Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

March 2025 Volume XVII Number 1 ISSN 2061-2079

Technically Co-Sponsored by

IEEE ComSoc™ IEEE Communications Society

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.
Infocommunications Journal is also included in the Thomson Reuters – Web of ScienceTM Core Collection, Emerging Sources Citation Index (ESCI)

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section

Supporters

FERENC VÁGUJHELYI – president, Scientific Association for Infocommunications (HTE)

The publication was produced with the support of the Hungarian Academy of Sciences and the NMHH

National Media and Infocommunications Authority • Hungary

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1051 Budapest, Bajcsy-Zsilinszky str. 12, Room: 502
Phone: +36 1 353 1027 • E-mail: info@hte.hu • Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Telecommunications and Media Informatics
Phone: +36 1 463 4189 • E-mail: pvarga@tmit.bme.hu

Subscription rates for foreign subscribers: 4 issues 13.700 HUF + postage

Publisher: PÉTER NAGY

HU ISSN 2061-2079 • Layout: PLAZMA DS • Printed by: FOM Media

Authors, co-authors of the March 2025 issue

Zoltán Belső, László Pap;
Arun Kumar, Aziz Nanthaamornphong;
Amalia Eka Rakhmania, Hudiono Hudiono,
Umi Anis Ro'isatin, Nurul Hidayati;
Lajos Nagy; Botond Tamás Csathó,
Bálint Péter Horváth; Shahad A. Hussein,
Suadad S. Mahdi, Alharith A. Abdullah;
Ákos Leiter, Döme Matusovits, László Bokor;
Mohammed Jabardi; Attila Zoltán Jenei,
Réka Ágoston, István Valálik







Effect of the Imperfect Channel Estimation on Achievable NOMA Rate

Zoltán Belső, and László Pap

***Abstract*—In recent years, Non-orthogonal Multiple Access (NOMA) has been proposed as an alternative to the more traditional Orthogonal Multiple Access (OMA) schemes for mobile communication. In the NOMA method, the resource domains (like power and bandwidth) are not split but shared between the users of the network. The non-orthogonality means that there is cross-talk between the signals of different users, and the interference is either cancelled by a method called successive interference cancellation (SIC) or treated as part of the noise.**

Comparing the achievable capacity region of OMA and NOMA schemes show that NOMA has advantage over OMA. The SIC method requires knowledge of the channel characteristic between the base station and the user. In the ideal case where all the channel conditions are precisely known, NOMA always performs better than or equal to OMA. In real application, the channel characteristic can only be estimated, which can be non-perfect.

In this paper, we will examine the effect of non-perfect channel estimation on the performance of NOMA and will find that in some cases, NOMA still perform better than OMA, but in other cases OMA would perform better.

***Index Terms*—Non-Orthogonal Multiple Access (NOMA), achievable capacity region, non-perfect channel estimation**

I. INTRODUCTION

In recent years, there has been increased discussion of NOMA (Non-orthogonal Multiple Access) as a better choice for multi-user communication in comparison to Orthogonal Multiple Access (OMA) schemes [1]–[4]. The basic working principle of NOMA is superposition coding (SC) and successive interference cancellation (SIC). In the case of downlink communication, the base station transmits the superposition of all the signals of all the active users. In the case of uplink communication, all the active users transmit at the same time, and the superposition of these signals is received by the base station.

The receiver, applying SIC, decodes the strongest signal from the superposition first, even when that signal was not meant to be for them. It then re-modulates the decoded signal, applies the known or rather estimated channel condition of that signal, and subtracts it (which is the interference now for the rest of the signals) from the received signals. It then repeats the process until it reaches the signal of interest for them.

It is well-established that in ideal conditions, NOMA performs at least equally, and in most cases better, in some cases much better than a competing OMA schemes, such as Frequency Division Multiplexing Multiply Access (FDMA). By ideal conditions, we mean only additive white Gaussian noise (AWGN) is present in the channel, and the channel condition (both the phase shift and attenuation) is estimated perfectly for all signals.

Many papers discuss the problem of estimating the receiving channel [5]–[7] in both OMA and NOMA cases. In [8], the authors examine the effect of imperfect channel state information (CSI) due to hardware impairments in a cooperative uplink NOMA environment, focusing on sum rate as a metric. The paper [9] discusses the impact of imperfect SIC due to mismatched cancellation order.

In this paper, we discuss the effect of imperfect channel estimation on the effectiveness of SIC, considering that some part of the stronger signal remains as interference after re-modulation and subtraction. Our main metric for the two-user scenario is the achievable capacity region. For the many-user scenario, this metric becomes impractical, so we use the sum of the achievable rates as a metric.

There are two cases we have to discuss: One is downlink communication, where a single base station communicates with several users. The users must share both the bandwidth and the power budget available to the base station. The other is uplink communication, where several users try to communicate with a single base station. The bandwidth is also shared in this case, but every user has its own power budget independent of the others.

In the first part of the paper, we are considering a two-user and a base station scenario, and we share the resources between these two users. For each user, we can calculate the capacity of the channel given the bandwidth and power allocated. The capacity of a band-limited channel with a given signal power level (P), bandwidth (W), and noise power spectral density (N_0) [10]:

$$C = W \cdot \log_2\left(1 + \frac{P}{N_0 \cdot W}\right) = W \cdot \log_2\left(1 + SNR\right) \quad (1)$$

where the noise power in the band is $N = N_0 \cdot W$ and SNR is the signal-to-noise ratio ($SNR = \frac{P}{N} = \frac{P}{N_0 \cdot W}$). Since the resources have to be allocated between the users, allocating more to one of the users (to increase the capacity of its channel) mean there remains less for the other user, hence its channel capacity will decrease. That means there is a region of achievable capacity for the two users.

In all of this cases, we will discuss how this region is changing when the successive interference cancellation (SIC) cannot be performed perfectly due to imperfect channel estimation. This means that some ϵ portion of the cancelled signal remains and considered as part of the noise while decoding the second

Submitted on 2024.11.22

Zoltán Belső and László Pap, Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary (E-mail: {belso,pap}@hit.bme.hu)

DOI: 10.36244/ICJ.2025.1.1

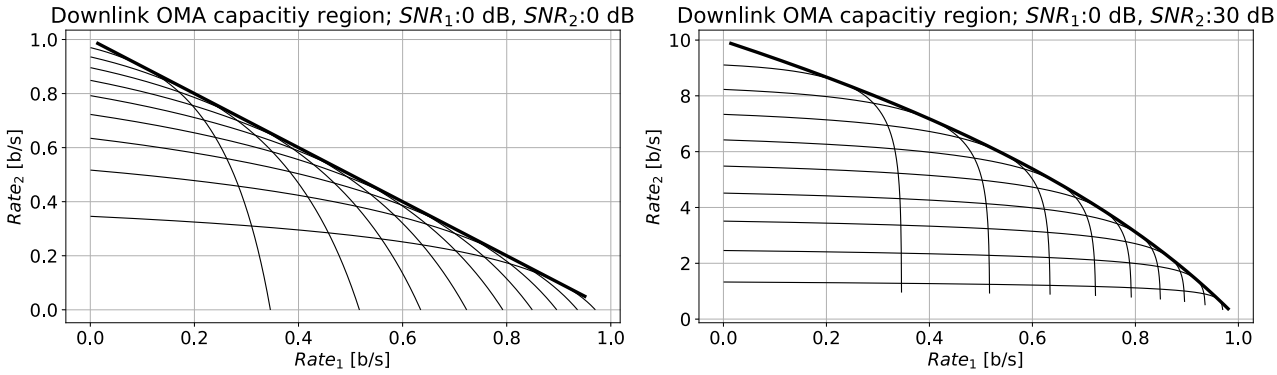


Fig. 1: Downlink OMA (FDMA) capacity region: user1 rate versus user2 rate for different SNR conditions. The thin lines are for fixed values of α ranging from 0.1 to 0.9 (from left to right) [11].

user's signal. As we will see, the effect depends not only on the extent to which the interfering signal remains, but also on what is the ratio of signal powers between the users.

After discussing the two-user case for both uplink and downlink, we extend our analysis by examining the sensitivity of the sum-rate in a multi-user scenario.

This paper is organized as follows: First, we discuss the downlink case, followed by the uplink case, both in a two-user scenario. In each case, we first consider the achievable capacity region in an OMA case (FDMA), then the ideal NOMA case, and finally, the case of imperfect channel estimation, where some interference remains. We then provide a sensitivity analysis for the multi-user scenario. Finally, we conclude with a summary of our findings.

II. Downlink channel of two users

In the downlink scenario, the base station is transmitting two separate signal, one for each user. The baseband signal is denoted by s_i $(i = 1, 2)$ with unity power: $E\left[|s_i|^2\right] = 1$. The transmit power for each user's signal is denoted by p_i $(i = 1, 2)$. The base station has to split its total transmit power budget between the users, so $p_{tot} = p_1 + p_2$. We can also denote a share coefficient $0 \leq \alpha \leq 1$:

$$\begin{aligned} p_1 &= \alpha \cdot p_{tot} \\ p_2 &= (1-\alpha) \cdot p_{tot} \end{aligned} \tag{2}$$

We denote the total bandwidth of the channel by W.

Each user's channel has a separate channel characteristic h_i $(i = 1, 2)$, which is assumed to be a complex number. The absolute value of h_i represents the channel gain, while the phase of h_i represents the phase shift of the channel. These characteristics are independent of each other.

A. OMA case

First, consider the Frequency Division Multiple Access (FDMA) case, where the available bandwidth is divided between the two users. Here, we consider an ideal case where no bandwidth is wasted. We can choose a parameter β, $(0 \leq \beta \leq 1)$, where one user occupies a $\beta \cdot W$ part of the channel bandwidth, while the other occupies the remaining $(1-\beta) \cdot W$ part, where W denote the total bandwidth of the channel. We consider the division perfectly orthogonal, so that there is no interference between the two users' signal.

The total transmit power of the base station also has to be split between the users.

Here, the maximal rate of communication of every OMA user is [1], [10]:

$$\begin{aligned} R_1 &= \beta \cdot W \cdot \log_2\left(1 + \frac{p_1 \|h_1\|^2}{\beta \cdot W \cdot N_0}\right) \\ &= \beta \cdot W \cdot \log_2\left(1 + \frac{p_1}{\beta \cdot W \cdot \frac{N_0}{|h_1|^2}}\right) \end{aligned} \tag{3}$$

$$\begin{aligned} R_2 &= (1-\beta) \cdot W \cdot \log_2\left(1 + \frac{p_2 |h_2|^2}{(1-\beta) \cdot W \cdot N_0}\right) \\ &= (1-\beta) \cdot W \cdot \log_2\left(1 + \frac{p_2}{(1-\beta) \cdot W \cdot \frac{N_0}{|h_2|^2}}\right) \end{aligned} \tag{4}$$

See Figure 1 for the case where both user has equal, 0 dB signal to noise ratio (SNR) and for the case where user1 has 0 dB signal-to-noise ratio, while user2 has a much better, 30 dB signal-to-noise ratio. Note that the shape of the convex hull of the capacity region is a straight line for the case of equal channel conditions but a curved line when the two users' conditions are significantly different. The exact shape is derived in [11].

Here, we consider as SNR the full-band noise $(W \cdot N_0)$ compared to the total transmit power of the base station as SNR: $SNR_i = \frac{p_{tot}|h_i|^2}{W \cdot N_0}$. That is the SNR for each user when the base station allocates all its power and all the bandwidth to this user, that is when the user is alone.

B. NOMA case

In the power domain NOMA case, both user occupies the whole channel bandwidth, and the base station's transmit power budget is distributed (at some proportion) between them: $p_1 + p_2 = p_{tot}$ or $p_2 = p_{tot} - p_1$, where p_{tot} is the given total transmit power of the base station [10], [12].

