JOURNAL OF ENGINEERING RESEARCH & SCIENCES

Special Issue

Computing, Engineering and Sciences

November 2023

www.jenrs.com
ISSN: 2831-4085

JENRS

EDITORIAL BOARD
(Special Issue)

Guest Editor

Prof. Paul Andrew
Department of Electrical Engineering, Universidade De São Paulo, Brazil

Editorial

The accelerating integration of computing technologies with engineering practices and scientific research has profoundly transformed the landscape of innovation in the 21st century. It is with great privilege that we present this special issue of the *Journal of Engineering Research and Sciences*, dedicated to the theme of *Computing, Engineering and Sciences*. This issue brings together a diverse array of contributions that collectively highlight the critical role of computational approaches in advancing engineering solutions and deepening scientific understanding.

The synergy between computing and engineering has become a fundamental driver of progress across multiple domains. Modern engineering challenges ranging from infrastructure development and energy systems to biomedical applications are increasingly reliant on computational tools for design, analysis, and optimization. Similarly, scientific discovery has been greatly enhanced by the ability to process vast datasets, simulate complex phenomena, and model intricate systems with remarkable precision. The research articles featured in this issue reflect this growing interdependence, showcasing how computing serves as a unifying framework that bridges theory, experimentation, and application.

A central theme emerging from this collection is the transformative impact of data-driven methodologies. Advances in artificial intelligence, machine learning, and data analytics are enabling researchers to extract meaningful insights from complex and often unstructured datasets. These capabilities are being applied to improve predictive modeling, enhance system performance, and support evidence-based decision-making. The studies presented in this issue demonstrate how data-centric approaches are not only improving efficiency but also opening new avenues for innovation across engineering and scientific disciplines.

In addition, the issue highlights the importance of simulation and computational modeling as indispensable tools in modern research. Through high-fidelity simulations, researchers are able to explore scenarios that may be impractical or impossible to investigate experimentally. This capability is particularly valuable in fields such as fluid dynamics, materials science, environmental modeling, and structural analysis. By integrating simulation with empirical validation, the contributions in this issue exemplify how computational techniques can accelerate development cycles and reduce uncertainty in complex systems.

The advancement of intelligent and automated systems also features prominently in this special issue. The integration of embedded systems, robotics, and smart technologies is reshaping industries and enabling more responsive and adaptive solutions. From smart manufacturing and autonomous systems to healthcare technologies and urban infrastructure, the research presented here illustrates the far-reaching implications of intelligent systems in enhancing productivity, safety, and sustainability.

Equally important is the emphasis on interdisciplinary collaboration, which underpins many of the contributions in this issue. The intersection of computing, engineering, and sciences fosters a collaborative environment where diverse expertise converges to address multifaceted challenges. Such collaboration not only enriches the research process but also leads to more robust and comprehensive solutions that are capable of meeting the demands of an increasingly complex world.

The editorial team expresses its sincere gratitude to all authors for their valuable contributions and to the reviewers for their rigorous and thoughtful evaluations. Their collective efforts have ensured the high quality and relevance of the research presented in this special issue, reinforcing the journal's commitment to academic excellence.

As we present this issue, we hope it will serve as both a reflection of current advancements and a catalyst for future innovation. The continued integration of computing with engineering and scientific research holds immense potential to address global challenges and drive sustainable development. We encourage readers to engage with the work presented herein and to contribute to the ongoing advancement of this dynamic and rapidly evolving field.

Guest Editor

Prof. Paul Andrew

JOURNAL OF ENGINEERING RESEARCH AND SCIENCES

Special Issue **November 2023**

CONTENTS

Received: 06 September 2023, Revised: 25 November 2023, Accepted: 26 November 2023, Online: 30 November 2023

DOI: https://doi.org/10.55708/js0211001

Neural Synchrony-Based State Representation in Liquid State Machines, an Exploratory Study

Nicolas Pajot, Mounir Boukadoum*

Department of Computer Science, University of Quebec at Montreal, Quebec, Canada
*Corresponding author: Mounir Boukadoum, boukadoum.mounir@uqam.ca

ABSTRACT: Solving classification problems by Liquid State Machines (LSM) usually ignores the influence of the liquid state representation on performance, leaving the role to the reader circuit. In most studies, the decoding of the internally generated neural states is performed on spike rate-based vector representations. This approach occults the interspike timing, a central aspect of biological neural coding, with potentially detrimental consequences on the LSM performance. In this work, we propose a model of liquid state representation that builds the feature vectors from temporal information extracted from the spike trains, hence using spike synchrony instead of rate. Using pairs of Poisson-distributed spike trains in noisy conditions, we show that such model outperforms a rate-only model in distinguishing two spike trains regardless of the sampling frequency of the liquid states or the noise level. In the same vein, we suggest a synchrony-based measure of the separation property (SP), a core feature of LSMs regarding classification performance, for a more robust and biologically plausible interpretation.

KEYWORDS: Liquid state machine, state representation, temporal decoding, separation property, classification

1. Introduction

Liquid State Machines (LSMs) are generic models of computation inspired by biological cortical circuits. As such, they are well suited for real-time, online and anytime computations, and they can under certain circumstances exhibit unlimited computational power [1]. One important aspect of LSMs is their conceptual simplicity and relative ease of implementation in comparison to multilayer networks with error backpropagation training. LSMs have been tested with relative success in a variety of experimental contexts: identification of spoken words, voice, phonemes [2], [3], [4] and musical instruments [5], [6], robotics [7], movement prediction [8]; classification of musical styles [9], seismic data for military vehicles [10], nuclear stockpile data [3]; recognition of signature counterfeits [11], and even the study of biological neurons [12]. However, the LSM performance for classification tasks is variable [13], [14], [15], [16], potentially due to the randomly connected and untrained liquid regardless of application. Several authors have investigated liquid optimization approaches to raise the performance of LSMs, including Genetic Algorithms [13], [17], Separation Driven Synaptic Modification [18], Reinforcement Learning [17], Particle Swarm Optimization [19], and a number of different learning methods for temporal representations of artificial spiking neural networks (statistics, Hebbian learning, gradient estimation, linear algebra formalisms, etc.) that are reviewed in [20].

Still, the research on LSMs rarely focuses on the influence of the liquid state representations on performance. In most studies, the decoding of the discrete input spike trains relies on rate-based feature vectors used as inputs to a classifier. The rate coding typically counts the number of spikes in arbitrary time bins and filters the results with an exponential kernel due to its shape resemblance to the postsynaptic currents in biological neurons [21]. This ignores the spike timing, with potentially detrimental consequences on LSM classification performance. In this work, we propose a model of the liquid states that is based

on temporal information extracted from the input spike trains, aiming to improve classification performance without increasing the liquid's dimensionality. We show that this model outperforms a rate-based model at classifying Poisson-distributed spike trains in noisy conditions regardless of the sampling frequency of the liquid states. We therefore suggest a synchrony-based measure of the Separation Property (SP) of LSMs, for a more robust and biologically plausible interpretation.

This paper is divided as follows: the next section provides a brief description of the LSM model, with examples of use in classification problems. In section III, we focus on the common rate-based representation of liquid states to show that it leaves apart critical temporal information about spike trains. We also review the known measures of SP and underline the absence of synchrony-based methods to quantify it. Section IV proposes a novel representation of the liquid state based on spike metrics, as well as a composite state model that incorporates both rate and temporal representations in a composite feature vector. We also describe the methodology used to test performance hypotheses about synchrony in the context of classifying Poisson-distributed input spike trains, as well as the correlation of performance with SP measures. We present the simulation results in Section V before a discussion and conclusions.

2. The Liquid State Machine Model (LSM)

Figure 1 summarizes the processing steps in the LSM model, showing the input signal encoding (a, b), the spatio-temporal propagation within the liquid (c), the liquid state vector coding (d) and the interpretation by a readout mechanism (e). The core of the LSM is the neural liquid, or microcircuit, which consists in a grid of interconnected artificial spiking neurons, usually in 3-dimensional space by analogy to biological cortical columns (see Figure 1.c). The neurons occupy the nodes of this structure and are usually connected by randomly generated synapses. Due to its inputs and recurrent connections, the liquid forms a dynamic system endowed with the memory of previous states [4], [22], [23]. As a result, it continuously projects input signals onto a high-dimensional space. This mapping fosters the emergence of spatiotemporal patterns (trajectories, or "*time-varying changes in the active state*" [24], p. 114) that may be identified by simple, memoryless readouts [13], [3], as long as two major constraints are enforced: the Approximation Property (AP) and the Separation Properties (SP); AP guarantees that the readout can approximate any function of the liquid states to an arbitrary level of accuracy, whereas SP ensures that trajectories produced in the liquid by different input stimuli are well differentiated.

The readout maps liquid states to meaningful outputs; in the case of classification, it projects state vectors onto classes. Similar to Support Vector Machines, the capability to separate complex trajectories by linear discriminators is guaranteed by the fact that the "*the dimension of the state space exceeds the 'complexity' of the trajectories*" [24].

The readout is the only LSM element that requires training, which may confer an advantage to the model over alternative architectures, since training a recurrent network of artificial spiking neurons can be a hard problem. However, the way the liquid states are encoded to serve as input feature vectors to the readout may play a role in achieving peak performance. This subject seems to have been largely ignored by the research community and most studies represent the liquid states by the same approach:

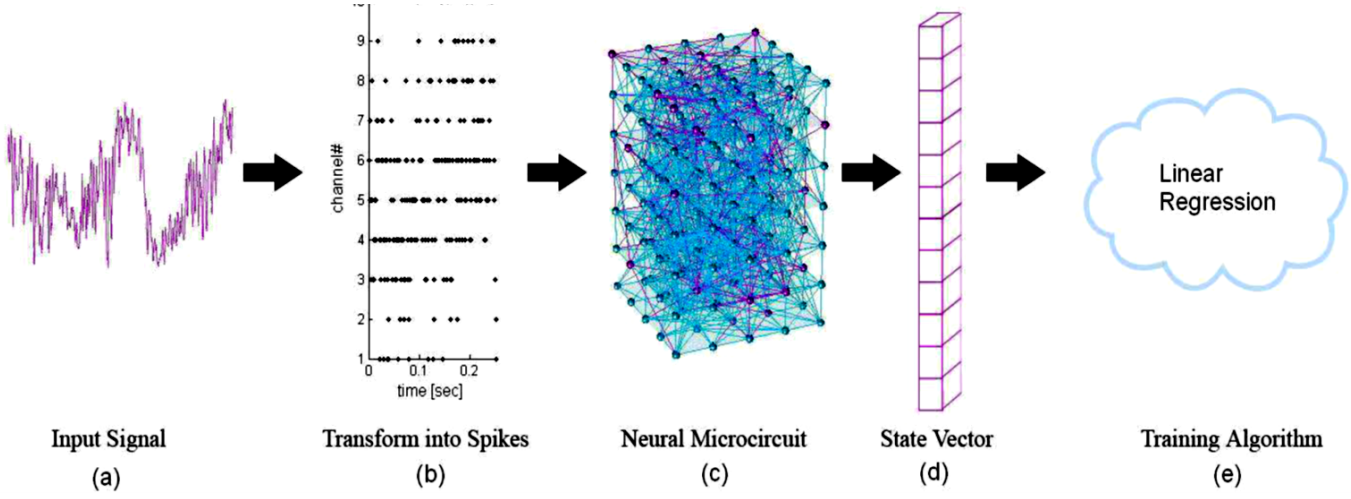


Figure 1: Typical Liquid State Machine model

measuring the spike rate in preset time intervals. Therefore, any potential phase information benefit is lost.

2.1. Models of Liquid States

Two broad approaches to represent the liquid states can be seen in the literature: sampling of analog signals [25], [10], such as the postsynaptic potential or the neural membrane voltage [26] and spike train decoding [23], [27], [28], [29], [8], [24]. T the typical representation in the latter is a state matrix obtained by filtering the discrete spike trains generated within the liquid after exposure to a stimulus ([3], [13], [30], [27], [10]). The filtering step is typically performed by convolution with an exponential kernel due to its resemblance to the shape of postsynaptic currents [21]. As a result, the spike firings within a liquid composed of N neurons are converted into an N-dimensional continuous signal [24], and each element of the state matrix stores the sampled values of this signal for a given observation interval, sampling rate, and neuron in the liquid. This representation is in effect a "rate code" (whereas small bin sizes can turn the representation into a "coincidence detector" [21]).

Although widely used to represent the liquid states, rate-based feature vectors miss a key aspect of the neural code: the actual timing of spike emissions, hence neglecting phase information [20], [31], [32]. On the other hand, it is now accepted that neural coding cannot be fully understood by only examining the rate of spike firing [33], [18], [34], [35], [36], [32], [31], and that the spike timing also encodes information. Hence, the synchrony between spike trains may be a cornerstone for understanding neural codes, as "*temporal codes employ those features of the spiking activity that cannot be described by the firing rate.*" [32]

The temporal decoding of liquid states has already been suggested before. In [29], the authors underline the potential power of the temporal relationships between the spike trains in the liquid, while [3] advocates the use of readouts that incorporate spike timings, but no study has examined the respective efficiency of synchrony-based and rate-based approaches. Similarly, SP models that consider the synchrony between spike trains in the liquid are rare [18].

2.2. Separation Property

The Separation Property evaluates the amount of "*separation between the trajectories of internal states*" [23] that are triggered in the liquid by two different input stimuli. The more separation, the easier it is for a readout endowed with the approximation property to distinguish between two different state trajectories in the liquid. This macroscopic property of the liquid can thus contribute to LSM classification performance.

While SP is widely regarded as a crucial predictor of performance, there is little consensus on how to measure it. The literature reveals different views, including statistical methods [13], [3], [37], and linear algebra formalisms [14] or vector distances between filtered firing rates [23], [1], [37]. For instance, Maass [23], [1] expresses SP as the Euclidean distance between the filtered state vectors of each neuron (Gaussian kernel), Dockendorf [37] uses both the Van Rossum [38] metric – which exploits the notion of distance between filtered states – and a custom measure based on the cross-correlation of spike times, Legenstein and Maass [14] link SP to the number of linearly independent variables in the state matrix, suggesting that the rank of the matrix is a good measure of separation, and Goodman and Ventura [3] and Hourdakis and Trahanias [13] use statistical methods to measure SP, with the former relying on centroids and the latter on Fisher's Discriminant Ratio (FDR). In the following section, we describe our synchrony matrix-based approach to liquid state representation, the methodology used to investigate the ensuing effect on classification performance, and we underline the relationship between the Separation property and LSM performance.

3. Classification with temporal liquid state representations

The proposed liquid state representation is based on the synchrony level between spike trains during a given time window, as quantified by metrics that evaluate the temporal similarity between the spike trains emitted by neuron pairs in the liquid. Thus, the metrics operate on spike timings rather than counts and, as stated in [39], p. 146, "*If a spike metric leads to a high-fidelity representation, then the temporal features that it captures are candidates for neural codes.*"

3.1. Synchrony matrix representation of liquid states

The synchrony matrix is constructed thanks to the Adaptive Spike Distance (ADS) metric ([40], [41], [42], [43], [44], [45]), although any other bivariate spike metric may be employed. We chose ADS for its sensitivity to spikes coincidence and the fact that it does not rely on a time-scale parameter.

Given two spike trains #1 and #2 of duration T, ADS is calculated by averaging their instantaneous "dissimilarity profiles", which measure how coincident the two spike trains are at any point in time. We have:

$$Ds = \frac{1}{T}\int_{t=0}^{T} S(t)dt \quad (1)$$

where *Ds* quantifies the overall dissimilarity between the two spike trains over *T*, and *S*(*t*) provides a joint measure of their instantaneous dissimilarity profiles at each time *t*. *S*(*t*) is defined by:

$$S(t) = \frac{S_1(t)x_{ISI}{}^{(2)}(t) + S_2(t)x_{ISI}{}^{(1)}(t)}{2\langle x_{ISI}{}^{(n)}(t)\rangle_n^2} \quad (2)$$

where $x^{(i)}{}_{isi}(t)$ stands for the instantaneous interspike interval of spike train #*i*, $\langle x_{ISI}{}^{(n)}(t)\rangle_n$ is the mean interspike interval for both spike trains, and $S_1(t)$ and $S_2(t)$ are given by:

$$S_i(t) = \frac{\Delta t_P{}^{(i)}(t)x_F{}^{(i)}(t) + \Delta t_F{}^{(i)}(t)x_P{}^{(i)}(t)}{x_{ISI}{}^{(i)}(t)}, \quad i = 1,2 \quad (3)$$

In the previous equation, for spike train #*i*, $x^{(i)}{}_P(t)$ and $x^{(i)}{}_F(t)$ are the time latency to the closest previous spike and closest following spike at time *t*, respectively, and $\Delta t_P{}^{(i)}$ and $\Delta t_F{}^{(i)}$ are the same latency from one of these spikes to the nearest one in the other spike train. For example, $\Delta t_P{}^{(1)}$ is defined as:

$$\Delta t_P{}^{(1)}(t) = min(|t_P{}^{(1)}(t) - t_i{}^{(2)}|) \quad (4)$$

where $t_P{}^{(1)}(t)$ is the time of occurrence of the closest previous spike in spike train #1 at time *t*, and $t_i{}^{(2)}$ is the time of occurrence of the *i*th spike in spike train #2. A more comprehensive description of ADS with graphical illustrations can be found in [46]. Table 1 provides the example of a synchrony matrix for a liquid made of 4 neurons, where *d*(*a*, *b*) evaluates the dissimilarity between the spike trains of neurons *a* and b as with equation (1).

Table 1: Example of a synchrony matrix for a 4-neuron LSM.

Neuron	**1**	**2**	**3**	**4**
1	0	*d*(1,2)	*d*(1,3)	*d*(1,4)
2	*d*(2,1)	0	*d*(2,3)	*d*(2,4)
3	*d*(3,1)	*d*(3,2)	0	*d*(3,4)
4	*d*(4,1)	*d*(4,2)	*d*(4,3)	0

Given the symmetry of the synchrony matrix and its zero diagonal, the final representation may consist only in the lower triangular matrix expressed as a vector. For a liquid composed of *N* neurons, the size *E* of this vector is:

$$E = \sum_{n=1}^{N} n = \frac{(1+N)N}{2} - N \quad (5)$$

3.2. Composite-state vector

Theoretically, rate and synchrony represent complementary information, since they encode two aspects of the spiking within a liquid (see [39], p.148, for an *in vivo* example). Therefore, taking inspiration from [10], we can enhance the representation of liquid states by combining filtered rates and synchrony information. Then, the size of the features vector extracted from the liquid would increase by *N*. For example, *N*=8 would lead to a vector of 8 rate elements (one for each neuron) and 28 synchrony elements (from equation 5). We expect this composite state representation to lead to better classification results than by only using rate-based or synchrony-based representations.

A. SP quantification with spike metrics

Using the hypothesis that the spike trains generated for different classes of input signals are significantly dissimilar (i.e., distant or unsynchronized), SP expresses the average spike train dissimilarities in the liquid. To build this synthetic measure out of spike distance metrics, we proceed similarly to getting the Fisher's discriminant ratio (FDR) of a 2-class classification problem:

1) For each stimulus belonging to one of the classes, we build the list of corresponding liquid states, composed of the spike trains of each neuron in the liquid for the duration of the experiment.
2) we then sample pairs of liquid states by including one from one class and one from the other.
3) we measure the dissimilarity for each sampled pair with a distance metric. The separation measure is the mean of the obtained results.

In this paper, we test cost-based measures (Victor-Purpura distance [47], [48], vector embedding (Van Rossum distance [38], Schreiber Similarity [49], Hunter-Milton Reliability [50]), scale-free measures (Spike Synchronization, ISI Distance, Spike Distance [51], [52], [40], [41], [42], [43], [44], [45] and statistical methods (Jolivet Coincidence [53]. The methodology for verifying the efficiency of these new approaches in described next.

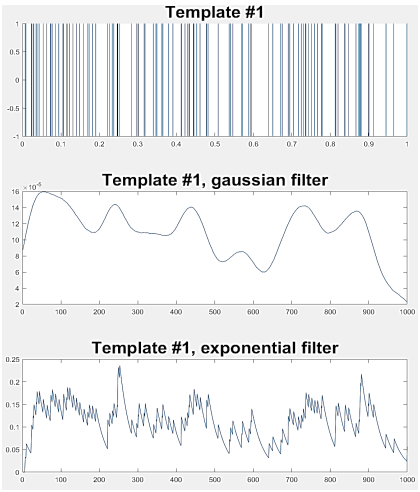
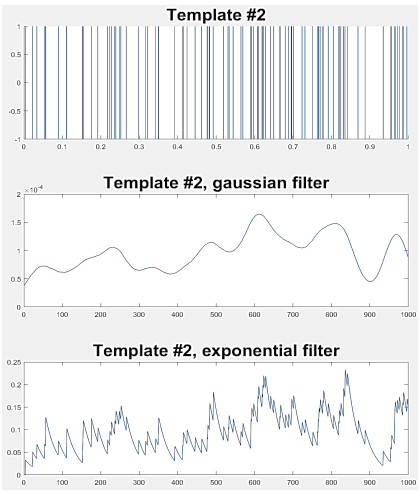
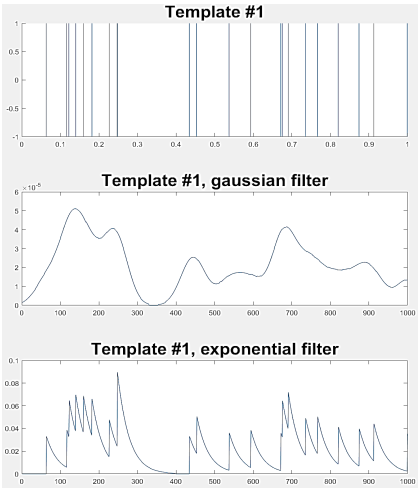
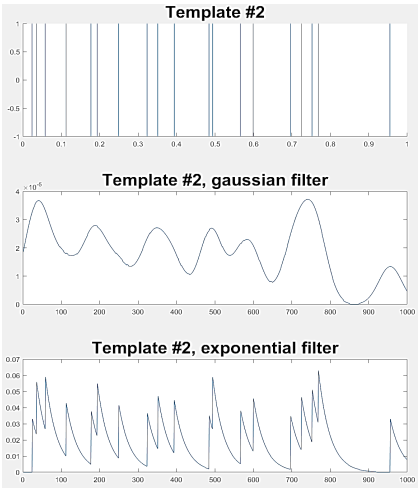
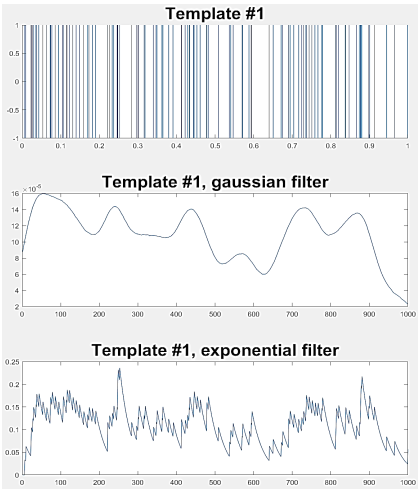
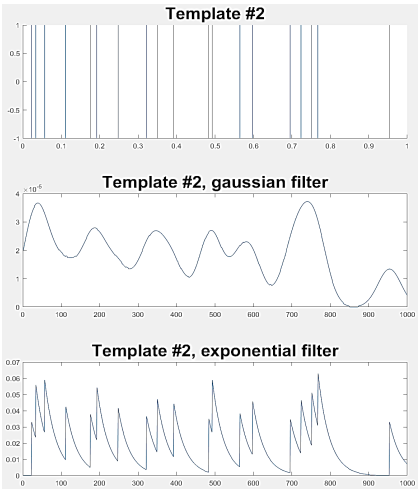
4. Methodology of Testing

To test the performance of each model, we aggregate and compare the error rates of randomly generated LSMs at classifying Poisson-distributed spike trains, using rate-based, synchrony-based and composite liquid state representations. As a corollary, we also quantify how the synchrony-based SP measures correlate with the error rate.

4.1. Experimental setup

Three hundred random liquids are generated and fed with as many jittered versions of two template spike trains. These input signals are generated by adding temporal noise ("jitter") to each template. Hence, each spike of the template is randomly shifted in time by an amount drawn from a uniform distribution with mean 0 and standard deviation equal to the desired jitter level. The values used in this paper

Table 2: Classification experiments and corresponding spike train templates

Experiment	First spike train before and after filtering	Second spike train before and after filtering
1: Two high-frequency spike trains ($\lambda_1 = \lambda_2 = 100$)		
2: Two low-frequency spike trains ($\lambda_1 = \lambda_2 = 20$)		
3: One high-frequency spike train ($\lambda_1 = 100$) and one low-frequency spike train ($\lambda_2 = 20$)		

(1 ms, 4 ms, 10 ms) are loosely inspired by [23] who used 4 ms and 8 ms in their "high jitter" experimental contexts.

the original spike train pair, of which 200 are used to train the readout and 100 to collect the testing error.

The spike output from the LSM is recorded, sampled, and processed in three different ways to represent the liquid states:

1. Sampling and exponential filtering by convolving the spike trains with an exponential kernel of width 0.3 (rate coding);
2. Sampling and calculating the synchrony matrix (synchrony coding);

Aggregating the two previous representations into a composite one (composite coding).

Three different experiments are conducted to evaluate the LSM classification performance, each one involving a pair of random Poisson distributed spike trains. Table 2 indicates the ƛ value of each train (taken from [40]) with example realizations before and after filtering for rate coding. Each experiment considers 300 jittered variations of

4.2. Neural Microcircuit

The liquid is organized as a 2x2x2 column (two layers of 2x2 neural grids), composed of 80% excitatory and 20% inhibitory Leaky Integrate and Fire (LIF) neurons as in [54], with dynamic synapses. A single input neuron connects to the liquid pool, and static synapses are set randomly based on the physical distance between pairs of neurons. The probability of establishing a connection is as follows [30]:

$$p = Ce^{-\left(\frac{D(a,b)}{\lambda}\right)^2} \quad (6)$$

where λ is a connection parameter that controls both the average number of connections and the average distance between connected neurons [45], $D(a, b)$ is the Euclidean distance between two neurons a and b, and C is a scaling parameter influenced by the excitatory or inhibitory effect of the connected neurons. In this study, λ was set to 2 so that any pair of neurons in the liquid column might be connected, and C was set to different values taken from previous studies ([23], [30]) and based on measures taken from cortical brain areas [30]. By default, they are set at 0.3 (for a connection between a pair of Excitatory-Excitatory neurons), 0.2 (Excitatory-Inhibitory), 0.4 (Inhibitory-Excitatory), and 0.1 (Inhibitory-Inhibitory). All other parameters are left untouched.

The chosen dimensions for the liquid allow for a large range of performance levels throughout the tests, depending only on the neuron type and connection topology of the liquids. We can easily generate drastically different liquid configurations, which in turn endow each LSM with largely different performance levels. An example of a 2x2x2 microcircuit is presented in Figure 2, where the neurons shown in cyan are excitatory, and those in magenta are inhibitory. The input neuron is the one located at position (0,0,0).

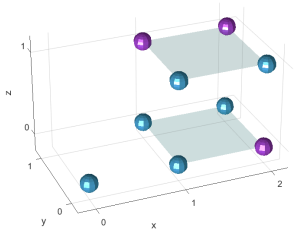


Figure 2: 2x2x2 neural microcircuit

The neural dynamics are set by the following membrane voltage equation (from [55] and [56]):

$$\tau_m \frac{dV_m}{dt} = -\left(V_m - V_{resting}\right) + R_m\left(I_{syn}(t) + I_{inject} + I_{noise}\right) \quad (7)$$

where τ_m represents the membrane time constant, V_m the membrane potential, $V_{resting}$ the resting membrane potential, R_m the input resistance, I_{syn} the current supplied by the synapses (also called "post-synaptic current" or PSC), I_{inject} an optional background current and I_{noise} is a Gaussian random variable with mean 0 and a given variance. At time t = 0, V_m is set to a default voltage V_{init}. During the course of the simulation, a spike is emitted if V_m exceeds a threshold V_{thresh}: the membrane potential is then reset to a given voltage V_{reset} and remains at this level for the duration of the absolute refractory period $T_{refract}$. Table 3 summarizes the values of these parameters during simulations. They are the default values of the simulation tool (see simulation tool subsection below).

Table 3: default parameters of the neural membrane model

Parameter	Description	Value
Cm	Membrane capacitance	3e-08 F
Rm	Membrane resistance	1e+06 Ω
Vthresh	Spike threshold	-0.045 V
Vresting	Membrane voltage at rest	-0.06 V
Vinit	Initial voltage condition	-0.06 V
Vreset	Post-spike voltage	-0.06 V
Trefract	Maximum refraction period	0.003 s
Inoise	Standard deviation of added noise	0 A
Iinject	Injected current	0 A

The amplitude of the postsynaptic current (I_{syn}) depends on previous spike activity, which constitutes a form of short-term plasticity (the model is described in [57]). The total post-synaptic current going into a neuron j from all the neurons i connected to it after a spike is described by the following equation:

$$I_{syn}(t) = \sum_i I_{ij}(t) \quad (8)$$

where

$$I_{ij}(t) = w_{ij}.exp(\frac{-t}{\tau_{syn}}) \quad (9)$$

and I_{ij}(t) is the post-synaptic current flowing from neuron i and j; w_{ij} is the synaptic strength of the connection, and τ_{syn} is the synaptic time constant.

Contrary to static synapses, w_{ij} varies with previous spike trains for dynamic synapses as follows [57]:

$$w_{ij} = W.r_n.u_n \quad (10)$$

where W is the absolute synaptic efficiency (or weight); r_n and u_n quantify the short-term depressing and facilitating effects after spike n has been fired. The two variables are described by the following two equations [57]:

$$r_{n+1} = r_n(1 - u_{n+1}).\exp\left(\frac{-\delta t}{\tau_{rec}}\right) + 1 - \exp\left(\frac{-\delta t}{\tau_{rec}}\right) \quad (11)$$

and

$$u_{n+1} = u_n.\exp\left(\frac{-\delta t}{\tau_{facil}}\right) + U(1 - u_n.exp\left(\frac{-\delta t}{\tau_{facil}}\right)) \quad (12)$$

As indicated in [45], the absolute synaptic weights for a given connection are drawn from a gamma distribution with mean W and standard deviation $(W.SHW)$. δt is the elapsed time between spike n and n+1; τ_{facil} and τ_{rec} are the time constants of the facilitating and depressing plasticity effects respectively. U is a constant describing the fraction of the absolute synaptic efficiency used. The values of these parameters are indicated in Table 4.

2. *Sampling, memory and decoding liquid states* – The liquid states must be sampled in time in order to construct the input state vectors to the readout, with the sampling time window having an impact on classification performance. Although LSMs are known to possess "*the capability to hold and integrate information from past input segments over several hundred ms*" ([40], p.8), shorter sampling intervals typically provide less spike information. The scarcity of spike information in these samples can make classification harder, for an increased classification error rate. Thus, the memory span capability of LSMs plays a crucial role in providing liquid state samples with information from the past, effectively enriching them through the temporal integration of input stimuli: any state of the liquid will then keep some memory of previous states, which allows the readout to be memoryless.

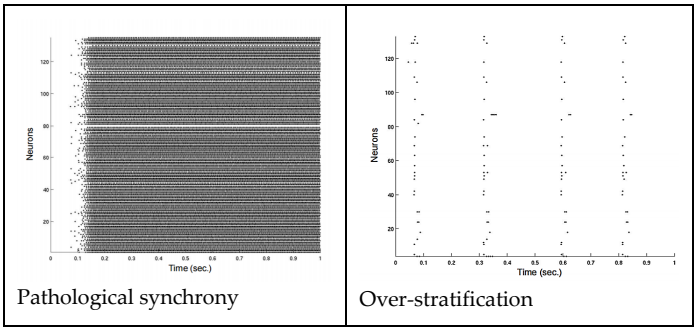


Figure 3. Example of two pathological liquids (from [46])

The role of memory in the classification performance of LSMs is relatively under-studied, but two extreme spike train patterns can emerge from pathological liquids: over-stratification ([15], [3]) and pathological synchrony [15]. The former happens when spikes are not propagated for a long enough time, resulting in a lack of memory capacity; the latter is the result of infinite feedback loops within the neurons, effectively spiking in synchrony and obfuscating the "real", important states. Figure 3 shows an example of each case.

Another important parameter to consider is *Tau*, the synaptic time constant controlling the time required for the postsynaptic response to fade to zero after being injected with current. Figure 4 illustrates the examples of an input spike train, the postsynaptic response and the resulting spike train for three different values of *Tau*.

In this work, *Tau* is set to 0.25, a value that increases the overall performance across all tests and all liquid state representations while avoiding temporal stratification and pathological synchrony. The choice of sampling rates (10 Hz, 4 Hz and 2 Hz) was inspired by similar values used in the literature [3], [2], [8], [29], [10].

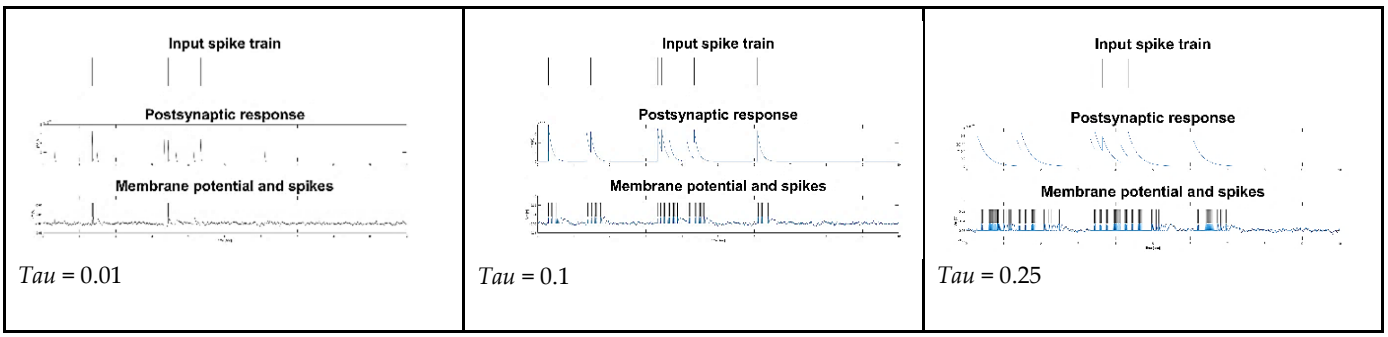


Figure 4. Illustration of the effect of different values of the synaptic time constant on spike trains.

4.3. Readout

A simple linear regression was chosen to avoid the potential impact of more advanced techniques such as multi-layer perceptrons, Support Vector Machines, etc., which may provide hard to interpret results due to their own capabilities. Thus, the predicted class value $\hat{y}_i$ of a feature state vector $\boldsymbol{x_i}$ is given by:

$$\hat{y}_i = \boldsymbol{m} \,.\, \boldsymbol{x_i} + b \tag{13}$$

where m, the slope of the separating hyperplane and b, the intercept, are estimated using the least squares method. The output of this simple readout is passed through a step function for binary classification: for a given value n returned by the regression, the class $y[n]$ is given by the following equation:

$$y[n] \; = \; \{-1, n <= 0;\; 1, n > 0\} \tag{14}$$

4.4. Simulation tool

We use CSIM (Circuit SIMulator, [58]), a neural network simulator that can handle LSM models with different neuron and synapse models. CSIM is built in C++ for performance considerations and provides an easy-to-use MATLAB interface. This tool uses the exponential Euler method of numerical integration, with a default time step for the update of 0.1 ms. A thorough description and a comparison to other simulators are provided in [55] and [59]. Each stimulus is simulated for 1000ms, and we used the default parameter values indicated in Table 3 for the neural dynamics (equation 7) and in Table 4 for the synaptic dynamics (equations 8 to 12)

4.5. Aggregate measure of error

The classification error of each simulation is measured as the MAE (Mean Average Error) of the training and testing steps (only the testing results are presented herein). The aggregate error for all 300 LSMs is the median of the individual errors. We chose the median over the mean because it mitigates outlier effects due to the large variations of performance observed in different LSMs, thus providing a "truer" portrait of performance. These measures are compared for equality at a 99% confidence level by a Wilcoxon rank-sum test for medians (the mean MAEs are also recorded and tested for equality using a *t*-Test for validation purposes). The results obtained for the "filtered rates" representation serve as a baseline for the other tests.

In all, 27 experiments are performed, each with a unique combination of the following parameters:

- 3 sample rates (10 Hz, 4 Hz, 2 Hz);
- 3 levels of jitter (1 ms, 4 ms, 10 ms);
- 3 stimuli pairs frequency patterns (100-100 Hz, 20-20 Hz and 100-20 Hz).

Table 4: Default parameters of the dynamic synapse model

Parameter	Description	Value
W(EE)	Mean synaptic weight (Excitatory-Excitatory connection)	30e-9
W(EI)	Mean synaptic weight (Inhibitory-Excitatory connection)	-19e-9
W(II)	Mean synaptic weight (Inhibitory-Inhibitory connection)	-19e-9
W(IE)	Mean synaptic weight (Excitatory-Inhibitory connection)	60e-9
SHW	Multiplier of the standard deviation of synaptic weights	0.7
U(EE)	Synaptic efficacy utilization (Excitatory-Excitatory connection)	0.5
U(EI)	Synaptic efficacy utilization (Inhibitory-Excitatory connection)	0.25
U(II)	Synaptic efficacy utilization (Inhibitory-Inhibitory connection)	0.32
U(IE)	Synaptic efficacy utilization (Excitatory-Inhibitory connection)	0.05
τ_{rec}(EE)	Time constant for depression (Excitatory-Excitatory connection)	1.1 s
τ_{rec}(EI)	Time constant for depression (Inhibitory-Excitatory connection)	0.7 s
τ_{rec}(II)	Time constant for depression (Inhibitory-Inhibitory connection)	0.144 s
τ_{rec}(IE)	Time constant for depression (Excitatory-Inhibitory connection)	0.125 s
τ_{facil}(EE)	Time constant for facilitation (Excitatory-Excitatory connection)	0.05 s
τ_{facil}(EI)	Time constant for facilitation (Inhibitory-Excitatory connection)	0.02 s
τ_{facil}(II)	Time constant for facilitation (Inhibitory-Inhibitory connection)	0.06 s
τ_{facil}(IE)	Time constant for facilitation (Excitatory-Inhibitory connection)	1.2 s

5. Results

We begin with the efficiency of the temporal coding of liquid states and the roles of sampling rate, jitter and memory capabilities. Then, we consider the problem of larger liquids and finally present the results on the new models of SP.

5.1. Filtered rates vs. Synchrony matrix vs. composite representations.

Figure 5 reports the MAE results for the three types of liquid state representations averaged across all 27 experiments and sample frequencies. It shows a significantly better

performance of the composite-state approach over the other state representations (36% better than using filtered rates).

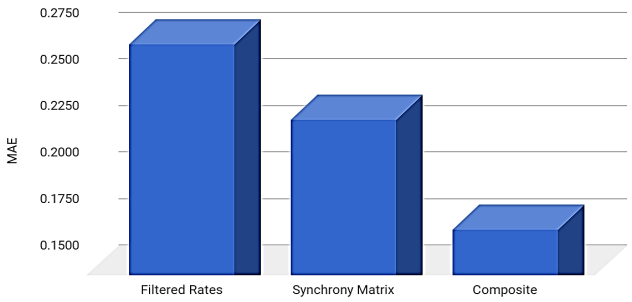


Figure 5. Classification performance as a function of liquid state representation.

The synchrony-matrix state representations ranked second behind the composite-state approach on average, but the synchrony-only representations performed slightly worse than the rate-based ones when the stimuli pairs were of different spike emission rates (i.e., 100 Hz and 20 Hz). In this case, the difference between the rates of emission led to clearly distinct rate-based representations, resulting in better performance as seen in Figure 6.

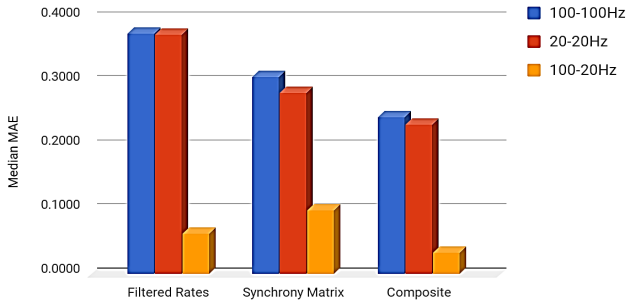


Figure 6. Classification performance as a function of input frequency type.

5.2. Impact of the sampling frequency

Figure 7 shows that higher sampling rates degrade the classification performance. This can be explained by the fact that less information is then conveyed by each sample, albeit temporal resolution is increased.

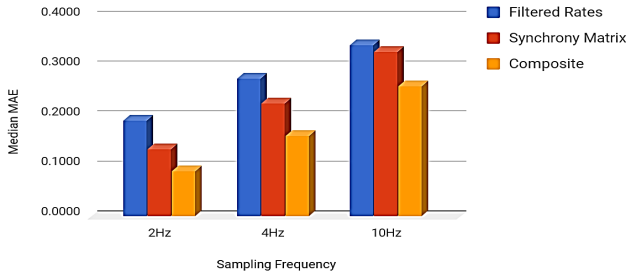


Figure 7: Classification performance as a function of input sampling rate.

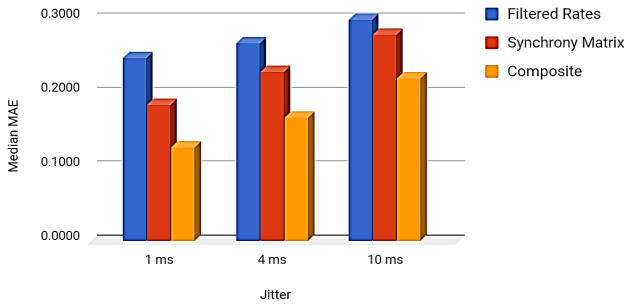


Figure 8. Classification performance for different spike train jitter levels

5.3. Impact of jitter on performance

Here also, increased levels of temporal noise decrease performance as may be expected. The degradation is significant as shown in Figure 8, but it appears to be less important than when using different sampling frequencies. Even a 10 ms jitter (representing between 10% and 50% of the stimulus base frequency) has a limited effect on classification error.

B. Effect of a bigger cortical column

The previous observations remain valid for a 3x3x3 LSM. Figure 9 shows the relative performance of each liquid state representation and for each of the three pairs of input stimuli. Again, the composite-state approach shows a significant improvement over the classical and the synchrony matrix methods. There are, however, caveats to the extensibility of the synchrony matrix approach as discussed in the next section.

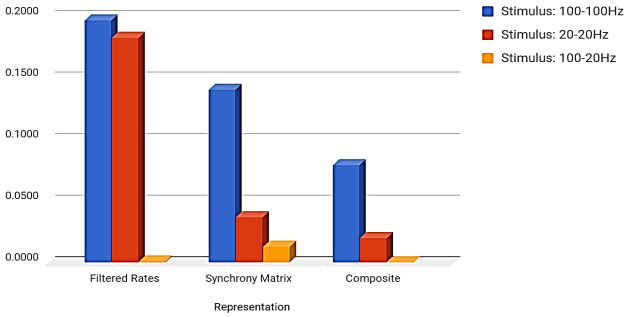


Figure 9. Median MAE for a 3x3x3 liquid

5.4. Role of memory

As expected, the value of *Tau* had a deep impact on performance. The difference between best and worst performance across a few sample points corresponding to *Tau* values ranging from 0 to 1.5 can be as high à 22.81% in the cast of the composite-state representations (see Table 4). It is also clear from Figure 10 that the relationship between memory and performance is nonlinear and dependent on the type of liquid state representation.

Table 4: Impact of memory on performance

	Filtered Rates	Synchrony Matrix	Composite
Minimum Error	0.3073	0.3005	0.2309
Maximum Error	0.3607	0.3560	0.2992
Variation	14.81%	15.59%	22.81%

5.5. Separation measures

We expect a negative correlation between the Separation measure and classification error. The magnitude of this correlation is an indication of how good a predictor of performance SP is. As shown in Figure 11, FDR performs best among the "classical" SP measures (Centroids, Ranks, Van Rossum distance, Gaussian distance). On the other hand, the newly introduced separation measures based on spike metrics seem to correlate even more with the classification performance of LSMs. The measures based on spike synchronization, ISI-distance (and its adaptive variant), adaptive rate-independent spike distance, and Schreiber correlation all correlate below the -0.5 mark.

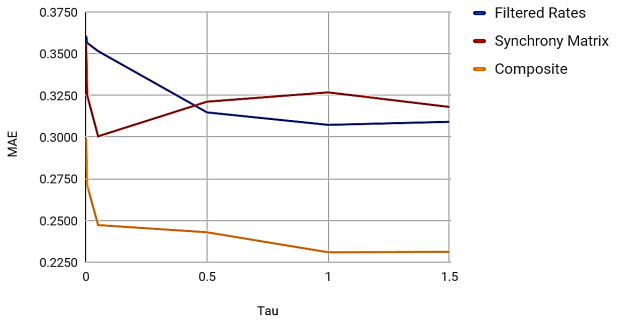


Figure 10. Performance as a function of Tau, the synaptic time constant

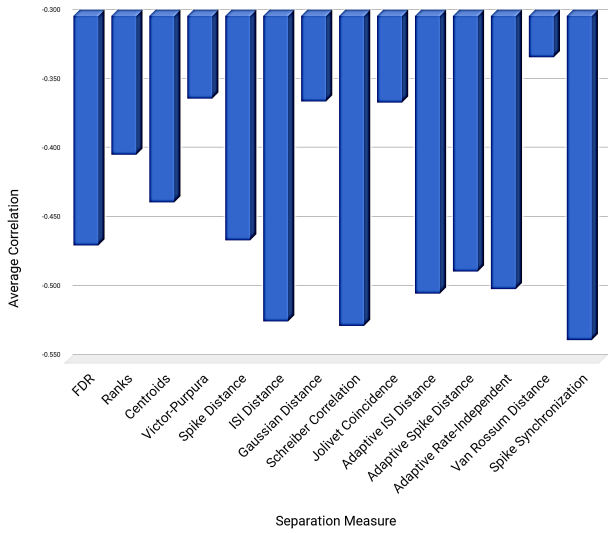


Figure 11. Correlation of SP measurement type with performance

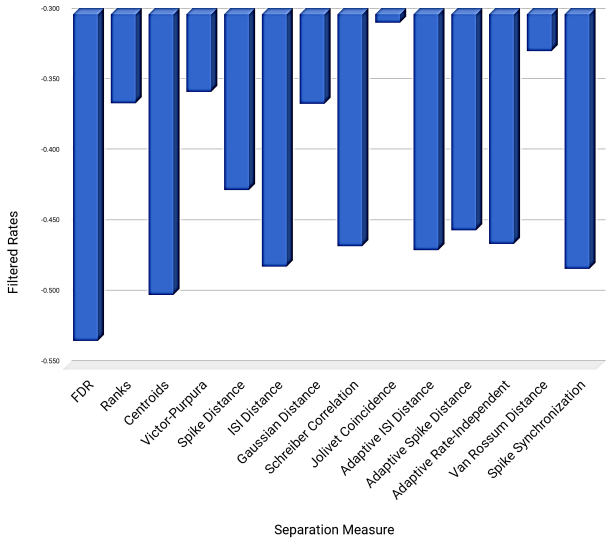


Figure 12. Correlation of SP measures with performance for a rate-based state representation

However, this synthetic chart shown in Figure 11 hides the local discrepancies between the different measures, since some of them perform significantly better in certain situations. For instance, Figure 12 shows that FDR correlates generally better with rate-based representations, but Figure 13 shows the opposite for synchrony-based representations (either using a synchrony matrix or a composite-state representation).

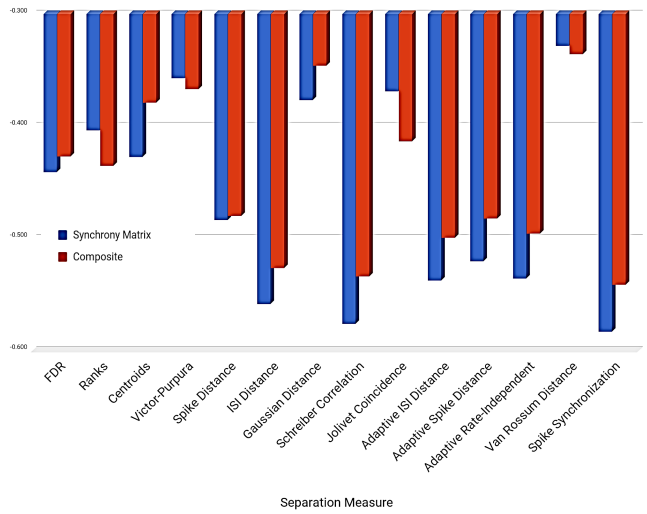


Figure 13. Correlation of SP measures with performance for synchrony-based and composite representations

6. Discussion and future research

At the core of each LSM study lies a neuron/synapse simulator for validation. It is not only a model of computation, but also a simplified model of the cortical columns found in real-life neural systems. This biological inspiration suggests that neuroscientific discoveries

regarding the understanding of the "neural code" may also enhance the computational model. We discuss below our findings on rate and temporal coding, the role of SP as a performance predictor, and the influence of sampling and memory on performance before ending this paper with a short review of the results of other comparable studies and concluding.

6.1. Rate and temporal decoding

The temporal decoding of liquid states can easily match the levels of LSM classification performance reached by rate-decoding approaches, while composite state feature vectors go beyond: they allowed to raise the overall performance without increasing the LSM dimensionality, even in the context of highly noisy inputs. These findings point to the critical role of phase information and temporal decoding in LSM classification, encouraging more research to explore temporal coding and decoding schemes. For instance, more encoding strategies based on phase information or absolute spike timings ought to be investigated (see [60] for inspiration), as well as more advanced spike rate estimation mechanisms (several of them are presented in [61]). With regards to performance, this study also found that the liquid state sampling frequency (in the light of its relationship to memory capabilities) and the topology of connections are critical factors, and they could thus become targets of optimization for peak performance. But, as the next subsection shows, the currently proposed measures of SP do not correlate well enough with LSM classification performance.

On the negative side, one major limitation of the synchrony matrix representation of liquid states is that the matrix grows with liquid size, with a quadratic impact on the synchrony vector size (equation 5). For example, for a 2x2x2 liquid, the number of elements of the feature vector is 28 as already mentioned, and for a 4x4x4 liquid, this number reaches 2080 (from equation 5 with N=64). Large feature vectors tend to promote overfitting: for instance, the linear regression that we use as a classifier quickly becomes overwhelmed with them. Several ideas can be put forward to cope with large-size liquids and reduce the dimensionality of the feature vectors. They include:

- Principal Component Analysis of the synchrony matrices for dimensionality reduction.
- Hierarchical construction of liquids: connecting a large liquid to gradually smaller ones until a suitably small liquid is found.
- Stochastic sampling of pairs of neurons in the liquid and subsequent reconstruction of the synchrony and rate information (also suggested in [5]).
- Temporal filtering and fuzzification: creating aggregate spike trains from liquid subregions (for example, we filter 2x2x2 regions in a 6x6x6 liquid to create a 3x3x3 "aggregate" liquid) and applying a form of temporal filtering, replacing multiple spikes emitted within a certain "fuzzy" timeframe by a single one at the mean of the that interval.

However, although temporal decoding may be a promising technique, it does not solve the performance variability across liquids. One way to address this problem is to determine performance predictors and engineer more efficient liquids.

6.2. Performance predictors and liquid optimization

The differences between "classical" SP and synchrony-based measures seem to highlight some complementarity between the two approaches. On average, our measure of SP roughly explained 50% of the performance of an artificial classification experiment (the absolute amount of correlation between SP and generalization error was, at best, slightly over 0.5). So, a question remains: what is missing from SP measures that could explain the missing 50%? Two very broad hypotheses can be put forward:

a) Other performance predictors should be used instead of or in conjunction with SP;

b) Our evaluation of SP is incomplete or deficient: some crucial information may not be captured by statistical methods, linear algebra or spike distance metrics.

While more SP measures (Bray-Curtis [62] and other vector distances [63], etc.) should be tested, we think that a custom, composite SP measure built out of rate and synchrony information should also be investigated. Indeed, synchrony-based or hybrid measures tend to correlate better with synchrony-based state representations while statistical measures perform better with rate-based representations. These results seem to hint that the ideal separation measure could be a hybrid of FDR and synchrony metrics, exploiting the idea that phase and rate information are complementary representations of the same spiking activity.

6.3. Memory and state sampling frequency

Our results also show that the memory capacity of a liquid has an impact on classification performance, particularly when using composite-state representations

(over 22% gap between best and worst performance, significantly higher than for rate-based or synchrony-only representations). Determining the optimal memory parameters and sampling frequency to extract enough information while preserving temporal resolution and avoiding pathological memory conditions deserves attention. The standard approach to divide the time axis into same-duration bins can create two problems: a) empty states, b) window boundary issues, especially for synchrony calculation (spikes fired just before or after the time window boundaries are not accounted for).

We observed a significant degradation of performance when the state sampling frequency was increased (illustrated in Figure 7), attributed to the decline of the information content of each sample and the increased time resolution. The reduction of the width of the sampling time window has two consequences:

a) More state samples required to cover the entire simulation , with the compounded sampling error most likely increased (i.e., the probability of missing relevant spikes at the boundaries of each time window).

b) Lower average number of spikes that can be captured in each sample, which in turn augments the impact of missed spikes on the total count of spike firings.

These two combined effects can lead to less reliable feature vectors. Low spiking activity also happens during network "warm up", as discussed in [6] and [17]. A solution to these problems may be to use interleaved sampling with window smoothing as done in automatic voice recognition systems.

6.4. Related work on LSM performance

Several authors have explored the question of the classification performance of the LSM model and highlighted its strengths and limitations in various contexts, but there seems to be a relative dearth of results comparable to ours, as methodologies and data vary significantly across studies. It is also worth noting that we have deliberately left out research on readout performance, although works such as [64] show that this crucial element of the LSM model can also be improved.

Putting aside the previous caveats, the conclusions we draw from this work are very much in line with those of the seminal work of Maass and al. [23] who validated the role of SP in LSM performance using a globally comparable method. In our work, we looked at this problem from the angle of liquid state representations, but numerous studies have focused on other aspects and suggested techniques to increase both the separability and the generalization properties of an LSM. The list includes:

- Optimizing the connection topology and the synaptic weights ([65], [19], [13], [17], [18]);
- Careful selection or mixtures of neuron types ([23], [66], [16], [1]);
- Addition of parallel columns [23];
- Construction of hierarchical liquids [16];
- Correct choice or composition of liquid state representations [10];
- Selection of the right memory parameters [3];
- Usage of ensemble techniques [5].

In addition, the problem of quantifying SP remains open. In [62] and [37], the authors indicate that their own custom measure of SP outperforms those based on either the ubiquitous Gaussian distance or the Van Rossum metric, but they did not provide correlation results with actual performance. Similarly, compelling evidence of a strong correlation between classification performance and SP is reported in [14]. The measures proposed by [3], [18] and [13] correlate to levels up to 0.79, 0.68 and 0.86 respectively, whereas our own results show an average correlation of only slightly above the 0.5 mark, a discrepancy than can be explained by the differences in the methodologies and validation contexts.

7. Conclusion

In this paper, we have shown through simulation experiments that the temporal decoding of spike trains by evaluation of the synchrony between pairs of neurons in the liquid can improve the LSM performance for classification tasks. We have also shown that the Separation Property, a fundamental characteristic of LSMs, can reliably be measured by spike metrics.

While there is a strong consensus in the research community that the classification performance of the LSM model can be raised, no definitive solution has yet emerged. We believe that more research is needed to improve the aforementioned approaches and/or combine their strengths to improve the LSM core performance in classifying time-varying input data. We also think that the results presented here should be tested and validated on a larger scale. Moreover, if the temporal decoding of liquid states improves classification performance, its efficiency in a less artificial context remains to validate.

References

[1] W. Maass, "Liquid State Machines: Motivation, Theory, and Applications," *World Scientific Review*, Vol. March 25/2010, doi: 10.1142/9781848162778_0008

[2] D. Verstraeten, B. Schrauwen, B., D. Stroobandt, "Isolated word recognition with the liquid state machine: a case study," *Proc. 13th European Symposium on Artificial Neural Nets (ESANN)*, vol. 95(6), pp. 435–440, 2005.

[3] E. Goodman, E., Ventura, D. " Spatiotemporal pattern recognition via liquid state machines," *International Joint Conference on Neural Networks (IJCNN)*, pp. 3848–3853, 2006.

[4] W. Maass, T. Natschlager, H. Markram, "Fading memory and kernel properties of generic cortical microcircuit models,"*Journal of Physiology (Paris)*, vol. 98, pp. 315–330, 2004.

[5] J. De Gruijl, M. Wiering, "Musical Instrument Classification using Democratic Liquid State," *Proc. 15th Belgian-Dutch Conference on Machine Learning, 2006*.

[6] L. Pape, R. Jornt, J. De Gruijl, M. Wiering, "Democratic Liquid State Machines for Music Recognition", *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks. Volume 83b, Berlin, 2008*.

[7] J. Hertzberg, H. Jager, F. Schonherr. F., "Learning to ground fact symbols in behavior-based robot," *Proc. 15th European Conference on Artificial Intelligence.*, pp. 708–712, Amsterdam, 2002.

[8] H. Burgsteiner, M. Kroll, A. Leopold, G., Steinbauer, "Movement prediction from real-world images using a liquid state machine," *Appl. Intell.* vol. 27 (2), pp. 99–109, 2007.

[9] H. Ju, J. Xu, A. VanDongen, A. "Classification of Musical Styles Using Liquid State Machines," *Proc. International Joint Conference on Neural Networks s (IJCNN*), 2010, doi: 10.1109/IJCNN.2010.5596470.

[10] F. Rhéaume, D. Grenier, E. Bosse, "Multistate combination approaches for liquid state machine in supervised spatiotemporal pattern classification," *Neurocomputing*, vol. 74, pp. 2842–2851, 2011.

[11] M. Aoun, M. Boukadoum, "Chaotic Liquid State Machine," *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 9(4), 1-20, October-December 2015

[12] J. Huang, H. Fang, Y. Wang, " Directional classification of cortical signals using a liquid state machine," *Proc. 17th World Congress of The International Federation of Automatic Control*, Seoul, Korea, July 6-11, 2008.

[13] E. Hourdakis, P. Trahanias , "Use of the separation property to derive Liquid State Machines with enhanced classification performance," *Neurocomputing*, vol. 107, pp40–48, 2013.

[14] R. Legenstein, W. Maass, "Edge of chaos and prediction of computational performance for neural circuit models, *Neural Networks*, vol 20(3), pp. 323–334, 2007.

[15] D. Norton, D. Ventura, "Preparing More Effective Liquid State Machines Using Hebbian Learning," *Proc. International Joint Conference on Neural Networks (IJCNN)*, 2006.

[16] J. Matser, " Structured liquids in liquid state machines," (Master Thesis, Utrecht University, 2010).

[17] S. Kok, "Liquid State Machine Optimization," (Master Thesis, Utrecht University, 2007).

[18] D. Norton, D. Ventura, "Improving the performance of liquid state machines through iterative refinement of the reservoir," *Neurocomputing*, vol. 73, pp. 2893–2904, 2010.

[19] J. Huang, Y. Wang, J. Huang, "The Separation Property Enhancement of Liquid State Machine by Particle Swarm Optimization," *International Symposium on Neural Networks (ISNN 2009: Advances in Neural Networks)*, pp 67-76, 2009.

[20] A. Kasinski, F. Ponulak, "Comparison of supervised learning methods for spike time coding in spiking neural networks,".*Int. J. Appl. Math. Comput. Sci.*, vol. 16(1), pp. 101–113, 2006.

[21] T. Kreuz, T. (2001). Measures of spike train synchrony, *Scholarpedia*, vol. 6(10):11934, 2001.

[22] *http://www.scholarpedia.org/article/Measures_of_spike_train_synchrony*

[23] W. Maass, T. Natschlager, H. Markram, H. "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14(11), pp. 2531–2560, 2002.

[24] D. Buonomano, W. Maass, "State-dependent computations: Spatiotemporal processing in cortical networks," *Nature Reviews in Neuroscience*, vol. 10, no. 2, pp. 113–125, 2009.

[25] J. Mayor, W. Gerstner, " Transient information flow in a network of excitatory and inhibitory model neurons: role of noise and signal autocorrelation," *Journal of Physiology (Paris)*, vol. 98, pp. 417–428, 2004.

[26] CSIM user manual: *http://www.lsm.tugraz.at/download/csim-1.1-usermanual.pdf*

[27] R. Legenstein, H. Markram, W. Maass, "Input prediction and autonomous movement analysis in recurrent circuits of spiking neurons," *Reviews in the Neurosciences (Special Issue on Neuroinformatics of Neural and Artificial Computation*), vol. 14, no. 1-2, pp. 5–19, 2003

[28] S. Hausler, H. Markram, W. Maass, "Perspectives of the high-dimensional dynamics of neural microcircuits from the point of view of low-dimensional readouts". *Complexity*, vol. 8(4), pp. 39–50, 2003.

[29] A. Oliveri, R. Rizzo, A. Chella, "An application of spike-timing-dependent plasticity to readout circuit for liquid state machine," *IEEE International Joint Conference on Neural Networks*, pp. 1441–1445, 2007.

[30] H. Ju et al., "Effects of synaptic connectivity on liquid state machine performance," *Neural Networks*, vol. 38, pp. 39–51, 2013.

[31] R. Brette, "Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain,". *Front. Syst. Neurosci.* 2015, doi: 10.3389/fnsys.2015.00151

[32] B. Meftah et al., "Image Processing with Spiking Neuron Networks," *Artif. Intell., Evol. Comput. and Metaheuristics, SCI 427*, 525–544, 2012.

[33] J. D. Victor, "Spike train metrics," *Current Opinion in Neurobiology*, vol. 15, pp. 585–592, 2005.

[34] T. Kreuz, "Measures of spike train synchrony," *Scholarpedia*, vol. 6(10), pp. 11934, 2001.

[35] W. Gerstner et al., *Neuronal Dynamics: From single neurons to networks and models of cognition*, Cambridge University Press, 2014, *http://neuronaldynamics.epfl.ch/online/Ch7.S6.html*.

[36] G. B. Stanley, "Reading and writing the neural code," *Nature Neuroscience*, vol. 16, pp. 259–263, 2013, doi:10.1038/nn.3330

[37] K. P. Dockendorf et al., "Liquid state machines and cultured cortical networks: the separation property," *Biosystems*, vol. 95(2), pp. 90–97, 2009.

[38] M. C. W. Van Rossum, "A novel spike distance," *Neural Comput.*, vol. 13, pp. 751–763, 2001.

[39] J. D. Victor, K. P. Purpura K.P., "Spike Metrics," *In: Analysis of Parallel Spike Trains. Ed. Stefan Rotter and Sonja Gruen. Springer,* pp. 129-156, 2010. http://www-users.med.cornell.edu/~jdvicto/pdfs/vipu10.pdf.

[40] E. Satuvurori et al., " Measures of spike train synchrony for data with multiple time scales," *J Neurosci. Meth,* vol. 287, pp. 25-38, 2017

[41] T. Kreuz et al., "Measuring spike train synchrony," *J Neurosci Meth.,* vol. 165, pp. 151–161, 2007.

[42] T. Kreuz et al., "Measuring synchronization in coupled model systems: a comparison of different approaches," *Phys D.,* vol. 225, pp. 29–42, 2007.

[43] T. Kreuz et al., "Measuring multiple spike train synchrony," *J. Neurosci Meth.,* vol. 183, pp. 287–299, 2009.

[44] T. Kreuz, D. Chicharro, M. Greschner, R. G. Andrzejak, "Time-resolved and time-scale adaptive measures of spike train synchrony," *J. Neurosci Meth.,* vol. 195, pp. 92–106, 2011.

[45] T. Kreuz et al., "Monitoring spike train synchrony," *J. Neurophysiol.,* vol. 109, pp. 1457-72, 2013.

[46] *http://www.scholarpedia.org/article/SPIKE-distance*

[47] J. D. Victor, K. P. Purpura, "Nature and precision of temporal coding in visual cortex: a metric-space analysis," *J. Neurophysiol.,* vol. 76, pp. 1310–1326, 1996.

[48] J. D. Victor, K. P. Purpura, "Metric-space analysis of spike trains: theory, algorithms and application, *Network,* vol. 8, pp. 127–164, 1997.

[49] S. Schreiber et al., "A new correlation-based measure of spike timing reliability," *Neurocomputing,* vol. 52, pp. 925–931, 2003.

[50] J. D. Hunter, G. Milton, G., "Amplitude and frequency dependence of spike timing: implications for dynamic regulation," *J. Neurophysiol.,* vol. 90, pp. 387-394, 2003.

[51] R. Quian Quiroga, T. Kreuz, P. Grassberger, "Event synchronization: a simple and fast method to measure synchronicity and time delay patterns," *Phys Rev E Stat Nonlin Soft Matter Phys, vol. 66:041904, 2002, doi: 10.1103/PhysRevE.66.041904.*

[52] T. Kreuz, M. Mulansky, N. Bozanic, "SPIKY: A graphical user interface for monitoring spike train synchrony," *J Neurophysiol.,* vol 113, pp. 3432, 2015.

[53] R. Jolivet et al., "A benchmark test for a quantitative assessment of simple neuron models," *J Neurosci Meth.,* vol 169, pp. 417–424, 2008.

[54] *http://www.lsm.tugraz.at/download/circuit-tool-1.1-manual.pdf*

[55] T. Natschlager, W. Maass, H. Markram. *Chapter 9, Computer Models and Analysis Tools for Neural Microcircuits, http://www.lsm.tugraz.at/papers/lsm-koetter-chapter-144.pdf*

[56] *http://www.lsm.tugraz.at/download/csim-1.1-usermanual.pdf*

[57] H. Markram, Y. Wang, M. Tsodyks, " Differential Signaling via the Same Axon of Neocortical Pyramidal Neurons," *Proc Natl Acad Sci U S A.*, 1998, doi: 10.1073/pnas.95.9.5323..

[58] CSIM: A neural Circuit SIMulator, *http://www.lsm.tugraz.at/csim/*

[59] R. Brette et al., "Simulation of networks of spiking neurons: A review of tools and strategies," *J. Comput Neurosci.,* vol. 23(3), pp. 349-98, 2007, doi 10.1007/s10827-007-0038-6

[60] B. Gardner, I. Sporea, A. Gruning, "Encoding Spike Patterns in Multilayer Spiking Neural Networks," *arxiv.org/abs/1503.09129v1,* 2015.

[61] S. Shinomoto, S. Grun, S. Rotter, S. (2010). *Chapter 2, Estimating the Firing Rate. Analysis of Parallel Spike Trains, in S. Grün, S. Rotter (eds.), Analysis of Parallel Spike Trains, Springer Series in Computational Neuroscience 7, 2010, doi 10.1007/978-1-4419-5675-0_2*

[62] G. M. Wojcik, M. Wazny, " Bray-Curtis Metrics as Measure of Liquid State Machine Separation Ability in Function of Connections Density," *Procedia Computer Science, Vol. 51*, pp. 2938–2942, *Proc. International Conference on Computational Science (ICCS), 2015.*

[63] M. Loginov, "Beyond decoding: representational similarity, analysis on fMRI data," (Master's Thesis, University of Tartu, 2015).

[64] S. Roy, A., Banerjee, A. Basu, "Liquid State Machine with Dendritically Enhanced Readout for Low-power, Neuromorphic VLSI Implementations," *IEEE Trans. Biomedical Circuits and Systems,* vol. 8(5), pp. 681-695, 2014. doi: 10.1109/TBCAS.2014.2362969

[65] H. Hazan, L. M. Manevitz "Topological constraints and robustness in liquid state machines," *Expert Systems with Applications,* vol. 39, pp. 1597–1606, 2012.

[66] B. J. Grzyb, E. Chinellato, G. M. Wojcik, W. A. Kaminski, W.A., "Which model to use for the Liquid State Machine?" *Proc. International Joint Conference on Neural Networks (IJCNN), pp. 1018-1024, 2009, doi: 10.1109/IJCNN.2009.5178822.*

NICOLAS PAJOT was a PhD candidate in cognitive informatics during the tenure of this work. He is the founder and CEO of Ultracine Studios, a Montreal company specialized in Cinematic Gaming on Mobile Platforms, building realistic representations of clothing, hair and other soft body objects in real time .

MOUNIR BOUKADOUM is a full professor of engineering at Université du Québec à Montreal (UQAM), Canada. He has an interdisciplinary background in physics, electrical engineering and computer science. He received the Ph.D. degree in EE from the University of Houston, TX, USA. His research activities include artificial intelligence in engineering analysis and design, with an emphasis on medical and biomedical applications.

Received: 04 October, 2023, Revised: 06 January, 2024, Accepted: 06 January, 2024, Online: 30 January, 2024

DOI: https://doi.org/10.55708/js0301002

Baggage Cart with Weighing Mechanism for Hotels and Airlines

Vishal Verma*[,1], Kuldeep Kumar[2], Rashmi Aggarwal[1], Tanvi Verma[1]

[1] Chitkara College of Hospitality Management, Chitkara University, Rajpura, 140401, Punjab India
[2] Chitkara Business School, Chitkara University, Rajpura, 140401, Punjab India
*Corresponding author: Vishal Verma, vishal.verma@chitkara.edu.in, +91 9041078993

ABSTRACT: In this article, it has been proposed a functional design of Baggage Cart with Weighing Mechanism for Hospitality Industry based on empirical observations. This design is expected to promote and become one of the high demand products which can later be used especially by hotels and respective airlines. It has been often observed that travelers are always concerned about the maximum weight of their baggage allowed in order to board a flight. Because as per norms of airlines one needs to carry a specific amount of baggage in the flight, in case of extra weighed baggage carried by a guest, he/she is charged extra amount as per KG by the respective airline. Keeping this point in mind, travelers are always conscious about the weight of their baggage before boarding a flight to avoid the last-minute hassle of paying huge extra money, and sometimes they need to drop the necessary items out from the baggage to adjust the weight of the baggage. In hotels, a guest request to weigh his/ her luggage is dealt with in the following ways: The front office associate/porter first needs to fetch the baggage from the guest room and bring it up to the bell desk or need to drag it up till time office / receiving area where it is weighed on a heavy-duty weighing scale. It is the traditional way of measuring baggage. The traditional way is time-consuming, uncomfortable, and less suitable to support the need of the traveler. Hence, to minimize the time and efforts, we wanted to invent a mechanism that will help and reduce the burden of managing the separate weighing machine, and therefore we came up with a baggage cart with a weighing mechanism in it.

KEYWORDS: Baggage Cart, Weighing Mechanism, Customer Satisfaction, Service Quality

1. Introduction

Passenger Satisfaction is a crucial metric in order to evaluate airport performance. International airports in various regions or countries usually do not compete with each other. Passengers generally do not have the choice to choose between airports, irrespective of the quality and price of airport services. However, Passenger's demand for airport services is likely to be fairly inflexible [1]. To minimize time, effort, and ease out the task, the design provides a mechanism to weigh the baggage as per the guest's request. While working in the hotel industry, it was a routine task to deal with guests and their needs and out of all to weigh their baggage was one of them. One genuine problem which is paid less attention and provided with an ordinary, less effective, or non-hospitable solution is the handling of a traveler's need of weighing his baggage. Moreover, Airports' key customers are passengers and their initial impressions of airport facilities and services can significantly impact their perceptions and evaluations of the airport. Therefore, it is vital for the airports to provide safe, comfortable and convenient services in a cost-effective manner [2]. Transactions involving goods and services are not the only aspects of business. Its capacity to last is significantly reliant on a variety of humanitarian aspects related to the management and upkeep of commercial relationships with clients, suppliers, and staff, particularly when conducting business internationally [3]. Keeping these in mind, Quality of Service as apparent by customers is a comparison amongst performance and expectations [4]. As per [5], The overall observation of service quality is determined by the modification between customers' expectations and their actual experiences. When a

customer's experience matches their expectation, service quality is perceived as upright. In other words, Quality of the service is measured by comparing customer expectations with their experiences and this comparison is based on the performance of service delivery. The goal of providing better service to clients is to meet their needs in addition to making sure they are completely satisfied with the service they are utilizing [6]. In the hospitality industry, it is a routine task to weigh the baggage of guests. The proposed design provides a mechanism to weigh the baggage as per the guest's request with minimal time, effort. Present solutions are not convenient and need the efforts of the porter to fetch the baggage to and fro between the guests' rooms and hotel utilities which is a time-consuming process.

2. Technical Background

At present, the hospitality industry is using the following mechanisms to weigh the baggage. Based on empirical observations and as shown in Figure 1 below, Earlier In a traditional way, a porter needs to fetch the bag from the guest room which is taken to the time office / receiving area where it is weighed on a heavy-duty weighing scale. The moment it's being measured, it is properly tagged and delivered to the respective guest room. The process needs to be executed immediately on the guest's demand as the guest has to check out and take a flight or train. Even a reasonable delay in the process upsets the guests and creates a bad impression of the hotel.

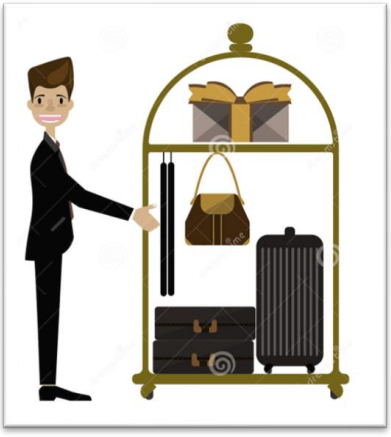


Figure 1: Traditional way of Measuring Baggage through the help of Porter

In another mechanism, as per empirical observations and as shown in Figures 2 & 3 below, some hotels use a weighing device with a lifting hook. The porter has to hold the device and lift the bag hooked to the device to weigh it. Sometimes the process may result in a muscular sprain in the porter's hand or back or lead to dropping the bag, which could again lead to a bad impression of the hotel.

In a few hotels, as per empirical observations and as shown in Figure 4 below, the guest is prompted to measure his / her baggage on his own by using the weighing scale kept in their bathroom. However, this system is not suitable as it is meant for weighing human body weight. In case one holds a bag and stands on the machine, it gives total weight which is not an appropriate method. it's not even feasible to keep a large piece of baggage on the surface of a weighing scale.

Figure 2: Traditional way of Measuring Baggage through a weighing device with a lifting hook (top view)

Figure 3: Traditional way of Measuring Baggage through a weighing device with a lifting hook (front view)

Figure 4: Traditional way of Measuring Baggage through the weighing scale kept in their bathroom

3. Related Work

3.1. Prior Art

As per empirical observations, figure 5 represents the traditional way of measuring the baggage. It would be easy and convenient for the front office associate/porter to use efficient ways to reduce their efforts and save time.

As per empirical observations, the use of the proposed mechanism in baggage cart is not restricted or limited to just hotels only, even the respective airlines can use it during a passenger's arrival inside the airport, before

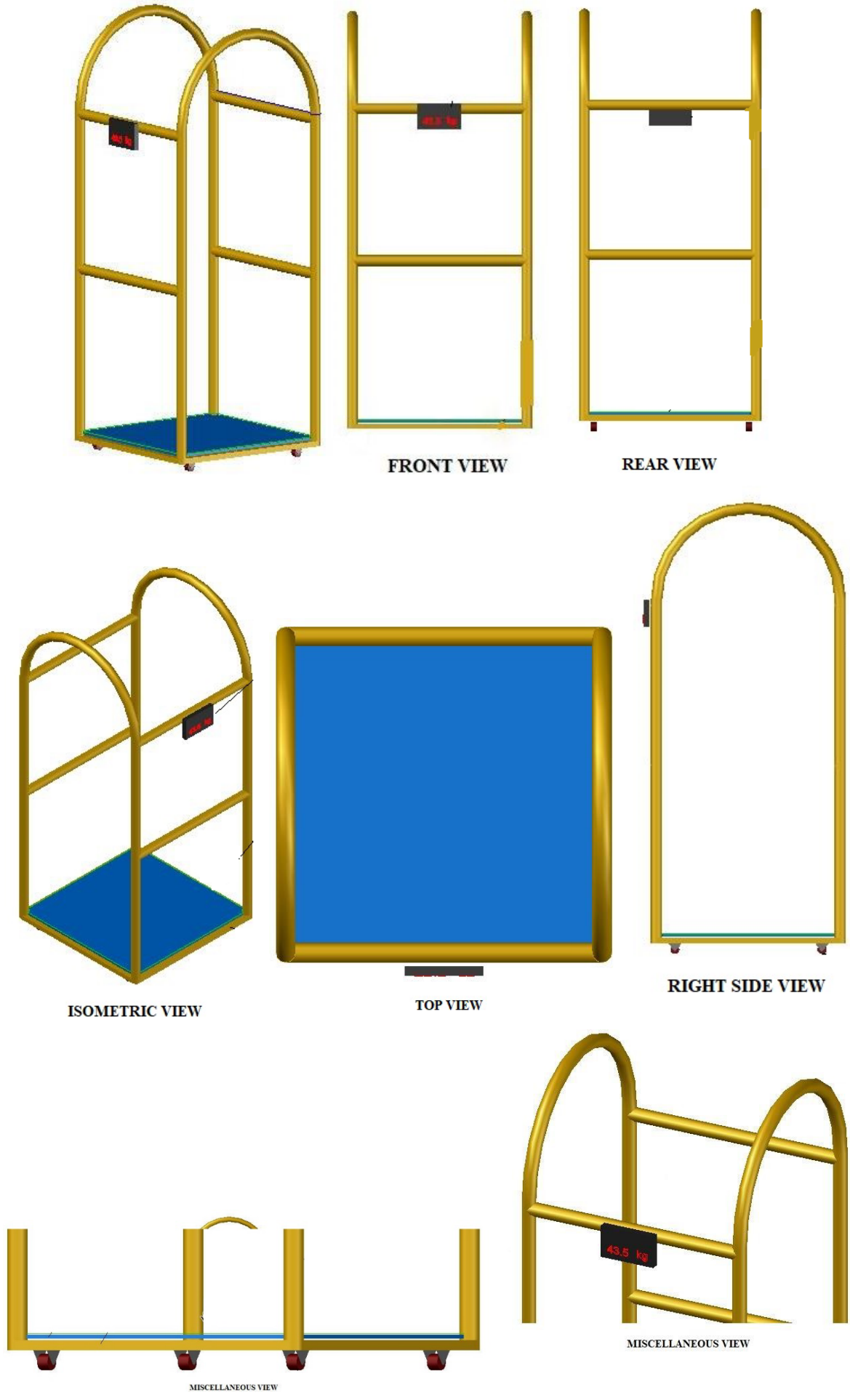


Figure 6: Proposed design for Baggage Cart with Weighing Mechanism for Hotels and Airlines

moving towards their flight one needs to finish check-in formalities including the weighing of their bags. Using this mechanism would enable them to get the results faster than the existing ones.

Figure 5: Traditional way of Measuring Baggage through the weighing scale

3.2. *Advantage of the proposed system in comparison with prior art*

- **Time Effective:** It saves time as it is always available with Front Office / Bell Desk.
- **Easy Handling:** It is always handy and in use by Bell Desk Staff.
- **Reduced Efforts:** No more to-and-fro efforts, as a handy baggage cart with a weighing mechanism, is available all the time.

3.3. *Proposed Solution*

The mechanism was designed to minimize the time and effort required in the traditional weighing method through a separate weighing machine. The proposed design allows installing the weighing mechanism on the baggage cart. The proposed design for Baggage Cart with weighing mechanism for hotels and airlines are shown in figure 6.

4. Hardware Implementation

As per empirical observations, the surface of the baggage cart has a mechanism to measure the weight. Upon placing a bag on the cart's surface, it will display the weight on the digital display mounted on the cart. The baggage cart will appear like a regular baggage cart when the digital meter is turned off.

5. Conclusions

The product is designed to eliminate the existing flaws used by respective hotels and airlines of guests for weighing their luggage, especially at the eleventh hour of checking out of the hotel. As per the enclosed pictures from 1-4, traditionally these are the methods of fetching baggage and measuring them too which are not at all convenient and time-effective. The most significant advantage of the baggage cart with a weighing mechanism is that it can be used by both hotels and airlines for their purposes. This product is expected to find high acceptance by hospitality professionals because of lesser effort, easy handling, and being time effective.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. Doganis, The Airline Business, Routledge, 2005, doi:10.4324/9780203596807.

[2] R. Rendeiro Martín-Cejas, "Tourism service quality begins at the airport," *Tourism Management*, vol. 27, no. 5, pp. 874–877, 2006, doi:10.1016/j.tourman.2005.05.005.

[3] B.K. Das, S. Sharma, "Influence of Relationship & Behavioural Aspects on Customer Retention & Loyalty in B2B Flexible Packaging Industry in Bangladesh," *International Journal of Applied Business and Economic Research*, vol. 15, no. 22, pp. 71–81, 2017.

