 - x2d)}}{2S_2} \qquad (15)$$

where

$$a = -4\lambda\left(LC\lambda - \frac{L}{R}\right)$$

Our command described by equation (15), is show in the Figure 6.

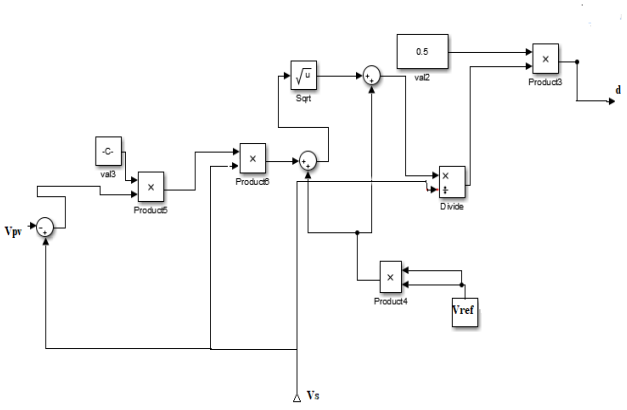

Figure 6: Sliding Mode Control Matlab/SIMULINK

The SMC is characterized by its simple design and robustness regardless of the types of disturbance. The best features of this approach are: [22,23]

- The classification of the switching principle introduced us to a dynamic comportment of the system.
- In closed buckle, the response is unfeeling to the types of perturbations and to changes in parameters.
- Choice of a fairly large area according to the case. It's faster while still having the opportunity to grow in an unstable state.

5.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization is a biologically inspired method for solving optimization problems. Like artificial neural networks, genetic algorithms, or ant colony algorithms, the PSO is a bio-inspired algorithm.

The objective function of the PSO algorithm has been defined as the power delivered by our PV system.

The equations (15) and (16) represent the movement of particles through each iteration of the program and its objective functions (fitness) are calculated to be capable to measure the best position of all the Pg.

$$v_i(k+1) = wv_i(k) + c_1r_1\big(P_{best} - x_i(k)\big) + c_2r_2\big(g_{best} - x_i(k)\big) \qquad (16)$$

$$x_i(k+1) = x_i(k) + \; v_i(k+1) \qquad (17)$$

where k is the number of iterations, i The number of the particle, $v_i(k)$ is the speed of the particle at the precedent step, $v_i(k+1)$ is the new value the studied velocity of the particle, c1 and c2 are positive parameters specific to cognitive actions and to the algorithm of individual social particles, r1 and r2 are taken randomly between [0,1], P_{best} is the better position acquired by the measured element in the objective function, g_{best} is the better position obtained amongst all the particles evaluated in the objective function, w presents to an inertial mass which directly affects the current velocity of the particle with respect to the rate which precedes [24,25]

The updating of the Pb and Pg is carried out at each iteration as described by the organization chart of Fig.4. The steps are repeated as long as the stopping criterion is not reached.

- Stage I: Initialize coefficients c1 and c2, the inertia coefficient (w).
- Stage II: Random construction of the initial particle group and calculation of the fitness of each element (P_{best}): better position of the number i in the current group; (P_{gbest}): Better positioning in all group (best of the best).
- Stage III: Based on the formula, we again calculate the velocity and location of each particle.
- Stage IV: In the current state, we calculate the best fitness of the group and compare it with the last one to have the better of all the groups (P_{gbest}).
- Stage V: Increment K by 1 so k = k+ 1.
- Stage VI: If the stop condition is obtained, we go to the 7th step. Other, go back to stage III.
- Stage VII: the optimal solution will be saved in (P_{gbest})

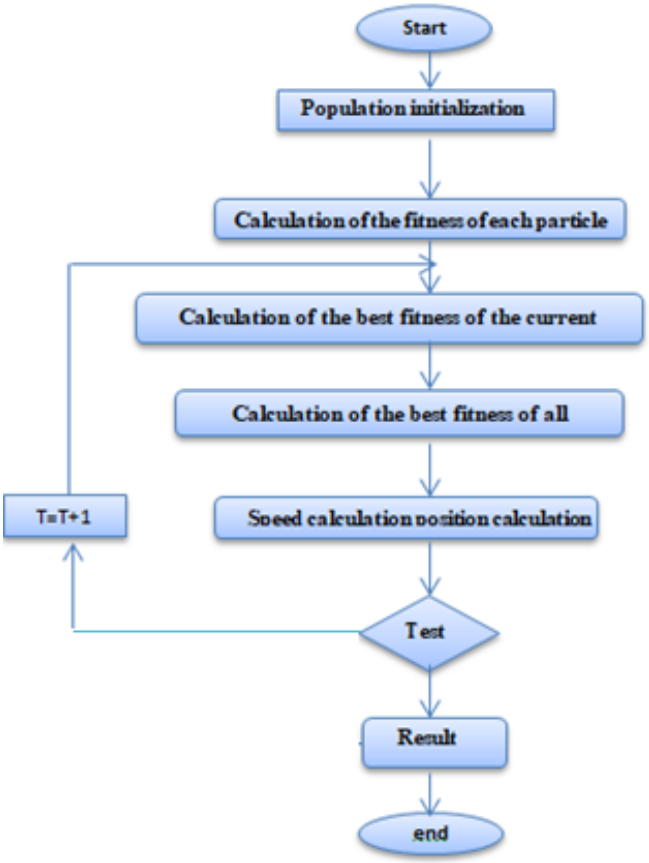

Figure 7: Flowchart for the PSO algorithm.

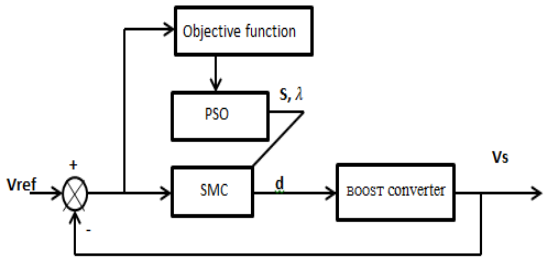

Figure 8: Block diagram of the PSO-SMC method

5.3 The Proposed PSO-SMC Controller

This part presents the sliding mode method to optimize with the evolutionary algorithm PSO as seen previously the SMC method can be better improve its performance, the parameters (S, λ) can be obtained using the PSO algorithm. After initialization the values of the particles. Our PSO algorithm tests the search space based on equations (16) and (17). Each particle keeps its best position (P_{gbest}) which will be compared each time with the best position of the other particles. And finally to have a minimum value of objective function, the optimal parameters are $S_{optimum}$ and $\lambda_{optimum}$ [26].

6. Simulation Results and Discussion

To know the stability and follow the dynamic response of our two MPPT methods: SMC and PSO. We implemented the model of our PV system proposed in the MATLAB / Simulink. We vary the irradiations and with a constant temperature equal to 25°C [27].

Table 2: The parameters of the PSO method

Parameters	Values
C1	2
C2	1.8
W	0.8
Kmax	10
Number of particles	3

Figure 9 and Figure 10 shows the simulation results of the two tested methods. Figure.10 present the output voltage value and the Figure.9 present the power of boost chopper.

The simulation showed us that the SMC method gave us an acceptable result from a stability point of view. But if we compare it with the PSO method we find it more efficient and faster. And for that we are going to combine the two methods together and we will find the result in the figure 11. We notice that the PSO-SMC method is correct our new method and faster and stable and more exact folds as shown in the table.

Table 3: Comparison of MPPT techniques

Algorithm	PSO	SMC	PSO-SMC
Complexity	Low	Low	Low
Setting Time ts	10.034	10.044	10.034
Peak time tp	10.011	10.005	10.011
D	2.726	7	5

7. Conclusion

In this article we worked on a photovoltaic system controlled by two techniques PSO and SMC. After comparative with the two methods, the PSO is very robust in relation to SMC.

So, we have optimized the parameters of the SMC by the PSO algorithm to obtain a new SMC-PSO method. We can conclude that the SMC method when combined with PSO has become faster and more stable.

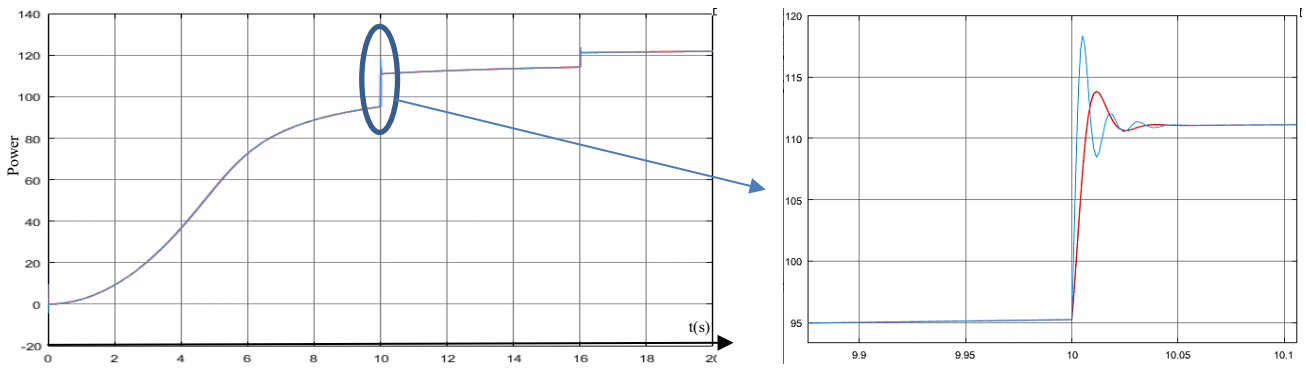

Figure 9: Optimum power of the PV system for varying conditions of irradiation

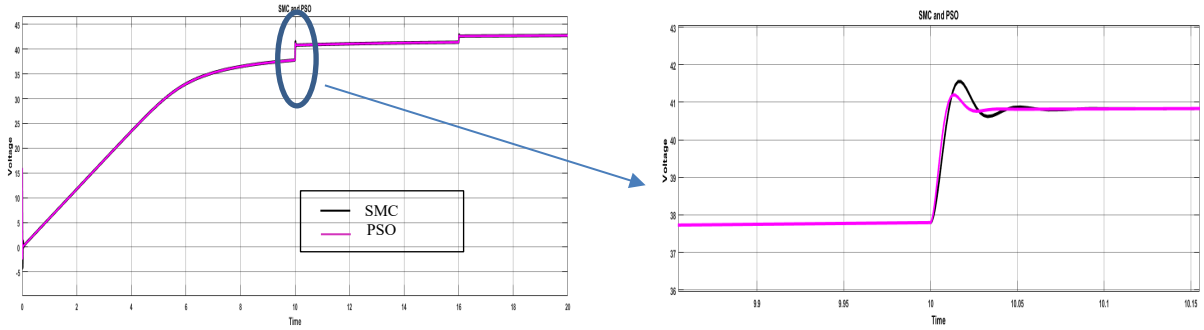

Figure 10: Optimum voltage of the PV system for varying conditions of irradiation.

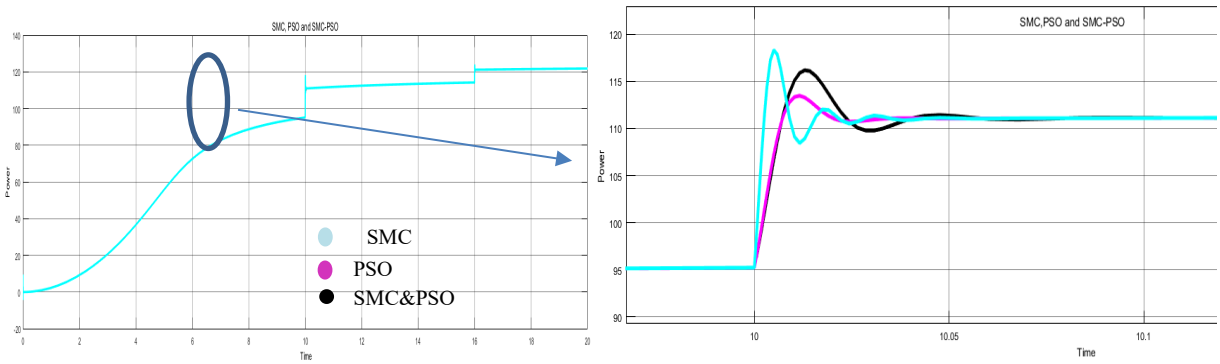

Figure 11: Optimum power of the PV system for varying conditions of irradiation

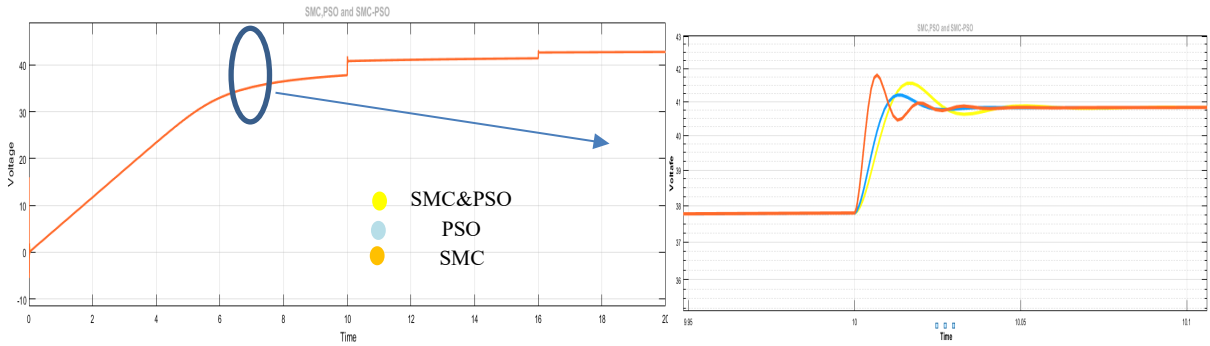

Figure 12: Optimum voltage of the PV system for varying conditions of irradiation

References

[1] J. Sedo, S. Kascak, "Control of single-phase grid connected inverter system," ELEKTRO 2016 - 11th International Conference, Proceedings, 207–212, 2016, doi:10.1109/ELEKTRO.2016.7512066.

[2] I. Bouchriha, A. ben Ghanem, K. Nouri, "A Comparison of Sliding Mode Control and Particle Swarm Optimization for Photovoltaic System under Partially Shaded Condition," Undefined, 310–314, 2020, doi:10.1109/IC_ASET49463.2020.9318254.

[3] E. Fossas, A. Ras, "Second order sliding mode control of a buck converter," Proceedings of the IEEE Conference on Decision and Control, **1**, 346–347, 2002, doi:10.1109/CDC.2002.1184516.

[4] A. Al-Gizi, S. Al-Chlaihawi, A. Craciunescu, "Efficiency of Photovoltaic Maximum Power Point Tracking Controller Based on a Fuzzy Logic," Undefined, **2**(3), 1245–1251, 2017, doi:10.25046/AJ0203157.