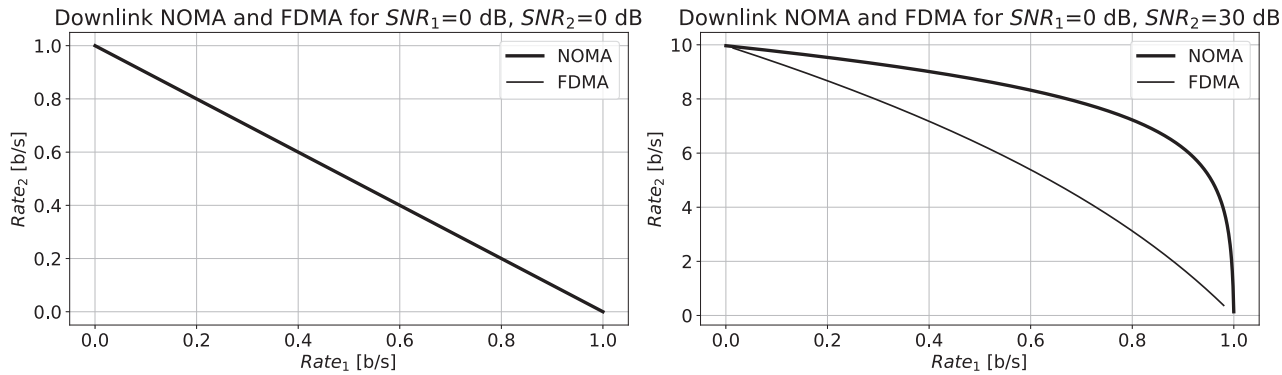


Fig. 2: Downlink NOMA capacity region: user1 rate versus user2 rate for different SNR conditions. For reference, the boundary of the OMA capacity region also plotted.

The transmitted signal by the base station is the sum of the two users' signal:

$$x = \sqrt{p_1}s_1 + \sqrt{p_2}s_2 \tag{5}$$

The received signal by each user $(i = 1, 2)$:

$$y_i = h_i \cdot x + w_i \tag{6}$$

where h_i is the complex channel characteristic between the user i and the base station. w_i is the noise sample at the receiver i, assumed to be Gaussian distribution with a mean of 0 and a power spectral density of N_0.

The optimum order of decoding is based on the signal-to-noise ratio (SNR) of the individual signals at each user's receiving end: $|h_i|^2/N_0$. In this decoding order, each user can successively decode any stronger (better SNR) signals and remove them from the received signal (cancellation by re-modulation). The ith user proceeds with successively decoding and cancelling the other user's signal until it reaches its own signal. All the remaining weaker signals are considered as noise or interference.

In the case of only two users, this mean that the first user with better channel conditions receives the other user's signal at a higher power (because it is transmitted at higher proportion of the base station's power budget in order to reach the farther user at a decodable level). Therefore, it decodes the other user's signal first, re-modulates it, and removes it from the original signal. Then, it decodes the remaining signal. During this process, it is assumed that the user can decode the other user's signal without error, but that does not mean that during the cancellation phase it can eliminate it perfectly because during re-modulation, it must consider the effect of the channel on the signal. If the real channel characteristic ($\hat{h_i} = h_i$) were known perfectly, the cancellation could be perfect. If there were some remaining error in the estimated value ($\hat{h_i} \neq h_i$), there would be some interfering signal remaining, proportional to the receiving power of the interfering signal.

1) Perfect channel estimation: In the case of perfect channel estimation ($\hat{h_i} = h_i$), the maximal rate of communication for every NOMA user in a channel with bandwidth W is [1], [10], [12], [13]:

$$\begin{aligned} R_1 &= W \cdot \log_2\left(1 + \frac{p_1|h_1|^2}{W \cdot N_0}\right) \\ &= W \cdot \log_2\left(1 + \frac{p_1}{W \cdot \frac{N_0}{|h_1|^2}}\right) \end{aligned} \tag{7}$$

$$\begin{aligned} R_2 &= W \cdot \log_2\left(1 + \frac{p_2|h_2|^2}{W \cdot N_0 + p_1|h_2|^2}\right) \\ &= W \cdot \log_2\left(1 + \frac{p_2}{W \cdot \frac{N_0}{|h_2|^2} + p_1}\right) \end{aligned} \tag{8}$$

Figure 2 shows the boundary of the achievable capacity region for the NOMA scheme. The first diagram shows the case where both user has equal, 0 dB signal-to-noise ratio (SNR). Note that in that case, we get the same rate limit as in the OMA (FDMA) case. The second diagram shows the case where one of the users has a better, 30 dB SNR. For reference we also plot the boundary for the OMA case. For an exact comparison, when calculating the SNR, we consider the noise power in the total bandwidth ($W \cdot N_0$) compared to the same total transmit power of the base station as in the OMA (FDMA) case, although we get the different rate pairs on the figure by allocating the total base station power divided between the individual users. The difference of in the SNR of the two users represents either the difference in the channel conditions or the local power of the additive Gaussian noise.

2) Imperfect channel estimation: In the case when the channel estimation is not perfect, that is $\hat{h_i} \neq h_i$, after the first user receives and demodulates the stronger signal of the second user, the re-modulation and the cancellation of the received stronger signal cannot be done perfectly. This means that even for the first user, some part of the second user's signal remains as interference. We consider this as if some $\epsilon > 0$ part of the interfering signal power were added to the ever present Gaussian white noise ($W \cdot N_0$):

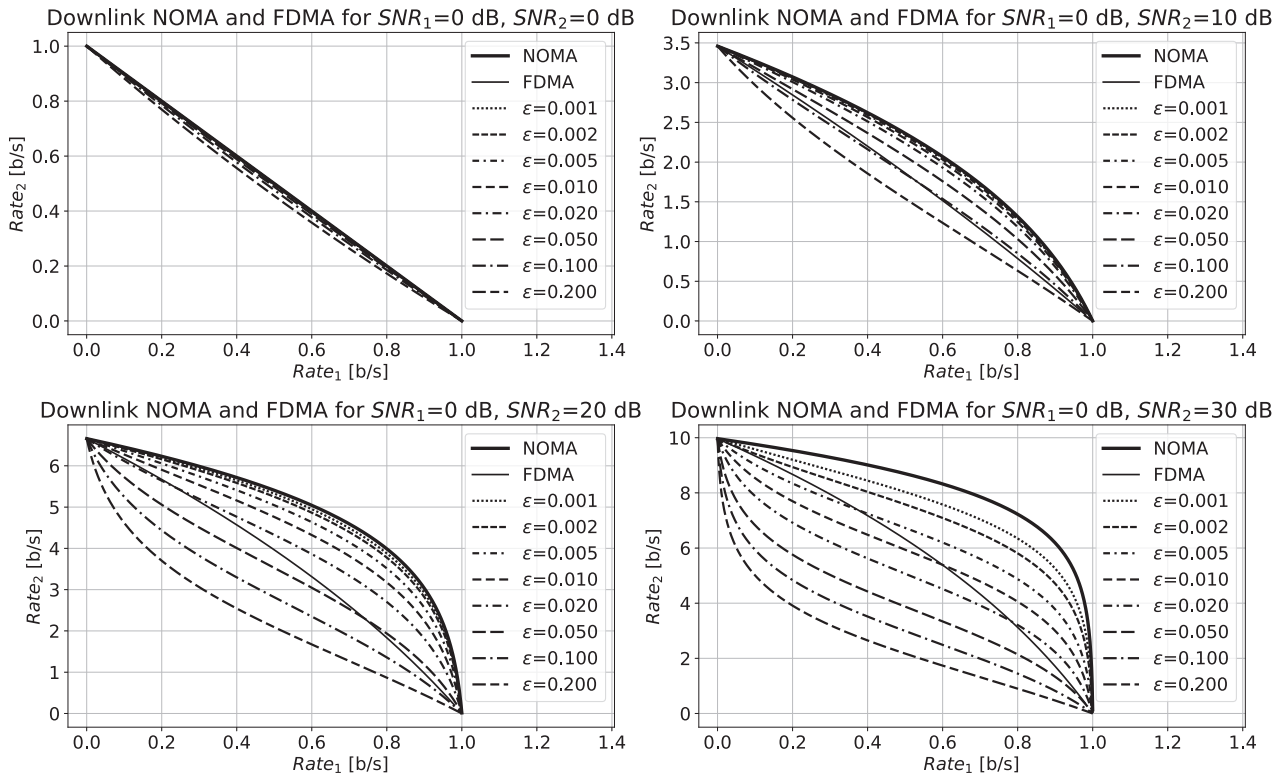


Fig. 3: Downlink NOMA capacity region with imperfect channel estimation: user1 rate versus user2 rate for different SNR conditions. ϵ is the proportion of the remaining interference signal after imperfect cancellation. For reference, the boundary of the OMA capacity region also plotted.

$$R_1 = W \cdot \log_2\left(1 + \frac{p_1|h_1|^2}{W \cdot N_0 + \epsilon \cdot p_2|h_1|^2}\right) \tag{9}$$
$$= W \cdot \log_2\left(1 + \frac{p_1}{W \cdot \frac{N_0}{|h_1|^2} + \epsilon \cdot p_2}\right)$$
$$R_2 = W \cdot \log_2\left(1 + \frac{p_2|h_2|^2}{W \cdot N_0 + p_1|h_2|^2}\right) \tag{10}$$
$$= W \cdot \log_2\left(1 + \frac{p_2}{W \cdot \frac{N_0}{|h_2|^2} + p_1}\right)$$

It is easy to predict that the gain of NOMA will decrease as ϵ increases. See Figure 3 for the achievable rates depending on the value of ϵ. In the first diagram, both users have a signal-to-noise ratio (SNR) of 0 dB. In that case there was no gain for NOMA, so for any $\epsilon > 0$, the NOMA rate limit will go below the FDMA rate limit. The second diagram shows the case where one user has a better SNR of 10 dB, while the other has the same poor SNR of 0 dB. In the other two diagrams, one of the users has an even better SNR of 20 dB and 30, respectively. In those cases, NOMA can benefit from the great difference in the power level of the two signals: the interference caused by the weak signal on the decoding of the strong signal is minimal, and the cancellation of the strong signal helps a lot in decoding the weak signal. However, if the cancellation is imperfect, the small portion that is interfering from the strong signal decreases the rate limit of the weak signal because even a small portion of the much stronger signal causes great interference.

III. Uplink channel of two users

In the uplink scenario, the users transmit independently to the base station. The base station receives the superposition of the users' signals. Let's denote the baseband signal of the two users by s_i $(i = 1, 2)$ with unity power: $E\left[|s_i|^2\right] = 1$. The transmit power of each user is independent of the other and is denoted by p_i $(i = 1, 2)$. We denote the total bandwidth of the channel by W again.

Each user's channel has a separate channel characteristic h_i $(i = 1, 2)$, assumed to be a complex number. The absolute value of h_i represents the channel gain, while the phase of h_i represents the phase shift of the channel. These characteristic are independent of the other user's value.

The received signal at the base station is:

$$y = h_1\sqrt{p_1}s_1 + h_2\sqrt{p_2}s_2 + w \tag{11}$$

where w is the noise sample at the receiver, assumed to be a Gaussian distribution of mean 0 and power spectral density of

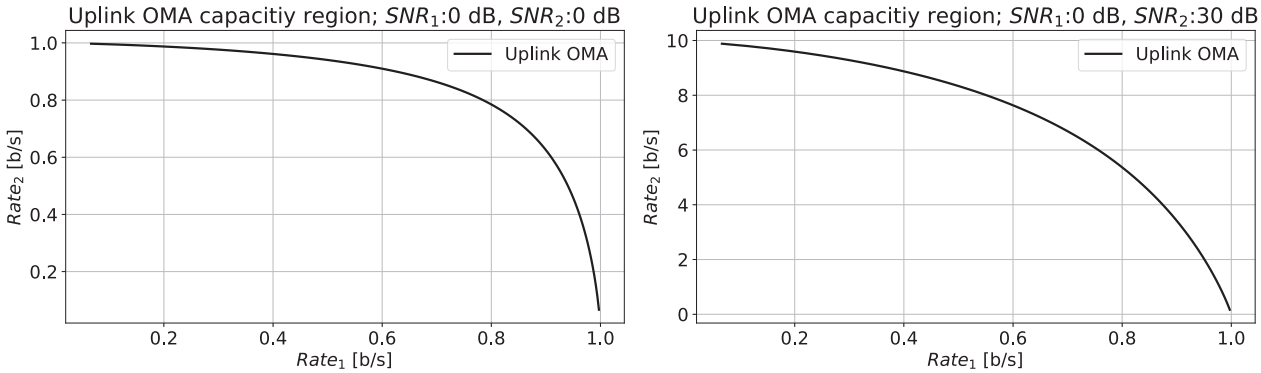


Fig. 4: Uplink OMA (FDMA) capacity region: user1 rate versus user2 rate for different SNR conditions.