[4] A. Parasuraman, V.A. Zeithaml, L.L. Berry, "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, vol. 49, no. 4, pp. 41–50, 1985, doi:10.1177/002224298504900403.

[5] C. Gronroos, Service management and marketing: managing the moments of truth in service competition, Lexington Books, 1990.

[6] J. Sharma, J. Singh, A. Singh, "Impact of E-Banking Service Quality on Customer Satisfaction," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 5, pp. 2296–2300, 2020, doi:10.35940/ijrte.E5841.018520.

VISHAL VERMA is an Experienced hotel professional with 8 years of hotel experience with 5-star hotel brands, like Oberoi Hotels & Resorts, Leela Kempinski, Carlson Rezidor, Wyndham, Post that it's been 8 years since he's working with Chitkara University as an Assistant Professor. He has publications in various reputable journals with exceptional orientation of research. His area of interest is Revenue Management and research skills in Tourism, Travel & Hospitality with expertise in Room Division, Sales and Marketing as well.

KULDEEP KUMAR has a decade of experience in the hospitality industry, working with top brands such as Marriott International, The Oberoi Hotels and Resorts, Kempinski Hotels, and Wyndham Hotels. For the last five years, he has been sharing his industry insights with students at the prestigious Chitkara University, while also pursuing research and innovation in the same field.

RASHMI AGGARWAL is a Professor and Dean Commerce at Chitkara Business School, Chitkara University, Punjab. She is PhD, M. Phil, M. Com and UGC-NET qualified, with a vast experience of 23 years in teaching and administration at various positions. Her areas of specialization are Accounting, Finance and Management. She has publications in many reputable journals with excellent research orientation. She has been awarded with the excellence Award for Most Collaborating Employee. She also got the letters of appreciation for strenuous and meticulous efforts by her in guiding and mentoring the students. She is an effective communicator along with the exceptional

ability of relationship management. She is always committed to creating an inclusive learning environment for the students where they can explore their potential.

TANVI VERMA holds a doctorate from University Business School, Panjab University, in Finance and Strategic Management. She possesses a decade of experience in the fields of Accountancy, Performance Management, Financial Reporting, Strategic Management, and subjects related to the Association of Chartered Certified Accountants (ACCA). At present, she holds the esteemed position of Assistant Professor at Chitkara University in Punjab. She has a strong focus on research and a solid list of papers to her name. Her work has been published in a multitude of prestigious journals. In several of the national and international conferences where she has presented, Ms. Verma has been honored with the Best Paper Award.

Received: 04 January, 2024, Revised: 09 February, 2024 Accepted: 14 February, 2024, Online: 29 February, 2024
DOI: https://dx.doi.org/10.55708/js0302004

Robust Localization Algorithm for Indoor Robots Based on the Branch-and-Bound Strategy

Huaxi (Yulin) Zhang[1], **Yuyang Wang**[2], **Xiaochuan Luo**[*,3], **Baptiste Mereaux**[4], **Lei Zhang**[5]

[1]LTI, Université de Picardie Jules Verne, Saint Quentin, 02100, France
[2]Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, 201210, China
[3]College of Information Science and Engineering,Northeastern University, Shenyang, 110819, China
[4]Independent Researcher, Saint Quentin, 02100, France
[5]euroDAO S.A.S., Saint Quentin, 02100, France
*Corresponding author: Xiaochuan Luo, Northeastern University, Shenyang, 110819, China, Contact Email: luoxch@mail.neu.edu.cn

ABSTRACT: Robust and accurate localization is crucial for mobile robot navigation in complex indoor environments. This paper introduces a robust and integrated robot localization algorithm designed for such environments. The proposed algorithm, named Branch-and-Bound for Robust Localization (BB-RL), introduces an innovative approach that seamlessly integrates global localization, position tracking, and resolution of the kidnapped robot problem into a single, comprehensive framework. The process of global localization in BB-RL involves a two-stage matching approach, moving from a broad to a more detailed analysis. This method combines a branch-and-bound algorithm with an iterative nearest point algorithm, allowing for an accurate initial estimation of the robot's position. For ongoing position tracking, BB-RL uses a local map-based scan matching technique. To address inaccuracies that accumulate over time in the local maps, the algorithm creates a pose graph which helps in loop-closure optimization. Additionally, to make loop-closure detection less computationally intensive, the branch-and-bound algorithm is used to speed up finding loop constraints. A key feature of BB-RL is its Finite State Machine (FSM)-based relocalization judgment method, which is designed to quickly identify and resolve the kidnapped robot problem. This enhances the reliability of the localization process. BB-RL's performance was thoroughly tested in real-world situations using commercially available logistics robots. These tests showed that BB-RL is fast, accurate, and robust, making it a practical solution for indoor robot localization.

KEYWORDS Branch-and-bound, Global localization, Position tracking, Robot kidnapping

1. Introduction

The growing demand for mobile robots in tasks such as repair, transportation, and cleaning necessitates the development of efficient techniques for robot localization [1]–[5]. Particularly in known environments, robots should be able to localize themselves within a prebuilt map, enabling them to position themselves based on data collected from various sensors. The problem of localizing mobile robots in indoor environments can be categorized into three sub-problems: position tracking, global localization, and the kidnapped robot problem [6, 7]. This paper proposes a fast, robust, and accurate algorithm to achieve indoor localization of mobile robots, effectively solving the three localization sub-problems simultaneously in real-world applications.

Recent advancements in indoor robot localization research have shown significant progress, yet challenges remain in simultaneously addressing three critical localization issues. The first issue involves global localization. Often, an initial pose is determined by observing the robot's approximate position in the environment to reduce computational effort and maintain localization stability. Despite this, without an initial estimate, achieving desirable global localization remains difficult.

The second issue is position tracking. Here, the challenge lies in the timely elimination of accumulated errors. To address this, two main strategies are employed. The first is simultaneous localization and mapping (SLAM), which involves frontend scan matching and backend optimization. While effective, SLAM methods are computationally demanding and rely on loop-closure detection to correct errors. The second strategy involves odometry, such as visual or LiDAR odometry, which calculates the robot's relative pose incrementally using adjacent data. However, these methods are prone to error accumulation over time, making them suitable primarily for short-term tracking.

Finally, the third issue is the kidnapped robot problem. This occurs when a robot, initially well-localized, is unexpectedly moved to an unknown location. This problem can be split into two scenarios: real kidnapping, where the robot is physically relocated by external forces such as

human intervention or an accident, and perceived kidnapping, which results from localization failures. Addressing this issue effectively remains a challenge for most existing approaches.

Considering these aspects, we propose a robust and accurate robot localization algorithm, which consists of three parts: global localization, position tracking, and relocalization judgment.

- The *global localization algorithm*, which is used to determine the robot's initial pose, can be divided into two stages. In the coarse matching stage, the branch-and-bound algorithm based on depth-first search (DFS) is used to promptly identify the absolute position of the robot on the map without any initial estimate. In the fine matching stage, the iterative nearest point algorithm is used to perform iterative optimization to determine the optimal initial pose of the robot. This algorithm can rapidly converge anywhere on the map, and the robot pose exhibits global optimality.
- The *position tracking algorithm* is used for the continuous localization of the robot when the initial pose is known. A local map-based scan matching method is used to estimate the relative pose of the robot and simultaneously build a local map. Moreover, a global pose graph optimization algorithm is used to eliminate the accumulated errors between local maps. Additionally, to ensure that the computing time is nonintractable, a DFS-based branch-and-bound algorithm is used to accelerate the process of identifying the loop constraints.
- The *relocalization judgment algorithm* is used to address the problem of robot kidnapping and eliminate the accumulated errors of the robot. We propose an FSM-based relocalization judgment method based on confidence calculation and dual-threshold judgment to effectively monitor the localization status of the robot. When the calculated confidence is less than the minimum threshold, the global localization algorithm is invoked for localization recovery.

The main contributions of this research can be summarized as follows:

1. *Development of a Two-Stage Global Localization Algorithm*: We introduce a novel two-stage global localization algorithm that combines the broad search capabilities of the branch-and-bound algorithm with the local optimization efficiency of the iterative closest point algorithm. This ensures the robot quickly identifies the globally optimal initial pose without relying on any preliminary estimates.
2. *Establishment of a Position Tracking Algorithm*: Our research incorporates a position tracking algorithm that integrates frontend local map-based scan matching with backend pose graph optimization. This approach provides a highly accurate state estimation of the robot, crucial for precise navigation.
3. *Creation of an FSM-based Relocalization Judgment Algorithm*: We have developed an innovative FSM-based relocalization judgment algorithm that utilizes an inflated occupancy grid map to minimize the impact of sensor measurement noise. This algorithm is adept at efficiently detecting instances of robot kidnapping, thereby safeguarding against localization failures in diverse scenarios and ensuring swift and effective localization recovery.
4. *Proposal of a Joint Localization Algorithm*: The research culminates in a comprehensive joint localization algorithm capable of concurrently addressing the challenges of global localization, position tracking, and robot kidnapping in indoor settings. The efficacy of this algorithm has been rigorously validated using commercial logistics robots, demonstrating its successful application in real-world environments.

2. Related work

Consistent and efficient localization is a core concept of indoor robot navigation, as knowledge of the robot position is crucial in deciding future actions [8]. In recent years, several researchers have focused on indoor robot localization [9]. However, most of the existing approaches focus on solving a specific problem of localization (such as global localization), which is fundamentally different from the motivation of our work.

Localization refers to the procedure of determining the robot pose with respect to its environment by using various noisy sensors. According to the type of measurement data, the sensors used in the process of robot localization can be divided into two classes: proprioceptive sensors and exteroceptive sensors. Proprioceptive sensors (such as encoders and IMUs) measure the robot motion by using deduced reckoning to calculate the relative robot displacement [10]–[12]. Since such sensors consider the instantaneous speed or acceleration to estimate the robot state, the integrated error in the localization process increases in a nonbounded manner over time. Hence, such sensors are usually used in combination with exteroceptive sensors that can determine the absolute positions to enhance the robot's ability in managing uncertainties [13]–[16]. Proprioceptive sensors address position tracking issues due to their inability to sense environmental information.

In addition to the methods based on proprioceptive sensors for localization, several approaches use exteroceptive sensors to recognize the environment around a robot to estimate the robot location. Among these methods, SLAM is widely used. In terms of the primary type of adopted sensor, the SLAM algorithm can be divided into two classes: visual SLAM and LiDAR SLAM. Visual SLAM aims to address the pose estimation of cameras with visual information. This method has evolved from the use of monocular cameras [17] to stereo cameras [18] and depth cameras [19]. The classic variants of monocular SLAM include ORB-SLAM [20], DSO [21], LSD-SLAM [22], and SVO [23]. Certain researchers, [24] adapted ORB-SLAM to a fisheye camera, tightly coupled visual information and IMU data to robustly estimate the camera pose and used the multimap technology to effectively solve the problem

of localization failure. In another study [25], a rolling-shutter camera and IMU were tightly coupled to minimize the photometric error to estimate the robot pose. Other researchers [26, 27, 28] used deep neural networks to eliminate the scale ambiguity of monocular cameras and extract high-level semantic features to enhance the system robustness and accuracy. The classic variants of stereo SLAM include ORBSLAM2 [29], ORBSLAM3 [30], PL-SLAM [31], and SOFT2 [32]. An event camera [33] was used to address the problems of high dynamics and low light, and the depth estimation of multiple viewpoints was merged in a probabilistic manner to build a semidense point cloud map. Notable research on RGB-D SLAM includes that on the RTAB-MAP [34], bundle fusion [35], and Kintinuous [36]. Moreover, a lightweight semantic network model was proposed [37], which integrates multiple technologies such as VIO, pose graph optimization, and semantic segmentation, to achieve the high-precision reconstruction of the three-dimensional environment. Deep learning techniques have also been employed in visual SLAM to extract features, enhancing the algorithm's ability to interpret and understand the visual information as in LIFT-SLAM [38] and Object-Fusion [39]. Because depth cameras can directly obtain the depth information of the environment, their use has been widely considered [40]. However, processing of the depth data is computationally expensive, and it is difficult to satisfy the real-time operation requirements of the CPU. Moreover, the frontend odometry aspects of visual SLAM can only estimate the relative pose of the robot, and back-end loop-closure detection can only achieve relocalization in similar scenes. Therefore, this approach cannot realize global localization.

LiDAR SLAM can be divided into 2D LiDAR SLAM and 3D LiDAR SLAM according to the type of LiDAR used. The classic 3D LiDAR SLAM algorithms include LOAM [41], HDL graph slam [42], and SuMa++ [43]. LOAM exhibits a high performance on the KITTI dataset, and thus, many improved versions of this algorithm have been proposed. In [44], the distinctive edge features and planar features were extracted to achieve two-step Levenberg–Marquardt optimization. In [45], the LiDAR and IMU data were tightly coupled. The IMU preintegration factor was introduced in the pose graph optimization to update the bias of the IMU, and the accumulated errors were corrected through loop-closure detection. Moreover, excellent schemes for 2D LiDAR SLAM have been proposed in recent years. The classic filter-based algorithms include Fast SLAM [46] and Gmapping [47], and graph-based algorithms include Karto SLAM [48] and Cartographer [49]. Cartographer, developed by Google engineers, has been proven to be a complete SLAM system that integrates localization, mapping, and loop-closure detection. At the frontend of this algorithm, the relative pose of the robot is calculated using the scan-to-submap matching method, which has a significantly lower accumulated error than the scan-to-scan matching method [50]. Additionally, compared with the scan-to-map matching method [51], it is considerably less computationally intensive and can run in real time. Similarly, since the origin of the robot localization is determined when initializing the algorithm, LiDAR SLAM is essentially an odometry technique and cannot solve the problems of global localization and robot kidnapping. To realize indoor localization, 2D LiDAR has been widely used due to its cost and accuracy. Certain researchers [52] and [53] attempted to enhance the accuracy of their localization system by using the extended Kalman filter to achieve multisensor fusion. However, these approaches cannot solve the problems of global localization and robot kidnapping. In [54], a quasistandardized 2D dynamic time warping (QS-2DDTW) method was proposed to solve the problem of robot kidnapping. The approach uses scan data for two consecutive ranges to obtain the geometric shape similarity of the environment to determine the robot state. Nevertheless, this approach cannot solve the position tracking problem. However, other studies [55]–[59] addressed the three major localization problems by using the adaptive Monte Carlo localization algorithm. Notably, using only ultrasonic sensors, the localization accuracy of the order of decimeters can be achieved.

In addition to the two types of exteroceptive sensors for localization, several wireless devices (such as WiFi, UWB, Bluetooth, and RFID) can be deployed indoors to realize reliable localization. In [60] and [61], the Kalman filter was used to fuse IMU and UWB data to obtain a relatively accurate robot pose. However, these approaches could not solve the problems of global localization and robot kidnapping. In addition, high accuracy localization was achieved using commercial WiFi devices [62]. The robust principal component analysis for extreme learning machine algorithm (RPCA-ELM) could suppress the effect of measurement noise in the localization process. In [63] and [64], to enhance the robustness of localization, UHF radio frequency identification technology was adopted. However, the system accuracy depended on the RFID tag, and global localization could not be realized at arbitrary positions. Furthermore, localization was realized in [65] and [66] by deploying a set of photoresistor sensors on a robot to collect information regarding an LED array in the environment. However, high-precision position tracking could not be realized. In addition, a robot localization system based on asynchronous millimeter-wave radar interference was proposed [67], which used the interference between multiple millimeter-wave radars with known positions in the environment to calculate the position of the robot. However, the system exhibited limited localization accuracy.

In summarizing the state of the art in indoor robot localization, it is clear that researchers have made significant strides using a variety of methodologies and sensor technologies. From SLAM implementations—both visual and LiDAR-based—to sophisticated sensor fusion techniques leveraging proprioceptive and exteroceptive sensors, including the use of wireless technologies like WiFi, UWB, Bluetooth, and RFID to enhance localization capabilities, each method aims to address specific facets of the complex challenge of localization, focusing on global localization, position tracking, or resolving the kidnapped robot problem.

Despite these advances, a comprehensive solution that simultaneously addresses all three critical challenges of indoor robot localization remains elusive. Existing studies tend to focus on optimizing specific aspects of localization rather than offering a unified algorithm capable of han-

dling global localization, precise position tracking, and the kidnapped robot scenario in an integrated manner. This gap in the research landscape underscores the innovative potential of the proposed BB-RL algorithm, which aims to provide a holistic approach to the multifaceted problem of autonomous indoor navigation. By doing so, BB-RL aspires to establish a new method in the field, offering a more robust, accurate, and comprehensive solution to indoor robot localization than currently available methods.

Some literature mentions studies that attempt to address all three major localization challenges simultaneously using the Adaptive Monte Carlo Localization algorithm [55]–[59]. These studies illustrate the potential of multi-sensor fusion and intelligent algorithms to enhance indoor localization accuracy and robustness. However, despite offering a composite solution, these approaches may still face limitations in practical application, such as dependency on specific types of sensors, the impact of environmental complexity on localization accuracy, and challenges in maintaining high precision in dynamic and unknown environments.

The Self-Adaptive Monte Carlo Localization (SA-MCL) method represents an advancement in addressing the inherent challenges of robot localization, including global localization, position tracking, and the "kidnapping" problem, where a robot is moved to an unknown location. Previous studies have shown that by employing the adaptive Monte Carlo localization algorithm, significant strides can be made in solving these three major localization challenges. These methods, however, are predominantly based on 2D environments and utilize ultrasonic sensors for sensing.

Transitioning from 2D to 3D environments introduces new challenges for the Monte Carlo Localization (MCL) algorithm. In [68], the authors propose a pure 3D MCL localization algorithm to address these challenges directly. Meanwhile, other approaches, such as the one by [69], adapt 2D MCL for localization in 3D maps. These methods illustrate the diversity of strategies being explored to solve localization problems in three-dimensional spaces using the MCL framework in 3D Map. The demand for computational resources and memory usage significantly increases in 3D Monte Carlo localization due to the necessity to process and track a much larger number of particles to accurately estimate a robot's pose in three-dimensional space. Each particle's position, orientation, and weight must be maintained, leading to escalated memory requirements as the particle count increases. Furthermore, without prior knowledge of the robot's approximate location, distributing particles effectively throughout the three-dimensional space to ensure comprehensive coverage and, by extension, the accuracy of the localization process, presents a considerable challenge. This challenge underscores the complexity of initializing the algorithm in 3D spaces, which is vital for the successful application of Monte Carlo localization methods in more complex environments.

In [70], the authors developed a branch-and-bound (BnB)-based 2D scan matching technique utilizing hierarchical occupancy grid maps of varying sizes. While this approach provides accurate and fast global localization on 2D maps, its processing time significantly increases when applied to 3D maps. In [71], the authors advanced this research by introducing a BnB-based method for 3D global localization, which more effectively addresses the challenges of extending the work of [70] to three-dimensional environments. However, these studies primarily focus on global localization issues without offering an integrated solution.

In summary, despite attempts to address the three major localization challenges simultaneously and the existence of various methods focusing on solving specific issues, there remains a significant research gap in developing an accurate and robust comprehensive localization system. This highlights the importance and innovative value of proposing new algorithms, such as the BB-RL algorithm introduced in this paper, aimed at enhancing the performance of indoor robot localization. The BB-RL algorithm seeks to overcome the limitations of existing solutions through innovative techniques and methods, providing a more comprehensive and effective solution to meet the demands of complex indoor environments for robot navigation.

3. System overview

3.1. Hardware setting

The hardware settings are shown in Figure 1. The adopted autonomous mobile robot (AMR) is a commercial differential wheeled logistics robot, model IR300, which is equipped with an Intel NUC8BEH minicomputer as the computing platform of the robot; two SICK TIM561 LiDAR for range measurements, which are deployed diagonally on the left front and bottom right of the robot and have a measurement frequency of up to 10 Hz; an inertial measurement unit model LMPS-be1, which is used for high-frequency linear acceleration and angular velocity measurement and can exhibit a measurement frequency of up to 200 Hz; two wheel encoders, which are used to measure the wheel speed with a measurement frequency of up to 50 Hz.

3.2. System architecture

The system architecture of the proposed algorithm is shown in Figure 2. The algorithm is composed of three parts: global localization, position tracking, and relocalization judgment.

Figure 1: IR300 robot, manufactured by Sunspeed Robotics Ltd, Co.

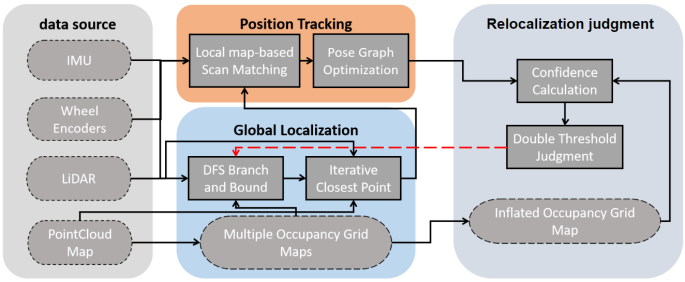


Figure 2: System architecture

In the localization process, we first obtain the robot's initial pose in the environment through the global localization algorithm, which is a two-stage matching algorithm composed of a branch-and-bound algorithm and an iterative closest point algorithm. After determining the robot's initial pose, we implement the position tracking algorithm, which uses the initial pose as the robot's initial state to realize local map-based scan matching. To effectively eliminate the accumulated errors between local maps, we maintain a global pose graph at the backend of the algorithm. When a valid loop closure is detected, the algorithm is implemented to correct the accumulated errors. Finally, we use a single thread to implement the relocalization judgment algorithm to monitor whether the robot can be located correctly when the position tracking algorithm is used. When the confidence of the current localization result is less than the set dual-thresholds, the global localization algorithm is called to reinitialize the algorithm.

4. Global Localization

Global localization, as an indispensable part of our algorithm, is mainly used to determine the robot's initial pose and ensure localization recovery when the robot is kidnapped. When the algorithm is implemented, we first convert the prebuilt point cloud map into multiple occupancy grid maps with different resolutions. Subsequently, we use the DFS-based branch-and-bound algorithm to accelerate the matching of the current LiDAR data with the occupancy grid maps. Finally, the iterative nearest point algorithm is used to continue the optimization on the computational results of the DFS-based branch-and-bound algorithm and ensure rapid convergence to obtain the optimal pose of the robot.

4.1. Global search using the branch-and-bound algorithm

We formulate global localization as a search problem on the occupancy grid map. The linear and angular search window sizes can be easily determined according to the map size. To ensure the search accuracy, we set the linear step size as the grid size. The angular step size can ensure that the farthest LiDAR point s_{max} moves once without exceeding the map resolution r. Thus, the angular step size ε can be estimated using the following equation:

$$\varepsilon \quad \arccos 1 - \frac{r^2}{2s_{max}^2} \tag{1}$$

Furthermore, the integral number of steps covering the set linear and angular search window sizes can be computed as:

$$s_x \quad \lceil \frac{S_x}{r} \rceil, s_y \quad \lceil \frac{S_y}{r} \rceil, s_\theta \quad \lceil \frac{S_\theta}{\varepsilon} \rceil \tag{2}$$

where S_x and S_y are the linear search window sizes in the x- and y-directions, respectively. S_θ is the angular search window size. s_x and s_y are the integral numbers of the linear steps in the x- and y-directions, respectively, and s_θ is the integral number of the angular steps. If the center of the occupancy grid map is assumed to be the origin of the search process, the search set can be defined as:

$$\boldsymbol{W} \quad \{-\frac{1}{2}s_x, ..., \frac{1}{2}s_x\} \times \{-\frac{1}{2}s_y, ..., \frac{1}{2}s_y\} \times \{-\frac{1}{2}s_\theta, ..., \frac{1}{2}s_\theta\} \tag{3}$$

Because the time to search an occupancy grid map increases exponentially with increasing map size, we apply the branch-and-bound algorithm to accelerate the search process. In practical applications, we build a global search tree to determine the initial pose for a given occupancy grid map, where each node in the tree represents a search result. The map search process is converted into node transversal in the search tree, and the target is to identify the leaf node with the best score.

In contrast to the breadth-first search-based branch-and-bound algorithm, which traverses most of the nodes in the search tree to identify the leaf node with the best score, we use the DFS-based branch-and-bound algorithm to promptly evaluate the nodes by performing a layer-by-layer search on multiple occupancy grid maps with low to high resolutions and prune the intermediate nodes that do not meet the boundary conditions and all the corresponding subnodes. Therefore, only a few nodes need to be traversed to identify a leaf node with the best score. The flow of the algorithm is illustrated in Figure 3.

Schematic of the DFS-based Branch-and-Bound Method (Search Tree Depth d 3). The root node is implicitly divided into different subnodes to form a set $\boldsymbol{N_0}$, and a node $\boldsymbol{n_0}$ is extracted to illustrate the algorithmic process.

First, we use the prebuilt point cloud map to create multiple occupancy grid maps with high to low resolutions. Specifically, we first rasterize the point cloud map according to the required highest resolution r_0. The probability value of each grid is averaged according to the number of point clouds in the grid, and the resulting occupancy grid map is denoted as map_0. Subsequently, according to the depth d of the global search tree, map_0 is downsampled $d-1$ times. The resolution of $map_i i$ $1, ..., d-1$ obtained by each downsampling is doubled to $2^i r_0$. Finally, we save these maps from low to high resolutions.

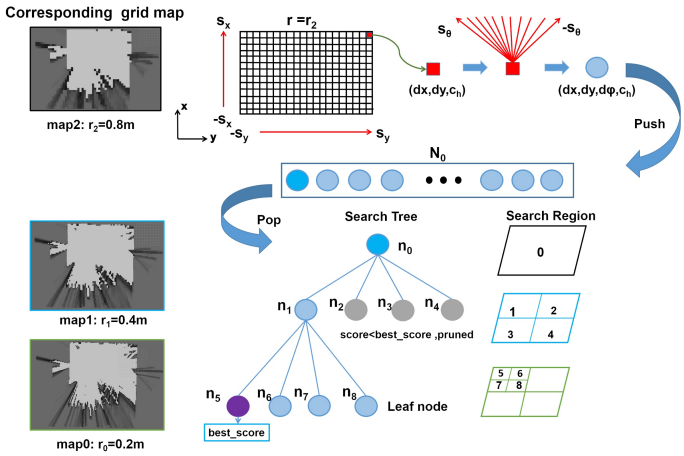


Figure 3: Schematic of the DFS-based branch-and-bound method (search tree depth d 3). The root node is implicitly divided into different subnodes to form a set $\boldsymbol{N_0}$, and a node $\boldsymbol{n_0}$ is extracted to illustrate the algorithmic process.

Second, we consider the search strategy. In the global search tree, the root node corresponds to the set of all possible solutions. We do not explicitly express this node but only branch it into a series of child nodes, which can be denoted as the set $\boldsymbol{N_0}$ of all possible solutions searched on map_{d-1}. The leaf nodes represent a possible solution searched on map_0. Each node $\boldsymbol{n_i}$ in the tree is represented as a tuple of integers:

$$\boldsymbol{n_i} \quad dx, dy, d\varphi, c_h \tag{4}$$

where dx and dy represent the position offsets in the x- and y-directions relative to the origin of the search process, respectively. $d\varphi$ is the rotation offset relative to the positive direction of the search process, and c_h represents the height of the search tree in which the node is located. Each node in the search tree is defined as a search area with a certain boundary.

Each node with $c_h > 1$ can branch into four child nodes of height $c_h - 1$:

$$\boldsymbol{N_n} \quad \{2dx, 2dx \;\; 1\} \times \{2dy, 2dy \;\; 1\} \times \{d\varphi\} \times \{c_h - 1\} \tag{5}$$

For each leaf node with c_h 0, branching cannot continue to generate new nodes. Thus, the search pose corresponding to the leaf node is a possible solution. When the leaf node with the best score is found, the optimal solution to the problem can be expressed as

$$\boldsymbol{\xi_n^*} \quad r_0 dx, r_0 dy, \varepsilon d\varphi \tag{6}$$

Finally, the upper bound calculation strategy is implemented. An excellent upper bound can help promptly identify the optimal solution to the problem. To ensure the accuracy of the upper bound, when building multiple occupancy grid maps with low to high resolutions, the probability value of each grid in $map_i i$ $1, ..., d-1$ is the maximum probability value of the corresponding $2^i \times 2^i$ grids in map_0. Therefore, the grids on the occupancy grid map with a lower resolution have a higher probability value:

$$Score\boldsymbol{n} \quad \sum_{i1}^{N} F_{Multimap}^{c_h} \boldsymbol{T_{\xi_n} s_i} \tag{7}$$

where $F_{Multimap}^{c_h}$ transforms the LiDAR point to the map frame to obtain the probability of the corresponding grid according to the prebuilt multiple occupancy grid maps. The search process is essentially a table lookup process, and thus, the computational complexity of the algorithm is always maintained in a constant range. The specific steps of the algorithm are shown in Algorithm 1.

Algorithm 1: Branch-and-bound Algorithm Based on Depth First Search.

Input: current period t, current scan $\boldsymbol{S_t}$, point cloud map m_p
Parameters: search tree depth d, search window sizes S_x, S_y, S_θ, occupancy grid map highest resolution r_0
Output: robot initial guess $\boldsymbol{\xi_n^*}$

Convert point cloud map m_p to multiple occupancy grid maps;
$s_x \leftarrow \lceil S_x r_{d-1} \rceil$;
$s_y \leftarrow \lceil S_y r_{d-1} \rceil$;
$\varepsilon \leftarrow \arccos 1 - r_0^2 2 s_{max}^2$;
$s_\theta \leftarrow \lceil S_\theta \varepsilon \rceil$;
$best\_score \leftarrow 0$;
$c_h \leftarrow d - 1$;
for $j_x \leftarrow -s_x$ **to** s_x **do**
  for $j_y \leftarrow -s_y$ **to** s_y **do**
    for $j_\theta \leftarrow -s_\theta$ **to** s_θ **do**
      $\boldsymbol{n} \leftarrow j_x, j_y, j_\theta, c_h$;
      Push $\boldsymbol{n}$ into $\boldsymbol{N_0}$;
    end
  end
end
initialization;
Initialize a priority queue $\boldsymbol{N}$ to save each node in $\boldsymbol{N_0}$ according to the score;
while $\boldsymbol{N} \neq$ *empty* **do**
  Pop the node $\boldsymbol{n}$ with the beat score from $\boldsymbol{N}$;
end
if $Score\boldsymbol{n} > best\_score$ **then**
  if $\boldsymbol{n}$ *is a leaf node* **then**
    $\boldsymbol{\xi_n^*} \leftarrow \boldsymbol{\xi_n}$;
    $best\_score \leftarrow Score\boldsymbol{n}$;
  end
  else
    Split $\boldsymbol{n} \rightarrow \boldsymbol{N_n}$:
    $\{2dx, 2dx \;\; 1\} \times \{2dy, 2dy \;\; 1\} \times \{d\varphi\} \times \{c_h - 1\}$;
    Compute the score of each node in $\boldsymbol{N_n}$;
    Store each node in $\boldsymbol{N_n}$ into $\boldsymbol{N}$ according to the score;
  end
end

4.2. Optimization of the initial pose using the iterative nearest point algorithm

Although the pose $\boldsymbol{\xi_n^*}$ specified by the DFS-based branch-and-bound algorithm has global optimality, the final search accuracy is inevitably limited by the highest resolution of the occupancy grid map. Hence, we use the iterative closest

point algorithm to further optimize $\boldsymbol{\xi}_n^*$.

The iterative nearest point algorithm calculates the rigid transformation matrix between the two sets of point clouds in an iterative manner. We convert the matching problem between the two sets of point clouds into a nonlinear least squares problem and iteratively compute the rigid transformation matrix around the initial value $\boldsymbol{\xi}_n^*$. We assume that the robot pose in the iterative process is $\boldsymbol{\xi}$ x, y, φ, the point in the point cloud map is $\boldsymbol{p}_i^{'}$, the current LiDAR point is $\boldsymbol{p}_i$, and the error function e is defined as

$$e\boldsymbol{\xi} \quad argmin\frac{1}{2}\sum_{i1}^{N}\left\|\boldsymbol{p}_i - exp\boldsymbol{\xi}^{\wedge}\boldsymbol{p}_i^{'}\right\|_2^2 \tag{8}$$

where $exp\cdot$ represents the exponential mapping of $so3 \rightarrow SO3$. We can use iterative algorithms (e.g., Gauss–Newton and Levenberg–Marquardt) to solve this problem. The Jacobian matrix of the iterative update process can be expressed as follows:

$$\frac{\partial e}{\partial \delta\xi} \quad -\boldsymbol{I} \quad exp\boldsymbol{\xi}^{\wedge}\boldsymbol{p}_i^{'\wedge} \tag{9}$$

The convergence speed of the iterative nearest point algorithm is affected by the maximum number of iterations and the robot pose difference calculated by two consecutive iterations. When the algorithm is used on hardware-constrained robot platforms, and it is necessary to consider the operating efficiency and localization accuracy, the convergence conditions can be alleviated. Our algorithm provides a satisfactory initial guess. Additionally, the number of point clouds involved in the matching is small. Hence, the convergence condition can be met after several iterations.

5. Position Tracking

The position tracking algorithm is of significance in enhancing the performance of the localization algorithm, especially in challenging circumstances such as those involving map expiration or environmental changes due to dynamic obstacles. In this paper, we use a scan matching method that aligns the current LiDAR data with the local map. The local map contains a certain number of LiDAR frames, which are expressed in an occupancy grid map. The map is updated continuously with each new LiDAR data. When the local map is built, it is added to the backend pose graph for optimization. The accumulated errors are corrected with the introduction of loop constraints to ensure the accuracy of the position tracking algorithm.

5.1. Frontend local map-based scan matching

The matching process involves inserting the current LiDAR data into the appropriate position in the local map. We formulate this process as a local nonlinear optimization problem, in which the LiDAR pose is optimized relative to the current local map. The problem is solved using the Gauss–Newton method. By iteratively optimizing the error function, a LiDAR pose with the highest probability is identified. In the optimization problem, $\boldsymbol{T}_{\xi}$ denotes the transformation matrix that transforms the LiDAR data into the local map. The error function can be expressed as:

$$E\boldsymbol{\xi} \quad argmin\sum_{i1}^{N}1 - F\boldsymbol{T}_{\xi}\boldsymbol{s}_i{}^2 \tag{10}$$

where $F : R_2 \rightarrow R$ represents a bicubic interpolation function that smooths the sum of the probability values of each LiDAR point in the local map. Specifically, we assume that $\boldsymbol{T}_{\xi}\boldsymbol{s}$ is defined as a point x, y in the two-dimensional plane. In this case, the bicubic interpolation function is:

$$Fx, y \quad \sum_{i1}^{3}\sum_{j0}^{3} fx_i, y_j Wx - x_i Wy - y_j \tag{11}$$

where fx_i, y_j is the probability of the four neighborhoods x_i, y_j around the point x, y, and $W\cdot$ represents the weight of the x_i, y_j interpolation on x, y, computed as:

$$Wx \begin{cases} a\ 2|x|^3 - a\ 3|x|^2\ 1 & for|x| \le 1 \\ a|x|^3 - 5a|x|^2\ 8a|x| - 4a & for1 < |x| < 2 \\ 0 & otherwise \end{cases} \tag{12}$$

where a takes values in the range $-0.75, -0.5$. Solving $E\boldsymbol{\xi}$ is a local nonlinear optimization problem. Thus, a satisfactory initial guess is critical. Before scan matching, we use a two-stage pose prediction method to obtain this initial guess. First, we use the extended Kalman filter (EKF) algorithm to fuse the wheel odometry and IMU data. The process uses these two types of data as observation information to update the state of the moment, as in [12].

Second, we use a multilocal-map-based scan matching method to further optimize the fusion result. The specific process is shown in Algorithm 2.

In the beginning, we perform a 2× downsampling on the local map to generate multiple local maps with resolutions ranging from high to low. Subsequently, we intend to find a LiDAR pose that maximizes the probabilities at the current LiDAR data in the lowest resolution local map. The initial pose is provided by the fusion result. Moreover, to ensure the matching accuracy, the pose obtained by matching against this local map is used as the initial value of the subsequent matching. This process is repeated until the matching against the highest resolution local map is realized, and the optimal initial guess is obtained.

After identifying the appropriate position, we insert the LiDAR data into the local map. This process updates the probability value of the corresponding grid. Each insertion of the LiDAR data is equivalent to adding an observation, and the result of the observation is saved using a hit set and miss set. According to the ray-tracing model, we use the projected LiDAR point as the hit point and save the grid point closest to this hit point in the hit set. Each grid point passing through the rays between the hit point and LiDAR data origin is saved in the miss set.

When the grid in the local map has never been observed previously, the probability is zero. When the grid is observed for the first time, it is assigned a probability value determined by its set (hit set or miss set). Each subsequent observation is based on the following formula to update the probability value of the grid:

Algorithm 2: Multilocal-map-based Scan Matching.

Input: current local map map_t, current scan $\boldsymbol{S_t}$
Parameters: search window sizes S_x, S_y, S_θ, downsampling times *num*
Output: ekf predicted pose $\boldsymbol{\xi_t^{ekf}}$, current predicted matching pose $\boldsymbol{\xi_t^{mul}}$

initialization;
2 times downsampling the current local map map_t *to form a set* $\{map_t^1, map_t^2, ..., map_t^{num}\}$;
$count \leftarrow 0$;
$best\_score \leftarrow 0$;
while $count \leq num$ **do**
  $r_{cur} \leftarrow$ resolution of $map_t^{num-count}$;
  $s_x \leftarrow \lceil S_x r_{cur} \rceil$;
  $s_y \leftarrow \lceil S_y r_{cur} \rceil$;
  $\varepsilon \leftarrow \arccos 1 - r_0^2 2 s_{max}^2$;
  $s_\theta \leftarrow \lceil S_\theta \varepsilon \rceil$;
  for $j_x \leftarrow -s_x$ **to** s_x **do**
    for $j_y \leftarrow -s_y$ **to** s_y **do**
      for $j_\theta \leftarrow -s_\theta$ **to** s_θ **do**
        $score \leftarrow \sum_{k1}^{K} F\boldsymbol{T}_{\boldsymbol{\xi_t^{ekf}} r_{cur} j_x, r_{cur} j_y, r_{cur} j_\theta} \boldsymbol{h_k}$;
        if $score > best\_score$ **then**
          $\boldsymbol{\xi_t^{ekf}} \leftarrow \boldsymbol{\xi_t^{ekf}} \; r_{cur} j_x, r_{cur} j_y, r_{cur} j_\theta$;
          $best\_score \leftarrow score$;
        end
      end
    end
  end
  $count \leftarrow count \; 1$;
end
$\boldsymbol{\xi_t^{mul}} \leftarrow \boldsymbol{\xi_t^{ekf}}$;

$$S \;\; S^- \;\; LogMeas \tag{13}$$

where S is the probability value of grid s after observation z, S^- is the probability value of grid s before observation, and $LogMeas$ represents the measurement model of the update process, which can be defined as

$$S \;\; logOdds \mid z \tag{14}$$

$$S^- \;\; logOdds \;\; log\frac{ps \;\; 1}{ps \;\; 0} \tag{15}$$

$$LogMeas \;\; log\frac{pz \mid s \;\; 1}{pz \mid s \;\; 0}, z \in \{0, 1\} \tag{16}$$

where the $logOdd$ function converts the product operation between the probability values into an addition operation, $ps \;\; 1$ is the probability that grid s is occupied before the observation, and $ps \;\; 0$ is the probability that grid s is free before the observation. According to the value of z, $LogMeas$ has two states. The specific value is determined by the sensor characteristics.

5.2. Backend pose graph optimization

The local map-based scan matching method can only decrease the short-term accumulated errors. However, the built local maps also accumulate errors over time, which can be optimized by building a global pose graph in the backend. In this process, we first use LiDAR frames that satisfy both rotation and translation conditions as key frames. Subsequently, we add all the keyframes and local maps to the pose graph as nodes to be optimized. Finally, the estimated trajectory is smoothed according to the constraints between the keyframe nodes and local map nodes. The optimization process of the pose graph is shown in Figure 4.

After a new loop constraint is constructed in the backend of the algorithm, we optimize the pose graph. We formulate the optimization process as a nonlinear least squares problem, in which the error term describes the error between the measured and estimated values. We consider the keyframe i and local map j as examples. The pose of keyframe i in the world frame is $\boldsymbol{\xi_i^s}$, and the pose of local map j in the world frame is $\boldsymbol{\xi_j^l}$. The error term can be expressed as

$$\boldsymbol{e}_{ij} \;\; z_{ij} - h\boldsymbol{\xi_i^s}, \boldsymbol{\xi_j^l} \tag{17}$$

where z_{ij} is the relative pose measured between keyframe i and local map j, calculated through loop-closure detection. $h\boldsymbol{\xi_i^s}, \boldsymbol{\xi_j^l}$ is the relative pose estimated between keyframe i and local map j, which represents the result of the local map-based scan matching.

The algorithm involves two types of constraints, namely, internal and loop constraints. The internal constraints are generated by keyframes and local maps that have subordination relationships. Specifically, the keyframes are inserted in the local map. In contrast, the loop constraints are generated by keyframes and other local maps, that is, the keyframes are associated with historical local maps. When more local maps are added to the pose graph, the time to identify the loop constraints gradually increases. Therefore, a DFS-based branch-and-bound algorithm is used to accelerate the search for loop constraints.

The process of loop-closure detection is similar to that of the DFS-based branch-and-bound algorithm used in global localization, except that the search range is changed from a global map to historical local maps. Hence, the search window no longer contains the prebuilt map but a partial area inside the local map. Because the frontend provides the current pose estimation $\boldsymbol{\xi_{front}}$ of the robot, we use the pose as the search origin to traverse the search space around it. The result $\boldsymbol{\xi_{loop}}$ is defined as

$$\boldsymbol{\xi_{loop}} \;\; \boldsymbol{\xi_{front}} \;\; r_0 dx, r_0 dy, \varepsilon d\varphi \tag{18}$$

If k represents the constraint between local map i and keyframe j, the error function can be expressed as

$$argmin \sum_{k1}^{K} \boldsymbol{e}_k^T \boldsymbol{\xi_k^s}, \boldsymbol{\xi_k^l} \boldsymbol{\Sigma_k^{-1}} \boldsymbol{e}_k \boldsymbol{\xi_k^s}, \boldsymbol{\xi_k^l} \tag{19}$$

where $\boldsymbol{\Sigma_k^{-1}}$ is the information matrix of the error term formed by keyframe i and local map j. The objective of optimizing the error function is to adjust $\boldsymbol{\xi^s}$ and $\boldsymbol{\xi^l}$ to minimize the trajectory errors formed by all nodes. Since no constraint relationship exists between each local map and keyframe in the pose graph, in solving the nonlinear optimization problem, considerable time is not required to calculate the Hessian matrix and only the pose increment needs to be solved via the Cholesky decomposition.

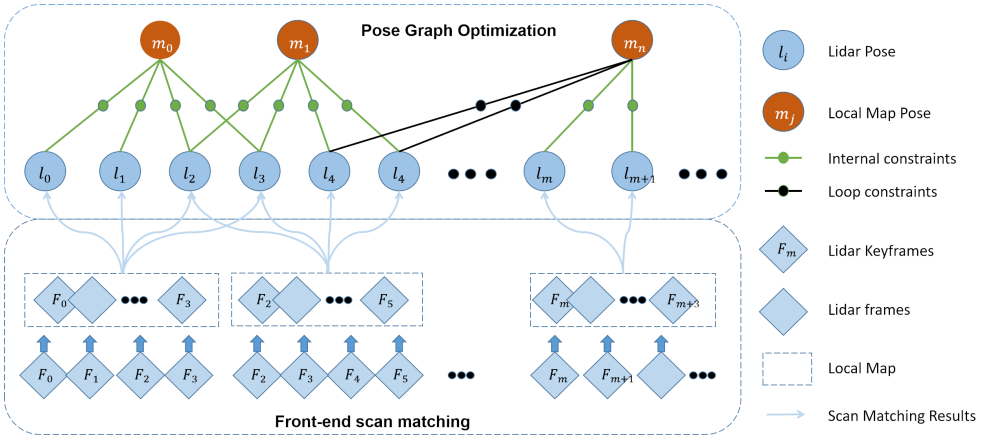


Figure 4: Schematic of the optimization of the backend pose graph.

6. Relocalization judgment

In numerous practical application scenarios, such as in warehouse logistics, robots are required to accomplish specific tasks within extensive workspaces. Due to the inability to form loop closures within short periods, robots tend to accumulate errors gradually. Furthermore, challenges arise when incorrect observational data leads to 'robot kidnapping', making it arduous to achieve localization recovery solely through position tracking algorithms.

In light of these challenges, we introduce an FSM (Finite State Machine)-based relocalization judgment algorithm. This algorithm initiates by acquiring the confidence level of the robot's current pose through the alignment of current LiDAR data with an inflated occupancy grid map. Subsequently, based on pre-set dual-threshold conditions, we assess the necessity to engage the global localization algorithm for timely localization recovery.

6.1. Confidence calculation and dual-threshold judgment

We use a method similar to the calculation of scores in scan matching to verify the pose $\boldsymbol{\xi_{pt}}$ obtained by the position tracking algorithm. In contrast to the point cloud registration algorithm that adopts the Euclidean distance to calculate the matching score between the two point clouds, we use the pose $\boldsymbol{\xi_{pt}}$ to project the current LiDAR data $\boldsymbol{S}$ onto the occupancy grid map and calculate the sum of the probability values of each LiDAR point $\boldsymbol{s_i}$ falling on the corresponding grid:

$$Score_{\boldsymbol{\xi_{pt}}} \quad \frac{1}{N}\sum_{i1}^{N} M\boldsymbol{T}_{\boldsymbol{\xi_{pt}}}\boldsymbol{s_i} \tag{20}$$

where $\boldsymbol{T}_{\boldsymbol{\xi_{pt}}}$ converts the current LiDAR data $\boldsymbol{S}$ from the LiDAR frame to the map frame, and $M\cdot$ is used to calculate the probability value of each LiDAR point projected onto the occupancy grid map.

In this process, the occupancy grid map is converted from the prebuilt point cloud map. The resolution of this map is the same as that of the local map generated by the position tracking algorithm.

In practical applications, since there are relatively few valid points in the LiDAR frame, the measurement error of each valid point affects the confidence calculation results. Considering this aspect, we use an inflated occupancy grid map instead of the original occupancy grid map to suppress the impact of LiDAR measurement errors.

In contrast to the cost map used to set the expansion areas to avoid robot collisions, we use the inflated occupancy grid map to reduce the error caused by noisy LiDAR measurements. When designing the inflated occupancy grid map, we first set the expansion radius r_{inf} according to the sensor range accuracy and extend it outward from the obstacle to obtain the expansion area according to r_{inf}. The grid probability in the expanded area is

$$P_{inf}x, y \quad e^{-k\delta d} \tag{21}$$

where δd is the distance between grid x, y and the obstacle, and k is the scale factor. When k is large, the grid probability $P_{inf}x, y$ decreases rapidly. The probability of $P_{inf}x, y$ is limited to the range $0, 1$. The process of generating an inflated occupancy grid map is shown in Figure 5. We update the confidence calculation formula as follows:

$$Score_{\boldsymbol{\xi_{pt}}} \quad \frac{1}{N}\sum_{i1}^{N} M^{InfMap}\boldsymbol{T}_{\boldsymbol{\xi_{pt}}}\boldsymbol{s_i} \tag{22}$$

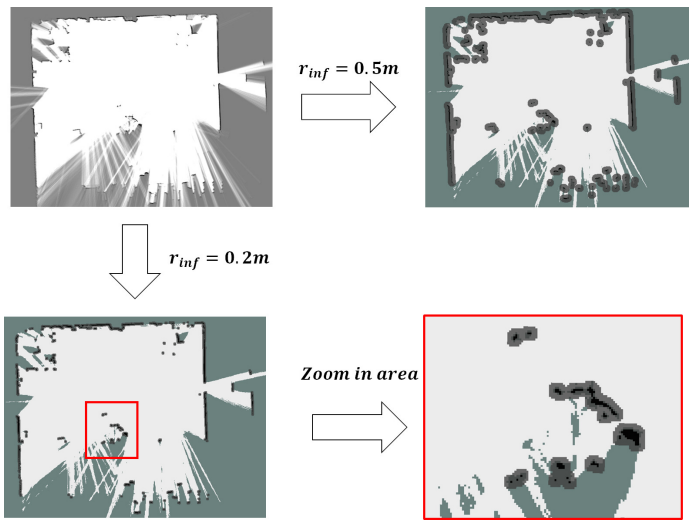


Figure 5: Process of generating an inflated occupancy grid map.

After calculating the confidence according to the above

formula, we use the dual-threshold judgment to evaluate the pose $\boldsymbol{\xi_{pt}}$.

1. When the confidence is greater than the set threshold T_{h2}, the LiDAR data are projected inside the expansion area of the map, and the errors of the confidence calculation are generated by the noisy LiDAR measurements.

2. When the confidence is between the two thresholds T_{h1} and T_{h2} $T_{h2} > T_{h1}$, the accumulated errors exceed expectations, and the robot kidnapping problem does not occur. Therefore, we call the global localization algorithm to complete the search in the local area near the pose $\boldsymbol{\xi_{pt}}$ to correct the accumulated errors.

3. When the confidence is less than the set threshold T_{h1}, the robot kidnapping problem is considered to occur. We invoke the global localization algorithm to search the whole map. Specifically, the search window of the branch-and-bound algorithm covers the occupancy grid map to complete the localization recovery.

6.2. Relocalization judgment based on finite state machine

To monitor the localization state in real time, we use the idea of a finite state machine to model the relocalization judgment process. The mathematical model for a certain finite state machine can be defined as

$$\boldsymbol{M} \quad \boldsymbol{Q}, \boldsymbol{\Sigma}, \delta, q_0, \boldsymbol{F} \tag{23}$$

where $\boldsymbol{Q}$ is a nonempty set consisting of a finite number of states. According to the results of the position tracking algorithm, the states of the whole algorithm are divided into three categories: normal localization q_{norm}, large localization error q_{err}, and localization failure q_{kid}, which correspond to three cases of the dual-threshold judgment. Therefore, $\boldsymbol{Q}$ can be defined as

$$\boldsymbol{Q} \quad q_{norm}, q_{err}, q_{kid} \tag{24}$$

where $\boldsymbol{\Sigma}$ represents the set of all inputs that can be accepted by each state, that is, the set of trigger conditions that cause the state transition. In this algorithm, we use the result of the dual-threshold judgment as the trigger condition. Additionally, we use e_{norm}, e_{err} and e_{kid} to represent the inputs of the algorithm in the transition between q_{norm}, q_{err} and q_{kid}. At this time, $\boldsymbol{\Sigma}$ is defined as

$$\boldsymbol{\Sigma} \quad e_{norm}, e_{err}, e_{kid} \tag{25}$$

where $\delta : \boldsymbol{Q} \times \boldsymbol{\Sigma} \rightarrow \boldsymbol{Q}$ represents the state transition function, which is mainly based on the current trigger condition e to complete the state transition of the algorithm from the current state q_{cur} to the second state q_{sec}:

$$q_{sec} \quad \delta q_{cur}, e \tag{26}$$

where q_0 is the initial state. $\boldsymbol{F}$ is the set of termination states, which is a subset of $\boldsymbol{Q}$ that represents that the algorithm is acceptable in this state (for instance, q_{norm}).

At the beginning of the algorithm operation, the robot is normally located. We first define the initial state q_0 as the state q_{norm} and subsequently determine the trigger condition according to the result of the confidence calculation.

1. If the result of the confidence calculation is greater than T_{h2}, the condition e_{norm} is triggered. The algorithm maintains the state q_{norm} and outputs the result of the position tracking algorithm.

2. When the result of the confidence calculation is between T_{h1} and T_{h2} $T_{h2} > T_{h1}$, the condition e_{err} is triggered. The algorithm executes the function $\delta q_{norm}, e_{err}$ to achieve the transition from q_{norm} to q_{err}, that is, the global localization algorithm is called to perform a search in the local range.

3. When the result of the confidence calculation is less than T_{h1}, the condition e_{kid} is triggered. The algorithm executes the function $\delta q_{norm}, e_{kid}$ for the transition between the two states of q_{norm} to q_{kid}. Specifically, the global localization algorithm is invoked to perform a search on the global map.

The state transition relationship in the finite state machine is shown in Figure 6. The specific steps of the FSM-based relocalization judgment algorithm are shown in Algorithm 3.

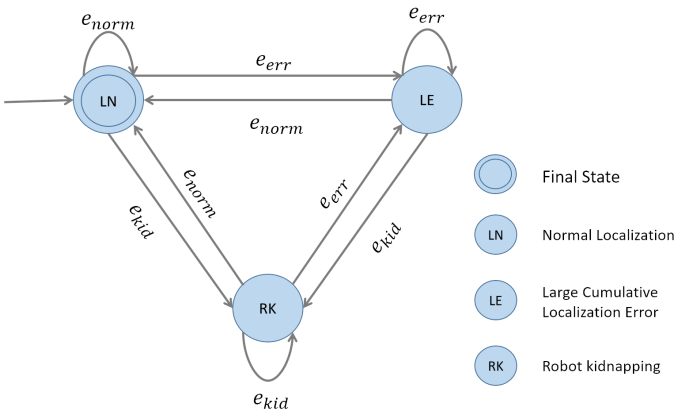


Figure 6: State transition relationship in the finite state machine.

This algorithm is expected to solve the problem of robot kidnapping. Hence, it is necessary to search for the best matches on the global map. To ensure the stability of the algorithm, we limit the number of calls to the global localization algorithm to manage the errors in the confidence calculation caused by environmental changes (such as dynamic environments). Our confidence calculation method averages the matching probabilities of each LiDAR point participating in the scan matching. The method exhibits a certain degree of robustness in scenarios involving slight environmental changes; however, its performance is limited in cases involving severe environmental changes. Thus, it is preferable to limit the number of calls to global localization. When the set maximum number of times is reached, the relocalization judgment algorithm is automatically terminated.

7. Experiments

As described in this section, we validate the robustness and accuracy of our algorithm through extensive experiments. First, we present the implementation details, including the experimental environment and preparation steps. Second, we describe the evaluation of our algorithm in a simulated laboratory environment and analysis of the performance of different parts. Finally, we assess the performance of our algorithm in an actual workshop environment.

Algorithm 3: Relocalization judgment based on finite state machine.

Input: current period t, current scan $\boldsymbol{S_t}$, inflated occupancy grid map m, position tracking pose $\boldsymbol{\xi_{pt}}$
Parameters: confidence thresholds T_{h1}, T_{h2}, Number of relocalization N_{rel}
Output: optimal robot pose $\boldsymbol{\xi_t^*}$

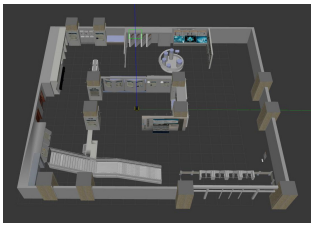
initialization;
$score \leftarrow 0$;
$StatusFlag \leftarrow false$;
$count \leftarrow 0$;
while $StatusFlag$ $false$ **do**
  $score \leftarrow Score\boldsymbol{\xi_{pt}}$;
  if $score < T_{h2}$ **then**
    if $T_{h1} \leq score \leq T_{h2}$ **then**
      $\boldsymbol{\xi_t^*} \leftarrow GlobalLocalization\boldsymbol{\xi_{pt}}, S_t$;
      $count \leftarrow count$ 1;
    end
    else
      $\boldsymbol{\xi_t^*} \leftarrow GlobalLocalizationS_t$;
      $count \leftarrow count$ 1;
    end
  end
  else
    $\boldsymbol{\xi_t^*} \leftarrow \boldsymbol{\xi_{pt}}$;
    $count \leftarrow 0$;
  end
  if $count \geq N_{rel}$ **then**
    $StatusFlag \leftarrow true$;
  end
end

7.1. Implementation Details

Using the Gazebo physical simulation platform, we build a virtual laboratory environment that mimics the layout and dimensions of the real-world laboratory. In such a typical structured environment, we use a simulated jackal robot with basic sensors (e.g., 2D LiDAR, IMU, and wheel encoders) to perform the experiments. To perform the assessment in an actual workshop environment, we use the IR300 commercial logistics robot to conduct the experiments. The environments are shown in Figure 7(a) and Figure 7(b).

In the preparation stage, we use an open-source 2D LiDAR SLAM algorithm to build a point cloud map of the environment. The process can be divided into three stages:

1. Data preprocessing: Raw sensor data for time synchronization are collected to alleviate the errors caused by the difference in the working frequency of different sensors;
2. Mapping: The handle is used to ensure that the robot can traverse the complete environment to build a point cloud map in real time;
3. Postprocessing: The built point cloud map is filtered to eliminate anomalies and outliers.



(a) Simulated laboratory environment with dimensions of 20 m×20 m.

(b) Actual workshop environment with dimensions of 30 m×60 m

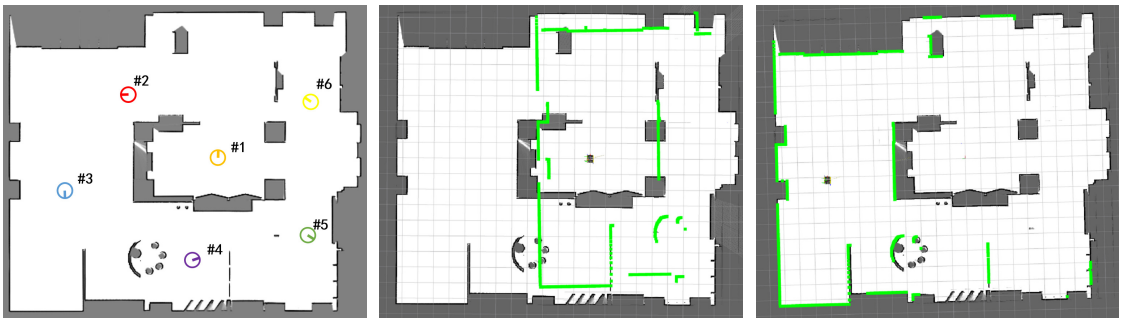
Figure 7: Experiment environment.

7.2. Localization experiment in the simulated laboratory environment

We first test the global localization in the simulated laboratory environment. The size of the laboratory is approximately 20 m×20 m; thus, we set the linear search window sizes in the x- and y-directions as 30 m, respectively, and the angular search window size is set as 2π. The depth of the search tree in the branch-and-bound algorithm is 7. Correspondingly, there exist seven built occupancy grid maps, in which the highest resolution of the occupancy grid map is r_0 0.4 m. To ensure that the iterative nearest point algorithm can achieve the highest accuracy, we set the maximum number of iterations as 100 and maximum tolerance of two consecutive iterations as 10^{-13}.

In the experiment, we select six positions on the map to test the performance of the algorithm. To uniformly cover the free space of the environment, the selected adjacent positions are separated by Δd 5 m, and the orientations of each position are uniformly distributed in $-\pi, \pi$, as shown in Figure 8(a). When the robot starts operating, it automatically implements the global localization algorithm to obtain the robot's initial pose based on the current LiDAR data, as shown in Figure 8(b) and Figure 8(c).

e_x and e_y denote the position error between the real position and estimated position of the robot, and e_φ represents the difference between the real and estimated orientations. In addition to these standard criteria, we consider the runtime and success rate of the algorithm. The runtime refers to the time from the beginning of the algorithm to the time at which the final result is obtained. The success rate describes the probability of successful localization at the specified position. When the error between the real position and estimated position of the robot is less than 0.05 m and the orientation error is less than 2°, the localization is considered successful. We perform 20 experiments for each specified position.



(a) Positions selected on the map for the global localization experiment (b) Before global localization at #3 (c) Global localization at #3

Figure 8: Evaluation of the global localization algorithm in the simulated laboratory environment.

In the test, we verify the performance of the proposed algorithm. Hence, a comparison experiment is not conducted for the following reasons:

1. The localization result is compared with the real position of the robot;

2. Few open-source algorithms can achieve global localization. Actual results for the few algorithms that can accomplish this function have been extensively reported. Therefore, the details do not need to be presented.

According to Table 1, the average orientation error is less than 0.2°, the average position errors in the x- and y-directions are less than 0.03 m and 0.01 m, respectively. As described in Section 4, the search accuracy of the branch-and-bound algorithm is limited by the highest resolution of the occupancy grid map (0.4 m). However, the two-stage matching algorithm achieves a localization accuracy that is higher than that of algorithms that use an occupancy grid map with a resolution of 0.05 m for scan matching. Moreover, we achieve a 100% localization success rate in each position.

The runtime varies considerably across positions (Figure 9). According to the runtime of each stage in the global localization algorithm, the most notable time consumption pertains to the determination of the initial pose by the branch-and-bound algorithm. In contrast, the runtime of the iterative closest point algorithm is stable and occupies only a small proportion. Although the runtime does not meet the requirements of real-time localization, considering the actual size of the map used in the search process, our algorithm can promptly find the global pose of the robot and dramatically decrease the time associated with redundant calculations.

In the algorithm, when the depth (d 7) is constant, the resolution r_0 of map_0 used in the branch-and-bound algorithm directly influences the localization accuracy and runtime. We analyze the impact of the different resolutions r_0 on the algorithm at position 4 $-6.33, 1.23, -45°$. The results are shown in Table 2. When r_0 is small, although the solution obtained by the branch-and-bound algorithm is closer to the optimal solution, the search time is large. In contrast, the proposed algorithm achieves a reasonable balance between the efficiency and localization accuracy. The localization result obtained by the proposed algorithm does not considerably fluctuate with the change in r_0, and the runtime is exponentially decreased.

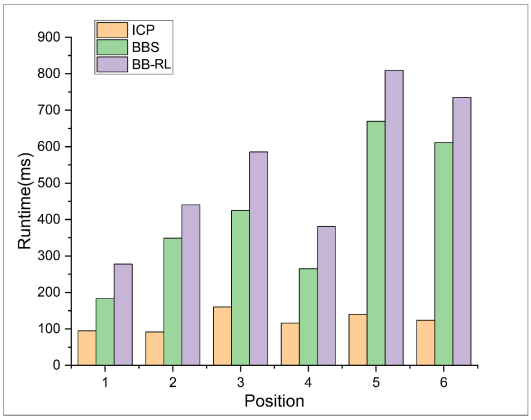


Figure 9: Runtime distribution for specific positions (BBS: branch-and-bound algorithm, ICP: iterative closest point algorithm, BB-RL: proposed algorithm).

To assess the accuracy of our algorithm, we conducted 50 experimental runs in the simulation environment, and for each run, we randomly selected a position on the map to measure the error. The results are shown in Figure 10. The average position errors in the x- and y-directions are 0.02037 m and 0.00648 m, respectively. The average orientation error is 0.00129 rad, and the average runtime is 576.35 ms. Additionally, the maximum position error in the x- and y-directions are 0.0317 m and 0.0185 m, respectively. The maximum orientation error is 0.00353 rad, and the maximum runtime is 1027.67 ms.

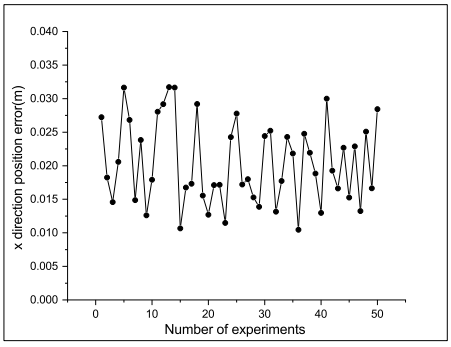
Next, we conduct the position tracking experiment. We assume that the robot's initial pose is known (automatically obtained by Gazebo). In the test, the robot moves in a circle around the indoor environment. The starting and ending points coincide. We evaluate the error of the robot between the starting and ending points. The process is shown in Figure 11. As a reference, we compare the AMCL and Cartographer frameworks to verify the accuracy of the algorithm.

In the parameter settings, the number of local maps for multi-local-map-based scan matching was established as 3. For loop-closure detection, the linear search window size was set at 20 m and the angular search window size for loop detection at 2π radians. Additionally, the search depth was defined as 7.

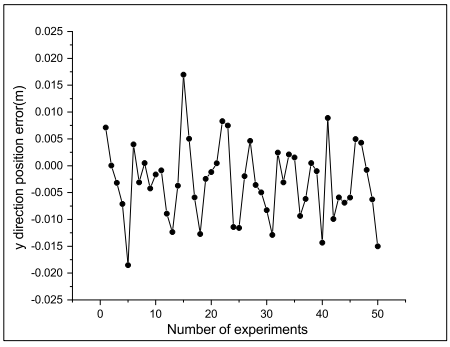
The trajectory of each algorithm is shown in Figure 12.

Table 1: Global localization results for specific positions in the simulated laboratory environment.

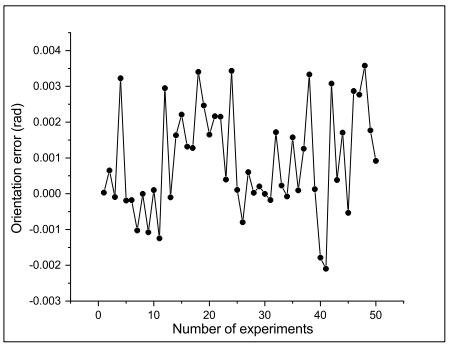
Position	$e_x m$	$e_y m$	$e_\varphi rad$	Runtimems	Success Rate%
#1	0.0249	-0.00631	0.000119	277.806	100
#2	0.0130	-0.00529	0.00123	440.500	100
#3	0.0191	-0.00455	0.000581	585.314	100
#4	0.0236	-0.00971	0.00267	380.872	100
#5	0.0246	-0.00933	0.000609	809.102	100
#6	0.0180	-0.00303	0.000237	734.903	100



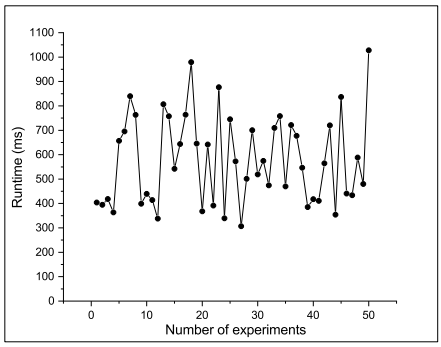
(a) Position error in the x-direction



(b) Position error in the y-direction

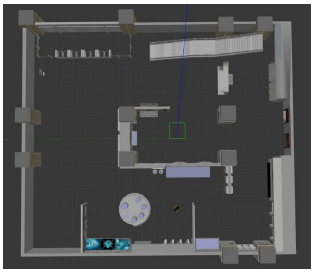


(c) Orientation error

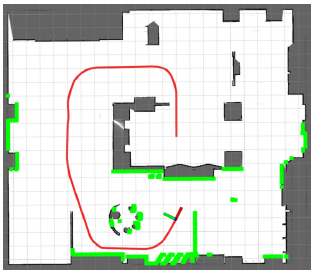


(d) Runtime

Figure 10: Experimental results of 50 positions randomly selected for global localization in the simulated laboratory environment.



(a) Operation of the robot in the simulated laboratory environment



(b) Real-time trajectory of the robot shown on the map

Figure 11: Evaluation of the position tracking algorithm in the simulated laboratory environment.

Table 2: Experimental results of global localization algorithm at position 4 with different resolutions r_0.

r_0		$e_x m$	$e_y m$	$e_\varphi rad$	Runtimems
0.5m	BB-RL	0.02319	0.00626	0.00426	329.356
	BBS	0.33	0.27	0.03045	262.846
0.4m	BB-RL	0.02236	0.00546	0.00271	411.043
	BBS	0.07	0.03	0.01378	295.019
0.3m	BB-RL	0.02332	0.00480	0.00609	647.166
	BBS	0.03	0.03	-0.03621	531.401
0.2m	BB-RL	0.02331	0.00633	0.00209	1069.53
	BBS	0.07	0.03	0.00163	1016.26
0.1m	BB-RL	0.02019	0.00332	0.00584	6069.81
	BBS	0.03	0.03	-0.00288	5980.87
0.05m	BB-RL	0.02148	0.00370	0.00198	31851
	BBS	0.02	0.02	0.00211	31808.9

Results obtained using AMCL, Cartographer, and the proposed algorithm are relatively close to the ground truth because the sensor data obtained in the simulation environment are ideal, and no sensor failures or other emergencies occur. However, according to the analysis of trajectory details, the proposed algorithm fits the ground truth more closely. According to the trajectory error comparison shown in Table 3, the proposed algorithm outperforms the compared algorithms in terms of the accuracy. The Figure 13 shows the time-based error of the position on both the x-axis and y-axis, as well as the orientation error during the position tracking experiment.

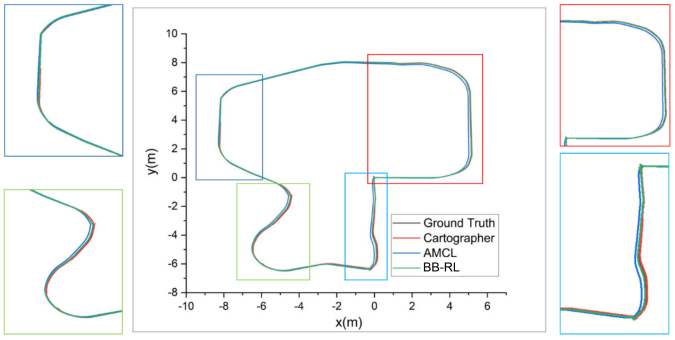
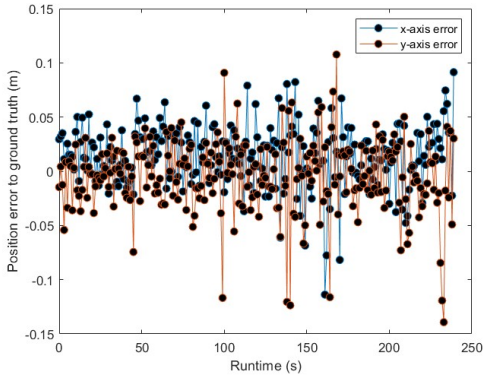


Figure 12: Comparison of trajectories of different position tracking algorithms in a simulated laboratory environment.

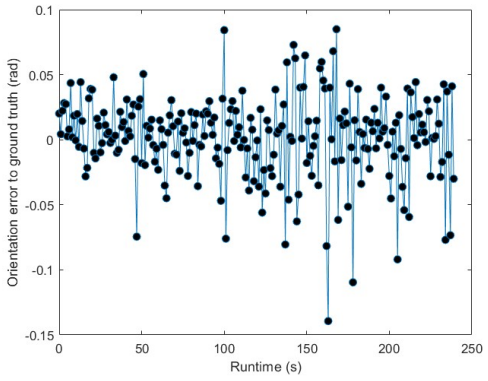
Finally, the relocalization experiment is conducted. Since the correction of the accumulated errors is reflected in the experimental results of the position tracking, we test only the localization recovery ability of the algorithm in the case of robot kidnapping. First, we initialize the robot and control it to move in the environment. Second, we suddenly move the robot to positions A, B, C, and D (Figure 14) to artificially create a robot kidnapping situation to verify the effectiveness of the relocalization. Due to only a few existing open-source algorithms can solve the robot kidnapping problem. Additionally, no uniform standard for the experimental procedure exists. Hence, we do not conduct a comparison experiment in this test.

Table 3: Comparison of the position tracking error in the simulated laboratory environment.

	$e_x m$	$e_y m$	$e_\varphi rad$
BB-RL	**0.01398**	**-0.00176**	**0.00108**
AMCL	-0.06515	-0.070497	-0.0278
Cartographer	-0.05141	0.043688	0.00232



(a) Position error in the x-direction and y-direction

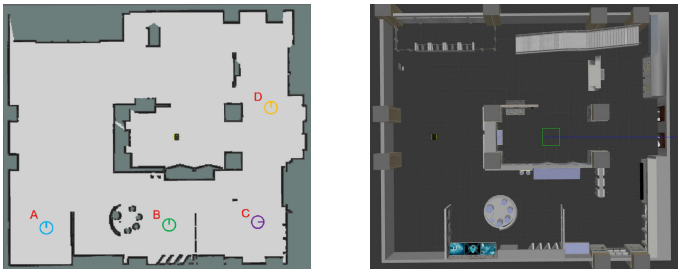


(b) Orientation error in the y-direction

Figure 13: Time-based error analysis of the BB-RL position tracking algorithm compared to ground truth data.

Before the test, to ensure that the map has the same resolution as that of the local map used in the position tracking

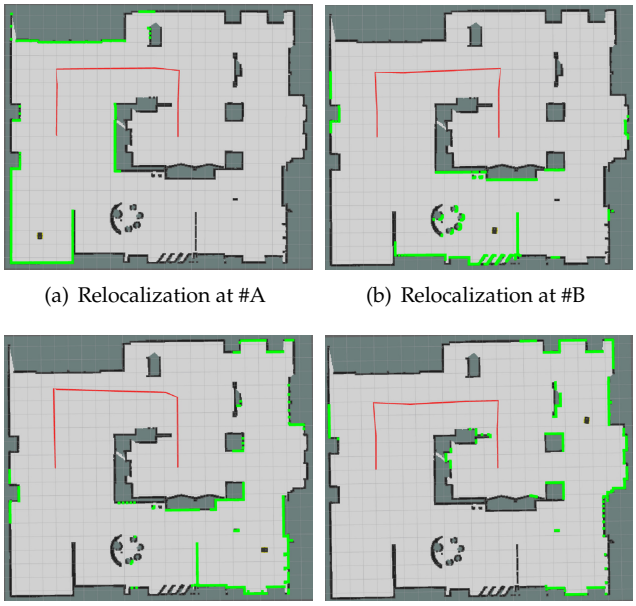
algorithm, we convert the prebuilt point cloud map into an occupancy grid map with a resolution of 0.05 m. According to the range accuracy of the LiDAR, the expansion radius r_{inf} is set as 0.1 m, the scale factor is set as 1, and the thresholds T_{h1} and T_{h2} are set as 0.5 and 0.8. The experimental results are shown in Figure 15.



(a) Specific positions to which the robot travels

(b) Position at which robot kidnapping occurs

Figure 14: Process of relocalization experiment in the simulated laboratory environment.



(a) Relocalization at #A

(b) Relocalization at #B

(c) Relocalization at #C

(d) Relocalization at #D

Figure 15: Relocalization results for positions A, B, C, and D following a robot-kidnapping.

The quantitative results are shown in Table 4. Among these results, the average position errors in the x- and y-directions are less than 0.03 m and 0.01 m, respectively, the orientation error is less than 0.1°, the runtime is within 600 ms, and the success rate at each position is consistent with global localization, remaining at 100%. From the overall perspective, the relocalization results are similar to those of the global localization in the simulation environment. When the relocalization judge algorithm is used to identify if the robot is kidnapped, localization recovery can be effectively realized by calling the global localization algorithm.

7.3. Localization experiment in the actual workshop environment

A global localization experiment is conducted in the actual workshop environment. In this experiment, the parameters of the branch-and-bound algorithm are changed. Because the size of the workshop is approximately 60 m ×30 m, we set the linear search window sizes in the x- and y-directions as 70 m and 40 m, respectively. All other parameter settings are the same as those in the global localization experiment in the simulated laboratory environment.

Similarly, we select six positions on the map to analyze the performance of the algorithm. Each adjacent position is separated by Δd 15 m, and the orientation of each position is uniformly distributed in $-\pi, \pi$. The selected positions are shown in Figure 16. The localization process of position 3 is shown in Figure 17 and Figure 18. The evaluation criteria and number of experiments are the same as those in the global localization experiment in the simulated laboratory environment.

Figure 16: Positions selected on the map for the global localization experiment.

Figure 17: Robot-kidnapping on position #3.

Figure 18: Global localization on position #3.

According to Table 5, the average position errors in the x- and y-directions are less than 0.032 m and 0.02 m, respectively, and the average orientation error is less than 1.2°. Compared with the experimental results of global localization in the simulated laboratory environment, the error of global localization in the workshop environment is significantly larger. The sensor noise and interference of the dynamic environment in the actual environment are more

Table 4: Relocalization results for specific positions in the simulated laboratory environment.

Position	$e_x m$	$e_y m$	$e_\varphi rad$	Runtimems	Succuess Rate%
A	0.02448	0.000870	-0.000313	401.946	100
B	0.02017	-0.00485	0.000605	376.736	100
C	0.02452	0.00723	-0.00116	538.637	100
D	0.02205	0.00374	0.00114	329.169	100

Table 5: Global localization results for specific positions in the actual workshop environment.

Position	$e_x m$	$e_y m$	$e_\varphi rad$	Runtimems	Succuess Rate%
#1	0.02261	-0.00402	0.00711	459.686	100
#2	0.02289	-0.00326	0.00884	605.213	100
#3	0.02382	0.01919	0.01673	633.351	95
#4	0.03064	0.01212	0.01193	1260.876	90
#5	0.02819	0.00895	0.00416	833.633	95
#6	0.02775	-0.01226	0.01938	671.117	95

unpredictable than those in the simulation environment and directly affect the localization accuracy.

The success rate is slightly decreased at positions 3-6 because the current LiDAR data tend to produce mismatches with the occupancy grid map. The runtime associated with each stage in the global localization algorithm (Figure 19) shows that the overall runtime at each position increases. Especially, at position 4, the overall running time is 1260.87 ms, 1226.81 ms of which correspond to the branch-and-bound algorithm. This finding demonstrates that most of the time consumed by the global localization algorithm pertains to the branch-and-bound algorithm implementation.

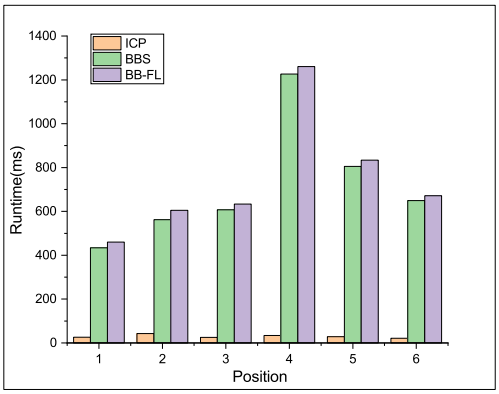


Figure 19: Runtime distribution for specific positions.

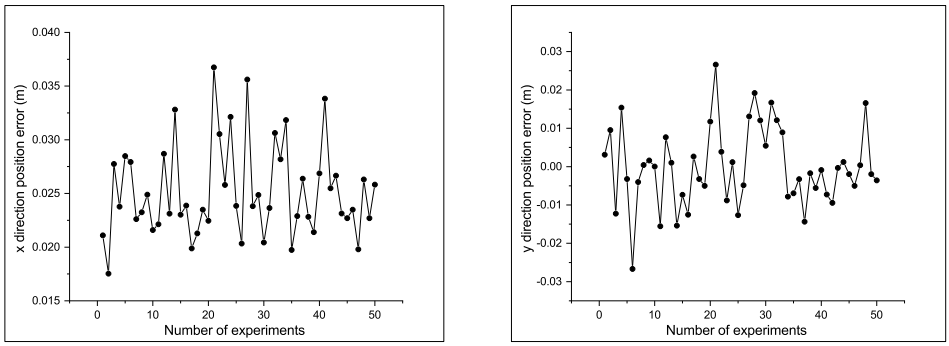
Moreover, in this experiment, we test the impact of different resolutions r_0 of map_0 used in the branch-and-bound algorithm on the localization accuracy and runtime when the depth (d 7) remains unchanged. The experimental results at position 1 $-0.43, -0.365, 0°$ are shown in Table 6. Compared with the results of the simulated laboratory environment, the runtime at different resolutions r_0 is higher due to the increased size of map_0. However, the proposed algorithm exhibits similar localization accuracies at different resolutions r_0. Therefore, we can choose r_0 0.4 m to balance localization efficiency and accuracy.