[5] E.M. Zakaria, S.H. Arafa, M.N. Fahmy Nashed, S.G. Ramadan, "Fuzzy Logic Control Management with Stand Alone Photovoltaic – Fuel Cell System," Undefined, **5**(1), 424–430, 2020, doi:10.25046/AJ050154.

[6] V.I. Utkin, "Sliding Modes in Control and Optimization," Sliding Modes in Control and Optimization, 1992, doi:10.1007/978-3-642-84379-2.

[7] S.Z. Hassan, H. Li, T. Kamal, M. Nadarajah, F. Mehmood, "Fuzzy embedded MPPT modeling and control of PV system in a hybrid power system," in 2016 International Conference on Emerging Technologies (ICET), IEEE: 1–6, 2016, doi:10.1109/ICET.2016.7813236.

[8] S.Z. Hassan, H. Li, T. Kamal, U. Arifoğlu, S. Mumtaz, L. Khan, "Neuro-Fuzzy Wavelet Based Adaptive MPPT Algorithm for Photovoltaic Systems," Energies, **10**(3), 394, 2017, doi:10.3390/en10030394.

[9] Y. Jung, J. So, G. Yu, J. Choi, "Improved perturbation and observation method (IP&O) of MPPT control for photovoltaic power systems," Conference Record of the IEEE Photovoltaic Specialists Conference, 1788–1791, 2005, doi:10.1109/PVSC.2005.1488498.

[10] K. Sundareswaran, V. Vignesh kumar, S. Palani, "Application of a combined particle swarm optimization and perturb and observe method for MPPT in PV systems under partial shading conditions," Renewable Energy, **75**, 308–317, 2015, doi:10.1016/J.RENENE.2014.09.044.

[11] K. Chennoufi, M. Ferfra, "Fast and Efficient Maximum Power Point Tracking Controller for Photovoltaic Modules," Undefined, **5**(6), 606–612, 2020, doi:10.25046/AJ050674.

[12] V. Repecho, D. Biel, J.M. Olm, E. Fossas, "Sliding Mode Control of Power Converters with Switching Frequency Regulation," Studies in Systems, Decision and Control, **115**, 385–409, 2018, doi:10.1007/978-3-319-62896-7_16.

[13] S.Z. Hassan, "Hermite-Wavelet Based Intelligent MPPT Adaptive Control Paradigm for PV in a Standalone/Grid Connected Hybrid Power System," MS Thesis, COMSATS Univeristy, Pakistan, 2014.

[14] S.Z. Hassan, S. Mumtaz, T. Kamal, L. Khan, "Performance of grid-integrated photovoltaic/fuel cell/ electrolyzer/battery hybrid power system," in 2015 Power Generation Systems and Renewable Energy Technologies, PGSRET 2015, 1–8, 2015, doi:10.1109/PGSRET.2015.7312249.

[15] T. Kamal, S.Z. Hassan, M.J. Espinosa-Trujillo, H. Li, M. Flota, "An optimal power sharing and power control strategy of photovoltaic/fuel cell/ultra-capacitor hybrid power system," Journal of Renewable and Sustainable Energy, **8**(3), 035301, 2016, doi:10.1063/1.4948926.

[16] A. Jouila, K. Nouri, "Applications to function approximation and system control based on wavelet networks," 2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, STA 2017 - Proceedings, **2018-January**, 426–431, 2018, doi:10.1109/STA.2017.8314864.

[17] A. Jouila, K. Nouri, "Stabilization of Neuro-Control Structure using Lyapunov Functional Based Approach," 2019 International Conference

on Signal, Control and Communication, SCC 2019, 81–86, 2019, doi:10.1109/SCC47175.2019.9116173.

[18] S. Jebri, K. Nouri, L. Bouslimi, "Speed tracking control performance with frequency and amplitude dependence," Proceedings of International Conference on Advanced Systems and Electric Technologies, IC_ASET 2017, 454–459, 2017, doi:10.1109/ASET.2017.7983736.

[19] A. Azzabi, K. Nouri, "Design of a robust tracking controller for a nonholonomic mobile robot based on sliding mode with adaptive gain:," Https://Doi.Org/10.1177/1729881420987082, **18**(1), 2021, doi:10.1177/1729881420987082.

[20] M.G. Villalva, J.R. Gazoli, E. Ruppert Filho, "Modeling and circuit-based simulation of photovoltaic arrays," 2009 Brazilian Power Electronics Conference, COBEP2009, 1244–1254, 2009, doi:10.1109/COBEP.2009.5347680.

[21] A. Azzabi, K. Nouri, "Design of a robust tracking controller for a nonholonomic mobile robot based on sliding mode with adaptive gain," International Journal of Advanced Robotic Systems, **18**(1), 2021, doi:10.1177/1729881420987082.

[22] J. Sedo, S. Kascak, "Control of Single-Phase Grid Connected inverter system," in: J. Sedo and S. Kascak, ", eds., in 11th international ELEKTRO Conference, Strbske Pleso, Slovakia, 2016, doi:10.1109/ELEKTRO.2016.7512066.

[23] A.H. Alqahtani, V.I. Utkin, "Self-optimization of photovoltaic system power generation based on sliding mode control," IECON Proceedings (Industrial Electronics Conference), 3468–3474, 2012, doi:10.1109/IECON.2012.6389342.

[24] L. Hang, S. Liu, Z. Lu, M. Castilla, "A novel robust sliding mode controller for Half-bridge converter," 2010 International Power Electronics Conference - ECCE Asia -, IPEC 2010, 1618–1621, 2010, doi:10.1109/IPEC.2010.5543578.

[25] H. Renaudineau, F. Donatantonio, J. Fontchastagner, G. Petrone, G. Spagnuolo, J.P. Martin, S. Pierfederici, "A PSO-based global MPPT technique for distributed PV power generation," IEEE Transactions on Industrial Electronics, **62**(2), 1047–1058, 2015, doi:10.1109/TIE.2014.2336600.

[26] M. Arsalan, R. Iftikhar, I. Ahmad, A. Hasan, K. Sabahat, A. Javeria, "MPPT for photovoltaic system using nonlinear backstepping controller with integral action," Solar Energy, **170**, 192–200, 2018, doi:10.1016/J.SOLENER.2018.04.061.

[27] K. Nouri, R. Dhaouadi, N.B. Braiek, "A New Adaptive Control Scheme Using Dynamic Neural Networks," Undefined, **20**(1), 171–177, 2008, doi:10.20965/JRM.2008.P0171.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 1, 107-124 (2022)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Efficient Publicly Verifiable Proofs of Data Replication and Retrievability Applicable for Cloud Storage

Clémentine Gritti*, Hao Li

Computer Science and Software Engineering Department, University of Canterbury, Christchurch, 8011 New Zealand

ARTICLE INFO

Article history:
Received: 13 December, 2021
Accepted: 09 February, 2022
Online: 28 February, 2022

Keywords:
Proofs of Retrievability and Reliability
Verifiable Delay Functions
Cloud Storage Services

ABSTRACT

Using Proofs of Retrievability (PORs), a file owner is able to check that a cloud server correctly stores her files. Using Proofs of Retrievability and Reliability (PORRs), she can even verify at the same time that the cloud server correctly stores both her original files and their replicas. In 2020, a new PORR combined with Verifiable Delay Functions (VDFs) was presented by Gritti. VDFs are special functions whose evaluation is slow while verification is fast. Therefore, those functions help guarantee that the original files and their replicas are stored at rest. Moreover, an important feature of the 2020 PORR solution is that anyone can verify the cloud provider's behaviour, not only the file owner. This paper extends Gritti's version. In particular, a realistic cloud framework is defined in order to implement and evaluate accurately. Results show that this PORR solution is well suitable for services provided for cloud storage.

1 Introduction

Cloud data storage has exploded over the last decade, for both business and personal purposes. The latter have offered the opportunity to delegate the storage of individuals' files, through various services that have been developed and diversified with competitive fees for data owners (e.g. copies of files stored in different storage locations). However, due to its complex structure, many unfortunate events have happened over the last few years. In 2015, lightning strikes engendered data loss on Google centers[1]. The startup *Front Edge CNC'* used to keep its online production data in Tencent Cloud Storage but realised in 2018 that it was completely lost[2]. More recently, according to McAfee, 99% of misconfiguration incidents occurring in a public cloud environment are not detected, thus exposing enterprises and organisations to a huge risk of undetected data breaches[3]. Therefore, it has become urgent to carefully design cloud storage solutions to overcome all possible unfortunate events such as the aforementioned ones.

An ideal solution will enable a simple deal between a client, who owns some files, and a cloud provider (also called cloud server), who offers safe and attractive cloud storage services. For instance, the cloud provider may propose to store copies of a file in addition to the file itself, and spread them across multiple servers, avoiding single point of loss. The cloud provider may also offer to decrease fees per copy, to encourage the client to adopt the replication process as much as possible. More precisely, both the client and cloud provider are financially incentivized: the client wants to save fees as much as possible while uploading files as many as possible, while the cloud provider wants to earn as much as possible while saving storage as much as possible. Those assumptions clearly motivate their behaviours. The rational cloud provider may claim to store the client's files while actually not, and thus offering the free storage space to other clients. The malicious client may claim to upload copies of the files, with lower fees, but rather uploading different files where fees would have higher if behaving honestly.

This work is an extension of [1], in which a new solution was proposed for secure and efficient cloud storage for a client who owns some files stored on a cloud provider. Security is enhanced by allowing detection of a misbehaving cloud provider that does not store the client's files correctly and prevention against a malicious client who tries to save fees. Informally, the cloud provider is asked to create copies of the client's files stored across various locations and can be challenged to prove the client that all files and their copies are entirely stored. The client uploads encrypted files

*Corresponding Author: Clémentine Gritti, CSSE Department, University of Canterbury, clementine.gritti@canterbury.ac.nz
[1] https://www.bbc.com/news/technology-33989384
[2] https://medium.com/genaro-network/tencent-was-claimed-ten-million-for-data-loss-due-to-cloud-hard-drive-glitch-344a26449fe2
[3] https://www.computerweekly.com/news/252471175/Enterprises-exposed-to-data-loss-by-cloud-configuration-errors

to the provider but does not have the responsibility of creating the copies, avoiding the client to upload fake copies (i.e. files that are not at all linked to the original ones). Efficiency is conserved even with improved security. In particular, experimental results presented for the first time in this paper show that the mechanisms used to preclude malicious behaviours incur low costs on the computational and communication processes.

1.1 Problem Statement

Let a client subscribe to data storage services offered by a cloud provider. Proof of Data Possession (PDP) and Proofs of Retrievability (POR) are cryptographic protocols proposed to enable the client to verify that the cloud provider actually proceeds as agreed, that is correctly stores the client's data.

The latter protocol, POR [2], guarantees the client that her data is available in its entirety. The former protocol, PDP [3, 4], allows the client to verify that the stored data has not undergone any modifications. Recent PDP works [5, 6, 7] offer a replication feature: the cloud provider creates copies of the data and the client can check in a single instance that both the original and copied data are all correctly stored.

However, the aforementioned solutions do not reflect the actual framework. Limited bandwidth is offered for free and costs grow quickly with the increasing bandwidth. In most of existing works, the replication and uploading processes rely on the client, imposing huge burden and fees to the latter. In addition, fees for storage of extra replicas of a given file are less consequent than fees for storage of different files. Hence, a *malicious* client may attempt to upload different files while claiming that they are replicas of the same file. The cloud provider cannot observe such a trick since data is assumed to encrypted before being uploaded, and thus are not readable.

Limitations also raise on the cloud provider's side. The latter may appear to be financially motivated and consequently act maliciously. For instance, it attempts to not store all file replicas and offers the resulting unoccupied storage space to other potential customers. Then, such a *rational* cloud provider creates the missing replicas on the fly when the client who owns the original file asks for storage verification. In addition, the client may put tacit trust on its cloud provider and neglect to carefully read the service level agreements [8]. Likely, the client does not take the time and effort to verify how the cloud provider deals in storing her data.

The design of an appropriate solution for secure data storage in the cloud is clearly encouraged by the aforementioned financial concerns. Clients are motivated by reducing the costs due to bandwidth and storage while the cloud provider is stimulated by partitioning data to make profit. Those concerns have been carefully studied and overcome in [9]. The authors presented an extension of POR, called Proof of Retrievability and Reliability (PORR), to encompass the correct storage verification of both file and its replicas at once. However, the main obstacle of the PORR instantiating in [9] is the private feature of verification: only the client is able to check that the cloud provider correctly stores her file and its replicas.

In [1], we proposed a new PORR protocol overcoming all the aforementioned challenges at once. Along with getting over malicious clients and rational providers, we overcame the likely laziness: the client can delegate the task of checking that the cloud provider has been acting honestly on her data.

1.2 Idea

Our solution is designed to enable a client to upload her file and the cloud provider to generate the replicas of this file. By doing so, we prevent malicious clients to upload different files and claim that there are all replicas of the same file.

Moreover, we offer the client the guarantee that the cloud provider correctly stores both the original file and its replicas at rest. By using slow functions, we prevent rational cloud providers to compute replicas of an original file on the fly when challenged to prove correct storage. Indeed, evaluating a slow function is noticeably slow while verifying its unique evaluation output is fast and easy. Therefore, if the cloud provider tries to generate a replica on demand, it will take a noticeable time to evaluate the slow function attached to the replica and output the unique solution, rather than just storing the replica along with the output from the evaluated slow function.

The cloud provider is challenged by a *verifier*, that can be anyone on behalf of the client. The verification allows a client to ensure that the cloud provider stores the original file and all replicas in their entirety.

Our PORR solution with public verification is a combination of the publicly verifiable POR protocol with RSA signatures from Shacham and Waters [10, 11] and of the slow exponentiation-based Verifiable Delay Functions (VDFs) in finite groups from [12].

We suggest an extension of the POR scheme built by Shacham and Waters [10, 11], where replicas of the challenged file are contained in the verification mechanism. Public verification allows anyone, on behalf of the client, to request the cloud provider a proof of correct storage. Such a feature overcomes a possible lack of data integrity awareness from the client. In order to stop the cloud provider to generate the file replicas on demand when being challenged by the verifier, we incorporate slow functions from [12], namely VDFs. Of course, storing VDF outputs must result into storage costs that are at least as high as storing replicas as required. In fact, rational cloud providers are expected to commit minimal computation resources when generating a correct response. Therefore, the cost of replying to a challenge on the fly should be more than the cost of storing the replicas correctly.

To our knowledge in [1], the author presented the first PORR protocol that both delegates the construction of the replicas to the cloud provider and that permits anyone to verify that the latter correctly stores the files and their replicas.