N_0 shaped by the receiving filter to the receiving bandwidth. Since there is only one receiver (the base station) in this case, only a single additive noise component applies to the superposition of the two users' signal.

For each user's signal, there are two terms that affect the receiving level: the channel gain and the transmit power. Therefore, without loosing generality, we can assume that the transmitted power is the same for both users ($p_1 = p_2 = p$) and account all the differences in the receive level due to the channel gain.

A. OMA case

For the Orthogonal Multiple Access (OMA) case, let us consider the Frequency Division Multiple Access (FDMA) again: In this scheme, the available bandwidth is divided between the two users. We can choose a parameter $\beta,\ (0 \leq \beta \leq 1)$, where one user occupies a $\beta \cdot W$ part of the channel bandwidth while the other occupies the remaining $(1-\beta) \cdot W$ part, where W denotes the total bandwidth of the channel. We assume the division perfectly orthogonal, so that there is no interference between the two users' signal. The power is not divided between the users; both users are transmitting at full power but only use the allocated part of the bandwidth.

Here, the maximal rate of communication of every OMA user is [1], [10]:

$$R_1 = \beta \cdot W \cdot \log_2\left(1 + \frac{p|h_1|^2}{\beta \cdot W \cdot N_0}\right) \tag{12}$$
$$= \beta \cdot W \cdot \log_2\left(1 + \frac{p}{\beta \cdot W \cdot \frac{N_0}{|h_1|^2}}\right)$$

$$R_2 = (1-\beta) \cdot W \cdot \log_2\left(1 + \frac{p|h_2|^2}{(1-\beta) \cdot W \cdot N_0}\right) \tag{13}$$
$$= (1-\beta) \cdot W \cdot \log_2\left(1 + \frac{p}{(1-\beta) \cdot W \cdot \frac{N_0}{|h_2|^2}}\right)$$

Figure 4 shows the boundary of the capacity region achievable with a classical FDMA case. There are two cases shown: in the first diagram, the users have the same signal-to-noise ratio (SNR): 0 dB, while in the second diagram, one of the users have 30 dB better SNR than the other. Here we consider the in band noise ($\beta \cdot W \cdot N_0$) for calculating the SNR.

B. NOMA case

In the power domain NOMA case, both users occupy the entire channel bandwidth. The base station first decodes the signal of one of the users while considering the interference caused by the other signal as part of the noise. Then, it can re-modulate the decoded signal, apply the channel characteristic of the user, and subtract this from the received signal. In an ideal case, it fully eliminates the decoded signal, and the base station can decode the other signal as if it were the only signal. This is called successive interference cancellation (SIC).

In this case, it is not useful to just consider both users transmitting at full power, as it would give us a single point on the capacity plane. Instead, we can trade the channel capacity between the users by scaling the transmit power. Of course, this gives the same result as if we consider the channel conditions as a parameter.

It is usually assumed that the stronger signal is decoded first because eliminating that could help a lot to decode the weaker signal. But that is not the only possibility. Depending on the goal, one may choose to decode and eliminate the weaker signal first. For example, if the goal is to maximize the achievable bit rate of the user with the stronger signal, while letting the weaker user communicating at some lower rate without interfering with the other, that can be achieved by decoding and cancelling the weaker signal first.

As in the downlink case, we can consider two sub-cases: first, when the channel characteristics are perfectly known (perfect channel estimation); and second, when the channel estimation (denoted by $\hat{h_i}$) is imperfect and not equal to the real channel parameter.

1) Perfect channel estimation: If it is the first user's signal that is decoded first, during decoding its signal, the other user's signal is considered as noise [1], [10], [13]:

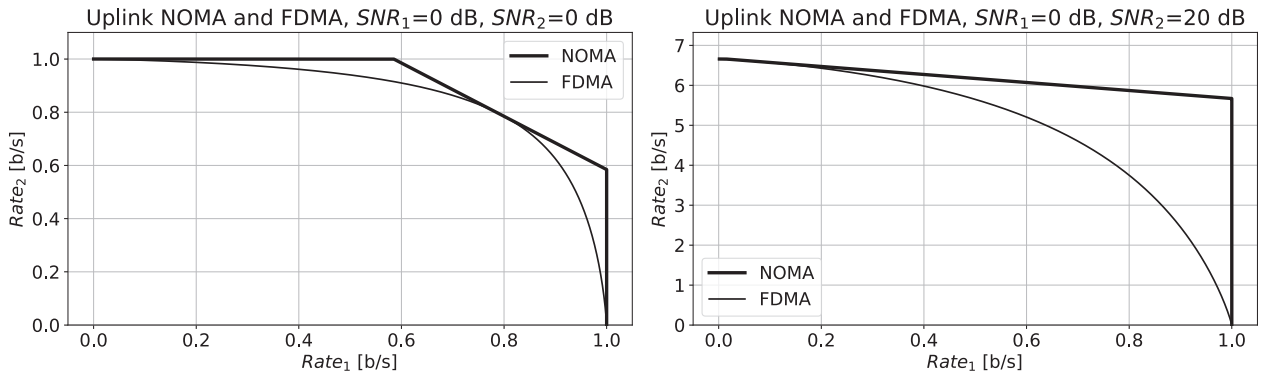


Fig. 5: Uplink NOMA capacity region: user1 rate versus user2 rate for different SNR conditions. For reference, the boundary of the OMA capacity region also plotted.

$$R_1 = W \cdot \log_2\left(1 + \frac{p_1|h_1|^2}{W \cdot N_0 + p_2|h_2|^2}\right) \tag{14}$$

$$R_2 = W \cdot \log_2\left(1 + \frac{p_2|h_2|^2}{W \cdot N_0}\right) \tag{15}$$

If it is the second user's signal that is decoded first, the case is the opposite:

$$R_1 = W \cdot \log_2\left(1 + \frac{p_1|h_1|^2}{W \cdot N_0}\right) \tag{16}$$

$$R_2 = W \cdot \log_2\left(1 + \frac{p_2|h_2|^2}{W \cdot N_0 + p_1|h_1|^2}\right) \tag{17}$$

In both cases, the sum rate $(R_1 + R_2)$ is limited by the sum of the two power (scaled by the channel conditions), which is the total capacity of the channel:

$$R_1 + R_2 = W \cdot \log_2\left(1 + \frac{p_1|h_1|^2 + p_2|h_2|^2}{W \cdot N_0}\right) \tag{18}$$

The achievable capacity region is limited by three factors: for both users, their own maximum power limits the capacity achievable by that user, even in the absence of the other user. That gives us a horizontal and a vertical line in our capacity diagram. At the same time, the sum of the achievable rate of the two users is limited by the sum of their power. That gives us a diagonal line in our capacity diagram. Since all of the conditions must be fulfilled, the capacity region is the convex hull marked by these three fraction lines.

Figure 5 shows the capacity region for two uplink NOMA users. In the first diagram, the two users has equal, 0 dB signal-to-noise ratio (SNR), while on the second diagram one of the users has 20 dB better SNR condition. For reference, we have also plotted the limit of the OMA (FDMA) case from the previous section. Note that in both cases there is one point where the two limits coincides, every other case, the NOMA outperforms the OMA case.

Note, that the two corners of these fraction lines are representing the case where (14), (15) and (16),(17) fulfils with equality, respectively. On the horizontal part of the fraction lines R_2 is constant since the first user is completely eliminated. Similarly on the vertical part R_1 is constant since here the second user is decoded first and completely eliminated, so the capacity of the first user's channel does not depend of the transmitted power of the second user's signal.

2) Imperfect channel estimation: In the case when the channel estimation is not perfect, that is $\hat{h_i} \neq h_i$, after the base station demodulates the signal of the first user, the re-modulation and the cancellation of the received signal cannot be done perfectly. This means that for the second user, there remains some part of the first user's signal as interference. We consider this as if some $\epsilon > 0$ part of the interfering signal power is added to the ever-present Gaussian white noise $(W \cdot N_0)$:

$$R_1 = W \cdot \log_2\left(1 + \frac{p_1|h_1|^2}{W \cdot N_0 + \epsilon \cdot p_2|h_2|^2}\right) \tag{19}$$

$$R_2 = W \cdot \log_2\left(1 + \frac{p_2|h_2|^2}{W \cdot N_0 + p_1|h_1|^2}\right) \tag{20}$$

Please refer to Figure 6 for the effect of imperfect channel estimation. In the first diagram, both users have an equal signal-to-noise ratio (SNR) of 0 dB. In this case the situation is symmetrical, and the imperfect cancellation slightly decreases the achievable rate pairs. In the second diagram, one of the users has a better SNR of 10 dB. In this case, imperfect cancellation has a larger effect on the achievable rate pairs. The other two diagrams shows the case where one of the users has an even better SNR of 20 dB and 30 dB, respectively. Please note that these are the cases where NOMA can gain a lot compared to the OMA case.

One thing can be noticed is that the limits for the individual users, which were a horizontal and vertical lines previously, are not straight lines anymore. That is because the transmitted power of the other user, that is decoded first, affects the user whose signal is decoded second due to the imperfect

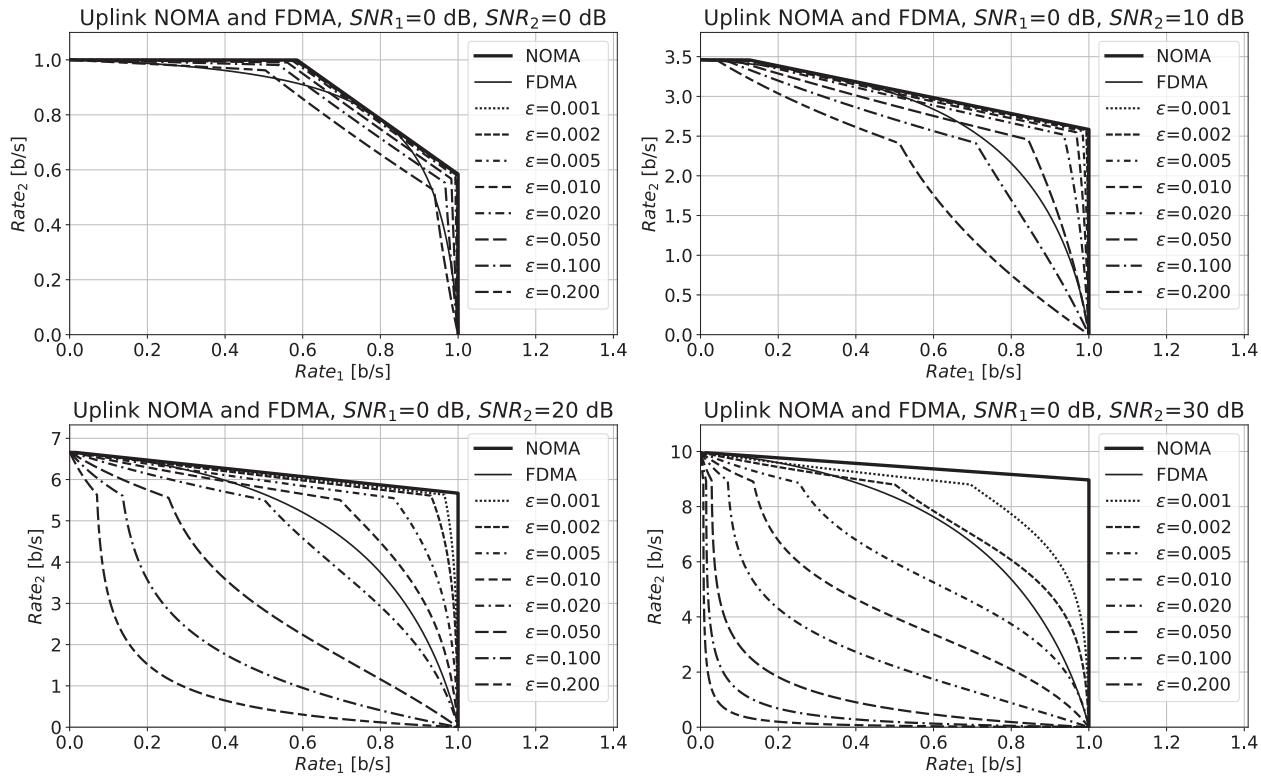


Fig. 6: Uplink NOMA capacity region with imperfect channel estimation: user1 rate versus user2 rate for different SNR conditions. ϵ is the proportion of the remaining interference signal after imperfect cancellation. For reference, the boundary of the OMA capacity region also plotted.

cancellation. We can see that even a small imperfection causes NOMA to partially under-perform the OMA case, and for larger ϵ values, there are hardly any case where NOMA is better. Even in that case, for some regions, NOMA can outperform OMA, but there are large regions where OMA can win.