Table 6: Experimental results of global localization algorithm at position 1 with different resolutions r_0.

r_0		$e_x m$	$e_y m$	$e_\varphi rad$	Runtimems
$0.5m$	BB-RL	0.02313	-0.00623	0.00505	398.397
	BBS	-0.145	0.31	0.0333	369.458
$0.4m$	BB-RL	0.02281	-0.00172	0.00545	532.387
	BBS	-0.145	0.11	0.0133	507.262
$0.3m$	BB-RL	0.02273	-0.00901	0.00462	934.256
	BBS	-0.145	-0.09	0.03	910.888
$0.2m$	BB-RL	0.02205	-0.01520	0.00489	1878.96
	BBS	0.055	-0.09	0.00667	1864.33
$0.1m$	BB-RL	0.01898	-0.01287	0.00465	14806.3
	BBS	0.055	0.01	0.0133	14796.1
$0.05m$	BB-RL	0.02157	-0.00126	0.00537	296631
	BBS	0.005	0.01	0.01167	296612

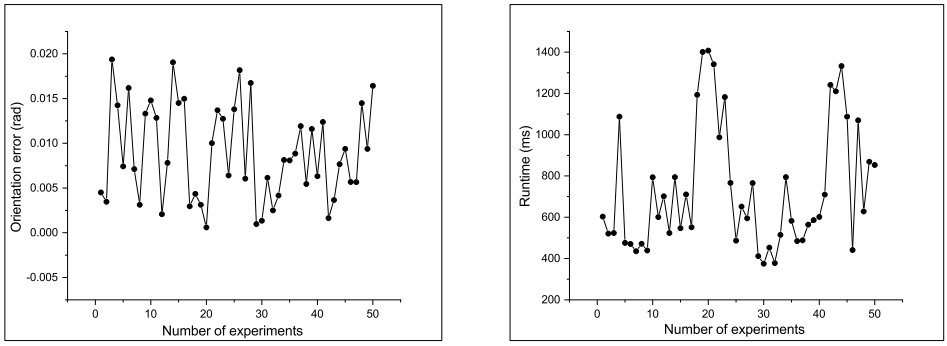
To assess the accuracy of the global localization algorithm, we conducted 50 experiments in a real-world environment. For each experiment, we randomly selected a position on the map to evaluate the error. The results are shown in Figure 20. The average position errors in the x- and y-directions are 0.02516 m and 0.0079 m, respectively. The average orientation error is 0.0089 rad, and the average runtime is 734.18 ms. Additionally, the maximum position errors in the x- and y-directions are 0.03675 m and 0.02669 m, respectively. The maximum orientation error is 0.0193 rad, and the maximum runtime is 1407.48 ms. Compared with the results in the simulated laboratory environment, the position error and runtime are higher, although the actual engineering needs can still be satisfied.

Next, we perform the position tracking experiment. First, we assume that the robot's initial pose is the origin of the map in this experiment. Subsequently, we control the robot to move in a circular path in the workshop to return to the starting point. Finally, the error between the starting point and ending point is calculated as the accuracy criterion. As a reference, we compare the results of EKF fused with IMU and wheel odometry, AMCL, and Cartographer to verify the accuracy of the algorithm. The experiment is shown in Figure 21 and Figure 22 .



(a) Position error in the x-direction

(b) Position error in the y-direction



(c) Orientation error

(d) Runtime

Figure 20: Experimental results of 50 positions randomly selected for global localization in the actual workshop environment.

Figure 21: Operation of the robot in the actual workshop environment

Figure 22: Real-time trajectory of the robot shown on the map in the actual workshop environment.

All the parameter settings are the same as those in the position tracking experiment in the simulated laboratory environment. The trajectory of each algorithm is shown in Figure 23. Notably, (1) the trajectory error associated with the EKF fusion is the largest; (2) there exists a certain deviation in the local details between each trajectory; and (3) the trajectory of the AMCL near the starting point is not closed.

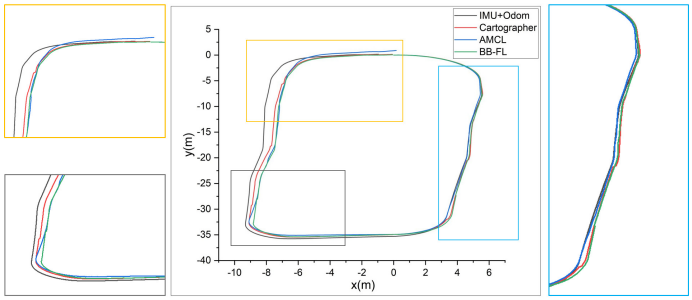


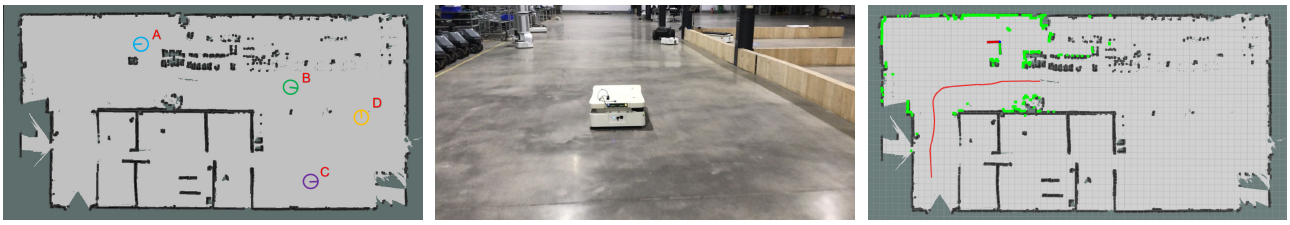
Figure 23: Comparison of trajectories of different position tracking algorithms in an actual workshop environment.

According to Table 7, the position error in the x-direction of EKF fusion is approximately 1 m, the position error in the y-direction of AMCL is approximately 0.9 m, and the orientation error of AMCL exceeds 4.5°. In contrast, the proposed algorithm achieves satisfactory results in all aspects: the position errors in the x- and y-directions are both less than 0.05 m, and the orientation error is less than 1°.

Table 7: Comparison of position tracking errors in the actual workshop environment.

	$e_x m$	$e_y m$	$e_\varphi rad$
BB-RL	**-0.03780**	**0.04696**	0.01111
AMCL	0.16055	0.87677	0.08407
Cartographer	-0.08248	0.12581	**0.01017**
IMUOdom	-0.96486	0.16832	0.05129

Finally, a relocalization experiment is conducted in the



(a) Specific positions to which the robot travels (b) Position at which robot kidnapping occurs (c) Relocalization at #A

(d) Relocalization at #B (e) Relocalization at #C (f) Relocalization at #D

Figure 24: Relocalization results for positions A, B, C, and D (the red line represents the trajectory of the robot before robot kidnapping occurs, the measurement data of the LiDAR are represented by the green line, and the position of the robot is represented by the orthogonal coordinate axes)

Table 8: Relocalization results for specific positions in the actual workshop environment.

Position	$e_x m$	$e_y m$	$e_\varphi rad$	Runtimems	Succuess Rate%
A	0.02819	0.00895	0.00416	904.561	95
B	0.02686	-0.00881	0.00632	630.238	95
C	0.02246	-0.00195	0.00937	374.754	100
D	0.02302	-0.00736	0.00207	434.718	100

actual workshop environment. The same experimental method as that of the relocalization experiment in the simulated laboratory environment is followed: First, the robot is controlled to move in the workshop through the handle. Second, a robot kidnapping situation is created by artificially moving the robot to positions A, B, C, and D (Figure 24(a) and Figure 24(b)). Finally, the position error and orientation error of different positions are calculated. In terms of the parameter settings, the thresholds T_{h1} and T_{h2} are set as 0.5 and 0.75, respectively. The resolution of the inflated occupancy grid map is 0.05 m, the expansion radius r_{inf} is set as 0.2 m, and the scale factor k is set as 1. The experimental results are shown in Figure 24(c), Figure 24(d), Figure 24(e) and Figure 24(f).

According to Table 8, the error in the actual workshop environment is higher than that in the simulated laboratory environment. At position A, the position error in the x-direction exceeds 0.028 m, the orientation error exceeds 0.2°, and the runtime is close to 1 s. Additionally, the runtime at positions C and D is significantly decreased with values of only 374.754 ms and 434.718 ms, respectively. For the success rate, relocalization failures occur at positions A and B. However, overall, the success rate is maintained at each selected position.

Building on the introduction of the Branch-and-Bound for Robust Localization (BB-RL) algorithm, the experimental findings can be effectively summarized. The BB-RL algorithm offers a potent solution for indoor robot localization by harmoniously integrating position tracking, global localization, and the resolution of the kidnapped robot dilemma within a cohesive framework. The evaluation shows that BB-RL achieves a balance among speed, accuracy, and robustness, establishing it as an effective and practical choice for indoor robot localization scenarios.

In summary, the proposed trajectory aligns more closely with the ground truth compared to those generated by other compared algorithms. The BB-RL algorithm surpasses competing algorithms in accuracy. Regarding the kidnapping problem, robots equipped with BB-RL successfully overcome localization failures, maintaining a commendable success rate. The effectiveness of the BB-RL algorithm in solving the three core localization challenges has been confirmed in real-world settings, achieving sustained accuracy and an appropriate execution frequency. This underscores the algorithm's viability and efficiency in practical applications, particularly in navigating and localizing within indoor environments.

8. Conclusion and Future Work

A robust and accurate localization is crucial for effective path planning, precise motion control, and reliable obstacle avoidance in the field of autonomous robotics. Recognizing the need for accurate and robust localization in real-world applications, this paper presents a BB-RL (Branch-and-Bound for Robust Localization) algorithm for indoor mobile robots. Its novelty lies in the comprehensive and integrated approach to addressing the three key localization tasks: global

localization, position tracking, and the kidnapped robot problem.

The approach begins with a two-stage global localization algorithm to determine the robot's initial pose. A DFS-based branch-and-bound algorithm ensures the search solution is globally optimal. To achieve localization precision beyond grid resolution, the iterative closest point (ICP) algorithm refines this solution locally.

For continuous position tracking, a local map-based scan matching technique is used. To achieve reliable results, a two-tier prediction method combining an Extended Kalman Filter (EKF) with multi-local map-based scan matching is proposed, ensuring initial guesses converge to the global optimum. Additionally, a global pose graph is constructed to minimize accumulated errors across local maps, while a DFS-based branch-and-bound algorithm accelerates loop-closure detection.

Long-term stability of the algorithm is maintained through an innovative Finite State Machine (FSM)-based relocalization judgment method, which uses an inflated occupancy grid map to reduce LiDAR measurement noise effects on confidence calculations. A dual-threshold judgment strategy accurately identifies the robot's localization state, triggering the global localization algorithm as needed for timely localization recovery.

In conclusion, our algorithm shows out for its robustness, scalability, and practicality, underscored by its fast processing capabilities. Extensively tested in both simulated laboratory environments and real-world workshops, it has also been successfully implemented on a commercial logistics robot platform. This deployment demonstrates not only its high localization accuracy but also its robust and rapid performance in diverse operational contexts.

Finally, we have underscored the advantages of our localization framework, especially in indoor environments prone to localization difficulties, such as logistics warehouses and factory inspections. These environments require a robust and accurate localization algorithm. By integrating existing sensor data with advanced algorithms, our framework significantly improves localization accuracy and robustness in these complex scenarios.

In the future, our research will focus on utilizing a broader array of features for robot localization, including the features from 3D point cloud maps and camera sensors. These data types promise to enhance localization accuracy by providing a richer set of environmental information. However, incorporating these algorithms and features is expected to increase computational demands. A key direction for our future work will be to find a balance between integrating these diverse and multi-dimensional features and maintaining efficient processing speeds. We aim to integrate 3D point cloud features for improved relocalization without compromising computational efficiency.

Another aspect of our future work will address the challenges posed by complex, dynamic environments, such as scenarios where robots are surrounded by crowds. Identifying the cause of localization failures—whether due to actual kidnapping scenarios or temporary disruptions caused by dynamic environmental factors—and deciding whether to initiate relocalization presents a challenge we plan to address. This involves differentiating between true kidnapping situations and temporary conditions caused by dynamic environments, thereby guiding the decision on whether relocalization is necessary.

This comprehensive approach, leveraging a variety of data sources and technologies, is designed to ensure that localization challenges, even in the most demanding environments, can be effectively addressed. Our goal is to provide a more comprehensive and reliable solution for indoor robot localization, overcoming current limitations and preparing for future challenges.

Conflict of Interest The authors declare no conflict of interest.

References

[1] DeSouza, Guilherme N and Kak, Avinash C, "Vision for mobile robot navigation: A survey", *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[2] Georgiev, Atanas and Allen, Peter K, "Localization methods for a mobile robot in urban environments", *IEEE Transactions on Robotics*, vol. 20, no. 5, pp. 851–864, 2004.

[3] Alatise, Mary B and Hancke, Gerhard P, "A review on challenges of autonomous mobile robot and sensor fusion methods", *IEEE Access*, vol. 8, pp. 39830–39846, 2020.

[4] Altan, Aytaç and Bayraktar, Köksal and Hacıoğlu, Rıfat, "Simultaneous localization and mapping of mines with unmanned aerial vehicle", *2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1433–1436, 2016, doi:10.1109/SIU.2016.7496019.

[5] Aytaç Altan and Rıfat Hacıoğlu, "Model predictive control of three-axis gimbal system mounted on uav for real-time target tracking under external disturbances", *Mechanical Systems and Signal Processing*, vol. 138, p. 106548, 2020, doi:https://doi.org/10.1016/j.ymssp.2019.106548.

[6] Cox, Ingemar Johansson and Wilfong, Gordon Thomas, *Autonomous robot vehicles*, Springer, Schweiz, 1990.

[7] Thrun, Sebastian and Beetz, Michael and Bennewitz, Maren and Burgard, Wolfram and Cremers, Armin B and Dellaert, Frank and Fox, Dieter and Haehnel, Dirk and Rosenberg, Chuck and Roy, Nicholas and others, "Probabilistic algorithms and the interactive museum tour-guide robot minerva", *The international journal of robotics research*, vol. 19, no. 11, pp. 972–999, 2000.

[8] Alqahtani, Ebtesam J and Alshamrani, Fatimah H and Syed, Hajra F and Alhaidari, Fahd A, "Survey on algorithms and techniques for indoor navigation systems.", *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pp. 1–9, 2018, doi:10.1109/NCG.2018.8593096.

[9] Zafari, Faheem and Gkelias, Athanasios and Leung, Kin K, "A survey of indoor localization systems and technologies", *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019, doi:10.1109/COMST.2019.2911558.

[10] Liu, Wenxin and Caruso, David and Ilg, Eddy and Dong, Jing and Mourikis, Anastasios I and Daniilidis, Kostas and Kumar, Vijay and Engel, Jakob, "Tlio: Tight learned inertial odometry", *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5653–5660, 2020, doi:10.1109/LRA.2020.3007421.

[11] Brossard, Martin and Bonnabel, Silvere, "Learning wheel odometry and imu errors for localization", *2019 International Conference on Robotics and Automation (ICRA)*, pp. 291–297, 2019, doi:10.1109/ICRA.2019.8794237.

[12] Moore, Thomas and Stouch, Daniel, "Intelligent autonomous systems 13", pp. 335–348, Springer, Padua, 2016, doi:10.1007/978-3-319-08338-4_25.

[13] He, Chengyang and Tang, Chao and Yu, Chengpu, "A federated derivative cubature kalman filter for imu-uwb indoor positioning", *Sensors*, vol. 20, no. 12, p. 3514, 2020, doi:10.3390/s20123514.

[14] Liu, Fei and Li, Xin and Wang, Jian and Zhang, Jixian, "An adaptive uwb/mems-imu complementary kalman filter for indoor location in nlos environment", *Remote Sensing*, vol. 11, no. 22, p. 2628, 2019, doi:10.3390/rs11222628.

[15] Cui, Jishi and Li, Bin and Yang, Lyuxiao and Wu, Nan, "Multi-source data fusion method for indoor localization system", *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 29–33, 2020, doi:10.1109/ICCC49849.2020.9238826.

[16] Yang, Xiaofei and Wang, Jun and Song, Dapeng and Feng, Beizhen and Ye, Hui, "A novel nlos error compensation method based imu for uwb indoor positioning system", *IEEE Sensors Journal*, vol. 21, no. 9, pp. 11203–11212, 2021, doi:10.1109/JSEN.2021.3061468.

[17] Davison, Andrew J and Reid, Ian D and Molton, Nicholas D and Stasse, Olivier, "Monoslam: Real-time single camera slam", *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007, doi:10.1109/TPAMI.2007.1049.

[18] Pire, Taihú and Fischer, Thomas and Castro, Gastón and De Cristóforis, Pablo and Civera, Javier and Berlles, Julio Jacobo, "S-ptam: Stereo parallel tracking and mapping", *Robotics and Autonomous Systems*, vol. 93, pp. 27–42, 2017, doi:10.1016/j.robot.2017.03.019.

[19] Kerl, Christian and Sturm, Jürgen and Cremers, Daniel, "Dense visual slam for rgb-d cameras", *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2100–2106, 2013, doi:10.1109/IROS.2013.6696650.

[20] Mur-Artal, Raul and Montiel, Jose Maria Martinez and Tardos, Juan D, "Orb-slam: a versatile and accurate monocular slam system", *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015, doi:10.1109/TRO.2015.2463671.

[21] Engel, Jakob and Koltun, Vladlen and Cremers, Daniel, "Direct sparse odometry", *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017, doi:10.1109/TPAMI.2017.2658577.

[22] Engel, Jakob and Schöps, Thomas and Cremers, Daniel, "Lsd-slam: Large-scale direct monocular slam", *European conference on computer vision*, pp. 834–849, 2014, doi:10.1007/978-3-319-10605-2_54.

[23] Forster, Christian and Pizzoli, Matia and Scaramuzza, Davide, "Svo: Fast semi-direct monocular visual odometry", *2014 IEEE international conference on robotics and automation (ICRA)*, pp. 15–22, 2014, doi:10.1109/ICRA.2014.6906584.

[24] Campos, Carlos and Elvira, Richard and Rodríguez, Juan J Gómez and Montiel, José MM and Tardós, Juan D, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam", *IEEE Transactions on Robotics*, 2021, doi:10.1109/TRO.2021.3075644.

[25] Schubert, David and Demmel, Nikolaus and von Stumberg, Lukas and Usenko, Vladyslav and Cremers, Daniel, "Rolling-shutter modelling for direct visual-inertial odometry", *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2462–2469, 2019, doi:10.1109/IROS40897.2019.8968539.

[26] Li, Guangqiang and Yu, Lei and Fei, Shumin, "A deep-learning real-time visual slam system based on multi-task feature extraction network and self-supervised feature points", *Measurement*, vol. 168, p. 108403, 2021, doi:10.1016/j.measurement.2020.108403.

[27] Kang, Rong and Shi, Jieqi and Li, Xueming and Liu, Yang and Liu, Xiao, "Df-slam: A deep-learning enhanced visual slam system based on deep local features", *arXiv preprint arXiv:1901.07223*, 2019.

[28] Li, Ruihao and Wang, Sen and Long, Zhiqiang and Gu, Dongbing, "Undeepvo: Monocular visual odometry through unsupervised deep learning", *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 7286–7291, 2018, doi:10.1109/ICRA.2018.8461251.

[29] Mur-Artal, Raul and Tardós, Juan D, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras", *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017, doi:10.1109/TRO.2017.2705103.

[30] Campos, Carlos and Elvira, Richard and Rodríguez, Juan J Gómez and Montiel, José MM and Tardós, Juan D, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam", *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[31] Gomez-Ojeda, Ruben and Moreno, Francisco-Angel and Zuniga-Noël, David and Scaramuzza, Davide and Gonzalez-Jimenez, Javier, "Pl-slam: A stereo slam system through the combination of points and line segments", *IEEE Transactions on Robotics*, vol. 35, no. 3, pp. 734–746, 2019, doi:10.1109/TRO.2019.2899783.

[32] Cvišić, Igor and Marković, Ivan and Petrović, Ivan, "Soft2: Stereo visual odometry for road vehicles based on a point-to-epipolar-line metric", *IEEE Transactions on Robotics*, pp. 273–288, 2023, doi:10.1109/TRO.2022.3188121.

[33] Zhou, Yi and Gallego, Guillermo and Shen, Shaojie, "Event-based stereo visual odometry", *IEEE Transactions on Robotics*, 2021, doi:10.1109/TRO.2021.3062252.

[34] Labbé, Mathieu and Michaud, François, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation", *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019, doi:10.1002/rob.21831.

[35] Dai, Angela and Nießner, Matthias and Zollhöfer, Michael and Izadi, Shahram and Theobalt, Christian, "Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration", *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, p. 1, 2017, doi:10.1145/3072959.3054739.

[36] Whelan, Thomas and Kaess, Michael and Fallon, Maurice and Johannsson, Hordur and Leonard, John and McDonald, John, "Kintinuous: Spatially extended kinectfusion", 2012.

[37] Rosinol, Antoni and Abate, Marcus and Chang, Yun and Carlone, Luca, "Kimera: an open-source library for real-time metric-semantic localization and mapping", *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1689–1696, 2020, doi:10.1109/ICRA40945.2020.9196885.

[38] Bruno, Hudson and Colombini, Esther, "Lift-slam: a deep-learning feature-based monocular visual slam method", *Neurocomputing*, vol. 455, 2021, doi:10.1016/j.neucom.2021.05.027.

[39] "Objectfusion: Accurate object-level slam with neural object priors", *Graphical Models*, vol. 123, p. 101165, 2022, doi:https://doi.org/10.1016/j.gmod.2022.101165.

[40] Qing Li and Rui Cao and Jiasong Zhu and Hao Fu and Baoding Zhou and Xu Fang and Sen Jia and Shenman Zhang and Kanglin Liu and Qingquan Li, "Learn then match: A fast coarse-to-fine depth image-based indoor localization framework for dark environments via deep learning and keypoint-based geometry alignment", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 169–177, 2023, doi:https://doi.org/10.1016/j.isprsjprs.2022.10.015.

[41] Zhang, Ji and Singh, Sanjiv, "Loam: Lidar odometry and mapping in real-time.", *Robotics: Science and Systems*, vol. 2, 2014.

[42] Koide, Kenji and Miura, Jun and Menegatti, Emanuele, "A portable 3d lidar-based system for long-term and wide-area people behavior measurement", *IEEE Trans. Hum. Mach. Syst*, 2018, doi:10.1177/1729881419841532.

[43] Chen, Xieyuanli and Milioto, Andres and Palazzolo, Emanuele and Giguere, Philippe and Behley, Jens and Stachniss, Cyrill, "Suma++: Efficient lidar-based semantic slam", *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4530–4537, 2019, doi:10.1109/IROS40897.2019.8967704.

[44] Shan, Tixiao and Englot, Brendan, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain", *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4758–4765, 2018, doi:10.1109/IROS.2018.8594299.

[45] Shan, Tixiao and Englot, Brendan and Meyers, Drew and Wang, Wei and Ratti, Carlo and Rus, Daniela, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping", *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5135–5142, 2020, doi:10.1109/IROS45743.2020.9341176.

[46] Montemerlo, Michael and Thrun, Sebastian and Koller, Daphne and Wegbreit, Ben and others, "Fastslam: A factored solution to the simultaneous localization and mapping problem", *Aaai/iaai*, vol. 593598, 2002.

[47] Grisetti, Giorgio and Stachniss, Cyrill and Burgard, Wolfram, "Improved techniques for grid mapping with rao-blackwellized particle filters", *IEEE transactions on Robotics*, vol. 23, no. 1, pp. 34–46, 2007, doi:10.1109/TRO.2006.889486.

[48] Konolige, Kurt and Grisetti, Giorgio and Kümmerle, Rainer and Burgard, Wolfram and Limketkai, Benson and Vincent, Regis, "Efficient sparse pose adjustment for 2d mapping", *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 22–29, 2010, doi:10.1109/IROS.2010.5649043.

[49] Hess, Wolfgang and Kohler, Damon and Rapp, Holger and Andor, Daniel, "Real-time loop closure in 2d lidar slam", *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1271–1278, 2016, doi:10.1109/ICRA.2016.7487258.

[50] Konolige, Kurt and Grisetti, Giorgio and Kümmerle, Rainer and Burgard, Wolfram and Limketkai, Benson and Vincent, Regis, "Efficient sparse pose adjustment for 2d mapping", *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 22–29, 2010, doi:10.1109/IROS.2010.5649043.

[51] Kohlbrecher, Stefan and Von Stryk, Oskar and Meyer, Johannes and Klingauf, Uwe, "A flexible and scalable slam system with full 3d motion estimation", *2011 IEEE international symposium on safety, security, and rescue robotics*, pp. 155–160, 2011, doi:10.1109/SSRR.2011.6106777.

[52] Lv, Wenjun and Kang, Yu and Qin, Jiahu, "Indoor localization for skid-steering mobile robot by fusing encoder, gyroscope, and magnetometer", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 6, pp. 1241–1253, 2017, doi:10.1109/TSMC.2017.2701353.

[53] Jiang, Ping and Chen, Liang and Guo, Hang and Yu, Min and Xiong, Jian, "Novel indoor positioning algorithm based on lidar/inertial measurement unit integrated system", *International Journal of Advanced Robotic Systems*, vol. 18, no. 2, p. 1729881421999923, 2021, doi:10.1177/1729881421999923.

[54] Ismail, Zool H and Bukhori, Iksan, "Efficient detection of robot kidnapping in range finder-based indoor localization using quasi-standardized 2d dynamic time warping", *Applied Sciences*, vol. 11, no. 4, p. 1580, 2021, doi:10.3390/app11041580.

[55] Zhang, Lei, "Self-adaptive markov localization for single-robot and multi-robot systems", Ph.D. thesis, Université Montpellier II-Sciences et Techniques du Languedoc, 2010.

[56] Zhang, Lei and Zapata, Rene and Lepinay, Pascal, "Self-adaptive monte carlo localization for mobile robots using range finders", *Robotica*, vol. 30, no. 2, pp. 229–244, 2012, doi:10.1017/S0263574711000567.

[57] Zhang, Lei and Zapata, René and Lépinay, Pascal, "Self-adaptive monte carlo localization for mobile robots using range sensors", *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1541–1546, 2009, doi:10.1109/IROS.2009.5354298.

[58] Zhang, Lei and Zapata, René, "A three-step localization method for mobile robots", *Proceedings of International Conference on Automation, Robotics and Control Systems (ARCS 2009)*, pp. 50–56, 2009.

[59] Zhang, Lei and Zapata, René, "Probabilistic localization methods of a mobile robot using ultrasonic perception system", *2009 International Conference on Information and Automation*, pp. 1062–1067, 2009, doi:10.1109/ICINFA.2009.5205075.

[60] Yu, Wei and Li, Jie and Yuan, Jing and Ji, Xi, "Indoor mobile robot positioning based on uwb and low cost imu", *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, pp. 1239–1246, 2021, doi:10.1109/IAEAC50856.2021.9390754.

[61] Liu, Jianfeng and Pu, Jiexin and Sun, Lifan and He, Zishu, "An approach to robust ins/uwb integrated positioning for autonomous indoor mobile robots", *Sensors*, vol. 19, no. 4, p. 950, 2019, doi:10.3390/s19040950.

[62] Cui, Wei and Liu, Qingde and Zhang, Linhan and Wang, Haixia and Lu, Xiao and Li, Junliang, "A robust mobile robot indoor positioning system based on wi-fi", *International Journal of Advanced Robotic Systems*, vol. 17, no. 1, p. 1729881419896660, 2020, doi:10.1177/1729881419896660.

[63] Motroni, Andrea and Nepa, Paolo and Buffi, Alice and Tellini, Bernardo, "Robot localization via passive uhf-rfid technology: State-of-the-art and challenges", *2020 IEEE International Conference on RFID (RFID)*, pp. 1–8, 2020, doi:10.1109/RFID49298.2020.9244884.

[64] Bernardini, Fabio and Buffi, Alice and Fontanelli, Daniele and Macii, David and Magnago, Valerio and Marracci, Mirko and Motroni, Andrea and Nepa, Paolo and Tellini, Bernardo, "Robot-based indoor positioning of uhf-rfid tags: The sar method with multiple trajectories", *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2020, doi:10.1109/TIM.2020.3033728.

[65] Al-Forati, Israa Sabri Abdulameer and Rashid, Abdulmuttalib, "Design and implementation an indoor robot localization system using minimum bounded circle algorithm", *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*, pp. 1–6, 2019, doi:10.1109/ICMSAO.2019.8880404.

[66] Alfurati, IS and Rashid, Abdulmuttalib T, "An efficient mathematical approach for an indoor robot localization system", *Iraqi Journal of Electrical and Electronic Engineering*, vol. 15, no. 2, pp. 61–70, 2019, doi:10.37917/ijeee.15.2.7.

[67] Albuquerque, Daniel F and Gonçalves, Edgar S and Pedrosa, Eurico F and Teixeira, Francisco C and Vieira, José N, "Robot self position based on asynchronous millimetre wave radar interference", *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2019, doi:10.1109/IPIN.2019.8911809.

[68] Perez-Grau, Francisco J and Caballero, Fernando and Viguria, Antidio and Ollero, Anibal, "Multi-sensor three-dimensional monte carlo localization for long-term aerial robot navigation", *International Journal of Advanced Robotic Systems*, vol. 14, no. 5, p. 1729881417732757, 2017.

[69] Chen, Xieyuanli and Läbe, Thomas and Nardi, Lorenzo and Behley, Jens and Stachniss, Cyrill, "Learning an overlap-based observation model for 3d lidar localization", *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4602–4608, IEEE, 2020.

[70] Wolfgang Hess and Damon Kohler and Holger Rapp and Daniel Andor, "Real-time loop closure in 2d LIDAR SLAM", *2016 IEEE International Conference on Robotics and Automation, ICRA 2016, Stockholm, Sweden, May 16-21, 2016*, pp. 1271–1278, IEEE, 2016, doi:10.1109/ICRA.2016.7487258.

[71] Koki Aoki and Kenji Koide and Shuji Oishi and Masashi Yokozuka and Atsuhiko Banno and Junichi Meguro, "3d-bbs: Global localization for 3d point cloud scan matching using branch-and-bound algorithm", *CoRR*, vol. abs/2310.10023, 2023, doi:10.48550/ARXIV.2310.10023.

Received: 28 November, 2023, Revised: 20 February, 2024 Accepted: 21 February, 2024, Online: 19 March, 2024

DOI: https://doi.org/10.55708/js0303001

"Greenwashing" or "Helping": ESG Performance and Chinese Firm Total Factor Productivity

Jing Zhang, Ziyang Liu*

Graduate School, Kyonggi University, Suwon 16227, South Korea, zhangjing@kgu.ac.kr (J.Z.)

*Corresponding author: Ziyang Liu, Kyonggi University, Suwon Korea, victor@kgu.ac.kr

ABSTRACT: As the market economy has continued to develop, businesses have consistently prioritized profits, excessively emphasizing income and financial gains while neglecting ecological conservation and financial fraud. Consequently, the phenomenon of "greenwashing" has emerged. How to prevent this "greenwashing" phenomenon while pursuing economic benefits and enabling high-quality business development has become a focal point. Therefore, this paper analyzes whether the ESG (Environmental, Social, and Governance) performance of listed companies has an impact on the enhancement of the Total Factor Productivity (TFP) of enterprises. This study aims to explore how companies, while striving to maximize economic interests, can more proactively undertake environmental protection and social responsibility, thereby promoting the green transformation of enterprises. Using A-share listed companies from 2012 to 2022 as the sample, through an empirical examination of the correlation between the ESG performance of listed companies in China and the TFP of enterprises, the following conclusions are drawn: (1) ESG performance significantly promotes the TFP of enterprises, indicating that higher ESG performance corresponds to higher TFP; (2) Through intermediary effect tests, it is found that corporate reputation plays a role in enhancing the TFP of enterprises. That is, through good ESG performance, a company's reputation is improved, thereby leading to higher TFP; (3) Heterogeneity analysis demonstrates that the impact of good ESG performance on the enhancement of TFP is more significant in large-scale enterprises and state-owned enterprises.

KEYWORDS: ESG performance；Total Factor Productivity (TFP)；Corporate reputation

1. Introduction

Globally, issues concerning Corporate Social Responsibility (CSR) and Environmental, Social, and Governance (ESG) in sustainable development have garnered widespread attention. This attention originates not only from academia but also from governments, investors, media, and the general public. With shifts in policy orientation across the globe, more and more countries and regions are encouraging companies to fulfill their social and environmental responsibilities. ESG disclosure has become an integral part of a company's daily operations. For instance, the United Nations Sustainable Development Goals (SDGs) present a global consensus that demands active engagement from companies in addressing social and environmental issues. This is reflected in legislative and regulatory frameworks in numerous countries, offering policy support for corporate sustainability [1]. The proliferation of ESG disclosure is not merely a compliance requirement but has profound implications on business operations and investment decisions. Previous studies indicate that ESG disclosure aids in enhancing a company's reputation, reducing investment risks, and attracting more socially responsible investors [2]. This transformative shift is not confined to traditional CSR realms but encompasses operational performance closely associated with Total Factor Productivity (TFP). This paper's focal point is the relationship between ESG performance and TFP, with corporate reputation as the mediating variable. TFP is a crucial metric for assessing a company's performance, considering the comprehensive efficiency of multiple production factors[3,4]. The impact of ESG performance on TFP might be conveyed through various channels, with corporate reputation being a vital intermediary

mechanism. However, current research on how ESG performance affects TFP remains limited, especially in exploring mediating mechanisms [5]. This study aims to bridge this knowledge gap by delving into how ESG performance influences TFP through corporate reputation, providing practical guidance for corporate decision-makers, investors, and policymakers.

In recent years, more studies have started to focus on the relationship between ESG performance and TFP. In [6], the authors found a positive correlation between high ESG performance and higher TFP. They suggested that enhancing ESG performance might reduce environmental and social risks, enhance corporate reputation, and consequently affect TFP. Similarly, the research findings of [7] also indicate that companies with strong ESG performance exhibit higher TFP. However, these studies, while providing evidence, have certain limitations. Primarily, existing research often leans towards qualitative analysis or employs small samples, lacking large-scale quantitative analysis. This might limit the universality and robustness of the conclusions. Moreover, although some studies suggest that ESG performance may influence TFP through reputation channels, in-depth research into related mediating mechanisms remains relatively limited [8].

The innovative aspects of this study design will encompass several areas. Firstly, we aim to conduct comprehensive quantitative analysis using a large-scale dataset to address the deficiencies in current research [9]. This will contribute to a more comprehensive understanding of the relationship between ESG performance and TFP. Secondly, we will extensively investigate how ESG performance influences TFP through corporate reputation, exploring the details and pathways of the mediating mechanism. Finally, we will employ various methods, including robustness checks and heterogeneity analysis, to validate the robustness and universality of the research results, thereby enhancing the credibility of the study [10]. In summary, this paper aims to fill the research gap regarding the relationship between ESG performance and TFP, explore its mechanisms, and offer practical guidance for corporate decision-making and policy formulation. Through this research, we anticipate better comprehension of the connection between ESG and corporate operational performance, providing further insights for sustainable economic development.

2. Literature Review and Research Hypotheses

2.1. ESG Performance and Total Factor Productivity

According to the theory of information asymmetry, investors often require a substantial amount of relevant information to make decisions. In situations of severe information asymmetry, investors might make erroneous decisions, thereby reducing the allocation efficiency of the capital market. This leads to high-quality companies finding it challenging to obtain substantial funding for innovative activities, while less efficient companies can continue to acquire funds for less efficient production models. In [11], the authors demonstrated that the ESG rating of listed companies can reflect a company's level of corporate governance and effectively indicate the quality of information disclosure. There exists a significant positive correlation between these two aspects. Additionally, higher levels of corporate governance can effectively reduce situations where management sacrifices the company's long-term interests for personal gains due to agency problems and significantly curb the interests' encroachment by major shareholders [11–13]. By fully disclosing relevant information, companies enable investors to better understand their developmental status, facilitating the movement of capital in the capital market toward high-quality companies. Consequently, this enhances the ESG rating of listed companies, enabling companies to obtain more substantial resources for production activities, thereby increasing the total factor productivity of enterprises [14].

From a stakeholder theory perspective, if a company demonstrates excellent ESG performance, it can gain the trust of various stakeholders, thereby fostering the improvement of relationships among groups of stakeholders such as shareholders, employees, consumers, creditors, and the media. This leads to additional funding sources for external investments and internal research and development. Firstly, internal research and development, as a driving force for technological progress, enables companies to increase their knowledge reserves, optimize the combination of production factors, improve and upgrade production technology to enhance production efficiency. Additionally, research and development can help companies achieve technological innovation, providing new impetus for company development. Secondly, regarding external investments, limited resources can be invested in more efficient projects, thereby reducing the risk of resource misallocation and increasing total factor productivity [15]. Based on this, the following hypothesis is proposed:

H1: ESG performance contributes to enhancing the total factor productivity of enterprises.

2.2. ESG Performance, Corporate Reputation, and Total Factor Productivity

In the 1970s, signal transmission theory, stemming from information asymmetry theory, became a focal point for many scholars. Signal transmission theory is employed to alleviate information asymmetry. It posits that companies should timely disclose relevant internal information to enable outsiders to better understand the

company's actual operational status, thereby reducing information asymmetry. Moreover, to acquire external resources more effectively, internal personnel actively disclose information about the company to convey the company's operational situation to the external environment. Companies often use signals such as profits and incentives as financial indicators, while in the current scenario, ESG reports, non-financial indicator disclosures, have become an aspect that investors particularly focus on. Therefore, actively engaging in ESG disclosure essentially informs the market that the company considers the interests of various stakeholders during its operations, maintaining a positive stakeholder relationship, thus enabling the acquisition of more external resources [16–18].

Compared to companies that do not undertake social responsibility, socially responsible companies can transmit positive information about their good operational management to the outside world, thereby gaining the trust of more stakeholder groups, aiding the establishment of higher social reputation for the company [19]. Furthermore, corporate reputation is an intangible asset that allows companies to obtain capital from the outside, thereby alleviating the financing constraints of the company. Simultaneously, it increases the probability of attracting high-quality talents, which are crucial elements for company growth and also ensure the investment in human resources [19,20].

Hence, the following hypothesis is formulated:

H2: Corporate reputation acts as a mediator in the impact of ESG performance on the total factor productivity of enterprises.

3. Research Methods

3.1. Sample Selection

This study primarily investigates the impact of ESG performance of A-share listed companies on the total factor productivity of enterprises. Data were collected and organized through the Guotai An (CSMAR) database and the Wind Information (WIND) database. Using the Huazheng ESG rating data, the study period was set from 2012 to 2022, considering the availability of other key research variables. The data selection followed these criteria: (1) ST and *ST companies were removed; (2) companies from the financial and insurance industries were excluded; (3) samples with missing values were eliminated, resulting in an observed sample of 5993. To mitigate the influence of outliers on empirical results, a 1% winsorization was applied to continuous variables.

3.2. Variable Definitions

Dependent Variable: The total factor productivity (TFP) of enterprises was measured using methods such as OP, LP, OLS, etc. The study used the [21] and [22] methods to measure TFP through the LP method.

Explanatory Variable: The explanatory variable in this study is ESG performance. The study constructed the Huazheng ESG rating index by referencing mainstream ESG rating systems both domestically and internationally. Each indicator's applicability was deliberated to exclude unsuitable or unattainable data. Based on the rating criteria, ESG ratings were divided into 8 levels, from low to high: C, CC, CCC, B, BB, BBB, A, AA, and these ratings were used as the explanatory variable to measure a company's ESG performance.

Mediating Variable: After consulting relevant literature on corporate reputation, various foreign methods for measuring corporate reputation were examined, such as the "Most Admired Companies in America," "Global Most Admired Companies," and "Reputation Index" published by Fortune magazine. However, due to differences in cultural aspects between the East and the West, these methods might not be suitable for research on Chinese companies. Hence, building upon the work in [23], this study measured corporate reputation using intangible asset data.

Control Variables: Existing data suggests that factors such as company size, financial status, debt-paying ability, profitability, and corporate governance all impact the improvement of total factor productivity. Hence, referring to the methods of reference [24], the following control variables were selected: (1) Company Size (Size), (2) Company Age (Age), (3) Debt-to-Asset Ratio (Lev), (4) Return on Assets (ROA), (5) Fixed Asset Ratio (Fixed), (6) Company Growth (Growth), (7) Company Market Value (TobinQ), (8) Management Shareholding Ratio (Mshare), (9) Equity Concentration (Top10), and (10) Independent Director Ratio (Outdir). Definitions for related variables are detailed in Table 1.

3.3. Model design

To examine the impact of ESG performance on total factor productivity and explore the relationship between ESG performance, corporate reputation, and total factor productivity, a fixed-effects model based on the mediation effect testing method proposed [25] is established. To test the impact of ESG performance on overall factor productivity, the following model is designed:

$$TFP_{it}=\alpha_0+\alpha_1 ESG_{it}+\sum\alpha_j Controls_{it} + u_i + \lambda_t + \varepsilon_{it} \quad (1)$$

1. This model aims to investigate how ESG performance affects total factor productivity by considering various control variables.
2. Using corporate reputation as the dependent variable, the model aims to investigate the influence of ESG

performance on corporate reputation. The model is represented as:

$$Rep_{it} = \beta_0 + \beta_1 ESG_{it} + \sum \beta_j Controls_{it} + u_i + \lambda_t + \varepsilon_{it} \quad (2)$$

This model examines the impact of ESG performance on corporate reputation, incorporating control variables.

3. Incorporating both ESG performance and corporate reputation, this model examines whether corporate reputation mediates the relationship between ESG and total factor productivity:

$$TFP_{it} = \gamma_0 + \gamma_1 ESG_{it} + \gamma_2 Rep_{it} + \sum \gamma_j Controls_{it} + u_i + \lambda_t + \varepsilon_{it} \quad (3)$$

This model investigates the potential mediation effect of corporate reputation in the relationship between ESG and total factor productivity.

Here, I represents the enterprise, t represents the year, **Controls represents** the control variables, u_i represents individual fixed effects, λ_t **represents time fixed effects, and** ε_{it} represents the random disturbance term.

Table 1. Variable definition explanations

Variable type	Variable Names:	variable symbol	variable definition
Dependent Variable	High-Quality Development of Enterprises	*TFP*	Measurement of Total Factor Productivity using the LP method
Independent Variable	Enterprise ESG Performance	*ESG*	Assignment of values from 1-8 based on the Huazheng ESG rating from low to high
Mediating Variable	Corporate Reputation	*Rep*	Natural logarithm of Intangible Asset Net Value
Control Variable	Company Size	*Size*	Natural logarithm of Total Assets of the enterprise
	Company Age	*Age*	Age of the enterprise from its establishment to the specific period
	Asset-Liability Ratio	*LEV*	Total Liabilities to Total Assets ratio of the enterprise
	Asset Yield/Asset Profitability Ratio	*ROA*	Net Profit to Total Assets ratio
	Fixed Asset Ratio	*Fixed*	Net Fixed Assets to Total Assets ratio
	Company Growth Potential	*Growth*	Growth in Total Assets of the enterprise
	Company Market Value	*TobinQ*	Market Value to Total Assets ratio
	Management Holding Percentage	*Mshare*	Quantity of Management-held Shares to Total Shares
	Equity Concentration	*Top*10	Sum of the shareholding proportion of the top 10 shareholders of the enterprise
	Proportion of Independent Directors	*Outdir*	Proportion of Independent Directors among the total number of Board Members

Table 2: Descriptive statistics of variables

Variable name	Sample size	Mean value	Standard deviation	Minimum value	Maximum value
TFP	5993	8.634	1.115	4.706	13
ESG1	5993	3.965	1.162	1	8
Size	5993	22.608	1.313	18.524	28.293
Lev	5993	0.493	0.207	0.01	1.957
ROA	5993	0.03	0.075	-0.894	0.517
Cashflow	5993	0.047	0.078	-0.556	0.661
FIXED	5993	0.225	0.177	0	0.929
Growth	5993	0.276	2.175	-0.985	87.484
Indep	5993	0.374	0.059	0.167	0.714
Top10	5993	0.534	0.152	0.106	0.952
TobinQ	5993	2.025	2.16	0.674	76.82
Age	5993	3.058	0.244	1.946	3.761
Mshare	5993	0.036	0.111	0	1.694
Rep	5993	18.935	1.877	7.458	24.398

4. Analysis of empirical results

4.1. Descriptive Statistics

Descriptive statistical analysis, as depicted in Table 2, provides an overview of the collected data. This study gathered a total of 5993 valid samples, encompassing 14 variables. Among these variables, the total factor productivity (TFP) ranges from a minimum of 4.706 to a maximum of 13, with a standard deviation of 1.115. These values indicate significant differences in development across various industries in China.

The ESG performance (ESG1) of companies varies substantially, ranging from a minimum of 1 to a maximum of 8, with an average value of 3.965 and a standard deviation of 1.162. This highlights the considerable diversity in ESG performance among different companies.

Regarding corporate reputation (Rep), the average value is 18.935 with a standard deviation of 1.877. These figures illustrate significant differences in reputation among different enterprises. Such variability among the samples enhances the effectiveness of the empirical model analysis. The differences between samples enable more comprehensive matching in regression analysis, resulting in more credible and robust outcomes.

4.2. Correlation Analysis

The correlation analysis reveals significant relationships among the variables. From Table 3, it is evident that the correlation coefficient between ESG and total factor productivity is 0.276. This indicates a substantial positive association between ESG performance and total factor productivity, and the research findings are statistically significant at the 1% level. This result provides initial support for hypothesis H1, suggesting that ESG performance significantly enhances total factor productivity.

Furthermore, the analysis shows a strong correlation between corporate reputation and ESG performance, with a correlation coefficient of 0.153. Additionally, the correlation coefficient between corporate reputation and total factor productivity stands at 0.361, both achieving statistical significance at the 1% level. These results suggest that ESG performance elevates total factor productivity through the enhancement of corporate reputation, thereby providing preliminary validation for hypothesis H2.

Table 3: Correlation analysis of variables

Variables	TFP	ESG1	Size	Lev	ROA	Cashflow	FIXED	Growth	Indep	Top10	Tobin Q	Age	Mshare	Rep ut
(1) TFP	1.000													
(2)ESG1	0.276***	1.000												
(3) Size	0.737***	0.312***	1.000											
(4) Lev	0.361***	0.044***	0.415***	1.000										
(5) ROA	0.179***	0.189***	0.086***	0.357***	1.000									
(6)Cashflow	0.100***	0.047***	0.040***	0.205***	0.405***	1.000								
(7) FIXED	0.197***	0.050***	-0.004	-0.017	0.041***	0.232***	1.000							
(8) Growth	0.038***	0.041***	0.010	0.014	0.070***	0.006	-0.026*	1.000						
(9) Indep	0.016	0.096***	0.038***	0.000	-0.036**	0.040***	-0.034**	0.004	1.000					
(10) Top10	0.273***	0.084***	0.333***	0.083***	0.133***	0.085***	0.031**	0.048***	-0.036**	1.000				
(11)Tobin Q	0.301***	0.122***	0.391***	0.209***	0.013	0.018	0.074***	-0.011	0.029**	0.124***	1.000			
(12) Age	0.101***	0.095***	0.130***	0.060***	0.089***	-0.045***	-0.126***	0.023	0.040***	-0.019	0.026*	1.000		
(13) Mshare	-0.022	-0.008	0.088***	0.139***	0.099***	0.059***	-0.121***	0.053***	-0.003	0.006	0.007	-0.176***	1.000	
(14) Rep	0.361***	0.153***	0.560***	0.105***	0.075***	0.116***	0.220***	-0.026*	0.005	0.153***	-0.212***	-0.004	-0.027*	1.000

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Variance Inflation Factor (VIF) Test

Variables	VIF	1/VIF
Size	1.78	0.561
Lev	1.54	0.65
ROA	1.32	0.759
ESG1	1.22	0.82
Age	1.17	0.855
Top10	1.15	0.87
TobinQ	1.1	0.909
Mshare	1.07	0.933
FIXED	1.05	0.949
Indep	1.02	0.982
Growth	1.00	0.998

In order to accurately mitigate the interference of multicollinearity among the variables, a VIF test was conducted. The analysis results, as presented in Table 4, reveal that all variables have VIF values less than 5. This indicates that there is no evidence of multicollinearity among the variables. Hence, the selection of variables is reasonable, and the analytical data results possess a higher level of reliability.

4.3. Hausman Test

Prior to conducting regression analysis, a Hausman test was employed to determine whether a fixed effects model or a random effects model should be selected. The test results, as displayed in Table 5, indicate that the test statistics for each variable are less than 0.01. This signifies that for this research, a fixed effects model is more suitable compared to a random effects model.

4.4. Baseline Regression Analysis

Following the research approach, a baseline regression was conducted to explore the relationship between a company's ESG performance and Total Factor Productivity (TFP). The results of the model are presented in Table 6. Model (1) includes Total Factor Productivity (TFP) and Company ESG1 performance alone, while Model (2) incorporates Total Factor Productivity (TFP), Company ESG1 performance, and control variables.

The analysis indicates that the value of ESG1 in Model (1) is 0.0454, and in Model (2), it stands at 0.0484. Both values are statistically significant at the 1% level, indicating a positive impact of a company's ESG performance on its Total Factor Productivity. In other words, as a company's ESG performance improves, its Total Factor Productivity tends to increase.

Table 5: Hausman Test

variable	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
ESG1	0.008	0.011	-0.004	0.001
Size	0.525	0.562	-0.037	0.006
Lev	0.276	0.333	-0.057	0.014
ROA	1.46	1.548	-0.088	0.009
FIXED	-1.076	-1.018	-0.058	0.023
Growth	0.001	0.001	0.001	0.001
Indep	0.167	0.127	0.04	0.03
Top10	0.04	0.045	-0.005	0.023
TobinQ	0.004	0.004	0	0.001
Age	0.608	0.252	0.355	0.104
Mshare	0.351	0.281	0.07	0.032

b = Consistent under H0 and Ha; obtained from xtreg.

B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 333.59

Prob > chi2 = 0.0000

Model (1) has an adjusted R-squared of 0.107, whereas Model (2), which includes control variables, exhibits an adjusted R-squared of 0.275. The higher adjusted R-squared in Model (2) compared to Model (1) suggests that the inclusion of control variables enhances the model's fit.

Table 6: Baseline Regression Results

	TFP	TFP
	(1)	(2)
ESG1	0.0454***	0.0484***
	(4.9122)	(6.3509)
Lev		1.0163***
		(18.1972)
ROA		2.1160***
		(20.5147)
FIXED		-1.2914***
		(-17.2327)
Growth		-0.0000
		(-0.9690)
Indep		0.0791
		(0.5050)
Top10		0.8605***
		(11.7103)
TobinQ		-0.0148***
		(-9.1943)
Age		1.2318***
		(36.8100)
Mshare		0.4258***
		(4.0982)
cons	8.3609***	3.9117***
	(225.0124)	(29.0693)
N	5993	5993
adj. R2	0.107	0.275

4.5. Mediation Analysis

The results of the mediation effect for company reputation are depicted in Table 7. Model (1) represents the regression model between the dependent variable and independent variable along with control variables in the absence of Reputation. Model (2) includes the regression model of the independent variable, control variables, and Reputation. Lastly, Model (3) displays the regression model between the dependent variable and independent variable after adding control variables and Reputation.

The regression outcomes reveal a significant impact of a company's ESG performance on the enhancement of Total Factor Productivity under the mediation effect of company reputation. Both Model (1) and Model (3) exhibit significant results for ESG performance at a 1% confidence level. Specifically, when a company's ESG1 performance increases by 1 unit, it leads to a 0.0471 unit increase in high-quality development for the enterprise. In Model (2), the mediator variable, company Reputation, shows a significant positive correlation with ESG ratings at a 1% level. This signifies the transmission role of company reputation between ESG ratings, indicating that an improvement in ESG performance enhances company reputation, further boosting Total Factor Productivity.

Table 7: Company Reputation Mediation Effect

	TFP	Rep	TFP
	(1)	(2)	(3)
ESG1	0.047***	0.070***	0.043***
	(6.07)	(4.05)	(5.50)
Rep			0.070***
			(10.12)
Lev	0.903***	0.667***	0.856***
	(14.16)	(4.75)	(13.55)
ROA	1.891***	0.067	1.887***
	(17.65)	(0.28)	(17.82)
FIXED	-1.027***	0.441**	-1.058***
	(-12.16)	(2.37)	(-12.67)
Growth	0.020***	0.004	0.020***
	(7.18)	(0.63)	(7.17)
Indep	0.175	-0.905**	0.239
	(1.05)	(-2.47)	(1.45)
Top10	0.849***	1.406***	0.750***
	(10.47)	(7.87)	(9.29)
TobinQ	-0.033***	-0.029***	-0.031***
	(-7.92)	(-3.18)	(-7.51)
Age	1.197***	-0.446	1.228***
	(5.64)	(-0.95)	(5.86)
Mshare	0.313***	1.023***	0.241**
	(2.79)	(4.14)	(2.17)
Y			0.070***
			(10.12)
cons	4.191***	18.230***	2.910***
	(6.50)	(12.83)	(4.48)
N	4783	4783	4783
R2	0.405	0.207	0.419
adj. R2	0.318	0.091	0.334

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The resolution of images should not be less than 118 pixels/cm when width is set to 16 cm. Images must be scanned at 1200 dpi resolution and submitted in jpeg or tiff format. Graphs and diagrams must be drawn with a line weight between 0.5 and 1 point. Graphs and diagrams with a line weight of less than 0.5 point or more than 1 point are not accepted. Scanned or photocopied graphs and diagrams are not accepted.

4.6. Robustness Tests

4.6.1 Replacing the Dependent Variable

To test the robustness of the model, the dependent and independent variables were replaced. In the previous basic regression analysis, the ESG ratings from Huazheng were assigned numerical values according to their order. For robustness testing, following the methodology from previous research by [26] and [27], a reassignment of the Huazheng ESG ratings was done. Publicly available data indicate that these ratings can be categorized into three levels: A-rated companies are ESG performance leaders, B-rated companies represent average ESG performance, and C-rated companies lag behind. In this context, the ESG2 variable was assigned values according to these definitions to replace ESG1. Additionally, the alternative measurement methods for the dependent variable, Total Factor Productivity, were considered. Robustness testing included using the OP method, OLS method, and fixed-effect method to reevaluate Total Factor Productivity and use these new measurements as dependent variables.

In Table 8, Model (1) represents the replacement of the explanatory variable ESG1 with ESG2. Models (2), (3), and (4) represent the results of regressions using OLS, fixed-effect, and OP methods with different measurements of the dependent variable. The results show that ESG performance remains highly significant in all models at a 1% significance level. This consistency in results suggests that the model is robust.

4.6.2 Instrumental Variable

The baseline regression results indicate that the better the ESG performance, the stronger the positive effect on enhancing a company's total factor productivity. However, this result could also be due to the fact that companies with higher total factor productivity are more willing to actively improve their own ESG rating, potentially leading to a two-way causal endogeneity issue. To address this potential endogeneity issue, this paper employs lagged ESG performance and the industry average ESG performance as instrumental variables in a two-stage least squares regression. In the first-stage regression of the two-stage least squares method, the study uses the firm's ESG performance as the explained variable, the lagged ESG performance, and industry average ESG performance as explanatory variables, alongside the control variables From Model (1), controlling for industry and time effects in a least squares regression. In the second stage, the study employs the total factor productivity measured by the OP and LP methods as the explained variables and ESG performance as the explanatory variable for regression.

Table 8: Replacing the Dependent Variable

	TFP_LP (1)	TFP_OLS (2)	TFP_FE (3)	TFP_OP (4)
ESG1		0.0709***	0.0770***	0.0191***
		(8.4224)	(8.8641)	(2.6220)
ESG2	0.1248***			
	(7.0238)			
Lev	1.2301***	1.2503***	1.3103***	0.7495***
	(19.9518)	(20.2767)	(20.6031)	(14.0487)
ROA	2.3734***	2.3542***	2.4035***	1.9902***
	(20.8115)	(20.6719)	(20.4631)	(20.1982)
FIXED	-0.2301***	-0.2165***	-0.0589	-1.0492***
	(-2.7761)	(-2.6171)	(-0.6898)	(-14.6566)
Growth	-0.0001*	-0.0001*	-0.0001**	-0.0000
	(-1.8598)	(-1.9281)	(-2.0534)	(-0.7751)
Indep	-0.0789	-0.1321	-0.1822	0.2694*
	(-0.4559)	(-0.7635)	(-1.0210)	(1.8001)
Top10	1.2332***	1.2340***	1.3121***	0.6541***
	(15.1693)	(15.2100)	(15.6809)	(9.3184)
TobinQ	-0.0180***	-0.0182***	-0.0190***	-0.0123***
	(-10.1472)	(-10.2735)	(-10.3803)	(-8.0163)
Age	1.6894***	1.6823***	1.7672***	1.0993***
	(45.5806)	(45.5339)	(46.3758)	(34.3910)
Mshare	0.2287**	0.2391**	0.2112*	0.3907***
	(1.9898)	(2.0845)	(1.7849)	(3.9369)
cons	4.5909***	4.5490***	4.7424***	2.8117***
	(30.8029)	(30.6185)	(30.9489)	(21.8732)
N	5878	5878	5878	5878
adj. R^2	0.299	0.301	0.305	0.225

The regression results are presented in Table 9. It can be observed that in the first stage of Table 9, the regression coefficients of L.ESG and IndESG are both significantly greater than 0 at the 1% significance level. This indicates a high correlation between these two instrumental variables and the explanatory variables. The results of the weak instrumental variable test and over-identification test also demonstrate the effectiveness of the instrumental variable selection in this study. Looking at columns (2) and (3) of Table 9, the regression coefficients for ESG are 0.101 and 0.185, both significant at the 1% level. This demonstrates that, even after using instrumental variables to address potential endogeneity issues in the two-stage least squares regression, the conclusions of this study still hold, affirming that a strong ESG performance in companies can contribute to increased total factor productivity.

Table 9: Results of Instrumental Variable Regression

	First ESG (1)	Second TFP_OP (2)	Second TFP_LP (3)
ESG		0. 101***	0. 185***
		(17. 47)	(26. 61)
L. ESG	0. 774***		
	(145.65)		
IndESG	0. 066***		
	(3. 94)		
Control	YES	YES	YES
YEAR	YES	YES	YES
INDUSTRy	YES	YES	YES
_cons	0. 924***	5. 125***	6. 580***
	(7. 64)	(71.22)	(76.07)
N	5878	5878	5878
adj. R2	0. 614	0. 484	0. 552

4.6.3 Controlling for Individual Effects

To further address estimation biases caused by endogeneity, this study undertook regression analysis controlling for individual effects. Table 10 provides regression results simultaneously controlling for individual, industry, and yearly effects. From Table 10, it is observed that after controlling for individual effects, the coefficients for ESG are 0.0629, 0.0654, 0.1142, and 0.1194, all significant at the 1% level. This indicates that strong ESG performance in companies can enhance total factor productivity and affirms the prior research conclusion even after addressing potential endogeneity by controlling for individual effects.

4.6.4 Sample Selection Issue

This paper employs the Heckman two-stage regression to address potential sample selection issues. In the first stage Probit regression, the dependent variable DESG is a dummy variable. When ESG is greater than the mean, DESG is 1; when ESG is less than the mean, DESG is 0. The exogenous instrument variable remains the industry average of ESG performance (IndESG). The choice of the industry average of ESG performance as an exogenous instrument variable is because a company's ESG performance can be influenced by industry ESG performance, especially the industry environment, social responsibility awareness, and corporate governance, which can affect a company's ESG performance. Other control variables remain consistent with Model (1). The inverse Mill's ratio is calculated from the first-stage regression results, and it is included in the model for the second-stage regression. The regression results, as shown in the second column and third column of Table 11, demonstrate that the ESG regression coefficients for both OP and LP measures of total factor productivity are significant at the 1% level. This indicates that even after employing Heckman's two-stage regression to overcome potential sample selection issues, the study's conclusion remains valid, demonstrating the robustness of the research findings.

Table 10: Results of Regression Controlling for Individual Effects

variable	(1) TFP_OP	(2) TFP_OP	(3) TFP_LP	(4) TFP_LP
ESG	0. 0629***	0. 0654***	0. 1142***	0. 1194***
	(15. 22)	(16. 12)	(23. 60)	(24. 98)
Control	YES	YES	YES	YES
YEAR	NO	YES	NO	YES
INDUSTRY	NO	YES	NO	YES
FIRM	NO	YES	NO	YES
_cons	4. 8033***	4. 8910***	6. 4325***	6. 4697***
	(105. 99)	(80. 20)	(121. 30)	(90. 03)
N	5878	5878	5878	5878
Adj. R2	0. 4488	0. 5001	0. 5426	0. 5793
F	1713. 11	526. 8621	2495. 51	725. 0600

Table 11: Heckman Two-Stage Regression Results

variable DESG	(1) TFP_OP	(2) TFP_LP	(3)
ESG		0. 0796***	0. 1436***
		(19. 53)	(29. 47)
IndESG	0. 0935**		
	(2. 08)		
imr		1. 6762***	2. 1807***
		(10. 77)	(11. 72)
Control	YES	YES	YES
YEAR	YES	YES	YES
INDUSTRY	YES	YES	YES
_cons	0. 9302***	4. 6803***	6. 2011***
	(2. 76)	(71. 93)	(79. 72)
N	5787	5878	5878
Adj. R2		0. 4869	0. 5559
F		524. 3927	691. 3780

4.7. Further Research

4.7.1 Analysis Based on Scale Heterogeneity

Larger companies tend to hold leading positions within their industries due to their robust risk resilience and extensive funding sources, providing a secure foundation for the full realization of ESG's role. This subsequently aids companies in alleviating financing constraints and increasing research and development investments, thus enhancing total factor productivity. Accordingly, this study anticipates that the positive impact of excellent ESG performance on the total factor productivity of large-scale enterprises will be greater than that on small and medium-sized enterprises (SMEs). Therefore, In [28, 29], papers measures company size using the natural logarithm of total assets and groups the sample based on the mean of this indicator. Companies larger than the mean are categorized as large-scale enterprises, while those smaller than the mean are classified as SMEs. Subsequently, group regression and inter-group coefficient difference tests are conducted based on Model (1). Columns (1) and (2) in Table 12 provide the results of the group regression, showing that while the ESG coefficients in both sample groups are significantly positive, the regression coefficients in the large-scale enterprise sample (0.0744 and 0.1291) are higher than those in the SME sample (0.0095 and 0.0398). Moreover, the inter-group coefficient difference test also indicates significance at the 1% level, confirming the hypothesis that excellent ESG performance has a greater positive impact on the total factor productivity of large-scale enterprises compared to SMEs.

4.7.2 Heterogeneity Analysis Based on Property Rights

For state-owned enterprises, they possess stronger talent and technological advantages and tend to have easier access to funding sources during financing. Consequently, state-owned enterprises' research and development investments often enjoy more robust financial support, contributing to an improved total factor productivity. In this scenario, state-owned enterprises are better positioned to leverage the advantages of excellent ESG performance, translating it into impetus for increased research and development investments to enhance total factor productivity. Therefore, this study expects that excellent ESG performance has a greater positive impact on the total factor productivity of state-owned enterprises compared to non-state-owned enterprises.

Table 12: Heterogeneity Test Results Based on Scale

	(1)	(2)	(3)	(4)
variable	TFP_OP	TFP_OP	TFP_LP	TFP_LP
ESG	0. 0744***	0. 0095*	0. 1291***	0. 0398***
	(12. 95)	(1. 90)	(19. 83)	(7. 26)
Control	YES	YES	YES	YES
YEAR	YES	YES	YES	YES
INDUSTRY	YES	YES	YES	YES
_cons	5. 4618***	5. 6829***	7. 8087***	7. 4120***
	(51. 23)	(85. 13)	(64. 67)	(101. 60)
N	5787	5878	5787	5878
Adj. R2	0. 3843	0. 2829	0. 4481	0. 3346
F	150. 8827	121. 3907	195. 9741	154. 4751

Coefficient difference test between groups
Chi2(1) = 71. 15
Prob>Chi2 = 0. 0000
Chi2(1) = 106. 61
Prob>Chi2 = 0. 0000

To test the heterogeneous effects of ESG performance on the improvement of total factor productivity for state-owned and non-state-owned enterprises, this study uses a dummy variable based on the ownership nature. State-owned enterprises are represented as 1, while non-state-owned enterprises are represented as 0. Group regression is conducted based on Model (1), and inter-group coefficient difference tests are performed. Columns (3) and (4) in Table 13 present the regression results for the sub-samples of state-owned and non-state-owned enterprises.

The regression coefficients for ESG in state-owned enterprise results are 0.1072 and 0.1721, both significant at the 1% level. In the non-state-owned enterprise results, the ESG coefficients are 0.0527 and 0.1084, also significant at the 1% level. While both are significantly positive at the 1% level, the regression coefficient for state-owned enterprises is higher than that for non-state-owned

Table 13: Heterogeneity Test Results Based on Property Rights

variable	(1) TFP_OP	(2) TFP_OP	(3) TFP_LP	(4) TFP_LP
ESG	0. 1072***	0. 0527***	0. 1721***	0. 1084***
	(14. 51)	(10. 74)	(19. 56)	(18. 58)
Lev	1. 3074***	1. 4923***	1. 7763***	2. 2212***
	(23. 99)	(45. 49)	(27. 39)	(56. 92)
Control	YES	YES	YES	YES
YEAR	YES	YES	YES	YES
INDUSTRY	YES	YES	YES	YES
_cons	4. 8553***	5. 1079***	6. 6820***	6. 6744***
	(41. 67)	(68. 11)	(48. 18)	(74. 81)
N	4564	5878	4564	5878
Adj. R2	0. 5018	0. 4409	0. 5588	0. 5040
F	184. 4533	289. 8876	231. 6766	373. 3346

enterprises. Moreover, the inter-group coefficient difference test is significant at the 1% level, indicating that excellent ESG performance has a more substantial positive effect on the total factor productivity of state-owned enterprises compared to non-state-owned enterprises.

5. Conclusion and Enlightenment

5.1. Research Conclusions

This study focuses on A-share listed companies from 2012 to 2022, utilizing the Huazheng rating index to measure corporate ESG performance and validating the correlation between ESG performance and total factor productivity of listed companies. The main conclusions drawn from this research are as follows:

ESG Performance and Total Factor Productivity: ESG performance significantly enhances a company's total factor productivity. This implies that higher ESG performance correlates with increased total factor productivity.

Effect of Corporate Reputation: Empirical analysis demonstrates that a company's reputation significantly influences the enhancement of its total factor productivity. This indicates that improving a company's reputation through good ESG performance further elevates its total factor productivity. Furthermore, these conclusions remain robust after various sensitivity tests such as altering the computation method of the explained variable, alternative assignment methods for explanatory variables, implementing instrumental variable methods to alleviate potential endogeneity issues, and utilizing the Heckman two-stage method to correct sample selection biases.

Heterogeneity Analysis: The research shows that the impact of good ESG performance on enhancing total factor productivity is more pronounced in large-scale enterprises and state-owned enterprises.

5.2. Policy Insights

Based on the conclusions drawn in the previous sections, the paper proposes policy recommendations in the following areas:

Policy Formation: Governments should expedite the formulation of laws and regulations governing corporate ESG performance to establish a robust institutional framework, directing policies to steer companies toward high-quality development. This includes fostering environmentally friendly practices and preventing inadequate or untimely disclosures, thus discouraging unethical behaviors born from insufficient information, such as illegal waste disposal or opportunistic practices.

Enhanced Regulatory Oversight: Regulatory authorities should intensify their focus on corporate ESG performance by implementing stricter laws, establishing robust regulatory mechanisms, and employing more robust measures to maintain market order. Introducing third-party audit mechanisms for comprehensive assessment and evaluation of a company's ESG performance would ensure its quality and credibility.

Balanced Perspective on ESG Performance: Companies should adopt a balanced perspective on the unique role of ESG performance. They should invest efforts into achieving a win-win scenario in terms of economic and social benefits based on their specific circumstances. Acknowledging the necessity of improving ESG performance for sustainable development, companies can align economic and social benefits, promoting organizational legitimacy for their high-quality development.

This research provides practical policy implications to promote sustainable business practices and organizational growth.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] X. Deng, W. Li, X. Ren. "More Sustainable, More Productive: Evidence from ESG Ratings and Total Factor Productivity among Listed Chinese Firms." Finance Research Letters, vol. 51, no. 3, pp. 789–795, 2023, doi.org/10.1016/j.frl.2022.103439.

[2] N. Li, X. Wang, Z. Wang, X. Luan. "The Impact of Digital Transformation on Corporate Total Factor Productivity." Frontiers in Psychology, vol. 13, no. 5, pp. 1544–1553, 2022, doi: 10.3389/fpsyg.2022.1071986.

[3] G. Ge, X. Xiao, Z. Li, Q. Dai. "Does ESG Performance Promote High-Quality Development of Enterprises in China? The Mediating Role of Innovation Input." Sustainability, vol. 14, no. 7, pp. 3843, 2022, doi.org/10.3390/su14073843.

[4] Q. Xia, Y. Liu, F. Wei. "How Can ESG Funds Improve Their Performance? Based on the DEA-Malmquist Productivity Index and fsQCA Method." zgkxjsdxxb, vol. 53, no. 4, pp. 0803–0809, 2023, doi：10.52396/justc-2023-0017.

[5] X. Su, S. Wang, F. Li. "The Impact of Digital Transformation on ESG Performance Based on the Mediating Effect of Dynamic Capabilities." Sustainability, vol. 15, no. 18, pp. 13506, 2023, doi.org/10.3390/su151813506.

[6] L. Sun, N.A.M. Saat. "How Does Intelligent Manufacturing Affect the ESG Performance of Manufacturing Firms? Evidence from China." Sustainability, vol. 15, no. 4, pp. 2898, 2023, doi: 10.3390/su15042898.

[7] G. Sun, C. Guo, J. Ye, C. Ji, N. Xu, H. Li. "How ESG Contribute to the High-Quality Development of State-Owned Enterprise in China: A Multi-Stage fsQCA Method." Sustainability, vol. 14, no. 23, pp. 15993, 2022, doi.org/10.3390/su142315993.

[8] F. Wang, Z. Sun. "Does the Environmental Regulation Intensity and ESG Performance Have a Substitution Effect on the Impact of Enterprise Green Innovation: Evidence from China." International Journal of Environmental Research and Public Health, vol. 19, no. 14, pp. 8558, 2022, doi.org/10.3390/ijerph19148558.

[9] Y. Shen, H. Zheng, H. Cai, X. Chen, Y. Liu, S. Ma, X. Zhao. "ESG Performance, R&D Innovation and High Quality Development of Corporate: A Perspective Based on Firm Performance." Industrial Engineering and Innovation Management, vol. 5, no. 6, pp. 23–34, 2022, https://www.nature.com/articles/s41565-019-0603-y.

[10] P. Yang, X. Hao, L. Wang, S. Zhang, L. Yang. "Moving toward Sustainable Development: The Influence of Digital Transformation on Corporate ESG Performance." Kybernetes, vol. 53, pp. 1544–1553, 2023, doi.org/10.1126/sciadv.1501122.

[11] R. Yao, Y. Fei, Z. Wang, X. Yao, S. Yang. "The Impact of China's ETS on Corporate Green Governance Based on the Perspective of Corporate ESG Performance." International Journal of Environmental Research and Public Health, vol. 20, no. 3, pp. 2292, 2023, doi.org/10.1002/ani.202116068.

[12] J. Zheng, Y. Jiang, Y. Cui, Y. Shen. "Green Bond Issuance and Corporate ESG Performance: Steps toward Green and Low-Carbon Development." Research in International Business and Finance, vol. 66, pp. 102007, 2023,doi.org/10.1016/j.ribaf.2023.102007 .

[13] Y.P. Chen (Vincent), Z. Zhuo, Z. Huang, W. Li. "Environmental Regulation and ESG of SMEs in China: Porter Hypothesis Re-Tested." Science of The Total Environment, vol. 850, pp. 157967, 2022, doi.org/10.1016/j.scitotenv.2022.157967.

[14] D. Zhang, L. Liu. "Does ESG Performance Enhance Financial Flexibility? Evidence from China." Sustainability, vol. 14, no. 18, pp. 11324, 2022, doi.org/10.3390/su141811324.

[15] D. Zhang. "Does Green Finance Really Inhibit Extreme Hypocritical ESG Risk? A Greenwashing Perspective Exploration." Energy Economics, vol. 121, pp. 106688, 2023, doi.org/10.1016/j.eneco.2023.106688 .

[16] D. Zhang, L. Meng, J. Zhang. "Environmental Subsidy Disruption, Skill Premiums and ESG Performance." International Review of Financial Analysis, vol. 90, pp. 102862, 2023, doi.org/10.1016/j.irfa.2023.102862.

[17] P. Moskovics, P. Wanke, Y. Tan, A.M. Gerged. "Market Structure, ESG Performance, and Corporate Efficiency: Insights from Brazilian Publicly Traded Companies." Business Strategy and the Environment, doi:10.1002/bse.3492.

[18] S. Li, E. Xie. "The Effect of Economic Growth Target Constraints on ESG." Applied Economics Letters, vol. 0, pp. 1–4, 2023, doi: 10.1002/bse.3492.

[19] A. Babkin, E. Shkarupeta, L. Tashenova, E. Malevskaia-Malevich, T. Shchegoleva. "Framework for Assessing the Sustainability of ESG Performance in Industrial Cluster Ecosystems in a Circular Economy." Journal of Open Innovation: Technology, Market, and Complexity, vol. 9, pp. 100071, 2023, doi: 10.1016/j.joitmc.2023.100071.

[20] C. Li, S. Ba, K. Ma, Y. Xu, W. Huang, N. Huang. "ESG Rating Events, Financial Investment Behavior and Corporate Innovation." Economic Analysis and Policy, vol. 77, pp. 372–387, 2023, doi.org/10.1016/j.eap.2022.11.013.

[21] Y. Luo, Z. Lu, C. Wu, C.N. Mensah. "Environmental Regulation Effect on Green Total Factor Productivity: Mediating Role of Foreign Direct Investment Quantity and Quality." International Journal of Environmental Research and Public Health, vol. 20, no. 4, pp. 3150, 2023, doi.org/10.3390/ijerph20043150, .

[22] L. Chen, M.U. Khurram, M.Z. Abedin, yuyang Gao, B.M. Lucey. "ESG Disclosure and Technological Innovation Capabilities of the Chinese Listed Companies." 2022, doi.org/10.1016/j.ribaf.2023.101974.

[23] W. Ma, Y. Li, L. Ding. "Does Marine Financial Policy Affect Total Factor Productivity of Marine Enterprises? An Empirical Evidence Based on Chinese First Guidance on Strengthening Finance for Marine Economy." Marine Pollution Bulletin, vol. 195, pp. 115493, 2023, doi.org/10.1016/j.marpolbul.2023.115493.

Jing Zhang got master's degree from Kyonggi University, and currently a doctoral student. He is producer of China Education Television, and my research direction is ESG and brand management. I have published several related papers so far.

Ziyang Liu received the PhD in management from the Kyonggi University and am currently an assistant professor at the University, specializing in brand management and sustainability.

Received: 23 January, 2024, Revised: 29 February, 2024 Accepted: 11 March, 2024, Online: 19 March, 2024
DOI: https://dx.doi.org/10.55708/js0303003

Missile Guidance using Proportional Navigation and Machine Learning

Mirza Hodžić, Naser Prljača*
Control systems and Robotics, Faculty of electrical engineering, University of Tuzla, Tuzla, 75000, Bosnia and Herzegovina
*Corresponding author: Prljača Naser, Franjevačka 2 Tuzla 75000, naser.prljaca@fet.ba

ABSTRACT: Variants of proportional navigation (PN) are perhaps mostly used guidance laws for tactical homing missiles. PN aims to generate commanding missile lateral acceleration proportional to line of sight (LOS) angular rate, so that missile velocity vector rotates in such a way to assure interception of a target. In order to generate commanding lateral accelerations, the guidance system needs measurements of LOS angular rate and the closing velocity between the missile and the target, or the missile velocity. A device which provides guidance information is referred to as the missile seeker. In the case of imaging based seekers (visible light (EO), infrared light (IIR)), LOS rate is estimated using imaging sensor, while closing or missile velocity is measured using appropriate sensors or guess estimated. In this paper, we present the design and simulation of a missile homing system which includes: true PN guidance law, linear multiloop acceleration autopilot, and gimbaled imaging based missile seeker. Target seeker uses advanced deep machine learning object detection YOLO (You only look once) model, for target detection and tracking as well as LOS rate estimation. Comprehensive simulation model, consisting of full 6DOF missile and controls dynamics, 3D world and camera model, is developed. Intensive simulation results show performances of the proposed missile homing system.

KEYWORDS Missile guidance, YOLO, Missile seeker, Proportional navigation, 6DOF

1. Introduction

Many missile guidance and control laws have been developed in the past. Reported missile autopilot designs have ranged from classical linear to the most advanced ones [1]–[4]. The most widely used guidance law among them is a variant of proportional navigation(PN) [5]–[8]. Homing missiles using a variant of proportional navigation (PN) guidance law require measurement of LOS angular rate and measurement of closing velocity or missile velocity. This implies that missile seeker is able to detect and track targets during engagement. In active homing missiles, radar based missile seeker is used to detect the target, track the target and measure LOS rate and closing velocity. In passive homing missiles, strap-down or gimbaled imaging sensor (EO/IIR) based missile seeker is used to detect the target, track the target and estimate LOS angular rate, while missile velocity is measured with an appropriate sensor (IMU). In this work, a two axis (pitch and yaw) gimbal carrying an imaging device keeps the target in the camera field of view (FOV) by means of a gimbal control tracking loop. Accurate measurement of tracking error is of essential importance for tracking accuracy and LOS rate estimation. Tracking error is obtained by locating targets within the image using appropriate image processing techniques. Thanks to rapid advances in deep machine learning(convolutional neural networks), high accuracy and high speed deep learning object detection models have emerged. The most prominent among them is the YOLO (You only look once) family [9]. This work investigates application of YOLO deep machine learning object detector for tracking error measurements and consecutive LOS angular rate estimation in missile seeker. YOLO performance assessment in guidance loop is realized by means of a comprehensive Matlab/Simulink simulator, which includes nonlinear 6DOF missile and controls dynamics as well as 3D world (scene) and imaging models. This work was motivated by the fact, that to the best of the authors knowledge, no published work has been found which deals with similar simulation study.

2. Missile 6-DOF dynamic model

In order to simulate missile guidance, a high-fidelity nonlinear model is needed. Cruciform missile can be modeled as a rigid body which is governed by a system of coupled nonlinear first order differential equations with constant coefficients:

$$\dot{P} = L/I_x \tag{1}$$

$$\dot{Q} = PR\left(I_z - I_x\right)/I_y + M/I_y \tag{2}$$

$$\dot{R} = PQ\left(I_x - I_y\right)/I_z + N/I_z \tag{3}$$

$$\dot{u} = vR - wQ + F_x/m \tag{4}$$

$$\dot{v} = wp - uR + F_y/m \tag{5}$$

$$\dot{w} = uQ - vP + F_z/m \tag{6}$$

Where u, v and w are missile velocities along x, y and z axes of the missile body frame. P, Q and R are rotational angular velocities about x, y and z axes of the missile body frame. F_x, F_y and F_z are forces which are acting along the respected body axes and L, M and N are moments which are acting along the respected body axes. I_x, I_y and I_z are moments of inertia about the missile body axes. Inertia cross products are zero since the missile airframe is symmetric. For a cruciform missile, aerodynamic forces can be approximated as follows [10]:

$$\begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \begin{bmatrix} C_{x_0} + C_{x_2}\left(\alpha^2 + \beta^2\right) \\ C_N\beta \\ C_N\alpha \end{bmatrix} \tag{7}$$

Total force on acting along the missile body axes can:

$$\begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \begin{bmatrix} T \\ 0 \\ 0 \end{bmatrix} + mg \begin{bmatrix} -\sin\theta \\ \sin\phi\cos\theta \\ \cos\phi\cos\theta \end{bmatrix} - qS \begin{bmatrix} C_{x_0} + C_{x_2}\left(\alpha^2 + \beta^2\right) \\ C_N\beta \\ C_N\alpha \end{bmatrix} \tag{8}$$

Where $q = \rho v_m^2/2$ is the dynamic pressure, q is the density of air at given altitude and v_m is the total velocity of the missile. The gravitational force is naturally given in earth reference frame so it has to be transformed to body axes frame. In the previous equation, C_{x_0} and C_N are aerodynamic coefficients calculated at the various angles of attack α and sideslip angles β. Angle of attack and sideslip angle can be calculated as follows:

$$\alpha = \arctan\frac{w}{v} \tag{9}$$

$$\beta = \arctan\frac{v}{u} \tag{10}$$

Furthermore, it is important to define Euler RPY angles which describe the orientation of the missile with respect to the inertial reference frame. Euler angles can be integrated from the following equation:

$$\begin{bmatrix} \dot\phi \\ \dot\theta \\ \dot\psi \end{bmatrix} = \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ 0 & \cos\phi & -\sin\phi \\ 0 & \frac{\sin\phi}{\cos\theta} & \frac{\cos\phi}{\cos\theta} \end{bmatrix} \begin{bmatrix} P \\ Q \\ R \end{bmatrix} \tag{11}$$

Moments acting on the missile are a consequence of rotation of actuator surfaces and of forces acting on the missile. Aerodynamic forces are acting on the point called the center of pressure, while the gravitational force is acting on the center of gravity of the missile. This causes a moment which rotates the missile. Therefore, the missile moments can be calculated as follows:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0 \\ -r_X F_{AZ} \\ r_X F_{AY} \end{bmatrix} + \frac{qS}{v_m} \begin{bmatrix} C_{LP}P \\ C_{MQ}Q \\ C_{NR}R \end{bmatrix} + qS \begin{bmatrix} C_{L\delta_E}\delta_E \\ C_{M\delta_V}\delta_V \\ C_{N\delta_P}\delta_P \end{bmatrix} \tag{12}$$

In the previous equation, r_x is the distance between center of gravity and center of pressure and F_{AY} and F_{AZ} are aerodynamic forces acting along the missile body axes. C_{LP}, C_{MQ} and C_{NR} are rolling, pitching and yawing aerodynamic coefficients respectively. $C_{L\delta_E}$, $C_{L\delta_V}$ and $C_{L\delta_P}$ are aerodynamic coefficients with respect to the angular displacement of actuator surfaces denoted by δ_E, δ_V and δ_P. Table 1 shows used missile parameters [11].