We give an overview of the PORR protocol in Figure 1. A client encodes and processes a file M as M^*, and outsources it to the cloud provider. The latter commits to store M^* entirely across a set of r storage nodes. This means that M^* is updated to contain the original copy of M and r replicas. On inputs the public parameters, the verifier (possibly the client as in the figure) can efficiently launch challenge requests and verify the responses from the cloud provider. The cloud provider must store both the original file and its replicas at rest, at r storage nodes.

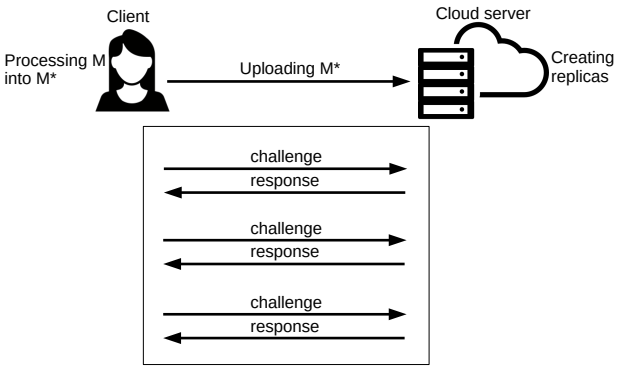

Figure 1: **PORR protocol:** First the client prepares her file M by encrypting and encoding it, resulting into M^*. Second, she uploads M^* to the cloud server along with some additional elements, such as a tag T. The cloud server generates the replicas based on the additional information. Later, and multiple times, the client and cloud server enter into a challenge-response protocol to allow the former to check whether the latter correctly stores her file and its replicas.

1.3 Related Work

We first present existing works on Provable Data Possession and Proof of Retrievability. We then move on Proof of Reliability, where several variants have been proposed based on different mechanisms, namely data replication, erasure codes, proof of location and proof of space. Finally, we review slow functions that allow to generate puzzles, and in particular Verifiable Delay Functions.

1.3.1 Provable Data Possession and Proof of Retrievability.

Provable Data Possession: The authors in [3] introduce the concept of Provable Data Possession (PDP), which enables a client to privately verify the integrity of her files stored at an untrusted cloud provider without the need to retrieve the entire files. Thereafter, the authors in [13] improve the efficiency of the previous PDP scheme [3] by using symmetric encryption. Subsequently, PDP constructions for static data have been proposed in the literature [14]–[17]. PDP schemes with additional properties, such as data dynamicity, have been suggested in [18]–[23]. Several works have been published recently [24]-[30], proposing PDP schemes preserving data privacy and publicly verifying data integrity.

Proof of Retrievability: In [2], the author defined Proof of Retrievability (POR), that is a similar concept to PDP. In POR, the client can correct existing erros and recover the whole files, in addition to check that the cloud provider stores her files in their entirety. In [10], the author proposed two POR schemes: a first one with public verification built on BLS signatures, and a second one with private verification based on pseudo-random functions. In [31], the author improved Shacham-Waters POR schemes by achieving only a constant number of communication bits per verification rather than linear in the number of sectors. Subsequent works on distributed systems follow in [32]–[34], as well as POR protocols with data dynamicity in [35, 36].

1.3.2 Proof of Reliability.

Based on Data Replication: The authors in [5] extend the scheme from [3] by enabling the file owner to verify that at least k replicas are stored by the cloud provider. In [37, 38], the author propose a mechanism with a tunable masking feature in order to let the cloud provider be responsible of the repair operations and the client manage the process of repair. In [39], cross-client deduplication is enabled at the file level by extending [38]. In [6, 7], clients are able to modify and add some pieces of a given file, while to check the correctness of replications of this file. In [40], the replication process is made transparent in a distributed system for cloud storage by extending [18]. The aforementioned solutions, a similar framework is encountered: replicas are prepared by the client, and further sent to the cloud provider with the original files.

More recently, the Proof of Retrievability and Reliability (PORR) protocol is defined for the first time in [9] to illustrate the case where the cloud provider prepares the replicas rather than the client. Tunable puzzles are used for replicating the file, guaranteeing the the replicas are stored at rest. Nevertheless, only a private verifying process is available.

Based on Erasure Codes: In [33], the author provides a high availability and integrity layer for cloud storage. By using erasure codes, data retrievability and reliability are guaranteed among distributed storage servers, and clients can detect and repair file corruption. The authors in [41] achieve a more efficient mechanism for repairs by extending HAIL [33] and moving bulk computations to the cloud provider. The authors in [42] present a scheme for remote data checking in network coding-based distributed storage systems. The scheme minimizes the communication overhead of the repair component compared to the erasure code-based approaches. The repair mechanism in [42] is improved in [22], such that the computation cost for the client is reduced. Following the idea of adding a third party auditor [22], in [43] the author offer a network coding-based scheme such that the repair mechanism is executed between the auditor and the cloud provider, and the client is let apart.

A proof of fault tolerance, as a protocol based on erasure codes, is proposed in [44]. The scheme uses technical characteristics of rotational hard drives to build a time-based challenge, such that the client can check that her encoded files are stored at multiple storage nodes within the same cloud provider. Thereafter, the authors in [45] construct POROS for erasure code-based cloud storage systems by combining POR with time-constrained operations. In order to set a time threshold for the response generation and thus to force the cloud provider to store replicas at rest, rotational hard drive-based ideas are leveraged similar to the ones in [44].

More recently, in [46] the author uses erasure codes to guarantee data reliability, and ensures that the burden of replica generation as well as the repair operations are shifted to the cloud provider. Nevertheless, contrary to [44, 45], the solution in [46] does not rely on the technicity of rotational hard drive technology. Indeed, such method is inevitably dependent on the parameters of the underlying erasure code system. However, in [46], only the client can check the behaviour of the cloud provider in storing her files and replicas.

Proof of Location: Proofs of Location (PoL) [47, 48] have been introduced to prove the geographic position of data, such that whether certain servers in a given country store such data. A PoL scheme is presented in [48] as a combination of geo-location mechanisms and the Shacham-Waters POR schemes [10]. The PoL model looks similar to the PORR one, such that files are uploaded only once by the client to the cloud provider. Then, the cloud provider generates the file tags, prepares the replicas and scatters the latter across servers that are geo-located at different places. Individual POR instances are invoked by the client with all servers to verify whether the latter correctly store the files. On the contrary, PORR schemes must enable clients to launch a single instance to check that all file replicas are correctly stored instead of one instance per replica.

Proof of Space: Proofs of Space (PoS) [49] allow the prover to reply correctly as long as a given amount of space or time per execution is invested. In our case, we require that the prover invests a minimum amount of space. In addition, we need to get a single instance to check that all file replicas are correctly stored, that is not made possible with PoS. Hence, the notion of PoS does not suit our requirements.

1.3.3 Slow Functions.

Time-lock puzzles [50], benchmarking [51], timed commitments [52] and client puzzles [53, 54] rely on the sequential nature of large exponentiation in a finite group.

In [53], the author design a slow function by extracting modular square roots. No algorithm is known for computing modular exponentiation which is sub-linear in the bit-length of the exponent. Nevertheless, the puzzle difficulty is limited to $O(\log p)$, meaning that a very large prime p (being the order of the group) must be chosen to produce a difficult puzzle. In [54], the author consider the sequential nature of Dwork-Naor solution to suggest chaining a series of such puzzles together (e.g. for lotteries). However, this construction does not provide asymptotically efficient verification, hence it can rather be seen as a pseudo-Verifiable Delay Function (VDF).

Time-lock puzzles [50] involve computing an inherently sequential function, by repeating squaring in a RSA group. Used in the context of POR with data reliability guarantees, the cloud provider requires an apparent amount of time to solve those puzzles while the latter are efficiently solved by the client using a trapdoor. Moreover, storing the solution of a time-lock puzzle incurs extra storage costs that are at least as large as the required storage for replicas. The difficulty of a RSA-based puzzle can be easily adjusted by the client (who creates the puzzle) to cater for variable length and different cost metrics. However, such puzzles are not guaranteed to be publicly verifiable: the client (or a dedicated verifier) uses a secret element to prepare each puzzle and to verify the solution.

A given number of sequential steps are required for VDFs, resulting into a unique output verified in an efficient and public way. The construction of a VDF with a trusted setup was proposed [55]. The authors in [12] observe that the trusted setup can be discarded by choosing a sufficiently large random number N, such that $N = pq$ with high probability, for p, q two large prime factors. Nevertheless, the adversary would benefit for parallelizing the arithmetic due to the large size of N, and the running time of the verifier would then increase. The authors in [12] rather suggest to build a VDF from exponentiation, based on the assumption that the adversary cannot run a long pre-computation from the publication of the public parameters to the evaluation of VDF. Hence, this solution achieves proofs that are shorter than the proofs in [55] at a similar level of security.

1.3.4 Work Comparison.

In Table 1, we compare our work with main other ones, based on the specific features that enable secure cloud storage against malicious servers and clients. We omit the Proof-of-Space (PoS) mechanism for Proof of Reliability since it is too far-off our work.

Table 1: Comparing existing protocols similar to PORR

Type	Work	Proof	Replicas? Where?	Public? Private?
Original PDP	[3]	PDP	no replicas	private
	[13]	PDP	no replicas	private
Original POR	[2]	POR	no replicas	private
	[10]	POR	no replicas	private (pseudo)
	[10]	POR	no replicas	public (BLS)
Proof of Reliability (data replication)	[5]	POR	replicas at client's side	private
	[39]	POR	replicas at client's side	private
	[9]	POR	replicas at server's side	private
	ours	POR	replicas at server's side	public
Proof of Reliability (erasure codes)	[33]	other	replicas at client's side	private
	[46]	POR	replicas at server's side	private
Proof of Reliability (proof of location)	[48]	POR	replicas at server's side but individual POR instances	private

Most of the compared works are based on PDP and POR proofs. When replica services are offered, we consider proof of correct data storage in a single instance for both original files and their copies (otherwise specified). Replica options are an important service to be offered by cloud providers, in order to avoid unfortunate data loss due to technical issues (e.g. single point of failure).

In almost all works, only private verification is offered. Private verification means that only the client can challenge the server for correct storage. We think that such an assumption is too strong as clients may be too lazy to do so. We suggest that delegating such a checking process to a third party will guarantee a better storage experience for both clients and cloud providers.

The closest work to ours is definitely the work from [9]. Indeed, the authors aim to develop a cloud storage solution to prevent both rational cloud providers and malicious clients. To do so, they combine POR instances with RSA-based puzzles [50].

One limitation in [9] is from the private aspect of POR instance verification. Indeed, either the client or someone in possession of the secret key can perform the verification of integrity of stored data. We aim to develop a publicly-verifiable PORR scheme, that allows everyone, using only public elements, to check that the cloud provider correctly stores original files and their replicas.

2 Contributions and Road Map

In [1], the first Proof of Retrievability and Reliability (PORR) with public verification was proposed. The original file of the client is first uploaded to a cloud provider. The replica plan has been agreed by both the client and cloud provider. Following that plan, replicas of the original file are prepared by the cloud provider. Anyone (and possibly the client) can challenge the cloud provider to check the integrity of the stored files and replicas.

In [1], the publicly verifiable RSA-based POR scheme from Shacham and Waters [10, 11] was combined with the exponentiation-based VDF construction from [12]. This allows us to define a common parameter setting, roughly a RSA-based one, while benefiting all the properties offered by the two schemes. Namely, the verifier can efficiently launch a large number of challenge requests and checks the cloud provider's responses using only public elements, while the cloud provider is forced to store the client's original file along with all the replicas at rest (otherwise being timely noticed).

In this paper, we provide the experimental analysis of our publicly verifiable PORR protocol. Implementation and evaluation results reveal realistic applications. Communication overhead between the client, the cloud provider and the verifier remains fair. Computation costs are affordable for the client and verifier, while they are made such that the cloud provider does not gain in computing the replicas on the fly from challenge requests. We also compare our solution with MIRROR [9], a PORR prototype with private verification, built upon Waters and Shacham's POR [10] and RSA-based puzzles [50].

Therefore, our solution fits cloud-related bandwidth and storage requirements, while preventing malicious clients and rational cloud providers from being successful.

In the following Sections 3 and 4, we recall the definition and construction of Shacham-Waters POR [10, 11] and of VDF from [12]. The reader can directly jump to Section 5 if being familiar with concepts of POR and VDF. In Section 5, our publicly verifiable PORR protocol presented in [1] is described. In Section 6, we depict the implementation framework of our PORR and analyse the experimental results. We conclude this paper in the last section.

3 RSA-based POR with Public Verification

In this section, we recall the publicly verifiable RSA-based POR construction from [10, 11]. Since our choice of slow functions [12] is based on RSA and also publicly verifiable, such POR construction is welcomed. The RSA-based POR construction in [11] is an extension of the RSA-based PDP construction in [4].

3.1 Definition

A Proof Of Retrievability (POR) comprises five algorithms. The client runs the two first algorithms. A public and secret key pair (pk, sk) are generated by running the key generation algorithm. For each file M, the client computes a corresponding tag T. Using the secret key sk, the file and tag generation algorithm processes the file M, resulting into M^*, and computes the tag T.

The client challenges the cloud provider when she wants to verify her files are stored in their entirety. A challenge set Q, selected when running the challenge generation algorithm, defines the file blocks that the cloud provider must prove to store correctly. From this challenge, the latter replies by running the response generation algorithm and shares the response $resp$ with the client. The latter can then verifies $resp$ by running the verification algorithm. The output of this algorithm tells the client whether or the cloud provider stores her files in their entirety.

In the above description, we suggest that the client verifies the cloud provider's response. However, in a publicly verifiable case, anyone can verify $resp$ since only public elements are required to run the verification algorithm. In a privately verifiable case, only someone with a secret element can check $resp$, the client in general would have such competency, but a private verifier can also be given a secret key. We only consider the first case in our paper.

POR protocols must be proved correct and sound. Correctness requires that for all key pairs (pk, sk) output by the key generation algorithm, for all file $M \in \{0,1\}^*$, and for all pair (M^*, T) output by the file and tag generation algorithm, the verification algorithm accepts with challenge and response respectively output by the challenge generation algorithm and the response generation algorithm. Soundness states that if a cheating cloud provider convinces the verifier that it is storing the file M, then it is actually storing the file M. Shacham and Waters formalize the notion of an extractor algorithm that interacts with the cheating cloud provider using the POR protocol.

3.2 Notations

The processed file M' (obtained from applying an erasure code on the file M) is split into n blocks, and each block is then split into s sectors. Each sector $m_{i,j}$, for $i \in [1, n]$ and $j \in [1, s]$, is an element of $\mathbb{Z}_N$. For instance, for a processed file M' that is b-bit long, there are $n = \lceil b/s \log(N) \rceil$ blocks.

A challenge is an l-element set $Q = \{(i, v_i)\}$. The size l of the set Q is a system parameter. Each tuple $(i, v_i) \in Q$ can be described as follows:

- The value i is a block index in $[1, n]$.