The diagrams are arranged such that the user with the stronger maximal transmit power is on the vertical axis. One may wonder that at first glance it may seam like the imperfect cancellation affects the stronger user more. To interpret the diagrams correctly, one must keep in mind that they show the boundary of the achievable region, which in most cases corresponds to one of the user not transmitting at maximal power. The imperfect cancellation means that the stronger signal causes strong interference to the weaker signal, so in order to reach a relatively high capacity for the weaker user, the stronger one must decrease power, limiting their own achievable capacity.

IV. Sensitivity

Extending the discussion to more than two users means that the capacity region becomes multidimensional, which makes visualization challenging. A more useful approach is to investigate how sensitive the achievable rate of the users is to small changes in cancellation imperfection.

A. Sensitivity in multi-users case

The inequalities for the multi-user case are (without loss of generality, assuming that the users are numbered in the order of decoding):

$$R_i = W \cdot \log_2 \left(1 + \frac{p_i |h_i|^2}{W \cdot N_0 + \sum_{j=i+1}^{N} p_j |h_j|^2} \right) \tag{21}$$

There is one such equation for each user. For the first user, the signals from all other users act as interference. For the last user, the summation in the denominator is empty since all other users' signals have been canceled.

Considering the imperfection in cancellation, the equations become:

$$R_i = W \cdot \log_2 \left(1 + \frac{p_i |h_i|^2}{W \cdot N_0 + \sum_{k=1}^{i-1} \epsilon_k p_k |h_k|^2 + \sum_{j=i+1}^{N} p_j |h_j|^2} \right) \tag{22}$$

For the first user, the summation term in the denominator that contains ϵ is empty, as there is no signal for the first user to (imperfectly) cancel.

To simplify the following discussion, let's introduce some notation: let $X_i = \frac{R_i}{W} \ln 2$ represent the normalized rate, and $A_i = \frac{p_i |h_i|^2}{W \cdot N_0}$ represent the normalized received power or signal-to-noise ratio (SNR) without interference. Using this notation, the general equation becomes:

$$X_i = \ln\left(1 + \frac{A_i}{1 + \sum_{k=1}^{i-1} \epsilon_k A_k + \sum_{j=i+1}^{N} A_j}\right) \tag{23}$$

To evaluate the sensitivity of the channel to the ϵ_i factors ($i \in [1 \dots N-1]$), which represent the imperfections in cancellation during SIC processing, we can use the sum rate of the users as a metric:

$$X_\sum = \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} \ln\left(1 + \frac{A_i}{1 + \sum_{k=1}^{i-1} \epsilon_k A_k + \sum_{j=i+1}^{N} A_j}\right) \tag{24}$$

We are interested in quantifying the change $dX_\sum$ in response to a small change $d\epsilon_i$. This can be expressed through the partial derivatives of $X_\sum$:

$$\begin{aligned} dX_\sum &= \sum_{l=1}^{N-1} \left(\frac{d}{d\epsilon_l} X_\sum\right) d\epsilon_l \\ &= \sum_{l=1}^{N-1} \frac{d}{d\epsilon_l} \left[\sum_{i=1}^{N} \ln\left(1 + \frac{A_i}{1 + \sum_{k=1}^{i-1} \epsilon_k A_k + \sum_{j=i+1}^{N} A_j}\right)\right] d\epsilon_l \end{aligned} \tag{25}$$

$$\begin{aligned} &= \sum_{l=1}^{N-1} \sum_{i=l+1}^{N} \frac{1}{\left(1 + \sum_{k=1}^{i-1} \epsilon_k A_k + \sum_{j=i+1}^{N} A_j\right)^2} \\ &\quad \cdot \frac{-A_i A_l}{1 + \frac{A_i}{1 + \sum_{k=1}^{i-1} \epsilon_k A_k + \sum_{j=i+1}^{N} A_j}} d\epsilon_l \end{aligned} \tag{26}$$

Since the sensitivity represents the change in the rate in response to a small imperfection in cancellation, we need to evaluate the derivative at $\epsilon_k = 0$:

$$dX_\sum = \sum_{l=1}^{N-1} \sum_{i=l+1}^{N} \frac{1}{1 + \frac{A_i}{1 + \sum_{j=i+1}^{N} A_j}} \frac{-A_i A_l}{\left(1 + \sum_{j=i+1}^{N} A_j\right)^2} d\epsilon_l \tag{27}$$

Depending on the relative power levels and whether we are in a high SNR regime ($A_i \gg 1$) or a low SNR regime ($A_i \cong 1$), the imperfection in SIC may change the optimal cancellation order.

B. Sensitivity in two users case

In the case of two users, the sum rate $X_\sum$ contains only two terms, and there is only a single ϵ factor:

$$X_\sum = \ln\left(1 + \frac{A_1}{1 + A_2}\right) + \ln\left(1 + \frac{A_2}{1 + \epsilon A_1}\right) \tag{28}$$

The sensitivity in this case is:

$$\begin{aligned} \frac{d}{d\epsilon} X_\sum &= \frac{d}{d\epsilon} \ln\left(1 + \frac{A_2}{1 + \epsilon A_1}\right) \\ &= \frac{1}{1 + \frac{A_2}{1 + \epsilon A_1}} \frac{-A_2 A_1}{(1 + \epsilon A_1)^2} \end{aligned} \tag{29}$$

Since we are considering small imperfections in SIC processing, we take the derivative at $\epsilon = 0$:

$$dX_\sum = \frac{-A_1 A_2}{1 + A_2} d\epsilon \tag{30}$$

In the high SNR regime, where $A_2 \gg 1$, so $1 + A_2 \simeq A_2$ it can be further approximated as:

$$dX_\sum \cong -A_1 d\epsilon \tag{31}$$

In the low SNR regime, where $A_2 \cong 1$, we can approximate as:

$$dX_\sum \cong \frac{-A_1}{2} d\epsilon \tag{32}$$

The sensitivity is practically determined by the power level of the user whose signal we are canceling.

V. CONCLUSION

We have seen that in multi-user communication, we have some degree of freedom in allocating resources (bandwidth and power) to users, which leads to different achievable channel capacities. We can speak of optimal resource allocation in the sense of maximizing the achievable rate of one user while providing some rate for the other. We have discussed the achievable capacity region for both uplink and downlink communication, for both OMA and NOMA schemes. We have seen that a NOMA scheme is attractive and outperforms (or at least equal to) the theoretically optimal OMA case (i.e.: perfect orthogonality, no guard bandwidth needed, no inter signal interference). However, when the channel estimation is imperfect, the successive cancellation of the interfering stronger signal will also be imperfect, leading to a reduced achievable capacity rates for some or both users. In some cases, the degradation due to the imperfection of the channel estimation may result in capacity rates achievable with NOMA being even lower than those achievable with a more traditional OMA scheme.

REFERENCES

[1] Linglong Dai, Bichai Wang, Zhiguo Ding, Zhaocheng Wang, Sheng Chen, and Lajos Hanzo. A survey of non-orthogonal multiple access for 5g. *IEEE communications surveys & tutorials*, 20(3):2294–2323, 2018. **DOI**: 10.1109/COMST.2018.2835558.

[2] Linglong Dai, Bichai Wang, Yifei Yuan, Shuangfeng Han, I Chih-Lin, and Zhaocheng Wang. Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends. *IEEE Communications Magazine*, 53(9):74–81, 2015. **DOI**: 10.1109/MCOM.2015.7263349.

[3] Yuanwei Liu, Zhijin Qin, Maged Elkashlan, Zhiguo Ding, Arumugam Nallanathan, and Lajos Hanzo. Nonorthogonal Multiple Access for 5G and Beyond. *Proceedings of the IEEE*, 105(12):2347–2381, 2017. **DOI**: 10.1109/JPROC.2017.2768666.

[4] Yuya Saito, Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, Anxin Li, and Kenichi Higuchi. Non-orthogonal multiple access (noma) for cellular future radio access. In *2013 IEEE 77th vehicular technology conference (VTC Spring)*, pages 1–5. IEEE, 2013. **DOI**: 10.1109/VTCSpring.2013.6692652.

[5] Yizhi Tan, Jingrong Zhou, and Jiayin Qin. Novel channel estimation for non-orthogonal multiple access systems. *IEEE Signal Processing Letters*, 23(12):1781–1785, 2016. **DOI**: 10.1109/LSP.2016.2617897.

[6] Yang Du, Binhong Dong, Wuyong Zhu, Pengyu Gao, Zhi Chen, Xiaodong Wang, and Jun Fang. Joint channel estimation and multiuser detection for uplink grant-free noma. *IEEE Wireless Communications Letters*, 7(4):682–685, 2018. **DOI**: 10.1109/LWC.2018.2810278.

[7] Adam Knapp and László Pap. Statistical based optimization of number of pilot signals in lte/lte-a for higher capacity. In *IEEE EUROCON 2015 – International Conference on Computer as a Tool (EUROCON)*, pages 1–5, 2015.

[8] Sudhir Kumar Sa and Anoop Kumar Mishra. An uplink cooperative noma based on cdrt with hardware impairments and imperfect csi. *IEEE Systems Journal*, 17(4):5695–5705, 2023.

[9] Talgat Manglayev, Refik Caglar Kizilirmak, Yau Hee Kho, Nurzhan Bazhayev, and Ilya Lebedev. Noma with imperfect sic implementation. In *IEEE EUROCON 2017 – 17th International Conference on Smart Technologies*, pages 22–25, 2017.

[10] Thomas M Cover and Joy A Thomas. Information theory and statistics. *Elements of information theory*, 1(1):279–335, 1991. **DOI**: 10.1002/047174882X.

[11] Zoltán Belső and László Pap. On the convex hull of the achievable capacity region of the two user fdm oma downlink. *Infocommunications Journal*, XV(1):9–14, 2023. **DOI**: 10.36244/ICJ.2023.1.2.

[12] David Tse and Pramod Viswanath. Fundamentals of wireless communication. *Cambridge university press*, 2005. **DOI**: 10.1017/CBO9780511807213.

[13] Kenichi Higuchi and Anass Benjebbour. Non-orthogonal multiple access (noma) with successive interference cancellation for future radio access. *IEICE Transactions on Communications*, 98(3):403–414, 2015. **DOI**: 10.1587/transcom.E98.B.403.

Zoltán Belső graduates from the Eötvös Loránd University, Faculty of Science as a Computer Scientist (M.S. degree) in 1995. He also graduated as an Electrical Engineer (M.S. degree) from the Technical University of Budapest, Faculty of Electrical Engineering, Branch of Telecommunications in 2007. He is working at the Technical University of Budapest, Faculty of Electrical Engineering, Department of Telecommunications since graduated there as a part-time lecturer. He has worked on Unmanned Aerial Vehicle (UAV) communications systems and Quantum Key Distribution Systems (QKD).