Table 1: Missile parameters

I_x	I_y, I_z	m	C_{x_2}
0.024 kgm^2	0.958 kgm^2	11.25 kg	0.484
C_{x_0}	$C_{N\delta_P}$	$C_{M\delta_V}$	$C_{L\delta_E}$
2.04	0.0905	0.0905	0.0905
C_N	C_{MQ}	C_{NR}	C_{LP}
3.298	−10	−10	0.0905
T	r_x	ρ	S
750 N	−0.119 m	1.225 kg/m^3	0.0314 m^2

3. Proportional navigation

The basic intuition of proportional navigation(PN) [12] is to generate commanding lateral accelerations which are proportional to the line of sight rate (LOS) angular rate as follows:

$$a_c = N\dot\lambda v_c \tag{13}$$

Where a_c is the commanded lateral acceleration normal to the line of sight, $N \geq 2$ is the effective navigational ratio and v_c is the relative velocity between the missile and the target. These commanded accelerations are then forwarded to the autopilot which makes sure that the missile achieves these lateral accelerations. In recent works, authors simulated this guidance method on 6-DOF missile simulator in Matlab in [13] and [14]. For even more insight into missile guidance and control, the reader is referred to [10], [11], [15]–[21]. As is shown in figure 1, in order to calculate λ or $\dot\lambda$ it is needed to measure relative distance and relative velocity between the missile and the target.

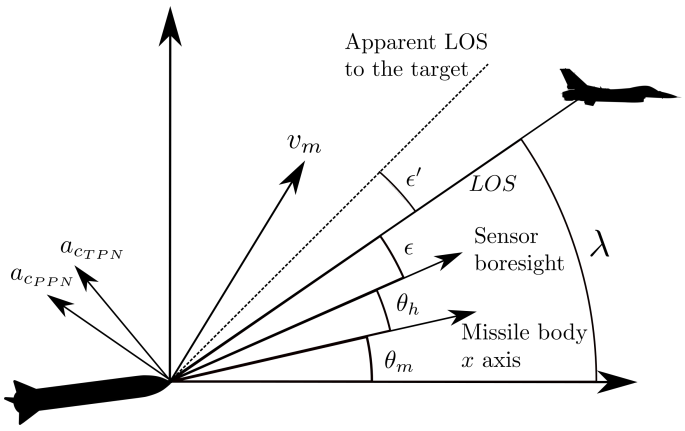


Figure 1: Geometry of proportional navigation

Unfortunately, these values cannot be measured without an active radar which most tactical missiles do not have. Similarly, the closing velocity v_c cannot be measured without an active radar. That is why this article implements another variant of PN called the pure proportional navigation whose commanding accelerations are normal to the missile velocity [12] and are given by:

$$a_c = N\dot\lambda v_m \tag{14}$$

Missile seeker is the device that has to measure LOS rate. The seeker is made out of two independent gimbal axes, two

servo motors and an imaging (EO/IIR) sensor. These two gimbals move across the vertical and the horizontal plane in order to lock onto the target. Once locked, the seeker can measure the tracking error, which is noted with ϵ in figure 1. Earlier work [14], [22]–[25] explain in great detail on how to obtain LOS and LOS rate from the measured tracking error. The LOS angular rate in vertical plane can be obtained using the following equation:

$$\dot{\lambda}_{z_m} = \epsilon_z \frac{s}{\tau s + 1} + \dot{\theta}_h + \dot{\theta}_m \tag{15}$$

Where θ_m is the Euler pitch angle of the missile relative to the inertial frame of reference, θ_h is the angular position of the gimbal servo relative to the missile centerline in vertical plane and ϵ_z is the measured tracking error in the vertical plane. Angular velocity $\dot{\theta}_m$ is easy to obtain using rate gyro as is explained in [14] and $\dot{\theta}_h$ has to be calculated from measured angular rates around the body axes as is explained in [13]. Similar to equation 15, LOS rate in the horizontal plane can be obtained as follows:

$$\dot{\lambda}_{y_m} = \epsilon_y \frac{s}{\tau s + 1} + \dot{\psi}_h + \dot{\psi}_m \tag{16}$$

Where ψ_m is the Euler yaw angle of the missile relative to the inertial frame of reference, ψ_h is the angular position of the gimbal servo relative to the missile centerline in the horizontal plane and ϵ_y is the measured tracking error in the horizontal plane. Most tactical missiles have infrared homing seeker. Because of this, tracking error cannot be measured directly but it is assumed to be indicated angular position of the target relative to the camera centerline or boresight [10]. This paper aims to simulate a seeker sensor which utilizes Matlab 3D camera, which would measure the tracking error from obtained images. Matlab interface for Unreal Engine allows simulations of 3D vehicles as well as taking images from Unreal Engine world. Also, implementation, simulation and control of gimbal servo system are presented in this article since it is guidance requirement to keep the tracking error as low as possible.

4. Autopilot design

If the missile lateral accelerations are exactly equal to the commanded accelerations commanded by the guidance subsystem, then the interception is guaranteed. Missile subsystem which will ensure this is referred to as the autopilot. From the missile dynamic model, it can be shown that there will be no cross-coupling between yaw and pitch motion if the missile is not rolling. Therefore, it is imperative that the autopilot ensures zero roll angle of the missile. Once that is achieved, motion in yaw and pitch plane can be controlled independently. Therefore, it is enough to implement three PID controllers. One PID controller to ensure roll stability and two PID controllers for two lateral accelerations. Figure 2 shows a roll controller design.

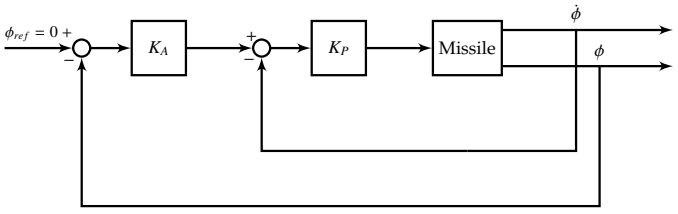


Figure 2: Roll controller design

Roll controller employs a cascade error regulation. The inner loop which feeds back roll derivative is a damping loop which ensures minimum oscillations in the roll channel. In order to achieve lateral acceleration tracking, an expression for calculating lateral acceleration is needed. In order to calculate them, a derivative of a rotating vector is needed and that vector has to be transformed into coordinate system whose x-axis coincides with the velocity vector. Such coordinate system is also called the wind frame. Therefore, accelerations in the wind frame are given by:

$$\begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} \cos\alpha\cos\beta & -\sin\beta & \sin\alpha\cos\beta \\ \sin\beta\cos\alpha & \cos\beta & -\sin\alpha\sin\beta \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix} \left(\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{w} \end{bmatrix} + \begin{bmatrix} P \\ Q \\ R \end{bmatrix} \times \begin{bmatrix} u \\ v \\ w \end{bmatrix} \right) \tag{17}$$

Where a_y and a_z are missile horizontal and vertical acceleration components normal to the velocity vector and a_x is the missile acceleration component along the velocity vector. This acceleration is usually zero since the missile does not have any fuel left during the terminal guidance phase. Figure 3 shows a controller for lateral accelerations. Similar design for horizontal and vertical acceleration components is employed. The inner loop reduces response oscillations and the outer loop ensures that the lateral acceleration follows the demanded lateral acceleration.

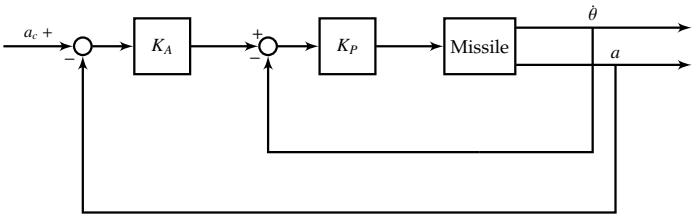


Figure 3: Lateral acceleration controller design

Missile aerodynamics coefficients change with respect to the Mach number, angle of attack, angle of sideslip, altitude and angular velocities, so it might be advisable to design multiple controllers for various operating points. This approach is referred to as gain scheduling [26, 27]. In his article fixed gain linear (PD) controllers are used regardless of missile homing operational conditions, due to their simplicity and robustness.

5. YOLO neural network

In order to calculate the tracking errors ϵ_y and ϵ_z, it is first required to localize object in the image. Furthermore, it is needed to classify the object in the obtained image. There is a handful of requirements that the localization algorithm

has to satisfy in missile guidance applications. First, during simulations images are obtained from Unreal Engine environment so the algorithm must be able to work with artificial images. Second, the algorithm has to be very fast since camera can generate images with very high frequency. Most suitable localization algorithm in this case is You Only Look Once(YOLO) convolutional neural network. YOLO was first introduced in 2016 where object detection was reframed as a single regression problem straight from image pixels to bounding box and class probabilities [9]. This made sure that YOLO needed only one passage through the network to obtain localization result which made it much faster then existing algorithms which were based on region proposals. Over time, a number of object detectors based on YOLO were developed. Most notably, YOLO9000 [28], YOLOv3 [29] and YOLOv4 [30] and YOLOv8 [31]. This article uses YOLOv4 since it is latest available in Matlab. YOLOv4 in Matlab is pretrained on Common Objects in Context(COCO) dataset [32] which consists of 80 classes and over 330000 images.

6. Camera model

Let us assume the pinhole camera model as is shown in the Figure 4 [33].

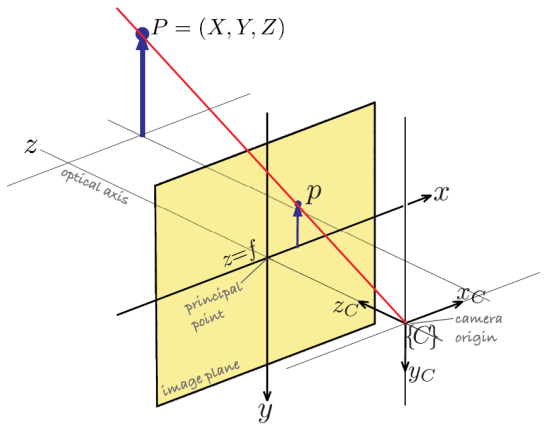


Figure 4: Pinhole camera model

Let X and Y denote the object size and let x and y denote size of the image in the image plane. Let f_x and f_y denote horizontal and vertical focal length measured in pixels which is also the distance of image plane from the CCD sensor. Inspecting Figure 4, it can be seen that tracking error angle in vertical plane can be evaluated as follows:

$$\epsilon_z = \arctan \frac{y}{f_y} \tag{18}$$

For vertical tracking error angle, the equation is analogous:

$$\epsilon_y = \arctan \frac{x}{f_x} \tag{19}$$

It is straightforward to implement calculation of the tracking errors. YOLO detector can obtain bounding box around the detected target with coordinates of the upper left corner(denoted with x_t and y_t) and its width and height(denoted with w and h). Now expressions for vertical and horizontal tracking errors are as follows:

$$\epsilon_z = \arctan \frac{y_t + h/2}{f_y} \tag{20}$$

$$\epsilon_y = \arctan \frac{x_t + w/2}{f_y} \tag{21}$$

7. Seeker camera gimbal servo controls

The camera is mounted on the head of the missile and it can rotate about the vertical and the horizontal axes on gimbal (yaw, pitch) using two DC motors. Rotational motions of these two motors is independent. Figure 5 shows mechanical structure of a missile seeker.

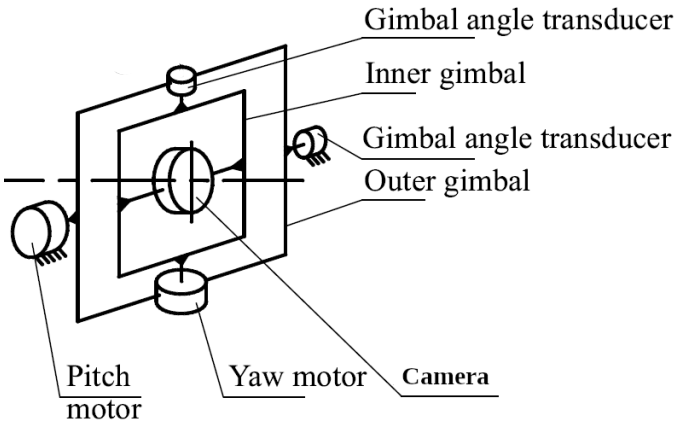


Figure 5: Seeker gimbal mechanical structure

The goal of these DC motor based servo systems is to nullify the tracking error so that the target is always at the center of the obtained image. Therefore, the target tracking problem can be viewed as a control problem where the tracking error is considered an output value with desired reference value equal to zero. This way, radome refraction error is minimized and better measurements are available. It is important to note that the gimbal servo control systems have to be sufficiently fast in order to quickly eliminate tracking errors. The servo motor can be modeled as a second order linear system as follows:

$$G(s) = \frac{\omega_n^2}{s\,(s + 2\xi\omega_n)} \tag{22}$$

Where ξ and ω_n are damping coefficient and system angular frequency respectively. The tracking loop dynamics bandwidth is very wide, usually about 100 rad/sec [10]. A standard PID controller is used for each gimbal axis. The controllers are tuned so that they achieve acceptable settling time and overshoot.

8. Missile guidance system synthesis

Each previous section presents a single part of the missile guidance system. The typical missile guidance system consists of guidance system, missile seeker, autopilot and

the missile itself. Figure 6 shows the block diagram of the missile guidance system.

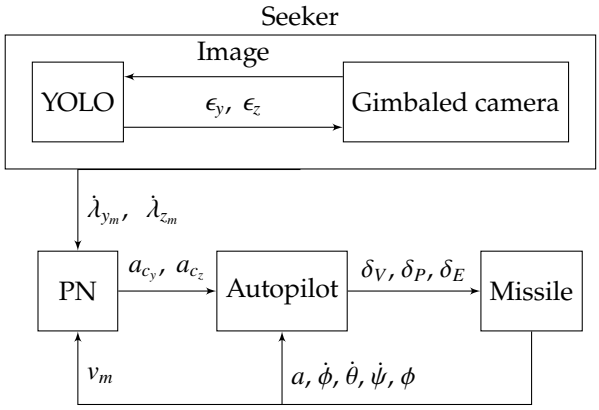


Figure 6: Missile guidance system design

Missile guidance systems design showed in the figure 6 only measures the image of the target. The remaining measurements can be obtained from the missiles internal sensors. Missile linear and angular velocities can be measured using inertial measurement units (IMU), from which Euler angles and their derivatives can be obtained using sensor fusion technique. Gimbal angles are measured by appropriate angle transducers.

9. Simulation

Previously described missile guidance system was implemented in Matlab and Simulink implementing previously described equations. Missile and the target position and orientation are fed into Unreal Engine blocks. Matlab offers Simulation 3D camera block to obtain images from the Unreal Engine scene. Simulation 3D camera block outputs an image which is fed to the YOLO block(implemented as a Matlab system) which calculates tracking errors given by equations 18 and 19. Camera generates 100 frames per second(fps) which is fast enough to capture tracking error dynamics. Obtained tracking errors are then fed into two PID controllers as error inputs. Controller outputs are fed into gimbal systems as inputs and gimbal positions are fed into camera as inputs to rotate the camera relatively to the missile body. Based on the equation 15, the seeker subsystem estimates the line of sight angular rate in the vertical and the horizontal plane. Subsystem for proportional navigation calculates the required lateral accelerations in vertical and horizontal plane. The autopilot subsystem calculates required elevator, rudder and aileron deflections.
Figure 7 shows trajectories of the target and the missile during the terminal guidance phase. The target is at the initial relative height of 200 meters and an initial relative distance of 500 meters. The target is escaping with the velocity 100m/s and moving sideways with the velocity 20m/s. It can also be seen that the missile is aiming at future intercept point rather than following the target. This means that PN requires lower commanded acceleration comparing to more traditional guidance methods.
Figure 8 shows one frame which is processed by YOLO detection algorithm. Here it can be seen that YOLO finds the bounding box around the object and that the object is in the center of the image.

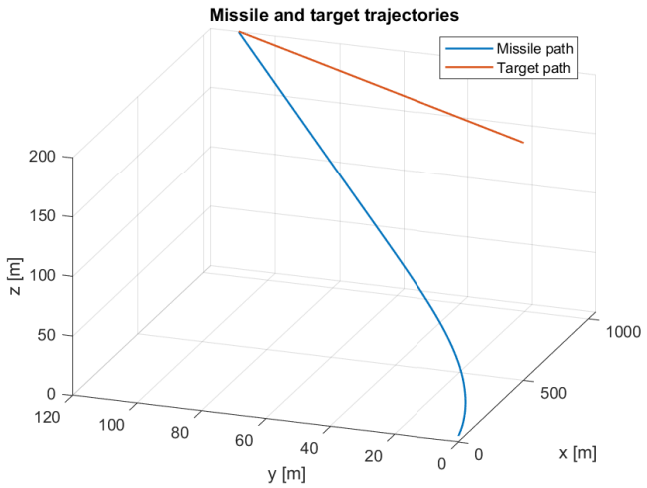


Figure 7: Missile and the target trajectories

Figure 8: Example of one image frame

Figure 9 shows obtained tracking errors. It can be seen that there are some initial oscillations due to the target moving which can be interpreted as output disturbances in the tracking errors.
Figures 10 and 11 show estimated and real LOS rates in the vertical and the horizontal plane respectively. It can be seen that the seeker estimates LOS rates fairly quickly and that they approach zero as the distance shortens which is a requirement for interception.
Figure 12 shows the missile Euler angles. It can be seen that the roll angle is always zero because the roll controller stabilizes the missile roll channel. The autopilot steers the missile so the missile has to keep an almost constant yaw and pitch angle in order to achieve a hit.
Figures 13 and 14 show commanded and achieved lateral accelerations. It can be seen that the autopilot subsystem ensures that the missile lateral accelerations are equal to the demanded PN accelerations.

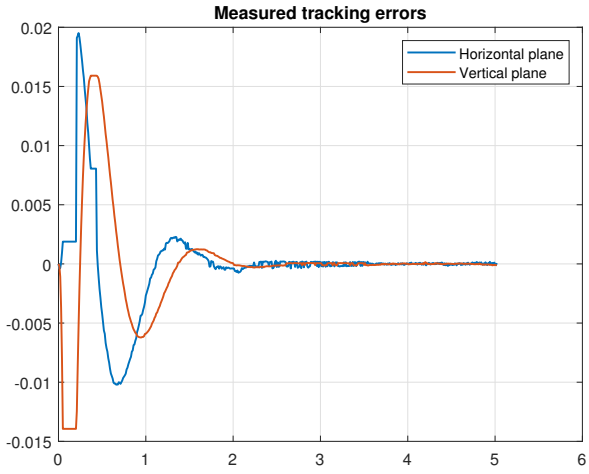


Figure 9: Tracking errors

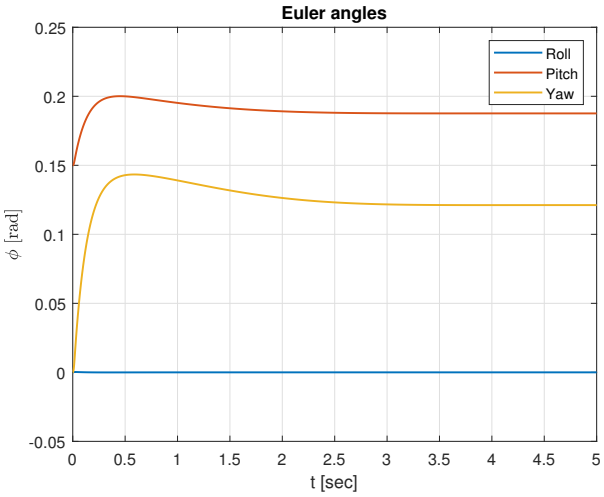


Figure 12: Missile Euler angles

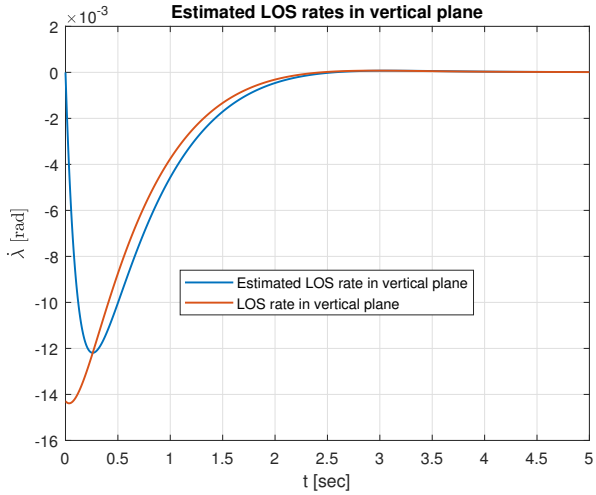


Figure 10: Estimated LOS rate in vertical plane

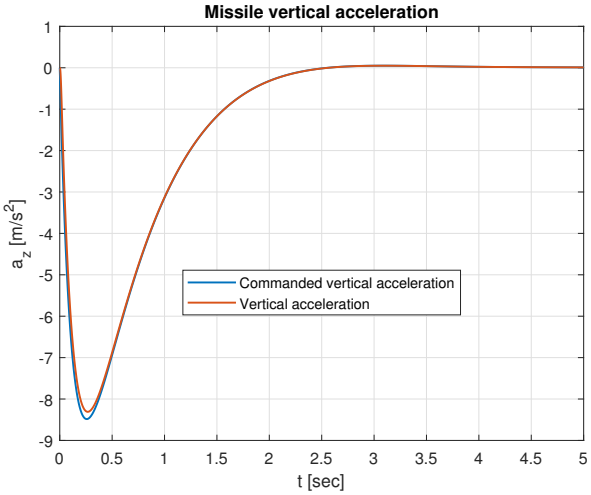


Figure 13: Missile vertical lateral acceleration

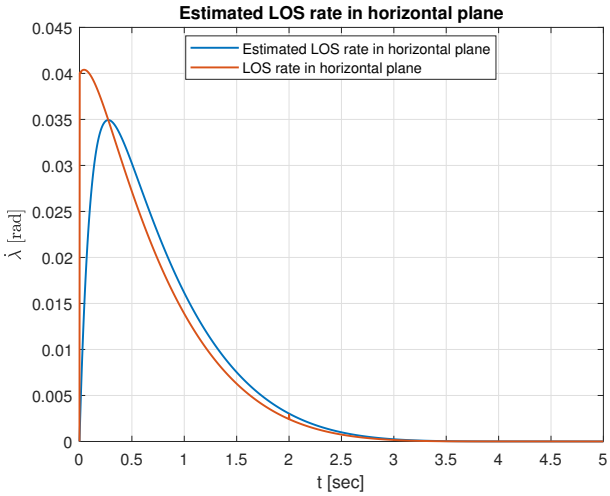


Figure 11: Estimated LOS rate in horizontal plane

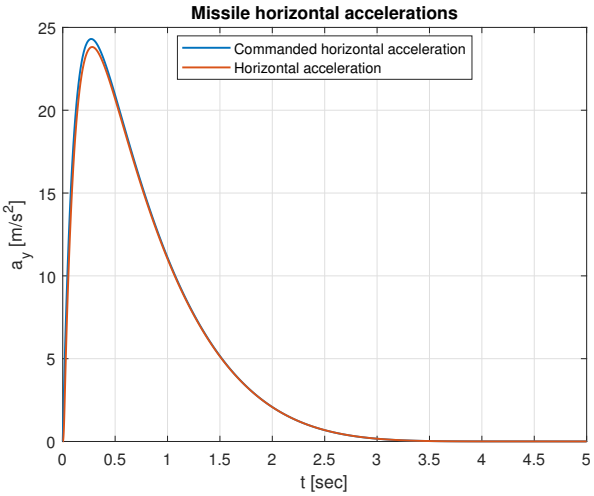


Figure 14: Missile horizontal lateral acceleration

Table 2: Missile performance table

Initial yaw heading error	Initial pitch heading error	Target horizontal acceleration	Target vertical acceleration	Miss distance	Maximum required horizontal acceleration	Maximum required vertical acceleration
−0.15 rad	0.15 rad	0	0	0 m	1.0384 m/s^2	9.81 m/s^2
−0.15 rad	0.3 rad	0	0	0 m	1.2325 m/s^2	9.6998 m/s^2
0 rad	0 rad	0.15g	0	0 m	1.8232 m/s^2	43.6711 m/s^2
0 rad	0 rad	0.3g	0	0 m	3.6392 m/s^2	43.6711 m/s^2
0 rad	0 rad	0.45g	0	0 m	5.4415 m/s^2	43.6711 m/s^2
0 rad	0 rad	0.6g	0	2.7509 m	7.2246 m/s^2	43.6711 m/s^2
0 rad	0 rad	0	0.15g	0 m	22.7714 m/s^2	43.6562 m/s^2
0 rad	0 rad	0	0.3g	0 m	23.2802 m/s^2	43.6562 m/s^2
0 rad	0 rad	0	0.45g	13.3609 m	23.2802 m/s^2	43.6562 m/s^2

10. Performance analysis

Proportional navigation can guarantee interception of a target moving with constant velocity. If the target is maneuvering or accelerating, there can be nonzero miss distance. Proportional navigation can ensure interception even if there is an initial heading error, although this requires the missile to have more energy available and larger demanded accelerations. Table 2 shows performance for various heading errors and target accelerations. Final miss distance and maximum required accelerations are chosen as a measure of performance. It can be seen that proportional navigation can nullify the miss distance due to the initial heading error given that the missile can achieve demanded lateral acceleration. If the target is accelerating, the proportional navigation may not be able to nullify the miss distance.

11. Conclusion

This article presents design and simulation of missile homing system. The missile homing system consist of guidance law, autopilot and imaging based target tracking subsystem (missile seeker). In this work the guidance law is based on true proportional navigation. Autopilots stabilize roll angle and regulate horizontal and vertical accelerations. They are standard linear multi loop controllers. Two axes (pitch, yaw) gimbaled seeker use standard PID controllers to track the target. The seeker uses state-of-art deep machine learning YOLO (You Only Look Once) object detector for tracking error measurements and LOS rate estimation. Comprehensive homing missile simulator is developed, and simulation results show robust accuracy of the proposed scheme to target velocity, launch heading errors and low level target maneuvering. Our future research work will focus on design and simulation of advanced guidance and control laws, as well as design and simulation of advanced machine learning algorithms.

References

[1] M. Abd-elatif, L. jun Qian, Y. ming Bo, "Optimization of three-loop missile autopilot gain under crossover frequency constraint", *"Defence Technology"*, vol. 12, no. 1, pp. 32–38, 2016, doi:https://doi.org/10.1016/j.dt.2015.08.006.

[2] B. Zhang, Q. Lv, Y. Lei, "The application of pid controller in missile longitudinal loop system and its simulation", *"Proceedings of the 5th International Conference on Advanced Design and Manufacturing Engineering"*, pp. 2113–2118, Atlantis Press, 2015/10, doi:10.2991/icadme-15.2015.395.

[3] A. Awad, H. Wang, "Roll-pitch-yaw autopilot design for nonlinear time-varying missile using partial state observer based global fast terminal sliding mode control", *"Chinese Journal of Aeronautics"*, vol. 29, no. 5, pp. 1302–1312, 2016, doi:https://doi.org/10.1016/j.cja.2016.04.020.

[4] A. Thukral, M. Innocenti, "A sliding mode missile pitch autopilot synthesis for high angle of attack maneuvering", *"IEEE Transactions on Control Systems Technology"*, vol. 6, no. 3, pp. 359–371, 1998, doi:10.1109/87.668037.

[5] C.-D. Yang, C.-C. Yang, "A unified approach to proportional navigation", *"IEEE Transactions on Aerospace and Electronic Systems"*, vol. 33, no. 2, pp. 557–567, 1997, doi:10.1109/7.575895.

[6] S. A. Murtaugh, H. E. Criel, "Fundamentals of proportional navigation", *"IEEE Spectrum"*, vol. 3, no. 12, pp. 75–85, 1966, doi:10.1109/MSPEC.1966.5217080.

[7] M. Guelman, "A qualitative study of proportional navigation", *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-7, no. 4, pp. 637–643, 1971, doi:10.1109/TAES.1971.310406.

[8] M. Guelman, "Proportional navigation with a maneuvering target", *"IEEE Transactions on Aerospace and Electronic Systems"*, vol. AES-8, no. 3, pp. 364–371, 1972, doi:10.1109/TAES.1972.309520.

[9] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection", *https://arxiv.org/abs/1506.02640*, 2016.

[10] G. M. Siouris, *Missile Guidance and Control Systems*, Springer-Verlag New York, 2004.

[11] J. Harris, N. Slegers, "Performance of a fire-and-forget anti-tank missile with a damaged wing", *"Mathematical and Computer Modelling"*, vol. 50, no. 1, pp. 292–305, 2009.

[12] U. Shukla, P. Mahapatra, "The proportional navigation dilemma-pure or true?", *"IEEE Transactions on Aerospace and Electronic Systems"*, vol. 26, no. 2, pp. 382–392, 1990.

[13] M. Hodžić, N. Prljača, "Simulation of short range missile guidance using proportional navigation", *"2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)"*, pp. 1–6, 2021.

[14] M. Hodžić, N. Prljača, "Los rate estimation techniques for proportional navigation guided missiles", *"2022 21st International Symposium INFOTEH-JAHORINA (INFOTEH)"*, pp. 1–6, 2022.

[15] P. Zarchan, *Tactical and Strategic Missile Guidance*, no. v. 219 in AIAA Tactical Missile Series, American Institute of Aeronautics and Astronautics, 2007.

[16] R. Yanushevsky, *Modern Missile Guidance*, Taylor & Francis, 2007.

[17] N. Shneydor, *Missile Guidance and Pursuit: Kinematics, Dynamics and Control*, 1, Elsevier Science, 1998.

[18] M. Drela, *Flight Vehicle Aerodynamics*, FLIGHT VEHICLE AERODYNAMICS, MIT Press, 2014.

[19] S. Graovac, *Automatsko vođenje objekata u prostoru*, Akademska Misao, 2005.

[20] C.-F. Lin, *Modern Navigation, Guidance and Control Processing*, Prentice Hall, 1991.

[21] M. V. Cook, ed., *Flight Dynamics Principles (Third Edition)*, Butterworth-Heinemann, third edition ed., 2013.

[22] N. F. Palumbo, R. A. Blauwkamp, J. M. Lloyd, "Basic principles of homing guidance", *"Johns Hopkins Apl Technical Digest"*, vol. 29, pp. 25–41, 2010.

[23] Q. Zaikang, L. Defu, *Design of Guidance and Control Systems for Tactical Missiles*, CRC Press, 1st ed., 2019.

[24] S. Mondal, S. Sadhu, A. Banerjee, "Platform motion disturbances attenuation in a missile seeker subsystem using internal model control", *"2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE)"*, pp. 1–4, 2013.

[25] S. He, Y. Liang, J. Tang, Z. Bai, K. Li, "Homing guidance law design against maneuvering targets based on ddpg", *"International Journal of Aerospace Engineering"*, vol. 2023, p. 4188037, 2023.

[26] D. White, J. Wozniak, D. Lawrence, "Missile autopilot design using a gain scheduling technique", *"Proceedings of 26th Southeastern Symposium on System Theory"*, pp. 606–610, 1994.

[27] A. Hiret, G. Duc, J. Bonnet, "The application of gain-scheduling h controllers for a missile autopilot", *"IFAC Proceedings Volumes"*, vol. 31, no. 21, pp. 59–64, 1998, 14th IFAC Symposium on Automatic Control in Aerospace 1998, Seoul, Korea, 24-28 August 1998.

[28] J. Redmon, A. Farhadi, "Yolo9000: Better, faster, stronger", *https://arxiv.org/abs/1612.08242*, 2016.

[29] J. Redmon, A. Farhadi, "Yolov3: An incremental improvement", *https://arxiv.org/abs/1804.02767*, 2018.

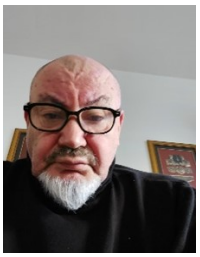
[30] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection", *https://arxiv.org/abs/2004.10934*, 2020.

[31] D. Reis, J. Kupec, J. Hong, A. Daoudi, "Real-time flying object detection with yolov8", 2023.

[32] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, "Microsoft coco: Common objects in context", *https://arxiv.org/abs/1405.0312*, 2015.

[33] P. Corke, *Robotics, Vision and Control - Fundamental Algorithms in MATLAB®*, vol. 73 of *Springer Tracts in Advanced Robotics*, Springer, 2011.

MIRZA HODŽIĆ was born in Tuzla, Bosnia and Herzegovina in 1997. He finished his bachelor's degree from the Faculty of electrical engineering in Tuzla in 2020. He has done his master's degree from Faculty of electrical engineering Tuzla in 2022 at the department of control systems and robotics. Besides working in the industry he is also working at the Faculty of electrical engineering at the department of control systems , robotics and industrial informatics. His research interests are in robotics, control systems, machine vision and missile guidance. Currently, he is working on his PhD in control systems at the Faculty of electrical engineering at the University of Tuzla.



NASER PRLJAČA is full professor and head of department of Control Systems, Robotics and Industrial Informatics at the Faculty of Electrical Engineering, University of Tuzla. He holds BSc, MSc and PhD in Electrical and Electronics Engineering. He has been taking part in numerous research and industrial projects, and is author of a number of academic publications. His research and teaching interests include control systems, embedded systems, robotics, machine learning and computer vision.

Received: 21 January, 2024, Revised: 22 March, 2024, Accepted: 23 March, 2024, Online: 16 April, 2024

DOI: https://doi.org/10.55708/js0304001

Estimation of Elbow Joint Movement Using ANN-Based Softmax Classifier

Abdullah Y. Al-Maliki [*,1], **Kamran Iqbal** [1], **Gannon White** [2]

[1]Department of Electrical and Computer Engineering, University of Arkansas at Little Rock, Little Rock, 72204, USA

[2]Department of Kinesiology, Colorado Mesa University, Grand Junction, 81501, USA

*Corresponding author: Dr. Abdullah Al-Maliki, ayalmaliki@gmail.com

ABSTRACT: Estimating the natural voluntary movement of human joints in its entirety is a challenging problem especially when high accuracy is desired. In this paper, we build a modular estimator to estimate the elbow joint motion including angular displacement and direction. Being modular, this estimator can be scaled for application to other joints. We collected surface Electromyographic (sEMG) signals and motion capture data from healthy participants while performing elbow flexion and extension in different arm positions and at different effort levels. We preprocessed the sEMG signals, extracted features array, and used it to train an ANN-based Softmax classifier to estimate the angular displacement and movement direction. When compared against the motion cap-ture data, the classifier achieved estimation accuracy ranging from 80% to 90% with a resolution of 5°, which translates into Pear-son Correlation Coefficient (PCC) ranging from 0.91 to 0.95. Such high PCC values in mimicking the voluntary movement of the upper limb may help toward building intuitive prostheses, exoskeletons, remote-controlled robotic arms, and other Human Ma-chine Interface (HMI) applications.

KEYWORDS: Elbow Angle Estimation, ANN, Softmax Classifier, sEMG, Signal Processing

1. Introduction and Literature Review

An Individual's intent to perform any physical movement that involves the contraction of skeletal muscles can be estimated by analyzing the corresponding muscles' electromyography (EMG) signal [1,2]. An EMG signal is the electrical manifestation of the neuromuscular activation associated with a muscle contraction [3]. EMG signals are characterized as a stochastic and nonlinear signal due to a large number of motor units and their different firing rates, as it carries the brain command to create a muscle contraction [4]. Since the EMG signal has a low signal-to-noise ratio in addition to its stochastic nature, it is hard to fully utilize the information embedded in it with high accuracy without implementing various filtering, dimension reduction, and/or pattern recognition techniques [4,5]. EMG can be measured by using one of two methods; surface EMG or sEMG and intramuscular EMG. sEMG is measured on the skin above the muscle of interest using an electrode to measure the muscle's electrical activity i.e. EMG signal [6]; therefore, this technique is categorized as noninvasive [6,7]. Intramuscular EMG, on the other hand, is measured from within the muscle of interest by inserting thin metal electrodes into the muscle and referenced to the surface electrode placed above the muscle [8]. Consequently, it requires more caution than sEMG to perform and it is considered an invasive procedure [9,10]. As a result, sEMG is the dominant method to study how the human body controls skeletal muscles as it is safer, easier to perform, and can be implemented in end products with less complexity [7,10].

This work focuses on estimating the direction and angular displacement of the elbow's joint flexion and extension movement using the information embedded within sEMG data of the related muscles. Below we mention some of the related research addressing the same problem using various methods. The authors trained an Artificial Neural Network (ANN) to estimate the upper limb joint angles from related muscles' sEMG data, in their experiment setup the upper arm was held fixed at the shoulder level [11]. Their ANN was able to follow the

target angle in training data, and they showed low average Mean Square Error (MSE) for eight participants. The authors measured the upper limb joint angles, muscles EMG and set a constant load to estimate the angle-torque relation during constant muscle activation in able-bodied humans using ANN [12]. The authors showed the effectiveness of Kalman Filter in linearizing a set of twelve time-domain features. They estimated the elbow joint angle during elbow flexion with a fixed load and position which in turn increased the elbow angle estimation accuracy [13]. The authors used least squares support vector regression to estimate the knee and hip joint angles using sEMG signals [14]. Their data was from able-bodied participants performing treadmill and leg extension exercises. They were able to achieve single-digit Root Mean Square Error (RMSE) between measured and estimated joint angles in training data. The authors were able to estimate the elbow joint angle in able-bodied participants, using multiple time-delayed features of the sEMG and random forests [15]. They were able to achieve single-digit MSE but they had a fixed time window, with strictly one degree of freedom, fixed arm effort, and fixed arm orientation. The same authors from [15] used grey feature weighted least square support vector machine algorithm to estimate the elbow joint angle [16]. Similar to [15], the estimation was for one degree of freedom while the elbow was fixed on the table, and the starting angle between the forearm and the upper arm was 90°and the wrist joint was kept along the forearm and again all the movements were done with empty hand i.e. single effort level. The authors used EMG to model muscle activation with the help of a feed-forward ANN and Gaussian process to estimate the fingers' joint angles of ten able-bodied participants [17]. Their estimation involved more than one degree of freedom, but was held constant at one effort level. The authors used Radial Basis Function ANN to estimate the elbow joint angle for a flexion and extension movement in healthy participants [18]. Their experimental setup included fixed elbow location, fixed elbow orientation, and a fixed effort level. They achieved correlation coefficient values between actual and estimated angle ranging from 0.76 to 0.91 depending on the movement speed.

The above literature review shows that most, if not all, of the related research, did not account for the following factors when trying to estimate the elbow's joint movement:

1. Different limb orientation when performing the movement
2. Carried load changes between trials of the same movement
3. Movement speed variation between trials

Therefore, the goal of this work is to estimate the human elbow joint displacement and direction using sEMG signals for one degree of freedom at two different limb orientations, two different load levels, and variable movement speed, thus expanding on the previously mentioned studies. By developing such a versatile joint angle estimator, we broaden the utilization of sEMG signal data in different Human Machine Interface (HMI) application. For an instant prosthesis development has, in a way, surpassed the current limb action estimation techniques; as modern prostheses have multiple degrees of freedom, actuated by motors capable of fine movements [19], while the majority of the cur-rent literature focuses on the estimation of the limb torque during a muscle contraction and not so on the final location of the limb [19–21]. The estimator developed in this work is designed with a modular approach from the software point of view; so it can be scaled to other joints and their corresponding muscle groups, e.g., the wrist or the shoulder.

2. Elbow Joint Movement Estimation

In this section, we describe the biomechanics of the elbow joint and which muscles to focus on to estimate the elbow flexion and extension movement, we also discuss the sEMG signal characteristics and the procedure used in collecting sEMG data for this study. Then we show the methods used to preprocess this data and the building blocks of the elbow angle estimator, and by the end of this section, we show the estimation results of the developed estimator and the correlation factor between the actual and estimated elbow angles.

2.1. Elbow Joint Anatomy and Biomechanics

Healthy humans take their elbow joint function for granted, but as researchers try to tackle this joint, they realize how complicated this joint is. The elbow joint consists of and is a pathway to a lot of different types of bone structures, tissues, vessels, and fibers [22,23], however in this work we are focusing on the muscles responsible for flexing and extending the joint which can be divided into two groups [23].

1. The prime movers and heavy lifters of the forearm are [23]:
   - A. Biceps (Elbow Flexion)
   - B. Brachialis (Elbow Flexion)
   - C. Triceps (Elbow Extension)
2. The secondary movers and supporting muscles are [23]:
   - A. Brachioradialis (Elbow Flexion)
   - B. Pronator teres (Elbow Flexion)
   - C. Anconeus (Elbow Extension)

The location of the three heavy lifting muscles in number 1 above with respect to the upper limb is shown in Figure 1 [22].

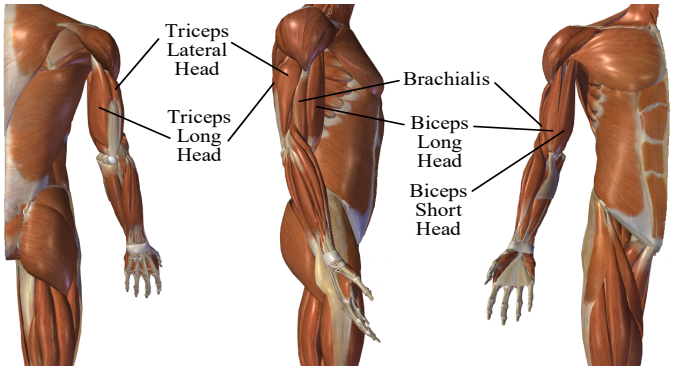


Figure 1: Shows the location of elbow joint muscles targeted in this study [22].

What is interesting is that the body utilizes the elbow muscles in a selective, highly case-dependent manner [22]. For unconstraint and unloaded forearm moving at medium speed, and when the arm is on the side of a standing hu-man, biceps, and brachialis are mainly used to raise and lower the arm, and triceps is mainly used to stabilize the elbow. This movement is known as biceps curls. And the opposite is true when the hand is raised above the shoulder and the arm is extended to the top. In this case, the triceps is mainly used to extend the arm, biceps and brachialis are used to stabilize the elbow, this movement is known as triceps extension [23].

2.2. sEMG Signal Limitations

Many researchers agree that sEMG is the most common and most feasible way to study limbs' movements and to control prosthetic limbs [24]. Nevertheless, getting useful data from the stochastic sEMG signal has its own challenges [25]. Common disadvantageous characteristics and noise sources of sEMG signal are [26–30]:

1. sEMG signal is Quasi-Random in nature
2. High noise to signal ratio (NSR)
3. Cross-talk noise, which can be mitigated by using smaller footprint sensors and by accurately targeting the muscles in question
4. Inherent equipment and electrode noise (can be estimated to a certain degree from resting trials of sEMG data re-cording session)
5. Inherent electromagnetic noise from the environment e.g. electric power, (part of which can be figured out from the power line frequency, 60 Hz in the USA)
6. Motion artifact noise (usually high amplitude, comparable to sEMG signal and low frequency 1-10 Hz)
7. Heartbeat artifact noise, which is usually very consistent with Electrocardiographic (ECG) signal.
8. Other biological noise sources, e.g. blood flow velocity, fat index, and skin temperature.
9. Fatigue resulting from task repetition and/or increasing the load has its toll on the sEMG signal.
10. sEMG signal distribution is highly non-Gaussian at low and high levels of force, whereas the distribution has its maximum gaussianity at the mid-level of maximum voluntary contraction (MVC). Therefore some techniques may work well at certain levels of the contraction power and fail at different levels [28].

It is obvious that most of the above-mentioned drawbacks are biological noises, some of which can be dealt with using different kinds of filters, and others will need different approaches to get more accurate data from the sEMG.

2.3. Data Collection

Since we can reflect the sEMG signal processing and classification from able-bodied to amputees with not much loss of accuracy [31,32], in this study we will only collect data from volunteer able-bodied participants. The data collection was done in accordance with the Institutional Review Board (IRB) guidelines. Our study was approved by IRB with protocol number 18-134-R1. After describing the whole procedure to the volunteers and signing the consent form, four silver/silver chloride (Ag/AgCl) electrodes were put on the muscles responsible for flexing and extending the elbow joint. The first channel was assigned to the biceps between the long-head and short-head, the second to brachialis, the third to triceps lateral, and the fourth and last channel was assigned to triceps long head. Retro-reflective markers were placed on the right arm of the participants to utilize the Vicon motion capture system to measure the elbow joint angular displacement. Noraxon TeleMyo Direct Transmission System (Noraxon USA Inc) was used to record the sEMG and motion data. The sEMG sampling frequency was set to 1500 Hz. Each participant was asked to perform fifteen arm flexing and extension movements at their own pace in three different arm positions:

A. The arm is relaxed by the participant side, a.k.a biceps curl movement
B. The arm abducted to shoulder height to the side of the participant, in which the elbow flexes and extends in the horizontal plane, perpendicular to the gravitational force.
C. Arm above the shoulder, i.e. overhead arm extension movement a.k.a triceps extension.

All three of the above movements were done with an empty hand and while holding a 5-pound weight (two effort levels). We also recorded sEMG data for each arm position while it's idle i.e., with no movement. Each trial lasted approximately 20 seconds which is the average time the participants took to complete 15 repetitions.

We allocate 80% of the acquired data for training the classifier and 20% for testing it. We also implement a five-fold cross-validation method on the whole data to verify the effectiveness of the classifier across all sections of the data.

2.4. Signal Preprocessing

As mentioned in Section 2.2, sEMG signal suffer from different kinds of noise; to address some of these signal imperfections, preprocessing is a must to obtain useful information out of the signal. In this case, we have implemented the following:

1. Removal of DC bias.
2. Notch filters to filter out power line fundamental and harmonic frequencies.
3. Bandpass filter allowing frequencies ranging from 30-400Hz to pass and has a low band stop frequency of 25Hz and a high band stop frequency of 450Hz.
4. Scaling up the signal by a factor of two.

2.5. Elbow Joint Motion Estimator

In this section, we design the estimator which will estimate the elbow joint's angular displacement magnitude in degrees, as well as its movement direction. Furthermore, it needs to do so while being insensitive to the elbow's angu-lar speed, orientation, or load changes.

The designed estimator is a classifier at its core, and to be more specific it is an ANN-based Softmax classifier. The softmax function is shown in (1) [33]:

$$P(c_r|x) = \begin{cases} \frac{P(x|c_r)P(c_r)}{\sum_{j=1}^{k} P(x|c_j)P(c_j)} & , (General\ Form) \\ P(c_r|x) = \frac{e^{a_r}}{\sum_{j=1}^{k} e_j^a} & , \begin{pmatrix} Normalized \\ Exponential\ form \end{pmatrix} \end{cases} \quad (1)$$

Where

$a_r = \ln\big(P(x|c_r)P(c_r)\big)$

$P(x|c_r)$ is the conditional probability of the sample x given class r and it has to satisfy the following two conditions:

$0 \le P(c_r|x) \le 1 \quad , \quad \sum_{j=1}^{k} P(c_j|x) = 1$

$P(c_r)$ is the class prior probability

and the ANN is trained using the scaled conjugate gradient algorithm, [34] provides an elaborate explanation on this algorithm. The selected loss function for this training is cross-entropy which is given in (2) [33]:

$$E = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} t_{ij}\ln(y_{ij}) + (1 - t_{ij})\ln(1 - y_{ij}) \quad (2)$$

Where

E is the loss function

N is the number of samples

K is the number of classes

$t_{i,j}$ means the i^{th} sample belongs to the j^{th} class

y_{ij} is ANN response for sample i for class j

As mentioned in section 2.2 sEMG has some random properties and inherited noises associated with it; thus, before feeding the sEMG data to the estimator it needs to undergo a feature extraction process. This process serves two key purposes:

1. Dimension reduction and normalization.
2. Amplify useful data and suppress noises.

Then the classifier will use the extracted features array instead of the sEMG data to do the classification.

We have chosen the following time-domain features:

1. Mean Absolute Value; $MAV = \frac{1}{N}\sum_{n=1}^{N}|x_n|$
2. Number of Zero Crossings; $NZC = \sum_{n=1}^{N-1} s(x_i, x_{i+1})$

   Where $s(x,y) = \begin{cases} 1\ if (x \cdot y) < 0 \\ 0\ if\ (x \cdot y) \ge 0 \end{cases}$
3. Number of Slope Sign Changes; $SSC = \sum_{n=2}^{N-1} ss(Sl_i, Sl_{i+1})$

   Where $ss(x,y) = \begin{cases} 1\ if (x \cdot y) < 0 \\ 0\ if\ (x \cdot y) \ge 0 \end{cases}$ and $Sl_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$
4. Root Mean Square Value; $RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2}$
5. Variance within each Channel; $ChVar = \frac{1}{N-1}\sum_{n=1}^{N}|x_n - \mu|^2$

   Where $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$
6. Variance Across Channels; $AChVar = \frac{1}{nCh-1}\sum_{n=1}^{nCh}\left|x_{Ch_n} - \mu\right|^2$

   Where $\mu = \frac{1}{nCh}\sum_{n=1}^{nCh} x_{Ch_n}$
7. RMS Difference between Agonist (Ag) and Antagonist $(Antg)$ Muscles; $RMS_{Diff} = \sum_{n=1}^{N_{Ag}} RMS_n - \sum_{n=1}^{N_{Antg}} RMS_n$
8. Integration of absolute value of sEMG; $I|EMG| = \sum_{n=1}^{N}|x_n| * t_s$

   Where: $t_s = \frac{1}{sF}$; sF is the sampling frequency

9. Signal Power; $SP = \frac{\|x\|^2}{N}$

   Where $\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2}$

10. Average of the Signal RMS Envelope; $RMS_{mean} = \frac{1}{N}\sum_{n=1}^{N} RMS_n$

And the following frequency-domain features:

1. Mean Signal Frequency; $mF = \frac{\sum_{n=1}^{F} SP_n \cdot f_n}{\sum_{n=1}^{F} SP_n}$
2. Average Waveform Length; $W_L = \frac{1}{mF}$

From section 2.3 and the previously mentioned features, we have a 4×12 feature array for each time window. This array has a fixed size regardless of the time window length and therefore helps the classifier to be indifferent to the joint motion speed. Thus, the input layer of the ANN classifier is set to be 48. One positive side effect of having multiple signal features is the compensation for a relatively low number of sEMG signal electrodes [35], as we examined only four muscles in this study.

Since the angular range of motion of a healthy human's elbow joint flexion-extension movement is [0°,145°] [36], we have selected the classes to be 0°, -5°, and +5° to keep the estimation resolution at 5° or 3.45% at a theoretical 100% classification accuracy. Where 0° represents a no-movement, -5° represents 5° flexion movement, and +5° represents 5° extension movement. However, 100% classification accuracy is hard to achieve especially with sEMG data, therefore, we have added two more classes -10°, and +10° to boost the overall classification accuracy. This is true because classes having the same direction but different displacements have more features similarity than classes with the same displacement and different direction. Therefore, the classifier is more likely to miss classify +5° as +10° than to miss classify +5° as -5° and vice versa. So, adding these extra two classes will aid in keeping the classification accuracy within 5°. As a result, the output layer of the ANN classifier is set to be 5, corresponding to 5 distinct classes. We have chosen Matlab R2020a to build the estimator and hence the classifier; Figure 2 shows a block diagram of the designed ANN classifier, Figure 3 shows the data flow throughout the elbow joint motion estimation system.

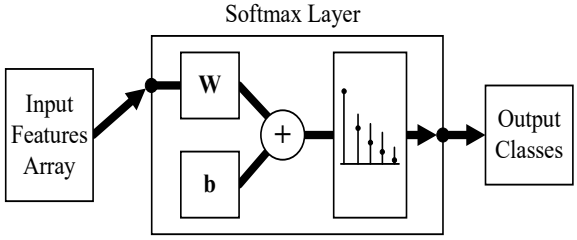


Figure 2: The ANN softmax classifier block diagram.

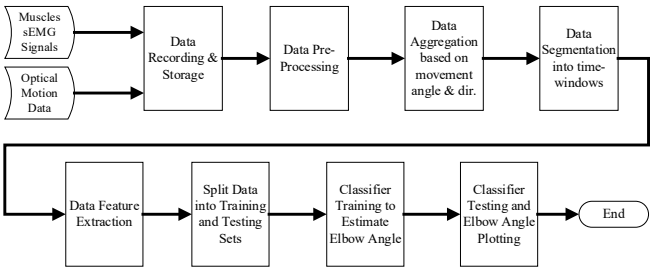


Figure 3: The elbow joint motion estimator flow chart.

2.6. Estimation Results

This section shows the results of the elbow joint movement estimation. We first begin by training the ANN soft-max classifier for the training data set for a maximum of 4000 epochs. Training data and testing data of biceps curl motion confusion matrices are shown in Figure 4 and Figure 5 respectively.

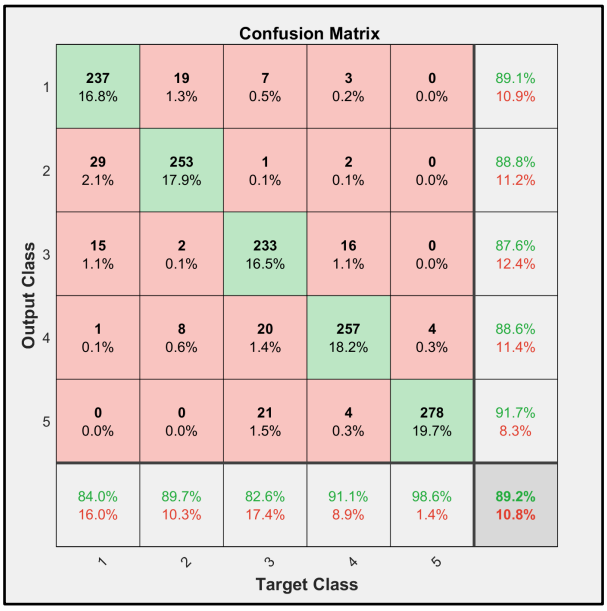


Figure 4: Classifier's confusion matrix for the training data.

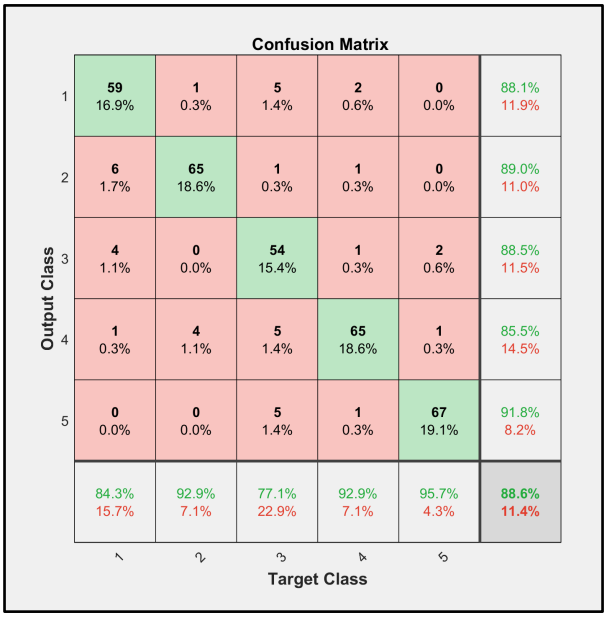


Figure 5: Classifier's confusion matrix for the testing data.

The order of the classes at the output layer of the classifier is as follows +5°, +10°, -5°, -10°, and 0°. If we look closer into Figure 4 and Figure 5 we can notice that most of the confusion happens between the same direction motions. So, if we do a simple probability analysis on the testing data, we can see that adding the two longer displacement classes helps in keeping the error within 5°.

$$k_{+5^\circ} = 88.1\% \; ; k_{+10^\circ} = 89\%$$
$$k_{-5^\circ} = 88.5\% \; ; k_{+10^\circ} = 85.5\%$$
$$k_{+5^\circ or +10^\circ} = 88.1\% + 89\% - (88.1\% \times 89\%) = 98.69\%$$
$$k_{-5^\circ or -10^\circ} = 88.5\% + 85.5\% - (88.5\% + 85.5\%) = 98.33\%$$

Adding two longer displacement classes helps in increasing the accuracy of the classifier in determining the direction to more than 98%. This is true because most of the confusion happens between the same direction movement as they have more similarity across their features.

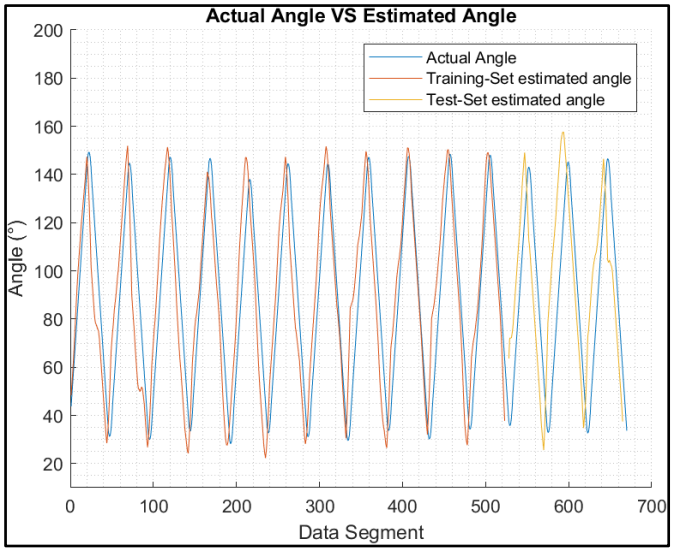


Figure 6: Biceps curls with a 5 pound weight target angle versus estimated angle (confusion for training and testing data are 89.2% and 88.6% respectively).

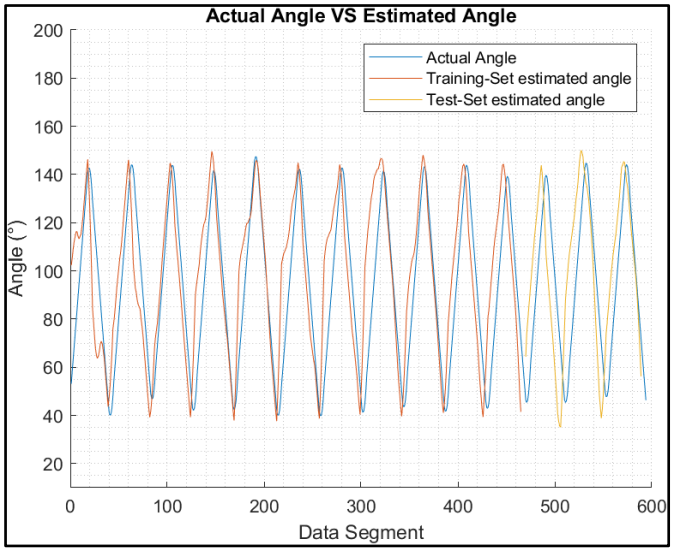


Figure 7: Biceps curls with empty-handed target angle versus estimated angle (confusion for training and testing data are 88.5% and 80% respectively).

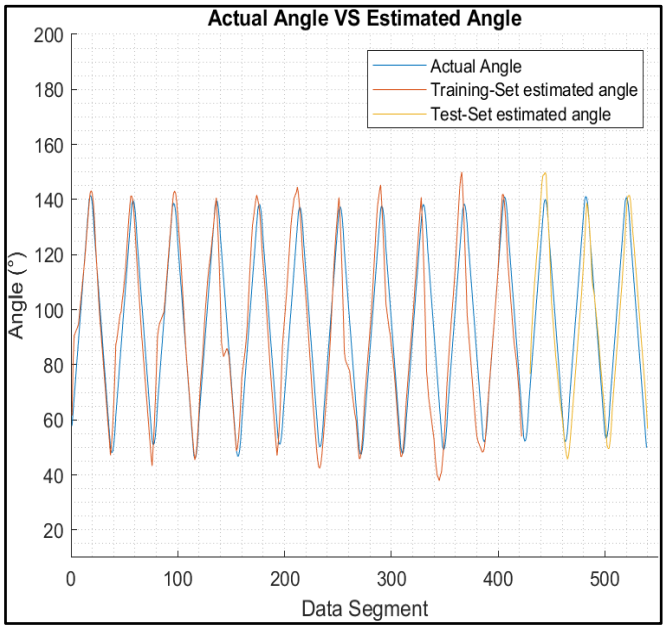


Figure 8: Arm flexion and extension at shoulder level in a horizontal plane target angle versus estimated angle (confusion for training and testing data are 90.4% and 83.5% respectively).

Figure 6, Figure 7, and Figure 8 show the actual joint angle versus estimated angle, training and testing data, for biceps curl with 5 pounds weight, biceps curl empty-handed and arm at shoulder level flexion and extension respectively.

We have chosen Pearson Correlation Coefficient (PCC) as a performance index when comparing the estimated elbow angle against the actual angle because unlike other commonly used performance indices, it is proven to be more accurate in representing the statistical association between two continuous variables as it employs the covariance between the two variables in its calculation [37]. PCC value ranges from +1 to -1, the closest it gets to +1 the more the two vectors are similar to each other. PCC can be found using (9) [16]:

$$PCC(Sig_1, Sig_2) = \frac{cov(Sig_1, Sig_2)}{\sigma_{Sig_1} \cdot \sigma_{Sig_2}} \quad (9)$$

Where

$cov(Sig_1, Sig_2)$ is the covariance of Sig_1 and Sig_2 which is given by (10)

$$cov(A, B) = \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \mu_A)^* (B_i - \mu_B) \quad (10)$$

μ_x is the mean of x, which is given by $\mu_x = \frac{1}{N} \sum_{n=1}^{N} x_n$, and $(*)$ denotes the complex conjugate.

σ_{Sig_1} and σ_{Sig_2} are the standard deviation of Sig_1 and Sig_2 respectively and it's given by (11)

$$S = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} |A_n - \mu|^2} \quad (11)$$

Table 1 list the PCC values for the estimated elbow joint movements for both training and testing sets for the designated movement.

Table 1: Elbow joint movement estimation correlation coefficient.

Movement	Training / Testing	PCC
Biceps curl with an empty hand	Training	0.92
	Testing	0.92
Biceps curl with 5 lbs.	Training	0.94
	Testing	0.95
Triceps extension with an empty hand	Training	0.92
	Testing	0.91
Triceps extension with 5 lbs.	Training	0.95
	Testing	0.95
Arm flexion at shoulder level in a horizontal plane with an empty hand	Training	0.92
	Testing	0.92
Arm flexion at shoulder level in a horizontal plane with 5 lbs.	Training	0.94
	Testing	0.92

Figure 9 shows the histogram of time-window length in seconds for weighted biceps curls (i.e., 5 lbs) which in turn shows that the estimator was able to work with a variety of time-window sizes for the same angular displacement (and thus variable speeds) and still maintain high accuracy, i.e. +80% in classification and 0.91 in correlation.

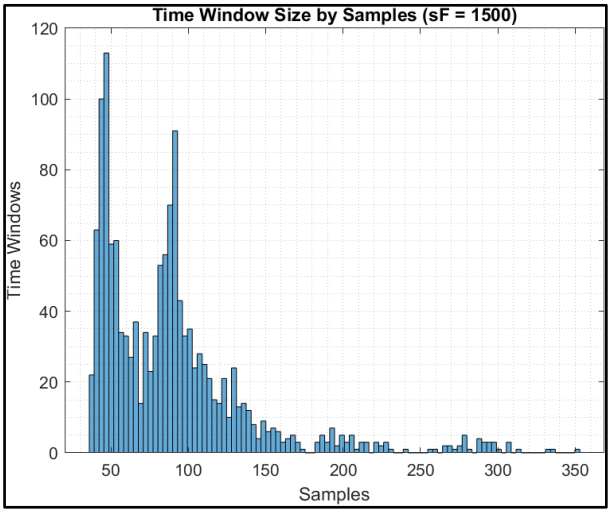


Figure 9: Time-window length in seconds.

3. Discussion

This work describes human's upper limb movement estimation using sEMG data. We have designed an estimator that can estimate elbow joint movement immediately after the neural command sensed through sEMG signals at the muscles. Three muscles were targeted in this work, biceps brachialis, and triceps. The raw sEMG signal underwent preprocessing before it was fed to the estimator. The estimator used a combination of time and frequency domain features extracted for each sEMG channel. The features array is then fed to an ANN-based softmax classifier to be trained to classify five classes in joint angle displacement. The classifier had an accuracy ranging from 80- 90% for different effort levels and limb orientation. The classification was easier for the classifier when the participant carried weight in his hand, which increase the accuracy by 5-7%. The selection of the classes and the accuracy of the classifier translates to PCC between actual and estimated angles ranging from 0.91 to 0.95.

This study aimed to estimate the voluntary movement of upper extremities at the joint level. Since the appendicular skeleton has 126 bones [22], it's not feasible to estimate all the joints movements in one study. To compensate for that we build an estimator that is modular from the software point of view. We believe the developed estimator can work to estimate the movement of any joint in the appendicular skeleton with no modification if we use 4 sEMG channels per joint per degree-of-freedom; assuming that we can properly distinguish agonistic and antagonistic muscles for the chosen joint.

We have chosen the elbow joint for this study because of three main reasons: first, ease of access to its agonistic and antagonistic muscle groups; second, its movement can be done in different planes with respect to the torso; and third, we can vary the joint load easily by asking the participant to hold weights. The last two reasons are essential to check the versatility of the estimator and to make sure it has consistent estimation in different movement conditions. We see this as a necessity because all the related literature reviewed in this study did not account for such variation during joint movement estimation [11–16]. Furthermore, we validate the estimator across the whole dataset by implementing a five-fold cross-validation technique when training and testing the estimator.

Being based on ANN, the softmax classifier took a relatively long time to be trained, and it averaged about 30 minutes for the whole data set. But, on the other hand, the average estimation time per segment was about 1 microsecond, and according to the research in [38], this speed is more than sufficient for real-time applications. We used Windows 10 Pro 64-bit PC with Intel Core i7-4790K CPU and 32 GB RAM and Matlab R2020a to train and test the estimator, which can be a valid option for some but not all Human Machine Interface (HMI) applications targeted to benefit from this study.

The developed estimator can predict the current data segment joint angle with high accuracy regardless of the previous and the next data segment angle estimation accuracy. In other words, misestimation in a single data segment doesn't affect the consecutive data segments. Moreover, due to the selected movement classes, any misestimation will stay within the set 5° resolution which is reflected by the high PCC values shown in Table 1. The choice of features across time-domain and frequency-domain, the choice of classes, and the choice of the classifier all contributed to high estimation accuracy with

0.91 to 0.95 correlation between actual and estimated elbow angle by using only four sEMG sensors on the joint's agonistic and antagonistic muscles.

4. Conclusion

The designed estimator achieved higher accuracy in joint movement estimation regardless of the joint orientation, speed, or effort level than the similar work reviewed in this paper. Which introduces the ability to continuously track the intended movement of the elbow joint with more than 90% correlation. Although the proposed estimator in this work doesn't have 100% joint movement estimation accuracy, section 2.6 and Figures 5 to 8 show that the estimator has over 98% accuracy in predicting the direction of the movement even if it makes a mistake whether the movement is 5° or 10°, as its error margin to predict the wrong direction of movement is less than 2%.

And since the sEMG signals from able-bodied can be used to simulate the human limbs response in amputees [17,39,40], this work can be of great help towards a more natural myoelectric-prosthesis action, and it can also be used to other HMI applications.

References

[1] G. Hefftner, W. Zucchini, G.G. Jaros, "The electromyogram (EMG) as a control signal for functional neuromuscular stimulation. I. Autoregressive modeling as a means of EMG signature discrimination," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 4, pp. 230–237, 1988, doi:10.1109/10.1370.

[2] G. Hefftner, G.G. Jaros, "The electromyogram (EMG) as a control signal for functional neuromuscular stimulation. II. Practical demonstration of the EMG signature discrimination system," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 4, pp. 238–242, 1988, doi:10.1109/10.1371.

[3] J. V. Basmajian, C.J. De Luca, Muscle Alives: Their Functions Revealed by Electromyographic, 5th ed, Williams & Wilkins, Baltimore, 1985.

[4] Z. Ju, G. Ouyang, H. Liu, "EMG-EMG correlation analysis for human hand movements," *Proceedings of the 2013 IEEE Workshop on Robotic Intelligence in Informationally Structured Space, RiiSS 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013*, no. Mi, pp. 38–42, 2013, doi:10.1109/RiiSS.2013.6607927.

[5] P.C. Doerschuk, D.E. Gustafson, A.S. Willsky, "Upper Extremity Limb Function Discrimination Using EMG Signal Analysis," *IEEE Transactions on Biomedical Engineering*, vol. BME-30, no. 1, pp. 18–29, 1983, doi:10.1109/TBME.1983.325162.

[6] M. Asghari Oskoei, H. Hu, "Myoelectric control systems-A survey," *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275–294, 2007, doi:10.1016/j.bspc.2007.07.009.

[7] M. Dyson, K. Nazarpour, "Data Driven Spatial Filtering Can Enhance Abstract Myoelectric Control in Amputees," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3770–3773, 2018.

[8] L.A. Johnson, A.J. Fuglevand, "Mimicking muscle activity with electrical stimulation," *Journal of Neural Engineering*, vol. 8, no. 1, pp. 1–25, 2011, doi:10.1088/1741-2560/8/1/016009.

[9] L.H. Smith, L.J. Hargrove, "Comparison of surface and intramuscular EMG pattern recognition for simultaneous wrist/hand motion classification," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 4223–4226, 2013, doi:10.1109/EMBC.2013.6610477.

[10] E.N. Kamavuako, J.C. Rosenvang, R. Horup, W. Jensen, D. Farina, K.B. Englehart, "Surface versus untargeted intramuscular EMG based classification of simultaneous and dynamically changing movements," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 6, pp. 992–998, 2013, doi:10.1109/TNSRE.2013.2248750.

[11] Y.M. Aung, A. Al-Jumaily, "Estimation of upper limb joint angle using surface EMG signal," *International Journal of Advanced Robotic Systems*, vol. 10, , pp. 1–8, 2013, doi:10.5772/56717.

[12] T. Uchiyama, T. Bessho, K. Akazawa, "Static torque — angle relation of human elbow joint estimated with artificial neural network technique," *Journal of Biomechanics*, vol. 31, , pp. 545–554, 1998.

[13] Triwiyanto, O. Wahyunggoro, H.A. Nugroho, Herianto, "Evaluating the linear regression of Kalman filter model on elbow joint angle estimation using electromyography signal," in *AIP Conference Proceedings*, 020004, 2018, doi:10.1063/1.5054408.

[14] Q.L. Li, Y. Song, Z.G. Hou, "Estimation of Lower Limb Periodic Motions from sEMG Using Least Squares Support Vector Regression," *Neural Processing Letters*, vol. 41, no. 3, pp. 371–388, 2015, doi:10.1007/s11063-014-9391-4.

[15] F. Xiao, Y. Wang, Y. Gao, Y. Zhu, J. Zhao, "Continuous estimation of joint angle from electromyography using multiple time-delayed features and random forests," *Biomedical Signal Processing and Control*, vol. 39, no. December, pp. 303–311, 2018, doi:10.1016/j.bspc.2017.08.015.

[16] F. Xiao, Y. Wang, Y. Gao, Y. Zhu, J. Zhao, "Continuous estimation of elbow joint angle by multiple features of surface electromyographic using grey features weighted support vector machine," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 3, pp. 574–583, 2017, doi:10.1166/jmihi.2017.2054.

[17] J.G. Ngeo, T. Tamei, T. Shibata, "Continuous and simultaneous estimation of finger kinematics using inputs from an EMG-to-muscle activation model," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, pp. 1–14, 2014, doi:10.1186/1743-0003-11-122.

[18] S. Wang, Y. Gao, J. Zhao, T. Yang, Y. Zhu, "Prediction of sEMG-based tremor joint angle using the RBF neural network," *2012 IEEE International Conference on Mechatronics and Automation, ICMA 2012*, pp. 2103–2108, 2012, doi:10.1109/ICMA.2012.6285668.

[19] D. Van Der Riet, R. Stopforth, G. Bright, O. Diegel, "An overview and comparison of upper limb prosthetics," *2013 Africon*, vol. 9, no. 2, pp. 1–8, 2013, doi:10.1109/AFRCON.2013.6757590.

[20] K. Nurhanim, I. Elamvazuthi, P. Vasant, S. Parasuraman, "Determination of Mathematical Model and Torque Estimation of s-EMG Signals based on Genetic Algorithm," vol. 4, no. 5, pp. 135–139, 2013.

[21] D. Shin, J. Kim, Y. Koike, "A Myokinetic Arm Model for Estimating Joint Torque and Stiffness From EMG Signals During Maintained Posture," *Journal of Neurophysiology*, vol. 101, no. 1, pp. 387–401, 2009, doi:10.1152/jn.00584.2007.

[22] R.L. Drake, A.W. Vogl, A.W.M. Mitchell, GRAY'S ANATOMY FOR STUDENTS, THIRD EDIT, CHURCHILL LIVINGSTONE ELSEVIER, Philadelphia, Pa, 2015.

[23] S. Fornalski, R. Gupta, T.Q. Lee, "Anatomy and biomechanics of the elbow joint," *Techniques in Hand & Upper Extremity Surgery*, vol. 7, no. 4, pp. 168–178, 2003, doi:10.1055/s-0033-1361587.

[24] M. Gazzoni, N. Celadon, D. Mastrapasqua, M. Paleari, V. Margaria, P. Ariano, "Quantifying Forearm Muscle Activity during Wrist and Finger Movements by Means of Multi-Channel Electromyography," *PLoS ONE*, vol. 9, no. 10, pp. e109943, 2014, doi:10.1371/journal.pone.0109943.

[25] M.B.I. Reaz, M.S. Hussain, F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification and applications," *Biological Procedures Online*, vol. 8, no. 1, pp. 11–35, 2006, doi:10.1251/bpo115.

[26] Y. Zhang, P. Xu, P. Li, K. Duan, Y. Wen, Q. Yang, T. Zhang, "Noise - assisted multivariate empirical mode decomposition for

multichannel EMG signals," *Biomedical Engineering Online*, vol. 16, no. 1, pp. 1–17, 2017, doi:10.1186/s12938-017-0397-9.
[27] M. Haris, P. Chakraborty, B.V. Rao, "EMG signal based finger movement recognition for prosthetic hand control," *International Conference Communication, Control and Intelligent Systems, CCIS 2015*, pp. 194–198, 2016, doi:10.1109/CCIntelS.2015.7437907.
[28] R.H. Chowdhury, M.B.I. Reaz, M.A. Bin Mohd Ali, A.A.A. Bakar, K. Chellappan, T.G. Chang, "Surface electromyography signal processing and classification techniques," *Sensors (Switzerland)*, vol. 13, no. 9, pp. 12431–12466, 2013, doi:10.3390/s130912431.
[29] S. Muceli, A.T. Boye, A. Avella, D. Farina, "Identifying Representative Synergy Matrices for Describing Muscular Activation Patterns During Multidirectional Reaching in the Horizontal Plane," *Journal of Neurophysiology*, vol. 103, no. 3, pp. 1532–1542, 2010, doi:10.1152.
[30] N. Nazmi, M. Abdul Rahman, S.-I. Yamamoto, S. Ahmad, H. Zamzuri, S. Mazlan, "A Review of Classification Techniques of EMG Signals during Isotonic and Isometric Contractions," *Sensors*, vol. 16, no. 8, pp. 1304, 2016, doi:10.3390/s16081304.
[31] T.A.K. Guanglin Li, Aimee E. Schultz, "Quantifying Pattern Recognition—Based Myoelectric Control of Multifunctional Transradial Prostheses," *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING*, vol. 18, no. 2, pp. 185–192, 2010, doi:10.1109/TNSRE.2009.2039619.
[32] E. Scheme, K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *Journal of Rehabilitation Research & Development*, vol. 48, no. 6, pp. 643–660, 2011, doi:10.1682/JRRD.2010.09.0177.
[33] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[34] J.R. Shewchuk, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, 1994.
[35] N. Miljković, M.S. Isaković, "Effect of the sEMG electrode (re)placement and feature set size on the hand movement recognition," *Biomedical Signal Processing and Control*, vol. 64, , pp. 102292, 2021, doi:10.1016/j.bspc.2020.102292.
[36] C.D. Bryce, A.D. Armstrong, "Anatomy and Biomechanics of the Elbow," *Orthopedic Clinics of North America*, vol. 39, no. 2, pp. 141–154, 2008, doi:10.1016/j.ocl.2007.12.001.
[37] J. Benesty, J. Chen, Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008, doi:10.1109/TASL.2008.919072.
[38] A.J. Young, L.J. Hargrove, T.A. Kuiken, S. Member, "The Effects of Electrode Size and Orientation on the Sensitivity of Myoelectric Pattern Recognition Systems to Electrode Shift," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2537–2544, 2011, doi:10.1109/TBME.2011.2159216.
[39] E. Campbell, A. Phinyomark, A.H. Al-timemy, R.N. Khushaba, G. Petri, "Differences in EMG Feature Space between Able-Bodied and Amputee Subjects for Myoelectric Control," *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 33–36, 2019.
[40] V. Ravindra, C. Castellini, "A comparative analysis of three non-invasive human-machine interfaces for the disabled," *Frontiers in Neurorobotics*, vol. 8, no. October, pp. 1–10, 2014, doi:10.3389/fnbot.2014.00024.

Received: 15 March, 2024, Revised: 10 April, 2024, Accepted: 11 April, 2024, Online: 26 April, 2024

DOI: https://doi.org/10.55708/js0304002

Comprehensive E-learning of Mathematics using the Halomda Platform enhanced with AI tools

Philip Slobodsky[1], Mariana Durcheva*[2,3]

[1]Halomda Educational Software, Rishon LeZion, Israel
[2]Sami Shamoon College of Engineering, Mathematics Department, Ashdod, Israel
[3]TU-Sofia, Informatics Department, Sofia, Bulgaria
*Corresponding author: Mariana Durcheva, Rishon LeZion, Israel, mdurcheva66@gmail.com

ABSTRACT: The method of assessment affects on learning by determining how students manage their time and prioritize subjects. It is widely accepted that students may demonstrate different skills in different assessment formats. The authors demonstrated how e-assessment through the Halomda educational platform can not only improve student learning outcomes, but also enrich their learning experiences. In addition, it is shown how ChatGPT integrated with two new math exploration tools into proprietary Chat-Mat™ module, can help students learn at home and in the classroom, as well as support teachers in their daily work of reviewing student assignments. The outcomes of teaching courses with Halomda not only reveal impressive student performance on final exams but also illustrate a strong correlation between exam scores and weekly assignment grades.

KEYWORDS: E-assessment, E-learning, Halomda educational platform, ChatGPT.

1. Introduction

Like other learning activities, various methods of e-learning are available, tailored to the specific knowledge and skills being achieved. As for the assessment, it is conducted considering the context and purpose, acknowledging that learners may demonstrate varying strengths in different assessment formats and that when students are actively engaged in the activities, it results in deeper thinking and long-term retention of learned concepts [1]. The method of assessment plays a key role in shaping learning outcomes, impacting how students allocate their time and which areas they prioritize to focus on [2-3]. Assessment is typically categorized into the following types: (1) formative assessment, which serves developmental purposes and occurs during the learning process, (2) summative assessment, which aims to evaluate whether students have acquired the expected knowledge or skills, (3) continuous assessment, which means that both formative and summative assignments are used during the course. The feedback provided through formative assessment assists students in tracking their learning progress and serves as a motivational tool.

E-assessment finds broad application, especially in the realms of formative and continuous assessment. Several authors [4] highlighted the fact that e-assessment employs traditional assessment methods but facilitated via online processes and digital tools. However, it could be pointed out that e-learning and e-assessment are not only digital tools, for instance, "many students take e-assessments primarily for the marks, but in doing so, they learn from the questions and especially the feedback" [5]. As digital technology becomes more accessible in education, educators worldwide are increasingly tasked with addressing its potential impact on the academic success of the students. In [6] is argued that when interacting with dialog-based AI models, the goal and context of the desired content should be clearly and concisely defined, and guidance should be provided by adding the necessary instructions and parameters for the appropriate outcome. The authors agree with the authors of [7], that there are two main approaches in technology-assisted assessment, namely - assessment with technology and assessment through technology.