- The value v_i is an element in $E \subseteq \mathbb{Z}_N$. Let $E \subseteq \mathbb{Z}_N$ be the set of coefficients v_i chosen for verification requests. For instance, $E = \mathbb{Z}_N$, such that coefficients v_i are randomly chosen from $\mathbb{Z}_N$.

3.3 Construction

Let κ be the security parameter. Let κ_1 be a bit length such that the difficulty of factoring a $(2\kappa_1 - 1)$-bit modulus fits the security parameter κ. Let $max(E)$ be the largest element in E. Let κ_2 be the bit length equal to $\lceil \log(l \cdot max(E)) \rceil + 1$ [10, 11].

The construction of publicly verifiable POR with RSA signatures given in [11] is as follows:

Key Generation: On input a security parameter κ, the randomized key generation algorithm outputs the public key and secret key pair (pk, sk) as follows. Let $(spk, ssk) \leftarrow \mathsf{S.KeyGen}(\kappa)$ be a random signing key pair. Choose two primes p and q at random such that $p, q \in [2^{\kappa_1 - 1}, 2^{\kappa_1} - 1]$. Let $N = pq$ be the RSA modulus such that $2^{\kappa_1 - 2} < N < 2^{\kappa_1}$. Let $G : \{0,1\}^* \rightarrow \mathbb{Z}_N^*$ be a full-domain hash function, seen as a random oracle. Pick at random a prime e of length $2\kappa_1 + \kappa_2$ bits, and set $d = e^{-1} \mod \phi(N)$.

Set the public key $pk = (N, e, G, spk)$ and the secret key $sk = (N, d, G, ssk)$.

File and Tag Generation: On inputs the secret key sk and the file M, the file and tag generation algorithm outputs a processed file M^* and the tag T as follows:

1. Apply the erasure code on M to obtain M'.
2. Split M' into n blocks for some integer n, each of them being s sectors long, resulting into a $n \times s$ matrix $\{m_{i,j}\}_{i \in [1,n], j \in [1,s]}$. Each sector $m_{i,j}$ is an element of $\mathbb{Z}_N$.
3. Choose a random file identifier $id \in \mathbb{Z}_N$.
4. Pick at random s random numbers $u_1, u_2, \cdots, u_s \in \mathbb{Z}_N^*$.
5. Let $T_0 = id \| n \| u_1 \| u_2 \| \cdots \| u_s$. Compute the file tag $T = T_0 \| \mathsf{S.Sig}_{ssk}(t_0)$.
6. For each $i \in [1, n]$, compute $\sigma_i = (G(id \| i) \cdot \prod_{j=1}^{s} u_j^{m_{i,j}})^d \mod N$.
7. The processed file is $M^* = (\{m_{i,j}\}_{i \in [1,n], j \in [1,s]}, \{\sigma_i\}_{i \in [1,n]})$.

Challenge Generation: On inputs the secret key sk and the tag T, the randomized challenge generation algorithm outputs the challenge set Q as follows:

1. Use the key spk to verify the signature on T. If the signature is invalid, then output 0 and halt.
2. Otherwise, recover $id, n, u_1, u_2, \cdots, u_s$.
3. Pick at random an l-element subset I from the set $[1, n]$. For each $i \in I$, choose a random element $v_i \in E$. Let $Q = \{(i, v_i)\}_{i \in I}$.

Response Generation: On inputs the processed file M^* and the challenge set $Q = \{(i, v_i)\}_{i \in I}$, the deterministic response generation algorithm outputs a response $resp$ as follows:

1. Compute $\mu_j = \sum_{(i,v_i) \in Q} v_i m_{i,j} \in \mathbb{Z}$ for $j \in [1, s]$ (note that there is no modular reduction).
2. Compute $\sigma = \prod_{(i,v_i) \in Q} \sigma_i^{v_i} \mod N$.
3. Set the response $resp = (\{\mu_j\}_{j \in [1,s]}, \sigma)$.

Verification: On inputs the response $resp$, the deterministic verification algorithm checks first whether each μ_j is in the range $[0, l \cdot N \cdot max(E)]$. If some values are not in the range, then halt and output 0. Otherwise, check whether the equation $\sigma^e = \prod_{(i,v_i) \in Q} G(id \| i)^{v_i} \times \prod_{j=1}^{s} u_j^{\mu_j} \mod N$. If so, then it outputs 1; otherwise, it outputs 0.

3.4 Security

The security for POR is linked to *unforgeability*, *extractability* and *retrievability*. Informally, the security proof is split in three parts such that:

- The first part shows that the verifier can not receive an illegitimate (i.e. forged) response from the adversary (i.e. the cheating cloud provider).
- The second part shows that there exists an extractor extracting a constant fraction δ of file blocks (previously encoded with an erasure code) as soon as there exists an adversary that succeeds the verification process a noticeable number of times. Indeed, all responses that have been checked should be legitimate.
- The last part shows that one can use an erasure code to rebuild the whole file M if the constant fraction δ of file blocks has been recovered.

The RSA-based POR construction has been proved correct and sound. We let the readers refer to [10, 11] for the formal security proof.

4 Verifiable Delay Functions From Exponentiation in a Finite Group

In order to prevent rational behaviour from cloud providers, we aim to use slow functions (also called puzzles). Slow functions are constructed such that:

- Puzzles and file blocks are combined, resulting into r correct replicas of the file M. Homomorphic properties required for compact proofs are preserved.
- The client can adjust the difficulty of the puzzles supply multiple cost metrics.
- The costs from the storage of the solution of the puzzle are at least as high as the storage needed for the replicas of the file M.
- The cloud provider needs more time to solve the puzzles than the client, since the latter has access to a trapdoor.

In this section, we recall the construction of VDFs [12] that we use in our PORR. In order to successfully combine Shacham-Waters POR with VDFs, we opt for the exponentiation-based version with bounded pre-computations in a RSA group. This version is secure against attackers with bounded pre-computations, from a generalization of exponentiation-based time-lock puzzles in groups of unknown order.

4.1 Definition

An algorithm is said to run with p processors in parallel time t if one can implement it on a PRAM machine with p parallel processors that run in time t. The total sequential time refers to the time required for computation on a single processor.

A Verifiable Delay Function (VDF) is composed of three algorithms, to set up the system, evaluate the slow function and verify a solution. On inputs a security parameter κ and a delay parameter t, the randomized setup algorithm outputs the public parameters pp containing an evaluation key ek and a verification key vk. This algorithm must be polynomial-time in κ. The delay parameter t must be sub-exponentially sized in κ.

On inputs the evaluation key ek and a puzzle x from some known sampleable set, the evaluation algorithm outputs the solution y and a (possibly empty) proof π. For all public parameters pp and for all puzzles x, the evaluation algorithm must run with $poly(\log(t), \kappa)$ processors in parallel time t. For a given puzzle x, there must be a unique output y whose verification will be correct.

On inputs the verification key vk, the puzzle x, the solution y and its proof π, the deterministic verification algorithm outputs either 1 if y is a valid solution for the puzzle x, or 0 otherwise. The verification algorithm must run in total time polynomial in $\log(t)$ and κ. This algorithm is much faster than the evaluation one.

The VDF is expected to be *sequential* where honest parties can compute (y, π) by running the evaluation algorithm in t sequential steps, while no parallel-machine adversary with a polynomial number of processors can make the distinction between the output y from a random value in many less steps. Moreover, the VDF is expected to be *efficiently computable*, where honest parties run the verification algorithm as fast as possible such that the total time should be $O(polylog(t))$. The VDF is finally expected to be *unique* where for all puzzles x, the value y is difficult to calculate such that the verification algorithm outputs 1 while y is not an output of the evaluation algorithm on inputs pp and x.

4.2 Notations

Given an integer n, let $[1, n]$ denote the set of integers $\{1, 2, \cdots, n\}$. Let $L = \{l_1 = 3, l_2 = 5, \cdots, l_t\}$ be the first t odd primes. The parameter t is the provided delay parameter. Let P be the product of the primes in L, i.e. $P = l_1 \cdot l_2 \cdots l_t$. The parameter P is a large integer with about $t \log t$ bits.

4.3 Construction

Here, we describe in details the exponentiation-based solution for VDF in [12].

Setup: The trusted setup process is the following:

1. Set a RSA modulus $N = pq$ (for instance, 4096 bits long) such that the prime factors p, q are strong primes. The factorization of N is only known by the trusted setup algorithm. Let $H : \mathbb{Z} \to \mathbb{Z}_N^*$ be a random hash function.
2. For a given pre-processing security parameter B (for instance, $B = 2^{30}$), do the following:
   - Compute $H(i) = h_i \in \mathbb{Z}_N^*$ and $g_i = h_i^{1/P} \in \mathbb{Z}_N^*$ for $i \in [1, B]$.
   - Set $ek = (\mathbb{Z}_N^*, H, g_1, g_2, \cdots, g_B)$ and $vk = (\mathbb{Z}_N^*, H)$.

While the parameters of the verifier are short, the ones of the evaluator are not.

Evaluation: Solving a puzzle x works as follows:

1. Map the puzzle x to a random subset $L_x \subseteq L$ of size κ and a random subset S_x of κ values in $[1, B]$, using a random hash function.
2. Let P_x be the product of all the primes in L_x and let $g = \prod_{i \in S_x} g_i$.
3. The puzzle solution is $y = g^{P/P_x}$.

The computation of the solution takes $O(t \log t)$ multiplications in $\mathbb{Z}_N^*$.

Verification: Verifying a solution y works as follows:

1. Compute P_x and S_x as in the evaluation algorithm on inputs ek and x.
2. Compute $h = \prod_{i \in S_x} H(i)$.
3. Output 1 if and only if $y^{P_x} = h$.

We observe that exactly one element $y \in \mathbb{Z}_N^*$ will be accepted as a solution for a puzzle x. The verification process takes only $\tilde{O}(\kappa)$ group operations.

4.4 Security

Security is defined in face of an attacker able to perform polynomially bounded pre-computations. A VDF scheme must satisfy:

Correctness and Soundness. Every output of the evaluation algorithm must be accepted by the verification algorithm. The solution y for a puzzle x is guaranteed to be unique because the evaluation algorithm evaluates a deterministic function on the sampleable set of puzzles. The proof π does not require to be unique but should be sound and a verifier cannot be convinced that some different output is the correct VDF outcome.

τ-Sequentiality. No adversary should be able to compute an output for the evaluation algorithm on a random puzzle in parallel time $\tau(t) < t$, even with up to many parallel processors, and after a potentially large amount of pre-computations.

We let the readers refer to [12] for more details on the security models for VDFs.

The above construction does not satisfy the definition of a secure VDF presented in [12]. More precisely, an adversary who is able to run a large pre-computation once the parameters pp are known can break the above construction. There are various possible pre-computation attacks requiring tB group operations in $\mathbb{Z}_N^*$ [12].

New parameters must be generated after B challenges; otherwise, the scheme is not secure. This is sufficient for our application of a VDF, for instance by choosing $B = 2^{30}$. Regarding experiments of other solutions [9, 45, 46], storage challenges never excess such B.

5 RSA-based PORR with Public Verification

In this section, we describe our RSA-based solution for Proof Of Retrievability and Reliability (PORR) with public verification, using exponentiation-based VDFs to prevent the cloud provider to generate replicas on the fly when being challenged.

A client encodes and then processes a file M into M^*, and outsources the latter to the cloud provider. The cloud provider then commits to store M^* entirely across a set of r storage nodes with reliability guarantee R. This means that M^* contain the original copy of M along with replicas.

A PORR protocol is executed between a client and a cloud provider provided by its k storage nodes. The goal of such protocol is to enable either the client or anyone else to check the integrity and reliability of the processed file M^*.

5.1 Definition

Informally, our PORR protocol combines the RSA-based POR scheme of Shacham and Waters [10] and the exponentiation-based VDF scheme from [12].

The *Setup* phase initiates the protocol. It corresponds to the key generation and file and tag generation algorithms of Shacham-Waters POR scheme and to the setup algorithm of the VDF scheme. The client generates the parameters of the protocol, corresponding to the ones found in the two underlying schemes. She prepares her to-be-stored M by encrypting and processing it, and by generating the tag T and the authenticators σ_i for $i \in [1, n]$ (where n is the number of blocks) as in the POR scheme [10]. According to the agreed number r of replicas that the cloud provider must store, the client also prepares the VDF puzzles $x_{i,j}^{(k)}$, for $i \in [1, n]$, $j \in [1, s]$ and $k \in [1, r]$. In other words, there is one challenge per sector per replica (we recall that there are s sectors per block).

Once the cloud provider receives the file-related elements to be stored, the Setup phase is over. It can then start the second phase, namely the *Replica Generation* phase, in order to create the r replicas of the original file M. To do so, the cloud provider evaluates the VDF puzzles $x_{i,j}^{(k)}$ by running the evaluation algorithm. It appends each solution $y_{i,j}^{(k)}$ with the corresponding sector replica $m_{i,j}^{(k)}$, for $i \in [1, n]$, $j \in [1, s]$ and $k \in [1, r]$.

The next three phases can be requested multiple times. There is an interaction between a verifier (the client herself or someone on her behalf) and the cloud provider. During the *Challenge Generation* phase, the verifier generates the challenge *chal* and sends it to the cloud provider, by running the challenge generation algorithm of Shacham-Waters POR scheme. During the *Response Generation* phase, the latter replies back to the client with a response *resp* by running the response generation algorithm of POR scheme. Finally, during the *Verification* phase, the verifier then checks *resp* using only public elements, using the verification algorithms of Shacham-Waters POR scheme and of the VDF scheme. Indeed, verification exactly contains two steps: one check for POR and one check for VDF. If the output is 1, then the verifier is guaranteed that the cloud provider stores the file in its entirety along with its r replicas.

5.2 Construction

Here, we describe in details our publicly verifiable RSA-based PORR solution.

Setup: This phase includes the POR-based key, file and tag generations along with the VDF-based setup process.

Let κ be the security parameter. Let $(\mathsf{S.KeyGen}, \mathsf{S.Sign}, \mathsf{S.Verify})$ be a digital signature scheme. Choose two primes p and q at random such that $p, q \in [2^{\kappa_1 - 1}, 2^{\kappa_1} - 1]$. Let $N = pq$ be the RSA modulus such that $2^{\kappa_1 - 2} < N < 2^{\kappa_1}$. Let $G : \{0,1\}^* \to \mathbb{Z}_N^*$ be a full-domain hash function, seen as a random oracle. Pick at random a prime e of length $2\kappa_1 + \kappa_2$ bits, and set $d = e^{-1} \mod \phi(N)$. Let t be the delay parameter and B be the security parameter for VDFs. Let $L = \{l_1 = 3, l_2 = 5, \cdots, l_t\}$ be the first t odd primes and $P = l_1 \times l_2 \times \cdots \times l_t$. Let $H : \mathbb{Z} \to \mathbb{Z}_N$ be a hash function, seen as a random oracle.