László Pap graduated from the Technical University of Budapest, Faculty of Electrical Engineering, Branch of Telecommunications. He became Dr. Univ. and Ph.D. in 1980, and Doctor of Sciences in 1992. In 2001 and 2007 he has been elected as a Correspondent and Full Member of the Hungarian Academy of Sciences. His main fields of the research are the electronic systems, nonlinear circuits, synchronization systems, modulation and coding, spread spectrum systems, CDMA, multiuser detection and mobile communication systems. His main education activity has covered the fields of electronics, modern modulation and coding systems, communication theory, introduction to mobile communication. Professor Pap had been Head of the Dept. of Telecommunications, the Dean of the Faculty of Electrical Engineering at Budapest University of Technology and Economics, and Vice Rector of the University.

Reducing the Peak to Average Power Ratio in Optical NOMA Waveform Using Airy-Special Function based PTS Algorithm

Arun Kumar, *Senior Member, IEEE*, and Aziz Nanthaamornphong, *Senior Member, IEEE*

***Abstract*— This paper introduces a novel Peak-to-Average Power Ratio (PAPR) reduction technique for Non-Orthogonal Multiple Access (NOMA) waveforms, leveraging an Airy function-based Partial Transmit Sequence (PTS) method. The proposed technique is evaluated on NOMA waveforms with subcarrier configurations of 64, 256, and 512, and its performance is benchmarked against conventional PTS, Selective Mapping (SLM), and Clipping and Filtering methods. Comprehensive analysis is conducted on key metrics, including PAPR, Bit Error Rate (BER), and Power Spectral Density (PSD). Results demonstrate that the Airy-based PTS method achieves substantial PAPR reduction across all subcarrier scenarios, consistently surpassing traditional approaches. Furthermore, the proposed method maintains competitive BER performance, particularly in high subcarrier scenarios, where conventional methods typically face limitations. PSD analysis further highlights the spectral efficiency of the Airy-based PTS method, exhibiting minimal out-of-band emissions. These findings position the Airy-based PTS technique as a promising solution for improving NOMA waveform performance in 5G and beyond, achieving an optimal balance between PAPR reduction, BER, and spectral efficiency.**

***Index Terms*— PAPR, NOMA, Airy-PTS, BER, PSD**

I. Introduction

OPTICAL Non-Orthogonal Multiple Access (NOMA) waveforms offer enhanced spectral efficiency and improved user connectivity, marking a significant advancement in the domain of optical wireless communication. Unlike traditional orthogonal multiple access techniques, which allocate distinct frequency and temporal resources to individual users, NOMA enables multiple users to share the same frequency and temporal resources by differentiating them based on power levels [1]. In optical systems, this is achieved by modulating the light signal power for different users, allowing for simultaneous transmission and reception. This capability is pivotal for addressing the massive connectivity demands of emerging applications such as the Internet of Things, where numerous devices must communicate efficiently within limited spectral resources.

Optical NOMA stands out by enhancing spectral efficiency and overall system capacity through the use of successive interference cancellation (SIC) at the receiver and the superposition coding principle [2]. As the demand for higher data rates and efficient spectrum utilization continues to grow, Optical NOMA is poised to play a vital role in the evolution of future communication systems. This is particularly true for integrating optical technologies with advanced wireless paradigms like 5G and beyond.

However, the high Peak-to-Average Power Ratio (PAPR) of Optical NOMA waveforms poses a significant challenge to system performance. In optical communication systems, high PAPR leads to nonlinear distortions due to the limited dynamic range of optical transmitters such as light emitting diodes and laser diodes. These nonlinearities result in spectral regrowth and intermodulation distortions, degrading signal quality and increasing the Bit Error Rate (BER). Moreover, high PAPR forces optical transmitters to operate at lower average power levels to avoid clipping, which reduces the effective Signal-to-Noise Ratio (SNR) and limits communication range and data throughput [3].

In NOMA systems, where multiple users share the same spectrum, high PAPR exacerbates inter-user interference, complicating signal separation at the receiver. This increases decoding complexity and reduces overall system capacity. Consequently, effective PAPR management is crucial for ensuring the reliability and efficiency of Optical NOMA systems [4].

High PAPR can significantly degrade system performance by causing non-linear distortion and power inefficiency during transmission. In scenarios with high subcarrier counts, conventional techniques such as Selective Mapping (SLM) and clipping struggle to effectively manage the increased complexity and power fluctuations.

SLM generates multiple candidate sequences to reduce PAPR; however, this approach becomes computationally expensive and less efficient as the number of subcarriers increases. As a result, the return on investment for PAPR reduction diminishes with higher subcarrier configurations. Similarly, clipping introduces signal distortion and spectral regrowth, which are particularly problematic in dense communication systems.

To address these challenges, the smoothed Airy-Partial Transmit Sequence (PTS) method provides a more effective power control mechanism, offering a scalable and efficient solution for high subcarrier systems. This approach achieves significant PAPR reduction without introducing substantial computational complexity, making it an ideal candidate for modern high-capacity communication networks [5].

The clipping and filtering (C&F) approach reduces out-of-band distortion by clipping signal peaks that exceed a predetermined threshold. Despite its simplicity, this method can result in out-of-band radiation and signal distortion. SLM creates multiple signal versions using several phase sequences and selects the version with the lowest PAPR for transmission. While effective, SLM requires additional signaling overhead to transmit the phase sequence information.

A. Kumar is with Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, India (E-mail: arun.kumar1986@live.com)

A. Nanthaamornphong is with College of Computing, Prince of Songkla University, Phuket, Thailand (E-mail: aziz.n@phuket.psu.ac.th)

DOI: 10.36244/ICJ.2025.1.2

PTS reduces PAPR by dividing the input data into sub-blocks, optimizing the phase of each sub-block, and then combining them. Although this technique provides good PAPR reduction, it is computationally intensive. These methods often utilize specific coding schemes to limit the occurrence of high PAPR sequences. However, such schemes may reduce data rates. Tone Reservation reserves specific tones (subcarriers) to cancel out high peaks, offering a balance between complexity and effectiveness [5].

In this study, we propose the Airy-based PTS algorithm, which leverages the properties of Airy functions to reduce PAPR in optical NOMA. The Airy functions are employed to create smoother phase transitions, optimizing the phase combinations of sub-blocks in the PTS algorithm. This results in a reduced peak power level and, consequently, lower PAPR.

The Airy-based PTS method overcomes PAPR issues in optical NOMA by preserving signal integrity and minimizing the likelihood of distortion and clipping—problems commonly encountered in optical communication systems. Additionally, the use of Airy functions enhances the efficiency of the phase optimization process, leading to improved power efficiency and signal quality. This makes the proposed method highly effective for enhancing the overall performance of optical NOMA systems.

The key contributions of this article are summarized as follows:

1) This paper introduces a novel algorithm for PAPR reduction within the framework of optical NOMA systems using PTS. The algorithm exploits the Airy special function for subcarrier configurations of 64, 256, and 512. This is the first time the Airy special function has been employed to generate a reduced PAPR, phase-optimized signal through PTS for optical NOMA.
2) The Airy-function-based PTS algorithm significantly enhances the performance of optical waveform NOMA by achieving a substantially lower PAPR compared to traditional methods such as PTS, SLM, and C&F. This improvement minimizes signal distortion induced by high-power amplifiers, resulting in higher signal fidelity and system efficiency. The proposed method is particularly suitable for real-time applications in optical communication systems, especially in resource-constrained environments.
3) The proposed method achieves significant PAPR reduction while maintaining the BER performance of the framework. However, it retains a computational complexity comparable to that of traditional algorithms.

II. LITERATURE REVIEW

The authors in [6] investigated PAPR reduction in Frequency-domain NOMA (F-NOMA) using the SLM method. Their study demonstrated that SLM effectively minimizes PAPR by generating multiple candidate signals and selecting the one with the lowest PAPR. However, the approach significantly increases computational complexity and requires the transmission of side information, which reduces spectral efficiency and overall system performance in F-NOMA networks.

In [7], the authors explored PAPR reduction in NOMA-OFDM Visible Light Communication systems using a combination of Precoder and Companding methods. Their study demonstrated effective PAPR reduction and enhanced system performance in terms of spectral efficiency. However, this approach introduced additional computational complexity due to the combined processing, and the potential signal distortion caused by the Companding process may negatively impact the system's overall signal quality and reliability.

The authors in [8] proposed an efficient PAPR reduction scheme for OFDM-NOMA systems by combining Dynamic Subcarrier Indexing (DSI) and Precoding methods. This hybrid approach effectively reduced PAPR while maintaining system performance. However, the integration of DSI and Precoding increased computational overhead, posing implementation challenges, particularly in real-time scenarios.

In [9], a low-complexity SLM technique was proposed to reduce PAPR in downlink Power Domain OFDM-NOMA systems. This method efficiently reduced PAPR while maintaining system performance. Nonetheless, the need for side information transmission and the associated signal overhead remain significant drawbacks, potentially increasing system complexity and reducing data rates in practical applications.

The authors in [10] presented a hybrid method that combines SLM, PTS, and C&F techniques to minimize PAPR in optical NOMA systems. This approach effectively reduced PAPR while maintaining signal quality. However, the integration of multiple techniques significantly increased computational requirements, making implementation more challenging.

Lastly, a low-complexity PTS-SLM-Companding hybrid method for PAPR reduction in 5G NOMA waveforms was proposed in [11]. This method achieved considerable PAPR reduction while keeping computational complexity within acceptable limits. However, the strategy has drawbacks, including potential signal distortion caused by the Companding process and increased system complexity due to the integration of multiple approaches, which could negatively affect overall system performance.

In [12], the authors proposed an innovative three-layer hybrid technique that incorporates clipping, precoding, and coding to address PAPR issues in Filter Bank Multi-Carrier VLC systems. This method achieves a balance between PAPR reduction and computational complexity while maintaining system performance. However, its inability to dynamically adapt to network conditions limits its applicability in practical scenarios. Dynamic thresholding and real-time optimization could enhance the robustness of this approach for various deployment scenarios.

The study in [13] introduced a novel approach to mitigate PAPR in NOMA systems, which is crucial for future wireless networks. The proposed hybrid technique combines SLM and precoding, yielding significant PAPR reductions with low computational complexity and minimal impact on system throughput. However, the method's limitations include the lack of performance evaluation in dynamic multipath environments, limited scalability for massive MIMO-NOMA configurations, and insufficient consideration of hardware impairments, which may hinder real-world implementation.

In [14], the authors discussed the use of a companding scheme to enhance the performance of optical OFDM systems, with particular attention to addressing high PAPR. The study provided a detailed analysis of companding and its impact on metrics such as BER and spectral efficiency. While the scheme improves system robustness, the increased computational complexity poses a significant drawback. Future work should focus on optimizing the algorithm to achieve a better balance between complexity and performance.

The authors in [15] proposed an advanced PAPR reduction method for DCO-OFDM systems based on multi-point constellations and a discrete particle swarm optimization (DPSO) algorithm. This method demonstrated efficient PAPR reduction while preserving signal quality, with substantial performance improvements over traditional methods. However, the reliance on computationally intensive DPSO optimization makes the method less feasible for real-time implementation. Additionally, its scalability to higher-order modulation schemes or large-scale systems remains unexplored, limiting its broader applicability.

In [16], a joint optimization approach was presented to enhance Quality of Service and reduce PAPR in energy-efficient massive MIMO systems. This technique integrates precoding and resource allocation strategies, achieving significant improvements in energy efficiency and signal quality. However, its reliance on idealized channel conditions may not reflect real-world multipath fading scenarios. Moreover, the computational complexity of the optimization process presents challenges for practical implementation in large-scale systems.

The authors in [17] employed optimization-based methods to improve PAPR reduction techniques for OFDM signals, enhancing wireless communication system performance. The study effectively measured clipping, tone reservation, and active constellation extension techniques in terms of PAPR and error performance improvements. However, the reliance on static channel conditions limits its applicability in dynamic environments. Additionally, the optimization algorithms introduce high computational complexity, which may render the approach unsuitable for real-time implementation in low-power or latency-sensitive applications.

Table I provides a comparative analysis of the PAPR reduction algorithms discussed in these studies.