Of course, e-learning and e-assessment has its own challenges. Cheating has become a greater problem with increased use of e-assessment [8-9], since cheating and

plagiarism undermine the validity and reliability of the e-assessment process, eroding students' trust in its integrity [10-12], and, of course, cheating poses a serious threat to important university exams whether they are conducted on campus or online [13]. Students' cheating and plagiarism with regards to e-assessment and using AI technologies are object of research of different authors [14-15], [8], [11]. One approach to mitigate cheating in e-assessment is through the utilization of student authentication and authorship checking systems, like the TeSLA system [8-9], [11], [16]. However, such strategies to combat cheating are particularly well-suited for humanities courses and offer limited effectiveness in engineering disciplines [17]. Hence, biometric and plagiarism-based checking systems represent just a few of the effective measures to mitigate cheating. One potential way to prevent online assessment cheating is to provide each student with unique test instructions, questions, and results. However, creating and administering personalized assessments for each student requires significant preparation time and may not be practical for widespread implementation. To overcome this problem, the use of computer-based assessment tools can help streamline the process and reduce the time required for setup. Various tools are available in widely used learning management systems (LMS) to create tests or multiple-choice assignments, allowing students to submit written assignments to the teacher for manual grading.

Following the increasing usage of AI technologies, and specifically ChatGPT, in education and learning, and exponentially growing number of research in this field, the Halomda developed a unique technology enabling students to create prompts including complex math expressions, to feed them into the built-in ChatGPT, and to use the answers in self-learning. The authors share the hope of the authors of [6] that by overcoming concerns about plagiarism and leveraging the potential of artificial intelligence tools to remove existing barriers to learning, "the educational community will flourish in an increasingly AI-enhanced world".

This paper is organized as follows: Section 2 discusses the related systems in the field of a math e-assessment; Section 3 presents the abilities of the Halomda system; Section 4 is devoted to the e-learning and e-assessment based on the Halomda platform; Section 5 draws conclusions.

2. E-assessment in higher mathematics courses

The use of e-assessment in higher mathematics courses has been shown to be successful in various studies. In [18], the authors stated that the implementation of e-assessment strategies, particularly using multiple-choice questions in Moodle, led to improvements in teaching and learning processes. In [19], the author further supported this, demonstrating the effectiveness of online quizzes as a form of assessment in pure mathematics. However, in [20] is highlighted the need for further research on the practice of assessment in proctored and unproctored math e-learning courses, particularly in relation to academic integrity. Some known systems that support online assignments in mathematics, are: STACK [21], WebAssign [22], WeBWorK [23], Numbas [24]. These systems primarily target basic mathematics, which imposes limitations on the complexity of problems that can be presented.

STACK utilizes a computer algebra system (CAS) called Maxima, which is capable of handling mathematical tasks such as manipulating mathematical expressions. The primary objective of the system is to evaluate students' mathematical responses, generate problems, provide feedback, assign numerical grades, and create internal notes for subsequent analysis. The strength of CAS lies in its capacity to verify algebraic equivalence. However, if it is crucial to differentiate between different forms of answers in a given exercise, the exercise creator must specify in advance which answer formats are acceptable. Once the exercises are created, they can be reused multiple times, thereby reducing the time and workload for teachers. Furthermore, since the system operates independently of teacher supervision, assistants, or classroom availability, students can access it at any time of day or from any location. It's evident that not all question types are suitable for implementation with CAS. Assessing skills such as proof writing and other forms of reasoning cannot yet be automated [25]. Challenges that students may face with STACK often arise from attempts to use unfamiliar syntax when entering answers [26].

WebAssign is a versatile web-based instructional system that offers students additional practice opportunities and provides convenient access to their performance assessments. The software underlying the platform identifies mathematically equivalent expressions in the background. The system offers several benefits, including immediate feedback, the ability for multiple attempts, ease of access, and access to online tutoring [26]. The primary drawbacks include the fact that technical difficulties experienced by some students, reduced interaction with instructors leading to possible neglect of assignments; the need to match the system's format precisely when entering answers, and a less user-friendly experience when asking questions in the tutoring center [26]. From our point of view, the utility of WebAssign is somewhat constrained as it primarily serves homework assignments. Additionally, its integration with specific textbooks means that it must be used in conjunction with the selected textbook, further limiting its applicability.

WeBWorK is an open-source online homework system for math and science courses that enables teachers to

publish problems for students to solve online. Supported courses encompass a range of subjects, including college algebra, discrete mathematics, probability and statistics, single and multivariable calculus, differential equations, linear algebra, and complex analysis. However, it is worth noting that the current engineering content available in the WeBWorK library is limited [27]. Furthermore, the utility of the WeBWorK system is constrained by its primary focus on homework assessment.

Numbas is an online assessment system specifically designed for mathematics subjects and offers several benefits [28]. The system excels at formative assessment and uses client-side grading, although it still requires a server to store grades. Numbas makes it easy to create questions using appropriate mathematical notation, and students can enter symbols that are displayed mathematically. The system supports the creation of complex questions with features such as randomization and automatic scoring and can easily integrate with learning management systems. However, there are limitations in using this system including concerns related to infrastructure, support, the required skills in both teachers and students, as well as time constraints [28].

3. The Halomda Platform

The primary didactic advantage of the Halomda platform over the aforementioned e-systems is its ability to integrate the three main methods of mathematical education - Learn, Train, and Test (Figure 1) ([29] – [31]).

The necessity for a universal, user-friendly, simple yet powerful math platform prompted Halomda Educational Software [29] to develop Math-Xpress, initially introduced in 2001 [30]. Subsequently, additional modules have been integrated into the program, forming a multifunctional Math package with a unified user interface.

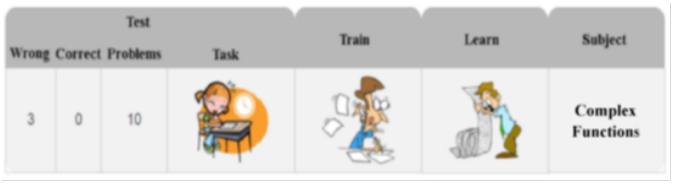


Figure 1 [29]: Modes of learning

These interconnected modules rely on proprietary computer algebra (CA) and interactive geometry algorithms, enabling linkage between modules that share common objects. The system comprises five modules:

- XPress-Editor - a formula editor that allows graphical WYSIWYG (What You See Is What You Get) editing of math expressions thus enables compiling of new items by unexperienced in programming people [31].
- XPress-Graph - a graph plotter.
- XPress-Geometry - a 2-D and 3-D geometry explorer.
- XPress-Evaluator.
- XPress-Tutor. For e-assessment purposes, it can be used independently; together with the XPress-Task Editor these modules form a comprehensive platform for e-teaching, e-learning, and e-assessment of mathematics courses.

Over the past few years, thousands of problems have been developed to cover courses in arithmetic, elementary algebra, and geometry for primary and intermediate schools, as well as algebra, trigonometry, and introduction to calculus for high schools. Additionally, courses in Calculus, Linear algebra, Differential equations, Probability, and Statistics have also been addressed. Throughout the course, students receive a comprehensive set of basic problems that thoroughly cover the curriculum. Typically, this set comprises 50 to 150 problems, arranged into 13 weekly assignments for one semester [31].

3.1. Typical student interaction with the platform

Throughout the course, students are provided with a comprehensive set of basic problems that cover the entire curriculum. Typically, this set comprises 50 to 150 problems, arranged into 13 weekly assignments for one semester [31]. In Learn mode, students are presented with a series of problems, each of which includes randomly selected parameters to ensure that different attempts produce different initial sets of parameters. Students have the opportunity to request help, which is provided at three levels: General help, which describes a solution method common to all problems within a particular topic; a List of steps to solve a problem with a description of each step; and the Result of each solution step. In Train mode, students are presented with multiple-choice options for the possible results of each step, facilitating extensive and interactive training. In Test mode, no help is provided, and the student is required to complete a series of tasks that simulate the conditions of a regular test.

Usually, students start with Test mode, attempting to solve problems without assistance. The system allows them to quit the test without penalty but limits the number of attempts for each task. If a student exits a Test mode due to difficulty, he can switch to Learn or Train modes to understand how to solve similar problems. Tasks in the Test mode differ from tasks in the Learn and Train modes only in the values of randomly generated parameters. The Learn and Train modes offer more than just problem solving; they include additional educational resources such as links to external files, websites, videos, etc. to enrich the learning experience.

3.2. Mitigating cheating through the Halomda platform

Adaptability of students' responses significantly reduces the likelihood of cheating. For example, requiring

students to enter their answers manually rather than choosing from multiple answer options greatly reduces the likelihood of dishonesty. Math-Xpress uses two approaches to evaluate student responses in exam mode: using so called "semi-intelligent" multiple choice questions (SI-MCQs) and open-ended (OA) mode. This approach is demonstrated with a problem shown in Figure 2:

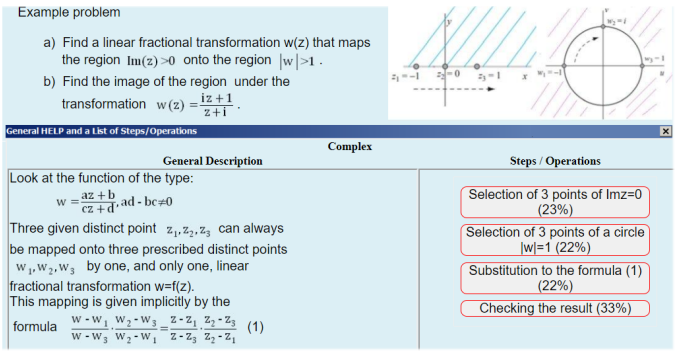


Figure 2 [29]: Task example

During the Learn mode, a student has the possibility to get the results of each step of the problem solution, one of which is the proper answer, and the other three – wrong answers (Figure 3, a). As in usual MCQ-mode, a student can select an answer by clicking on it (in our example: $z_1 = -1, z_2 = i, z_3 = 1$). In case of a wrong answer, a student is provided with short feedback as can be seen in Figure 4.

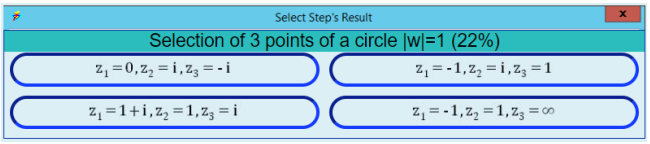


Figure 3 [29]: Step results

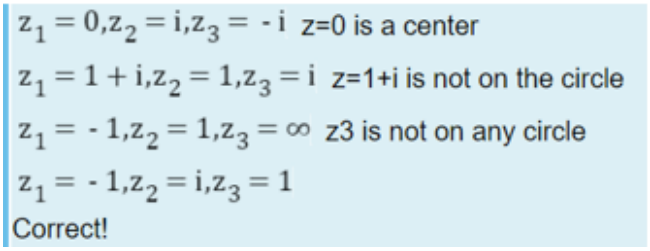


Figure 4 [29]: System comments

Due to the possibility of free-hand typing, the student may decide to enter another expression, for example, just: $(-1;\ i;\ 1)$, and the system will accept them. This ability of the system is provided by the task developer (teacher, instructor), who endeavors to anticipate the most expected options for the correct answer that students can enter.

In open access mode, the result windows are hidden, and students are prompted to manually enter the correct answer. Here, the task designer can choose between two evaluation options: the requirement of strict identity of expressions or their algebraic equivalence. In the first case, the student's answer is compared with all possible answers (both correct and incorrect) given in the task. In the latter case, the comparison is carried out using computer algebra methods to determine algebraic equivalence. For example, the following pairs of expressions are considered equivalent:

$$\frac{2}{\sqrt{2}}, \sqrt{2};\ \ 5^{-2x}, \frac{1}{25^x}\ \ ;\ \ y = 3x + 6, x = \frac{y}{3} - 2$$

An exam may consist of various tasks, each requiring a different assessment method. It is the responsibility of the teacher to determine the most suitable assessment approach for each problem.

The operation of any math assessment system within Moodle (or another Learning Management System (LMS)) can enhance cheating prevention through a combination of random test selection and parameter randomization for tasks. This entails integrating Math core functionalities (such as computer algebra operations, expression evaluation, and test result calculation) into the Moodle code and transferring student data into the Moodle LMS.

Halomda has developed a plugin that enables the system to operate seamlessly within Moodle, facilitating the transfer of results into the Moodle LMS.

3.3. *Integration of ChatGPT into the Chat-Mat™ module of Halomda*

In order to provide students with an easy access to ChatGPT, the last version of Halomda platform includes it as one of a new built-in features: **Graph-Men** (plot and explore graphs) (Figure 5), **Algebraic Calculator** (Equation Solver and Symbolic evaluator in 10 areas of mathematics**)** (Figure 6), **ChatGPT** and recommended **Key Prompts** (Figure 7).

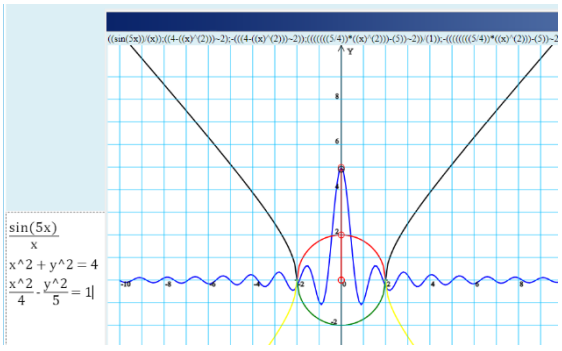


Figure 5: Graph Plotter

In addition to incorporating an AI bot into the system interface, Halomda developed a tool to graphically edit mathematical expressions and present them in the LaTeX format required by ChatGPT. For instance, to compose a prompt for ChatGPT including calculation of the limit: $\lim_{x\to 0} \frac{\sin x}{x}$, it is required to type:

lim_{x\rightarrow0}{(\frac{sinx}{x})}&&

The graphical editor of mathematical expressions offers students the ability to enter limits using the appropriate editor templates (Figure 8), making it easy to include results by copying and pasting into the ChatGPT message box allowing students to easily communicate with the bot. This feature makes communicating with the bot easier, allowing students to interact with it more effectively.

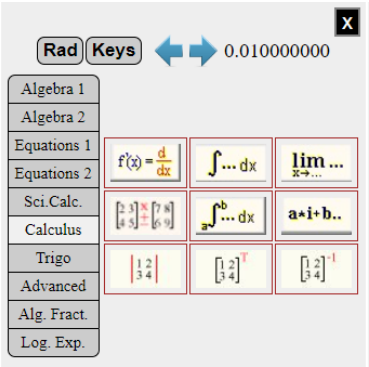


Figure 6: Algebraic Calculator

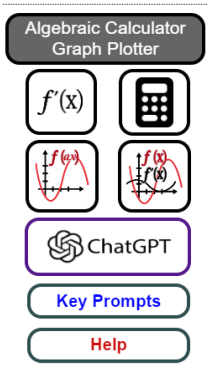


Figure 7: Chat-Mat module

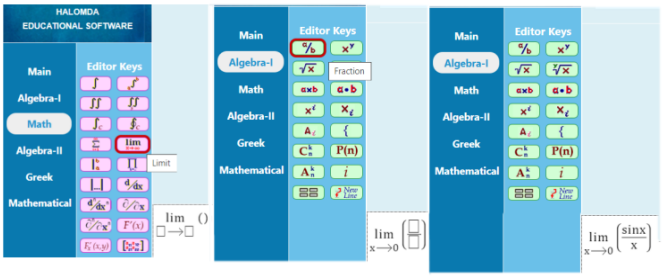


Figure 8: Graphical editing

2 Key Prompts as a new form of self-learning

The main didactical questions arose from most of the publications on using ChatGPT in education of mathematics ([32], [33], [34]) are:

a) How to effectively address the issue of frequently encountering incorrect responses from ChatGPT?

b) In what ways can students utilize ChatGPT as a math tutor beyond just direct problem-solving, considering the occasional inaccuracies?

The suggested answers, based on the teacher's experience and implemented by Halomda in Chat-Math module, can be summarized in the following statement: A student's knowledge of any subject in mathematics is directly proportional to the time spent studying the subject. The more activities offered to students, the greater success is expected!

Following this idea, the authors suggest that the communication with ChatGPT will stimulate the students by presenting other aspects of the problem solving beyond the algebraic technique provided by the Learn and Train modes of the system. To achieve this, the authors recommend that students ask the AI bot diverse questions. These questions (called *prompts*) should enhance their overall knowledge and understanding of the subject and also refine their problem-solving abilities, ultimately leading to greater success in exams. According to [35], "effective prompts are characterized by the following fundamental principles: clarity and precision, contextual information, desired format, and verbosity control, which means that "writing effective prompts is complicated for nontechnical users, requiring creativity, intuition, and iterative refinement".

To help students in composing appropriate prompts, a list of recommended "Key Prompts" is included with each problem in the Learn and Train sections of the system. It consists of 3 types of prompts reflecting the didactical ideas to engage students, including:

1. Solution of the equation; performing algebraic operations; solving integrals; finding derivatives, limits, etc., relevant to the topic.

Although the solution provided by ChatGPT may sometimes be incorrect and/or incomplete, its added value lies in the complementary verbal explanation it offers; the contextual Help recommends the students to verify the answer using the Algebraic Calculator or plot graph with Graph-Man - two other modules included in the Chat-Mat- package.

2. Reminding definitions, theorems, proofs relevant to the subject.

3. Asking broad questions about history, practical applications, "tricky" or amusing questions (like paradoxes, etc.).

Providing students with a list of possible (recommended) prompts not only teaches them to compose their own prompts but also enhances their learning experience by providing additional content required for a deeper comprehension of the subject matter.

The only tool to communicate with ChatGPT is a prompt. In the case where a student is interested in asking for help in solving math problems from class assignments, it is enough to copy the question in the ChatGPT message window, while writing the mathematical expressions in LaTeX format. To achieve this, the Chat-Mat module of the

Halomda platform includes a tool for editing and copying expressions in ChatGPT.

In addition to entering mathematical expressions for evaluation, it is useful to write them so that ChatGPT generates a concise but comprehensive answer. To demonstrate the students how to write math prompts, a short Instructions containing the rules and recommendations are included in the list of Key Prompts, thus completing the didactical need for effective use of ChatCPT.

The role of Key Prompts is two-fold: firstly, to demonstrate students how to formulate math prompts, and secondly, to enhance the overall value of the course, as previously outlined.

The combination of three exploration tools – Key Prompts, Algebraic Calculator and Graph Plotter - allows students to not only observe the solution (provided by ChatGPT), but also check its correctness using the Algebraic Calculator, receive explanations from ChatGPT, and visualize and check it using the Graph Plotter.

The following example demonstrates the use of the Chat-Mat™ module to solve the following:

Problem 1. Find the domain of the function

$$f(x) = 3\log_2(x+4) - 1 + \log_2(x-2)$$

and solve the equation:

$$3\log_2(x+4) = 1 - \log_2(x-2).$$

ChatGPT 3.5 gives a correct answer to the first question, including detailed explanations (Figure 9), but fails to solve the equation, giving the (correct!) comments for the incorrect calculations (Figure 10 and Figure 11), and final wrong answer (Figure 11).

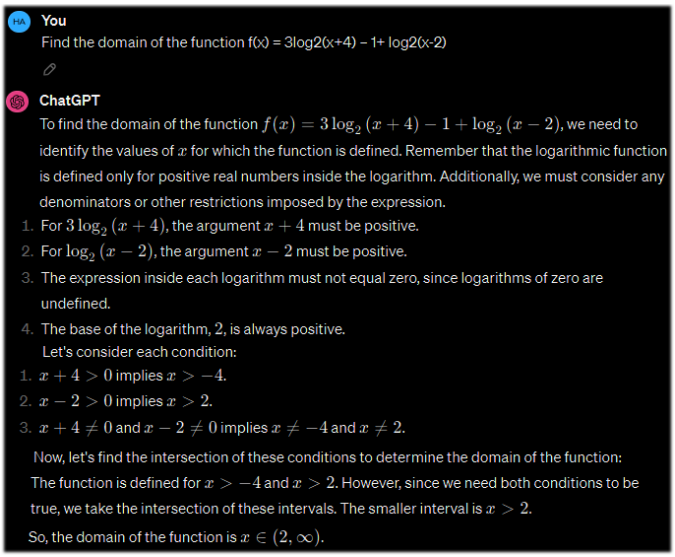


Figure 9: ChatGPT correct answer

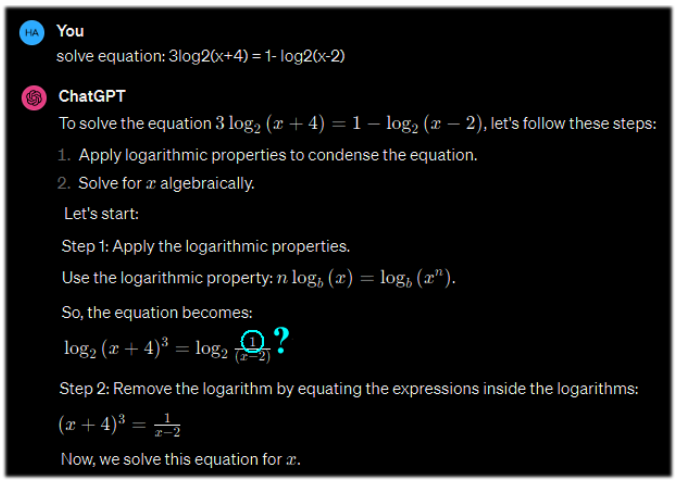


Figure 10: ChatGPT miscalculation

When students are aware of the bot's possible mistakes, they can follow the Instructions and try to check the result in another way, for example, visually - by plotting a corresponding graph using the graph module of Chat-Mat (Figure 12, a), or the Algebraic Calculator (Figure 12, b).

The intersection of the two curves shown in Figure 12,a), representing the graphs of the functions from two sides of the equation, shows that the solution given by the bot is wrong. This may encourage students to analyze the reason why this happened - this is in fact, the (hidden) didactical purpose of the Chat-Mat activity!

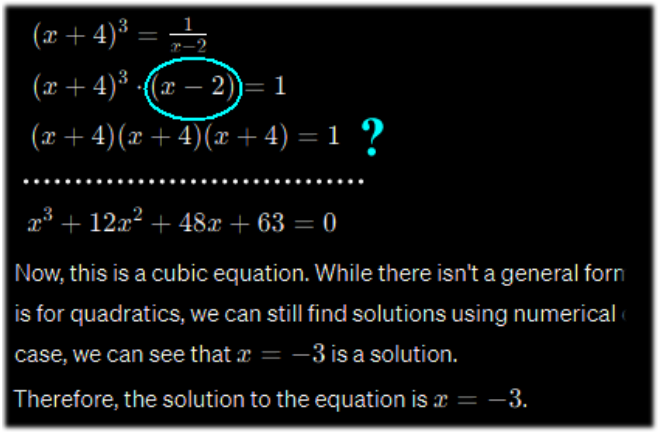


Figure 11: ChatGPT wrong answer

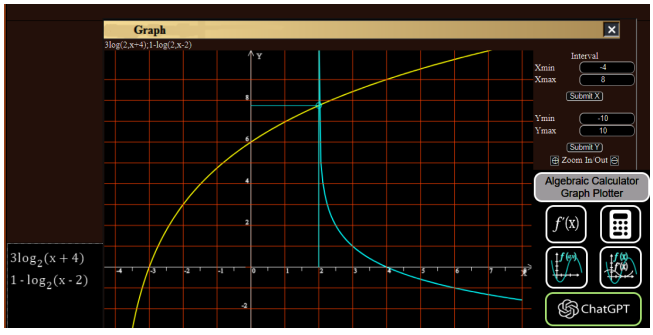


Figure 12 (a): Correct solution by Graph Plotter

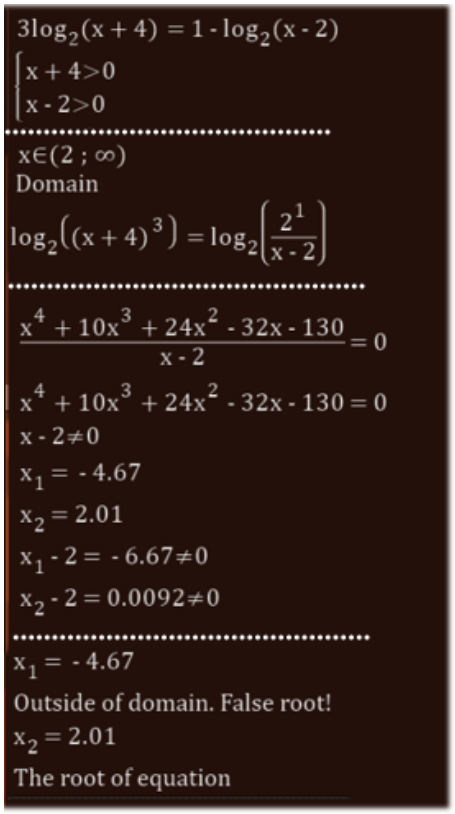


Figure 12 (b): Correct solution by Algebraic Calculator

Students can check each solution step of the bot and discover the mistake in the calculations (Figures 10 and 11), or/and use the Algebraic Calculator (Figure 12, b), compare the solutions and find the mistakes of ChatGPT. The AI abilities of ChatGPT enable them to continue the conversation and try to fix the problem (Figure 13).

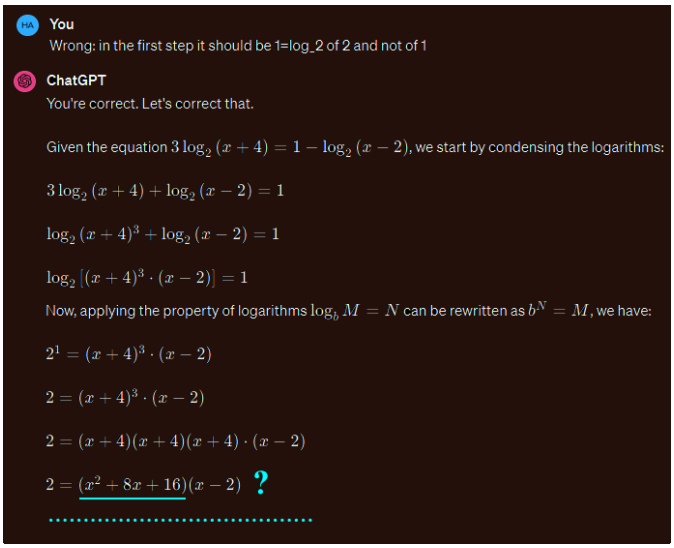


Figure 13: Conversation with ChatGPT

Such interactions between a student and three Chat-Mat tools – ChatGPT, Graph Plotter and Algebraic calculator provide the opportunity for a deeper understanding of the subject, despite the possibility of ChatGPT solving the problem incorrectly.

In order to help students in composing their own prompts, lists of Key Prompts have been created and included into several assignments in Mathematics and Physics, like: Complex Functions, Calculus-I, Transcendental Functions, Basics of Quantum mechanics and more.

4. E-assessment via the Halomda platform

The existence of software for solving mathematical problems has always been problematic due to the possible misuse of it by students during online examination. The Halomda platform does not allow students to get help neither from the built-in Algebraic Calculator or Help option of the system, nor to get access to ChatGPT: all such options are blocked during Test mode (i.e. during the exams).

Using ChatGPT or math-solver software during online exam or even asynchronous test is practically impossible, as long as writing math prompts in LaTeX format necessities using the Halomda's built-in tools (which are typically disabled during Test mode). Additionally, within the Test mode of the Halomda system, solution of a problem often requires answering intermediate steps. Furthermore, the randomization of task parameters helps mitigate cheating, which is crucial for ensuring trusted assessments.

So, the e-assessment offered by the Halomda platform won't interfere with its solving AI abilities.

The benefits of availability of three learning modes: Learn (providing the basic concepts of the subject), Train (allowing practice) and Test (providing opportunities for assessment, including self-assessment) can be summarized as follows:

(1) For students, the platform offers a comprehensive learning tool that includes an interactive e-textbook and extensive training on common problems encountered in the final exam. Combined with the ability to provide instant feedback, the Halomda platform significantly improves the learning process.
(2) For educators, the platform provides an interactive lesson presentation guide along with automatically generated trustworthy grades for midterm assignments.

Over the past decade, five basic Higher Mathematics courses at Ariel University in Israel have employed the Halomda educational platform for teaching and assessment, serving more than 3,000 students annually. The platform's e-assessment module proved indispensable for conducting online assessments during the pandemic [36]. Additionally, Calculus-1 and Elementary Math courses were conducted for 60 students at Talpiot and Orot teachers colleges, respectively. Each course was structured into 13 lessons, either face-to-face or online, followed by problem-solving seminars.

Consequently, students were assigned 13 weekly tasks (mandatory), based on Test mode of the Halomda platform, typically comprising 10 tasks with randomized parameters. Each task included detailed solutions and explanations, functioning as an interactive e-textbook for students and an instructional guide for teachers during online sessions. The Train mode facilitated students' necessary practice, while the Test mode encouraged them to tackle all typical problems, comprehensively covering the course curriculum.

Upon completing the course, students were invited to participate in an online questionnaire. The authors asked about their exam grades as well as the average grade for all Xpress-Tutor assignments. The final exam results revealed both high scores (Figure 14(a)) and a positive correlation between exam grades and weekly assignment grades (Figures 14 a), b)).

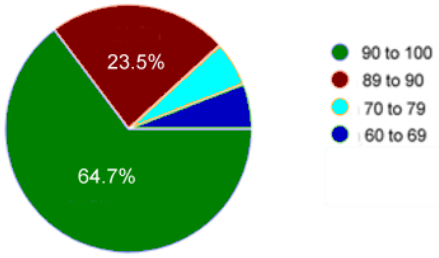


Figure 14 (a): Grades of the final exam [31]

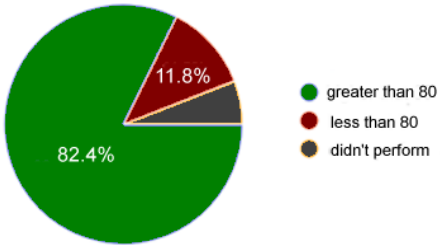


Figure 14 (b): The average grades of assignments [31]

The authors also queried participants about their preferred mode of preparation for the exam. The findings depicted in Figure 15 revealed that Test and Train modes were the most favored options (70.6 and 66.7% respectively).

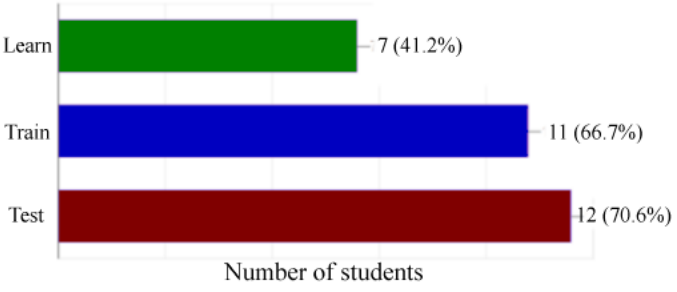


Figure 15: The preferred mode of learning [31]

Figures 16 and 17 display the responses of students to the following queries: To what extent did the system help you in preparing to class exam? (Figure 16) and To what extent was the operation of the system simple and clear? (Figure 17) (on both diagrams the horizontal axes show the rates of correspondent values).

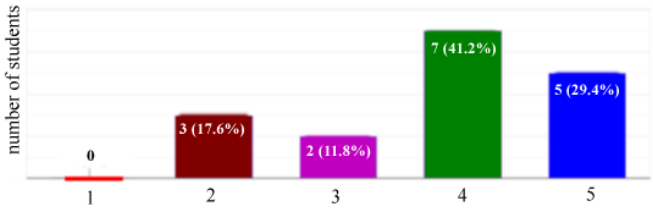


Figure 16: Rate of helpfulness [31]

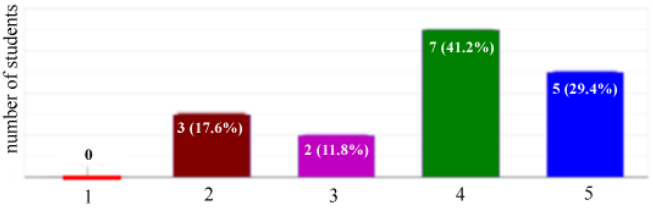


Figure 17: Rating of clarity of the system [31]

The results show that students felt that the system significantly helped them prepare for the exam. They found that the use of the system is simple and clear.

Using the Halomda platform has demonstrated increased student interest and engagement in the subject, as well as significant improvements in learning outcomes.

5. Conclusion

This paper analyzes the most recognized e-learning and e-assessment systems in the field of higher mathematics and highlights the limitations of each of them. Addressing the needs of both students and teachers, Halomda Educational Software has developed Math-Xpress, a versatile, user-friendly and reliable math platform. Math-Xpress offers various modes of interaction with students, including Learn, Train, and Test modes. The integration of the system into LMS-Moodle is a beneficial feature that helps teachers in their routine test checking.

To facilitate the communication of math students with ChatGPT, the latest version of the Halomda platform integrates it as a new built-in feature along with Graph-Men for plotting and exploring graphs, Algebraic Calculator for equation solving and symbolic evaluation, and recommended Key Prompts for various topics across several subjects. The authors believe that dialogue with ChatGPT through properly formulated Prompts can become a new and effective form of self-learning.

Finally, the survey results show a good correlation between final exam scores (which are proctored and administered on campus) and weekly assignment scores. This shows the advantages of using the Halomda platform for e-learning and e-assessment purposes, which can be considered in different directions: the ability of students to

learn from feedback, which significantly enhances students' learning activities (more than 70% of students rated the usefulness of the system as 4 or 5 points), andthe ability to get help from ChatGPT and compare its response with other functions of the Halomda platform.

The further development of the system and the ideas behind it is planned to be focused on the fast growing if AI tools, and particularly their abilities in solving math problems in different areas of mathematics. A study on using of ChatGPT by students of three teachers' colleges has begun and will be presented soon.

References

[1] P.E. Rawlusyk, "Assessment in Higher Education and Student Learning", *Journal of Instructional Pedagogies,* vol. 21, pp. 1-34, 2018. https://bit.ly/2vPtbzT.

[2] J. Biggs, C. Tang, Teaching for Quality Learning at University: What the student does, (3rd ed*).* Maidenhead: McGraw-Hill, 2007.

[3] G. Brown, J. Bull, M. Pendlebury, Assessing student learning in higher education, London: Routledge, 1997.

[4] R. De Villiers, J. Scott-Kennel, R. Larke, "Principles of effective e-assessment: A proposed framework", *Journal of International Business Education*, vol. 11, pp. 65–92, 2016.

[5] M. Greenhow, "Embedding e-assessment in an effective way", *MSOR Connections,* vol. 17, no. 3, – journals.gre.ac.uk., 2019.

[6] B. Eager, R. Brunton, "Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice", *Journal of University Teaching & Learning Practice*, vol. 20, no. 5, 2023, https://doi.org/10.53761/ 1.20.5.02.

[7] P. Drijvers, "Digital assessment of mathematics: Opportunities, issues and criteria", *Mesure et évaluation en éducation*, vol. 41, no. 1, 41-66, 2018, https://doi.org/10.7202/1055896ar.

[8] H. Mellar et al.," Addressing cheating in e-assessment using student authentication and authorship checking systems: teachers' perspectives," *Int. Journal for Educational Integrity,* vol. 14, no. 2, 2018. DOI 10.1007/s40979-018-0025-x.

[9] R. Peytcheva-Forsyth, L. Aleksieva, B. Yovkova, "The Impact of Technology on Cheating and Plagiarism in the Assessment – the teachers' and students' perspectives", *AIP Conference Proceedings* 2048, 020037, 2018, https://doi.org/10.1063/1.5082055.

[10] J. Dermo, "E-assessment and the student learning experience: A survey of student perceptions of e-assessment", *British Journal of Educational Technology*, vol. 40, no. 2, pp. 203–214, 2009, https://doi.org/10.1111/j.1467-8535.2008.00915.x.

[11] S. Kocdar, A. Karadeniz, R. Peytcheva-Forsyth, and V. Stoeva, "Cheating and Plagiarism in E-Assessment: Students' Perspectives", *Open Praxis*, vol. 10 issue 3, July–September 2018, pp. 221–235 (ISSN 2304-070X).

[12] M. E. Rodríguez, A.-E. Guerrero-Roldán, D. Baneres, I. Noguera, "Students' Perceptions of and Behaviors Toward Cheating in Online Education", *IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje*, vol. 16, no. 2, May 2021. Doi: 10.1109/RITA.2021.3089925.

[13] A. Vegendla, G. Sindre, "Mitigation of Cheating in Online Exams: Strengths and Limitations of Biometric Authentication", *Biometric Authentication in Online Learning Environments*, 2019, DOI: 10.4018/978-1-5225-7724-9.ch003.

[14] J. M. Bartley, Assessment is as Assessment does: A conceptual framework for understanding online assessment and measurement, Online assessment and measurement: Foundations and challenges, pp. 1–45, 2005, IGI Global.

[15] S. Stacey, "Cheating on your college essay with ChatGPT", *Business Insider*, https://www.businessinsider. com/professors-say-chatgpt-wont-kill-college-essays-make-education-fairer 2022-12.

[16] M. Ivanova, S. Bhattacharjee, S. Marcel, A. Rozeva, M. Durcheva, "Enhancing Trust in eAssessment - the TeSLA System Solution", *21st Conf. on e-assessment (TEA2018),*10-11 Dec 2018, Amsterdam, The Netherlands, Conf. Proc., https://www.teaconference.org/. arXiv:1905.04985.

[17] M. Durcheva et al., "Innovations in teaching and assessment of engineering courses, supported by authentication and authorship analysis system", *The 45th conference AMEE'19, AIP Conf. Proc.* 2172, 040004 (2019); https://doi.org/10.1063/1.5133514.

[18] J. M. Azevedo, E.P. Oliveira, P.D. Beites, "How Do Mathematics Teachers in Higher Education Look at E-assessment with Multiple-Choice Questions", *International Conference on Computer Supported Education*, 2017, DOI:10.21913/IJEI.V3I2.165.

[19] S. Zegowitz, "Evaluating the use of e-assessment in a first-year pure mathematics module", *Preprint*, 2022, https://arxiv.org/pdf/1908.01028.pdf.

[20] S. Trenholm, "An investigation of assessment in fully asynchronous online math courses", *The International Journal for Educational Integrity, 3*', 2007, DOI:10.21913/IJEI.V3I2.165.

[21] STACK, https://stack-assessment.org.

[22] WebAssign, https://webassign.com.

[23] WeBWorK, https://webwork.maa.org.

[24] Numbas. https://www.numbas.org.uk.

[25] C. J. Sangwin, Computer Aided Assessment of Mathematics. Oxford: Oxford University Press, 2013.

[26] D. Serhan and F. Almeqdadi, "Students' perceptions of using MyMathLab and WebAssign in mathematics classroom", *International Journal of Technology in Education and Science* (IJTES), vol. 4, no.1, pp. 12-17, 2020.

[27] A. G. d'Entremont, P. J. Walls, P. A. Cripton, "Student Feedback and Problem Development for WeBWorK in a Second-Year Mechanical Engineering Program". *Proc. 2017 Canadian Engineering Education Association (CEEA17) Conf.*

[28] V. Challis, R. Cook, and P. Anand, "Priming Numbas for formative assessment in a first-year mathematics unit", In Gregory, S., Warburton, S. & Schier, M. (Eds.), *Back to the Future – ASCILITE '21. Proc. ASCILITE 2021 in Armidale* (pp. 313–318), (2021), https://doi.org/10.14742/ascilite2021.0145.

[29] Halomda. https://halomda.org.

[30] S. Kornstein, "Xpress Formula Editor and Symbolic Calculator", *Mathematics Teacher*, vol. 94, no. 5, May 2001.

[31] P. Slobodsky, M. Durcheva, "E-assessment of Mathematics Courses using the Halomda Platform", *AIP Conf. Proc.* 2939, 050011, 2023, https://doi.org/10.1063/5.0178514.

[32] Y. Wardat, M. Tashtoush, R. AlAli, A. Jarrah, "ChatGPT: A revolutionary tool for teaching and learning mathematics", *Eurasia Journal of Mathematics, Science and Technology Education,* vol. 19, no. 7, em2286, 2023, https://doi.org/10.29333/ejmste/13272.

[33] S. Zhao, Y. Shen, Z. Qi, "Research on ChatGPT-Driven Advanced Mathematics Course", *Academic Journal of Mathematical Sciences,* vol. 4, no. 5, 2023, doi: 10.25236/AJMS.2023.040506.

[34] Rane, Nitin, "Enhancing Mathematical Capabilities through ChatGPT and Similar Generative Artificial Intelligence: Roles and Challenges in Solving Mathematical Problems", 2023, http://dx.doi.org/10.2139/ssrn.4603237.

[35] J. Velásquez-Henao, C.J. Franco-Cardona, L. Cadavid-Higuita, "Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering", *DYNA*, 90 (230), *Especial Conmemoración 90 años*, pp. 9-17, Nov 2023, ISSN 0012-7353, DOI: https://doi.org/10.15446/dyna.v90n230.111700.

[36] P. Slobodsky, A. Shtarkman, A. Kouropatov, "The implementation of e-training and e-assessment system Math-Xpres in development and practical use of system Math-Xpres in development and practical use of math courses at teachers' colleges in Israel during the Corona times", (Unpublished communication), 2022.

Received: 16 February, 2024, Revised: 14 April, 2024, Accepted: 15 April, 2024, Online: 18 May, 2024

DOI: https://doi.org/10.55708/js0305001

Mathematical Model of Optimum Management of the Customs Control Process and Expert System for Ensuring Data Reliability

Ilkhom Mukhtorov[1], Takhir Abduraxmonov[2], Abdusobir Saidov[*,3]

[1]Customs Committee of the Republic of Uzbekistan, First Deputy Chairman, Tashkent, Uzbekistan

[2]Customs Committee of the Republic of Uzbekistan, of Information and Communication Technologies and Cybersecurity, Tashkent, Uzbekistan

[3]Customs Institute, Department of Information Technology and Mathematics, Tashkent, Uzbekistan

*Corresponding author: Abdusobir Saidov, 100017, Chilanzar block, 20-17-6, Tashkent, Uzbekistan, Email: abdusobir59@gmail.com

ABSTRACT: The article considers the issue of modeling the multi-step process of customs clearance of goods in foreign trade. A mathematical model of control of the process under consideration has been developed. A brief review of existing methods for solving the linear programming problem with variable coefficients of the target function is given. The essence of customs risks has been studied and a method for identifying customs risks of reliability using threshold matrixes has been proposed. An algorithm for controlling the reliability of the customs value of goods is developed and the results of the implementation of this algorithm are given

KEYWORDS: customs clearance, mathematical modeling, linear optimization, objective function with a variable coefficient, customs risks, threshold matrix, reliability criteria

1. Introduction

International trade has long been considered the fundamental form of international economic relations. The pace of globalization in the first quarter of the XXI century confirms the role of international trade as the main driver of socio-economic development of countries. Analysis of the dynamics of international trade in the period from 2000 to 2015 shows that the export of developed countries increased from $4.243212 trillion to $8.613816 trillion (+103.00%), developing countries - from $2.059532 trillion to $7.344534 trillion (+256, 61%), countries with economies in transition - from 149.573 billion dollars to 525.571 billion dollars (+251.38%). On average, the volume of exports of world trade goods increased by 254.2% [1].

At the same time, the customs services of the countries participating in international trade play an important role in the international supply chain. There is a theory in the scientific literature, according to which any customs system successively passes several separate phases of its activity, characterized by the specificity of its relations both with foreign trade participants and with the state. Today, the customs systems of developed countries are in the "*customs for foreign trade participants*" phase, while for most developing countries the "*customs for the government*" phase is characteristic, and in a number of disadvantaged and underdeveloped countries, the "*customs for themselves*" phase is observed. [2].

Conversely, the faster a country's customs service approaches the "customs for foreign trade participants" phase of development, the faster the country approaches the level of developed countries. Today, the development of the activities of the customs service of each country is possible only through the use of modern information and communication technologies. Therefore, such requirements are imposed to the methods of customs service management as orientation of models on artificial intelligence, possibility of synthesis of adaptive control system and application to complex analysis of multilevel system.

From this point of view, the task of optimal management of the process of organizing customs control and customs clearance of non-trading goods is relevant.

2. The problem of optimal management of the customs clearance process

The purpose of any type of control is to change the state of the control object in accordance with a predetermined task. Control methods should answer the question, "how can we construct an algorithm that can control a given object in a way that achieves a predetermined goal?". To do this, the developer needs to know how the control object will respond to different influences, that is, the control object model is needed.

There are many definitions to the concept of "model". One of them is close to our question: "A model is an object that allows you to study the behavior of another object, which is called the original". The model and the original should be similar so that the conclusions drawn from studying the model can be applied to the original[3].

As noted above, in order to study the control object, it is necessary to know how it reacts to various influences on it. If we denote these influences as "input" signals for the control object, then the changes occurring under the influence of these "input" signals can be regarded as "output" signals. That is, the object interacts with the external environment using "incoming" and "outgoing" signals.

If the documents submitted to the customs authorities in this process are taken as "inputs" for modeling the multi-stage customs clearance process and denote them by X(t), then customs clearance will be carried out in accordance with these documents. In this case, you can take as an "output" signal the information received as a result of customs clearance, and designate it Y (t) (Fig. 1.)

here:
c(t), r(t), v(t)- set of impacts of customs clearance results to management;
c(t) - objects of organization of customs control (goods, vehicles, persons), documents, additional information flows, etc.;
r(t) – risks that may affect the process (information, human, financial and other resources);
v(t) - obstacles to achieving the goal (refusal, etc.).
This set of influences moves the customs clearance system Z(t) towards a given goal and generates a vector of output results Y(t). From a mathematical point of view, the function Y(t) is the reaction of the control object - the customs clearance process - to external influences.

Optimal management of the multi-stage customs clearance process can be carried out with respect to a number of objectives. In particular, optimization is envisaged in relation to one of the following goals:

- ensuring the completeness of revenues to the state budget, i.e. maximization of revenues to the state budget;
- minimize the amount of arrears that may not go to the state budget, that is, possible damage to the budget;
- Reduction of expenses of the entrepreneur - participant of foreign trade, i.e. minimization of damage caused to him in the process of customs clearance;
- minimize time for customs clearance processes;
- other purposes.

The purpose of this study is the optimal management of the customs clearance process, as well as minimizing the time spent on this process. Because it will ultimately lead to minimization of time for the customs clearance process, as well as to maximize revenues to the state budget and minimize the costs of the entrepreneur - participant of foreign trade.

It is known that in the general case the question of linear optimization can be expressed as follows [4]:

$$\max\left(\min\right) f(x) = \sum_{k=1}^{n} c_k x_k \tag{1.1}$$

$$\left.\begin{array}{lll} \sum_{k=1}^{n} a_{jk} x_k = b_j, & if & j = \overline{1, m_1} \\ \sum_{k=1}^{n} a_{jk} x_k \geq b_j, & if & j = \overline{m_1 + 1, m_2} \\ \sum_{k=1}^{n} a_{jk} x_k \leq b_j, & if & j = \overline{m_2 + 1, m} \\ x_k \geq 0,\ k = \overline{1, n} & & \end{array}\right\} \tag{1.2}$$

In this research paper, in the formation of a mathematical model of optimal control of the process of multi-stage customs clearance in relation to the time consumed, it is proposed to use the stages of customs clearance shown in Table 1.

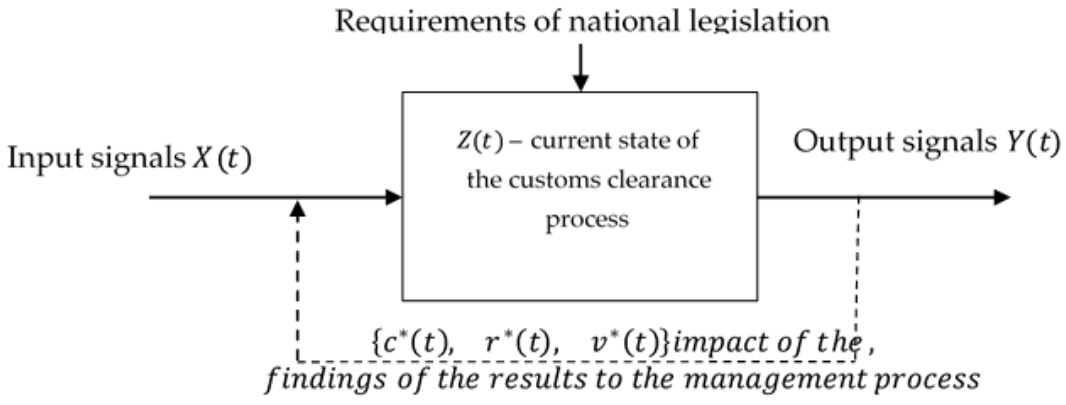


Figure 1: The main factors of the customs clearance process

Table 1. "Customs Clearance Process: Task Execution Matrix"

Estimated execution time	Function names	Implementer
t_1	Preparation of necessary primary documents for customs clearance	Foreign trade participant or customs broker
t_2	Determination of the code of goods according to the harmonized system Commodity nomenclature	
t_3	Calculation of the customs value of goods	
t_4	Calculation of customs payments	
t_5	Preparation of cargo customs declaration	
t_6	Ensuring customs payments	
t_7	Submit a preliminary declaration to the customs authorities prior to the arrival of the shipment	
t_8	Cargo delivery under customs control	Carrier
t_9	Storage related to customs inspection procedures	Customs warehouse jointly with a foreign trade participant and the customs service
t_{10}	Delay of goods due to inability to release due to technical or other reasons	
t_{11}	Storage initiated by the owner of the goods	
t_{12}	Request additional documents when necessary	Customs Service
t_{13}	Direct the cargo to the appropriate (red, yellow or green) customs control lane based on the analysis of submitted documents and risk profiles	
t_{14}	Implementation of procedures either according to the principles of red road customs	
t_{15}	Implementation of procedures either on the basis of the yellow road customs	
t_{16}	Implementation of procedures either according to the principles of green road customs	
t_{17}	Control over the completeness of receipt of customs payments	
t_{18}	Release of cargo into free circulation or for export in accordance with the established procedure	

The presented list of functions is grouped into four main blocks. These functions are performed by customs authorities, foreign trade participants and enterprises providing services in the customs field. The organization of customs clearance is characterized by a large number of operations performed, the complexity of which is determined by a combination of factors: type of goods, country of origin, declared customs value of the goods,

From the above table it follows that the process of multi-stage customs clearance covers 18 stages, of which stages 1-7 are performed by a foreign trade participant or a customs broker, stage 8 - by a cargo carrier, stages 9-11 - stages are performed by owners of customs warehouses, and for execution 12 - 18 - the stages are responsible for the employees of the customs service. That is, in formula (1.1), $n = 18$.

$$f(t) = \sum_{k=1}^{18} r_k t_k \rightarrow \min \qquad (1.3)$$

where: $r_k = r_k(X)$ – the level of risk of execution of the k-stage;

$X=X(x_1,x_2,\ldots,x_{58})$ –vector, x_i elements of which are determined depending on the value of the corresponding columns of the cargo customs declaration. In practice, the level of risk of customs clearance is determined depending on the documents submitted for customs clearance, including the cargo customs declaration.

At the same time, the estimated time of duration of the customs clearance process for responsible executors is determined by normative and directive documents, in particular, documents approved by the Cabinet of

Ministers of the Republic of Uzbekistan. If we denote them as b_1, b_2, b_3 and b_4, respectively, conditions (1.2) will come to the following form:

$$\left.\begin{aligned}
&0 < \sum_{k=1}^{n} a_{jk} t_k \le b_1;\ a_{jk} = 1 \quad if \quad j = \overline{1, m_1} \\
&0 < \sum_{k=1}^{n} a_{jk} t_k \le b_2;\ a_{jk} = 1 \quad if \quad j = \overline{m_1 + 1, m_2} \\
&0 < \sum_{k=1}^{n} a_{jk} t_k \le b_3;\ a_{jk} = 1 \quad if \quad j = \overline{m_2 + 1, m3} \\
&0 < \sum_{k=1}^{n} a_{jk} t_k \le b_4;\ a_{jk} = 1 \quad if \quad j = \overline{m_3 + 1, m_4} \\
&t_k \ge 0,\ k = \overline{1, n}
\end{aligned}\right\} \quad (1.4)$$

here: n=18, m_1=7, m_2=8, m_3=11, m_3=18.

a_{jk}=0 at the values of index j, which are not included in conditions (1.4)

The above formulas (1.3) and (1.4) give a mathematical model of the problem of optimal control of the process of multi-stage customs clearance.

3. Analysis of existing methods for solving the linear programming problem with variable coefficients

The obtained results (1.3) and (1.4) show that the mathematical model of the problem of optimal control of the process of multistage customs clearance has the form of a linear programming problem with variable coefficients of the objective function. Currently, there are a number of effective methods available for solving the linear programming problem.

In particular, for constant values of the coefficients (r_k) of the objective function (1.3) under the limiting conditions (1.4), a number of methods are used in practice to determine its minimum value. These include methods such as the simplex method, the deployment of a function on algebraic polynomials, Fourier series, the use of spline functions, and others. A sufficient number of computer programs for numerical solution of this problem have been implemented.

However, the features of the problem of optimal control of the multi-stage customs clearance process, which is described in (1.3) - (1.4) are variable coefficients of the target function. The functions $r_k = r_k(X)$ which represent the degree of risk of the k - process, is a function of the variables of the cargo customs declaration. This requires a specific approach to solve this problem.

In particular, the 3rd and 4th stages "*calculation of the customs value of goods*" and "calculation of the amounts of customs payments" of Table 1. are important stages in the customs clearance process in ensuring the fulfillment of fiscal tasks assigned to the customs authorities.

This fact indicates the relevance of improving mathematical modeling of the customs clearance process and its comprehensive study.

At the same time, the study of scientific papers on the study of similar problems showed that a sufficient number of studies have been carried out and certain methods have been developed for solving the problem of linear programming with variable coefficients.

For example, the work of [6] is one of the relatively early studies in this area. The problem of parametric programming of the following form is considered:

$$\left.\begin{aligned}
&f(X) = \sum_{j=1}^{n} c_j x_j, \quad X = \{x_j\}, \quad c_j \in \Im \\
&\sum_{k=1}^{n} a_{ij} x_j = b_i; \qquad i = \overline{1, m} \\
&x_j \ge 0,\ j = \overline{1, n}
\end{aligned}\right\} \quad (2.1)$$

here: c_j - elements of some ordered functional prospace $\Im$, a_{ij}, b_i - known, a x_j - unknown real numbers, X - plan of the problem (non-negative solution of the problem (2.1)) Plan X^* is optimal if for any plan X $f(X)<f(X^*)$. Note that the values of the target function (2.1) belong to the space $\Im$, in which the usual properties of numerical inequalities are known to hold.

The existence of a solution to problem (2.1) is proved in this paper by introducing the concept of resolving combinations of problem (2.1), considered as elements of the space $\Im$, are comparable with the zero of this space, i.e., if $x \in R$, then one and only one of three relations holds: $x>0$, $x<0$, and $x=0$. Under these conditions, the following theorem is proved:

Theorem. The problem (2.1) with a non-empty set of plans and a target function bounded from above, satisfying the condition: all solving combinations of coefficients of the target function are comparable to zero, has a solution.

However, the paper does not provide a methodology for determining the existing solution.

In [7], a parametric programming problem of the following kind is considered:

$$\left.\begin{aligned}
&\min\left(\max\right) z(x) = \sum_{j=1}^{n} c_j(x) x_j \\
&\sum_{j=1}^{n} a_{ij}(x) x_j \ge b_j(x), \qquad i = \overline{1, m}
\end{aligned}\right\} \quad (2.2)$$

here $a_{ij}(x)$, $b_i(x)$ and $c_j(x)$ – some piecewise constant argument functions $x = (x_1, x_2, \ldots, x_n)$.

The functions $a_{ij}(x)$, $b_i(x)$ and $c_j(x)$ are defined on the same set $G = \{x \in G \subset R_n\}$. There exists a finite partition $G = UG_k$, $(k=1, l)$ such that the functions are constant in each subset G_k, and G_k and G_{k+1} can intersect only along their boundaries.

By requiring that in problem (1)-(2) the target function $z(x)$ and the constraint functions:

$$f_i(x) = \sum_{j=1}^{n} a_{ij}(x)x_j - b_j(x), \qquad i = \overline{1,m}$$

be continuous and convex, by a simple enumeration of a finite number of regions G_k, in each the usual linear programming problem is solved.

The work [8] is devoted to the study of the problem of parametric programming of the following form:

here: $X=(x_1,\dots x_n)^T$ – n-dimensional vector of unknown variables, which satisfies the constraints (2.3), forming the set of admissible solutions of the proble;

$$\left.\begin{aligned} &f(X) = c_1(t)x_1 + c_2(t)x_2 + \dots + c_n(t)x_n \to \underset{x\in D}{extr} \\ &A(t)X \le b(t) \\ &x_j \ge 0,\ j = \overline{1,n} \end{aligned}\right\} \quad (2.3)$$

$b(t)=(b_1(t),b_2(t),\dots,b_m(t))^T$ and $C(t)=(c_1(t),\dots,c_n(t))^T$ – parametric vectors of free terms of constraints and coefficients of the target function, respectively;

$A(t)_{n\times m}=(a_{ij}(t))$, $i=1\dots n$, $j=1\dots m$ is an $n\times m$-dimensional matrix of parametric constraint coefficients. The functional dependence on the parameter t can be either linear or nonlinear.

In the work on the basis of simplex method and differential transformations the methods of solving linear programming problems with parametric coefficients of the target function and right parts of constraints are considered, allowing to organize simple iterative calculations and excluding the solution of systems of inequalities.

$$\left.\begin{aligned} &\max F = \max \sum_{j=1}^{n} c_j x_j \\ &\sum_{j=1}^{n} a_{ij}x_j \begin{Bmatrix} \le \\ = \\ \ge \end{Bmatrix} b_j, \qquad i = \overline{1,m} \\ &x_j \ge 0, \quad j = \overline{1,n} \end{aligned}\right\} \quad (2.4)$$

In this case, it is assumed that the parametric functions $X=(x_1,\dots x_n)^T$, $b(t)=(b_1(t),b_2(t),\dots,b_m(t))^T$ and $C(t)=(c_1(t),\dots,c_n(t))^T$ are sufficiently smooth, have smooth differentials, and have explicit expressions.

In the studies of the authors [9] and [10] the linear programming problem with variable parameters is considered, in which not only variables included in its composition, but also coefficients, as well as the right part and parameters (coefficients at variables) of the target function can change.

An approach is proposed that allows solving linear programming problems with interdependent variable coefficients using the simplex method.

The formulation of the problem is as follows:

here: m — number of restrictions,

n — number of variables,

and additional restrictions on variable coefficients:

$$\left.\begin{aligned} &a_{ij}^- \le f_{ij}a_{ij} \le a_{ij}^+ \\ &s_k^- \le \sum_{i=1}^{m} d_{kij}a_{ij} \le s_k^+; \quad k = \overline{1,K} \\ &\sum_{i=1}^{m} p_{lij}a_{ij} = r_l; \quad l = \overline{1,L} \\ &c_j^- \le f_{ij}c_j \le c_j^+ \end{aligned}\right\} \quad (2.5)$$

In the constraint system (2.4)-(2.5), all parameters $^-_{ij}$, a^+_{ij}, s^-_k, d_{kij}, s^+_k, p_{lij}, r_l, c^-_j, c^+_j, f_{ij} are constants set during the problem formulation process. Conditions (2.5) are sometimes called interval conditions and problem (2.4) is an interval linear programming problem with interdependent variable coefficients.

Under these conditions, the following lemma is proved by constructive method:

Lemma. Let in the linear programming problem (2.4) - (2.5) there are variable coefficients that depend on a parameter of the form $a_{ij}=a_{ij}(t_j)$, having a domain of definition on a certain interval $[\alpha_j; \beta_j]$ and continuously differentiable on it. Then the simplex method applied to solve such a problem converges if the above constraints are satisfied.

The proof is constructive in the sense that it substantiates the execution of all stages of the implementation of the simplex method, taking into account condition (2.5). The essence of the proposed method is that at each step the number of the column entering the basis at the next iteration of the simplex method algorithm is determined and calculations are performed on the points of minimum of the function $a_{ij}(t)$ In [11], the following parametric programming problem was considered:

$$\left.\begin{aligned} &\sum_{i=1}^{n} c_i x_i + c_0 \le Z, \\ &\sum_{k=1}^{n} a_{ik} x_k \le b_i, \qquad i = \overline{1,m} \\ &x_j \ge 0, \quad j = \overline{1,n} \\ &Z_{\min} - ?, \quad x_i(Z_{\min}) - ? \end{aligned}\right\} \tag{2.6}$$

To solve the linear parametric programming problem, it is assumed that the coefficients change insignificantly with respect to their average values, as, for example, the cost of goods depending on the exchange rate or inflation rate. In such cases, when the relative changes in the coefficients are of the order of 10% or less, it is proposed to use the asymptotic perturbation method. Its essence is reduced to the search for the decomposition of the desired functions into functional series, the rapidity of convergence of which depends on the "*smallness parameter*" of the relative change of the functions affecting the problem.

The algorithm for solving the problem is as follows. We assume solutions of the problem (2.6) in the form of series consisting of corrections of the corresponding order:

$$Z = Z^0 + Z^1 + Z^2 + \ldots;$$
$$x_i = x_i^0 + x_i^1 + x_i^2 + \ldots;$$

The values of Z^k and x_i^k are determined in an iterative way, assuming Z^0 and x_i^0 as the first approximation in the iteration process. The values of Z^0 and $x_i^0$are the solution of a simple linear programming problem (1), when their average values a_{ij}^0, b_i^0 and c_j^0 are taken instead of the variable coefficients $a_{ij}(x)$, $b_i(x)$ and $c_j(x)$.

For example, if $c_j(x)$ is continuous in the interval [0;1], then we take as its mean value:

$$c_j^0 = \int_0^1 c_j(x)dx \tag{2.7}$$

The above brief overview shows that the problem of linear programming with variable coefficients is a well-known problem that has sufficient applications and is being studied everywhere. A certain number of methods for solving this problem have been developed, which are successfully applied depending on the problem formulation and application area.

4. Trigger graph model of the process and customs risk criterion

Based on the results of the above analysis, we can say that all methods of solving the linear programming problem with variable coefficients imply certain requirements for the ratio of the coefficients of the target function. First, an explicit form and smoothness of the parametric function are required. Second, when applying some methods of solving this problem, additional conditions such as continuity and differentiability of the given function are required.

However, in the case of the problem of optimal control of the customs clearance process, the parametric coefficients of the target function do not have an explicit expression and the above conditions cannot be required. Consequently, there are certain difficulties in applying the available methods to solve the problem, and it is required to explore new approaches.

Based on the latter findings, to minimize the target function (1.3) under conditions (1.4), it is proposed to ensure minimization of the coefficients of the target function $r_k(X)$, which represent the degree of risk of the k – process. Without violating the conditions of the stated problem and generality, we can assume that:

$$r_k(X) \ge 0, \quad k = \overline{1,18} \tag{3.1}$$

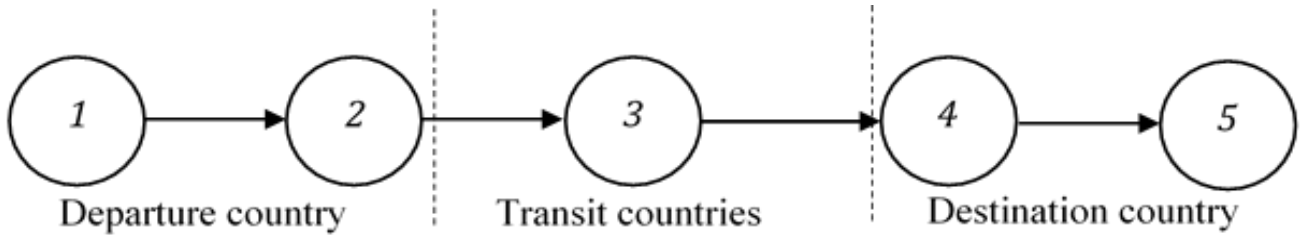
The task of minimizing the coefficients of the target function $r_k(X)$ gives rise to the task of investigating the essence of customs risks

The above studies have led to the fact that customs risks from the mathematical point of view is a function of many variables, is not presented in an explicit form, there are considerable uncertainties in its characteristics in terms of smoothness, continuity, differentiability and other qualities.

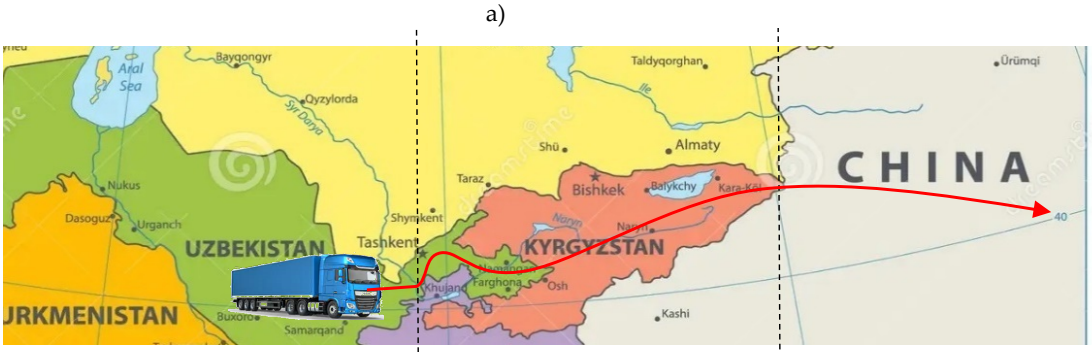
Before moving on to research on the specifics of the customs risk, it is important to consider the essence of the risk as a whole, since it is inherent in various fields of activity. There are many approaches to defining the concept of "*customs risk*". For example, the working group of the World Customs Organization, preparing the document "World Customs Organization Compendium on Customs Risk Management" in 2011, gave the following definition: "*Risk: the result of doubts arising in relation to objects*" [12]. In scientific literature it is defined as follows: "Risk is the probability of violation of customs legislation associated with evasion of payment of customs duties and taxes due" [13].

In these or other definitions of customs risk, which are observed in previously published scientific papers, the "*probability of violation of customs legislation*" passes as the main predicate. This means that the mathematical expression of customs risk must necessarily include the "*probability of violation of customs legislation*" model.

Based on the above, in order to develop a mathematical approach to this definition, it is necessary to study customs legislation both at the national level and at the



a)



b)

Figure 2: Simplified graph model (a) and explanatory scheme (b) of the process of foreign trade operations (author's development)

international level. At the same time, the importance of studying international conventions in the field of customs should be emphasized, as the process of a foreign trade operation is directly related to foreign partners.

Fig. 2. shows a simplified graph model (a) and an explanatory scheme (b) of the process of foreign trade operations.

An analysis of the customs legislation on the day of the present research shows the following. By the beginning of the second half of 2023, more than 2216 normative legal documents related to the regulation of the activities of customs authorities of the Republic of Uzbekistan were in force (Table 2).

Table 2: Information on legal documents related to the regulation of the activities of the customs authorities of the Republic of Uzbekistan

Type of normative document	**Amount**
Laws of the Republic of Uzbekistan	57
Codes of the Republic of Uzbekistan	12
Decrees of the President of the Republic of Uzbekistan	322
Resolutions of the President of the Republic of Uzbekistan	767
Resolutions of the Cabinet of Ministers of the Republic of Uzbekistan	1011
Orders of the Cabinet of Ministers of the Republic of Uzbekistan	11
legal acts registered with the Ministry of Justice of the Republic of Uzbekistan	36
Total	**2216**

(author's development)

Despite the fact that there is such an extensive customs legal framework, the main task of the customs authorities of the Republic of Uzbekistan is to protect the economic security of the country and almost all the rules of this framework are focused on the following two main tasks [14]:

a) fulfillment of the fiscal task - ensuring the completeness of customs payments collection;
b) prevention, detection and suppression of violations of customs legislation, including smuggling.

The study of customs legislation and practical experiments showed that between the above-mentioned main tasks of managing foreign trade operations, in terms of procedure, there is an irreconcilable contradiction. This contradiction is manifested in the following:

a) the key parameter for increasing customs payments is the time spent on customs clearance of foreign trade goods: the less time spent on customs clearance of a particular foreign trade cargo, the more cargo will be cleared for a certain period of time, hence, more customs payments will be made to the state budget during this period.

At the same time, the number of violations of customs legislation and the volume of goods of illegal circulation are increasing, since the time for a detailed study of the consignment of goods being processed remains minimal.

b) the key parameter for reducing customs law violations is also the time spent on customs clearance of foreign trade goods: the more time to study a specific consignment of foreign trade cargo, the less chance there is to commit violations of customs legislation.

At the same time, the amount of revenues to the state budget from customs payments decreases, as the foreign trade turnover for a certain period of time decreases.

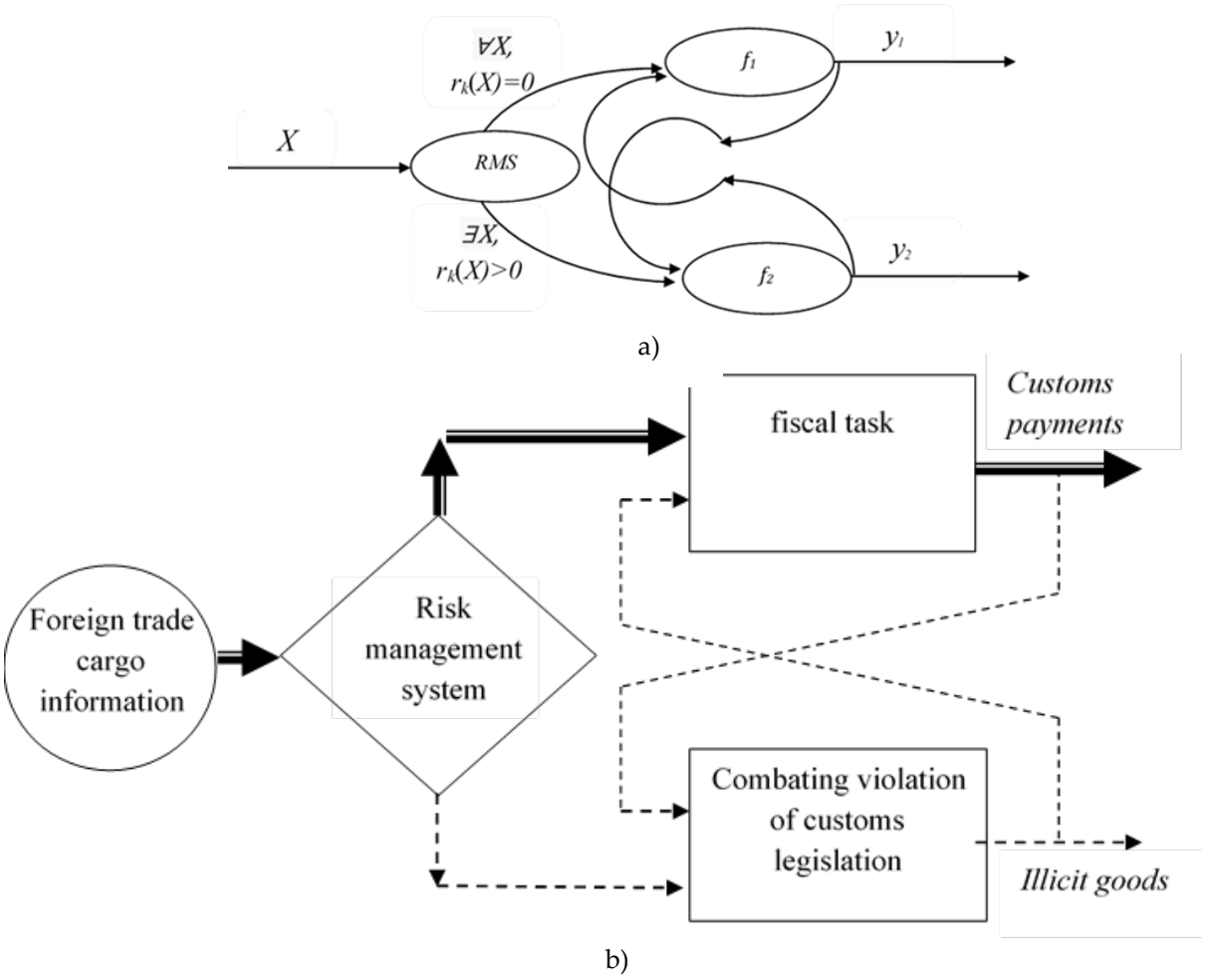


Figure 3: Trigger graph model (a) and its explanatory scheme (b) of the main tasks of foreign trade operations management (author's development)

To minimize irreconcilable contradictions between the main tasks of managing foreign trade operations in procedural terms, in practice, a customs risk management system is used. Trigger graph model of the main tasks of foreign trade operations management with application of risk management system is shown in Fig.3.

It should be noted that both functions - the function y_1, which reflects the volume of receipt of customs payments, and the function y_2, which reflects the volume of illegal goods (customs law violations) are functions of time, i.e. $y_i = y_i(t)$. Both functions are inversely proportional to the ratio of customs clearance time, i.e.

$$y_i = \varphi_i(1/t), \quad where \quad \varphi_i - linear\ functions, \quad i = 1,2 \qquad (3.2)$$

On the other hand, the conducted experiments showed that with an increase in the volume of goods of illegal circulation (offenses of the customs legislation), the receipts of customs payments decrease, i.e. functions y_1 and y_2 are inversely proportional to each other:

$$y_1 = \psi_i(1/y_2), \quad where \quad \Psi_i - linear\ function \qquad (3.3)$$

The mathematical contradiction reflected in (3.2) and (3.3) gives rise to the need for an optimization problem about the choice of time for customs clearance of foreign trade goods. It is required for each batch of goods to review the duration of customs clearance time and choose it so that there would be maximum receipt of customs payments and minimum volume of goods of illegal turnover (violations of customs legislation).

This task can be solved only if the following conditions are met:

a) choose the minimum time for customs clearance, in the absence of customs risk;
b) choose the time of customs clearance sufficient to ensure the minimization of the customs risk, if it is detected.

5. Risk assessment of the reliability of customs information

We mentioned above that in the scientific literature there is a theory according to which any customs system successively goes through several separate phases of its activity, characterized by the specifics of its relations both with foreign trade participants and with the state. Three phases of development are noted: the customs systems of developed countries are in the "*customs for foreign trade participants*" phase, in most developing countries the phase "*customs for the government*" is characteristic, and in

a number of underdeveloped countries there is a "*customs for themselves*" phase.

Proceeding from the fact that the Republic of Uzbekistan is carrying out large-scale works on transferring the customs service to the phase "customs for foreign trade participants", the authors of this paper study customs risks, categorizing them into three classes:

a) customs risks for business;
b) customs risks of economic security;
c) corruption risks.

a) when it comes to the customs risk for business, it means the submission of an unreliable customs declaration by the business to the customs authorities. Analysis of the database of violations of customs legislation for several years shows that every 4th fact about such violation is the result of false declaration.

Despite the fact that at the present stage of development of foreign trade, favorable conditions are created for a law-abiding participant in foreign trade, the laws react rather harshly towards them if they have submitted an unreliable customs declaration to the customs authorities. The consequence of such phenomena for them can sometimes be undesirable, severe and long-lasting.

Therefore, the primary task of the customs service of the Republic of Uzbekistan today is to minimize customs risks for business.

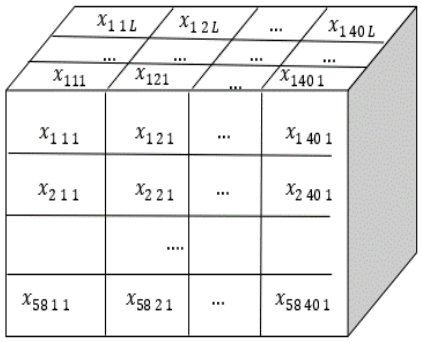
b) the customs risk of economic security is the probability of violation of customs legislation by a participant in foreign trade, associated with evasion of payment of due customs duties and taxes.

c) corruption risks shall mean abuse of official powers, receiving and giving bribes, bribery, mediation in bribery, commercial bribery or other illegal use by a customs officer of his/her official position contrary to the legitimate interests of the state, in order to obtain benefits for himself/herself or for third parties.

Thus, the first step in solving the problem of optimal management of the customs clearance process (1.3) - (1.4) is to minimize the implicit function $r_k(X)$, which represent the degree of risk of unreliable declaration of foreign trade goods, i.e. minimization of customs risks for business. To solve this problem it is necessary to assess the reliability of information about the goods on all its parameters, i.e. it is required to conduct a multivariate analysis of information about the goods. Information about the goods is fully reflected in the cargo customs declaration.

The research of the authors of this paper has shown that the customs cargo declaration is one of the fundamental documents of the customs clearance process. It is formalized in the form of a multidimensional matrix X, which is the source of state customs statistics [15].

$$X=$$



(4.1)

A brief characterization of the customs cargo declaration is as follows:

a) number of columns - 58;
b) the level of detail of each column of the cargo customs declaration is determined depending on the complexity of the task, but not more than 40;
c) L - the total number of cargo customs declaration per year.

It should be noted that each layer of this matrix corresponding to $l=l_0$ reflects a separate cargo customs declaration. It can be labeled as follows:

$$X_0 = \begin{pmatrix} x_{11l_0} & x_{12l_0}\cdots & x_{140l_0} \\ x_{21l_0} & x_{22l_0} & x_{240l_0} \\ \vdots & \ddots & \vdots \\ x_{581l_0} & x_{582l_0}\cdots & x_{5840l_0} \end{pmatrix} \quad (4.2)$$

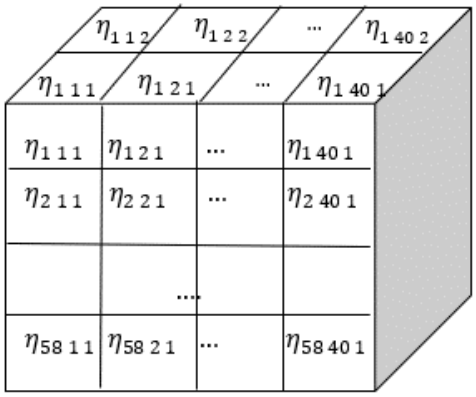
Assessing the validity of information about the goods for all its parameters requires controlling and assessing the validity of all elements of the matrix (4.2). The study of existing methods for solving this problem showed that the problem of identifying unreliable customs declarations is a special case of the general and, as you know, ancient problem of identifying false information, i.e. how to distinguish "truth" from "falsehood".

This famous problem is mentioned in many ancient writings, beginning with Aristotle (384 BC), who is the founder of logic as a science [16]. One of the great scientists who devoted his entire conscious life to the study of the task of distinguishing "truth" from "falsehood" is Imam al-Bukhari. His book Al-Jami'as-Sahih has been tested for over 11 centuries and is considered the most authentic book today [17].

The concept of information reliability has different meanings in philosophy, the theory of forensic evidence, epistemology, logic, probability theory, psychology, natural science and other areas. There is no single

definition of the term, although many famous philosophers have tried to give their own definition of the term. In logic and philosophy, reliability often acts as a synonym for the concept of "truth" and characterizes indisputable, firmly substantiated and demonstrative knowledge.

A "*threshold matrix*" (TM) is proposed to determine the concept of reliability of the elements of the customs cargo declaration (4.2.) (4.3.).

$$H=$$  (4.3)

here: η_{ij1}, η_{ij2} - some positive real numbers or textual information.

Definition 1. Each element of the matrix is $x_{ijl_0} \in X_0$ called reliable if the following condition is satisfied

$$\eta_{ij1} \le x_{ijl_0} \le \eta_{ij2} \qquad (4.4)$$

where: $1 \le i \le 58$, $1 \le j \le 40$.

Definition 2: If all elements of the matrix $x_{ijl_0} \in X_0$ are reliable, then the customs cargo declaration is called reliable.

Conditions (4.4) are called *criteria*, and the elements of the "Threshold Matrix" are called *indicators* of the reliability of the cargo customs declaration.

From Definition 1-2, the following statement is easily proved:

Statement 1. If at least one element of the matrix $x_{ijl_0} \in X_0$ does not satisfy the conditions (4.4), then the corresponding cargo customs declaration is unreliable.

To assess the reliability of a cargo customs declaration, the authors of this work propose the following function:

$$\rho_{ij} = \begin{cases} e^{\eta_{ij1} - x_{ijl_0}}, & if \quad \eta_{ij1} \ge x_{ijl_0} \\ 1, & if \quad \eta_{ij1} \le x_{ijl_0} \le \eta_{ij2} \\ e^{x_{ijl_0} - \eta_{ij2}}, & if \quad x_{ijl_0} \ge \eta_{ij2} \end{cases} \qquad (4.5)$$

here: $1 \le i \le 58$, $1 \le j \le 40$.