The client wishes to store a file $M \in \{0,1\}^*$ at the cloud. Without loss of generality, the file M is assumed to be encrypted and encoded (using the specific erasure code). Encryption guarantees confidentiality and encoding guarantees extractability and retrievability.

As in [10], the file M is first split into n blocks, and then split into s sectors. Let us denote a sector as $m_{i,j} \in \mathbb{Z}_N$, for $i \in [1, n]$ and $j \in [1, s]$. Bit representation of each sector $m_{i,j}$ includes a characteristic pattern (e.g. a sequence of zero bits), in order to guarantee extractability [9]. Pattern length and file size are dependent such that the former should be larger than $\log_2(n \cdot s)$.

The client runs the algorithm $\mathsf{S.KeyGen}(\kappa)$ and gets the signing and verification key pair (ssk, spk). She also chooses an identifier $id \in \mathbb{Z}_N$ for the processed file M^*. She then picks at random s non-zero elements $u_1, u_2, \cdots, u_s \in \mathbb{Z}_N$. The client computes $T_0 = id \| n \| u_1 \| u_2 \| \cdots \| u_s$ and then $T = T_0 \| \mathsf{S.Sign}_{ssk}(T_0)$. Moreover, the client calculates $\sigma_i = (G(id \| i) \cdot \prod_{j=1}^{s} u_j^{m_{i,j}})^d \mod N$ for $i \in [1, n]$. We notice that all operations are done in the multiplcative group $\mathbb{Z}_N^*$ of invertible integers modulo N.

Both the cloud provider and client have agreed to create r replicas of the file M and store all of them at rest. She com-

putes $h_i = H(i)$ and $g_i = h_i^{1/P}$ for $i \in [1, B]$. She also chooses the values $x_{i,j}^{(k)}$ for $i \in [1, n]$, $j \in [1, s]$ and $k \in [1, r]$.

Finally, the client uploads the processed file $M^* = (\{m_{i,j}\}_{i\in[1,n],j\in[1,s]}, \{\sigma_i\}_{i\in[1,n]})$ to the cloud provider. She also forwards the public parameters $params = (N, e, G, H, spk, L, P, T, \{g_w\}_{w\in[1,B]}, \{x_{i,j}^{(k)}\}_{i\in[1,n],j\in[1,s],k\in[1,r]})$ to the cloud provider and anyone interested in playing the role of the verifier.

The client keeps secret the tuple (N, d, G, ssk).

Replica Generation: This phase includes the evaluation process of the underlying VDF.

The cloud provider calculates the solution $y_{i,j}^{(k)}$ for each $x_{i,j}^{(k)}$, and then build the replica $m_{i,j}^{(k)}$ of the original sector $m_{i,j}$, for $k \in [1, r]$.

First, the cloud provider maps $x_{i,j}^{(k)}$ to $L_{i,j}^{(k)} \subseteq L$ of size κ and the random subset $S_{i,j}^{(k)}$ of κ values in $[1, B]$, using a random hash function. Second, it sets $P_{i,j}^{(k)}$ as the product of all primes in $L_{i,j}^{(k)}$ and computes $g_{i,j}^{(k)} = \prod_{w\in S_{i,j}^{(k)}} g_w$. Third, the cloud provider computes the solution $y_{i,j}^{(k)} = (g_{i,j}^{(k)})^{P/P_{i,j}^{(k)}} \in \mathbb{Z}_N$.

Finally, it computes $m_{i,j}^{(k)} = m_{i,j} + y_{i,j}^{(k)}$ as the k-th replica of the sector $m_{i,j}$.

Challenge Generation: This phase corresponds to the challenge generation of POR.

First, the verifier (possibly the client) generates the challenge $chal$. Given $T = T_0 \| \mathsf{S.Sig}_{ssk}(T_0)$, check that $\mathsf{S.Sign}_{ssk}(T_0)$ is a valid signature by running the algorithm $\mathsf{S.Verify}_{spk}$. If the signature is invalid then halt.

Thereafter, elements $id, n, u_1, u_2, \cdots, u_s$ are recovered the verifier. The latter then sets $I \subset [1, n]$ of l elements and randomly selects l elements $v_i \in \mathbb{Z}_N$, for $i \in I$. Then, let $Q = \{(i, v_i)\}_{i\in I}$ where i is defined as the index of the block m_i. A set $R \subset [1, r]$ is also set by the verifier. Finally, let $chal = (Q, R)$ be forwarded to the cloud provider.

Response Generation: This phase corresponds to the response generation of POR.

Upon reception of the challenge $chal$, the cloud provider creates its response $resp$. First, it computes $\mu_j = \sum_{(i,v_i)\in Q} v_i m_{i,j} \in \mathbb{Z}$ and $\sigma = \prod_{(i,v_i)\in Q} (\sigma_i \cdot \prod_{j=1}^{s} \prod_{k\in R} u_j^{m_{i,j}^{(k)}})^{v_i} \mod N$. It sets $resp = (\{\mu_j\}_{j\in[1,s]}, \sigma)$ and forwards it to the verifier.

Verification: This phase includes both POR- and VDF-based verification steps.

Upon reception of the response $resp$, the verifier verifies that whether each μ_j is in the range $[0, l \cdot N \cdot max(\mathbb{Z}_N)]$. The verifier halts and outputs) as soon as some values are not in the range. Otherwise, the verifier checks whether the following equation holds:

$$\sigma^e = \prod_{(i,v_i)\in Q} G(id\|i)^{v_i} \times \prod_{j=1}^{s} u_j^{\mu_j(1+|R|e)} \times \Big(\prod_{(i,v_i)\in Q} \prod_{j=1}^{s} \prod_{k\in R} u_j^{v_i (h_{i,j}^{(k)})^{1/P_{i,j}^{(k)}}} \Big)^e \mod N \quad (1)$$

The verifier outputs 1 if the equation holds. She outputs 0 otherwise.

5.3 Security

In this section, we first describe the possible misbehaving entities in PORR, w.r.t. the cloud provider and client. We then present the security goals of a PORR scheme and define the correctness requirements, according to the ones in [9]. Our PORR solution must guarantee three security goals, namely:

- **Extractability:** The client can recover the original file M in its entirety.
- **Soundness of replica generation:** The replicas of the original file M must be correctly generated.
- **Storage allocation commitment:** The cloud provider utilizes at least as much storage as required to store the original file M and its replicas.

Traditional security notions are either embedded in those security notions (e.g. integrity), or assumed to be guaranteed by default (e.g. confidentiality).

We also sketch the security proofs of our scheme according to their respective security goals. We let the reader reach [9, 10, 11, 45, 46] for more details about the PORR security models and proofs. Indeed, the extractability proof is based on the security of the original POR scheme [10, 11]. Proofs for soundness of replica generation and storage allocation commitment follow the same arguments than in MIRROR [9]. Those security goals are however made guaranteed based on the security of the VDF puzzles [45, 46] in our case rather than of the RSA-based puzzles [50] as in MIRROR.

5.3.1 Misbehaving Entities.

Two entities, namely a cloud provider and a client, participate in our PORR protocol. Both are assumed to attempt malicious behaviours.

Rational Cloud Provider: If the cloud provider cannot save any costs by misbehaving, then it is likely to simply behave honestly. The advantage of an adversarial cloud provider depends on the ratio between costs for storage and accessibility to various resources (e.g. computing) and their availability. Hence, a cloud provider is restricted to a bounded number of concurrent threads of execution.

Therefore, we say that a cloud provider is *rational* when it can achieve cost savings by cheating. For instance, the cloud provider may attempt to get some storage space saved while

the overall cost for operations has not increased. Such cost is limited to the number of storage servers along with limited computational and storage capacities of each server. If supplying computational resources incurs additional costs, then the cloud provider invests in extra computing resources if such a strategy would result in lower total costs (including the underlying costs of storage).

Let us assume that a rational cloud provider achieves to generate a valid response while not reliably storing the client's data. If in order to reach such a behaviour, the cloud provider either provides more resources for storage or provides more resources for computing than resources when it follows the protocol honestly, then the cloud provider likely decides to not behave maliciously.

The client is protected from a misbehaving cloud provider that is not storing the file in its entirety by considering the *extractability* notion. The client is also protected from a misbehaving cloud provider that is not committing enough storage to keep all the file replicas by considering the *storage allocation* notion. Both properties guarantee both the integrity of the original file and its replicas. This means that the cloud provider invests enough redundancy to keep the client's data safe.

Malicious client: We say that a client is *malicious* when she can encode additional data in the replicas by cheating. This additional data cannot be found in the original file.

A client may attempt to abuse on storing more data in the file replicas than what has been approved between the client and cloud provider. In particular, replicas may have a lower cost than their original files (e.g. Amazon S3). Therefore, additional data may be inserted into replicas by a malicious client.

The cloud provider is protected from such a misbehaving client by considering the *correct replication* notion.

The security model used to build our security proofs comes from [9, 45, 46]. We do not consider the confidentiality of the outsourced file: we simply assume that the client encrypts it before the start of the PORR protocol.

We first show the correctness of our PORR scheme. We then move forward to the security properties with relation to cheating cloud provider and client, namely extractability, storage allocation and correct replication.

Informally, our PORR must achieve the following properties:

- *Extractability:* The file can be recovered in its entirety if and only if at least a fraction δ is stored at the cloud provider.
- *Storage allocation:* Misbehavior will be detected with overwhelming probability if less than a fraction δ is stored at the cloud provider.
- *Correct replication:* The client does not participate in the replica generation. Moreover, the size of the file is independent of the size of the parameters needed to create the replicas.

5.3.2 Correctness

If both the cloud provider and the verifier are honest, then on input the challenge *chal* sent by the verifier (output by the challenge generation algorithm), the response generation algorithm (run by the cloud provider) generates a response *resp* such that the verification algorithm outputs "1" with probability 1. This means that an honest cloud provider should always be able to pass the verification of proof of data reliability. From it, the PORR scheme is said to be correct.

During the verification phase, if both the verifier and cloud provider are honest, then on input $Q = \{(i, v_i)\}_{i\in I}$ generated by the verifier, the cloud provider should output a response *resp* such that the Equation (1) holds with probability 1.

Proof. Let N be the modulus, e be the public exponent and d the private exponent. The elements $P_{i,j}^{(k)}$ and $S_{i,j}^{(k)}$ are calculated as in the Replication Generation phase. We recall that $h_{i,j}^{(k)} = \prod_{w\in S_{i,j}^{(k)}} H(w)$ and so:

$$(y_{i,j}^{(k)})^{P_{i,j}^{(k)}} = ((g_{i,j}^{(k)})^{P/P_{i,j}^{(k)}})^{P_{i,j}^{(k)}} = ((h_{i,j}^{(k)})^{1/P})^P = h_{i,j}^{(k)} \mod N$$

From a challenge set $Q = \{(i, v_i)\}_{i\in I}$, with $\mu_j = \sum_{(i,v_i)\in Q} v_i m_{i,j}$ and $\sigma = \prod_{(i,v_i)\in Q} \sigma_i^{v_i}$, we get the following mod N:

$$\begin{aligned}
\sigma &= \prod_{(i,v_i)\in Q} (\sigma_i \cdot \prod_{j=1}^{s}\prod_{k\in R} u_j^{m_{i,j}^{(k)}})^{v_i} \\
&= \prod_{(i,v_i)\in Q} \sigma_i^{v_i} \times \prod_{(i,v_i)\in Q}\prod_{j=1}^{s}\prod_{k\in R} (u_j^{m_{i,j}^{(k)}})^{v_i} \\
&= \prod_{(i,v_i)\in Q} (G(id||i) \cdot \prod_{j=1}^{s} u_j^{m_{i,j}})^{v_i d} \times \prod_{(i,v_i)\in Q}\prod_{j=1}^{s}\prod_{k\in R} (u_j^{m_{i,j}+y_{i,j}^{(k)}})^{v_i} \\
&= \prod_{(i,v_i)\in Q} (G(id||i)^{v_i} \cdot \prod_{j=1}^{s} u_j^{v_i m_{i,j}})^{d} \times \prod_{(i,v_i)\in Q}\prod_{j=1}^{s}\prod_{k\in R} u_j^{m_{i,j} v_i} \\
&\quad \times \prod_{(i,v_i)\in Q}\prod_{j=1}^{s}\prod_{k\in R} u_j^{y_{i,j}^{(k)} v_i} \\
&= \Big(\prod_{(i,v_i)\in Q} G(id||i)^{v_i} \times \prod_{j=1}^{s} u_j^{(\sum_{(i,v_i)\in Q} v_i m_{i,j})}\Big)^d \\
&\quad \times \prod_{j=1}^{s}\prod_{k\in R} u_j^{(\sum_{(i,v_i)\in Q} v_i m_{i,j})} \times \prod_{(i,v_i)\in Q}\prod_{j=1}^{s}\prod_{k\in R} u_j^{v_i (h_{i,j}^{(k)})^{1/P_{i,j}^{(k)}}} \\
&= \Big(\prod_{(i,v_i)\in Q} G(id||i)^{v_i} \times \prod_{j=1}^{s} u_j^{\mu_j}\Big)^d \times \prod_{j=1}^{s} u_j^{|R|\mu_j} \\
&\quad \times \prod_{(i,v_i)\in Q}\prod_{j=1}^{s}\prod_{k\in R} u_j^{v_i (h_{i,j}^{(k)})^{1/P_{i,j}^{(k)}}}
\end{aligned}$$

and so

$$\begin{aligned}\sigma^e &= \prod_{(i,v_i)\in Q} G(id\|i)^{v_i} \times \prod_{j=1}^{s} u_j^{\mu_j} \times \Big(\prod_{j=1}^{s} u_j^{|R|\mu_j} \\ &\times \prod_{(i,v_i)\in Q} \prod_{j=1}^{s} \prod_{k\in R} u_j^{v_i (h_{i,j}^{(k)})^{1/P_{i,j}^{(k)}}}\Big)^e \\ &= \prod_{(i,v_i)\in Q} G(id\|i)^{v_i} \times \prod_{j=1}^{s} u_j^{\mu_j(1+|R|e)} \\ &\times \Big(\prod_{(i,v_i)\in Q} \prod_{j=1}^{s} \prod_{k\in R} u_j^{v_i (h_{i,j}^{(k)})^{1/P_{i,j}^{(k)}}}\Big)^e \mod N\end{aligned}$$

5.3.3 *Extractability*

The client is protected against a malicious cloud provider that is not storing the file in its entirety, through the notion of extractability.

An honest client should recover her file M with high probability. Following the notion of extractability in [10, 11], if a cloud provider can convince a honest client with high probability that it stores the file M^* (i.e. the processed version of the original file M), then there exists an extractor algorithm that, given enough interaction with the cloud provider, can extract the file M.