TABLE I
COMPARATIVE ANALYSIS OF PAPR ALGORITHMS

PAPR Algorithms	**PAPR at CCDF of 10^{-3}**	**SNR at BER of 10^{-3}**
DSI & precoding method [7]	4.9 dB	13 dB
SLM [8]	6.8 dB	19 dB
SLM-PTS-CT [9]	7.1 dB	21.1 dB
PTS-SLM [10]	4.9 dB	5.8 dB
Amplitude clipping-SLM based Lifting Wavelet Transform [11]	6.8 dB	4 dB
Salp Swarm Algorithm-based PTS [12]	5.8 dB	10 dB
Companding methods [13]	4.9 dB	11 dB
Multi-point constellation method-SLM [14]	6.2 dB	9.8 dB
Bidirectional long short-term memory autoencoder [15]	7 dB	Not Simulated
Gradient method [16]	6 dB	6.2 dB
DSI & precoding method [17]	6.8 dB	Not Simulated
Proposed Airy-based PTS for 64 sub-carriers	2.8 dB	Not Simulated
Proposed Airy-based PTS for 256 sub-carriers	4.7 dB	6 dB
Proposed Airy-based PTS for 512 sub-carriers	8.7 dB	7.1 dB

III. SYSTEM MODEL

In an optical NOMA system incorporating SIC and Super Coding, the block diagram typically comprises several key components. At the transmitter, multiple users' data are encoded onto a single optical signal. Super Coding is utilized to enhance data transmission efficiency by encoding the signals in a manner that facilitates improved recovery and error correction.

The encoded data are modulated onto the optical carrier using techniques such as amplitude modulation or phase modulation. The modulated signal is then transmitted through an optical channel, which may consist of fiber optics or free-space optical links. During transmission, the signal quality is influenced by attenuation and noise introduced by the channel.

At the receiver end, the optical signal is detected and converted back into an electrical signal. SIC is employed for decoding the signals. The receiver first decodes the most significant signal and then iteratively subtracts it from the received signal to decode the remaining lower-power signals, effectively mitigating interference. To restore the original data, the demodulated signals undergo decoding, where super coding techniques are applied to enhance data reliability and correct errors [18].

The fundamental principles of resource allocation and signal processing in NOMA systems are represented in the mathematical model of a NOMA waveform. These principles form the basis for optimizing the performance and efficiency of optical NOMA systems.

The proposed work addresses a critical challenge in optical NOMA systems, where high PAPR negatively impacts performance and power efficiency. Optimization strategies aimed at mitigating this issue should prioritize refining the parameters of the Airy-special function to achieve an effective balance between complexity and PAPR reduction. Incorporating machine learning algorithms to dynamically select optimal phase rotation factors in the PTS algorithm could significantly enhance adaptability across varying channel conditions.

Further improvements in PAPR reduction could be achieved by combining this approach with complementary techniques such as companding or precoding. From a practical perspec-

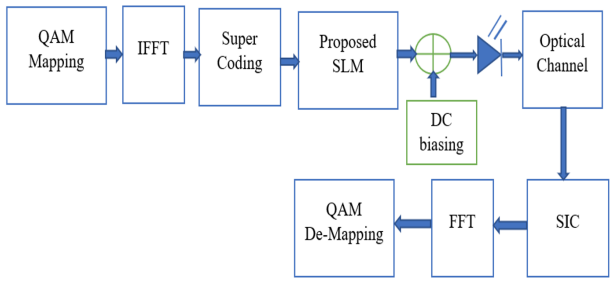


Fig. 1. Structure of Optical NOMA waveform

tive, lightweight computational methods must be developed to ensure real-time feasibility in resource-constrained optical networks. To alleviate computational overhead, hardware accelerators such as field-programmable gate arrays (FPGAs) or Graphics Processing Units (GPUs) could be integrated with the PTS algorithm.

Experimental validation of the PTS algorithm based on the Airy-special function is necessary for a variety of optical channel scenarios, including high-bandwidth and multi-user environments, to ensure robustness. Compatibility with existing optical NOMA standards and demonstration of interoperability would be essential to achieve broader acceptance of the proposed scheme.

Moreover, integration into system-level designs, such as visible light communication networks or 6G optical links, could significantly amplify its practical impact. These efforts could facilitate the development of efficient, scalable, and reliable optical communication systems, thereby advancing the state-of-the-art in optical NOMA technologies. Fig. 1 shows the optical NOMA's structure.

In NOMA systems, multiple users share the same frequency resources, but they are assigned different power allocations. Let $x(t)$ represent the transmitted optical signal, which is a superposition of signals from multiple users:

$$x(t) = \sum_{k=1}^{K} s_k \alpha_k(t) \exp\left(j\phi_k(t)\right), \tag{1}$$

where K is the number of users, $\alpha_k(t)$ denotes the amplitude of the signal, $s_k(t)$ represents the baseband signal, and $\phi_k(t)$ is the phase modulation for the k-th user.

The PAPR of the signal $x(t)$ is defined as:

$$\text{PAPR} = \frac{\max_t |x(t)|^2}{E\left[|x(t)|^2\right]}, \tag{2}$$

where $\max_t |x(t)|^2$ is the maximum instantaneous power, and $E\left[|x(t)|^2\right]$ is the average power.

After the signal is transmitted through the optical channel, the received signal $y(t)$ can be expressed as:

$$y(t) = x(t) \cdot h(t) + n(t), \tag{3}$$

where $h(t)$ is the channel impulse response, and $n(t)$ is the noise.

At the receiver, SIC is applied. The iterative decoding process begins by decoding the signal of the strongest user and subtracting it from $y(t)$, thereby reducing interference for decoding weaker signals. Super coding is employed to enhance reliability by encoding the signals with error-correcting codes. Let $c_k(t)$ represent the encoded signal for the k-th user:

$$s_k(t) = \text{Encode}\left(d_k(t)\right), \tag{4}$$

where $d_k(t)$ is the data to be transmitted, and Encode$(\cdot)$ denotes the encoding operation.

This formulation captures the fundamental concepts of power allocation, signal superposition, PAPR reduction, and interference cancellation in NOMA systems, emphasizing the critical role of SIC and super coding in improving system performance and reliability.

A. *Proposed airy-PTS Method*

The Airy function plays a critical role in enhancing the PAPR performance in the Airy-PTS method due to its ability to enable more precise phase optimization. Unlike conventional PTS techniques, where phase rotation factors are selected at random or based on predefined methods that may not ensure optimal PAPR reduction, the Airy function introduces unique mathematical properties that significantly improve the phase selection process. As a solution to the Airy differential equation, the Airy function exhibits oscillatory behavior with well-defined features. These oscillations facilitate smooth and continuous phase variation, leading to a more uniform distribution of signal power.

Analytically, the Airy function represents wave propagation and interference phenomena, allowing for refined phase selection compared to basic phase rotation techniques. By leveraging the Airy function, the Airy-PTS algorithm minimizes high peaks in the power spectrum of the signal, resulting in more evenly distributed power. This reduction in sharp power spikes leads to less distortion and greater efficiency, particularly in optical communication systems where power efficiency is of utmost importance.

The Airy-based PTS algorithm for PAPR reduction in NOMA waveforms utilizes Airy functions to optimize phase shifts and enhance PAPR performance. In this method, the signal is divided into multiple sub-blocks, and the Airy function, known for its smooth phase transition properties, is used to generate phase sequences for these sub-blocks [19]. Each phase sequence is applied to the sub-blocks to create different phase-adjusted versions of the signal. The version with the lowest PAPR is then selected for transmission.

This approach leverages the Airy function's capability to provide precise phase optimization, reducing peak power variations and improving signal uniformity. Consequently, the Airy-based PTS algorithm effectively mitigates PAPR issues while maintaining signal integrity and reducing computational complexity compared to conventional PTS methods. This results in enhanced performance in NOMA systems, particularly in optical communication scenarios.

The Airy-based Partial Transmit Sequence (PTS) method for PAPR reduction involves several mathematical steps. Let

the transmitted signal $x(t)$ be divided into M sub-blocks. The signal $x(t)$ can be expressed as:

$$x(t) = \sum_{m=1}^{M} x_m(t) \cdot \exp(j\theta_m), \tag{5}$$

where $x_m(t)$ represents the m-th sub-block, and θ_m is the phase adjustment applied to the sub-block.

The Airy function $\mathrm{Ai}(t)$ is utilized to generate the phase sequences ϕ_m for each sub-block, given by:

$$\phi_m = \mathrm{Ai}(\alpha_m), \tag{6}$$

where α_m is a parameter that controls the phase adjustment.

These phase sequences are then applied to each sub-block. The adjusted signal $x_{\text{adjusted}}(t)$, with Airy-based phases, is given as:

$$x_{\text{adjusted}}(t) = \sum_{m=1}^{M} x_m(t) \cdot \exp(j\phi_m). \tag{7}$$

Finally, the PAPR of the adjusted signal is calculated as:

$$\text{PAPR} = \frac{\max_t \left|x_{\text{adjusted}}(t)\right|^2}{\text{Avg}\left[\left|x_{\text{adjusted}}(t)\right|^2\right]}. \tag{8}$$

This formulation highlights the use of the Airy function for generating smooth and precise phase adjustments, which minimizes the PAPR effectively in NOMA systems, particularly in optical communication scenarios.

B. Complexity

The PTS method generates several phase sequences (or sub-blocks), with each sub-block undergoing a phase rotation. Let N denote the number of sub-blocks. For each sub-block, the algorithm performs a phase rotation with a complexity of $O(1)$. Consequently, the overall complexity for producing all possible sub-blocks in the PTS method is $O(N)$.

The proposed phase optimization method employs the Airy-special function. Let M represent the number of phase rotations evaluated for each sub-block. The evaluation of the Airy function is computationally intensive, as it often involves solving a differential equation. Assuming the evaluation cost of one Airy function is $O(A)$, the total cost for evaluating the Airy function across all phase rotations of all sub-blocks is $O(N \cdot M \cdot A)$.

The optimization step involves selecting the best phase combination that minimizes the PAPR. This step typically requires an exhaustive search over all possible phase combinations. If K denotes the number of possible phase combinations, the complexity of this exhaustive search is $O(K)$. However, advanced search techniques, such as gradient descent or evolutionary algorithms, can reduce the complexity of this step.

Combining all major operations, the overall complexity of the Airy-PTS method can be approximated as:

$$O(N \cdot M \cdot A + K). \tag{9}$$

For practical applications, particularly in large-scale systems, the values of N and M can result in high computational requirements. Complexity reduction is therefore crucial, and optimizations in the evaluation of the Airy function and the design of efficient search algorithms are necessary. Additionally, hardware acceleration techniques, such as the use of FPGAs or GPUs, can further expedite the evaluation of the Airy function and phase rotation processes. Table II provides a detailed complexity analysis of the PAPR reduction algorithms, highlighting the computational requirements associated with each method.

TABLE II
COMPLEXITY ANALYSIS OF THE PAPR ALGORITHMS [20]

PAPR Algorithms	Complexity	Remarks
Airy-PTS	$O(N \cdot M \cdot A + K)$	The Airy-Special Function-based PTS involves generating multiple phase sequences, evaluating the Airy function for phase optimization, and searching for the optimal combination. Complexity is high due to Airy function evaluations and exhaustive search.
PTS	$O(N \cdot M)$	Involves phase rotation and evaluation of the signal for each sequence, where N is the number of sub-blocks and M is the number of phase shifts. Simpler than Airy-PTS but still requires exhaustive search over possible combinations.
SLM	$O(N \cdot M)$	The complexity involves generating multiple candidate sequences (with N being the number of sub-blocks and M the number of phase shifts), then searching for the one with the lowest PAPR. Less computationally intensive than PTS but still requires phase optimization.
Companding	$O(N)$	This method involves compressing the amplitude of the signal to reduce PAPR. It is computationally efficient, involving a simple compression step, but might not be as effective in reducing PAPR as PTS or SLM.
SLM-PTS	$O(N \cdot M \cdot P)$	Combines both the SLM and PTS methods, leading to higher complexity than individual methods. P is the number of possible phase combinations. The method offers better PAPR reduction but at the cost of increased complexity.