The function ρ_{ij} can be estimated as follows: when the conditions $\eta_{ij1} \le x_{ijl_0} \le \eta_{ij2}$ are satisfied for all $1 \le i \le 58$, $1 \le j \le 40$ function value $\rho_{ij} = 1$; otherwise, - $\rho_{ij} > 1$. In other words, the function ρ_{ij} reflects the quantitative assessment of the reliability of the element $x_{ijl_0} \in X_0$ of the cargo customs declaration. Then the matrix ϱ (4.6) is the matrix of the reliability of the cargo customs declaration X_0.

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} \cdots & \rho_{140} \\ \rho_{21} & \rho_{22} & \rho_{240} \\ \vdots & \ddots & \vdots \\ \rho_{581} & \rho_{582} \cdots & \rho_{5840} \end{pmatrix} \qquad (4.6)$$

with the above notations, the following theorem is proved:

Theorem 1. In order for the cargo customs declaration X_0 to be reliable, it is necessary and sufficient to fulfill the following condition:

$$P = \prod_{i=1}^{58} \prod_{j=1} \rho_{ij} = 1 \qquad (4.7)$$

P is the coefficient of reliability of the customs cargo declaration X_0. It follows from (4.5) and (4.7) that the coefficient *P* takes on the values *P*=1 only if for all $1 \le i \le 58$, $1 \le j \le 40$ the conditions $\eta_{ij1} \le x_{ijl_0} \le \eta_{ij2}$ are satisfied, otherwise *P* >1.

6. Algorithm of control of risks of reliability of calculation of customs payments

To verify the above results, let's consider the tasks of controlling the risks of reliability of the calculation of customs payments. The amount of customs payments for the import of goods is determined as follows:

S=D+E+V

where: D- the amount of customs duty, E- the amount of excise tax, V- the amount of value-added tax on goods. They are determined mainly by the so-called "ad valorem rates". This means that the amount of each of the above types of customs payments is determined depending on the established rate in percentage terms. For example, the rate of value added tax in the Republic of Uzbekistan is set at 12% of the customs value of the goods.

The formulas for calculating them are as follows:

D = d c;
E = ec;
V= 0.12 (c +D +E).

where: c is the customs value of the goods, d is the rate of customs duty, e is the rate of excise tax.

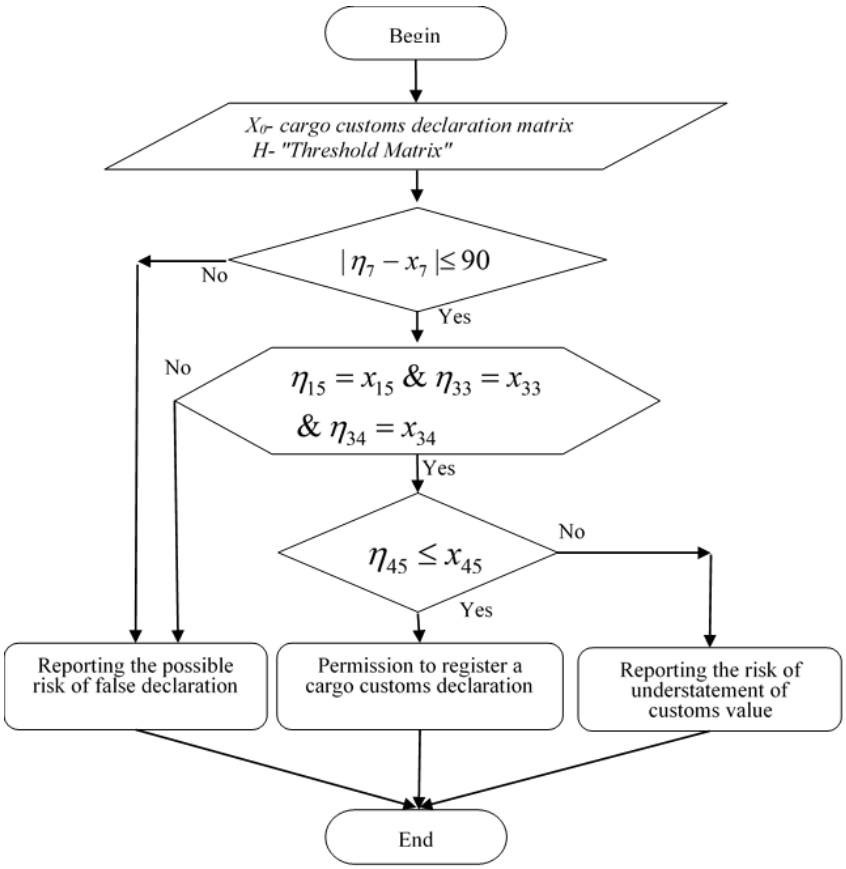


Figure 4: Algorithm for controlling the customs value of goods

After simple arithmetic transformations, you can get: *S=D+E+V= c (d+ e+0.12(1 + d + e)).*

Hence, it can be seen that the amount of customs duties on imports of a particular good depends directly on the customs value of the good c. The lower the customs value of the goods, the lower the receipt of customs payments to the state budget.

Considering this circumstance, the "Threshold Matrices" were formed in the form of "*Price Information Bulletins*". The table reflects the following data:

η_7- date of registration of the customs value of goods;
η_{33}- commodity code in accordance with the Harmonised System (HS);
η_{15} – code of the country of departure of the goods;
η_{34} – country of origin code;
η_{45}– customs value of goods.

The other elements of the "threshold matrix" for the considered criterion of the reliability of the customs value of goods are of little importance.

Then conditions (4.4) have the following form:

$$|\eta_7 - x_7| \le 90$$
$$\eta_{15} = x_{15}$$
$$\eta_{33} = x_{33}$$
$$\eta_{34} = x_{34}$$
$$\eta_{45} \le x_{45}$$

Taking into account the above designations, an algorithm for controlling the customs value of goods has been developed (Fig. 4)

It should be noted that the "*Price Information Bulletin*" is a characteristic feature of the national legislation of the Republic of Uzbekistan, and is not observed in the practice of the customs services of other countries.

7. Conclusion

In conclusion, I would like to note that the authors of this article conducted research on minimizing customs risks for business. This is due to the fact that large-scale work is being carried out in the Republic of Uzbekistan to organize the work of the customs service on the principle of "customs for participants in foreign trade". Because, the faster the country's customs service approaches the stage of development "customs for foreign trade participants," the faster the country approaches the level of developed countries.

When it comes to customs risk for a business, it means submitting an unreliable customs declaration to the customs authorities on the part of the business. An analysis of violations of customs legislation over several years shows that every 4th fact of such a violation is the result of an unreliable declaration. It is known that for an unreliable declaration, punishment is provided up to criminal. The consequences of such an incident can sometimes be undesirable, severe and long-term for business.

The research carried out within the framework of this article showed that the task of identifying unreliable customs declarations is a special case of the general and, as we know, ancient task of identifying false information, i.e. how to distinguish “truth” from “falsehood”.

To determine the reliability of a cargo customs declaration, the authors proposed a “threshold value matrix” method, which in fact forms a “knowledge base” of a production expert system that represents knowledge in the form of “IF-THEN” rules. The block diagram of the algorithm for one of these rules is shown above in Fig. 4.

The software of this expert system determines existing errors in the customs cargo declaration as they are received by the customs authorities via the Internet and automatically informs the foreign trade participant about this. No administrative or criminal sanctions will be applied to a foreign trade participant who promptly and voluntarily corrects errors.

Currently, 53 logical rules have been established in the knowledge base to control the reliability of the customs value of goods, which make it possible to localize such customs risks. As a result of the implementation of these rules in 2022, in 88 thousand 897 cases, the risks of “unreliability of the customs value of goods” and debts to the state budget in the equivalent of more than 9 million 968.8 thousand US dollars were prevented.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We thank the editor of the Journal of Engineering Research and Sciences and the anonymous reviewers for their valuable comments. All errors and omissions remain the responsibility of the authors.

References

[1]. A. O. Rudneva, "International trade: specifics and prospects of participation of developed, developing and transition countries," *MIR (Modernization. Innovation. Development)*, vol. 8, no. 3, 430-438, 2017, doi: 10.18184/2079–4665.2017.8.3.

[2]. A.D. Ershov, "Formation of customs services in foreign economic activity," *Scientific notes of the St. Petersburg branch of the Russian Customs Academy*, no. 1(23), 174-192, 2005.

[3]. S.V. Zvonarev, "Fundamentals of mathematical modeling: textbook," Yekaterinburg: Ural University Press, 112 p., 2019.

[4]. R.I. Ibyatov, "Optimization methods in problems of mathematical modeling," methodological guidelines, Kazan: Kazan State Agrarian University Publishing House, 32 p., 2016.

[5]. R. V. Fedorenko, "Methodology of management of service complexes in the customs sphere," *Dissertation for the degree of Doctor of Economics*, Samara, 2015. https://www.sseu.ru/wp-content/uploads/2015/06/Dissertatsiya-Fedorenko-R.V.pdf. (dissertation in Russian with an abstract in English)

[6]. A.G. Pinsker, "A linear programming problem with variable coefficients of the purpose function," *Sib Math J*, vol. 20, 466–468, 1979, doi: 10.1007/BF00969958. (article in Russian with an abstract in English)

[7]. M. O. Gavrilova, "On problems of linear programming with piecewise constant coefficients," *Scientific journal “Bulletin of the Perm State Technical University. Chemical technology and biotechnology”*, Perm, No. 9, 172-179, 2010.

[8]. A. G. Avetisyan, L. S. Gyulzadyan, "A method for solving problems of parametric linear programming based on differential transformations," *Scientific journal "Izvestia of the Tomsk Polytechnic University"*, vol. 324, No. 2, 25-30, 2014.

[9]. D.A. Salimonenko, "A method for solving a linear programming problem with variable coefficients in the form of parametric functions," *Scientific journal "Vestnik of the Bashkir University"*, Ufa, vol. 20, No. 1, 25–29, 2015.

[10]. D.A. Salimonenko, A.M. Ziganshin, V.A. Mudrov, Yu.D. Salimonenko, "On interdependent variable coefficients in linear programming problems," *Scientific journal "Mathematical Structures and Modeling"*, Omsk, vol. 2, no. 58,. 96–111, 2021, doi: 10.24147/2222-8772.2021.2.96-111.

[11]. S.P. Kravchuk, I.S. Kravchuk, O.V. Tatarnikov, E.V. Shved, "Perturbation method for solving linear programming problems with a parameter," *Scientific journal "Fundamental Research"*, Moscow, No. 5,.299-303, 2015.

[12]. WCO, "Customs Risk Management Compendium," Brussels, Belgium, June 2011, http://www.wcoomd.org.

[13]. S.E. Tamrazyan, “Customs Risks: Essence, Management and Evaluation,” *Scientific journal “Economics and Management in the XXI century: development trends”*, No. 23,.168-172, 2015.

[14]. A.A. Saidov, F.A. Khakimova, T.T. Abdurakhmonov, "The concept and model of the 'soft component' of the risk management system of customs authorities," *Scientific journal “Bulletin of the Russian Customs Academy”*, Moscow (Russia), No. 3, 100-109, 2022, doi: 10.54048/20727240 _2022_03_100 (article in Russian with an abstract in English).

[15]. A.A. Saidov, "Classical Methods of Controlling the Reliability of Information and Features of Their Application to Customs," *Monograph*, Tashkent, 498 p, 2021. (monograph in Russian with an abstract in English)

[16]. Aristotle, "Metaphysics," Translation from Greek by P. D. Pervov and V. V. Rozanov, Moscow: Institute of Philosophy, Theology and History of St, Thomas, 232 p., 2006.

[17]. Muhammad ibn Ismail al-Bukhari, "Al-Jami' as-sahih," Translation by Vladimir (Abdullah) Nirsha, Moscow (Russia), Umma Publishing House, 448 p., 2017.

ILKHOM MUKHTOROV - First Deputy Chairman of the Customs Committee under the Ministry of Economy and Finance of the Republic of Uzbekistan. Project Manager of the Customs Risk Management System. Scientific interests are connected with the intellectualization of the processes of organization of customs control. Author of the concept "Export in three steps", is the author of the monograph on factorial data analysis.

TAKHIR ABDURAXMONOV - Chief Inspector of the Department of Information and Communication Technologies and Cybersecurity of the Customs Committee of the Republic of Uzbekistan, participated in the work of the group on the development of the customs information system "Electronic Declaration". Her research interests include monitoring the reliability of customs information, digitalization of customs control processes and interdepartmental interaction within the framework of the customs expert information system..

ABDUSOBIR SAIDOV - Abdusobir Saidov - Head of the Department of Information Technology and Mathematics, Customs Institute of the Customs Committee under the Ministry of Economy and Finance of the Republic of Uzbekistan. Doctor of technical sciences, professor. Scientific research is devoted to ensuring the reliability of customs information, is the author of several monographs on this topic

Received: 14 March, 2024, Revised: 13 May, 2024, Accepted: 16 May, 2024, Online: 29 May, 2024

DOI: https://doi.org/10.55708/js0305003

Analyzing the Impact of Optical Wireless Communication Technologies on 5G/6G and IoT Solutions: Prospects, Developments, and Challenges

Ramsha Khalid[1,2], Muhammad Naqi Raza[1]

[1] Department of Electrical Engineering Technology, University of Sialkot, Sialkot, 51310, Pakistan

[2] Department of Electrical Engineering, University of Lahore, Lahore, 53720, Pakistan

*Corresponding author: Ramsha Khalid, University of Sialkot, Sialkot, Pakistan Email: ramshakhalid2404@gmail.com

ABSTRACT: The imminent 5G and 6G communication systems are projected to exhibit substantial advancements in comparison to the current 4G communication system. Several critical and prevalent concerns pertaining to the service quality of 5G and 6G communication systems encompass elevated capacity, extensive connectivity, minimal latency, robust security measures, energy efficiency, superior quality of user experience, and dependable connectivity. Undoubtedly, 6G communication is expected to offer markedly improved performance across these domains compared to 5G communication. The integration of the Internet of Things (IoT) within the framework of the tactile internet is anticipated to be a fundamental component of advanced communication systems, encompassing both 5G and beyond (5GB), such as 5G and 6G. Consequently, 5GB wireless networks will encounter various challenges in accommodating diverse types of heterogeneous traffic and meeting the specified parameters related to service quality. Optical wireless communication (OWC), alongside various other wireless technologies, emerges as a promising candidate to fulfill the requisites of 5G communication systems. This comprehensive review articulates the efficacy of OWC technologies, including Visible Light Communication (VLC), Light Fidelity (LiFi), Optical Camera Communication (OCC), and Free Space Optics (FSO) Communication, as a viable solution for the successful deployment of 5G/6G and IoT systems.

KEYWORDS: 5G, 6G, internet of things, heterogeneous traffics, wireless technologies, communication systems, Optical Wireless Communication

1. Introduction

In recent times, OWC technologies have garnered significant research attention owing to their notable features [1–5]. OWC designates wireless connectivity utilizing the optical spectrum. OWC has positioned itself as a favored complementary technology to Radio Frequency (RF)-based wireless technologies, particularly in the context of future communication networks, encompassing the 5G and 6G communication systems. OWC technologies exhibit several notable features, including broad spectrum coverage, high data rates, minimal latency, robust security, cost-effectiveness, and energy efficiency. These attributes effectively cater to the demanding specifications of 5GB communications, exemplified by 5G and 6G technologies. In addition to this, the IoT network is gaining significant importance, with a proliferation of end-user devices or sensors being interconnected within IoT. Furthermore, the integration of tactile internet will emerge as a pivotal aspect of future IoT, facilitating real-time communication systems across various societal, industrial, and commercial applications. In visualizing the concept of IoT, there is an exponential surge in the quantity of physical devices connected to the internet [6]. Hence, the IoT generates a substantial volume of data. OWC technologies assume a crucial role in sensing, monitoring, and facilitating resource sharing within the extensive device connectivity of IoT networks [2,6]. Additionally, OWC can effectively fulfill the low-power consumption and stringent security requisites of IoT.

The specifications for the 5G communication system have been finalized, and it is anticipated that 5G will be fully implemented by 2020 [7]. The forthcoming 5G communication infrastructure will introduce novel services characterized by exceptionally high Quality of

Service (QoS). Key attributes of 5G communication services will encompass unparalleled system capacity, minimal latency, enhanced security measures, extensive device connectivity, minimal energy consumption, and exceptional Quality of Experience (QoE) [7–11]. The introduction of the 6G communication system is projected to occur within the timeframe spanning 2027 to 2030. While the precise specifications for 6G have yet to be defined, numerous researchers are actively engaged in its development [12–16]. Research challenges encompassing capacity enhancement, augmented connectivities, latency reduction, heightened security, improved energy efficiency, elevated user QoE, and enhanced reliability are focal points addressed by both the 5G and prospective 6G communication systems. The forthcoming 6G communication infrastructure is anticipated to serve as a global communication cornerstone, offering service levels significantly superior to those of 5G.

RF currently serves as a prevalent choice for diverse wireless connectivity needs. However, RF-based wireless communication encounters significant hurdles, including spectrum limitations, susceptibility to interference, and stringent regulatory constraints. Sole reliance on RF technologies proves inadequate in meeting the demands of 5G and IoT networks. Consequently, researchers are diligently exploring alternative spectrums to address the escalating requirements. One particularly promising avenue involves leveraging a significantly expansive optical band. This strategic shift toward OWC holds considerable potential for advancing 5G and IoT networks, offering distinct advantages over conventional RF-based networks. These advantages encompass heightened data rates, diminished latency, enhanced security, and improved energy efficiency [1–3], [6]. Effective communication spans distances ranging from a few nanometers to over 10,000 kilometers, facilitated by the implementation of various OWC systems [2]. Key technologies integral to OWC networks comprise VLC [6][17–19], LiFi [20–22], OCC [23–27], and FSO [28–30]. A subsequent section provides a concise exploration of the distinctions and commonalities inherent in these technologies. Each of these technologies possesses unique strengths alongside certain limitations. Diverse OWC technologies present a spectrum of services catering to indoor, outdoor, and space communications. Consequently, OWC technologies assume a crucial role in realizing the objectives of 5G and IoT systems.

Our prior review paper concerning OWC [2] extensively examines and compares various optical wireless technologies, offering a comprehensive understanding of their distinctions. However, the primary objective of the current review paper diverges from providing a detailed explanation of OWC technologies. Instead, its focus is on illustrating how OWC technologies can serve as an effective solution for the seamless deployment of 5G/6G and IoT systems. Within this study, we delineate potential detailed solutions for 5G/6G and IoT utilizing diverse OWC networks. This paper's contributions can be succinctly outlined as follows:

1. Comprehensive examination of the key characteristics of 5G and IoT networks, with a brief presentation of potential 6G requirements.
2. Concise discussion of various OWC technologies within the context of 5G/6G and IoT systems.
3. Detailed exploration of the scope of OWC technologies in meeting the specific requirements of 5G/6G and IoT deployments.
4. Thorough survey of recent advancements in OWC technologies pertaining to 5G and IoT solutions, accompanied by a discussion on emerging research trends.
5. In-depth consideration of challenging issues associated with the deployment of OWC for 5G/6G and IoT solutions.

The subsequent sections of the paper are structured as follows: Section 2 furnishes a concise overview of the requirements associated with 5G, 6G, and IoT. Section 3 provides an in-depth description of various OWC technologies. In Section 4, the potential of OWC technologies to address the demands of 5G, 6G, and IoT systems is elucidated. Section 5 delves into several key challenging issues inherent in OWC-based 5G/6G and IoT solutions. Finally, Section 6 encapsulates the conclusion of this paper.

2. Concise Examination of the Requirements for 5G, 6G, and IoT

5G is anticipated to deliver a significant enhancement in key attributes compared to 4G, enabling efficient support for the burgeoning array of heterogeneous multimedia applications with varying requirements [11]. The specifications for 5G requirements have been delineated, with full deployment of the 5G system anticipated by 2020. The essential requirements of 5G can be succinctly summarized as follows:

- *High Traffic Volume:* The mobile data volume per unit area is projected to increase by a factor of 1000 in comparison to 4G wireless networks, accompanied by a surge in the number of connected wireless devices, which is expected to be 100 times higher.
- *Massive Connectivity:* 5G is designed to facilitate massive connectivity, with the capability to connect ten to 100 times more devices than the 4G communication system [11].

- *High User Data Rate Link:* The 5G networks are mandated to support exceptionally high user data rates, enabling users to achieve up to 10 Gbps, representing a ten to 100-fold increase compared to 4G.
- *Low-Energy Consumption:* Significantly reduced energy consumption is a pivotal requirement in the 5G communication system, aiming to achieve more than a 90% reduction, i.e., 10 times lower compared to 4G networks [11].
- *Extremely Low Latency:* Ensuring extremely low latency, with end-to-end latency levels ranging from sub-millisecond to a few milliseconds, is a critical objective for 5G networks [11].

Researchers are currently engaged in the standardization of requirements for 6G networks [12–16,31–34]. A pivotal requirement for 6G is anticipated to be ultra-high bit rates per device, ranging from tens of gigabits per second to terabits per second [12,31]. Furthermore, 6G is projected to exhibit 1000 times higher simultaneous wireless connectivity compared to 5G. Envisaged characteristics for 6G encompass ultra-long-range communication coupled with ultra-low-power consumption, ensuring user experiences with latency of less than 1 millisecond [13]. Other key anticipated features of 6G include spatial multiplexing, higher spectral efficiency at 100 bits per second per Hertz, ultra-high wireless security, exceptional reliability, ultra-low-power consumption, and the integration of massively connected complex networks.

The networks will possess distinct characteristics designed to accommodate the demands of 5G wireless communication systems. The essential features of future 5G and 6G networks can be encapsulated as follows:

- *Ultra-High-Density Network:* To ensure consistent QoE, accommodate massive connectivity, and meet high capacity demands, 5G networks are anticipated to exhibit significantly higher density, characterized by ultra-dense heterogeneous networks, compared to their 4G counterparts.
- *Small-Cell Networks:* The establishment of high-density small-cell networks is identified as a fundamental characteristic in the design of 5G communication systems.
- *Higher Spectral Efficiency:* 5G systems are poised to optimize frequency spectrum utilization through the incorporation of multiple-input and multiple-output techniques, advanced coding and modulation schemes, and innovative waveform design. The targeted spectral efficiency for 5G is set to be at least three times higher than that of 4G networks.
- *Low Cost:* A key objective for 5G systems is to achieve a 100-fold increase in efficiency compared to 4G systems, delivering a hundred times more data traffic using the same energy across the network. This necessitates the adoption of low-cost network equipment, reduced deployment expenses, and enhanced power-saving functionalities on both network and user equipment sides [35].
- *Offloading Heavy Traffic to Indoors:* Recognizing that nearly 80% of mobile traffic is generated indoors, a strategic characteristic of 5G and 6G networks involves offloading this substantial data volume to indoor small cells. This approach aims to alleviate the strain on macrocells, preserving valuable resources and enhancing overall network efficiency [36].

3. Brief Overview of OWC Technologies

The four primary OWC technologies, namely VLC, LiFi, OCC, and FSO, are regarded as promising solutions to address the requirements of 5G/6G and IoT networks due to their unique features. Figure 1 provides a concise depiction of the architectures of these technologies [37]. In terms of infrastructure, these technologies exhibit variations in transmitter types, receiver configurations, and communication media. VLC utilizes light-emitting diodes (LEDs) or laser diodes (LDs) as transmitters and photodetectors (PDs) as receivers, utilizing only visible light (VL) as the communication medium. LiFi, akin to Wireless Fidelity (WiFi) technology, offers high-speed wireless connectivity alongside illumination, employing LEDs or diffuse LDs as transmitters and PDs as receivers. While VL serves as the forward path medium, LiFi employs infrared (IR) for the return path communication, although VL can also be utilized for the return path. However, the uplink communication performance in both VLC and LiFi may be constrained as receiver devices in most user equipment, such as smartphones, are not equipped with high-power LEDs [38–40]. Furthermore, they exhibit limitations in return path performance when the uplink involves diffused light, facing significant interference from the downlink lights. OCC employs an LED array or light as a transmitter, with a camera or image sensor serving as the receiver. The inclusion of built-in complementary metal-oxide semiconductor cameras enhances the capability to capture photos and videos [41]. The camera can be of either global shutter or rolling shutter type [42]. OCC typically utilizes VL or IR as the communication medium, although the ultraviolet (UV) spectrum can also be employed. FSO technology commonly employs a LD and PD as the transmitter and receiver, respectively. However, heterodyne optical detection receivers are also utilized in FSO communication. Typically, it operates using IR as the communication medium but can also utilize VL and UV. Table 1 outlines a comparison of performance metrics

across various OWC technologies [37]. These technologies exhibit distinct differences, with each offering specific characteristics. Notably, VLC distinguishes itself by employing visible light as its communication medium. A LiFi system is required to support seamless mobility, bidirectional communication, point-to-multipoint, and multipoint-to-point communications. Among all OWC technologies, only the OCC system utilizes a camera or image sensor as a receiver. Leveraging the narrow beams of focused light from a LD transmitter, FSO systems can establish both long-distance and high-data-rate communication links. For further insight into the variances among OWC technologies, refer to our previous work [2].

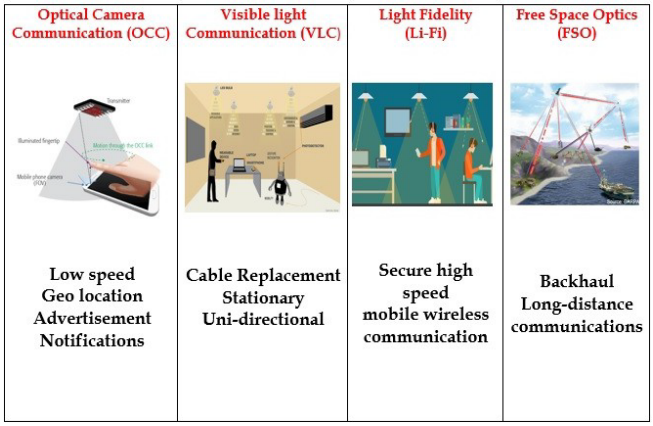


Figure 1: Taxonomies of OWC for 5G, 6G, and Internet of Underwater Things Communications.

4. OWC Technologies for the 5G, 6G, and IoT Solutions

4.1. Advantages of Opting for OWC Technologies

The RF band spans from 3 kHz to 300 GHz within the electromagnetic spectrum [2]. However, the range of 3 kHz to 10 GHz is predominantly utilized by existing wireless technologies due to its favorable communication properties. This spectrum is nearing exhaustion and falls short in meeting the high demands of 5G/6G and IoT networks. Additionally, it is subject to stringent regulations imposed by local and international authorities.

O WC emerges as a compelling alternative, offering outstanding features to address these stringent requirements. OWC finds application across a diverse range of scenarios, including machine-to-machine, device-to-device, chip-to-chip, vehicle-to-vehicle, vehicle-to-infrastructure, infrastructure-to-vehicle, point-to-point, and point-to-multipoint communications [2,6,29]. The inherent properties of light enable connectivity across a wide range, spanning from nanometers to over 10,000 km. This facilitates various communication scenarios, such as ultra-short-range inter-chip interconnects using FSO systems and in-body networks employing VLC, OCC, or LiFi systems. Other applications encompass short-range LiFi, vehicle-to-everything (V2X) communications, indoor positioning, medium-range inter-building networks, long-range inter-city backhaul connectivity, and extended-range satellite-to-satellite communications.

Furthermore, OWC technologies offer the capability to establish high-data-rate communication links. Key features of OWC encompass a wide unregulated bandwidth, enhanced security measures, low power consumption, cost-effectiveness in infrastructure and device deployment, absence of interference with RF devices and networks, high Signal-to-Noise Ratio (SNR), and seamless integration into existing lighting infrastructures. However, a notable limitation of OWC systems is the susceptibility to transmission blockage by obstacles.

The coexistence of RF and OWC networks presents an effective strategy for mitigating the limitations inherent in individual RF-based and optical wireless communication systems. Figure 2 showcases several notable 5G/6G and IoT platforms leveraging OWC technologies [37]. OWC networks have the capacity to support a myriad of applications across various aspects of daily life, including V2X communications, underwater communications, cellular connectivity support, space communication, smart shopping, eHealth, and smart home systems. This section elucidates how OWC networks can deliver effective solutions for the deployment of 5G, 6G, and IoT networks.

Table 1: Performance Metric Comparison Across Different Optical Wireless Communication Technologies [2,18,20,24,28,43]

Problem	**Parameter**	**VLC**	**Lifi**	**OCC**	**FSO**
Topology of communication	Direction	Uni or Bi-direction	Bi-direction	Uni-direction	Uni or Bidirection
Area of Communication	Distance	20m	10m	60m	10, 000km
Deployment	support for mobility	Not- compulsory	compulsory	Not- Compulsory	No
Effect on environment	Indoor/ Outdoor	No/Yes	No/Yes	No	Yes
Obstruction	Level of Interference	Low	Lows	Zero	Low
Speed of communication	Data rate	100Gbps using LD and 10Gbps using LED	100Gbps using LD and 10Gbps using LED	55Mbps	40Gbps
Network Performance	Security (related to data encryption and protection measures)	High	High	High	High

4.2. Achieving Service Quality Characteristics

Substantial Capacity Enhancement: Achieving thousand-fold capacity improvements in 5G networks necessitates a significantly broader bandwidth, a requirement readily met by the optical spectrum. Table 2 provides a comparison of RF and optical frequencies within the electromagnetic spectrum [44]. The RF band occupies merely 300 GHz of this vast spectrum, while the optical band (ranging from 300 GHz to 30 PHz) offers considerably greater potential. Currently, only a fraction of the optical spectrum, encompassing parts of visible light, near-infrared, and middle ultraviolet, is actively utilized. However, ongoing research aims to expand utilization across the optical spectrum and enhance its efficiency. Notably, the terahertz band (0.3–3 THz) within the infrared region is anticipated to play a crucial role in future high-data-rate cellular communications [31]. Leveraging the expansive optical spectrum through various OWC technologies presents an opportunity to accommodate the substantial data capacity requirements. Additionally, high-speed network connectivity is imperative to support the extensive connectivity demands of massive IoT deployments. Thus, the optical spectrum holds promise in handling the substantial data traffic generated by high-data-rate heterogeneous multimedia applications in 5G, 6G, and IoT networks.

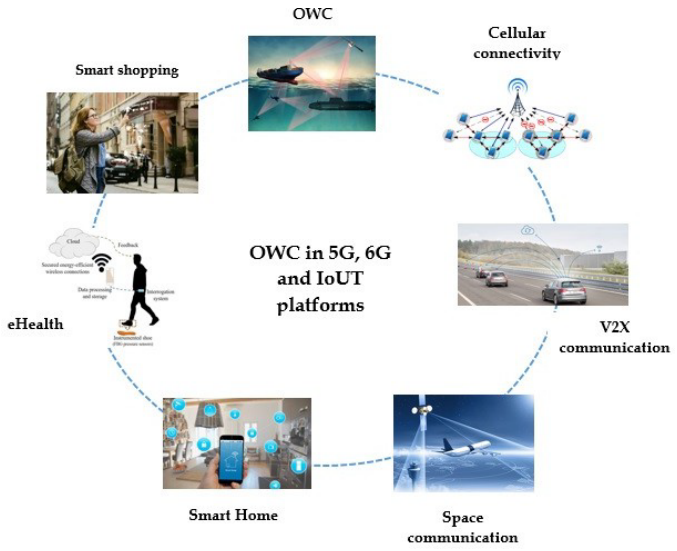


Figure 2: OWC networks for the 5G/6G and IoT platforms.

Ultra-High User Data Rate: The anticipated transmission rates for 5G mobile communication systems are projected to average around 1 Gbps, with a peak rate of 10 Gbps [8]. Subsequently, 6G is expected to support even higher bit rates ranging from tens of Gbps to Tbps per device. Notably, VLC and LiFi technologies demonstrate the capability to deliver exceptionally high-data-rate services at the user level. LiFi, in particular, offers comprehensive network support, encompassing point-to-multipoint, multipoint-to-point, and bidirectional communications akin to WiFi. VLC has already achieved a confirmed data rate of 100 Gbps [18,45]. FSO technology also excels in supporting high-data-rate services both indoors and outdoors, facilitating outdoor remote high-speed connectivity. Additionally, OWC utilizing the UV band extends its capabilities to offer high-data-rate, non-line-of-sight communications [4]. Ongoing research initiatives aim to further elevate data rates within OWC technologies. Consequently, OWC technologies emerge as valuable complementary solutions for enabling high-data-rate connectivity in 5G, 6G, and advanced communication systems. As illustrated in Figure 3, a diverse array of OWC technologies facilitates high-speed connectivity scenarios for both indoor and outdoor users, as well as in V2X communications, offering promising prospects for supporting advanced communication systems beyond 5G and 6G [44].

Table 2: Comparison of RF and optical spectra [2–6,21,29]

Property	**RF Spectrum**	**Optical Spectrum**
Frequency Range	Limited (3 kHz to 300 GHz)	Extensive (300 GHz to 30 PHz)
Bandwidth	Restricted	Broad
Utilized Spectrum	Primarily below 300 GHz	A small portion (Visible light, near-infrared, middle ultraviolet) actively used
Future Research	Limited expansion potential	Ongoing research to explore and expand utilization
Emerging Band	Terahertz band (0.3–3 THz) within infrared	Potential for high-data-rate cellular communications in terahertz range [31]
Communication Medium	Radio waves	Light waves
Interference Potential	Susceptible to interference due to crowded spectrum	Lower interference potential as optical spectrum is underutilized
Capacity Potential	Limited capacity due to spectrum congestion	High capacity potential due to broad spectrum availability

Ultra-low latency: Achieving low latency is a critical requirement for communication systems, especially in the context of 5G and beyond. OWC systems typically operate along line-of-sight (LOS) paths, resulting in minimal communication distance and no signal loss due to obstructions. In contrast, RF-based communications utilize both LOS and non-line-of-sight (NLOS) paths, encountering significant signal loss in NLOS scenarios and increased communication distances. Despite both optical and RF signals propagating at the speed of light, optical communication systems demonstrate faster communication due to rapid propagation. Furthermore, optical systems exhibit short processing times, enabling the provision of communication services with a fraction

of millisecond end-to-end delays. Consequently, OWC technologies emerge as a promising solution for 5G communication systems, delivering services with negligible latency.

Ultra-low-energy consumption: Energy efficiency stands out as a paramount requirement in the design of 5G, 6G, and IoT systems. OWC systems, predominantly structured around LEDs, align with this imperative. Currently deployed LEDs exhibit minimal power consumption, and ongoing global research endeavors are focused on further reducing their energy usage. Notably, LEDs serve a dual purpose by functioning as both illumination sources and communication transmitters, eliminating additional energy consumption when utilized for illumination. In comparison to RF sensors, LED sensors demonstrate significantly lower energy consumption. Consequently, OWC technologies present a compelling solution, offering communication systems with markedly low power consumption. This aligns seamlessly with the critical demand for energy-efficient communication systems in the deployment of 5G and IoT technologies.

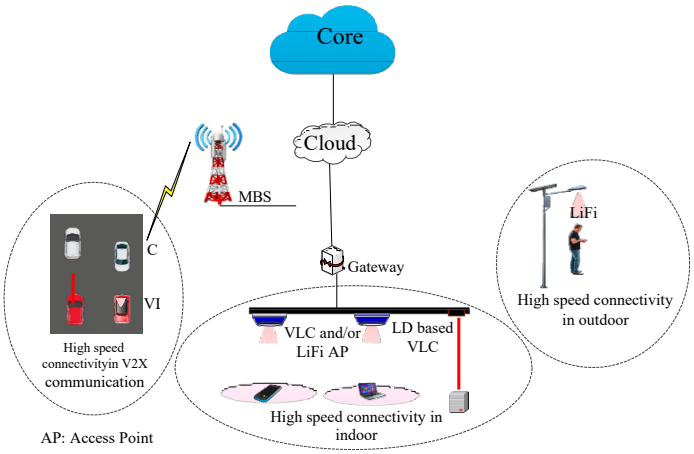


Figure 3: Achieving high-speed connectivity through various OWC technologies.

Reliable connectivity: Ensuring dependable connectivity stands as a pivotal criterion for any communication system. OWC systems offer a notably elevated SNR, particularly beneficial for indoor users. In outdoor scenarios, OCC ensures interference-free communication and a robust SNR, maintaining stable performance even with increased communication distances. FSO also exhibits commendable SNR levels for long-distance outdoor communication. Furthermore, OWC networks present an additional tier for indoor users, contributing to heightened communication system reliability. Consequently, OWC systems play a crucial role in enhancing connectivity reliability for users within the realms of 5G/6G and IoT networks.

Ultra-high security: OWC technologies, essential for the robust communication demanded by 5G, 6G, and IoT networks, ensure a high level of security. Due to the inability of OWC signals to penetrate obstacles, external entities are prevented from unauthorized access to sensitive information. The impervious nature of OWC technology prevents external network hacking devices from intercepting internal optical signals. This unparalleled security feature makes OWC systems particularly well-suited for the exchange of information in highly sensitive domains, such as healthcare. Consequently, OWC systems provide an elevated level of security for 5G/6G and IoT networks.

4.3. Fulfilling the Network and Infrastructure Characteristics

Network densification using highly dense heterogeneous networks: Network capacity enhancement can be achieved through three primary methods: network densification, spectrum efficiency optimization, and utilization of additional frequency spectra. Network densification involves the strategic addition of more cell sites to augment capacity, encompassing the deployment of small cells and the optimization of frequency utilization. This approach strategically places cell sites in capacity-constrained areas to augment overall capacity and alleviate traffic congestion on surrounding sites. Network densification is particularly relevant in densely populated areas with significant traffic volumes. The 5G/6G communication systems, characterized by high system capacity and per-user data rates, necessitate the densification of access networks and the deployment of supplementary network infrastructures. Increasing the number of small cells can boost traffic volume, while reducing the access network-to-user distance enhances achievable data rates. Consequently, network densification, specifically through the deployment of small cells, becomes imperative to fulfill the demands of 5G/6G paradigms. In dense deployments, a combination of macrocells, wide-area networks, and various indoor and outdoor optical or RF small cells is employed. Each indoor environment may host multiple optical small cells (e.g., VLC, LiFi, and OCC networks) alongside RF small cells. Outdoor applications, such as vehicular networks and street lighting, also utilize numerous optical small cells for communication. The dense deployment of OWC networks aligns with the network densification criterion, ensuring a high-capacity FSO backhaul connectivity. Figure 4 illustrates that the OWC-based small-cell networks, in conjunction with RF small cells, contribute to a highly dense network deployment.

Multi-tier architecture and convergence of heterogeneous networks: To address the evolving requirements of future communication, networks will leverage a multi-tier architecture comprising broader coverage satellite and/or macrocell networks supporting smaller cells housing RF small cells alongside optical VLC, LiFi, and OCC networks. In this architecture, VLC and LiFi technologies form a sub-tier below RF small cells. Illustrated in Figure 4 is an exemplary depiction of this multi-tier architecture

featuring macrocells, RF small cells, and optical small cells. The integration of optical small cells, including VLC and LiFi, presents an opportunity to augment high-capacity capabilities within multi-tier wireless heterogeneous networks. Consequently, the burden on costly satellite or macrocell networks can be alleviated through load offloading to small-cell networks. Indoor OWC systems can efficiently serve a significant number of users, enhancing the overall service quality provided by outdoor macrocell and satellite networks, which are often constrained by capacity limitations. Furthermore, the incorporation of OWC technologies within multi-tier heterogeneous networks addresses the limitations inherent in RF-based wireless communication systems. Optical and RF signals operate independently, mitigating interference effects within the multi-tier network infrastructure. In essence, OWC technologies will assume a pivotal role in the advancement of multi-tier heterogeneous networks, spanning across 5G, 6G, and future generations of communication systems.

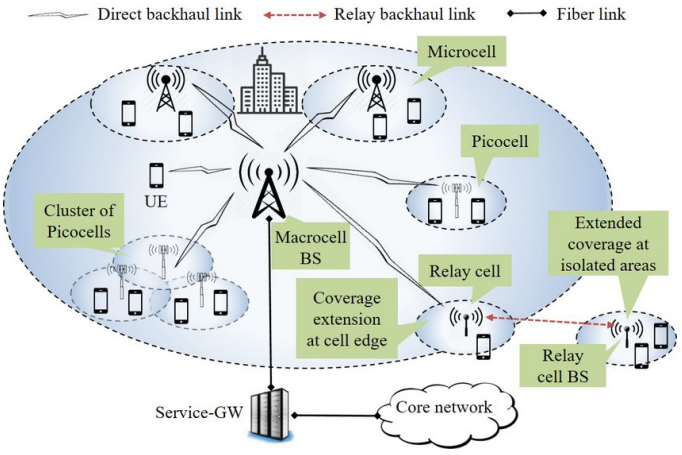


Figure 4: Scenario of heterogeneous multi-tier networks containing an RF microcell, many RF smallcells, and a large number of optical small cells.

Provision of hybrid network connectivity: Each of the distinct RF and optical wireless technologies possesses inherent limitations and advantages. The integration of heterogeneous networks, characterized by the coexistence of both RF and OWC technologies, offers an effective solution to overcome these limitations. The concurrent operation of two systems enhances link reliability and facilitates load balancing, thereby optimizing network performance. In outdoor applications, the hybrid system proves particularly advantageous in mitigating atmospheric effects. Figure 5 provides Performance Analysis of Hybrid Radio Frequency and Free Space Optical Communication Networks with Cooperative Spectrum Sharing. Collaboration between RF and optical links is leveraged to establish direct or relay-based connectivity from a source to a destination. The relay system incorporates optical links, connecting either from source-to-relay or relay-to-destination within the hybrid framework. Additionally, the simultaneous utilization of optical and RF links is possible in either or both of these connection scenarios. The configuration of forward and return communication links may vary based on application scenarios and the specific hybrid architecture. This can involve separate forward and return paths or the sharing of paths, where optical links handle the forward path, and RF links manage the return path. Consequently, OWC technologies assume a pivotal role in the strategic design of hybrid systems, effectively mitigating limitations and providing viable solutions within the context of 5G/6G networks.

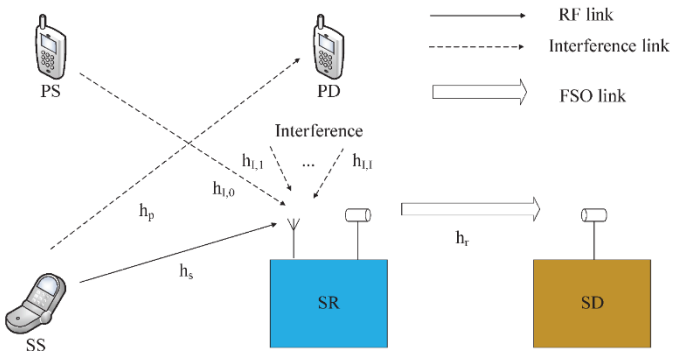


Figure 5: Performance Analysis of Hybrid Radio Frequency and Free Space Optical Communication Networks with Cooperative Spectrum Sharing

Massive device connectivity: Robust connectivity on a large scale stands as a pivotal attribute in the landscape of future communication systems. In the 5G era, the IoT is anticipated to interconnect a diverse array of up to 50 billion heterogeneous devices. This connectivity extends beyond mobile phones, encompassing applications in vehicles, household electronics, and medical equipment, contributing to the realization of a smart society [46]. The IoT, facilitated by massive connectivity, enables the integration of various sensors and physical devices, allowing them to communicate and interact autonomously, free from human intervention [47]. Projections for the 6G paradigm indicate an even broader scope, connecting a greater number of intelligent devices. OWC emerges as a pivotal enabler for achieving massive connectivity. The escalating use of LEDs is noteworthy due to their cost-effectiveness, low energy consumption, and extended lifespan. OCC, in particular, garners significant interest within the realm of IoT. Leveraging existing or minimally modified infrastructures, OCC presents economically viable solutions for a diverse range of IoT applications. Thus, OWC technology, employing low-power LEDs, has the potential to establish an extensive network of connections, aligning with the objectives of 5G/6G and IoT networks. Figure 6 illustrates various examples of widespread connectivity across different environments through diverse OWC technologies, supporting applications in homes, healthcare, transportation systems, remote connectivity, and smart grid systems [44]. In the context of smart grids, which integrate operational and energy-measuring devices like smart meters, appliances, renewable energy resources, and energy-efficient

technologies, OWC technologies facilitate extensive connectivity. Through these interconnected elements, smart grids serve as foundational components for effective energy management within a sustainable environment [47].

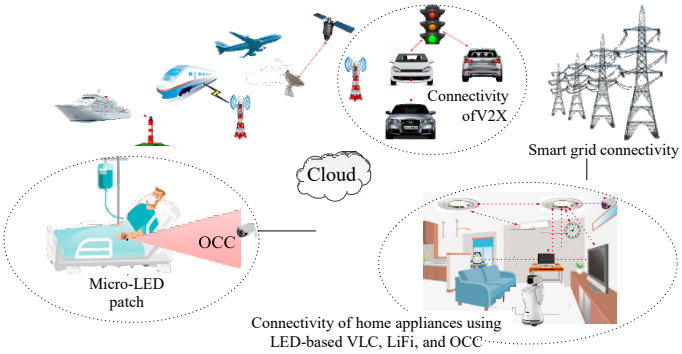


Figure 6: Several instances exemplify extensive connectivity facilitated by OWC technologies.

The IoT networks are characterized by several crucial requirements, including low device cost, low power consumption, economical deployment, heightened energy efficiency, robust security and privacy measures, and the ability to accommodate a large number of devices. LED-based OWC systems encompass all the essential features necessary to support the diverse needs of the IoT. Presently, key technologies employed for IoT connectivity include Zigbee, Bluetooth Low Energy (BLE), and WiFi. Zigbee, recognized for its cost-effectiveness and low-power attributes, serves as a prevalent wireless mesh network standard for IoT applications [48]. However, Zigbee faces limitations in terms of transmission rates and security levels, with interference emerging as a concern in densely populated Zigbee networks. BLE, designed as a low-energy variant of Bluetooth for short-range communication, operates in a single-hop topology (piconet) where one master device communicates with several slave nodes, alongside a broadcast group topology featuring an advertiser node broadcasting to multiple scanners [48]. On the other hand, WiFi lacks guaranteed QoS and is susceptible to interference due to shared unlicensed bands with Zigbee, Bluetooth, and various other devices in the Industrial, Scientific, and Medical (ISM) band. In contrast, OWC technologies exhibit superior capabilities in meeting the specific requirements of IoT networks compared to existing wireless technologies. The inherent advantages of OWC systems position them as a robust solution for addressing the multifaceted demands of IoT applications.

Small-cell networks: A highly effective approach for enhancing area spectral efficiency involves reducing the cell size in instances where a limited number of users are served by each cell [49]. In the evolution of communication systems, the third-generation system exclusively featured microcellular networks to support cellular connectivity. The 4G system introduced small-cell and microcell deployments alongside macrocellular networks, while the upcoming 5G system is anticipated to incorporate ultra-dense small-cells in addition to macro cellular networks [50]. The reduction in cell size presents an opportunity to allocate more spectra to each user. The integration of indoor small-cells or femtocells has significantly expanded possibilities in this regard. An indoor small-cell typically has a cell radius of around 10 meters, catering to five to six users [51]. This deployment strategy proves cost-effective and energy-efficient, meeting coverage and capacity requirements [52]. With the anticipated peak user data rates reaching 10 Gbps for 5G and 1 Tbps for 6G communication systems, the management of heavy data traffic, particularly generated indoors, becomes crucial. Consequently, the deployment of highly dense small-cell networks emerges as a key characteristic of 5G communication systems. Indoor VLC and LiFi technologies contribute to the creation of highly dense small cells. Each network formed under a single light source is considered a small cell, and in large indoor spaces, hundreds of VLC/LiFi-based small cells can be established. Therefore, OWC networks align with the criteria essential for the development and success of 5G/6G networks.

Seamless movement: Seamless mobility is a pivotal requirement for the incorporation of any technology into the 5G networks. The LiFi system stands out by providing comprehensive support for mobility, addressing the demands of both 5G and anticipated 6G communication systems. High-capacity backhaul networks play a crucial role in connecting the access network to the core network. Presently, backhaul networks predominantly utilize dedicated fiber, copper, microwave, mm Wave, and occasionally satellite links [8,53]. Satellite links for backhaul connectivity are contingent on alternative options. In the context of 5G systems, a high-capacity backhaul network is indispensable for facilitating the exchange of substantial data traffic between the access and core networks. Without a robust high-capacity backhaul network, even if the access networks support Gbps communication links to user equipment, the communication system remains incomplete, with a low-capacity backhaul network posing a potential bottleneck. To address this challenge, optical wireless networks, such as FSO systems, alongside wired optical fiber networks, present effective solutions. FSO systems exhibit remarkable features for establishing high-capacity, long-range outdoor backhaul links. Figure 7 illustrates Establishment of high-capacity backhaul connectivity for a ship connectivity, space communications, cellular BS & remote connectivity. FSO technology can also establish high-quality connectivity with Macrocellular Base Stations (MBSs), offering an alternative to existing backhaul network technologies. A comparative analysis in Table 3 highlights the achieved data rates and latencies of key backhaul technologies. While optical fiber currently boasts the highest throughput, FSO systems

demonstrate comparable throughput. Given their similar transmitter and receiver architecture, FSO systems have the potential to achieve throughput levels similar to optical fiber systems in the near future. Latency, calculated for transmission during backhaul connectivity, underscores the FSO network's potential as a valuable complementary solution to wired, microwave, and mmWave systems, supporting high-data-rate communications in 5G and 6G networks.

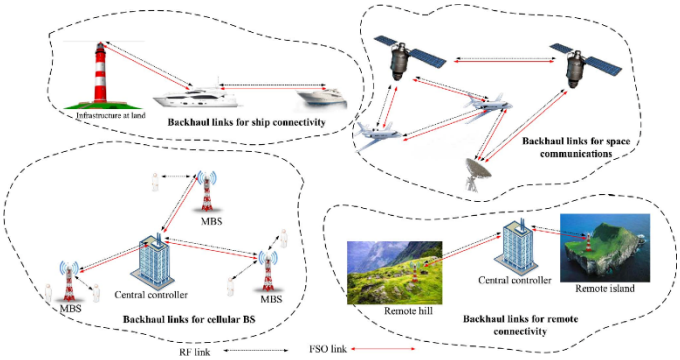


Figure 7: Establishing high-capacity backhaul connectivity for a ship connectivity, space communications, cellular BS & remote connectivity.

Table 4: Comparison of the achieved data rates and latencies in the existing important backhaul technologies [2,9,54]

Backhaul Technology	**Achieved Data Rates**	**Latency**
Optical Fiber	Highest throughput	Low
Copper	Moderate	Moderate
Microwave	Moderate	Moderate
mm Wave	Moderate	Moderate
Satellite	Varies	High

Green Communication: The realization of green communication in future 5G/6G and IoT networks relies on various factors, including energy-conscious network deployment, the selection of communication devices, and the design of communication network protocols. Achieving environmental sustainability necessitates energy-efficient communication methods, a goal that can be effectively met through the increased utilization of LED-based OWC technologies. OWC technologies are poised to handle a substantial portion of the overall wireless data volume, leading to significant energy savings when employed for indoor communication. By leveraging LEDs, which serve dual purposes as both communication devices and illumination sources, OWC networks have the potential to contribute significantly to energy efficiency. Furthermore, OWC systems can play a role in energy harvesting (EH), as demonstrated by integrating solar cells into VLC links. This integration allows solar cells to function not only as energy harvesters but also as optical receivers [55]. Consequently, OWC systems play a pivotal role in establishing environmentally conscious communication systems, a key characteristic integral to the development of 5G/6G and IoT networks.

Tactile internet support: The International Telecommunication Union characterizes the Tactile Internet as the forthcoming internet infrastructure merging ultra-low latency with exceptionally high levels of availability, reliability, and security. Representing the next phase of evolution for the IoT, the Tactile Internet will extend its scope to include interactions between humans and machines, as well as machine-to-machine interactions [56–59]. OWC technologies possess the capability to underpin the Tactile Internet. In a prior study [60], we introduced Human Bond Communication (HBC) as a concept facilitating continuous bidirectional communication among multiple users.

Intelligent transportation: Vehicular communication stands as a pivotal component of the modern era, promising pervasive connectivity with ultra-reliable and low-latency features [61]. V2X communications play a vital role in enhancing road safety, optimizing traffic efficiency, and ensuring the availability of infotainment services [62]. The Dedicated Short-Range Communication (DSRC) technology, operating in the 5.9 GHz band, is extensively utilized for supporting V2X communications, particularly in applications focused on vehicular safety [63]. In addition to DSRC, millimeter-wave (mm Wave) bands have gained prominence in V2X communications due to their ability to deliver Gigabits per second data rates, surpassing the capabilities of DSRC [63]. Furthermore, OWC technologies emerge as a promising option for ensuring reliable connectivity in LOS conditions. Specifically, VLC and LiFi can facilitate short-distance inter-vehicle communications, while OCC extends support for communication over a distance of 60 meters [64]. FSO communication, utilizing laser-based technology, offers the potential for even longer-distance communication in vehicular scenarios.

4.4. Surveys of OWC-Based 5G/6G and IoT Systems

Numerous researchers globally are actively engaged in exploring OWC for the development of future communication networks. An innovative approach introduced in [60] focuses on HBC utilizing head-mounted displays (HMDs). This method employs the camera of an HMD as a receiver and incorporates an IR light source as a transmitter, demonstrating the feasibility of HMDs for communication purposes. This HBC system enables efficient communication between users or devices using their respective HMDs. In [65], researchers propose an optical V2V communication system based on LED transmitters and camera receivers. This technology has the potential to emerge as a significant development for the Internet of Vehicles, where the LED transmitter in one vehicle communicates various data to camera receivers in

other vehicles using an optical communication image sensor. The LiFi/WiFi-integrated architecture presented in [66] is designed to meet the requirements of the 5G system, showcasing a comprehensive integration of LiFi and WiFi. A universal traffic management system detailed in [67] provides expressway and road information to vehicles. This system utilizes LED headlights as transmitters for the uplink and multiple PDs with lenses as receivers on the roadside. For the downlink, signals are transmitted from an LED on a roadside unit and received using an optical communication image sensor on the vehicle. Motivating factors for VLC usage in supporting highly dense users are discussed in [68]. The study explores VLC integration with RF technologies, emphasizing the importance of selecting suitable operating conditions for optimal outcomes in both RF and VLC solutions. Additionally, [69] introduces a relay-assisted VLC system where an amplify-and-forward relay is employed to forward signals while simultaneously transmitting its own signals. The relay terminal assists the source terminal in forwarding signals to the destination terminal, with signal allocation to even and odd subcarriers for source and relay terminals, respectively.

The integration of 5G New Radio (NR) with VLC downlink architecture is elucidated by the authors in 9-[70], showcasing the synergistic potential of these emerging wireless technologies. Specifically, the transmission of 5G NR frames over VLC is meticulously implemented, marking a significant stride in bridging these complementary technologies. Furthermore, in the context of a three-dimensional hybrid RF/VLC indoor IoT system described in [71], a homogeneous Poisson point process is employed to model terminal distribution. This study incorporates a light energy harvesting (EH) model alongside a LOS propagation model for VLC, enabling efficient energy utilization. Notably, the harvested energy from PDs at each device within the room is leveraged for transmissions over the RF uplink. The paper underscores pivotal advancements in OWC technologies, addressing future demands posed by 5G, 6G, and IoT systems, an aspect largely unexplored in existing review literature. By comprehensively examining various OWC technologies, the article delineates their potential contributions towards realizing the objectives of next-generation wireless systems.

5. Challenges of the OWC in the 5G/6G and IoT Solutions

Successfully deploying OWC technologies for 5G/6G and IoT solutions necessitates adeptly tackling a range of formidable challenges. Several critical challenges are succinctly examined below:

Frequent handover: Prospective communication systems will be characterized by heterogeneous small dense networks, leading to frequent handovers. These handovers will occur both within optical networks and between optical and RF networks. Given the diminutive size of optical cells, the likelihood of numerous superfluous handovers exists, necessitating the mitigation of unnecessary handovers and the associated ping-pong effect. Additionally, the distinctive properties of the physical and data-link layers in optical and RF-based wireless networks pose a significant challenge for ensuring effective mobility support in RF/optical hybrid systems.

Inter-cell interference: Effectively addressing the management of inter-cell optical interference emerges as a critical concern during the deployment of optical VLC and LiFi networks. The dense deployment of LEDs in OWC technologies has the potential to induce substantial interference within 5G/6G and IoT networks. Consequently, the mitigation of inter-cell optical interference stands out as a formidable challenge in this context.

Atmospheric loss: The efficacy of OWC technologies is susceptible to various atmospheric factors such as scattering, refraction, air absorption, free space loss, and scintillation. Outdoor settings introduce additional challenges, as fog and dust impede the transmission of optical signals from the transmitter to the receiver. Unfavorable atmospheric conditions contribute to degradation in the communication link quality for FSO. Consequently, mitigating atmospheric losses poses a considerable challenge, particularly in outdoor environments, in striving to achieve the objectives of 5G networks.

Limited uplink communication using OWC technologies: Many user equipment designs incorporate low-power LEDs to minimize power consumption. However, this presents challenges for VLC and LiFi systems in uplink communication. The use of low-power LEDs results in diffused, low-intensity light, making them susceptible to interference from downlink high-power lights and thereby constraining uplink communication performance. Additionally, the vulnerability of the uplink communication link is heightened by the slightest deflection or movement of the user equipment's receiver. Addressing these issues is crucial for the future enhancement of VLC and LiFi systems to efficiently support uplink communication.

Low data rate of the OCC system: A significant limitation of the current OCC system is its constrained data rate, primarily attributed to the low-frame rate cameras employed. Achieving a high data rate is challenging within this framework, as evidenced by the most recent recorded data rate of only 55 Mbps [27]. There is a pressing need to augment this data rate to meet the

burgeoning service requirements in the context of 5G/6G and IoT networks.

Flickering avoidance: Flickering refers to variations in the luminance of light perceivable by humans, posing a significant concern in OWC systems. Various modulation schemes employed in OWC systems may induce flickering, which can adversely impact human health. Effectively addressing this challenge involves modulating LEDs in a manner that mitigates flickering, adding a layer of complexity to the task.

Data rate improvement of the FSO backhaul system: The backhaul infrastructure within 5G/6G systems is tasked with managing a substantial volume of data traffic to facilitate high-data-rate services for end-users. Failure to address this efficiently may lead to bottleneck issues. Consequently, the challenging endeavor involves enhancing FSO backhaul capacity in response to the escalating traffic volumes.

Machine learning for OWC: The future landscape of 6G communication networks necessitates the incorporation of learning-based networking systems as a key requirement. Given the escalating complexity of network structures and diverse requirements, artificial control and decision-making become imperative in challenging environments. Supervised learning finds application in various OWC-based scenarios such as smart healthcare [72], smart home lighting [73], and OWC data mining. Unsupervised machine learning methods prove efficient for OWC data-based analysis, encompassing tasks such as correlation, ranking, spatial and temporal analysis, and flow prediction. Furthermore, reinforcement learning emerges as a valuable tool for optimizing data rates, implementing network switching, and managing traffic within ultra-dense OWC networks designed for 6G [14]. The integration of machine learning into 6G OWC networks facilitates intelligent network assignment, automated error correction, efficient decision-making, and network reassignment, among other functionalities. Notably, the application of the machine learning approach is integral in the context of indoor mobile robot-based dense OWC small networks, enabling swift and efficient task execution.

6. Conclusion

The introduction of 5G communication is anticipated to occur by 2020, followed by the projected launch of 6G communication between 2027 and 2030. Realizing the objectives of 5G/6G and the IoT through the tactile internet poses several challenges. Key among these challenges are the provision of high capacity, massive connectivity, low latency, high security, low-energy consumption, high QoE, and highly reliable connectivity for 5G communication systems. Solely relying on RF-based systems proves insufficient to meet the substantial demands of future 5G/6G and IoT networks. OWC technologies, including VLC, LiFi, OCC, and FSO communication, emerge as ideal complementary solutions to RF networks. The concurrent operation of RF and optical wireless systems holds the potential to achieve the ambitious goals set for these networks. This study provides a comprehensive examination of how OWC technologies contribute to the successful deployment of future 5G/6G and IoT networks. The characteristics of 5G, 6G, and IoT systems, as well as the features of OWC technologies such as VLC, LiFi, OCC, and FSO, are succinctly outlined. Each specification of 5G, 6G, and IoT is individually expounded upon, highlighting how OWC systems facilitate the realization of these features. Additionally, the paper offers a summary of existing OWC-related studies pertaining to 5G and IoT, making it a valuable resource for comprehending research contributions across various optical wireless systems in the context of future network deployment.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Z. Ghassemlooy, S. Arnon, M. Uysal, Z. Xu, J. Cheng, "Emerging optical wireless communications-advances and challenges," IEEE Journal on Selected Areas in Communications, vol. 33, no. 9, 1738–1749, 2015, doi:10.1109/JSAC.2015.2430821.

[2] M.Z. Chowdhury, M.T. Hossan, A. Islam, Y.M. Jang, "A comparative survey of optical wireless technologies: Architectures and applications," IEEE Access, vol. 6, 9819–10220, 2018, doi:10.1109/ACCESS.2018.2799852.

[3] M. Uysal, H. Nouri, "Optical wireless communications—An emerging technology," in Proceedings of the International Conference on Transparent Optical Networks, 2014, doi:10.1109/ICTON.2014.6876505.

[4] Z. Xu, R.B.M. Sadler, "Ultraviolet communications: Potential and state-of-the-art," IEEE Communications Magazine, vol. 46, no. 7, 67–73, 2008, doi:10.1109/MCOM.2008.4569743.

[5] J.B. Carruthers, Wireless Infrared Communications, Wiley Encyclopedia of Telecommunications, 2003.

[6] P.H. Pathak, X. Feng, P. Hu, P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," IEEE Communications Surveys \& Tutorials, vol. 17, no. 4, 2047–2077, 2015, doi:10.1109/COMST.2015.2444095.

[7] M. Shafi, A.F. Molisch, P.J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," IEEE Journal on Selected Areas in Communications, vol. 35, no. 6, 1201–1221, 2017, doi:10.1109/JSAC.2017.2695280.

[8] M. Jaber, M.A. Imran, R. Tafazolli, A. Tukmanov, "5G backhaul challenges and emerging research directions: A survey," IEEE

Access, vol. 4, 1143–1166, 2016, doi:10.1109/ACCESS.2016.2546541.

[9] J.G. Andrews, S. Buzzi, W. Choi, S. V Hanly, A. Lozano, A.C. Soong, J.C. Zhang, "What will 5G be?" IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, 1065–1082, 2014, doi:10.1109/JSAC.2014.2328098.

[10] W.A. Hassan, H.-S. Jo, T.A. Rahman, "The feasibility of coexistence between 5G and existing services in the IMT-2020 candidate bands in Malaysia," IEEE Access, vol. 5, 14867–14888, 2017, doi:10.1109/ACCESS.2017.2710159.

[11] A. Ijaz, L. Zhang, M. Grau, A. Mohamed, S. Vural, A.U. Quddus, M.A. Imran, C.H. Foh, R. Tafazolli, "Enabling massive IoT in 5G and beyond systems: PHY radio frame design considerations," IEEE Access, vol. 4, 3322–3339, 2016, doi:10.1109/ACCESS.2016.2546541.

[12] K. David, H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" IEEE Vehicular Technology Magazine, vol. 13, no. 3, 72–80, 2018, doi:10.1109/MVT.2018.2816618.

[13] F. Tariq, M. Khandaker, K.K. Wong, M. Imran, M. Bennis, M. Debbah, "A speculative study on 6G," ArXiv Preprint, 2019.

[14] S.J. Nawaz, S.K. Sharma, S. Wyne, M.N. Patwary, M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," IEEE Access, vol. 7, 46317–46350, 2019, doi:10.1109/ACCESS.2019.2900597.

[15] R.A. Stoica, G.T.F. Abreu, "6G: The wireless communications network for collaborative and AI applications," ArXiv Preprint ArXiv:1904.03413, 2019.

[16] W. Saad, M. Bennis, M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," ArXiv Preprint ArXiv:1902.10265, 2019.

[17] T. Nguyen, M.Z. Chowdhury, Y.M. Jang, "A novel link switching scheme using pre-scanning and RSS prediction in visible light communication networks," EURASIP Journal on Wireless Communications and Networking, vol. 2013, 1–17, 2013, doi:10.1186/1687-1499-2013-1.

[18] D. Tsonev, S. Videv, H. Haas, "Towards a 100 Gb/s visible light wireless access network," Optics Express, vol. 23, no. 2, 1627–1637, 2015, doi:10.1364/OE.23.001627.

[19] M.Z. Chowdhury, M.T. Hossan, M.K. Hasan, Y.M. Jang, "Integrated RF/optical wireless networks for improving QoS in indoor and transportation applications," Wireless Personal Communications, vol. 107, no. 3, 1401–1430, 2018, doi:10.1007/s11277-018-5482-8.

[20] H. Haas, L. Yin, Y. Wang, C. Chen, "What is LiFi?" Journal of Lightwave Technology, vol. 34, no. 6, 1533–1544, 2016, doi:10.1109/JLT.2016.2525829.

[21] S. Dimitrov, H. Haas, Principles of LED Light Communications: Towards Networked Li-Fi, Cambridge University Press, 2015.

[22] H.H. Lu, C.Y. Li, H.W. Chen, C.M. Ho, M.T. Cheng, Z.Y. Yang, C.K. Lu, "A 56 Gb/s PAM4 VCSEL-based LiFi transmission with two-stage injection-locked technique," IEEE Photonics Journal, vol. 9, 1–8, 2017, doi:10.1109/JPHOT.2017.2696500.

[23] M.K. Hasan, M.Z. Chowdhury, M. Shahjalal, Y.M. Jang, "Fuzzy based network assignment and link-switching analysis in hybrid OCC/LiFi system," Wireless Communications and Mobile Computing, 2018, doi:10.1155/2018/5703580.

[24] M. Hossan, M.Z. Chowdhury, M. Hasan, M. Shahjalal, T. Nguyen, N.T. Le, Y.M. Jang, "A new vehicle localization scheme based on combined optical camera communication and photogrammetry," Mobile Information Systems, 2018, doi:10.1155/2018/4609192.

[25] M. Shahjalal, M. Hossan, M. Hasan, M.Z. Chowdhury, N.T. Le, Y.M. Jang, "An implementation approach and performance analysis of image sensor based multilateral indoor localization and navigation system," Wireless Communications and Mobile Computing, vol. 2018, 2018, doi:10.1155/2018/5703580.

[26] Z. Ghassemlooy, P. Luo, S. Zvanovec, Optical camera communications, Springer: 547–568, 2016, doi:10.1007/978-3-319-48872-5_20.

[27] Y. Goto, I. Takai, T. Yamazato, H. Okada, T. Fujii, S. Kawahito, S. Arai, T. Yendo, K. Kamakura, "A new automotive VLC system using optical communication image sensor," IEEE Photonics Journal, vol. 8, 1–17, 2016, doi:10.1109/JPHOT.2016.2593718.

[28] A. Malik, P. Singh, "Free space optics: Current applications and future challenges," International Journal of Optics, vol. 2015, , 2015, doi:10.1155/2015/146591.

[29] M.A. Khalighi, M. Uysal, "Survey on free space optical communication: A communication theory perspective," IEEE Communications Surveys \& Tutorials, vol. 16, 2231–2258, 2014, doi:10.1109/SURV.2014.012214.00189.

[30] H. Kaushal, G. Kaddoum, "Optical communication in space: Challenges and mitigation techniques," IEEE Communications Surveys \& Tutorials, vol. 19, 57–97, 2017, doi:10.1109/COMST.2016.2618498.

[31] S. Mumtaz, J.M. Jornet, J. Aulin, W.H. Gerstacker, X. Dong, B. Ai, "Terahertz communication for vehicular networks," IEEE Transactions on Vehicular Technology, vol. 66, 5617–5625, 2017, doi:10.1109/TVT.2016.2639738.

[32] L. Lovén, T. Leppänen, E. Peltonen, J. Partala, E. Harjula, P. Porambage, M. Ylianttila, J. Riekki, "Edge AI: A vision for distributed, edge-native artificial intelligence in future 6G networks," in Proceedings of the 6G Wireless Summit, Levi, Finland, 2019.

[33] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic, P. Popovski, "From 5G to 6G: Has the time for modern random access come?" ArXiv Preprint ArXiv:1903.03063, 2019.

[34] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, M. Zorzi, "Towards 6G networks: Use cases and technologies," ArXiv Preprint ArXiv:1903.12216, 2019.

[35] 5G Requirements.

[36] Light Communications for Wireless Local Area Networking.

[37] D. Menaka, C.T. Sabitha Gauni, Manimegalai, K. Kalimuthu, "Vision of IoUT: advances and future trends in optical wireless communication," Journal of Optics, vol. 50, 439–452, 2021.

[38] P. Hu, P.H. Pathak, A.K. Das, Z. Yang, P. Mohapatra, "PLiFi: Hybrid WiFi-VLC networking using power lines," in Proceedings of the Workshop on Visible Light Communication Systems, New York, NY, USA, 2016.

[39] Z. Du, C. Wang, Y. Sun, G. Wu, "Context-aware indoor VLC/RF heterogeneous network selection: Reinforcement learning with knowledge transfer," IEEE Access, vol. 6, 33275–33284, 2018, doi:10.1109/ACCESS.2018.2847723.

[40] T. Koonen, "Indoor optical wireless systems: Technology, trends, and applications," Journal of Lightwave Technology, vol. 36, no. 7, 1459–1467, 2018, doi:10.1109/JLT.2018.2791740.

[41] C. Danakis, M. Afgani, G. Povey, I. Underwood, H. Haas, "Using a CMOS camera sensor for visible light communication," in Proceedings of the IEEE Globecom Workshops, Anaheim, CA, USA, 2012.

[42] H.M. Tsai, H.M. Lin, H.Y. Lee, "Demo: Rollinglight-universal camera communications for single LED," in Proceedings of the International Conference on Mobile Computing and Networking, Maui, HI, USA, 2014.

[43] L.Y. Wei, C.W. Chow, G.H. Chen, Y. Liu, C.H. Yeh, C.W. Hsu, "Tricolor visible-light laser diodes based visible light communication operated at 40.665 Gbit/s and 2 m free-space transmission," Optics Express, vol. 27, no. 18, 25072–25077, 2019, doi:10.1364/OE.27.025072.

[44] M.Z. Chowdhury, M. Shahjalal, M.K. Hasan, Y.M. Jang, "The Role of Optical Wireless Communication Technologies in 5G/6G and IoT Solutions: Prospects, Directions, and Challenges," Applied Sciences, vol. 9, 2019.

[45] C. Chang, "A 100-Gb/s multiple-input multiple-output visible laser light communication system," Journal of Lightwave Technology, vol. 32, no. 22, 4723–4729, 2014, doi:10.1109/JLT.2014.2361215.

[46] D. Zhang, Z. Zhou, S. Mumtaz, J. Rodriguez, T. Sato, "One integrated energy efficiency proposal for 5G IoT communications," IEEE Internet of Things Journal, vol. 3, 1346–1354, 2016, doi:10.1109/JIOT.2016.2593778.

[47] Z. Yan, O. Zhang, A. V Vasilakos, "A survey on trust management for internet of things," Journal of Network and Computer Applications, vol. 42, 120–134, 2014, doi: 10.1016/j.jnca.2014.01.012.

[48] M.R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, 510–527, 2016, doi:10.1109/JSAC.2016.2525478.

[49] X. Ge, J. Yang, H. Gharavi, Y. Sun, "Energy efficiency challenges of 5G small cell networks," IEEE Communications Magazine, vol. 55, no. 1, 184–191, 2017, doi:10.1109/MCOM.2017.1600493.

[50] Y. Hao, M. Chen, L. Hu, J. Song, M. Volk, I. Humar, "Wireless fractal ultra-dense cellular networks," Sensors, vol. 17, no. 4, 841, 2017, doi:10.3390/s17040841.

[51] M.Z. Chowdhury, Y.M. Jang, Z.J. Haas, "Cost-effective frequency planning for capacity enhancement of femtocellular networks," Wireless Personal Communications, vol. 60, no. 1, 83–104, 2011, doi:10.1007/s11277-010-9987-2.

[52] H.A.U. Mustafa, M.A. Imran, M.Z. Shakir, A. Imran, R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," IEEE Communications Surveys \& Tutorials, vol. 18, 419–445, 2016, doi:10.1109/ACCESS.2018.2847723.

[53] X. Artiga, A. Perez-Neira, J. Baranda, E. Lagunas, S. Chatzinotas, R. Zetik, P. Gorski, K. Ntougias, D. Perez, G. Ziaragkas, "Shared access satellite-terrestrial reconfigurable backhaul network enabled by smart antennas at mmWave band," IEEE Network, vol. 32, no. 5, 46–53, 2018, doi:10.1109/MNET.2018.1700241.

[54] F. Knobloch, "Delay analysis for optical wireless multihop networks," in Proceedings of the International Conference on Transparent Optical Networks (ICTON), Trento, Italy, 2016.

[55] S. Zhang, D. Tsonev, S. Videv, S. Ghosh, G.A. Turnbull, I.D.W. Samuel, H. Haas, "Organic solar cells as high-speed data detectors for visible light communication," Optica, vol. 2, no. 7, 607–610, 2015, doi:10.1364/OPTICA.2.000607.

[56] What is The Tactile Internet? Available online: https://5g.co.uk/guides/what-is-the-tactile-internet (accessed on 15 August 2019).

[57] G. Fettweis, "The tactile internet: Applications and challenges," IEEE Vehicular Technology Magazine, vol. 9, no. 1, 64–70, 2014, doi:10.1109/MVT.2013.2293016.

[58] A. Aijaz, M. Dohler, A.H. Aghvami, V. Friderikos, M. Frodigh, "Realizing the tactile internet: Haptic communications over next generation 5G cellular networks," IEEE Wireless Communications, vol. 24, no. 1, 82–89, 2017, doi:10.1109/MWC.2017.1600493.

[59] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, G. Fettweis, "5G-enabled tactile internet," IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, 460–473, 2016, doi:10.1109/JSAC.2016.2525478.

[60] M.T. Hossan, M.Z. Chowdhury, M. Shahjalal, Y.M. Jang, "Human bond communication with head-mounted displays: Scope, challenges, solutions, and applications," IEEE Communications Magazine, vol. 57, no. 2, 26–32, 2019, doi:10.1109/MCOM.2019.1800421.

[61] S.A.A. Shah, E. Ahmed, M. Imran, S. Zeadally, "5G for vehicular communications," IEEE Communications Magazine, vol. 56, no. 11, 111–117, 2018, doi:10.1109/MCOM.2018.1700241.

[62] S. Chen, J. Hu, Y. Shi, Y. Peng, J. Fang, R. Zhao, L. Zhao, "Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G," IEEE Communications Standards Magazine, vol. 1, no. 1, 70–76, 2017, doi:10.1109/MCOMSTD.2017.1700015.

[63] T.S. Rappaport, Y. Xing, G.R. MacCartney, A.F. Molisch, E. Mellios, J. Zhang, "Overview of millimeter wave communications for fifth generation (5G) wireless networks—with a focus on propagation models," IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, 6213–6230, 2017, doi:10.1109/TAP.2017.2734243.

[64] P. Luo, M. Zhang, Z. Ghassemlooy, H. Le Minh, H.M. Tsai, X. Tang, L.C. Png, D. Han, "Experimental demonstration of RGB LED-based optical camera communications," IEEE Photonics Journal, vol. 7, –12, 2015, doi:10.1109/JPHOT.2015.2457321.

[65] I. Takai, T. Harada, M. Andoh, K. Yasutomi, K. Kagawa, S. Kawahito, "Optical vehicle-to-vehicle communication system using LED transmitter and camera receiver," IEEE Photonics Journal, vol. 6, 1–14, 2014, doi:10.1109/JPHOT.2014.2361215.

[66] M. Ayyash, H. Elgala, A. Khreishah, V. Jungnickel, T. Little, S. Shao, M. Rahaim, D. Schulz, J. Hilt, R. Freund, "Coexistence of WiFi and LiFi toward 5G: Concepts, opportunities, and challenges," IEEE Communications Magazine, vol. 54, no. 2, 64–71, 2016, doi:10.1109/MCOM.2016.7432165.

[67] T. Yamazato, N. Kawagita, H. Okada, T. Fujii, T. Yendo, S. Arai, K. Kamakura, "The uplink visible light communication beacon system for universal traffic management," IEEE Access, vol. 5, 22282–22290, 2017, doi:10.1109/ACCESS.2017.2766759.

[68] M.B. Rahaim, T.D.C. Little, "Toward practical integration of dual-use VLC within 5G networks," IEEE Wireless Communications, vol. 22, 97–103, 2015, doi:10.1109/MWC.2015.7124872.

[69] Z. Na, Y. Wang, M. Xiong, X. Liu, J. Xia, "Modeling and throughput analysis of an ADO-OFDM based relay-assisted

VLC system for 5G networks," IEEE Access, vol. 6, 17586–17594, 2018, doi:10.1109/ACCESS.2018.2808258.

[70] L. Shi, W. Li, X. Zhang, Y. Zhang, G. Chen, A. Vladimirescu, "Experimental 5G new radio integration with VLC," in Proceedings of the IEEE International Conference on Electronics, Circuits and Systems (ICECS), 1–4, 2018, doi:10.1109/ICECS.2018.8611640.

[71] G. Pan, H. Lei, Z. Ding, Q. Ni, "3-D Hybrid VLC-RF indoor IoT systems with light energy harvesting," IEEE Transactions on Green Communications and Networking, vol. 3, 853–865, 2019, doi:10.1109/TGCN.2019.2908839.

[72] M.K. Hasan, M. Shahjalal, M.Z. Chowdhury, Y.M. Jang, "Real-time healthcare data transmission for remote patient monitoring in patch-based hybrid OCC/BLE networks," Sensors, vol. 19, 1208, 2019, doi:10.3390/s19051208.

[73] K.P. Pujapanda, "LiFi Integrated to power-lines for smart illumination cum communication," in Proceedings of the International Conference on Communication Systems and Network Technologies, 1–4, 2013.

RAMSHA KHALID has done her bachelor's degree from Lahore College for Women University, Lahore in 2018. She has done her master's degree from University of Lahore, Lahore in 2022.
Her area of interest includes Computer & Communication Networks, Machine Learning, Artificial Intelligence, Cyber Security, Control Systems and Renewable Energy Systems. Recently she has publish a conference paper in IEEE INMIC'23 held in University of Central Punjab, Lahore as a first author.

M. NAQI RAZA has done his bachelor's degree from University of Gujrat, Gujrat in 2018. Currently, he is doing his master's degree from University of Sialkot, Sialkot.
His area of interest includes Power Generation (Conventional and Renewable), Wind Power Generation and Utilization, Optimization of Wind Energy, Solar Energy, Solar Power Applications, Electric Vehicles (PHEVs).

Received: 30 April, 2024, Revised: 20 June, 2024, Accepted: 21 May, 2024, Online: 21 June, 2024

DOI: https://doi.org/10.55708/js0306003

Educational Applications and Comparative Analysis of Network Simulators: Protocols, Types, and Performance Evaluation

Nikolaos V. Oikonomou [1], Dimitrios V . Oikonomou [2]

[1] Department of Informatics & Telecommunications, University of Ioannina, Arta,47150, Greece

[2]Department of Regional & Cross Border Studies, University of Western Macedonia, Kozani,50100, Greece

*Corresponding author: Nikolaos V. Oikonomou, University of Ioannina Department of Informatics & Telecommunications, haikos13@gmail.com

ABSTRACT: This work explores the role of simulation in computer networks, discussing various network types, communication protocols, and the utilization of network simulators, with a focus on educational settings. We specifically analyze and compare five prominent network simulators: Cisco Packet Tracer, Riverbed Modeler Academic Edition, GNS3, NS-3, and Mininet. These tools are examined in terms of their functionality, user-friendliness, and suitability for educational purposes, assessing how they facilitate learning for students and trainees. The comparison extends to their operational capabilities, differences, effectiveness, and overall impact on networking education. The evaluation aims to highlight each simulator's strengths and weaknesses, providing insights into their practical applications in an academic context.