Sketch of Proof. Let us define a game between an adversary and an environment. The environment is a simulation of honest clients and honest verifiers. The adversary is allowed to submit requests to the environment to create new clients, along with their public and secret parameters, to let them prepare files and their replicas, and to verify correct storage. More precisely, the environment simulates honest client and verifier, and it further provides the adversary with oracles for the algorithms to set up the system, upload the file and replicas, generate the challenge and verify the response.

At the end, the adversary chooses a client with its file M and outputs a cloud provider who can execute the verification process with this client and the file M. A cloud provider is said to be ϵ-admissible if the probability that the verifier does not abort is at least ϵ. The adversary thus picks the client along with the client's secret tuple (N, d, G, ssk), the file M and the public parameters *params*, and simulates a cheating cloud provider. Let the latter succeed in making the verification algorithm yield "1" in an non-negligible ϵ fraction of PORR executions. We say that the PORR scheme meets the extractability guarantee, if there exists an extractor algorithm such that given sufficient interactions with the cheating cloud provider, it recovers M.

The computations executed to upload the file M and to verify correct storage are similar operations to the ones in the publicly verifiable POR scheme in [10, 11]. Therefore, the extractability arguments given in [10, 11] apply to our PORR solution directly. An additional assumption is made on the existence of an erasure coding mechanism applied to the file, to guarantee the entire recovery of the file M from any file fraction δ. We let the reader refer to [10, 11] to obtain additional information on the choice of the erasure codes.

5.3.4 *Soundness of Replica Generation.*

Contrary to extractability notion, soundness of replica generation aims to protect the cloud provider against a malicious client who tries to encode additional data in the replicas.

Sketch of Proof. The replica generation is said to be sound such that if the client is involved in the replica generation, then the cloud provider can get the assurance that the additional uploaded data represents indeed correctly built replicas that do not encode extra data. In our PORR, this situation is solved by not letting the client be fully involved in the replica generation. Indeed, while the client generates the puzzles whom solutions are created by the cloud provider, the latter is responsible of defining each replica with its puzzle solution.

The replica generation does not allow to encode a significant amount of extra data in the replicas. Indeed, the replication process takes as inputs elements whose size is independent of the file size. The replica generation is also said to be correct if replicas represent indeed copies of the uploaded file M. Indeed, the file M can be efficiently recovered from any replica M_k. There exists an efficient algorithm which given the file tags, the public parameters and any replica M_k outputs M.

5.3.5 *Storage Allocation Commitment.*

Similarly to the extractability notion, the storage allocation commitment property aims to protect a client against a cloud provider who does not commit enough storage to store all replicas.

The storage allocation commitment notion forces the cloud provider to store the outsourced file and its replicas at rest. Therefore, a cheating cloud provider that participates in the above extractability game [10, 11] and devotes only a fraction of its storage space to store the file and the replicas entirely, cannot convince the verifier to accept its response with overwhelming probability.

Sketch of Proof. The storage allocation ratio of the cloud provider is defined as follows:

$$\rho = \frac{|st|}{|M| + |M_1| + \cdots + |M_r|}$$

where st corresponds to the storage of the cloud provider that has been allocated for storing the original file M and its replicas $M_1, \cdots, M_r$. The file has been first encrypted and the replicas are copies of the processed file, thus they cannot be further compressed. We can assume that the cloud provider aims to save storage, thus it holds that $0 \leq \rho \leq 1$. The storage allocation commitment ensures that $\delta \leq \rho$ for a threshold $0 \leq \delta \leq 1$ agreed with the client (see the above extractability notion).

First, we want the cloud provider to use at least as much storage as needed to keep the file and its replicas. Second, we see our scheme as a POR protocol applied to both the original file and all the replicas. Let a challenge contain sectors that are not correctly stored. Then, our scheme guarantees that the cloud provider will fail with overwhelming probability unless the correct reconstruction of those sectors is possible.

Lastly, a malicious cloud provider should give a noticeable effort in reconstructing missing sectors. We can easily investigate such

an effort as follows. Let the cloud provider store the whole file but only parts of the replicas. The cheating cloud provider will require a significantly higher effort in recomputing missing replicas, compared to an honest service provider that has all the replicas stored in their entirety. By using slow functions (here the VDFs), the time in getting back the missing parts of replicas is noticeable from the verifier's point of view.

More precisely, the misbehaving cloud provider must compute the puzzle solutions $y_{i,j}^{(k)}$ in order to recompute the missing sectors. Since those elements $y_{i,j}^{(k)}$ are different for each replica, knowing (or reconstructing) one element $y_{i,j}^{(k)}$ from one replica sector does not help the cloud provider in deriving elements from other replica sectors. A rational cloud provider should thus require a significantly higher effort compared to an honest cloud provider in recomputing missing replicas.

Given that the VDF evaluation function requires a noticeable amount of time and effort compared to the associated VDF verification function, this incurs additional (significant) computational overhead on the cloud provider to compute the puzzle solutions on the fly rather than storing them at rest.

Let δ be the threshold selected by the client (see Extractability). Let us assume that less than a fraction δ of all sectors of a replica is stored at the cloud provider. Thus, for each element $y_{i,j}^{(k)}$ in the challenge, the probability to recompute it is at least $1 - \delta$. In addition, $l \cdot s$ sectors are contained in a challenge. Hence, the number of values to recompute is $l \cdot s \cdot (1 - \delta)$. Moreover, the costs from the time effort for recomputing these elements $y_{i,j}^{(k)}$ exceed the costs from the storage of those elements. This is made possible by having a number of challenges linear in the security parameter κ. If so, then the overall probability to avoid these computations is negligible in κ.

6 Implementation and Evaluation

We are interested in implementing and evaluating our PORR solution in a realistic cloud framework. We also wish to compare our results with the ones from [9], since their solution is the closest one to ours.

In this section, we first describe MIRROR, the PORR scheme presented in [9]. We then describe our implementation setting, and we discuss our results and compare them with the ones from [9]. We choose to compare our results with MIRROR's ones since, to our best of knowledge, MIRROR is the closest prototype to ours. Unlike existing schemes, the cloud provider replicates the data by itself in both MIRROR and our PORR. Therefore, expensive bandwidth resources are traded with cheaper computing resources.

As a summary, we evaluate an implementation of our PORR prototype within a realistic cloud setting and we compare the performance of it with MIRROR [9].

6.1 MIRROR

MIRROR [9] is the first scheme to enable the cloud provider to generate the replicas of the client's files by itself. Such move permits to trade expensive bandwidth resources with cheaper computing resources. Cloud providers and clients are likely to adopt it easily since storage services are improved while financial costs are lowered.

The authors in [9] present new PORR definition and security model. Their definition extends Shacham-Waters POR from [10]. Their security model encompasses security risks that have not been covered in previous multi-replica POR schemes, namely security against malicious clients and rational cloud providers. We have sketched the security proofs of our PORR based on their model.

The authors give a solution for PORR, called MIRROR, and prove it secure according to their new security model. Their motivations mostly rely on business matters for cloud providers and financial incentives for clients. They propose a tunable replication scheme by combining Linear Feedback Shift Registers and RSA-based puzzles [50]. By doing so, the burden incurred by the construction of the replicas is shifted to the cloud provider. The latter swaps higher resources for bandwidth with lower resources for computation. In addition, a realistic cloud framework has been defined for the implementation and evaluation of MIRROR. The results show that MIRROR is applicable to real cloud environments.

The main difference between MIRROR and our PORR solution is the nature of the verification process, private and public respectively. The authors in [9] propose to use privately verifiable puzzles to prevent rational behaviour from cloud providers. Therefore, such feature forces the puzzle creator, namely the client, to check herself the solution generated by the cloud provider. We claim that such requirement is too strict on the client's side. If we aim to design a solution for the general public, then we think about individuals that have limited knowledge on cloud storage security, thus that would not follow the process to prevent and/or detect malicious cloud provider behaviours. Instead, we suggest to use VDFs that offer public verification. This implies that anyone else on behalf of the client can verify that the cloud provider has been acting honestly.

6.2 Implementation

We implement our PORR scheme in order to analyze the computational efforts from honest and rational cloud providers, along with the ones from client and verifier.

We show that our protocol produces fair computational and communication overheads on the client, cloud provider and verifier, meaning that our solution is realistically applicable in cloud storage environments. Indeed, computation costs are affordable for the client and verifier, while they are made such that the cloud provider does not gain any advantage in computing the replicas on the fly from the verifier's challenge requests. Therefore, our solution fits cloud-related bandwidth and storage expectations, while preventing malicious clients and rational cloud providers from being successful.

Our implementation setting follows the one in [9] for a more accurate and legitimate comparison. Our code is written in Python 3.8. The selected hash function is SHA-256. The whole test environment is deployed on a PC running an Intel Core i7-9700 with 32GB of memory. We create four Virtual Machines (VMs) to design our test environment, namely:

- One VM represents the client that owns files to be stored on a cloud storage server.

- One VM represents the cloud provider (server) that offers cloud storage services.
- Two VMs represent storage nodes, linked to the cloud provider, that guarantee data replication.

We depict the test environment topology in Figure 2.

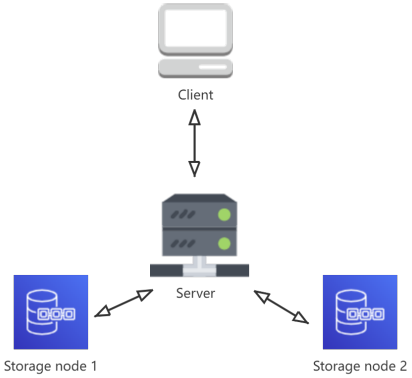

Figure 2: Test environment topology

To emulate a realistic Wide Area Network (WAN), the communication between the various VMs is bridged using a 100 Mbps switch. All traffic exchanged in the environment is shaped with NetEm [56]. The setting parameters are the following:

- We set the packet loss rate at 0.1%, and the correlation rate of lost packets at 0.001% [57, 58, 59].
- We set the delay rate to fit normal distribution with a mean of 20ms and a variance of 4ms [60].
- We set the packet corruption rate at 0.1% [61].
- We set the rate of the package being out of order at 0.2% [62].

Each data point is averaged in our plots by five independent measurements. Where appropriate, we include the corresponding 95% confidence intervals.

Parameter selection. We select our parameters similarly to [9] in order to achieve a better comparison. Let the modulus N be 2048-bit long (hence, 1024-bit primes p and q). We choose a block size equal to 8KB as such value offers the most balanced performance in average. This means that files of 64MB are split into n blocks where $n = 8,000$. Each block is then split into s sectors where $s = 32$. Then, r replicas are created for each file where $r = 2$. The cloud provider is thus expected to keep $8000 \times 32 \times 2 < 52 \cdot 10^4$ file elements, namely the original file sectors and the replica sectors. The number of blocks per challenge will be set to $|I| = 40$.

We consider three network storage protocols, namely File Transfer Protocol (FTP), Server Message Block (SMB) and Network File System (NFS), to enable the client to access her files stored in the cloud, through a computer network. NFS is used on Unix operating systems and SMB on Windows operating systems. The deployment of FTP is wider, as long as the communication port is open. In order to select one of those three network storage protocols, we test their uploading speed. Figure 3 shows the uploading speed values (in MB per second) for network storage protocols FTP, SMB and NFS. Due to the instability of network transmission, the uploading speed results show some ups and downs. The average FTP speed is the fastest, at 17.4 MB per second, which is around 1 second faster than NFS and 3 seconds faster than SMB. Hence, FTP was chosen as the default setting.

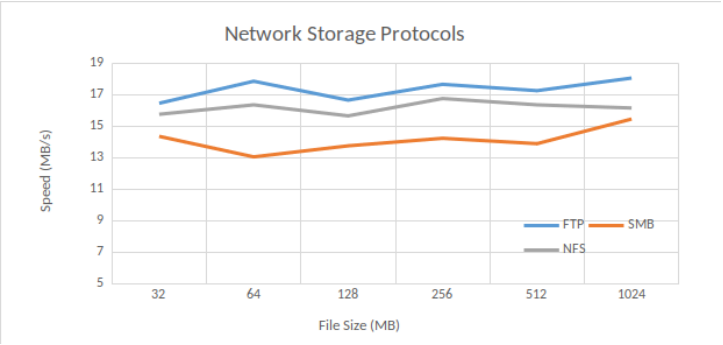

Figure 3: Network Storage Protocols

We summarize our parameter selection in Table **??**. When not specifically mentioned in Section 6.3, the default parameter values are kept as in Table **??** for our protocol evaluation.

Table 2: Parameter selection

Parameter	Default value
RSA modulus size $\|N\|$	2048 bits
RSA prime size $\|p\|$	1024 bits
RSA prime size $\|q\|$	1024 bits
Pre-processing parameter B	2^{30}
Delay parameter t	2^{16}
File size	64MB
Block size	8192 Bytes
Sector size	256 Bytes
Number of replicas r	2
Number of challenges $\|I\|$	40
Network Storage Protocol	FTP

VDF implementation. We recall that our PORR scheme is based on the exponentiation-based VDF scheme [12]. The latter has the following characteristics.

The VDF evaluation algorithm runs with $poly(\log(t), \kappa)$ processors in parallel time t, where κ is the security parameter and t is the delay parameter. We say that an algorithm runs with p processors in parallel time t if one can implement it on a PRAM machine with p parallel processors running in time t.

The VDF verification algorithm runs in total time polynomial in $\log(t)$ and κ. This algorithm should be much faster than the evaluation one.

Our PORR scheme uses VDFs in order to allow the verifier to detect a cheating cloud provider computing the puzzle solutions on the fly rather than keeping them at rest. Therefore, we aim to notice a difference between the response time of a honest cloud provider

and of a rational one. More precisely, a verifier should be able to observe a delay in responding from a cheating cloud provider. We choose a delay parameter t equal to 2^{16}.

The security of the VDF scheme based on exponentiation [12] relies on the assumption that the adversary cannot run a long pre-computation from the publication of the public parameters until the evaluation of the puzzles. Given the pre-processing parameter B, the VDF scheme is secure up to B puzzles. New public parameters must be generated once B puzzles have been used.

The authors in [12] suggest to set $B = 2^{30}$ for a reasonable trade-off between usability and security. Let us consider $B = 2^{30}$ and 64MB files. Up to 2000 files can be stored at the cloud with only one VDF instance. 2000 files roughly correspond to 128GB of data. Another VDF instance will be necessary if more files need to be stored. We recall that the setup algorithm runs in polynomial time in the security parameter κ, meaning that this algorithm is run easily and relatively fast.

We examine whether such condition on the pre-processing parameter B restricts the applicability of our PORR solution. In addition, we verify the costs incurred from computing and storing. Indeed, our results should convince us that our PORR solution fits real world requirements. We summarize our VDF implementation characteristics in Table **??**.