IV. SIMULATION RESULTS

This section presents the implementation of the Airy-based Partial Transmit Sequence (PTS) method using MATLAB 2016. The simulation parameters include 200,000 symbols, a 256-point FFT, 256-QAM modulation, sub-carrier configurations of 64, 256, and 512, and a NOMA waveform transmitted over a Rayleigh fading channel.

The CCDF of NOMA for 64 sub-carriers is shown in Fig. 2. At a CCDF of 10^{-3}, the PAPR achieved by the proposed Airy-PTS method is 2.8 dB, compared to 4.8 dB for the PTS, 6.2 dB for the SLM, 7.4 dB for the C&F technique, and 8.7

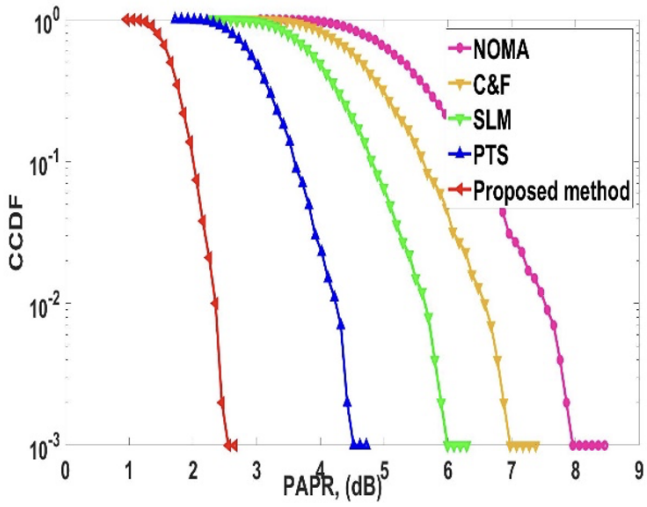


Fig. 2. PAPR analysis of 64 sub-carriers for NOMA waveform

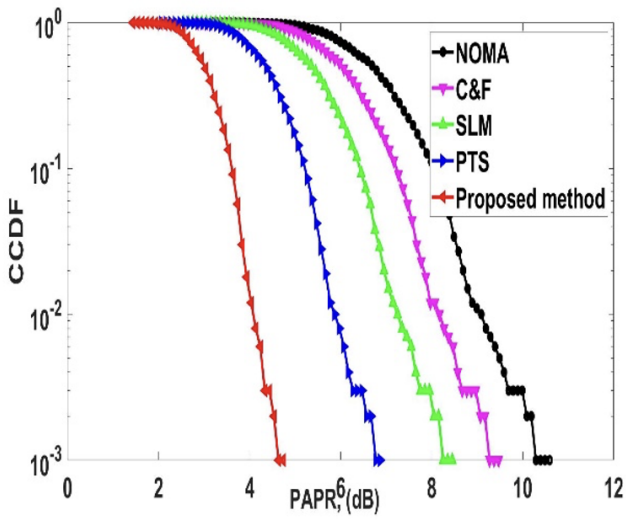


Fig. 3. PAPR analysis of 256 sub-carriers for NOMA waveform

dB for the original NOMA waveform. The proposed scheme demonstrates a PAPR gain of 2 dB to 5.9 dB compared to the conventional systems.

It is observed that the Airy special function efficiently determines the phase factor, which balances the NOMA symbol and reduces high peak signals. Therefore, it is concluded that NOMA systems with a smaller number of sub-carriers exhibit better PAPR performance compared to systems with a higher number of sub-carriers.

The PAPR curves for 256 sub-carriers in the NOMA waveform, both with and without precoding schemes, are depicted in Fig. 3. At a CCDF of 10^{-3}, the PAPR values of 4.8 dB, 5.7 dB, 6.9 dB, 8.3 dB, 9.4 dB, and 10.8 dB are achieved by the proposed Airy-PTS method, C&F, PTS, SLM, and the original NOMA signal, respectively.

It is observed that for higher sub-carrier configurations (256 sub-carriers), the PAPR is relatively high. However, the proposed Airy-PTS scheme effectively reduces the PAPR, resulting in a high-performance radio system. Therefore, it is concluded that the Airy-special function-based PTS method can be efficiently employed in large sub-carrier radio systems to achieve significant performance gains.

Fig. 4 illustrates the throughput performance of the NOMA waveform for 512 sub-carriers using various PAPR reduction algorithms. The primary objective is to mitigate the high peaks in the radio waveform for a large number of sub-carriers, enabling advanced radio systems to operate more effectively.

At a CCDF of 10^{-3}, the PAPR values achieved are 8.7 dB, 10 dB, 11.3 dB, 12.6 dB, and 14.2 dB for the proposed Airy-PTS method, C&F, PTS, SLM, and the original NOMA signal, respectively. The proposed Airy-based PTS method outperformed conventional schemes by reducing the PAPR by 1.3 dB to 4.7 dB compared to the traditional methods.

Thus, it can be concluded that the proposed Airy-PTS approach is well-suited for sophisticated radio systems employing a large number of sub-carriers. This enables the

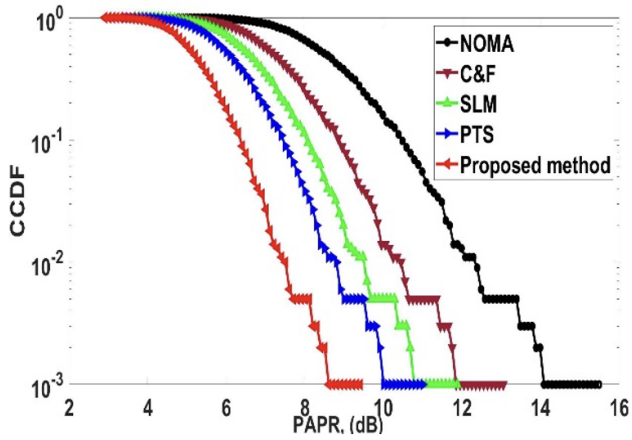


Fig. 4. PAPR analysis of 512 sub-carriers for NOMA waveform

achievement of high data rates, efficient spectrum utilization, and enhanced system capacity, meeting the demands of modern subscriber requirements.

Fig. 5 illustrates the BER curves for 256 sub-carriers, comparing the proposed Airy-based PTS method with conventional PAPR reduction algorithms. Evaluating throughput performance with a large number of sub-carriers is critical to assessing the effectiveness of the proposed algorithm. The BER of conventional algorithms shows degradation as the number of sub-carriers increases.

The proposed Airy-PTS method achieves a BER of 10^{-3} at an SNR of 6 dB, compared to 6.9 dB for the PTS, 8.1 dB for the SLM, 9.1 dB for the C&F technique, and 10.2 dB for the NOMA waveform. This indicates that the Airy-PTS method demonstrates significant BER performance improvements at lower SNR levels compared to traditional schemes.

Furthermore, the Airy-PTS method outperforms contempo-

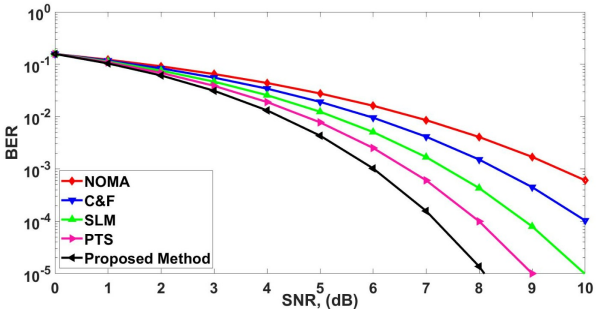


Fig. 5. BER analysis of 256 sub-carriers for NOMA waveform

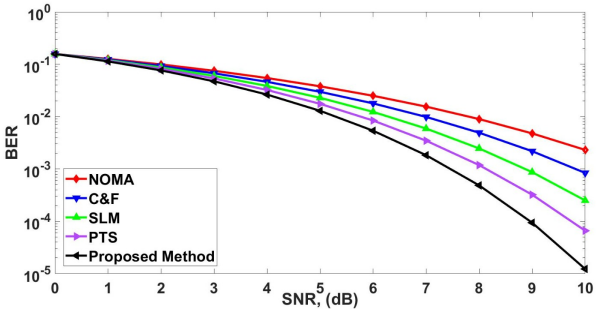


Fig. 6. BER analysis of 512 sub-carriers for NOMA waveform

rary algorithms by achieving SNR gains of 0.9 dB, 2.1 dB, 3.1 dB, and 4.2 dB over the PTS, SLM, C&F, and NOMA waveform techniques, respectively. These results highlight the efficiency of the proposed method in improving BER performance for systems with a large number of sub-carriers.

Fig. 6 presents the BER curves for 512 sub-carriers in NOMA systems. Evaluating the system's throughput in conjunction with PAPR performance is crucial for assessing the BER retention capability of the algorithms.

The proposed Airy-PTS method achieves a BER of 10^{-3} at an SNR of 8.1 dB, compared to 9 dB for the C&F technique, 10.1 dB for the PTS, 11.6 dB for the SLM, and 12.7 dB for the original NOMA signal. These findings demonstrate that the Airy-PTS method achieves SNR gains of 0.9 dB, 2 dB, 3.5 dB, and 4.6 dB over the C&F, PTS, SLM, and original NOMA techniques, respectively.

In conclusion, the proposed algorithm not only achieves optimal PAPR performance but also effectively retains BER performance, making it a suitable candidate for advanced NOMA systems.

V. CONCLUSION

This study evaluated the PAPR reduction of NOMA waveforms using the Airy-special function-based PTS method for subcarrier configurations of 64, 256, and 512. The proposed method's performance was compared against conventional PTS, SLM, and C&F techniques, considering metrics such as PAPR, BER, and PSD.

The Airy-based PTS method demonstrated significant PAPR reduction across all subcarrier configurations, consistently outperforming conventional techniques, particularly as the number of subcarriers increased. BER analysis revealed that the proposed method maintained better signal integrity compared to C&F, which is prone to BER degradation. Additionally, PSD analysis indicated that the Airy-PTS approach achieved a more favorable spectral profile by effectively reducing out-of-band emissions.

Despite its advantages, the study highlighted some limitations, including increased computational complexity and potential challenges in real-time implementations, especially for configurations with a high number of subcarriers. Future research could focus on optimizing the Airy-based PTS method to reduce computational complexity and exploring hybrid approaches that integrate Airy functions with other PAPR reduction techniques to further enhance performance.

REFERENCES

[1] A. Kumar, "Analysis of PAPR on NOMA waveforms using hybrid algorithm," *Wireless Personal Communications*, vol. 132, no. 3, pp. 1849–1861, 2023, DOI: 10.1007/s11277-023-10683-y.

[2] R. Sayyari, J. Pourrostam, and H. Ahmadi, "Efficient PAPR reduction scheme for OFDM-NOMA systems based on dsi & precoding methods," *Physical Communication*, vol. 47, p. 101 372, 2021, DOI: 10.1016/j.phycom.2021.101372.

[3] ——, "A low complexity PTS-based papr reduction method for the downlink of OFDM-NOMA systems," in *2022 IEEE wireless communications and networking conference (WCNC)*. IEEE, 2022, pp. 1719–1724, DOI: 10.1109/WCNC51071.2022.9771812.

[4] A. Kumar, K. Rajagopal, N. Alruwais, H. M. Alshahrani, H. Mahgoub, and K. M. Othman, "Papr reduction using SLM-PTS-CT hybrid papr method for optical NOMA waveform," *Heliyon*, vol. 9, no. 10, 2023, DOI: 10.1016/j.heliyon.2023.e20901.

[5] S. Singhal and D. K. Sharma, "A review and comparative analysis of PAPR reduction techniques of OFDM system," *Wireless Personal Communications*, pp. 1–27, 2024, DOI: 10.1007/s11277-024-11074-7.

[6] N. A. George and S. K. Mishra, "PAPR reduction in F-NOMA using selective mapping method," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2023, pp. 1–5, DOI: 10.1109/ICCCNT56998.2023.10308006.

[7] N. Sharan, S. Ghorai, and A. Kumar, "PAPR reduction using a precoder and compander combination in a NOMA-OFDM VLC system," in *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2022, pp. 1–4, DOI: 10.1109/AISP53593.2022.9760659.