KEYWORDS: Network Architecture, Network Simulation, Network Protocol, Emulator, Educational Technology, Performance Evaluation

1. Introduction

In today's interconnected world, networks encompass nearly every aspect of our global society, forming a vast web that spans the entire Earth. This extensive coverage is not just limited to geographic expansiveness but also permeates various sectors of human activity, ranging from economic operations and governmental infrastructure to personal communication and entertainment. Networks facilitate the seamless flow of information, enabling not just global connectivity but also driving advancements in technology and society. A computer network, by definition, is a system comprising both autonomous and non-autonomous computers or nodes interconnected through various means of communication. These networks are fundamental to the modern digital ecosystem, supporting an extensive array of devices beyond traditional computers, including mobile phones, printers, cameras, televisions, and even more sophisticated IoT (Internet of Things) devices like smart thermostats and security systems. The term "computer" in this context serves a formal role, recognizing any device capable of sending, receiving, and processing data as part of the network infrastructure. As networks have become ubiquitous, the functions and effectiveness of network simulations have similarly evolved to become indispensable tools in network design and management. These simulations are employed extensively across a spectrum of applications—from crafting robust architectures for large organizations and services to developing state defense mechanisms. They play a pivotal role in telecommunications, where they help in optimizing network performance and security under various scenarios without the need to physically alter the network during testing. Moreover, the role of network simulations extends into the realm of education, where they provide a practical learning experience for students and professionals alike. Through simulations, learners can explore complex network dynamics and interactions in a controlled environment, enhancing their understanding of network management, problem-solving, and strategic planning. These simulations are also crucial in the design and implementation of networks. They allow engineers and network designers to experiment with network configurations, simulate loads and attacks, and foresee how a network might behave

under stress or failure conditions. The predictive capabilities of network simulations help in preempting problems and designing networks that are both resilient and scalable. In summary, the development of network technology and simulations reflects our growing dependence on digital connectivity and the continuous need for advancements in network reliability, security, and efficiency. As we look towards future innovations—such as 5G networks, enhanced broadband capabilities, and more sophisticated cybersecurity measures—the role of network simulations will only grow in importance, shaping the backbone of our digital world and ensuring that networks not only cover the Earth geographically but also meet the evolving demands of a highly connected future [1].

2. Types of Networks

Networks are divided into wired and wireless in terms of their connectivity. Wired networks are defined as those that communicate with each other through physical means, namely through networking cables. Wireless networks communicate without any transmission medium such as any cable. In terms of their coverage, we have the 3 basic categories of networks:

Local coverage networks that usually connect devices within the same building or nearby building infrastructure and their coverage range does not exceed a radius of 1 km.

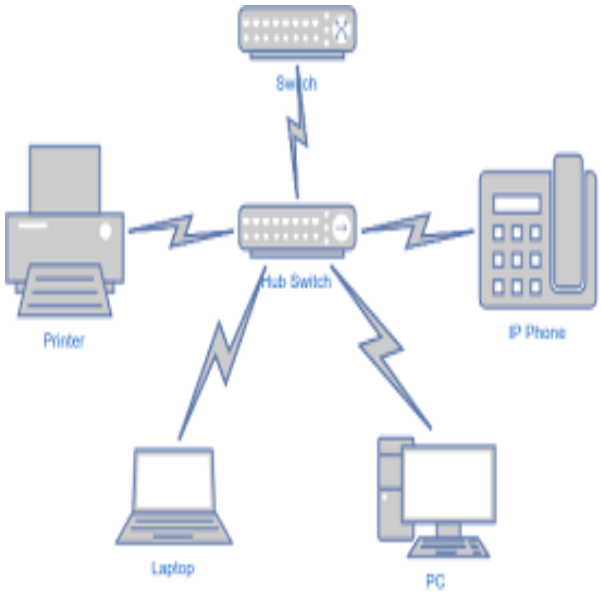


Figure 1: Example of a LAN network in a home or workplace.

Metropolitan area networks consist of many LAN networks together and usually cover the radius of an entire area, a campus, or a city and usually reach a radius of 50km.

Wide area networks provide much larger coverage than LAN & WAN as they are the sum of the above network types, and their coverage is considered unlimited because they cover the entire Earth [2].

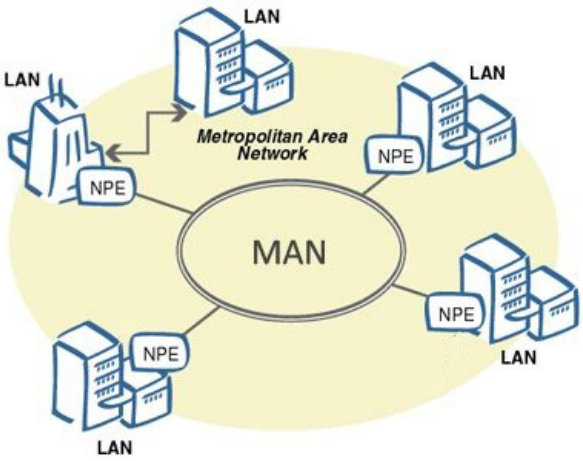


Figure 2: Example of a MAN network in a city.

Figure 3: The internet and global connectivity themselves are the ideal example of a WAN.

2.1. *Network communication protocols*

The most important network communication protocol as the entire internet is based on it. Divided into layers, it can manage all data transfer problems.

A basic data transfer protocol. Unlike TCP/IP, it does not have security functions but supports the transmission of information to multiple users simultaneously. It is mainly used in telecommunications (VoIP) and Online Videogames. FTP is widely used by TCP/IP for sending and receiving files. The 802-protocol family is used in LAN and MAN networks. This protocol allows the changing of IP addresses between computers to achieve accurate data transmission through the correct addressing.

It is the initial information transmission model which started as an idea in 1970 and was formalized in 1984. It divides the entire data transfer process into 7 layers, each undertaking a separate process in the facilitation of information transfer.

Figure 4: The OSI operating levels.

3. Introduction to the Network Simulations

In this chapter, the definition of simulation, its operation, why we perform a network simulation, as well as the types of simulators with their advantages and disadvantages will be discussed. Simulation is called the study of a system with the help of a computer through a numerical experiment modeling technique. Essentially, we input (input) the data we want to study to obtain the desired output (output) results by mimicking the real-world processes or system [3].

3.1. Network Simulation Operation

In the intricate world of network simulations, the operation harmoniously blends information technology and statistical processes. This synergistic integration allows for robust data collection, analysis, and the creation of detailed visual representations such as diagrams and charts. At the heart of these operations are sophisticated programs specifically designed to execute tasks such as sampling, processing data into actionable insights, and graphical representation. These tools are crucial in turning raw data into comprehensible outcomes that can inform strategic decisions and optimize network designs [4].

3.1.1. DES (Discrete Event Simulation)

Discrete Event Simulation (DES) represents a critical methodology in the realm of simulation technologies, where the flow of time is modeled as distinct, individual events. Each event occurs at a particular instant in time and marks a change of state in the system. This approach is widely utilized across various sectors, from complex industrial settings like factories and maritime installations to more streamlined applications such as the operational dynamics of household appliances. For instance, in large-scale industrial applications, DES can simulate the logistical operations of a factory floor, the scheduling and management of maritime ports, or the traffic flow of communication in large call centers. These environments benefit from DES's ability to model intricate systems where components interact at discrete points in time, allowing for detailed analysis of each interaction. On a smaller scale, DES can be applied to everyday applications such as the adaptive control of a radiator's fan speed based on temperature fluctuations or the intensity adjustments of a lamp in response to changes in ambient lighting conditions. These simulations help in refining product designs to enhance functionality and user experience [5].

3.1.2. ABS (Agent-Based Simulation)

Agent-Based Simulations (ABS) provide a dynamic framework where agents, autonomous decision-making entities with defined behaviors, interact within a simulated environment. This type of simulation is particularly suited to scenarios where complex interactions among agents lead to emergent behaviors and outcomes. ABS is extensively used in fields such as healthcare, where it can model the spread of diseases within a population, or in economics, where it can simulate market dynamics and consumer behaviors. Agents in these simulations are designed to mimic real-world behaviors, making ABS an excellent tool for studying social systems and organizational structures. For example, in health sector simulations, agents could represent individuals with varying susceptibility to a disease, allowing researchers to study the impacts of interventions on disease spread [6].

3.1.3. CS (Continuous Simulation)

Continuous Simulation (CS) deals with systems characterized by continuous state changes over time, modeled through differential equations. Unlike DES where changes occur at discrete intervals, CS provides a smooth and continuous description of system dynamics. This type of simulation is indispensable in fields like environmental science for studying climate change impacts or in engineering for assessing the stress and strain on materials over time. CS is particularly effective for simulations that require tracking of variables that change incrementally across every point in time, such as the growth of a population of animals or the dispersion of pollutants in an ecosystem. These simulations are crucial for long-term planning and forecasting in environmental management and urban planning [7].

3.1.4. HS (Hybrid Simulation)

Hybrid Simulations combine the features of both discrete and continuous simulations. This blend allows for the modeling of systems where both continuous and discrete processes exist simultaneously. For example, a

hybrid simulation might model a manufacturing process (a continuous flow of materials) alongside the maintenance schedules of the machinery (discrete events) [8].

3.1.4.1. Continuous Processes

Continuous simulations are used to model systems that change in a smooth, continuous manner over time. In the context of network simulations, this might involve the continuous flow of data through network channels, where parameters like bandwidth, latency, and error rates are modeled as continuous variables.

3.1.4.2. Discrete Events

Discrete event simulations, on the other hand, are used to model systems where changes occur at distinct points in time. In network simulations, this could involve modeling events such as packet arrivals, queue formations, and protocol state transitions.

3.1.4.3. Integration of Continuous and Discrete Models

Hybrid simulations integrate these two approaches to provide a more comprehensive modeling framework. For instance, in a network simulation, continuous models can simulate the overall data flow, while discrete models can handle specific events like packet drops or node failures.

3.1.4.4. Advantages of Hybrid Simulation

The primary advantage of hybrid simulation is its ability to capture the dynamic interactions between continuous processes and discrete events. This approach can provide more accurate and detailed insights into the behavior of complex systems, which is particularly useful in scenarios where both types of processes are significant.

Hybrid simulation techniques are increasingly used in various fields, including manufacturing, logistics, healthcare, and telecommunications. They offer a powerful tool for analyzing and optimizing systems where continuous and discrete dynamics interact.

3.2. Advantages/Disadvantages

The adoption of simulation technologies comes with its set of advantages and challenges. Below, we explore these aspects in detail, providing a comprehensive understanding of the potential benefits and limitations. [9].

3.2.1. Advantages

3.2.1.1. Risk-Free Testing Environment

3.2.1.1.1. Scenario Modeling

The foremost advantage of using simulations is their ability to model and test every conceivable scenario for network implementation in a risk-free, virtual environment. This allows for the examination of network behavior under various conditions without the risk of disrupting actual network operations.

3.2.1.1.2. Cost-Efficiency

Simulations help in identifying the most efficient, cost-effective, and robust network configurations. This pre-implementation testing can save significant costs associated with trial-and-error approaches in real-world deployments.

3.2.1.1.3. Safety

By testing network changes in a simulated environment, potential issues can be identified and resolved before they affect live systems, ensuring network stability and reliability.

3.2.1.2. Flexibility and Scalability

3.2.1.2.1. Scalable Models

Simulation tools can model networks of various sizes and complexities, from small local networks to large-scale global infrastructures. This scalability allows for comprehensive analysis of network performance and potential bottlenecks.

3.2.1.2.2. Customizable Scenarios

Users can customize simulations to match specific requirements, such as testing new protocols, evaluating network upgrades, or assessing security vulnerabilities.

3.2.1.3. Enhanced Understanding and Learning:

3.2.1.3.1. Educational Value

Simulation tools provide a valuable educational resource, allowing students and professionals to visualize and interact with complex network topologies and protocols. This hands-on experience enhances understanding and facilitates learning.

3.2.1.3.2. Predictive Analysis

By simulating future network scenarios, organizations can proactively identify potential issues and implement preventive measures, enhancing overall network resilience.

3.2.1.4. Efficient Resource Allocation:

3.2.1.4.1. Optimization

Simulations can identify the optimal allocation of network resources, such as bandwidth and hardware, ensuring efficient utilization and minimizing wastage.

3.2.1.4.2. Performance Evaluation

They enable the evaluation of network performance under different configurations, helping in making

informed decisions about resource investments and upgrades.

3.2.2. Disadvantages

3.2.2.1. Resource-Intensive Development

3.2.2.1.1. High Costs

The development and operation of advanced network simulators are resource-intensive, requiring significant investment in both computational resources and expert human capital. High-performance computing infrastructure and skilled personnel are essential for creating and managing detailed simulation models.

3.2.2.1.2. Complex Setup

Setting up simulation environments can be complex and time-consuming, particularly for large-scale or highly detailed simulations.

3.2.2.2. Limitations in Realism:

3.2.2.2.1. Approximation of Reality

Despite advances in simulation technology, the inherent unpredictability of real-world environments means that simulations, while highly indicative, cannot completely replicate all real-world variables. Factors such as human behavior, environmental changes, and unexpected network traffic patterns may not be fully captured.

3.2.2.2.2. Validation Required

Consequently, results from simulations should be interpreted as approximations. While they provide valuable insights, these results require careful consideration and validation against real-world data and experiences to ensure their accuracy and applicability.

3.2.2.3. Potential for Over-Reliance:

3.2.2.3.1. Overconfidence in Simulations

There is a risk that organizations may become overly reliant on simulation results, potentially overlooking the importance of real-world testing and validation. Simulations should complement, not replace, empirical testing and field trials.

3.2.2.3.2. Static Models

Simulation models can become outdated if they are not regularly updated to reflect changes in technology, network configurations, and usage patterns. Continuous maintenance is necessary to ensure simulations remain relevant and accurate.

3.2.2.4. Technical Challenges

3.2.2.4.1. Modeling Complexity

Accurately modeling complex networks and protocols can be technically challenging. Simplifications and assumptions made during the modeling process can impact the accuracy of simulation results.

Debugging and Troubleshooting: Identifying and resolving issues within simulation models can be difficult, particularly when dealing with intricate network interactions and behaviors.

In conclusion, while simulation technologies offer numerous advantages, including risk-free testing, cost-efficiency, and enhanced learning opportunities, they also present challenges such as high development costs, limitations in realism, and potential over-reliance. To maximize the benefits of simulations, it is crucial to balance their use with real-world testing and continuously validate simulation results against actual network performance. By doing so, organizations can leverage simulations to improve network design, optimize resource allocation, and enhance overall network resilience.

4. Network Simulation Tools

In this section, we will explore some of the most prevalent network simulation programs available today. These tools are crucial for modeling, analyzing, and optimizing network performance in various environments, ranging from educational settings to complex commercial deployments [10] , [11].

4.1. Opnet (Optical Micro-Networks)

Opnet, now known as Riverbed Modeler, is a discrete event simulation (DES) tool with graphical user interface (GUI) support. It is widely regarded as one of the most comprehensive and powerful network simulators available in the commercial market. Opnet's architecture allows it to be used in several sectors due to its ability to model a wide range of network components and behaviors.

4.1.1. Wireless Communications

Opnet can simulate various wireless technologies, including Wi-Fi, cellular networks, and satellite communications. It provides detailed models for radio frequency (RF) propagation, interference, and mobility patterns, enabling accurate performance analysis of wireless networks.

4.1.2. Wired Communications

The tool supports the simulation of traditional wired networks, including Ethernet, fiber optics, and other physical media. Users can model network topologies, link failures, and traffic patterns to study network performance under different conditions.

4.1.3. Protocols

Opnet includes a vast library of protocol models, covering all layers of the OSI model. This allows users to simulate the behavior of routing protocols, transport protocols (e.g., TCP, UDP), and application-layer protocols, providing insights into protocol interactions and performance.

4.1.4. Queues

The simulator offers detailed queue models, enabling the analysis of queuing behavior in network devices such as routers and switches. This helps in understanding delays, jitter, and packet loss under various traffic loads.

4.1.5. Microprocessors & Complex Hardware Types

Opnet can model the internal behavior of network devices, including the processing power of microprocessors and the performance of complex hardware components. This capability is essential for studying the impact of hardware configurations on network performance [12].

4.2. Cisco Packet Tracer

Cisco Packet Tracer is Cisco's primary network simulation and visualization tool, widely used in educational institutions, academies, and organizations for training and certification purposes. It provides a robust platform for learning, teaching, and testing network concepts and configurations.

4.2.1. Learning and Teaching

Packet Tracer is an integral part of Cisco's Networking Academy curriculum. It allows students to practice network configuration and troubleshooting in a virtual environment, reinforcing theoretical knowledge with hands-on experience.

4.2.2. Testing and Certifications

The tool supports the preparation for Cisco certification exams (e.g., CCNA, CCNP) by providing realistic simulation scenarios that mirror those encountered in real-world networks. This practical training is crucial for developing the skills required to manage and configure Cisco networks.

4.2.3. Features

Packet Tracer includes a variety of features, such as real-time and simulation modes, allowing users to visualize network behavior and performance. It supports a wide range of Cisco devices and protocols, enabling the simulation of complex network topologies and interactions.

4.2.4. Collaboration

The tool offers collaborative features, allowing multiple users to work on the same network simulation simultaneously. This is particularly useful in classroom settings, where instructors and students can interact and share insights in real-time [13].

4.3. GNS3 (Graphical Network Simulator-3)

GNS3 is an open-source network simulator that allows the simulation of complex networks using real network hardware images and virtualization technologies. It is highly popular among network professionals and enthusiasts for its flexibility and powerful features.

4.3.1. Real Hardware Emulation

Unlike other simulators that rely on abstract models, GNS3 uses real Cisco IOS, Juniper Junos, and other network operating system images to emulate actual hardware devices. This provides a highly realistic simulation environment.

4.3.2. Integration with Virtual Machines

GNS3 can integrate with VirtualBox, VMware, and other virtualization platforms, allowing the simulation of virtual machines alongside network devices. This is useful for simulating end-to-end network scenarios, including client-server interactions and multi-tier applications.

4.3.3. Extensibility

The simulator supports a wide range of plugins and third-party tools, such as Wireshark for packet capture and analysis. This extensibility makes GNS3 a versatile tool for network design, testing, and troubleshooting.

4.3.4. Community Support

GNS3 has a vibrant community of users and contributors who provide support, share configurations, and develop new features. This community-driven approach ensures continuous improvement and adaptation to emerging networking technologies. [14]

4.4. NS-3 (Network Simulator 3)

NS-3 is an open-source discrete-event network simulator designed for research and educational purposes. It provides a detailed simulation environment for networking protocols and internet systems.

4.4.1. Research Focus

NS-3 is widely used in academic and research settings to study the performance and behavior of networking protocols. It supports the simulation of a wide range of

network types, including wired, wireless, and satellite networks.

4.4.2. Realism and Accuracy

The simulator offers high fidelity models that closely mimic real-world network behavior. This accuracy makes it suitable for validating theoretical models and conducting performance evaluations.

Programming Interface: NS-3 provides a flexible programming interface, allowing users to extend and customize the simulator to meet specific research needs. It supports C++ and Python, making it accessible to a broad range of users.

4.4.3. Visualization Tools

While NS-3 itself is focused on the simulation engine, it integrates with various visualization tools, such as NetAnim and PyViz, to provide graphical representations of network topologies and traffic flows [15].

4.5. Mininet

Mininet is a network emulator that creates a realistic virtual network on a single machine. It is widely used for developing, testing, and demonstrating software-defined networking (SDN) applications.

4.5.1. SDN Focus

Mininet is specifically designed to support SDN and OpenFlow. It allows users to create and experiment with SDN topologies, controllers, and applications, making it an essential tool for SDN research and development.

4.5.2. Rapid Prototyping

The emulator can quickly instantiate network topologies, making it ideal for rapid prototyping and testing. Users can simulate large networks with hundreds of nodes using minimal resources.

4.5.3. Integration with Real Networks

Mininet can integrate with physical networks, enabling hybrid environments where virtual and real devices interact. This feature is valuable for testing SDN applications in realistic settings.

4.5.4. Educational Use

Mininet is widely used in educational settings to teach SDN concepts and practices. Its ease of use and flexibility make it a popular choice for classroom labs and assignments [16].

4.6. Presentation of Simulators

4.6.1. Analysis of Cisco Packet Tracer

Cisco Packet Tracer is an innovative educational tool that allows trainees to create networks with nearly unlimited capabilities, using a wide range of devices. This encourages diagnostics and troubleshooting, enhancing the learning experience.

From the Educator's Perspective: Packet Tracer allows the teaching of network device functions that operate in the background, visible in everyday life. For instance, it can simulate the operation of a router, showing what happens from the moment a user enters a password until the router grants access to the user's device. The simulation capabilities simplify the learning process by providing tables, diagrams, and visual representations of internal functions, such as dynamic data transfers. The simulation function reduces presentation time by replacing tables and static slides with real-time visual effects.

4.6.2. Benefits for Educators

- Provides visual displays of complex technologies with configuration capability.
- Allows customized, guided activities with immediate feedback.
- Facilitates various learning activities like lectures, lab activities, homework, assessments, and games.
- Supports network design, troubleshooting, modeling tasks, and case studies.
- Enables visualization, movement, and detailed modeling for exploration, research, and experimentation.
- Encourages learning outside the classroom.
- Supports social learning, collaboration, and healthy competition.
- Covers most protocols and technologies taught in Cisco seminars and courses.

From the Student-Trainee Perspective: Packet Tracer offers a practical way of learning through simulation, allowing better opportunities to gain practical skills and knowledge when working with real equipment. Trainees gain faster experience through troubleshooting in simulation and real-world scenarios, building confidence and leading to a more productive workforce.

4.6.3. Workspaces

4.6.3.1.1. Logical Workspace

Users can create logical network topologies by placing, connecting, and grouping virtual network devices.

4.6.3.1.2. Physical Workspace

Provides a graphic-physical dimension of the logical network, showing how devices like routers, switches, and

mainframes would appear in a real environment. It also includes geographical representations of networks, including cities, buildings, and cables.

4.6.4. Modes of Operation

4.6.4.1. Real-Time Mode

The network and its devices behave like real devices, offering immediate real-time response for all networks and subnets.

4.6.4.2. Simulation Mode

Allows users to control the timing, internal functions of data transfer, and data propagation in a network, helping students understand the fundamental concepts behind network operations.

4.6.5. Supported Protocols and Applications

4.6.5.1. Application

FTP, SMTP, POP3, HTTP, TFTP, Telnet, SSH, DNS, DHCP, NTP, SNMP, AAA, ISR, VOIP, SCCP config and calls ISR command support, Call Manager Express.

4.6.5.2. Transport

TCP/IP, UDP, TCP Nagle Algorithm & IP Fragmentation, RTP.

4.6.5.3. Network

BGP, IPv4, ICMP, ARP, IPv6, ICMPv6, IPSec, RIPv1/ v2/ng, Multi-Area OSPF, EIGRP, Static Routing, Route Redistribution, Multilayer Switching, L3 QoS, NAT, CBAL, Zone-based policy firewall, and Intrusion Protection System on the ISR, GRE VPN, IPSec VPN.

4.6.5.4. Network Access/Interface

Ethernet (802.3), 802.11, HDLC, Frame Relay, PPP, PPPoE, STP, RSTP, VTP, DTP, CDP, 802.1q, PAgP, L2 QoS, SLARP, Simple WEP, WPA, EAP.

4.6.6. Additional Features

- Modular devices simulation with graphical hardware allowing interface card insertion into routers.
- Multi-user functionality for collaborative construction of virtual networks through a real network.
- Step-by-step tutorials, advanced workshops, and a comprehensive help feature.
- Activity Wizard for creating custom learning activities with grading and feedback capabilities.
- Lab scoring function, international language support, and compatibility with all platforms (Windows and Linux).

4.7. Analysis of GNS3

GNS3 is an open-source network simulator highly regarded for its flexibility and powerful features, allowing the simulation of complex networks using real network hardware images and virtualization technologies.

4.7.1. From the Educator's Perspective

GNS3 is a valuable tool for teaching network concepts and configurations, offering hands-on experience with real network operating system images. This allows students to gain practical knowledge and skills by working with actual network environments.

4.7.2. Benefits for Educators

- Provides a realistic simulation environment using real hardware images.
- Supports the integration of virtual machines, enabling comprehensive network simulations.
- Allows the use of a wide range of plugins and third-party tools for enhanced functionality.
- Offers community support for troubleshooting and sharing configurations.

4.7.3. From the Student-Trainee Perspective

GNS3 provides a practical learning platform where students can experiment with complex network setups and troubleshoot issues in a safe environment. This hands-on experience is crucial for developing real-world networking skills.

4.7.4. Workspaces

4.7.4.1. Topology Workspace

Users can create and manage network topologies by adding, connecting, and configuring virtual network devices and links.

4.7.4.2. Virtual Machine Integration

Supports integration with virtualization platforms like VirtualBox and VMware, allowing the inclusion of virtual machines in network simulations.

4.7.5. Modes of Operation

4.7.5.1. Real-Time Mode

Provides real-time simulation of network operations, allowing users to observe and interact with the network as if it were live.

4.7.5.2. Simulation Mode

Offers control over simulation parameters, enabling detailed analysis of network behavior and performance.

4.7.6. Supported Protocols and Applications:

4.7.6.1. Application

Supports a wide range of networking protocols and applications through real network operating system images (e.g., Cisco IOS, Juniper Junos).

4.7.6.2. Transport

Capable of simulating various transport protocols and behaviors, providing a comprehensive understanding of network interactions.

4.7.6.3. Network

Supports detailed simulation of network protocols, including routing, switching, and security protocols.

4.7.6.4. Network Access/Interface

Emulates a variety of network interfaces and access technologies, providing flexibility in network design and testing.

4.7.7. Additional Features

- Real hardware emulation for high-fidelity network simulations.
- Integration with virtual machines for end-to-end network scenarios.
- Extensibility through plugins and third-party tools like Wireshark.
- Active community support for continuous improvement and problem-solving.

4.8. Analysis of Opnet (Riverbed Modeler)

Opnet, now Riverbed Modeler, is a comprehensive network simulation tool used for modeling a wide range of network components and behaviors across various sectors.

4.8.1.1. From the Educator's Perspective

Riverbed Modeler offers detailed simulation capabilities that are invaluable for teaching advanced network concepts and performance analysis. Its extensive library of models allows educators to cover a wide range of networking topics with high accuracy.

4.8.2. Benefits for Educators

- Provides detailed models for wireless and wired communications, protocols, and queuing behaviors.
- Enables the simulation of complex hardware components and microprocessors.
- Offers a comprehensive GUI for visualizing network simulations and results.
- Supports the creation of custom simulation scenarios for targeted learning objectives.

4.8.2.1. From the Student-Trainee Perspective

Riverbed Modeler provides a powerful platform for students to experiment with network configurations and analyze performance metrics. Its detailed simulation capabilities help students understand the intricacies of network operations and the impact of various factors on performance.

4.8.3. Workspaces

4.8.3.1. Modeler Workspace

Allows users to create detailed network models by adding and configuring nodes, links, and protocols.

4.8.3.2. Simulation Workspace

Provides tools for setting up and running simulations, as well as analyzing results through graphical and statistical outputs.

4.8.4. Modes of Operation

4.8.4.1. Real-Time Mode

Simulates network operations in real-time, providing immediate feedback and insights.

4.8.4.2. Simulation Mode

Allows users to control simulation parameters and analyze network behavior over time.

4.8.5. Supported Protocols and Applications

4.8.5.1. Application

Includes a vast library of application-layer protocols for comprehensive simulation of network traffic and interactions.

4.8.5.2. Transport

Supports detailed modeling of transport protocols like TCP and UDP, providing insights into performance and reliability.

4.8.5.3. Network

Offers extensive models for network-layer protocols, including various routing and switching protocols.

4.8.5.4. Network Access/Interface

Emulates a wide range of access technologies and network interfaces for flexible network design and testing.

4.8.6. Additional Features

- Detailed queue models for analyzing queuing behavior in network devices.
- Support for complex hardware modeling, including microprocessor performance.

- Comprehensive GUI for visualizing network simulations and results.
- Extensive library of protocol models covering all layers of the OSI model.

4.9. Analysis of NSE (Network Simulator 3)

NS3 is an open-source discrete-event network simulator, primarily used for research and educational purposes. It provides a realistic simulation environment for network protocols and internet systems.

From the Educator's Perspective: NS3 is a valuable tool for teaching network protocols and internet systems due to its detailed and accurate simulation capabilities. It allows educators to create realistic network scenarios and analyze the performance of various protocols.

4.9.1. Benefits for Educators

- Offers a realistic and flexible simulation environment.
- Provides extensive documentation and tutorials, aiding in the learning process.
- Supports a wide range of network protocols and models, allowing comprehensive educational coverage.

4.9.1.1. From the Student-Trainee Perspective

NS3 provides a detailed learning platform for students to understand network protocols and their behaviors in real-world scenarios. The hands-on experience with NS3 aids in developing a deep understanding of network operations.

4.9.2. Workspaces

4.9.2.1. Simulation Script Workspace

Users can write and execute simulation scripts in C++ or Python, allowing for detailed and customized network simulations.

4.9.3. Modes of Operation

4.9.3.1. Simulation Mode

Provides detailed and controlled simulation of network scenarios, allowing users to analyze protocol performance and network behavior.

4.9.4. Supported Protocols and Applications

4.9.4.1. Application

Includes support for a wide range of application-layer protocols for realistic simulation of network traffic and interactions.

4.9.4.2. Transport

Provides detailed modeling of transport protocols like TCP, UDP, SCTP, and more, offering insights into their performance and reliability.

4.9.4.3. Network

Supports extensive models for network-layer protocols, including various routing and switching protocols.

4.9.4.4. Network Access/Interface

Emulates a wide range of access technologies and network interfaces, enabling flexible network design and testing.

4.9.5. Additional Features

- Realistic and detailed simulation environment for network protocols.
- Extensible through custom scripts in C++ or Python, allowing for highly customizable simulations.
- Strong community support with extensive documentation and user-contributed models.
- Regular updates and active development to keep up with emerging networking technologies.

4.10. Analysis of Mininet

Mininet is an open-source network emulator that creates a realistic virtual network, running real kernel, switch, and application code on a single machine. It is widely used for developing and testing network applications and protocols.

4.10.1. From the Educator's Perspective

Mininet is an excellent tool for teaching software-defined networking (SDN) and network function virtualization (NFV) concepts. Its ability to emulate a complete network on a single machine makes it accessible for educational environments.

4.11. Benefits for Educators

- Provides a realistic and interactive environment for teaching SDN and NFV.
- Supports the creation of complex network topologies with minimal hardware requirements.
- Includes extensive documentation and a large collection of example scripts for various network scenarios.

4.11.1. From the Student-Trainee Perspective

Mininet offers a hands-on learning experience, allowing students to experiment with real network code and configurations. This practical approach helps in understanding the intricacies of network behavior and management.

4.11.2. Workspaces:

4.11.2.1. CLI Workspace

Users can interact with the Mininet environment through a command-line interface, allowing for the creation and management of network topologies.

4.11.2.2. Python API Workspace

Provides a Python API for scripting network configurations and behaviors, enabling automation and customization.

4.11.3. Modes of Operation

4.11.3.1 Emulation Mode

Emulates a complete network on a single machine, allowing for the testing and development of network applications and protocols in a controlled environment.

4.11.4. Supported Protocols and Applications:

4.11.4.1. Application

Supports the deployment of real applications within the emulated network, providing a realistic environment for testing network applications.

4.11.4.2. Transport

Emulates transport protocols like TCP and UDP, allowing for detailed analysis of their behavior and performance.

4.11.4.3. Network

Supports a wide range of network protocols, including SDN protocols like OpenFlow.

4.11.4.4. Network Access/Interface

Emulates various network interfaces and access technologies, enabling flexible network design and testing.

4.11.5. Additional Features

- Real kernel, switch, and application code running in the emulated environment.
- Supports integration with SDN controllers like OpenDaylight and Ryu.
- Extensible through custom scripts and plugins, allowing for advanced network emulation scenarios.
- Active community support with extensive documentation and user-contributed examples.

By leveraging these tools, network professionals and students can gain valuable insights, improve network designs, and enhance overall network performance. These simulators provide robust platforms for learning, testing, and optimizing networks, ensuring efficient and effective network management and development.

5. Results of the comparison & Conclusions

In this section, we present the results of our comparative analysis of three prominent network simulation tools: Cisco Packet Tracer, GNS3, and Riverbed Modeler Academic Edition (Opnet). Each tool's capabilities, strengths, and limitations were evaluated based on several key criteria: educational applicability, user experience, simulation depth, flexibility, and scalability. This comprehensive comparison aims to provide insights into how each simulator can be optimally utilized in various educational and professional contexts.

5.1. Comparative Criteria and Results

5.1.1. Educational Applicability:

5.1.1.1. Cisco Packet Tracer

- Strengths: Highly effective for beginners due to its intuitive interface and straightforward simulation capabilities. Integral to Cisco's Networking Academy curriculum, making it a staple in foundational network training.
- Limitations: Primarily focuses on Cisco devices and protocols, which may limit exposure to broader networking environments.

5.1.1.2. GNS3

- Strengths: Provides a realistic simulation environment using real network hardware images, which is beneficial for advanced learning and professional training. Supports a wide range of network devices and protocols.
- Limitations: The complexity of setup and use can be a barrier for beginners, requiring more advanced knowledge and skills.

5.1.1.3. Riverbed Modeler Academic Edition

- Strengths: Offers comprehensive modeling capabilities that are ideal for higher education and professional research. Detailed simulations and extensive protocol support make it suitable for in-depth studies and advanced network analysis.
- Limitations: Resource-intensive and complex, making it less accessible for beginners and smaller educational institutions.

5.1.1.4. NS3

- Strengths: Highly suitable for advanced research and educational purposes. Provides a realistic and detailed simulation environment for network protocols and internet systems.
- Limitations: Requires advanced technical knowledge and programming skills, making it less accessible for beginners.

5.1.1.5. Mininet

- Strengths: Ideal for teaching software-defined networking (SDN) and network function virtualization (NFV). Provides a practical, hands-on learning experience with real network code and configurations.
- Limitations: Limited to emulating rather than simulating, which may not fully capture the nuances of complex network interactions

5.1.2. User Experience:

5.1.2.1. Cisco Packet Tracer

- Strengths: User-friendly interface with visual aids and guided activities. Real-time and simulation modes enhance interactive learning.
- Limitations: Limited to Cisco-specific environments, which may not fully prepare users for multi-vendor network scenarios.

5.1.2.2. GNS3

- Strengths: Flexibility and realism due to the use of real network OS images. Strong community support provides additional resources and troubleshooting help.
- Limitations: The steep learning curve and technical setup requirements can be challenging for less experienced users.

5.1.2.3. Riverbed Modeler Academic Edition

- Strengths: Comprehensive GUI and detailed feedback mechanisms. Supports complex simulations with high accuracy.
- Limitations: High cost and complexity may deter smaller institutions and individual learners from using it.

5.1.2.4. NS3

- Strengths: Highly detailed and customizable simulation environment. Strong support for scripting in C++ and Python.
- Limitations: Requires significant programming expertise and setup time.

5.1.2.5. Mininet

- Strengths: Simple setup and easy-to-use CLI and Python API for network emulation. Strong integration with SDN controllers and virtualization platforms.
- Limitations: Less detailed than full network simulators like NS3 or Riverbed, as it focuses more on emulation.

5.1.3. Simulation Depth

5.1.3.1. Cisco Packet Tracer

- Strengths: Adequate for basic to intermediate network simulations. Provides essential tools for learning network concepts and troubleshooting.
- Limitations: Less detailed compared to GNS3 and Riverbed in terms of advanced protocol and network behavior simulations.

5.1.3.2. GNS3

- Strengths: High fidelity due to real hardware emulation. Supports a wide range of detailed network scenarios.
- Limitations: Requires substantial computing resources for complex simulations.

5.1.3.3. Riverbed Modeler Academic Edition

- Strengths: Offers in-depth and comprehensive simulation capabilities, including detailed queuing models and hardware behavior.
- Limitations: High complexity and resource requirements.

5.1.3.4. NS3

- Strengths: Provides highly detailed and accurate simulations of network protocols. Extensive support for custom simulation scripts and detailed statistical analysis.
- Limitations: Can be resource-intensive and complex to set up for large-scale simulations.

5.1.3.5. Mininet

- Strengths: Adequate for emulating network environments, particularly useful for SDN and NFV experiments.

- Limitations: Emulation rather than full simulation, which limits its ability to model network behavior in detail.

5.1.4. Flexibility and Scalability

5.1.4.1. Cisco Packet Tracer

- Strengths: Suitable for small to medium-sized network simulations. Allows for logical and physical workspace configurations.
- Limitations: Scalability is limited compared to GNS3 and Riverbed.

5.1.4.2. GNS3

- Strengths: Highly scalable and flexible, supporting integration with virtual machines and a wide range of plugins.
- Limitations: Requires advanced setup and configuration.

5.1.4.3. Riverbed Modeler Academic Edition

- Strengths: Capable of simulating large-scale networks with detailed feedback. Highly flexible in terms of protocol and hardware modeling.
- Limitations: Requires significant resources and expertise to manage large-scale simulations.

5.1.4.4. NS3

- Strengths: Extremely flexible and scalable, suitable for simulating a wide range of network scenarios. Strong support for custom extensions and detailed modeling.
- Limitations: Requires extensive programming knowledge and resources.

5.1.4.5. Mininet

- Strengths: Highly flexible for SDN and NFV emulations, allowing integration with real network controllers and virtual machines.
- Limitations: Limited scalability compared to full-scale simulators due to its focus on emulation.

5.1.5. 5.2 Comparative Insights

5.1.5.1. Cisco Packet Tracer

Best For: Beginners and intermediate learners, educational institutions focusing on foundational network training, and environments where Cisco certification is a priority.

5.1.5.1.1 Summary

Cisco Packet Tracer's simplicity and educational focus make it an excellent tool for introducing networking concepts. However, its limitations in multi-vendor simulations and advanced network behaviors mean it may not be suitable for more complex or diverse network studies.

5.1.5.2. GNS3

Best For: Advanced students, network professionals, and users seeking a realistic and flexible simulation environment.

5.1.5.2.1. Summary

GNS3 stands out for its realism and flexibility, providing a highly detailed simulation experience using real network OS images. It is ideal for users who need to simulate complex networks and advanced protocols, although its setup complexity may be challenging for some.

5.1.5.3. Riverbed Modeler Academic Edition

Best For: Higher education, research institutions, and professional training programs requiring detailed and comprehensive network simulations.

Summary: Riverbed Modeler excels in detailed and large-scale network simulations, offering extensive protocol support and detailed feedback. Its complexity and resource requirements make it more suitable for advanced users and institutional settings.

5.1.5.4. NS3

Best For: Researchers, advanced students, and professionals requiring detailed and programmable network simulations.

5.1.5.4.1. Summary

NS3 provides a highly detailed and customizable simulation environment suitable for research and advanced educational purposes. Its extensive support for custom simulation scripts and detailed statistical analysis makes it ideal for in-depth network studies. However, it requires significant programming expertise and setup time, which can be a barrier for beginners.

5.1.5.5. Mininet

Best For: Educators and students focusing on software-defined networking (SDN) and network function

virtualization (NFV), and professionals needing a practical, hands-on learning tool.

5.1.5.5.1. Summary

Mininet is ideal for teaching and learning SDN and NFV concepts. It offers a practical, hands-on experience by allowing users to experiment with real network code and configurations in an emulated environment. While it is excellent for SDN and NFV scenarios, its emulation approach may not capture the nuances of complex network behaviors as comprehensively as full network simulators.

5.2. Conclusions

The comparative analysis reveals that each network simulator has its unique strengths and is best suited for different educational and professional contexts. Cisco Packet Tracer is ideal for foundational learning and beginner training due to its simplicity and focus on Cisco devices. GNS3 offers advanced simulation capabilities and flexibility, making it suitable for professional training and complex network studies. Riverbed Modeler provides comprehensive and detailed simulations, ideal for higher education and research purposes.

Network simulators are invaluable tools in both educational and professional settings, offering platforms for students and professionals to engage with complex networking concepts in a practical, hands-on manner. The choice of the appropriate network simulator is crucial and should be guided by the specific needs and goals of the user, as each simulator brings its unique advantages and limitations to the table.

Cisco Packet Tracer is particularly well-suited for introductory and intermediate networking courses. Its user-friendly interface, combined with its integration into the Cisco Networking Academy curriculum, makes it an excellent choice for institutions aiming to provide foundational networking education. Its real-time and simulation models offer flexibility in teaching methods, enabling instructors to visualize network operations and troubleshoot scenarios in a controlled environment. However, its focus on Cisco-specific devices and protocols means that it may not provide the broad exposure necessary for students who need to learn about a wider range of network environments.

GNS3, with its ability to emulate real network operating systems and integrate with virtualization platforms, provides a high level of realism that is unmatched by other simulators. This makes it an ideal tool for advanced students and network professionals who require a deeper understanding of network operations and need to simulate complex, multi-vendor environments. The flexibility and extensibility of GNS3, supported by a strong user community, allows for detailed and customizable simulations. Nonetheless, its steep learning curve and technical setup requirements can be challenging, making it less suitable for beginners.

Riverbed Modeler Academic Edition stands out for its comprehensive simulation capabilities and detailed feedback mechanisms, making it a powerful tool for higher education and professional training. Its ability to simulate large-scale networks and provide in-depth analysis of network performance is invaluable for research and development projects. The extensive protocol support and complex hardware modeling capabilities of Riverbed Modeler enable users to conduct detailed studies and gain insights into the intricate workings of network systems. However, its complexity and resource requirements may limit its accessibility to smaller institutions and individual learners.

NS3 is highly regarded in the research community for its detailed and programmable network simulations. It is ideal for researchers and advanced students who require customizable simulation scripts and detailed statistical analysis. NS3's extensive support for custom simulations makes it particularly suited for in-depth network studies. However, it demands significant programming expertise and setup time, which can be a barrier for beginners.

Mininet is an excellent tool for teaching and learning software-defined networking (SDN) and network function virtualization (NFV). It offers practical, hands-on experience by allowing users to experiment with real network code and configurations in an emulated environment. Mininet's ability to emulate real network devices and its ease of use make it particularly beneficial for students and educators focusing on SDN and NFV. While it excels in these areas, its emulation approach may not capture the full complexity of network behaviors compared to full network simulators like GNS3 and Riverbed Modeler. [17] and [18]

5.3. Future Directions

Integration into Remote Learning: Exploring how these tools can be integrated into online courses and remote learning platforms to meet the growing demand for virtual education.

Real-World Data Incorporation: Enhancing simulations with real-world data to improve their realism and applicability.

Cross-Platform Interoperability: Investigating ways to improve interoperability between different simulation tools for a more seamless educational experience.

Summarizing, Cisco Packet Tracer, GNS3, Riverbed, NS3 and Mininet each offer distinct advantages that cater to different aspects of networking education and professional training. By leveraging the strengths of these

simulators, educators can create robust and effective learning environments that prepare students for the complexities of modern network management. As network technologies evolve, continuous improvement and adaptation of these tools will be essential in maintaining their relevance and utility in both educational and professional contexts [19].

References

[1] C. CasconeO'Connor, T. Vachuska, B. Davie, "Software-Defined Networks: A Systems Approach," 2021, p. 152.

[2] E. Altman, T. Jimenez, "NS Simulator for beginners," Lecture notes. [Online]. Available: http://www-sop.inria.fr/maestro/personnel/Eitan.Altman/COURS-NS/n3.pdf

[3] "Future Directions in Control, Dynamics and Systems," [Online]. Available: http://www.cds.caltech.edu/~murray/cdspanel, last accessed Mar. 2005.

[4] L. Globa, M. Skulysh, O. Romanov, M. Nesterenko, "Quality Control for Mobile Communication Management Services in Hybrid Environment," in *Advances in Information and Communication Technologies*, UKRMICO 2018, Lecture Notes in Electrical Engineering, vol 560, Springer, Cham, 2019, pp. 20-29. DOI: https://doi.org/10.1007/978-3-030-16770-7_4, Print ISBN: 978-3-030-16769-1.

[5] IETF RFC 7426, Request for Comments: 7426, ISSN: 2070-1721 EICT. Category: Informational, K. Pentikousis, Ed., 2015.

[6] J. Nilsson, "Real-time Control Systems with Delays," Ph.D. dissertation, Dept. Automatic Control, Lund Institute of Technology, Lund, Sweden, 1998. [Online]. Available: http://www.control.lth.se/documents/1998/nilj98dis.pdf

[7] J. Lam, S. Lee, H. Lee, Y. Oktian, "Securing SDN Southbound and Data Plane Communication with IBC," *Mobile Information Systems*, vol. 2016, Article ID 1708970, 12 pages, 2016. DOI: http://dx.doi.org/10.1155/2016/1708970.

[8] K. Phemius, M. Bouet, J. Leguay, "DISCO: Distributed Multi-domain SDN Controllers," Thales Communications & Security, arXiv:1308.6138v2 [cs.NI], Aug. 2013.

[9] O. I. Romanov, M. V. Oryschuk, Y. S. Hordashnyk, "Computing of influence of stimulated Raman scattering in DWDM telecommunication systems," in *2016 International Conference Radio Electronics & Info Communications (UkrMiCo)*, 2016. DOI: 10.1109/UkrMiCo.2016.7739622, eISBN: 978-1-5090-4409-2.

[10] O. Romanov, V. Mankivskyi, "Optimal Traffic Distribution Based on the Sectoral Model of Loading Network Elements," in *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, Oct. 2019, pp. 8-11. DOI: 10.1109/PICST47496.2019.9061296, eISBN: 978-1-7281-4184-8.

[11] O. I. Romanov, M. M. Nesterenko, N. O. Fesokha, V. B. Mankivskyi, "Evaluation of productivity virtualization technologies of switching equipment telecommunications networks," *Information and Telecommunication Sciences*, vol. 11, no. 1, pp. 53-58, 2020. DOI: https://doi.org/10.20535/2411-2976.12020.53-58.

[12] ONF specification, "OpenFlow Table Type Patterns," Version 1.0, Aug. 2014.

[13] ONF TR-508, "Requirements Analysis for Transport OpenFlow/SDN," Version 1.0, Aug. 2014.

[14] ONF TR-537, "Negotiable Datapath Model and Table Type Pattern Signing," Version 1.0, Sep. 2016.

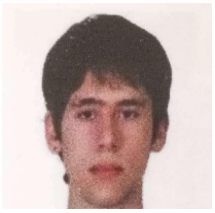
[15] ONF TR-539, "OpenFlow Controller Benchmarking Methodologies," Version 1.0, Nov. 2016.

[16] ONF TS-029, "MPLS-TP OpenFlow Protocol Extensions for SPTN," Version 1.0, June 2017.

[17] ONF, "OpenFlow Switch Specification Version 1.5.1 (Protocol version 0x06)," 2014. [Online]. Available: https://opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf

[18] ONOS, "Security and Performance Analysis," Report No. 2, Nov. 2018.

[19] Open Network Foundation, "Accelerating the Adoption of SDN & NFV," 2021. [Online]. Available: http://opennetworking.wpengine.com/wp-content/uploads/2013/03/ONF%20Overview%204-pager%20FINAL.pdf



NIKOLAOS V. OIKONOMOU has received his BSc degree from University of Ioannina, Department of Informatics and Telecommunications in 2021. He has received his MSc degree from the same institution in 2023. He is an academic researcher also working as a private tutor in the field of Computer Science and mathematics. He has years of experience as a Computer Engineer, IT specialist and Network consultant. He also taught in University of Ioannina and worked as an application developer.

DIMITRIOS V. OIKONOMOU has received his BSc from University of Western Macedonia, Department of Regional and Cross Border Studies in 2024. He is currently an active research member of the University of Western Macedonia and is about to begin his MSc studies.

Received: 05 April, 2024, Revised: 29 April, 2024, Accepted: 16 May, 2024, Online: 23 June, 2024

DOI: https://doi.org/10.55708/js0306001

Using Artificial Intelligence Models to Predict the Wind Power to be Fed into the Grid

Sambalaye Diop, Papa Silly Traore, Mamadou Lamine Ndiaye, Issa Zerbo

[1]Water, Energy, Environment, Industrial process (LE3PI) Polytechnic school of Dakar (ESP -UCAD) Dakar, Senegal

[2]Joseph KI-ZERBO University; Renewable Thermal Energy Laboratory (L.E.T.RE); Ouagadougou, Burkina Faso

*Corresponding author: Sambalaye Diop, Université Cheikh Anta Diop de Dakar, Sambalayediop7@gmail.com

ABSTRACT: The Taïba Ndiaye wind farm, connected to the SENELEC grid, plays a key role in offsetting shortfalls in electricity consumption, with an installed capacity of 158.7 MW. Moreover, as an intermittent power station, its production is highly dependent on the environmental conditions in the region. Bad weather can disrupt the electricity network, requiring forecasting methods to anticipate its production. This will make it easier to decide how much fossil energy to bring on stream to meet demand. The aim of this paper is to provide forecasts of wind generation at Taïba Ndiaye, subdividing the data into 80% for model training and 20% to assess its robustness to generalization to other situations. The aim is to quantify the energy produced and facilitate an optimal transition between intermittent and fossil energy sources. Two artificial intelligence models classified as machine learning (decision tree and random forest) are proposed in the study, with respective coefficients of determination of 0.92 and 0.938. The results, compared with the literature, demonstrate the reliability of the approach using only production data. These results promise significant benefits in terms of resource management.

KEYWORDS: Taïba Ndiaye, Wind power, SENELEC grid, forecast, machine learning, artificial intelligence models

1. Introduction

Prior to the integration of intermittent renewable energies into the power grid, the flow of energy followed a single direction, ensuring greater stability of the power system [1]. Today, however, with the injection of these energies, such as solar and wind power, the energy flow becomes bidirectional, which easily disrupts the grid when faced with rapid variations in meteorological parameters [2]. Furthermore, the injection of these intermittent energies must not exceed 30% of total energy demand in some countries [3]. This presents grid operators with a significant challenge in maintaining a consistent balance between production and consumption to avoid malfunctions, undesirable voltage and frequency variations, and costly imbalances [4]. Network operators must be able to anticipate the production of intermittent power plants in order to adjust the production of fossil fuels, thereby balancing customer consumption with production. Furthermore, in view of global concerns about the fight against climate change, the electricity grids of several countries continue are integrating intermittent energies into their electricity grids, despite the drawbacks [5]. Senegal is following a similar approach, with a 30% increase in the energy mix [6]. These include the Taïba Ndiaye wind farm, with a capacity of 158.7 MW, as well as Malicounda (20 MW), Diass (23 MW), Bokhol (20 MW), etc [7][8]. Against this backdrop of high penetration of intermittent renewable energies, forecasting has become essential to ensure the stability of the electricity network [9]. Several studies have focused on forecasting the potential of renewable resources, whether solar or wind. These studies mainly rely on artificial intelligence models to predict wind energy, given its complex characteristics of continuous production both day and night, which makes this difficult [10]. Indeed, operators face difficulties

due to the volatile nature of these sources, with weather parameters requiring constant monitoring to anticipate tasks linked to technical constraints [11]. To overcome these challenges, data science experts are working more closely with grid operators to collect data in order to accurately predict intermittent energy with artificial intelligence (AI) models. AI-based forecasting models are fed by data from sensors installed in the power plant. These models are currently significantly improving the prediction of intermittent power plant output with high accuracy [12][13]. Their reliability in predictive decision-making is no longer in question [14]. In fact, they enable production to be predicted over fairly short time horizons, thus enabling the SENELEC distributor to ensure the stability of network frequency and voltage [8]. Comparative studies have confirmed that these AI models outperform statistical models because of their very satisfactory predictive power [15]. This is evidenced by the studies conducted on the wind power production in Italy and the United States, as well as in Senegal on short-term solar irradiance [16][17]. Despite the robustness and relevance of AI models, their intensive use of data with several input variables to predict the target is not without consequences for computing resources, requiring considerable computing power. Some experts in the field have highlighted that their machines can sometimes overload, while others have mentioned that the latent time is sometimes too high to obtain optimal results [7]. To address this issue, we propose using two parameter predictors to forecast the short-term power output of the Diass wind power plant, using random forest models and decision trees. These models will be trained using only wind generation measured over a one-year period. The objective is to improve the prediction performance of the wind power plant by reducing the number of input parameters [7]. This paper is structured as follows: the presentation of the data as well as the wind power plant and the method discussed is provided in section II. Section III outlines the AI algorithms used. In Section IV, the results and discussion are presented. Finally, in Section V, the conclusion is provided.

2. Presentation of the plant and data

2.1. Classification of the Taïba Ndiaye wind farm

Wind energy is the kinetic energy generated by the movement of the wind, transformed into mechanical energy by wind turbines and then converted into electrical energy. The energy is given by equation (1).

$$E = \frac{1}{2} \times A \times \rho \times V^3 \times C_p \times \eta \quad (1)$$

where:

- E: is the wind energy produced (in watts or joules),

- A: is the area swept by the turbine blades (in square metres),

- ρ: is the density of the air (in kilograms per cubic metre),

- V: is the wind speed (in metres per second),

- Cp: is the power coefficient of the wind turbine (without unit, a typical maximum value is around 0.59),

- η: is the mechanical and electrical efficiency of the system (unitless, a typical value is around 0.85).

The value of the power coefficient Cp depends on the speed of rotation of the turbines and the angle of inclination of the blades. Wind turbines are classified into three groups according to propeller diameter and power output [18]. Table I shows a classification of wind turbines:

Table 1: Classification of wind turbines [18]

Group	Propeller diameter D_h	Power output P_w
Small wind turbine	$D_h \leq 12$ m	$P_w \leq 40$ kW
Average wind turbine	12 m < $D_h \leq 45$ m	40 kW < $P_w \leq 999$ kW
Large wind turbine	$D_h > 45$ m	$P_w > 1$ MW

According to this classification, our study plant, with a capacity of 158.7 MW, is classified as a large wind power plant. It is equipped with the necessary data collection equipment. These enable efficient planning of energy production by anticipating load variations to meet injection requirements. Careful analysis facilitates energy injection, minimizing waste and reducing the costs associated with fluctuations in production. It also facilitates the integration of forecasting models to guarantee operational stability.

2.2. Production data

Figure 1 illustrates the data obtained from the sensors installed at the wind farm. These data are related to various environmental factors that will be used in our forecasts. The measurements were taken every ten (10) minutes for one year and averaged by hours, days and months. This is a time series with repeating trends at the beginning and end of the year, probably due to favourable weather conditions [8]. Their associated temporal indices are of the order of minutes, hours, days and months. These parameters are crucial for modelling this type of problem. To achieve an accurate prediction, we will incorporate seasonal phenomena, including the temporal indices, into the data reduction process.

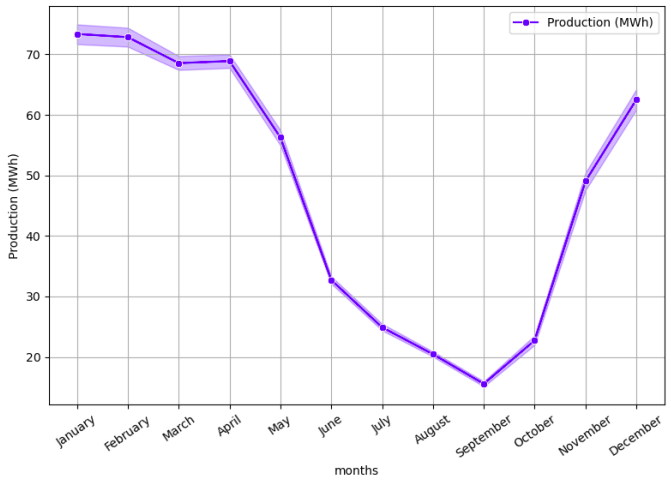


Figure 1: Production profile of the Taïba Ndiaye wind farm

Unlike solar power, wind power generates energy continuously, but this continuity is subject to unpredictable variations due to weather conditions. This intermittent nature of wind generation can sometimes pose complex challenges for electricity network managers. It is therefore important to keep a close eye on environmental parameters such as wind speed and direction, as they are closely linked to wind power generation. These variations can be rapid and significant, requiring proactive management to ensure grid stability. By understanding and anticipating these intermittencies, managers can take appropriate measures to maintain a reliable electricity supply.

2.3. *Wind speed data*

The power law also known as Murphy's law is a widely used approach to modelling wind speed [19],[20]. It states that the wind speed V at a given height above the ground is proportional to the power of the height h. Its mathematical relationship is given by equation (2) [19]:

$$V = V_{ref}\left(\frac{h}{h_{ref}}\right)^{a} \quad (2)$$

where:

- V: is the wind speed at height h,
- V_{ref}: is the wind speed at a reference height h_{ref},
- α: is the exponent of the power law, which depends on local site conditions and terrain characteristics.

From the modelling, the wind speed can be collected at the wind turbine installation site. The variation in mean wind speed at the Taïba Ndiaye site is shown in Figure 2.

It shows the typical fluctuations in wind speed, which are characterised by periods of rise and fall. These fluctuations are often influenced by specific weather conditions and are continuous throughout the day, month and year. This continuous variation in wind speed presents a significant challenge when predicting wind generation. It is particularly complex because of this variability. Indeed, this variability in wind speed can lead to rapid changes in energy production, requiring dynamic management of energy resources to maintain the stability of the power grid. This requires the use of advanced modelling and simulation techniques, as well as artificial intelligence algorithms capable of analysing large datasets and recognizing complex patterns.

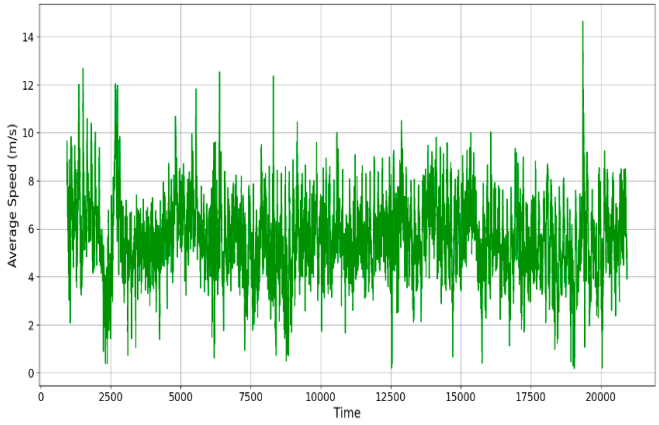


Figure 2: Wind speed profile for the Taïba Ndiaye area

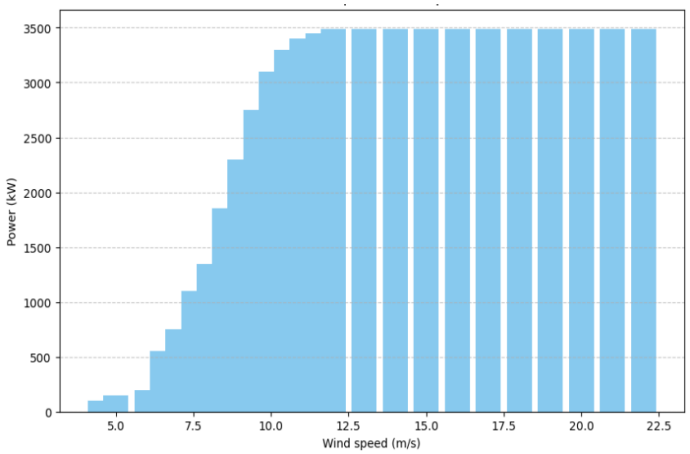


Figure 3: Power curve of the wind turbine installed at Taïba Ndiaye

Given the wind speed data, a wind turbine with the power curve shown in Figure 3 was chosen for the model.

The turbine's power increases until it reaches a speed of 12 m/s, where it remains constant until 22 m/s, which could correspond to the turbine's stall speed. This indicates that the turbine is designed to operate optimally within a predefined range of wind speeds. The turbine reaches its maximum rated output at a wind speed of 12 m/s. During this period, the turbine makes full use of the available kinetic energy of the wind. The turbine's power remains constant above the rated speed, up to 22 m/s. This mechanism is designed with an effective control system to prevent overloads and damage caused by excessively high wind speeds. In fact, the stall system protects the wind turbine and guarantees the durability of the components while stabilising the electricity. For accurate prediction purposes, it is important to take these wind fluctuations into account to provide a model capable of accurately predicting wind energy production. However, wind direction is one of the elements that creates turbulence, which is synonymous with wind fluctuations. It can have a positive influence on wind installations and their production.

2.4. *Wind direction data*

The wind direction mainly shows that the winds blow in the optimum directions. These predominant wind directions correspond to the periods of maximum

production for the wind power plant. Figures 4 and 5 show the predominant wind directions during the day and night respectively in the Taïba Ndiaye area. Figure 4 shows that the predominant wind direction from south-east to north-west during the day, it can be seen that the highest wind speeds are between 6 m/s and 8 m/s. Wind speeds of up to 10 m/s are fairly limited. At night, on the other hand, the prevailing winds blow from south to north at speeds of around 10 m/s. This observation shows that the wind farm's output is higher at night. It is therefore important to carefully monitor of these wind data in order to guarantee optimum energy feed-in to the power grid. By monitoring and anticipating variations in wind direction and speed, operators can adjust the plant's output accordingly. This not only optimises the production of wind energy, but also its smooth integration into the electricity grid, contributing to a more stable and reliable power supply for consumers.

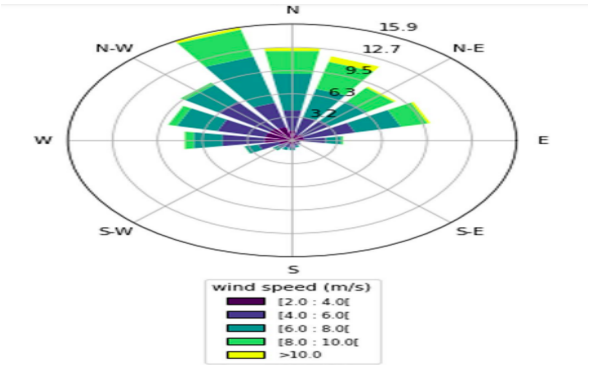


Figure 4: Wind rose for the Taïba Ndiaye power plant installation

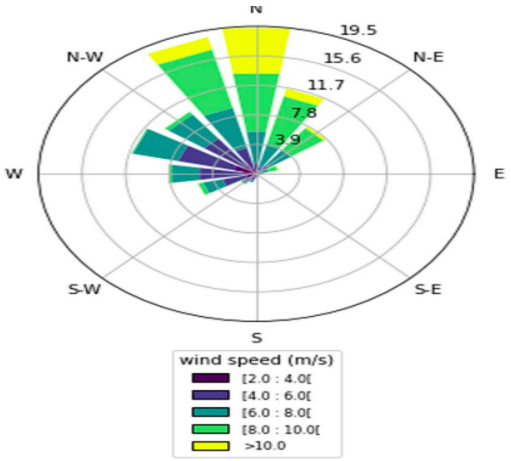


Figure 5: Wind rose in the Taïba Ndiaye power plant installation zone at night.

2.5. *Presentation of the plant*

At Taïba Ndiaye, the data collected come from the wind power plant, which is an impressive installation consisting of 46 Vesta V 126-3.45 wind turbines. The plant is equipped with a collector that feeds two 33/225 kV, 80/100 MVA step-up transformers, which gives it significant generating capacity [21]. The plant is strategically connected to the interconnected 225 kV network of the *Organisation pour la Mise en Valeur du Fleuve Sénégal* (OMVS), with an installed capacity of 158.7 MW. Figure 6 provides a clear illustration the installed wind farm and its characteristics. It is located in an open area and is well positioned to capture the wind. The meteorological data showed the dominant wind directions, as illustrated in Figures 4 and 5. It is important for the production of renewable energy and contributes significantly to the energy mix in the electricity grid.

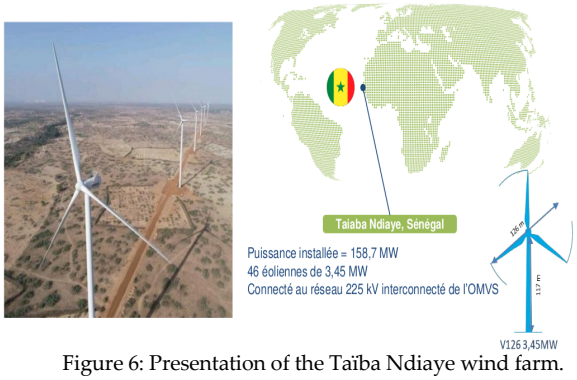


Figure 6: Presentation of the Taïba Ndiaye wind farm.

The importance of this data goes beyond simply monitoring solar production. By analysing this data, researchers may be able to gain a deeper understanding of the plant's current performance, as well as develop predictive models to anticipate seasonal and meteorological variations.

3. Prediction Algorithms

In this section, the two used prediction algorithm (the decision tree and the random forest algorithm) are presented.

3.1. *Decision Tree Model*

The decision tree is a classification and regression tree. The configuration of the tree is shown in Figure 7 and consists of the following elements:

- Root node: This represents the highest points in the figure 7.

- Internal nodes: These correspond to tests formulated in the form of questions on the characteristics of the parameters in relation to the target to be predicted.

- Branches: These present the results of the tests, and according to these answers, the subdivision is made as observed in figure 7.

- Leaf nodes: These nodes represent a decision.

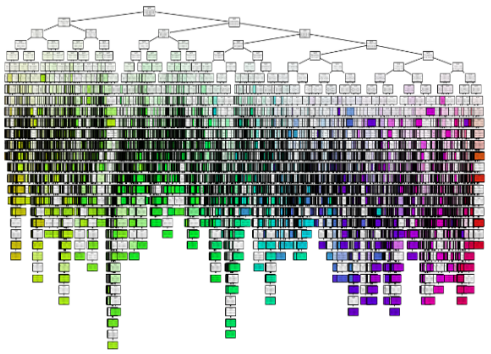


Figure 7: Illustration of the production data decision tree.

The partition equation of each node into two classes is given by (3) [14]:

$$\boldsymbol{X}: \boldsymbol{X_j} \leq \boldsymbol{s} = \boldsymbol{C1}(\boldsymbol{j}, \boldsymbol{s})$$
$$\boldsymbol{X}: \boldsymbol{X_j} > \boldsymbol{s} = \boldsymbol{C2}(\boldsymbol{j}, \boldsymbol{s}) \quad (3)$$

The couple (j, s) designate the partition limit of the data that we try to predict. Here, the goal is to find the boxes C1, ..., CJ that minimize the least squares criterion, represented by (4) according to [22]:

$$\sum_{j=i}^{k} \sum_{i:\in Rj}^{N} \left(Y_i - \widehat{Y}_t\right) = \text{SSE} \quad (4)$$

Where Y_i and $\hat{Y}_t$ refers to the actual and the predicted values respectively and SSE Residual Sum of Squares.

The forecasting task involves explaining the target variable Y (plant output) as a function of a set of explanatory variables X (measurement times, day number and month). Thus, the different modalities of the X explanatory variables are examined using the chi-square test to determine which variables are closely related to the Y target. When the p-value of the chi-square test is less than 0.05, we conclude that the variable is significantly associated with the target variable Y. This criterion is particularly important when the learning loop is interrupted, ensuring that all nodes have chi-square tests greater than 0.05, indicating the absence of a strong association between the explanatory variables X and the target variable Y.

3.2. Random Forest Model

This model consists of a collection of several decision trees trained using the Bagging method. The algorithm is applied in three stages:

-Bagging: this is a technique that involves grouping several decision trees together to obtain a final result, rather than relying on individual decision trees. Figure 8 shows its format.

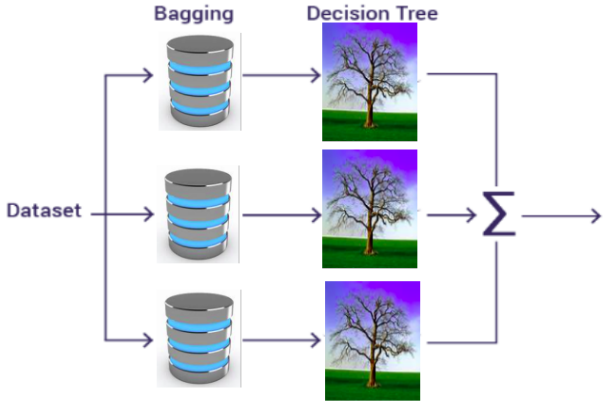


Figure 8: Bagging phase of the random forest algorithm

-Bootstrapping: This is a process that begins with the application of the bootstrap technique, which is a sampling method as shown in Figure 9. This approach involves creating random subsets from the initial dataset, using N samples. The N samples are selected with replacement, allowing the same sample to be included several times in the subset.

-Bagging aggregation: In the bagging aggregation phase, each random subset is subjected to a decision tree algorithm. The final result is obtained by taking the average of all the predictions generated by the different trees, as shown in Figure 10.

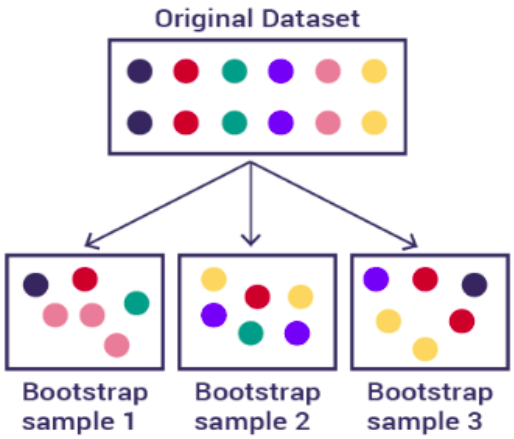


Figure 9: Bagging phase of the random forest algorithm

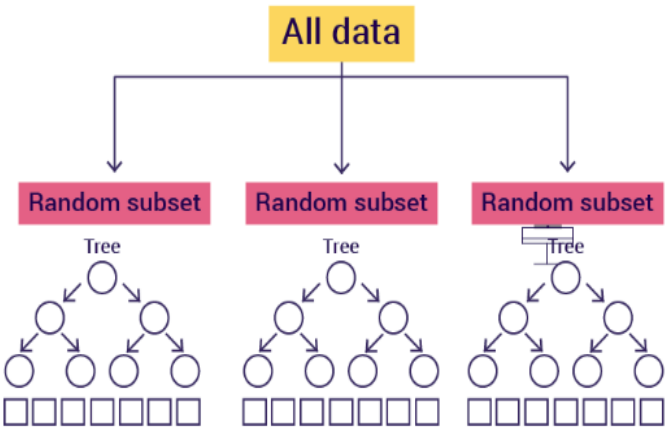


Figure 10: Random Forest Algorithm Bagging

3.3. Performance Evaluation Criteria

The evaluation of the performance of our forecasts is based on the criteria defined by equations (5), (6), (7) and (8), where N represents the total number of values contained in the data [16], [23]. These indices provide a basis for judging comparisons with a view to future model improvements. However, comparison between models remains complex due to differences in forecast horizons, number of input parameters and meteorological conditions. Nevertheless, the mean absolute error (MAE), as defined in equation (5), is particularly relevant for linear cost functions, providing a proportional measure of prediction errors. In contrast, the root-mean-square error (RMSE) (6) is more suitable for significant deviations between forecast and observation. On the other hand, the root mean square error (RMSE), as defined in equation (7), is very responsive to these

deviations, making it a valuable comparative parameter, particularly suitable for public applications [23]. It is worth noting that the lower the RMSE or MAE, the better the quality of the production forecast for our wind farm.

$$\frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_t\right| = \text{MAE} \quad (5)$$

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_t\right)^2} = \text{RMSE} \quad (6)$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_t\right)^2 = \text{MSE} \quad (7)$$

$$1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_t)^2}{\sum_{i=1}^{N} Y_i^2} = R^2 \quad (8)$$

All these equations evaluate the parameters used to measure the accuracy of the power predicted by the algorithms of the two models used.

3.4. Flowchart of the Artificial Intelligence Model Algorithm

The flowchart of the regression tree and forest type artificial intelligence algorithm is a representation of the sequences and decisions to be taken by the algorithm to predict numerical values of the wind production of the targeted Taïba Ndiaye. In this work, it is described as follows:

Begin

1. Enter the historical wind production data for Taïba Ndiaye.
2. Convert all data to hourly resolution by averaging.
3. Select the target variable to be predicted (energy produced per hour).
4. Apply WT decomposition (hierarchical multi-step decomposition) to historical target data (wind power).
5. Identify training (80%) and test (20%) data sets.
6. Verify tree convergence during model training.
7. Save the trees if the convergence condition is met (these saved trees are called Wind Production Forecasters).
8. If not, move on to the application of model hyper-parameters.
9. Recheck the convergence of the trees during training.
10. Save the trained shafts if the convergence condition is met.

4. Results and Discussion

TShort-term forecasting is of paramount importance in managing the distribution of wind generation throughout the year. It also offers managers the possibility of making real-time adjustments within the electricity network integrating intermittent renewable energies [24]. However, we have chosen to focus on the months of January and July, as they respectively encompass the most significant and least significant production of the year. The data was collected during this period. In fact, if the models manage to make a good prediction, then its generalisation to the other months of the year is quite obvious. Figures 11 and 12 illustrate the predictions generated by the two AI models for the month of January, when production rose. These graphical representations compare actual wind energy production with the one-hour forecasts. Indeed, a relevant method for evaluating the performance of a forecast consists of anticipating previously observed data based on the data that preceded it. By analysing these predictions for the month of January with the highest production provides an in-depth view of the models' ability to accurately anticipate variations in the wind power plant.

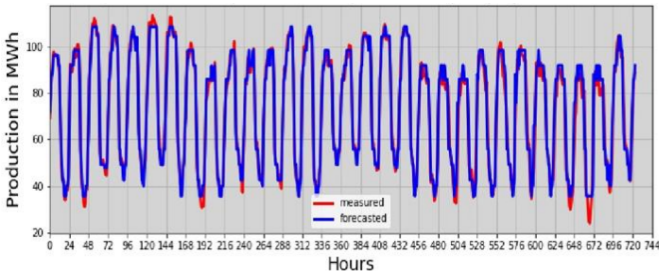


Figure. 11: Prediction in days for the month of January with

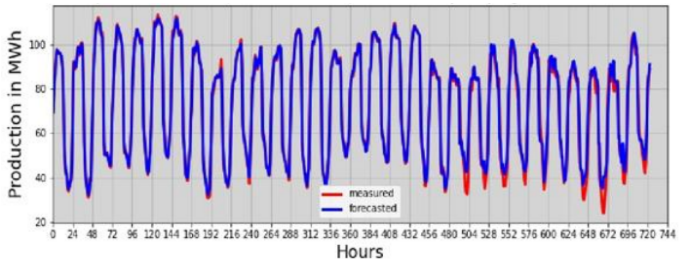


Figure. 12: Prediction in days for the month of January with the random forest.

The graphs above appear to show a potential correlation between the predictions (in blue) and the plant's actual output (in red) for the month of January, with RMSEs of 0.527 and 0.3332 Mwh/day respectively for the decision tree and random forest models. At the start of production, observations suggest that there may be occasional discrepancies between prediction and reality. These discrepancies are sometimes manifested by a much higher predicted production or, conversely, by an actual production at the lower limit of the prediction, for both models. It should be noted, for example, that except for day 27 (648 hours on the curve, Fig. 12), the observed values exceed the prediction of the random forest model from day 21 (504 minutes on the curve) to day 28 (672 hours on the curve), generally around 11pm. Despite a lower RMSE for the random forest model, these days show a better match between the predictions of the decision tree model and the actual observations (see Fig. 11). This

observation reveals some interesting nuances in the evaluation of performance with respect to the two models. This suggests that the RMSE metric alone may not fully capture the reliability of the models in specific situations. On the other hand, these differences do not appear uniformly for all the days of the month for which production is predicted. A general trend emerges, indicating that in most cases the random forest model provides more accurate forecasts than the decision tree model. It seems that the latter may be more effective in forecasting resources during low production hours. This could be attributed to the optimisation criterion favouring the homogeneity of the descendants with respect to the target variable. In other words, the variable tested in the node will be the one that maximises this homogeneity.

Furthermore, a particularly useful complementarity effect emerges in both models. During periods of increased production, the prediction of the random forest model stands out for its greater accuracy. The algorithm underlying the random forest performs its training on several trees formed from various subsets of data, thus conferring a complementarity that reinforces the effectiveness of the current hybrid models with better prediction. The coefficients of determination between the actual values and those predicted by the models reached 0.92 and 0.9382 respectively for the decision tree and the random forest during the month of January.

Fig 13 and 14 show a comparison between the values observed and predicted by the two models for the month of July. This is the month of the year when the plant supplies less energy to SENELEC. We also note that the predictions closely follow the actual production curve, with few systematic errors or apparent peaks (see Fig. 13 and Fig. 14). This consistency underlines the robustness of the models in predicting wind generation, irrespective of significant seasonal variations.

On the 26th day (624 hours of the curve) at around 11pm, a peak was observed for both models, although it did not affect the forecasts for the following hours.

The coefficients of determination were 0.76 for the decision tree and 0.794 for the random forest. These values are lower than those observed in January. When production falls, these coefficients show little variability, often attributable to unforeseen adverse weather conditions impacting production. It can sometimes be challenging to predict this with certainty.

A comprehensive examination of the error behaviour of each model over the month of July reveals slight differences (1.74 Mwh/m2/day for the tree model (see Fig.13) and 1.027 Mwh/m2/day (see Fig.14). These discrepancies can be attributed to the random nature of the seasonal variation in the study area and by the potential issue of underlearning.

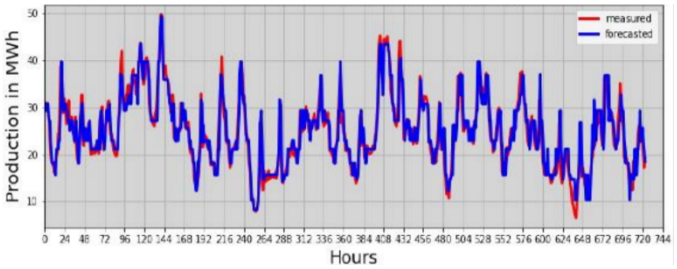


Figure. 13: Prediction in days for the month of July with the decision

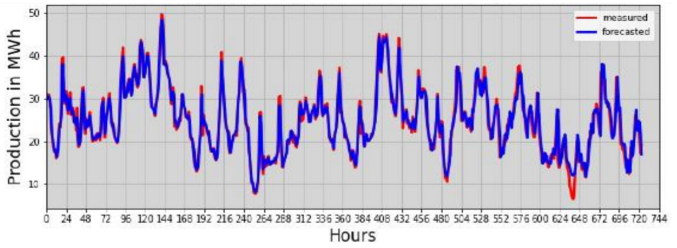


Figure. 14: Prediction in days for the month of July with the random forest.

The performance criteria, evaluated by the models [17] [25], [26] are used to examine the impact of the parameters and are applied to the test data to generalise the artificial intelligence models. In the study, a slight decrease in performance was observed for the different days of the predicted months. The summary of the performance parameters studied for the months of January and July are presented in Table II. The performance indices obtained are compared with those reported in the state of the art, with the aim of highlighting the limited number of input parameters used during model training. Despite this limitation, the learning techniques succeed in reducing the error, which illustrates the performance obtained. This performance is made possible in part by variations in tree depth.

Table 2: Comparison of performance indices

Model	*MAE*	*RMSE*	*R^2*	*Number of parameters*
Regression tree January (this work)	2.039	0.527	0.92	2
Regressive forest January (this work)	1.85	0.3332	0.938	2
Regression tree July (this work)	2.066	1.74	0.716	2
Regressive forest July (this work)	1.63	1.027	0.794	2
[17]	-	1.5	0.99	6

[25]	0.610	0.808	0.922	7
[26]	-	0.223	0.998	7

5. Conclusion

The strategy of increasing the share of renewable energies in the energy mix, while important for achieving sustainable development objectives, presents significant operational challenges. Indeed, this expansion leads to imbalances in the electricity network, causing excessive maintenance costs. Considering these challenges, it is becoming increasingly clear that accurate prediction of energy production is essential to guide decisions while anticipating operational requirements.

In order to achieve this goal, this article presents two artificial intelligence models, based on the decision tree and the random forest, with the intention of increasing the accuracy of forecasts for the Taïba Ndiaye power plant. The models were trained on the plant's production parameters over a one-year period. The results obtained demonstrate that, even in the absence of direct integration of meteorological parameters into the models, the proposed method allows for the robust prediction of wind power over a one-hour horizon. The coefficients of determination R2 were 0.92 and 0.938 respectively for the decision tree and random forest models. The root mean square error (RMSE) values of 0.3332 MWh and 0.527 MWh for the random forest model and decision tree respectively, reflect the considerable potential of AI models commonly referred to as machine learning in wind power forecasting. Overall, these results offer a promising prospect for optimising the penetration rate of intermittent energies such as wind power in the electricity grid. Nevertheless, we intend to utilise neural networks to enhance the plant's forecasts with the objective of further optimising the quality of the energy injected into the SENELEC electricity network.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] W. Zappa, M. Junginger, and M. van den Broek, "Is a 100% renewable European power system feasible by 2050?" *Applied Energy*, vol. 233–234, 1027–1050, 2019.

[2] A. Jamil, M. Imran, S. H. Ahmed, and A. Anpalagan, "Challenges and opportunities in wind power integration," *IEEE Access*, vol. 6, 63211-63233, 2018.

[3] A. Fischer, L. Montuelle, M. Mougeot, and D. Picard, "Statistical learning for wind power: A modeling and stability study towards forecasting," *Wind Energy*, vol. 20, no. 12, 2037-2047, 2017.