6.3 Evaluation and Discussion

Response generation. We measure the time that the cloud provider takes in order to generate its response, according to different numbers of challenges. Results are shown in Figure 4.

We consider two types of cloud providers: a rational cloud provider and a honest cloud provider. The former needs more time to generate its response since it requires to compute puzzle solutions on the fly. On the other side, since the honest cloud provider has already computed the puzzle solutions and stores them at rest, its response generation is noticeably faster. Therefore, we are guaranteed that the verifier (and thus the client) will notice a rational cloud server from a honest cloud provider since the time difference for response generation is consequent.

The average response time of an honest cloud provider is less than 2 seconds when $r = 2$ in [9], and a delay of almost 1 second is observed from a rational cloud provider. In our case, if the cloud provider tries to compute the 2 replicas on the fly, then it needs to compute the solutions y of 2 puzzles x, meaning that it runs in parallel time t with $poly(\log(t), \kappa)$ processors twice.

From Figure 4, we notice that a rational server needs almost twice the time than a honest server to generate its response, for a given number of challenges. For instance, for 40 challenges, a rational server generates its response in 0.63 seconds while a honest server in 0.31 seconds. Moreover, the response generation time increases with the number of challenges.

MIRROR's response generation depends on the bit size of the factors (that are elements needed to prepare the replicas) [9]. As long as the blind factors are greater or equal to 70 bits, the rational server should not gain any reasonable advantage in misbehaving. For instance, when the coefficients' size is equal to 70 bits, a rational server in MIRROR needs around 1 second to generate its response based on 40 challenges. Since we do not use blind factors, a rational server in our PORR only requires 0.63 seconds.

For a higher security level, that is a bigger blind factors' bit size (say 200 bits), a MIRROR rational server generates its response in roughly 2.8 seconds. In our case, in order to reach a higher security level, we can increase the number of challenges. For instance, for a number of challenges equal to 200, a rational server needs 4.86 seconds to generate its response. Such time cost happens since the rational server must calculate puzzle solutions on the fly.

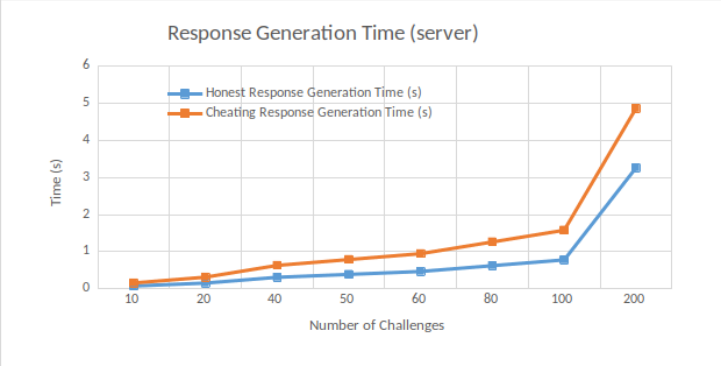

Figure 4: Response Generation Time

File preparation. We measure the time required to prepare a file before storing it on the cloud. Specifically, we measure the time for setup (i.e. key generation and tag generation) and replica generation (i.e. puzzle evaluation). Results are shown in Figure 5.

The time spent for preparing the file and its replicas increases exponentially with the size of the file. For instance, for a 16MB file and its 2 replicas, we need around 50 seconds while for a 256MB file, we require more than 700 seconds.

Moreover, file preparation takes around 180 seconds for a 64MB file and its 2 replicas in our case. We notice that setup and replica generation take around 500 and 700 seconds respectively in MIRROR [9], given the same file setting. Therefore, file preparation is almost 7 times faster in our PORR than in MIRROR for a 64MB file.

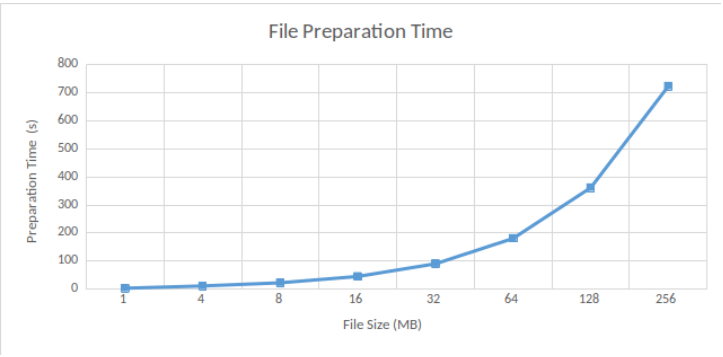

Figure 5: File Preparation Time

Verification. We measure the time required to verify correct storage at rest. Specifically, we measure the time for response generation and response verification. We consider multiple file sizes, and thus multiple numbers of challenges, since the latter depend

on the former. In particular, a large file means a bigger number of challenges. The number of challenges is calculated as the half of the file size (in MB). For example, 32 challenges are requested for a 64MB file. According to the default parameter settings, the verifier thus checks 128KB of data for each MB file data. Results are shown in Figure 6.

The verification time experienced by the client is 0.57 seconds in our case, while being 0.8 seconds in MIRROR [9]. The number of challenges increases with the file size, thus impacting the verification time. Given a file size of 128MB, the number of challenges is 64 and the verification time reaches 1.14 seconds. Note that the number of challenges in MIRROR is fixed at a value of 40, whatever the file size. Therefore, the verification time is constant, equal to 0.8 seconds.

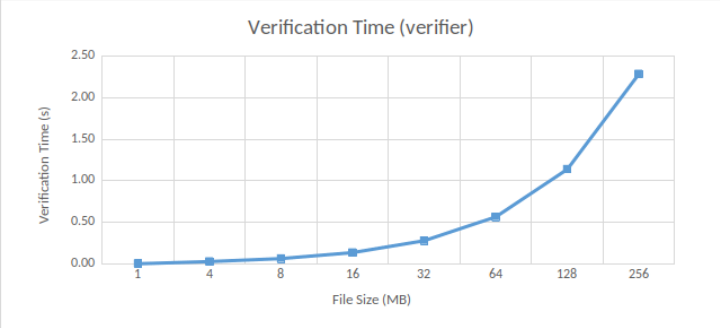

Figure 6: Verification Time

Puzzle evaluation. We measure the time spent for puzzle evaluation only. We are interested in such measurement since this part takes the longest time in file preparation. Results are shown in Figure 7.

Puzzle evaluations contribute to almost 3/4 of the file preparation. In the default configuration, that is given a 64MB file, file preparation time reaches 180 seconds, where around 130 seconds are spent on puzzle evaluations. Since the number of sectors increases with the file size, this leads to the linear growth of puzzle evaluation time.

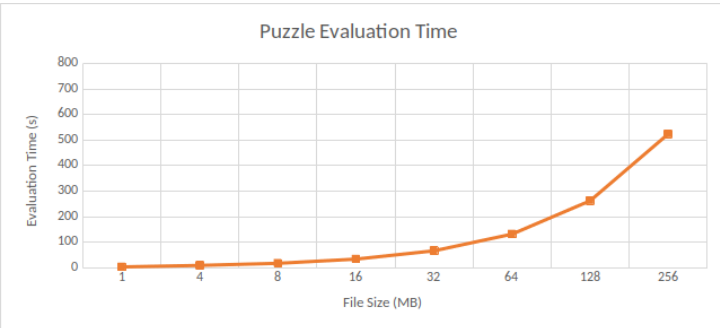

Figure 7: Puzzle Evaluation Time

Latency w.r.t. block sizes. We measure the time required for response generation and verification, according to the block sizes. We choose a sector size of 256 Bytes. Results are shown in Figure 8. We notice that a smaller block size gives a shorter time. Indeed, block size and challenge number are related, meaning that response generation and verification are faster as the block size decreases.

MIRROR [9] gets a similar trend to ours, but less obvious. In MIRROR, the verification time is around 1 second for a 1MB block size, and 1.1 seconds for a 2MB block size. In our case, the verification time is around 0.1 seconds for a 1MB block size, and 0.2 for a 2MB block size, thus an increase by 50%.

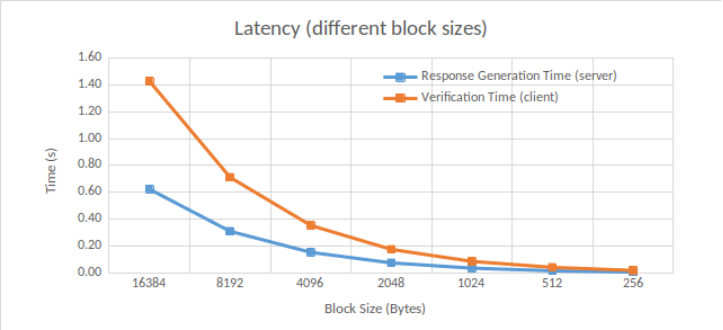

Figure 8: Latency (blocks)

Latency w.r.t. sector sizes. We measure the time required for response generation and verification, according to the sector sizes. We choose a block size of 8192 Bytes. For a given block size, a small sector size means a bigger number of sectors in that block. Therefore, the verification time is impacted since more data must be checked while the number of blocks and thus of challenges is fixed. Results are shown in Figure 9.

The authors in [9] do not explicitly analyze the effect of sector size of their protocol MIRROR. We expect that, for a given file, when the sector size decreases, its number increases, and thus impacts, with a rise, the server's response time and verifier's verification time.

Our default sector size is 256 Bytes, as a trade-off between efficiency and security. Verification is done in 0.36 seconds with such value. On the other size, a sector size equal to 32 Bytes results in getting a verification time 15 times slower.

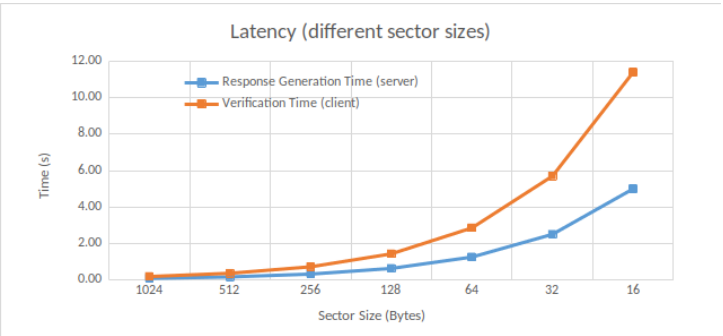

Figure 9: Latency (sectors)

Financial costs. We compare the financial costs between our PORR and MIRROR [9], w.r.t. file preparation. The replication process (as named in MIRROR) roughly corresponds to our file preparation. Results are shown in Figure 10.

The graph records those cost differences between the two protocols. The Amazon EC2 processor rental price is US$ 0.404 per hour. Such price is thus US$ 0.000112 per second, and has been multiplied with our file preparation time.

Given a file of size 384MB, our PORR and MIRROR will both incur a cost equal to 0.18 USD. Our PORR protocol seems more interesting from a financial point of view with file sizes smaller than 384 MB, saving from 2 to 15 times compared to MIRROR. On the other side, MIRROR seems more attractive with larger file sizes.

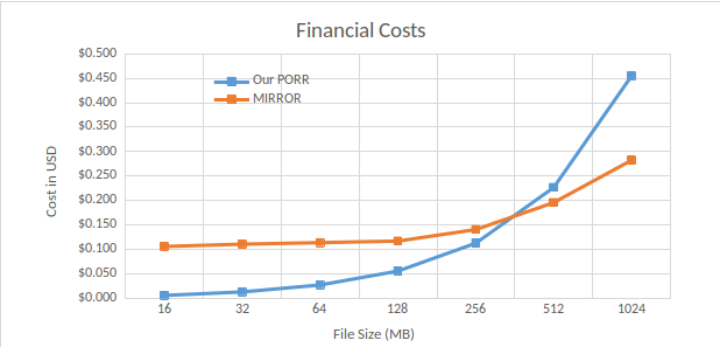

Figure 10: Financial Costs

Additional remarks. In MIRROR [9], replicas are computed as the product of the original file sector and two blind factors. However, the cloud provider may just keep the factors in reserve, and multiply by the sector when requested for verification. The gain may not be substantial and a cloud provider may just store the replicas.

Our replicas are computed as the addition of the original file sector and one blind factor. Similarly to [9], the cloud provider may or may not just keep the factor and add the sector when challenged. Future task will investigate whether a better design can prevent such behaviour from the cloud provider.

Summary. Our results show that our PORR prototype manages to trade expensive bandwidth resources with cheaper computing resources. Those results are likely to be well accepted by cloud providers and clients with the promise of better storage services and lower financial costs. Moreover, some of those results, especially on both technical and financial aspects, show that our PORR is more competitive with small file sizes, compared to MIRROR [9]. Therefore, our PORR solution may become one of the economically-viable, applicable systems that offer verifiable replicated cloud storage. In particular, our PORR fits the demands from indivudals and small businesses with lower file sizes to store on a cloud.

7 Conclusion

In this paper, we presented a recent PORR protocol, first proposed in [1], where one can check in a single instance that the cloud provider correctly stores an original file and its replicas. We combine our PORR with the slow function VDF to enable anyone to verify the cloud provider's behaviour, not only the file owner.

Implementation and evaluation of our solution have been carried out. Results show that such design is well applicable in realistic cloud environments. Multiple results, on technical and financial aspects, show that our PORR is more competitive with small file sizes, compared to MIRROR [9] and with the noticeable advantage of offering public verification.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment We thank the reviewers for their valuable comments.

References

[1] C. Gritti, "Publicly Verifiable Proofs of Data Replication and Retrievability for Cloud Storage," in 2020 International Computer Symposium (ICS), 431–436, 2020, doi:10.1109/ICS51289.2020.00091.

[2] A. Juels, B. S. Kaliski, Jr., "Pors: Proofs of Retrievability for Large Files," in Proceedings of the 14th ACM Conference on Computer and Communications Security, CCS '07, 584–597, ACM, New York, NY, USA, 2007, doi:10.1145/1315245.1315317.

[3] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, D. Song, "Provable Data Possession at Untrusted Stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, CCS '07, 598–609, 2007, doi:10.1145/1315245.1315318.

[4] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, D. Song, "Remote Data Checking Using Provable Data Possession," ACM Trans. Inf. Syst. Secur., **14**(1), 12:1–12:34, 2011, doi:10.1145/1952982.1952994.

[5] R. Curtmola, O. Khan, R. Burns, G. Ateniese, "MR-PDP: Multiple-Replica Provable Data Possession," in Proceedings of the 2008 The 28th International Conference on Distributed Computing Systems, ICDCS '08, 411–420, IEEE Computer Society, Washington, DC, USA, 2008, doi:10.1109/ICDCS.2008.68.

[6] A. F. Barsoum, M. A. Hasan, "Integrity Verification of Multiple Data Copies over Untrusted Cloud Servers," in Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (Ccgrid 2012), CCGRID '12, 829–834, IEEE Computer Society, USA, 2012, doi:10.1109/CCGrid.2012.55.

[7] A. Barsoum, M. Hasan, "Provable Multicopy Dynamic Data Possession in Cloud Computing Systems," IEEE Transactions on Information Forensics and Security, **10**, 485–497, 2015, doi:10.1109/TIFS.2014.2384391.

[8] G. S. Prasad, V. S. Gaikwad, "A Survey on User Awareness of Cloud Security," Int. J. of Engineering and Technology, **7**(2.32), 131–135, 2018, doi:10.14419/ijet.v7i2.32.15386.

[9] F. Armknecht, L. Barman, J.-M. Bohli, G. O. Karame, "Mirror: Enabling Proofs of Data Replication and Retrievability in the Cloud," in Proceedings of the 25th USENIX Conference on Security Symposium, SEC'16, 1051–1068, USENIX Association, USA, 2016.

[10] H. Shacham, B. Waters, "Compact Proofs of Retrievability," in Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology, ASIACRYPT '08, 90–107, Springer-Verlag, Berlin, Heidelberg, 2008, doi:10.1007/978-3-540-89255-7_7.

[11] H. Shacham, B. Waters, "Compact Proofs of Retrievability," Full version, 2008, https://hovav.net/ucsd/dist/verstore.pdf.

[12] D. Boneh, J. Bonneau, B. Bünz, B. Fisch, "Verifiable Delay Functions," CRYPTO 2018, 757–788, Springer-Verlag, Berlin, Heidelberg, 2018.

[13] G. Ateniese, R. Di Pietro, L. V. Mancini, G. Tsudik, "Scalable and Efficient Provable Data Possession," in Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, SecureComm '08, 9:1–9:10, ACM, New York, NY, USA, 2008, doi:10.1145/1460877.1460889.

[14] C. Wang, Q. Wang, K. Ren, W. Lou, "Privacy-preserving Public Auditing for Data Storage Security in Cloud Computing," in Proceedings of the 29th Conference on Information Communications, INFOCOM'10, 525–533, IEEE Press, Piscataway, NJ, USA, 2010.

[15] S. Yu, C. Wang, K. Ren, W. Lou, "Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing," in Proceedings of the 29th Conference on Information Communications, INFOCOM'10, 534–542, IEEE Press, Piscataway, NJ, USA, 2010.

[16] Z. Hao, S. Zhong, N. Yu, "A Privacy-Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability," IEEE Trans. on Knowl. and Data Eng., **23**(9), 1432–1437, 2011, doi:10.1109/TKDE.2011.62.

[17] Y. Yu, M. H. Au, Y. Mu, S. Tang, J. Ren, W. Susilo, L. Dong, "Enhanced privacy of a remote data integrity-checking protocol for secure cloud storage," International Journal of Information Security, 1–12, 2014, doi:10.1007/s10207-014-0263-8.

[18] C. Erway, A. Küpçü, C. Papamanthou, R. Tamassia, "Dynamic Provable Data Possession," in Proceedings of the 16th ACM Conference on Computer and Communications Security, CCS '09, 213–222, ACM, New York, NY, USA, 2009, doi:10.1145/1653662.1653688.

[19] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, S. S. Yau, "Dynamic Audit Services for Integrity Verification of Outsourced Storages in Clouds," in Proceedings of the 2011 ACM Symposium on Applied Computing, SAC '11, 1550–1557, ACM, New York, NY, USA, 2011, doi:10.1145/1982185.1982514.

[20] Y. Zhu, G.-J. Ahn, H. Hu, S. S. Yau, H. G. An, C.-J. Hu, "Dynamic Audit Services for Outsourced Storages in Clouds," IEEE Transactions on Services Computing, **6**(2), 227–238, 2013, doi:http://doi.ieeecomputersociety.org/10.1109/TSC.2011.51.

[21] C. Wang, Q. Wang, K. Ren, W. Lou, "Ensuring data storage security in cloud computing," in in Proc. of IWQoS'09, 2009.

[22] A. Le, A. Markopoulou, "NC-Audit: Auditing for Network Coding Storage," CoRR, **abs/1203.1730**, 2012.

[23] C. Wang, Q. Wang, K. Ren, N. Cao, W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Serv. Comput., **5**(2), 220–232, 2012, doi:10.1109/TSC.2011.24.

[24] B. Wang, B. Li, H. Li, "Oruta: privacy-preserving public auditing for shared data in the cloud," IEEE Transactions on Cloud Computing, **2**(1), 43–56, 2012, doi:http://doi.ieeecomputersociety.org/10.1109/TCC.2014.2299807.

[25] B. Wang, B. Li, H. Li, "Knox: Privacy-preserving Auditing for Shared Data with Large Groups in the Cloud," in Proceedings of the 10th International Conference on Applied Cryptography and Network Security, ACNS'12, 507–525, Springer-Verlag, Berlin, Heidelberg, 2012, doi:10.1007/978-3-642-31284-7_30.

[26] C. Wang, S. S. Chow, Q. Wang, K. Ren, W. Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," IEEE Transactions on Computers, **62**(2), 362–375, 2013, doi:http://doi.ieeecomputersociety.org/10.1109/TC.2011.245.

[27] X. Fan, G. Yang, Y. Mu, Y. Yu, "On Indistinguishability in Remote Data Integrity Checking," **58**(4), 823–830, 2015, doi:http://dx.doi.org/10.1093/comjnl/bxt137.

[28] B. Wang, B. Li, H. Li, "Panda: Public Auditing for Shared Data with Efficient User Revocation in the Cloud," IEEE T. Services Computing, **8**(1), 92–106, 2015, doi:10.1109/TSC.2013.2295611.

[29] C. Gritti, W. Susilo, T. Plantard, "Efficient Dynamic Provable Data Possession with Public Verifiability and Data Privacy," in E. Foo, D. Stebila, editors, Information Security and Privacy, 395–412, Springer International Publishing, Cham, 2015.

[30] C. Gritti, R. Chen, W. Susilo, T. Plantard, "Dynamic Provable Data Possession Protocols with Public Verifiability and Data Privacy," in J. K. Liu, P. Samarati, editors, Information Security Practice and Experience, 485–505, Springer International Publishing, Cham, 2017.

[31] J. Xu, E.-C. Chang, "Towards Efficient Proofs of Retrievability," in Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ASIACCS '12, 79–80, Association for Computing Machinery, New York, NY, USA, 2012, doi:10.1145/2414456.2414503.

[32] K. D. Bowers, A. Juels, A. Oprea, "Proofs of Retrievability: Theory and Implementation," in Proceedings of the 2009 ACM Workshop on Cloud Computing Security, CCSW '09, 43–54, ACM, New York, NY, USA, 2009, doi:10.1145/1655008.1655015.

[33] K. D. Bowers, A. Juels, A. Oprea, "HAIL: A High-availability and Integrity Layer for Cloud Storage," in Proceedings of the 16th ACM Conference on Computer and Communications Security, CCS '09, 187–198, ACM, New York, NY, USA, 2009, doi:10.1145/1653662.1653686.

[34] Y. Dodis, S. Vadhan, D. Wichs, "Proofs of Retrievability via Hardness Amplification," in Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography, TCC '09, 109–127, Springer-Verlag, Berlin, Heidelberg, 2009, doi:10.1007/978-3-642-00457-5_8.

[35] Q. Wang, C. Wang, J. Li, K. Ren, W. Lou, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing," in Proceedings of the 14th European Conference on Research in Computer Security, ESORICS'09, 355–370, Springer-Verlag, Berlin, Heidelberg, 2009.

[36] E. Shi, E. Stefanov, C. Papamanthou, "Practical Dynamic Proofs of Retrievability," in Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, CCS '13, 325–336, ACM, New York, NY, USA, 2013, doi:10.1145/2508859.2516669.

[37] B. Chen, R. Curtmola, "Towards Self-Repairing Replication-Based Storage Systems Using Untrusted Clouds," in Proceedings of the Third ACM Conference on Data and Application Security and Privacy, CODASPY '13, 377–388, Association for Computing Machinery, New York, NY, USA, 2013, doi:10.1145/2435349.2435402.

[38] B. Chen, R. Curtmola, "Remote data integrity checking with server-side repair," Journal of Computer Security, **25**(6), 537–584, 2017, doi:10.3233/JCS-16868.

[39] I. Leontiadis, R. Curtmola, "Secure Storage with Replication and Transparent Deduplication," in Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy, CODASPY '18, 13–23, Association for Computing Machinery, New York, NY, USA, 2018, doi:10.1145/3176258.3176315.

[40] M. Etemad, A. Küpçü, "Transparent, Distributed, and Replicated Dynamic Provable Data Possession," in M. J. J. Jr., M. E. Locasto, P. Mohassel, R. Safavi-Naini, editors, Applied Cryptography and Network Security - 11th International Conference, ACNS 2013, Banff, AB, Canada, June 25-28, 2013. Proceedings, volume 7954 of *Lecture Notes in Computer Science*, 1–18, Springer, 2013, doi:10.1007/978-3-642-38980-1_1.

[41] B. Chen, A. K. Ammula, R. Curtmola, "Towards Server-Side Repair for Erasure Coding-Based Distributed Storage Systems," in Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, CODASPY '15, 281–288, Association for Computing Machinery, New York, NY, USA, 2015, doi:10.1145/2699026.2699122.

[42] B. Chen, R. Curtmola, G. Ateniese, R. Burns, "Remote Data Checking for Network Coding-Based Distributed Storage Systems," in Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop, CCSW '10, 31–42, Association for Computing Machinery, New York, NY, USA, 2010, doi:10.1145/1866835.1866842.

[43] T. P. Thao, K. Omote, "ELAR: Extremely Lightweight Auditing and Repairing for Cloud Security," in Proceedings of the 32nd Annual Conference on Computer Security Applications, ACSAC '16, 40–51, Association for Computing Machinery, New York, NY, USA, 2016, doi:10.1145/2991079.2991082.

[44] K. D. Bowers, M. van Dijk, A. Juels, A. Oprea, R. L. Rivest, "How to tell if your cloud files are vulnerable to drive crashes," in Y. Chen, G. Danezis, V. Shmatikov, editors, Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS 2011, Chicago, Illinois, USA, October 17-21, 2011, 501–514, ACM, 2011, doi:10.1145/2046707.2046766.

[45] D. Vasilopoulos, K. Elkhiyaoui, R. Molva, M. Onen, "POROS: Proof of Data Reliability for Outsourced Storage," in Proceedings of the 6th International Workshop on Security in Cloud Computing, SCC '18, 27–37, Association for Computing Machinery, New York, NY, USA, 2018, doi: 10.1145/3201595.3201600.

[46] D. Vasilopoulos., M. Önen., R. Molva., "PORTOS: Proof of Data Reliability for Real-World Distributed Outsourced Storage," in Proceedings of the 16th International Joint Conference on e-Business and Telecommunications - Volume 2: SECRYPT,, 173–186, INSTICC, SciTePress, 2019, doi: 10.5220/0007927301730186.

[47] Z. N. J. Peterson, M. Gondree, R. Beverly, "A Position Paper on Data Sovereignty: The Importance of Geolocating Data in the Cloud," in Proceedings of the 3rd USENIX Conference on Hot Topics in Cloud Computing, HotCloud'11, 9, USENIX Association, USA, 2011.

[48] G. J. Watson, R. Safavi-Naini, M. Alimomeni, M. E. Locasto, S. Narayan, "LoSt: Location Based Storage," in Proceedings of the 2012 ACM Workshop on Cloud Computing Security Workshop, CCSW '12, 59–70, Association for Computing Machinery, New York, NY, USA, 2012, doi:10.1145/2381913.2381926.

[49] S. Dziembowski, S. Faust, V. Kolmogorov, K. Pietrzak, "Proofs of Space," in R. Gennaro, M. Robshaw, editors, Advances in Cryptology – CRYPTO 2015, 585–605, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[50] R. L. Rivest, A. Shamir, D. A. Wagner, "Time-Lock Puzzles and Timed-Release Crypto," Technical report, USA, 1996.

[51] J. . Cai, R. J. Lipton, R. Sedgewick, A. C. . Yao, "Towards uncheatable benchmarks," in [1993] Proceedings of the Eigth Annual Structure in Complexity Theory Conference, 2–11, 1993.

[52] D. Boneh, M. Naor, "Timed Commitments," in M. Bellare, editor, Advances in Cryptology — CRYPTO 2000, 236–254, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[53] C. Dwork, M. Naor, "Pricing via Processing or Combatting Junk Mail," in Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology, CRYPTO '92, 139–147, Springer-Verlag, Berlin, Heidelberg, 1992.

[54] A. K. Lenstra, B. Wesolowski, "A random zoo: sloth, unicorn, and trx," Cryptology ePrint Archive, Report 2015/366, 2015, https://eprint.iacr.org/2015/366.

[55] B. Cohen, K. Pietrzak, "Simple Proofs of Sequential Work," Cryptology ePrint Archive, Report 2018/183, 2018, https://eprint.iacr.org/2018/183.

[56] NetEm, "NetEm, the Linux Foundation," http://www.linuxfoundation.org/collaborate/workgroups/networking/netem, 2021, accessed on 11/04/2021.

[57] T. Flach, N. Dukkipati, A. Terzis, B. Raghavan, N. Cardwell, Y. Cheng, A. Jain, S. Hao, E. Katz-Bassett, R. Govindan, "Reducing Web Latency: The Virtue of Gentle Aggression," in Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM, SIGCOMM '13, 159–170, Association for Computing Machinery, New York, NY, USA, 2013, doi:10.1145/2486001.2486014.

[58] M. Dahlin, B. B. V. Chandra, L. Gao, A. Nayate, "End-to-End WAN Service Availability," IEEE/ACM Trans. Netw., **11**(2), 300–313, 2003, doi: 10.1109/TNET.2003.810312.

[59] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, A. Pescapè, "Broadband Internet Performance: A View from the Gateway," SIGCOMM Comput. Commun. Rev., **41**(4), 134–145, 2011, doi: 10.1145/2043164.2018452.

[60] D. Dobre, G. Karame, W. Li, M. Majuntke, N. Suri, M. Vukolić, "PoWerStore: Proofs of Writing for Efficient and Robust Storage," in Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS '13, 285–298, Association for Computing Machinery, New York, NY, USA, 2013, doi:10.1145/2508859.2516750.

[61] D. Zhuo, M. Ghobadi, R. Mahajan, K.-T. Förster, A. Krishnamurthy, T. Anderson, "Understanding and Mitigating Packet Corruption in Data Center Networks," SIGCOMM '17, 362–375, Association for Computing Machinery, New York, NY, USA, 2017, doi:10.1145/3098822.3098849.

[62] V. Paxson, "End-to-End Internet Packet Dynamics," SIGCOMM Comput. Commun. Rev., **27**(4), 139–152, 1997, doi:10.1145/263109.263155.