[4] M. Dieng, A. Ndoye, M. S. Sow, and A. Wane, "Energy consumption forecasting using artificial neural networks: A case study of Senegal," *2015 IEEE International Conference on Industrial Technology (ICIT)*, Seville, Spain, 2015, 2432-2437.

[5] K. Jha, & A. G. Shaik, "A comprehensive review of power quality mitigation in the scenario of solar PV integration into utility grid," *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 3, 100103, 2023. https://doi.org/10.1016/j.prime.2022.100103.

[6] A.S. Ba, (2018). "The energy policy of the Republic of Senegal.".

[7] S. Diop, P. S. Traore, & M. L. Ndiaye, "Power and Solar Energy Predictions Based on Neural Networks and Principal Component Analysis with Meteorological Parameters of Two Different Cities: Case of Diass and Taïba Ndiaye," *2022 IEEE International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, vol. 4, 1-6, IEEE, October 2022.

[8] S. Diop, P. S. Traore, and M. L. Ndiaye, "Wind Power Forecasting Using Machine Learning Algorithms," *2021 9th International Renewable and Sustainable Energy Conference (IRSEC)*, 2021, 1-6, doi: 10.1109/IRSEC53969.2021.9741109.

[9] M. Shafiullah, S. D. Ahmed, & F. A. Al-Sulaiman, "Grid Integration Challenges and Solution Strategies for Solar PV Systems: A Review," *IEEE Access*, vol. 10, 52233–52257, 2022. https://doi.org/10.1109/ACCESS.2022.3174555.

[10] A. Kaur, L. Nonnenmacher, and C. F. M. Coimbra, "Net load forecasting for a hybrid renewable energy system based on receding horizon optimization," *Energy*, vol. 150, 617-630, May 2018.

[11] J. Maldonado-Correa, M. Valdiviezo, J. Solano, M. Rojas, and C. Samaniego-Ojeda, "Wind energy forecasting with artificial intelligence techniques: A review," *International Conference on Applied Technologies*, December 2019, 348-362. Springer, Cham.

[12] S. Noman, A. A. Roya, K. Lauri, N. I. Muhammad, and R. Argo, "Forecasting Short Term Wind Energy Generation using Machine Learning," *Conference Paper*, October 2019, PSG 142. DOI: 10.1109/RTUCON48111.2019.8982365.

[13] T. Mahmoud, Z. Y. Dong, and J. Ma, "An advanced approach for optimal wind power generation prediction intervals by using self adaptive evolutionary extreme learning machine," *Renewable Energy*, vol. 126, 254–269, 2018.

[14] G. Notton, M. L. Nivet, C. Voyant, C. Paoli, C. Darras, F. Motte, and A. Fouilloy, "Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting," *Renewable and Sustainable Energy Reviews*, vol. 87, pp. 96-105, 2018.

[15] Y.-J. Ma, & M.-Y. Zhai, "A Dual-Step Integrated Machine Learning Model for 24h-Ahead Wind Energy Generation Prediction Based on Actual Measurement Data and Environmental Factors," *Applied Sciences*, vol. 9, no. 10, 2125, 2019. doi:10.3390/app9102125.

[16] W. M. Nkounga, M. F. Ndiaye, M. L. Ndiaye, O. Cisse, M. Bop, & A. Sioutas, "Short-term forecasting for solar irradiation based on the multi-layer neural network with the Levenberg-Marquardt algorithm and meteorological data: application to the Gandon site in Senegal," *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2018. doi:10.1109/icrera.2018.8566850.

[17] A. Clifton, L. Kilcher, J. K. Lundquist, & P. Fleming, "Using machine learning to predict wind turbine power output," *Environmental Research Letters*, vol. 8, no. 2, 024009, 2013. doi:10.1088/1748-9326/8/2/024009.

[18] M. F. Ndiaye, "Supervision de système multi-sources d'énergie," *Thèse Doctoral*, Université Cheikh Anta Diop de Dakar- École Supérieure Polytechnique, 2019.

[19] F. Pelletier, "Conception d'un site d'évaluation des performances d'éoliennes hors réseau en milieu complexe," *Doctoral dissertation*, École de technologie supérieure, 2003.

[20] R. Baïle, "Analyse et modélisation multifractales de vitesses de vent. Application à la prévision de la ressource éolienne," *Doctoral dissertation*, Université Pascal Paoli, 2010.

[21] S.A.A. Niang, M.S. Drame, A. Gueye, A. Sarr, M.D. Toure, D. Diop, S.O. Ndiaye, & K. Talla, "Temporal dynamics of energy production at the Taïba Ndiaye wind farm in Senegal," *Discov Energy*, vol. 3, art. 6, 2023. https://doi.org/10.1007/s43937-023-00018-0.

[22] T. R. Prajwala, "A Comparative Study on Decision Tree and Random Forest Using R Tool," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 1, January 2015.

[23] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, 569–582, 2017. doi: 10.1016/j.renene.2016.12.095.

[24] H. M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, 65-76, 2013.

[25] T. Brahimi, "Using Artificial Intelligence to Predict Wind Speed for Energy Application in Saudi Arabia," *Energies*, vol. 12, no. 24, 4669, 2019. doi:10.3390/en12244669.

[26] E. F. Alsina, M. Bortolini, M. Gamberi, and A. Regattieri, "Artificial neural network optimization for monthly average daily global solar radiation prediction," *Energy Conversion and Management*, vol. 120, 320–329, July 2016.

Sambalaye Diop - Data Scientist Engineer Data Science Certification Coach Trainer at GOMYCODE Renewable Energy and Environment Engineer PhD Student: Researcher at the Water, Energy, Environment and Industrial. Processes Laboratory (L3EPI) ESP-UCAD. Theme: Artificial Intelligence applied to renewable energies for high penetration in the electricity grid. Department: Electrical Engineering.

Papa Silly Traoré - PHD engineer in signal processing and embedded systems Co-ordinator of the professional degree course in Electrical Distribution and Automation (LPDEA) and the degree course in Electronics and Telecommunications Systems (EST) Lecturer-Researcher, Cheikh Anta Diop University (UCAD) Electrical Engineering Department of the Ecole Supérieure Polytechnique (ESP) Electrical Engineering Department of the Ecole Polytechnique Water, Energy, Environment and Industrial Processes Laboratory (LE3PI), Dakar, Senegal.

Mamadou Lamine NDIAYE
Full Professor (CAMES) - Cheikh Anta Diop University (UCAD)/ESP Electrical Engineering Design Engineer Doctorate in Signal Processing École Supérieure Polytechnique (ESP) -Département Génie Électrique- (UCAD) BP :5085 Dakar Fann

Issa Zerbo - Full Professor (CAMES) - Joseph KI-ZERBO University; Renewable Thermal Energy Laboratory (L.E.T.RE); Ouagadougou, Burkina Faso.
Doctor in Condensed Matter Physics
Director of the doctoral school université joseph ki-zerbo.

Received: 30 April 2024, Revised: 28 June 2024, Accepted: 30 June 2024, Online: 18 July 2024

DOI: https://doi.org/10.55708/js0307003

Comprehensive Analysis of Software-Defined Networking: Evaluating Performance Across Diverse Topologies and Investigating Topology Discovery Protocols

Nikolaos V. Oikonomou*[1], Dimitrios V. Oikonomou [2], Eleftherios Stergiou [1], Dimitrios Liarokapis [1]

[1]Department of Informatics & Telecommunications, University of Ioannina, Arta,47150, Greece

[2]Department of Regional & Cross Border Studies, University of Western Macedonia, Kozani,50100, Greece

*Corresponding author: Nikolaos V. Oikonomou, University of Ioannina Department of Informatics & Telecommunications, haikos13@gmail.com

ABSTRACT: Software-defined networking (SDN) represents an innovative approach to network architecture that enhances control, simplifies complexity, and improves operational efficiencies. This study evaluates the performance metrics of SDN frameworks using the Mininet simulator on virtual machines hosted on a Windows platform. The research objectives include assessing system performance across various predefined network topologies, investigating the impact of switch quantities on network performance, measuring CPU consumption, evaluating RAM demands under different network loads, and analyzing latency in packet transmission. Methods involved creating and testing different network topologies, including basic, hybrid, and custom, with the Mininet simulator. Performance metrics such as CPU and RAM usage, latency, and bandwidth were measured and analyzed. The study also examined the performance and extendibility of the OpenFlow Data Path (OFDP) protocol using the POX controller. Results indicate that balanced tree topologies consume the most CPU and RAM, while linear topologies are more efficient. Random topologies offer adaptability but face connection reliability issues. The POX controller and OFDP protocol effectively manage SDN network scalability. This research aims to analyze performance in a manner consistent with numerous previous studies, underscoring the importance of performance metrics and the scale of the network in determining the efficiency and reliability of SDN implementations. By benchmarking various topologies and protocols, the research offers a valuable reference for both academia and industry, promoting the development of more efficient SDN solutions. Understanding these performance metrics helps network administrators make informed decisions about implementing SDN frameworks to improve network performance and reliability.

Keywords: Network Architecture, Efficiency, SDN Controllers, Network Simulation, OpenFlow Protocol

1. Introduction

Networks are all around us, integral to our lives and daily routines. Most of the needs of a modern, technologically advanced society require strong and reliable networks. The demand for efficient networks is increasing exponentially with the passage of years and technological development. Nowadays, most people from all age groups use networks daily, and their quality of life depends on these networks, even if this is not immediately apparent. The COVID-19 pandemic that the world experienced from the beginning of 2020 changed many aspects of network usage. People were forced to spend much more time at home. This situation led people to find smart ways to meet most of their daily needs within the walls of their homes. Thus, the concept of the network in general came to everyone's doorstep in the form of the largest known global network, the internet. Young people had to be educated remotely using the internet. Adults were mostly required to work from a distance, and the elderly and vulnerable groups had to seek their care and support in a different way with the help of the internet. This situation led to an unprecedented increase not only in the number of internet users but also in the number of different devices each user employs to access it. As a result, some weaknesses in the existing global networks were revealed, and new ones were created. Network providers

and researchers realized that the capacity, speed, management, and reliability of networks needed to be increased due to the excessive load. Traditional networks that had been used for several years began to collapse because they primarily relied on physical infrastructures. Technologies such as SDN, which had been around for the last 10 years (mainly from 2013 onwards), began to be more extensively researched and implemented in global networks to strengthen them and ensure the smooth survival of the world. SDN enhances the capabilities of the network and simplifies its structure, with their main goal being more efficient network management. In this study, the architecture and structure of SDN networks will be examined. Network topologies will be created through simulations, and their operation and performance will be analyzed. Specifically, the performance between different topologies will be compared. It will be shown how the total number of switches, which are a fundamental pillar in the architecture of SDN, affects and burdens the overall performance of the network. The main measurements to be taken to draw accurate conclusions are CPU usage, RAM memory, and the delay in packet transfer between nodes. To achieve the above in the form of simulation, the Mininet simulator will be used on a computer with a Windows operating system. Additional software will be used in conjunction with Mininet to ensure the integrity and number of results. In this way, the behavior and adaptability of SDN networks will be studied. The controller used in Mininet will be POX, and further analysis will be done on the topology creation protocol OFDP, through which virtual networks will be studied below and their performance and scalability examined. The aim is to draw conclusions about the operation of SDN with the POX controller, specifically the use of the OFDP protocol. The reason for using random topologies is their effects on traditional networks and raises the question of how these random graphs can affect OpenFlow as its evolution into an even more modern network topology creation protocol. Given that network technologies have a strong relationship with network graphs and consequently with the concept of graph theory in mathematics, an analytical model for computation, comparison, and prediction is established through simulations on realistic platforms so that faster implementation in real-time networks can be achieved. The remainder of the article is as follows: Related search, SDN controllers, SDN protocols, Software & Hardware specifications, experiment specifications, analysis of results and finally the total conclusions from this research [1].

2. Related Search

Several research publications have been made on the aspect of SDNs, using the POX controller as well as OFDP for creating topologies. However, few utilize random topologies for interpreting SDN performance. In our previous research, we studied the performance results of Software-Defined Networking (SDN) tests conducted on standard network topologies using simulation. Concurrently, the performance of the standard topologies was compared with that of the random ones. Specifically, the performance measurements examined included: the setup and teardown time of the topology, the CPU and RAM usage of the system, and the delay in packet transfer between nodes. The entire study was conducted on a Windows computer using a virtual machine to run a Mininet simulator, similar to what we will use in the present work. From the meticulous analysis of the results, the following are worth mentioning: (i) the total number of switches in an SDN architecture has a significant impact on CPU load. (ii) RAM usage depends on the number of host computers and in cases of excessive load, it shows a much greater increase compared to CPU usage. (iii) The overall performance significantly depends on the type of topology and its properties. The experiments then were conducted on a typical and limited range of devices [2].

In his work, Guo created various types of network topologies for analysis, including ring topology, tree topology, and random Erdos-Renyi model topologies. In the randomly created networks, the probability of an edge between any two vertices was set at 0.4. Thus, these random networks were mostly dense networks with a short average path length. The ring networks had the longest average path length among the three topologies, while the tree topology was intermediate; all experiments were repeated 100 times. All nodes were subject to a common failure probability. Fifty nodes were used to create the three types of networks. The results show that the expected resilience of the network is inversely proportional to the average path length of the network topology, hence random topology networks perform better, and ring networks are less resilient [3].

In another study, the performance of the proposed discovery mechanism, which primarily relies on the OFDP protocol regarding the overall load, was analyzed. Various topologies were examined, focusing on random networks based on the Erdős–Rényi model. The study highlighted that researchers' efforts are concentrated on reducing the number of messages reaching the controller. However, the performance and scalability of SDN networks depend more on other factors, such as CPU load, memory usage, network topology, and the time required for topology discovery. The scalability of OFDP and OFDPv2 for a wide range of random networks based on the Erdős–Rényi model was tested. Experimental results showed that the protocols consume almost equal resources (CPU and RAM), while OFDP requires more

time for topology discovery than OFDPv2 for the same topology [4].

3. SDN Controllers

3.1. General SDN Information

Software-Defined Networking (SDN) is a networking approach that uses software-based controllers or APIs to communicate with the underlying hardware infrastructure and direct traffic within a network. This model differs from traditional networks, which solely utilize hardware devices (routers and switches) to manage network traffic. SDN can create and manage a virtual network or control a traditional network through software. While network virtualization allows the segmentation of different virtual networks on a single physical network and the connection of devices across various physical networks to form a single virtual network, SDN enables a new method of controlling data packet routing through a central server. Consequently, SDN achieves the successful and functional separation of the Control Plane from the Forwarding Plane in a network. Below in Figure 1 the SDN architecture is depicted.

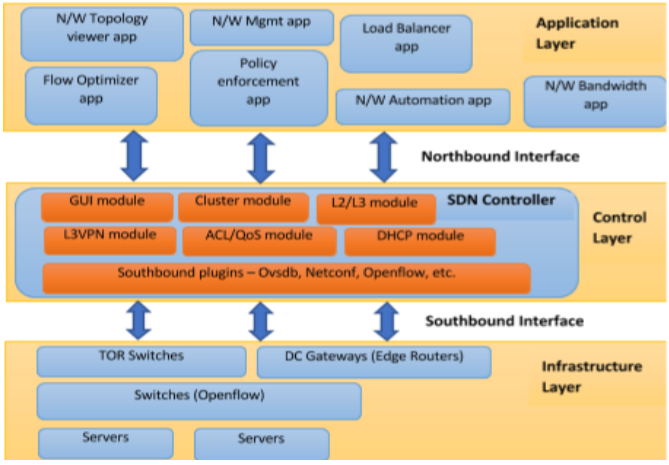


Figure 1: Schematic representation of the SDN architecture. Adapted from [5]

3.2. SDN Controller

A software-defined networking controller is a central element of the SDN architecture. It provides control over network elements in the managed domain. In networking, there are management, control, and data planes. An SDN controller offers management and control functions for network elements within the managed domain. This means that an SDN controller, based on network information and a set of predefined rules and policies, manages network elements and configures (or "programs") the data plane (i.e., directs data flow through the network). One of the key advantages of using an SDN controller is that it allows for more efficient network management, and changes to the network configuration can be applied from a central location instead of needing to manually configure each individual network element. Additionally, an SDN controller can automate certain tasks, such as traffic management and security, which can reduce the risk of human error and improve the overall reliability of the network. SDN controllers provide an API known as the northbound interface, through which external applications or systems such as orchestration platforms can interact with the network. In such cases, an SDN controller translates application-level requirements (e.g., high-level network configuration description) into configurations specific to the supported network elements. SDN controllers can manage both physical network devices and software elements that perform network functions [6].

In summary, the main functions of an SDN controller include:

- Managing data flow within the managed network
- Providing an API for applications and other components (e.g., orchestration platforms) to interact with the network.
- Providing visibility into the network, enabling network performance monitoring and troubleshooting
- Automating network management tasks, such as provisioning new network elements and reconfiguring network paths

More specifically, the controller provides the following capabilities:

Southbound Support: Defined as how a controller interacts with network devices to achieve optimized traffic flow. There are various southbound protocols that can be used, each with specific functionalities such as field matching, network discovery with different protocols, etc. When supporting the southbound interface, implementers must consider not only the characteristics of the protocol but also potential extensions, newer versions, etc.

Northbound Support: Northbound APIs are used for network integration and programming and can be utilized by orchestration systems that cater to customers and third-party applications. It is crucial to ensure that a controller is properly developed for orchestrating communications between layers. For example, the controller should support orchestration systems for applications such as cloud services, not only for open-source controllers and protocols but also those provided by various vendors. These applications could also include traffic engineering or applications that collect data used for network management tasks. As we can see below in Figure 2 the differences between the structure of

Centralized and Distributed control path are presented [7].

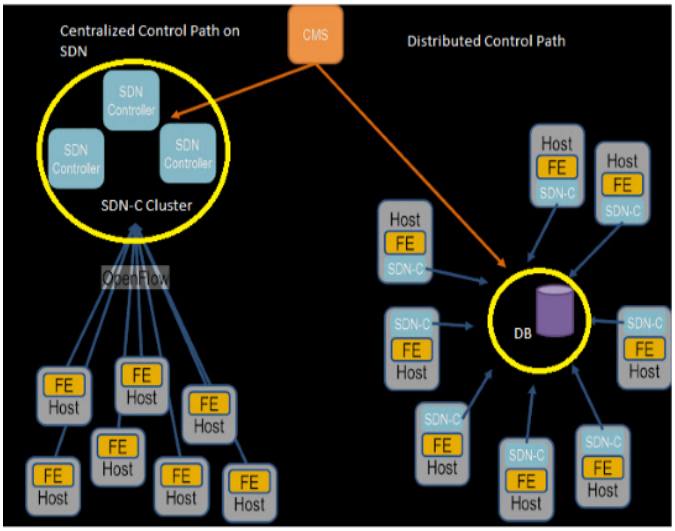


Figure 2: Comparison of Centralized and Distributed Architecture.

3.3. *NOX/POX Controller*

NOX was the first SDN controller. Initially developed by Nicira Networks, it was the first to support the OpenFlow protocol. Released to the research community in 2009, it laid the foundation for many SDN research projects. It was later expanded and supported at Stanford University with significant contributions from the University of California, Berkeley. Some popular NOX applications include SANE (an approach that represents the network as a file system) and Ethane. Today, NOX is considered inactive. Over the years, different versions of NOX have been introduced. These are known as NOX, NOX-MT, and POX. The new NOX only supports C++. It has a smaller application network compared to NOX but is much faster and has a much cleaner codebase. NOX-MT, introduced as a slightly modified version of the NOX controller, uses optimization techniques to introduce multi-threaded processing to improve the rate and response time of NOX. These optimization techniques include I/O batching to minimize general input/output overhead and others. POX is the latest version based on Python. The idea behind its development was to return NOX to its roots in C++ and develop a separate platform based on Python. It also features a Python OpenFlow interface, reusable element samples for path selection, topology discovery, etc. The primary goal of POX is research. Given that many research projects are by nature short-lived, the focus of POX developers is on good interfaces rather than API stability. In the current research, due to the multiple interfaces and the stability of the controller and because the Python language was used to create network topologies, POX was used [8].

Generally, the NOX controller provides a complete OpenFlow API using C++ and Python languages, uses asynchronous inputs/outputs (I/O), and is oriented towards operation on Linux, Ubuntu, and Debian systems. NOX is used both as a standalone controller and as a component-based framework for developing SDN applications. It is built on an event-based programming model and adopts a simple programming interface model that revolves around three pillars:

- Events
- Namespace
- Network view

Events can be generated either directly from OpenFlow messages or from NOX applications because of processing low-level events or other events generated by applications.

4. SDN Protocols

The SDN protocol is a set of standards and rules that define how SDN controllers and switches communicate with each other. Essentially, a protocol allows the SDN controller to configure the behavior of the switch, such as determining which packets should be forwarded to which ports and setting quality of service (QoS) parameters for different types of traffic. The most popular SDN protocol is OpenFlow.

4.1. OpenFlow Protocol

As mentioned, OpenFlow is the most widespread SDN protocol and defines the flow between the switch and the controller. It allows the controller to manage traffic forwarding between different network devices by controlling the switch's flow tables. This protocol was first developed by researchers at Stanford University in 2008 and was first adopted by Google in their backbone network in 2011-2012. It is now managed by the Open Networking Foundation (ONF). The latest version widely used in the industry is V1.5, while V2.0 is being refined. It is also often referred to as OFDP, meaning the OpenFlow Topology Discovery Protocol, because whether referred to as OpenFlow or OFDP, it automatically means the same function [9], [10].

OpenFlow is the standard southbound interface protocol used between the SDN controller and the switch. The SDN controller takes information from the applications and converts it into flow entries, which are fed into the switch via OpenFlow. It can also be used to monitor switch and port statistics in network management.

It is worth noting that the OpenFlow protocol is only installed between a controller and a switch. It does not affect the rest of the network. If a packet capture were to be taken between two switches in a network, both connected to the controller via another port, the packet capture would not reveal any OF messages between the switches. It is strictly for use between a switch and the controller. The rest of the network is not affected [11].

4.2. *NetConf Protocol*

NetConf is a protocol used in SDN for managing network devices such as routers and switches, providing a standardized way of configuring, monitoring, and managing these devices. It is an IETF standard and is based on XML data encoding and the SSH protocol for secure communication. The default TCP port assigned is 830. The NetConf server must listen for connections with the NetConf subsystem on this port.

With NetConf, network administrators can configure network devices programmatically using a standardized set of commands, rather than relying on proprietary interfaces for specific devices. This helps simplify network management and facilitates the automation of repetitive tasks such as deploying new network configurations or updates to hardware and software.

NetConf uses a client-server model, with the NetConf client sending requests to the device and the NetConf server responding with data or status updates. The protocol supports a range of functions, such as:

Retrieve: Retrieve specific data or configuration information from the device

Edit-config: Modify the device's configuration.

Commit: Apply changes to the device's configuration

Lock: Lock the device's configuration to prevent multiple managers from making conflicting changes

Unlock: Release the configuration lock

NetConf is often used in conjunction with YANG, a data modeling language that allows network administrators to describe network elements and their configurations in a structured and standardized way. Together, NetConf and YANG form a significant component of SDN, enabling greater automation and control in network management.

4.3. *Open vSwitch Database Management Protocol (OVSDB)*

OVSDB is a protocol used in SDN networks for managing Open vSwitch instances, which are software-based switches that can be used in an SDN environment. OVSDB provides a standard way to configure and manage Open vSwitch instances, allowing network administrators to deploy and manage switches more automatically and programmatically. It defines a set of functions that can be used for querying and modifying the configuration of Open vSwitch instances, including creating, deleting, and modifying ports, interfaces, and VLANs.

OVSDB is based on a client-server model, with the OVSDB client sending requests to the OVSDB server, which responds with data or status updates. The protocol uses JSON data encoding and supports secure communication using TLS.

In an SDN environment, Open vSwitch instances can be used to forward traffic between different network devices and allow network administrators to manage traffic flows using a central SDN controller. OVSDB provides a standardized way to configure and manage these switches, facilitating the development and management of large-scale SDN networks.

4.4. *Border Gateway Protocol (BGP)*

BGP is a routing protocol commonly used in SDN networking environments to exchange routing information between different autonomous systems in large-scale networks. BGP is a path vector routing protocol that uses a network of interconnected autonomous systems to route traffic between different parts of the network.

In an SDN environment, BGP can be used to facilitate communication between different network elements, such as the SDN controller and network devices, or between different SDN controllers. BGP provides a standardized way to exchange routing information, allowing network managers to manage traffic flows and optimize network performance.

BGP uses a hierarchical routing table system, with each autonomous system maintaining its own routing table and exchanging updates with neighboring autonomous systems. The protocol supports both internal and external routing, allowing more efficient routing within a single autonomous system and between different autonomous systems [12].

4.5. *Locator/Identifier Separation Protocol (LISP)*

LISP is a protocol used in SDN to separate the network location of a device (essentially its IP address) from its identity or identifier. LISP provides a way to

assign multiple IP addresses to a device and allows routing of traffic based on the device's identity rather than its physical location.

In an SDN environment, LISP can be used to simplify network management and enable more efficient routing of traffic between different devices. By separating the device's identity from its location, LISP allows network administrators to move devices between different network locations without changing their IP addresses, simplifying network management and reducing the likelihood of errors [13].

4.6. Simple Network Management Protocol (SNMP)

SNMP is a protocol used in software-defined networking (SDN) for monitoring and managing network devices. SNMP provides a standardized way for network administrators to monitor the performance of network devices, such as switches and routers, and to configure them remotely.

In an SDN environment, SNMP can be used to monitor and manage network devices from a central SDN controller. SNMP allows network administrators to monitor a range of device metrics, such as CPU usage, memory usage, and network traffic, and to receive alerts when performance issues arise.

SNMP is based on a client-server model, with SNMP agents running on network devices and SNMP managers running on the SDN controller. The agents collect performance data and send it to the managers, who can then analyze the data and take actions to improve network performance.

SNMP is a widely used protocol in network management and is supported by a range of network device suppliers and SDN controllers. It can be used in conjunction with all the above-mentioned SDN protocols to enable more effective and flexible network management.

4.7. Link Layer Discovery Protocol (LLDP)

LLDP is a Layer 2, vendor-neutral protocol used for discovering and advertising network device information on a local area network (LAN). It allows network devices to exchange information about their identity, capabilities, and connections [14].

LLDP operates by sending and receiving LLDP frames, which are multicast packets transmitted on every network interface. LLDP frames contain TLV (Type-Length-Value) elements that carry specific information about the transmitting device, such as system name, port description, system capabilities, and management addresses.

Key features and benefits of LLDP include:

Device Discovery: LLDP enables network devices to discover neighboring devices on the LAN, providing information about their identity, such as device type, vendor, and model.

Topology Discovery: By exchanging LLDP information, devices can gather details about the connections and topology of the network, including neighboring devices, port numbers, and connection speeds.

Automatic Configuration: LLDP can be used by network management systems to automatically configure network devices based on their discovered capabilities, simplifying network setup and reducing the efforts of manual configuration.

Troubleshooting and Monitoring: LLDP facilitates network troubleshooting by providing visibility into the network topology and device connectivity. It allows administrators to identify and locate devices, detect link failures, and monitor the status of connections.

LLDP is supported by a wide range of network devices, including switches, routers, wireless access points, and IP phones. It is often used in conjunction with other network protocols, such as SNMP, to enable comprehensive network management and monitoring.

It is important to note that LLDP is a Layer 2 protocol, and its functionality is limited to the local network segment. It does not route traffic nor provide visibility into the entire network.

4.8. Advantages of SDN

Software-defined networking (SDN) has emerged as a transformative approach to network architecture and management. By decoupling the control plane from the data plane and centralizing network control through software, SDN provides numerous benefits and impacts various industries. Key findings on SDN include:

- **Enhanced Network Flexibility**: SDN allows organizations to quickly provision, configure, and modify network services via software, leading to improved network flexibility. It enables dynamic allocation of network resources, making it easier to adapt to changing business needs and network traffic patterns [15].
- **Simplified Network Management**: SDN centralizes network management through a software-managed

controller, providing a single point of control and monitoring. This simplifies network management, reduces complexity, and enhances troubleshooting capabilities.

- **Scalability and Flexibility**: SDN offers scalability by abstracting network functionality from the underlying hardware. Organizations can more easily scale their networks by adding or reallocating resources according to needs. Furthermore, SDN allows flexibility in deploying new services and applications without significant changes to infrastructure.
- **Network Programmability**: SDN enables network programmability, allowing administrators to automate network functions and control network behavior through software. This programmability facilitates the development of innovative applications and services that can interact directly with the network.
- **Enhanced Security**: SDN provides enhanced security capabilities by leveraging centralized control and programmability. Security policies can be defined and enforced consistently across the network, making it easier to identify and respond to threats.
- **Cost Optimization**: SDN offers cost savings by reducing hardware dependencies and enhancing resource utilization. With the ability to dynamically control and distribute network resources, organizations can optimize their infrastructure, leading to better cost performance.
- **Innovation and Ecosystem Development**: SDN promotes innovation by enabling the development of new network services and applications. It encourages the development of an 'ecosystem' where vendors, developers, and researchers can collaborate to create new solutions and advance networking progress.
- **SD-WAN and Cloud Connectivity**: SDN plays a critical role in the adoption of software-defined wide area networks (SD-WAN) and in connecting on-premises networks to cloud environments. It simplifies the management of distributed networks, provides better visibility and control, and improves connectivity to cloud services.

4.8 challenges and issues

While SDN offers significant benefits, it also presents challenges, including interoperability among different SDN solutions, security concerns related to centralized control, the need for specialized personnel to manage and operate SDN environments, and the necessity for careful planning, testing, and collaboration with experienced vendors to overcome these challenges.

SDN Protocols:

- SDN protocols play a critical role in the implementation and operation of software-defined networking (SDN) environments. These protocols define the communication and interaction between different elements of an SDN architecture, facilitating network control and management.
- OpenFlow is one of the most widely adopted SDN protocols. It provides a standard interface between the control layer and forwarding devices (switches). OpenFlow enables centralized network control by separating control logic from switches and allowing the controller to program forwarding rules. It has significantly contributed to the development and deployment of SDN solutions.

SDN Controllers:

- SDN controllers serve as the central intelligence of software-defined network (SDN) architectures. They are responsible for managing and orchestrating network resources, facilitating communication between the control layer and the data layer, and enabling network programmability.

Table 1: below presents the network protocols along with their pros and cons.

Table 1: Network protocols.

PROTOCOLS	PROS	CONS
OpenFlow	Fully customizable, scalable	Complex
NetConf	Simplicity, management	Limited Performance
OVSDB	Customizable, management	Few complex options
BGP	Usable across different networks, routing	Recommended only for very large networks
LISP	Simplicity, efficient traffic control	Limited capabilities
SNMP	Advanced control	Complex
LLDP	Wide range of device compatibility	Limited only to LAN networks

5. Software & Hardware specifications

In this section, we will analyze each tool used for this work. Specifically, both the hardware and software components will be discussed.

5.1. Hardware Specifications

Compared to previous related research where high-performance laptops, low-performance desktops, or even workstations were used, this research utilized a new high-performance desktop computer. This system offers the capability to implement larger virtual networks as well as optimized management and distribution of physical resources, allowing for improved performance and more efficient scaling of the networks that will be created. In the heart of the computing system used for this research, the Gigabyte B550M AORUS PRO motherboard with an AMD B550 chipset lays the foundation. This motherboard was chosen for its robust support for modern connectivity standards such as PCI EXPRESS 4.0, which is pivotal for high-performance setups required in advanced simulations and experiments. The AMD Ryzen 5 5600X processor, featuring a 7nm FinFET technology with 6 cores and 12 threads, is selected for its ability to handle extensive computations more effectively than comparable models used in preceding studies. Its overclocking ability up to 4.7 GHz facilitates faster processing of complex tasks, crucial for developing larger virtual networks and conducting intensive data analysis.

Additionally, the system is equipped with 32GB of DDR4 RAM at 3600 MHz in dual-channel configuration, providing ample bandwidth and speed necessary for managing multiple operations simultaneously, which is essential when testing the limits of network simulations and other resource-intensive applications. The AMD Radeon RX 6750 XT graphics card with 12GB of GDDR6 memory ensures smooth rendering of complex graphics and supports the visualization demands of the research, including the manipulation and analysis of high-dimensional data sets.

Storage is handled by a Kingston KC3000 NVMe SSD with a capacity of 2TB, leveraging PCI Express 4.0 technology to offer rapid data access speeds of up to 7000 MB/s, significantly reducing load times and improving the overall efficiency of data processing tasks. This storage solution is vital for handling large volumes of data generated during simulations, ensuring quick retrieval and processing that are imperative for maintaining workflow continuity during the research.

Together, these hardware specifications are meticulously chosen not only for their individual capabilities but also for their synergy, which ensures a high-performance, stable, and reliable computing environment capable of supporting the sophisticated software tools and simulations utilized in this research. In Table: 2 we have the technical specifications of our systems.

Table 2: Simulation system specifications.

Component	Specification
CPU	AMD Ryzen 5 5600X, 6 cores/12 threads, 4.7 GHz, 45W
RAM	32GB DDR4, 3600 MHz, Dual Channel
GPU	AMD Radeon RX 6750 XT, 12GB, PCI Express 4.0
Storage	Kingston KC3000, NVMe, PCI Express 4.0, 7GB/s

5.2. Software Specifications

In this section, the specifications of the system software used are analyzed. It is crucial not only to conduct research to use the correct software that can deliver the desired results but also to ensure that all software can work harmoniously together. Cohesion, relevance, and repeated checks on the outcomes that will be extracted are necessary. For the software setup in this research, specific tools have been meticulously selected to complement the powerful hardware configuration and to meet the specialized requirements of the study. The primary operating system used is Windows 11 Pro for Workstations, which offers essential features like the ReFS file system for enhanced data resilience and support for advanced hardware configurations, critical for maximizing the potential of the system's physical components.

Oracle's VirtualBox plays a key role by allowing the deployment of multiple operating systems on a single physical machine, which is crucial for testing different network configurations and software interactions in a controlled, isolated environment. This flexibility is vital for reproducing and manipulating network scenarios in the development of software-defined networking (SDN) solutions.

Additionally, Visual Studio Code is employed as the primary code editor due to its robust support for multiple programming languages and its integrated development environment (IDE) features like debugging, code completion, and Git integration. These features enhance the efficiency of writing and testing code, particularly Python scripts used for creating network topologies in the research.

Gephi, an open-source network visualization software, is used to analyze and visualize complex network structures, which helps in understanding the interactions within the network and identifying key patterns and anomalies. The ability to dynamically model network traffic and topology changes in real-time using Gephi significantly aids in the exploratory phase of the research.

Furthermore, the inclusion of specialized tools like PuTTY for secure remote session management, WinSCP for secure file transfer, and Xming for running X Window System applications on Windows, consolidates the software environment.

Together, these software tools form a cohesive ecosystem that supports the rigorous demands of the research, enabling sophisticated simulations, extensive data analysis, and effective management of resources across different stages of the project. Table 3 contains an analysis of all the software used.

Table 3: Simulation software presentation.

Software	Brief Description
Windows 11 Pro for Workstations	Operating system designed for high-tech hardware and workloads, with additional features for enhanced performance and reliability.
VirtualBox	Open-source virtualization software that allows running multiple operating systems on a single physical machine.
Mininet	Network emulator that facilitates the simulation and testing of Software-Defined Networks (SDN).
X-Ming	Free X-Window-System server for Windows that enables remote graphical user interfaces over a network.
WinSCP	Free and open-source SFTP, FTP, and SCP client for Windows that enables secure file transfers between local and remote computers.
PuTTY	Free terminal emulator, serial console, and network file transfer application for Windows that supports multiple network protocols.
Visual Studio Code	Free, open-source code editor developed by Microsoft, supporting a wide range of programming languages and tools.
Gephi	Open-source software for visualizing and exploring graphs and networks, ideal for analyzing complex networks.

6. Experiment Specifications

6.1. Network Topologies

The term topology defines the geometric representation of the connections in a network. We examined three categories of topologies.

- Basic
- Hybrid
- Custom

Specifically, for the basic topologies, the bus topology was selected, for the hybrid topologies, the balanced tree topology was chosen, and for the Custom, the random topology was used [16], [17].

6.1.1. Basic Topologies

There are many basic network topologies commonly used in computer networking. These include:

Bus topology: All devices are connected to a single communication line or cable, known as the bus. Data travels in both directions along the bus and all devices on the network can receive the same message simultaneously. Figure 3 depicts bus topology.

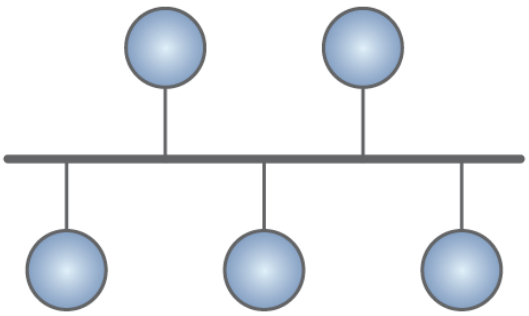


Figure 3: Example of bus topology.

Star topology: All devices are connected to a central hub or switch, and data flows through the hub or switch to reach its destination. Each device has an exclusive connection to the hub or switch, which can help reduce

network congestion and improve performance. Figure 4 depicts star topology.

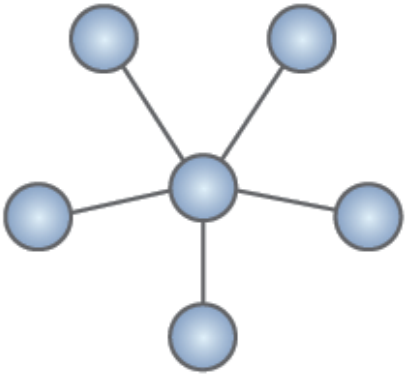


Figure 4: Example of star topology.

Ring topology: All devices are connected in a closed loop, with data flowing in one direction around the loop. Each device receives data from the previous device in the loop and sends data to the next device in the loop. The ring topology is depicted in Figure 5 below.

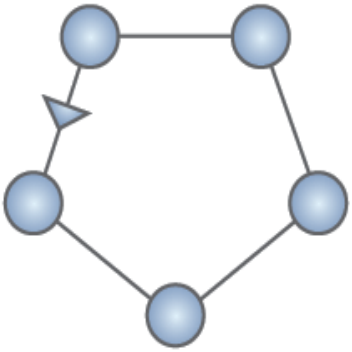


Figure 5: Example of ring topology.

Mesh topology: Each device is connected to every other device in the network, creating a fully interconnected network. This can provide high redundancy and fault tolerance but can be complex to manage and requires a lot of wiring. Figure 6 presents mesh topology.

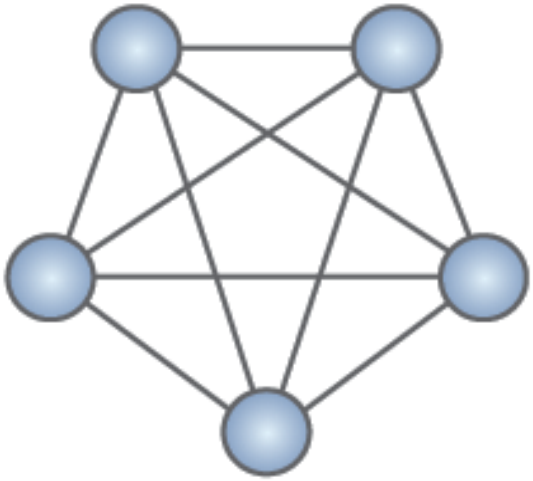


Figure 6: Example of mesh topology.

6.1.2. Hybrid Topologies

Hybrid Topology is a combination of two or more basic topologies, such as a star-bus topology or a ring-mesh topology. This can offer a balance between performance, redundancy, and ease of management.

Tree topology, also known as hierarchical topology, is a type of network topology based on a hierarchical structure. In this topology, multiple star topologies relate to a bus topology, creating a structure that resembles a tree. In a tree topology, the central bus acts as the main trunk of the tree, with multiple branches extending from it. Each branch is a separate star topology with a hub or switch at the center and multiple devices connected to it. This allows the creation of subnetworks within the larger network, with each subnet having its own exclusive hub or switch.

The main advantage of a tree topology is its scalability, as it can support many devices and subnetworks. It also provides a good balance between performance and redundancy, as each subnet can operate independently and problems in one subnet will not affect the rest of the network.

However, the main disadvantage of a tree topology is its complexity, as it requires a significant amount of cabling and configuration. It can also be difficult to troubleshoot and manage, as problems in one part of the network can affect the entire tree. Below in Figure 7 an example of hybrid topology can be found.

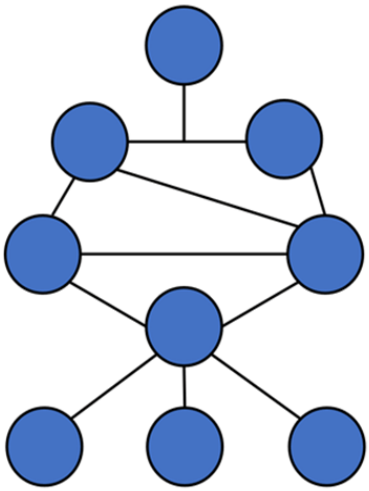


Figure 7: Example of hybrid tree topology.

A balanced tree topology is a specific type of tree topology where each branch of the tree has the same number of levels. This means that each subnet is of equal size and has the same number of devices connected to it. In a balanced tree topology, the central bus is connected to a set of level-1 switches, each of which is connected to

a set of level-2 switches, and so on, until the final level of switches is reached. Each switch in the tree has an equal number of branches connected to it, which helps balance the network traffic and avoid congestion.

6.1.3. *Custom Topologies*

Custom network topologies refer to network architectures designed to meet specific requirements or solve specific problems. They may be a combination of two or more basic topologies, or they may be entirely unique and tailored to a specific application or environment. Custom network topologies can be created by network designers and administrators using various networking devices and technologies, such as switches, routers, firewalls, load balancers, and others. These devices can be configured to implement specific routing protocols, VLANs, access control policies, and other features to achieve the desired network behavior and performance. Examples of custom network topologies include:

Mesh topology with adaptive routing: This topology can be used in large-scale wireless networks to provide high redundancy and fault tolerance. Adaptive routing protocols such as OLSR or B.A.T.M.A.N. may be used to optimize network performance and reduce congestion. Hub-and-Spoke topology with VPN: This topology can be used to connect multiple remote offices or branches to a central location using VPN tunnels. A hub router or firewall is used to manage the traffic flow and provide secure connectivity between the spokes.

Cluster topology with load balancing: This topology can be used to create a cluster of web or application servers for high availability. Load balancing devices are used to distribute traffic across multiple servers in the cluster, providing high performance and scalability.

Custom network topologies can offer unique advantages and solve specific problems, but they also require careful design and management to ensure effectiveness and security. Network administrators should consider the specific needs of their organization and consult experienced network designers to create a custom topology that meets these needs.

6.1.4. *Random Topology*

In computer networking, a random network topology refers to a network topology where connections between nodes are made in a random or stochastic manner. In such a topology, there is no predetermined plan or structure to the connections between nodes. Random network topologies are used in various applications, such as in the study of social networks, biological networks, and communication networks. They are also used in analyzing network properties, such as connectivity, robustness, and efficiency. It has been shown that they exhibit some interesting and unexpected behaviors, such as the emergence of small-world networks and scale-free networks.

6.1.5. *Erdős–Rényi Model*

An example of a random network topology is the Erdős–Rényi model. The Erdős–Rényi model, also known as the ER model, is a mathematical model for creating random graphs. Introduced to the field of mathematics by mathematicians Paul Erdős and Alfréd Rényi in 1959, the ER model creates a random graph with "n" nodes starting with "n" isolated nodes and then randomly connecting pairs of nodes with a certain probability "p". The edges between the nodes are independent and occur with probability "p". There are two variations of the ER model: the G(n,m) model, which creates a random graph with "n" nodes and m edges, and the G(n,p) model, which creates a random graph with "n" nodes and an edge between each pair of nodes with probability "p".

The ER model has been used to study various properties of random graphs, including the appearance of the giant component, the phase transition of connectivity, and the degree distribution of the graph. The model has also been applied in various fields such as social networks, computer networks, and biology. However, it should be noted that the ER model assumes a completely random and uniform distribution of edges, which may not always reflect the real structure of many networks. As a result, other network models, such as small-world networks and scale-free networks, have been proposed to better map the properties of real networks. In Figure 8 below we can see a custom random topology is presented [18].

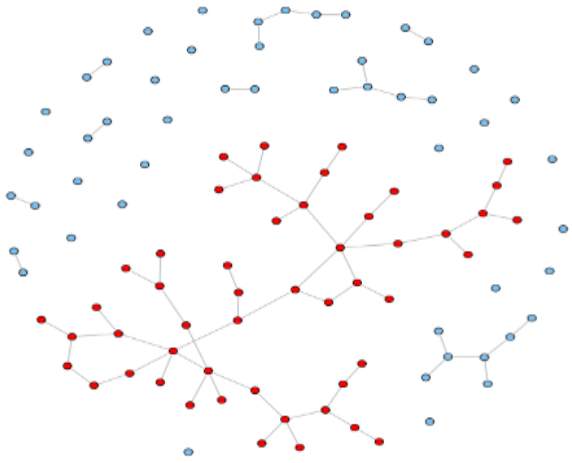


Figure 8: Example of Custom Random Topology using the Erdős–Rényi mathematical model

6.2. Experiment Specifications

The experiments implement topologies based on the random topology which in turn follows the Erdős–Rényi mathematical model. The SDN controller used is POX due to its compatibility with both topologies and the creation and parameterization of topologies through the PYTHON language. The topology creation protocol is OFDP, or otherwise OpenFlow [19].

The experiments examine the following:

- Comparison of system performance according to topologies.
- Comparison of system performance according to topology creation protocol.
- Comparison of system performance according to the number of switches and how the total number of switches affects performance.
- CPU usage.
- RAM usage
- The delay of packet transfer between network nodes.
- The time of creation and destruction of a topology

The above measurements will be compared:

Topologies:

- Linear
- Balanced Tree
- Random Topology

Creation Protocols

- OFDP
- LLDP
- BGP
- LSDP
- SNMP
- OVSDB

The number of switches will remain steadily increasing, and each switch will be connected to a host in the manner shown in the table below [20].

It is noted that a greater number of switches was achieved than in most similar studies. This fact alone allows for better interpretation of results and is primarily due to the available hardware resources. Table 4 contains the scale of the experiments depending on the number of switches and hosts.

Table 4: Scale of experiments conducted

SWITCHES	HOSTS
2	2
4	4
8	8
16	16
32	32
64	64
128	128
256	256
512	512
1024	1024
2048	2048
4096	4096
8192	8192

6.3. Collection of General Results

In this section, the statistical tables of the data collected from the above experiments will be presented. The controller used is POX and the topology creation protocol is OFDP. It is worth noting that each experiment was performed about a thousand times to ascertain the accuracy percentage of the results, and the deviations were minimal and consistent with the expected pattern. Therefore, the results presented are the overall average. Below the tables of experiment results are presented. Table 5 presents the results using random topologies, Table 6 presents the results using linear topologies and Table 7 presents the results using balanced tree topologies [21].

Table 5: Experiment results using random topologies.

CPU (%)	MEMORY (MB)	SWITCH	HOSTS	BW (Gbps)	SETUP TIME (sec)	TEAR TIME (sec)
1.9	150	2	2	41	0.092	0.085
2.6	170	4	4	42	0.145	0.136
6.4	210	8	8	48	0.326	0.413
11.2	250	16	16	48	1.256	1.646
13.5	290	32	32	47	2.719	6.167
18.1	330	64	64	38	12.752	22.39
22.4	390	128	128	38	18.393	29.712
27.8	440	256	256	36	26.715	39.513
33.6	625	512	512	33	39.212	58.004
39.2	1100	1024	1024	32	57.454	74.981
44.5	2000	2048	2048	28	83.757	119.046
47.3	3500	4096	4096	26	183.908	244.901
62.9	6800	8192	8192	27	274.483	368.271

Table 6: Experiment results using linear topologies.

CPU (%)	MEMORY (MB)	SWITCH	HOSTS	BW (Gbps)	SETUP TIME (sec)	TEAR TIME (sec)
1.2	300	2	2	45	0.098	0.067
1.9	340	4	4	44	0.182	0.224
4.8	380	8	8	49	0.295	0.313
9.3	420	16	16	47	0.542	0.621
10.3	480	32	32	48	0.894	1.128
14.5	560	64	64	481	1.889	2.359
19.3	680	128	128	38	3.319	4.858
23.6	1050	256	256	38	6.822	7.254
28.1	1680	512	512	39	14.841	18.952
33.6	2200	1024	1024	37	33.713	39.701
38.4	3500	2048	2048	36	55.915	62.113
42.5	7300	4096	4096	33	98.009	127.989
53.6	12300	8192	8192	31	181.411	229.410

Table 7: Experiment results using balanced tree topologies.

CPU (%)	MEMORY (MB)	SWITCH	HOSTS	BW (Gbps)	SETUP TIME (sec)	TEAR TIME (sec)
3.2	180	2	2	43	0.150	0.141
4.5	220	4	4	42	0.265	0.181
8.9	270	8	8	38	0.429	0.284
14.7	380	16	16	44	1.854	1.678
17.6	490	32	32	48	3.535	3.280
21,2	600	64	64	41	6.614	7.252
24.8	710	128	128	43	8.325	10.053
29.1	930	256	256	40	17.783	19.993
36.1	1450	512	512	39	26.977	41.900
44.9	2580	1024	1024	41	56.672	77.451
52.4	4310	2048	2048	37	128.334	168.513
59.9	8200	4096	4096	38	190.985	212.717
74.4	15200	8192	8192	30	260.511	332.557

6.3.1. *Collection of Latency Results*

Latency measurement will be done differently as each network is measured under similar conditions with a fixed packet size, increasing the number of packets and observing how this affects the network. The average and total transfer times are collected. The size of each packet is defined as 1024Bytes (1KB), and simulations will be executed with the corresponding packet numbers [1,10,50,100,500]. In previous studies, a usage limit of about 600 packets of this packet size was observed in Mininet. Below Table 8 is presented in which we can see the latency results of each topology.

Table 8: Latency results of experiments across all topologies.

PACKET NUMBER	TREE AVERAGE LATENCY (ms)	LINEAR AVERAGE LATENCY (ms)	ERDOS RENYI AVERAGE LATENCY (ms)
1	0.048	0.018	0.013
10	0.053	0.027	0.016
50	0.044	0.026	0.018
100	0.031	0.023	0.021
500	0.041	0.015	0.022
TOTAL AVERAGE	0.0434	0.0218	0.0181

7. Analysis of results

The results obtained in the present research are appropriately transformed into diagrams. On the vertical axis, each studied element (CPU, RAM, Bandwidth, Setup Time, Tear Time, Latency) is distributed, while on the horizontal axis there is the number of switches used, thus conclusions are drawn based on the quantity of Switches. In the Latency diagram, on the horizontal axis, the number of Switches is replaced by the number of packets.

7.1. *CPU Analysis*

The following section provides an in-depth analysis of CPU usage in relation to the number of switches in various network topologies. The data illustrates that CPU usage increases as the number of switches rises. The analysis is based on comparative data from three topologies: balanced tree, random, and linear.

7.1.1. *Balanced Tree Topology*

Highest CPU Consumption: The balanced tree topology consumes the most CPU resources among the three topologies. This is attributed to the complexity and structure of the tree-branches, which require more processing power to manage the available paths.

CPU Usage Increases with Switches: As the number of switches increases, CPU usage significantly rises. At the peak of 8192 switches, the CPU usage reaches 74.4%, which is 11.5% higher than the random topology and 20.8% higher than the linear topology.

7.1.2. *Random Topology:*

Close to Balanced Tree: The random topology's CPU consumption is slightly less than the balanced tree but higher than the linear topology. This is due to the adaptable nature of the random topology, which requires complex computations to manage dynamic connections.

Instabilities: Some instabilities in CPU usage are observed, caused by the probability of unsuccessful connections between nodes, which adds variability to the CPU load.

7.1.3. *Linear Topology*

Least CPU Usage: The linear topology demonstrates the least CPU usage due to its simple and straightforward

connectivity. The simplicity of managing linear connections results in lower processing requirements.

Stable Performance: The linear topology shows stable CPU performance, with less variability and lower overall CPU consumption compared to the other topologies.

Complexity and Resource Allocation: The balanced tree topology requires more CPU resources due to its hierarchical structure. Managing multiple levels and branches in the network involves more processing to maintain efficient routing and data flow. This complexity inherently increases the CPU load as the network scales.

Adaptability of Random Topology: While the random topology is designed for flexibility and adaptability, this also introduces challenges in maintaining stable connections. The CPU must handle dynamic routing and potential connection failures, leading to increased CPU usage and occasional spikes.

Efficiency of Linear Topology: The linear topology benefits from its simplicity, where each switch is directly connected in a straightforward path. This minimizes the processing required for routing decisions, leading to lower and more consistent CPU usage. The linear approach simplifies network management and reduces the computational burden on the CPU.

The analysis highlights that network topology significantly impacts CPU usage. The balanced tree topology, while offering robust and hierarchical structuring, imposes a high CPU load due to its complexity. The random topology, though adaptable, faces challenges with connection stability, leading to variable CPU consumption. Linear topology remains the most efficient in terms of CPU usage, owing to its simple and direct connectivity. These findings are crucial for network administrators and designers, emphasizing the need to consider topology choice based on the expected network load and performance requirements. Balancing complexity, adaptability, and efficiency is key to optimizing network performance and resource utilization. Figure 9 analyses the CPU usage in each experiment.

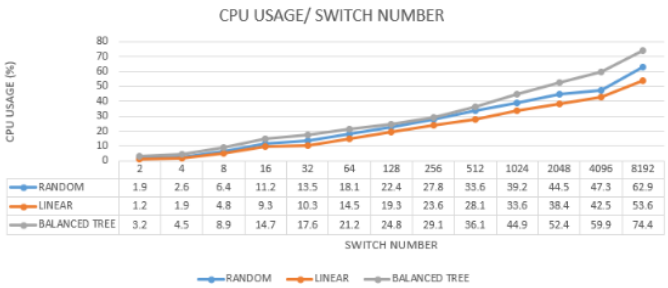


Figure 9: Comparative CPU usage diagram for the three topologies.

7.2. RAM Analysis

The following section provides an in-depth analysis of RAM usage in relation to the number of switches in various network topologies. The data illustrates that RAM usage varies significantly with the topology used and the number of switches in the network. This analysis is based on comparative data from three topologies: linear, balanced tree, and random.

7.2.1. Linear Topology

Initial High Memory Consumption: Initially, the linear topology consumes more RAM compared to the other two topologies, despite using less CPU than the balanced tree. This higher initial memory usage can be attributed to the straightforward but memory-intensive nature of maintaining direct connections between each switch.

Memory Usage Trends: As the number of switches increases, the memory usage grows but at a predictable and steady rate due to the simple structure of the linear topology.

7.2.2. Balanced Tree Topology

High Memory Consumption with Increased Switches: While the number of switches increases, the balanced tree topology eventually consumes the most memory. This is due to the complexity of managing a hierarchical tree structure, which requires more memory to store the state and routing information for multiple levels and branches.

Complexity Impact: The tree-branch structure inherently requires more memory to maintain the hierarchical relationships and efficient routing, resulting in higher memory usage as the network scales.

7.2.3. Random Topology

Lowest Memory Consumption: The random topology consistently shows lower RAM usage compared to the other two topologies. This is largely due to its customization and the retrospective improvements made to its implementation in Mininet, which optimize memory usage.

Efficiency of Customization: Due to its adaptable nature and optimized design, the random topology reduces memory consumption by about 45-55% compared to the linear and balanced tree topologies.

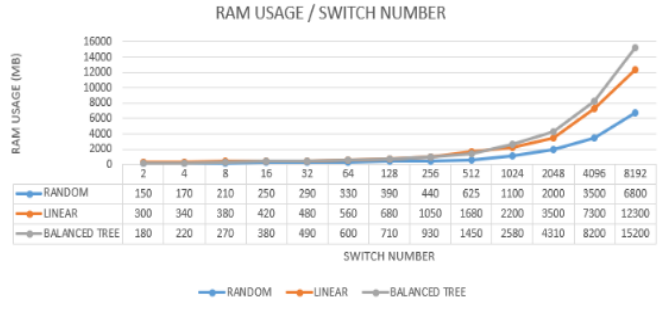
Initial Memory Usage in Linear Topology: The linear topology, despite its simplicity, requires substantial memory initially to establish and maintain direct

connections between each switch. This direct approach, while less CPU-intensive, places a higher initial burden on RAM.

Increasing Complexity in Balanced Tree Topology: As the network grows, the balanced tree topology's memory requirements increase significantly. This is because the hierarchical structure demands more memory to store the details of each level and branch, ensuring efficient data routing and network management.

Optimized Memory Usage in Random Topology: The random topology benefits from its customized and optimized implementation in Mininet. This design reduces unnecessary memory usage and streamlines the management of random connections, leading to significantly lower RAM consumption. The flexibility and adaptability of the random topology also contribute to its efficient memory usage.

The analysis highlights that network topology significantly impacts RAM usage. The linear topology, while simple, initially demands more memory but grows predictably. The balanced tree topology, due to its hierarchical structure, consumes the most memory as the network expands. The random topology, with its optimized and adaptable design, demonstrates the most efficient memory usage. These insights are crucial for network administrators and designers, emphasizing the need to consider topology choice based on the expected network load and performance requirements. Balancing complexity, adaptability, and efficiency is key to optimizing network performance and resource utilization, particularly in terms of memory usage. Figure 10 analyses the RAM usage in each experiment.



	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192
RANDOM	150	170	210	250	290	330	390	440	625	1100	2000	3500	6800
LINEAR	300	340	380	420	480	560	680	1050	1680	2200	3500	7300	12300
BALANCED TREE	180	220	270	380	490	600	710	930	1450	2580	4310	8200	15200

Figure 10: Comparative RAM usage diagram for the three topologies.

7.3. Bandwidth Analysis

The following section provides an in-depth analysis of RAM usage in relation to the number of switches in various network topologies. The data illustrates that RAM usage varies significantly with the topology used and the number of switches in the network. This analysis is based on the comparative data from three topologies: linear, balanced tree, and random.

7.3.1. Linear Topology

Initial High Memory Consumption: Initially, the linear topology consumes more RAM compared to the other two topologies, despite using less CPU than the balanced tree. This higher initial memory usage can be attributed to the straightforward but memory-intensive nature of maintaining direct connections between each switch.

Memory Usage Trends: As the number of switches increases, the memory usage grows but at a predictable and steady rate due to the simple structure of the linear topology.

7.3.2. Balanced Tree Topology

High Memory Consumption with Increased Switches: While the number of switches increases, the balanced tree topology eventually consumes the most memory. This is due to the complexity of managing a hierarchical tree structure, which requires more memory to store the state and routing information for multiple levels and branches.

Complexity Impact: The tree-branch structure inherently requires more memory to maintain the hierarchical relationships and efficient routing, resulting in higher memory usage as the network scales.

7.3.3. Random Topology

Lowest Memory Consumption: The random topology consistently shows lower RAM usage compared to the other two topologies. This is largely due to its customization and the retrospective improvements made to its implementation in Mininet, which optimize memory usage.

Efficiency of Customization: Due to its adaptable nature and optimized design, the random topology reduces memory consumption by about 45-55% compared to the linear and balanced tree topologies.

Initial Memory Usage in Linear Topology: The linear topology, despite its simplicity, requires substantial memory initially to establish and maintain direct connections between each switch. This direct approach, while less CPU-intensive, places a higher initial burden on RAM.

Increasing Complexity in Balanced Tree Topology: As the network grows, the balanced tree topology's memory requirements increase significantly. This is because the hierarchical structure demands more memory to store the

details of each level and branch, ensuring efficient data routing and network management.

Optimized Memory Usage in Random Topology: The random topology benefits from its customized and optimized implementation in Mininet. This design reduces unnecessary memory usage and streamlines the management of random connections, leading to significantly lower RAM consumption. The flexibility and adaptability of the random topology also contribute to its efficient memory usage.

The analysis highlights that network topology significantly impacts RAM usage. The linear topology, while simple, initially demands more memory but grows predictably. The balanced tree topology, due to its hierarchical structure, consumes the most memory as the network expands. The random topology, with its optimized and adaptable design, demonstrates the most efficient memory usage. These insights are crucial for network administrators and designers, emphasizing the need to consider topology choice based on the expected network load and performance requirements. Balancing complexity, adaptability, and efficiency is key to optimizing network performance and resource utilization, particularly in terms of memory usage. Figure 11 analyses the bandwidth of each experiment.

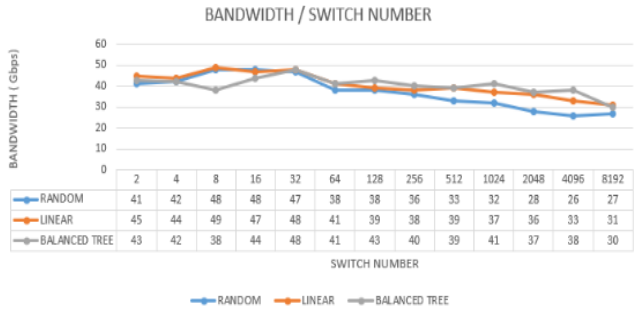


Figure 11: Comparative Bandwidth diagram for the three topologies.

7.4. Setup Time Analysis

The setup time refers to the duration required to create a network topology, measured from the moment the creation command is initiated. The analysis compares the setup times for three different network topologies: random, balanced tree, and linear. The results indicate significant differences in the time taken to establish each topology, highlighting the efficiency and complexity involved in their creation.

7.4.1. Random Topology

Longest Setup Time: The random topology consistently shows the longest time required to create a topology. This is due to its inherent complexity and the need for random connections between nodes, which involves additional computational overhead to ensure successful creation and connectivity.

Marginally Longer: Among the topologies with long setup times, the random topology takes slightly longer than the balanced tree, indicating higher variability and complexity in establishing random connections.

7.4.2. Balanced Tree Topology

Long Setup Time: The balanced tree topology also exhibits a long setup time, slightly less than the random topology. The hierarchical structure requires careful planning and execution to ensure all branches and levels are correctly established, which adds to the setup time.

Complexity Contribution: The structured nature of the balanced tree, with multiple levels and branches, contributes to the extended time needed for its creation.

7.4.3. Linear Topology

Shortest Setup Time: The linear topology shows a significantly reduced setup time compared to the other two topologies. This is due to its straightforward design, where each node is directly connected to the next in a simple chain.

Efficiency in Large Networks: In very large networks, the linear topology is approximately 30-40% faster to set up than the random and balanced tree topologies. This efficiency is attributed to the minimal complexity in establishing direct connections sequentially.

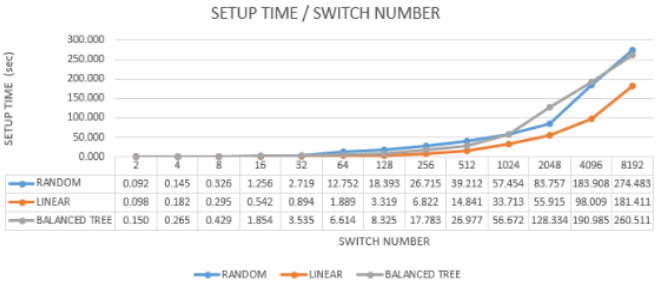
Complexity and Overhead: The random and balanced tree topologies require more time to create due to their inherent complexity. Random topology involves the creation of non-deterministic connections that need verification and correction, while the balanced tree requires a hierarchical setup with multiple levels, each adding to the overall setup time.

Linear Topology Efficiency: Linear topology's setup process is inherently simpler. Each new node is added in a straightforward manner, reducing the time required for planning and establishing connections. This simplicity translates to a significant reduction in setup time, especially as the network scales.

Scalability and Performance: As the network size increases, the difference in setup times becomes more pronounced. The linear topology's efficient setup process becomes increasingly advantageous in larger networks, where the time savings are substantial compared to the more complex topologies.

Practical Implications: For practical applications, the choice of topology can significantly impact the time required to deploy a network. In scenarios where rapid deployment is critical, the linear topology offers a clear advantage. Conversely, if the network's structural complexity and adaptability are priorities, the additional setup time for random or balanced tree topologies may be justified.

The setup time analysis underscores the importance of topology selection based on deployment time requirements and network complexity. The linear topology offers the fastest setup time, making it suitable for scenarios requiring quick deployment and straightforward management. The random and balanced tree topologies, while taking longer to set up, provide more complex and potentially more resilient network structures. Understanding these trade-offs is essential for network administrators and designers to optimize network deployment strategies and achieve the desired balance between setup efficiency and structural complexity. Figure 12 show the results of setup time of the experiments.



	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192
RANDOM	0.092	0.145	0.326	1.256	2.719	12.752	18.393	26.715	39.212	57.454	83.757	183.908	274.483
LINEAR	0.098	0.182	0.295	0.542	0.894	1.889	3.319	6.822	14.841	33.713	55.915	98.009	181.411
BALANCED TREE	0.150	0.265	0.429	1.854	3.535	6.614	8.325	17.783	26.977	56.672	128.334	190.985	260.511

Figure 12: Comparative setup time diagram for the three topologies

7.5. Tear Time Analysis

The tear time refers to the duration required to dismantle a network topology, measured from the moment the destruction command is initiated. The analysis compares the tear times for three different network topologies: linear, balanced tree, and random. The results indicate significant differences in the time taken to dismantle each topology, highlighting the efficiency and complexity involved in their destruction.

7.5.1. Linear Topology

Shortest Tear Time: As expected, the linear topology takes the least amount of time to tear down. This is due to its straightforward structure, where nodes are connected in a simple chain, making it easy to dismantle.

Efficiency in Large Networks: In large networks, the linear topology shows about 30-40% faster tear times compared to the random and balanced tree topologies. This efficiency is attributed to the minimal complexity involved in breaking the direct sequential connections.

7.5.2. Random Topology

Longest Tear Time: The random topology consistently shows the longest tear time among the three topologies. This is due to the complexity and unpredictability of its connections, which require additional time to ensure all links are properly dismantled.

Peak Tear Times: The tear time peaks higher in the random topology, reflecting the inherent variability and instability in its structure.

7.5.3. Balanced Tree Topology

Long Tear Time: The balanced tree topology also exhibits a long tear time, like the random topology but slightly less. The hierarchical structure requires careful dismantling of multiple levels and branches, adding to the overall tear time.

Deviations in Linearity: Some deviations in the linearity of tear time are observed in the balanced tree topology. These deviations are due to the changes in the tree structure as different branches and levels are dismantled.

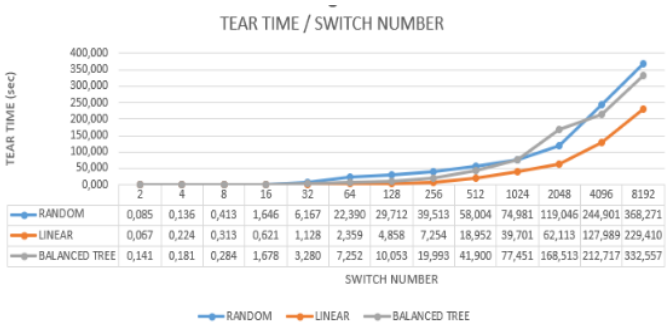
Efficiency of Linear Topology: The linear topology's simplicity extends to its tear-down process. Each node is directly connected to its predecessor and successor, making it easy to break these connections in sequence. This straightforward dismantling process results in consistently lower tear times.

Complexity in Random Topology: The random topology's longer tear time is attributed to its complex and unpredictable nature. The random connections between nodes mean that each dismantling process is unique and requires more time to ensure all links are effectively broken. This variability results in higher and more inconsistent tear times.

Structured Dismantling in Balanced Tree Topology: The balanced tree topology requires careful dismantling of its hierarchical structure. Each branch and level must be carefully broken down, which increases the overall tear time. The deviations in linearity are due to the varying complexity of dismantling different parts of the tree.

Instabilities and Variability: Both the random and balanced tree topologies show instabilities and variability in tear times. These instabilities are natural given the complexity of the structures and the need for careful dismantling to avoid leaving residual connections.

The tear time analysis highlights the impact of network topology on the efficiency of dismantling processes. Linear topology offers the fastest and most efficient tear-down times, making it suitable for scenarios requiring quick and straightforward network reconfiguration. The random and balanced tree topologies, while offering more complex and potentially more resilient structures, require significantly more time to dismantle. Understanding these differences is crucial for network administrators and designers in optimizing network management strategies, particularly in environments where frequent reconfiguration is necessary. Figure 13 show the results of tear time for each experiment.



	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192
RANDOM	0,085	0,136	0,413	1,646	6,167	22,390	29,712	39,513	58,004	74,981	119,046	244,901	368,271
LINEAR	0,067	0,224	0,313	0,621	1,128	2,359	4,858	7,254	18,952	39,701	62,113	127,989	229,410
BALANCED TREE	0,141	0,181	0,284	1,678	3,280	7,252	10,053	19,993	41,900	77,451	168,513	212,717	332,557

Figure 13: Comparative tear time diagram for the three topologies.

7.6. *Latency Analysis*

Latency measurement refers to the time taken for an information packet to travel from one network node to another, measured in milliseconds (ms). For this analysis, the packet size was set to 1024 Bytes (1 Kilobyte), and the maximum number of transferred packets was capped at 500, a limit identified in previous Mininet research for reliable measurements.

7.6.1. *Balanced Tree Topology*

7.6.1.1. *Highest Delay*

The balanced tree topology exhibits the highest latency among the three topologies. This significant delay is due to the complexity of its hierarchical structure, which requires packets to traverse multiple levels and branches before reaching their destination.

7.6.1.2. *Impact of Complexity*

The structured nature of the balanced tree increases the distance and processing time for packets, leading to higher latency.

7.6.2. *Linear and Random Topologies*

7.6.2.1. *Similar Delays*

Both linear and random topologies show similar latency measurements, but still lower than the balanced tree topology. These topologies have less complex routing paths, which reduces the overall transmission time.

7.6.2.2. *Comparative Analysis*

Although their delays are similar, the linear topology generally maintains a slightly more predictable and stable latency due to its straightforward path structure, while the random topology may experience more variability due to its non-deterministic connections.

7.6.3. *Latency Comparison*

Double the Delay in Balanced Tree: The latency in the balanced tree topology is at least double that of the other two topologies. This stark difference underscores the impact of hierarchical complexity on network performance.

7.6.4. *Balanced Tree Topology*

7.6.4.1. *Hierarchical Routing*

The balanced tree's multi-level structure means that packets often need to travel through several intermediary nodes (branches) before reaching their target. Each additional hop adds to the overall delay, resulting in the highest latency.

7.6.4.2. *Increased Processing Time*

Managing and routing through the hierarchical levels introduces additional processing delays, further contributing to the higher latency.

7.6.5. *Linear Topology*

7.6.5.1. *Direct Pathways*

The linear topology benefits from direct, sequential connections between nodes. This straightforward routing minimizes the number of hops and processing required, leading to more predictable and lower latency.

7.6.5.2. *Stable Performance*

The linear nature of the topology ensures consistent performance, with each packet following a clear and defined path.

7.6.6. *Random Topology*

Variable Pathways: The random topology features non-deterministic connections, meaning packets may traverse different paths depending on the network state. This variability can introduce occasional increases in latency, although the average delay remains lower than the balanced tree topology.

Adaptability and Efficiency: Despite its variability, the random topology's design aims to balance load and optimize pathways, helping to maintain relatively low latency overall.

The latency analysis highlights the significant influence of network topology on transmission delays. The balanced tree topology, with its complex hierarchical structure, results in the highest latency, making it less suitable for applications requiring rapid data transmission. Linear topology, with its direct and predictable pathways, offers the lowest latency and stable performance, ideal for time-sensitive applications. The random topology, while variable, maintains lower latency than the balanced tree and can adapt to different network conditions effectively. These insights are crucial for network administrators and designers to optimize network performance based on specific latency requirements and application needs. Figure 14 presents the latency results.

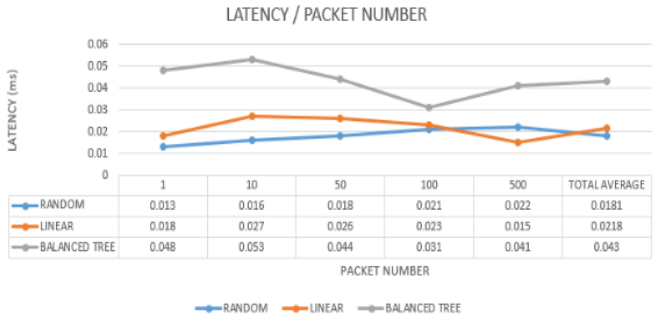


Figure 14: Comparative latency diagram for the three topologies.

7.7. Future Research

The content of this specific postgraduate work is a fundamental pillar of research on SDN networks and extends existing research in the field of computer science, networks, and telecommunications. Future research could be expanded on an even larger scale with the aid of supercomputers from major academic structures to show how a total shift in networking towards SDN would affect the internet and the world in general. With the right available resources, even more realistic simulations would be possible, aiming for direct integration, improvement, and gradual adaptation, initially in academic structures and subsequently in society, aiming for a stronger global network that would be more efficient, reliable, and capable of withstanding the continuously increasing needs of modern society. Lastly, as an extension of what was studied, the combination of currently active protocols to create a new improved one is feasible.

8. Conclusions

8.1. Performance of Topologies

Through experimental procedures, we can understand how SDN functions best and the operation of distinct topologies. Large-scale networks are created, and their characteristics are studied. This postgraduate work achieves an understanding of these network structures in real-time and how their effective application is possible in real-time.

The results indicate that the balanced tree topology consumes the most CPU resources due to its complexity, followed by the random topology. Linear topologies showed the least CPU usage. RAM consumption was highest in the balanced tree topology, while the random topology demonstrated lower RAM usage due to its customized nature. Latency measurements revealed that the balanced tree topology had the highest delay, while the linear and random topologies performed better with less delay. The random topology achieves improved results due to its adaptability and the ability to be parameterized. However, there is always the possibility of unsuccessful connections in the random topology, which affects performance but adds a more "realistic" application. The linear topology remains simple and maintains top performance. In contrast, the balanced tree topology, due to its architecture, reduces performance as it is burdened and expanded because of the complexity and calculations required for the successful creation of the tree. The use of the POX controller in collaboration with the OFDP protocol facilitated the expansion of SDN network sizes through parameterization.

References

[1] B. A. Nunes, M. Mendonca, N. Nguyen, K. Obraczka and T. Turletti, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," in IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1617-1634, Third Quarter 2014, doi: 10.1109/SURV.2014.012214.00180.

[2] N. V. Oikonomou, S. V. Margariti, E. Stergiou, and D. Liarokapis, "Performance Evaluation of Software-Defined Networking Implemented on Various Network Topologies," in 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Preveza, Greece, 2021, pp. 1-6, doi: 10.1109/SEEDA-CECNSM53056.2021.9566213.

[3] A. Zacharis, S. V. Margariti, E. Stergiou, and C. Angelis, "Performance evaluation of topology discovery protocols in software defined networks," in 2021 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Heraklion, Greece, 2021, 135-140, doi: 10.1109/NFV-SDN53031.2021.9665006.

[4] M. Guo and P. Bhattacharya, "Controller Placement for Improving Resilience of Software-Defined Networks," in 2013 Fourth International Conference on Networking and Distributed

Computing, Los Angeles, CA, USA, 2013, 23-27, doi: 10.1109/ICNDC.2013.15.

[5] IETF RFC 7426, "Request for Comments: 7426, ISSN: 2070-1721 EICT. Category: Informational," K. Pentikousis, Ed., 2015.,doi:10.20535/2411-2976.12021.24-32

[6] D. Kreutz, F. M. V. Ramos, P. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," arXiv, 2014, [Online], doi:/10.48550/arXiv.1406.0440

[7] A. Nayak, A. Reimers, N. Feamster, and R. Clark,"Resonance: Dynamic access control in enterprise networks," in Proc. Workshop: Research on Enterprise Networking, Barcelona, Spain, 2009, 1-6, doi:10.1145/2602204.2602219

[8] A. Voellmy and P. Hudak, "Nettle: Functional reactive programming of OpenFlow networks," in Proc. Workshop on Practical Issues in Programming, 2009, pp. 1-6,doi: 10112206

[9] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, "ElasticTree: Saving energy in data center networks," in Proc. 7th USENIX Symposium on Networked Systems Design and Implementation, 2010, 1-6,doi: 10.5555/1855711.1855728

[10] R. Wang, D. Butnariu, and J. Rexford, "OpenFlow-based server load balancing gone wild," in Hot-ICE, 2011, 1-6,doi: 10.5555/1972422.1972438

[11] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. Stuart, and A. Vahdat, "B4: Experience with a globally deployed software defined WAN," in ACM SIGCOMM, 2013, pp. 1-6,doi: 10.1145/2534169.2486019

[12] T. Koponen, M. Casado, N. Gude, J. Stribling, L. Poutievski, M. Zhu, R. Ramanathan, Y. Iwata, H. Inoue, T. Hama, and S. Shenker, "Onix: A distributed control platform for large-scale production networks," in OSDI, vol. 10, 1–6, 2010,doi: 10:351-364

[13] X. Zhao, S. S. Band, S. Elnaffar, M. Sookhak, A. Mosavi, and E. Salwana, "The Implementation of Border Gateway Protocol Using Software-Defined Networks: A Systematic Literature Review," IEEE Access, vol. 9, 112596-112606, 2021, doi: 10.1109/ACCESS.2021.3103241.

[14] I. F. Akyildiz, A. Lee, P. Wang, M. Luo, and W. Chou, "A roadmap for traffic engineering in SDN-OpenFlow networks," Elsevier Computer Networks, vol. 71, pp. 1–30, 2014,doi:10.1016/j.comnet.2014.06.002

[15] ONF TR-537, "Negotiable Datapath Model and Table Type Pattern Signing," Version 1.0, Sep. 2016, pp. 1-6.

[16] N. Handigol, M. Flajslik, S. Seetharaman, N. McKeown, and R. Johari, "Aster*x: Load-balancing as a network primitive," in ACLD '10: Architectural Concerns in Large Datacenters, 2010, 1-6,doi: 10.1109/GREE.2014.9

[17] R. Sherwood, G. Gibb, K.-K. Yap, G. Appenzeller, M. Casado, N. McKeown, and G. Parulkar, "Can the production network be the testbed?" in Proc. 9th USENIX OSDI, Vancouver, Canada, 2010, 1-6, doi: 10.5555/1924943.19249691

[18] A. Rodriguez-Natal, M. Portoles-Comeras, V. Ermagan, D. Lewis, D. Farinacci, F. Maino, and A. Cabello, "LISP: a southbound SDN protocol?" IEEE Communications Magazine, vol. 53, 201-207, 2015, doi: 10.1109/MCOM.2015.7158286.

[19] K. Greene, "TR10: Software-defined networking," MIT Technology Review, 2009, [Online]. Available: http://www2.technologyreview.com/article/412194/tr10-software-defined-networking/, doi:10.1109/COMST.2016.2633579

[20] M. Casado, M. J. Freedman, J. Pettit, J. Luo, N. McKeown, and S. Shenker, "Ethane: Taking control of the enterprise," in ACM SIGCOMM '07, 2007, 1-6, doi:10.1145/1282427.1282382

[21] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, "NOX: Towards an operating system for networks," ACM SIGCOMM Computer Communication Review, vol. 38, no. 3, 105–110, 2008, doi:10.1145/1384609.1384625

NIKOLAOS V. OIKONOMOU has received his BSc degree from the University of Ioannina, Department of Informatics and Telecommunications in 2021. He received his MSc degree from the same institution in 2023. He is an academic researcher also working as a private tutor in the field of Computer Science and mathematics. He has years of experience as a Computer Engineer, IT specialist and Network consultant. He also taught at the University of Ioannina and worked as an application developer.

DIMITRIOS V. OIKONOMOU has received his BSc from University of Western Macedonia, Department of Regional and Cross Border Studies in 2024. He is currently an active research member of the University of Western Macedonia and is about to begin his MSc studies.

ELEFTHERIOS STERGIOU has received his Meng Degree from National University of Athens, Department of Electrical Engineering and Informatics. He received his MSc in Telematics from the University of Sheffield Department of Computer Science and his PhD degree in the field of Computer Science from University of Patras. Currently he is an Associate Professor at the Department of Informatics and Telecommunications, University of Ioannina. His research interests include mainly Performance issues of computer networks and Telecommunication systems.

DIMITRIOS LIAROKAPIS has received his Meng Degree from University of Patras, Department of Computer Engineering and Informatics. He received his MSc and PhD degree in the field of Computer Science from the University of Massachusetts Boston. Currently he is a professor of

practice Professor at Department of Informatics and Telecommunications, University of Ioannina since 2012. His research interests include mainly databases and programing languages.