[8] R. Sayyari, J. Pourrostam, and H. Ahmadi, "Efficient PAPR reduction scheme for OFDM-NOMA systems based on dsi & precoding methods," *Physical Communication*, vol. 47, p. 101 372, 2021, DOI: 10.1016/j.phycom.2021.101372.

[9] M. Mounir, M. I. Youssef, and A. M. Aboshosha, "Low-complexity selective mapping technique for papr reduction in downlink power domain OFDM-NOMA," *EURASIP Journal on Advances in Signal Processing*, vol. 2023, no. 1, p. 10, 2023, DOI: 10.1186/s13634-022-00968-y.

[10] A. Kumar, N. Gaur, and A. Nanthaamornphong, "PAPR reduction using model-driven hybrid algorithms in the 6G NOMA waveform," *Internet Technology Letters*, vol. 7, no. 6, p. e515, 2024, DOI: 10.1002/itl2.515.

[11] A. Kumar, "A low complex PTS-SLM-companding technique for PAPR reduction in 5G NOMA waveform," *Multimedia Tools and Applications*, vol. 83, no. 15, pp. 45 141–45 162, 2024, DOI: 10.1007/s11042-023-17223-7.

[12] E. S. Hassan, "Three layer hybrid PAPR reduction method for NOMA-based FBMC-VLC networks," *Optical and Quantum Electronics*, vol. 56, no. 5, p. 890, 2024, DOI: 10.1007/s11082-024-06724-w.

[13] A. A. M. Inam Abousaber and H. F. Abdallah, "Novel techniques for efficient PAPR reduction in NOMA systems for future wireless networks," *Journal of Information and Telecommunication*, vol. 0, no. 0, pp. 1–17, 2025, DOI: 10.1080/24751839.2025.2454056.

[14] P. Kanjilal, A. Kumar, S. Bhowmick, J. P. Maroor, and A. Nanthaamornphong, "Implementation of companding scheme for performance enhancement of optical OFDM structure," *Journal of Optical Communications*, no. 0, 2024, DOI: 10.1515/joc-2024-0095.

[15] V. Aydin and G. Hacioglu, "Enhanced PAPR reduction in DCO-OFDM using multi-point constellations and DPSO optimization," *Neural Computing and Applications*, vol. 36, no. 11, pp. 5747–5756, 2024, DOI: 10.1007/s00521-023-09409-9.

[16] S. Bolla and M. Singh,"Joint optimization-based QoS and PAPR reduction technique for energy-efficient massive MIMO system," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 235, 2024, DOI: 10.1007/s44196-024-00648-9.

[17] S. Elaage, A. Hmamou, M. E. Ghzaoui, and N. Mrani, "PAPR reduction techniques optimization-based ofdm signal for wireless communication systems," *Telematics and Informatics Reports*, vol. 14, p. 100137, 2024, DOI: 10.1016/j.teler.2024.100137.

[18] A. Kumar, N. Gour, H. Sharma, and R. Pareek, "A hybrid technique for the PAPR reduction of NOMA waveform," *International Journal of Communication Systems*, vol. 36, no. 4, p. e5412, 2023, DOI: 10.1002/dac.5412.

[19] R. W. Ibrahim and D. Baleanu, "Ona new linear operator formulated by airy functions in the open unit disk," *Advances in Difference Equations*, vol. 2021, no. 1, p. 366, 2021, DOI: 10.1186/s13662-021-03527-1.

[20] R. Niwareeba, M. A. Cox, and L. Cheng, "Low complexity hybrid SLM for PAPR mitigation for ACO OFDM," *ICT Express*, vol. 8, no. 1, pp. 72–76, 2022, DOI: 10.1016/j.icte.2021.10.002.

Arun Kumar received his Ph.D. in electronics and communication engineering from JECRC Univer- sity, Jaipur, India. He is an Associate Professor in Electronics and Communication Engineering at New Horizon College of Engineering in Bengaluru, India. Dr. Kumar has a total of 10 years of teaching experience and has published more than 85 research articles in SCI-E and Scopus Index journals. His research interests are advanced waveforms for 5G mobile communication systems and 5G-based smart hospitals, PAPR reduction techniques in the multi-carrier waveform, and spectrum sensing techniques. Dr. Kumar has successfully implemented different reduction techniques for multi-carrier waveforms such as NOMA, FBMC, UFMC, and so on, and has also implemented and compared different waveform techniques for the 5G system. Currently, he is working on the requirements of a 5Gbased smart hospital system. He is a senior member of the IEEE and a reviewer for many refereed, indexed journals.

Aziz Nanthaamornphong is Associate Professor and serves as the Dean at the College of Computing at Prince of Songkla University's Phuket campus in Thailand. He earned his Ph.D. degree from the University of Alabama, USA. With an extensive academic background, Dr. Nanthaamornphong specializes in empirical software engineering and data science, among other areas. His research significantly contributes to the development of scientific software and leverages data science in the field of tourism. In addition to his core focus, he is also deeply engaged in the study of human-computer interaction, pioneering innovative approaches to foster a beneficial interplay between humans and technology. For correspondence, he can be reached at aziz.n@phuket.psu.ac.th.

Channel Estimation Methods in Massive MIMO: A Comparative Review of Machine Learning and Traditional Techniques

Amalia Eka Rakhmania, Hudiono, Umi Anis Ro'isatin, and Nurul Hidayati

***Abstract*—Massive Multiple Input Multiple Output (MIMO) has emerged as a crucial technology in 5G and future 6G networks, offering unprecedented improvements in capacity, energy efficiency, and spectral efficiency. A key challenge for Massive MIMO systems is accurate and efficient channel estimation, which significantly impacts system performance. Traditional channel estimation methods such as Least Squares (LS) and Minimum Mean Square Error (MMSE) have been widely employed, but their limitations, particularly in complex and dynamic environments, have led to the exploration of more sophisticated approaches, including machine learning (ML)-based techniques. This review aims to compare traditional channel estimation methods with modern machine learning-based techniques in Massive MIMO systems, providing insights into their performance, computational complexity, and scalability. Furthermore, this paper outlines potential future research directions, emphasizing the integration of machine learning, optimization techniques, and hardware-friendly designs for enhanced performance.**

***Index Terms*—comparative study, machine learning, massive MIMO, traditional methods.**

I. Introduction

Massive MIMO (Multiple Input Multiple Output) is a revolutionary technology in wireless communication that enhances the capacity and efficiency of networks. It involves the deployment of a large number of antennas at the base station, allowing for the simultaneous transmission and reception of data to multiple users within the same frequency band. This capability significantly improves spectral efficiency and overall network performance, making it a key component in modern wireless systems, especially in the context of 5G and beyond.

The concept of Massive MIMO is built on the principles of spatial multiplexing and beamforming, which enable the base station to serve multiple users by exploiting the spatial dimensions of the wireless channel. By using hundreds of antennas, Massive MIMO can create highly directional beams that focus energy toward specific users, thereby reducing interference and improving signal strength. This technology not only increases capacity but also enhances energy efficiency, as it can adaptively allocate resources based on user demand and channel conditions [1]. However, this benefit comes with challenges, particularly in terms of accurate channel estimation, which is essential for the success of beamforming and resource allocation algorithms [2].

Channel estimation in Massive MIMO is inherently difficult due to the large number of antennas and the complexity of the wireless channel in high-mobility and dense environments. Traditional estimation methods such as LS [3] and MMSE [4] offer basic solutions but fail to cope effectively with increasing system complexity. These limitations have led to the application of machine learning-based techniques, which leverage large datasets and complex models to learn the channel's characteristics and provide more robust solutions.

Recent advancements in deep learning, particularly convolutional neural networks, have shown promise in enhancing channel estimation accuracy by capturing spatial correlations and temporal dynamics more effectively than traditional methods. Moreover, the integration of reinforcement learning approaches has opened new avenues for adaptive channel estimation, allowing systems to dynamically adjust their parameters based on real-time feedback from the environment.

This paper provides a comprehensive review of both traditional and ML-based methods for channel estimation in Massive MIMO, discusses their strengths and weaknesses, and explores possible future developments.

II. Traditional Channel Estimation Methods

A. Least Squares (LS)

The Least Squares (LS) method is a widely used approach for channel estimation in wireless communication systems, particularly in OFDM and MIMO configurations, due to its simplicity and computational efficiency [5], [6]. While LS is simple and computationally efficient, it suffers from high mean square error, especially at low signal-to-noise ratios [7]. These approaches aim to improve estimation accuracy while maintaining low complexity.

This work was supported by State Polytechnic of Malang, Indonesia under the DIPA Funding under the contract SP DIPA-023.18.2.77606/2024. *Corresponding author: Amalia Eka Rakhmania.*

Amalia Eka Rakhmania is with the Department of Electrical Engineering, State Polytechnic of Malang, Indonesia (e-mail: amaliaeka.rakhmania@polinema.ac.id).

Hudiono is with the Department of Electrical Engineering, State Polytechnic of Malang, Indonesia (e-mail: hudiono@polinema.ac.id).

Umi Anis Ro'isatin is with the Department of Mechanical Engineering, State Polytechnic of Malang, Indonesia (e-mail: umi.anis@polinema.ac.id).

Nurul Hidayati is with the Department of Electrical Engineering, State Polytechnic of Malang, Indonesia (e-mail: nurulhid8@polinema.ac.id).

DOI: 10.36244/ICJ.2025.1.3

Least squares methods are pivotal in channel estimation for MIMO (Multiple Input Multiple Output) and Massive MIMO systems, where accurate channel state information (CSI) is essential for optimizing transmission strategies. These methods focus on estimating the characteristics of communication channels by minimizing the sum of the squares of the differences between observed and estimated values, which is particularly relevant in MIMO systems that utilize multiple antennas at both the transmitter and receiver to enhance communication performance.

In MIMO systems, the least squares method serves as a low-complexity design approach for parameter estimation, allowing for effective retrieval of channel state information even when pilot sequences are limited. The linear least squares problem, often referred to as regression analysis, provides a closed-form solution that is beneficial for estimating parameters in these complex systems. This is crucial because the performance of MIMO systems heavily relies on accurate channel estimation to mitigate the effects of noise and interference, which can significantly degrade transmission quality. Moreover, the application of least squares methods in Massive MIMO systems is particularly advantageous due to the large number of antennas involved. These systems can leverage the additional antennas to compensate for the reduced number of pilot signals, thus maintaining reliable channel estimation. The estimation error analysis is also vital, as it evaluates the accuracy of the channel estimates obtained through least squares methods, helping to improve overall system performance and reliability.

Spatial multiplexing, a technique that allows multiple data streams to be transmitted simultaneously over the same channel, further illustrates the importance of least squares methods in maximizing data rates in MIMO systems. Accurate channel estimation is essential for effectively implementing spatial multiplexing, as it directly impacts the system's ability to handle multiple independent data streams without interference. LS assumes that the channel is static and deterministic. It estimates the channel coefficients by solving a system of linear equations. The solution that minimizes the sum of squared errors between the estimated and actual received signals is chosen as the channel estimate.

In Massive MIMO systems, the LS estimator is commonly employed for uplink channel estimation, although its performance is highly dependent on the choice of training pilots and is sensitive to outlier measurements. To address these challenges, techniques such as sparse Bayesian learning (SBL) have been integrated with LS to improve estimation accuracy by exploiting channel sparsity and separating impulsive noise from the signal of interest [8]. Additionally, combining LS with singular value decomposition (SVD) has been proposed to enhance channel estimation accuracy by using SVD to calculate the initial channel matrix, followed by LS signal detection to refine the channel state information (CSI) [9], [10]. In MIMO-OFDM systems, LS estimation is used alongside adjustable phase shift pilots (APSPs) to reduce pilot overhead and improve the mean square error (MSE) of the channel estimate [11], [12].

Furthermore, LS methods are favored in 5G wireless communications for their practicality and ease of implementation, despite being less accurate than minimum mean square error (MMSE) methods, which require channel statistics [13]. The LS method's performance can be enhanced by increasing the number of base station antennas, which improves the bit error rate (BER) [14].

The Least Squares (LS) method in channel estimation for Massive MIMO systems offers several advantages and disadvantages. One of the primary advantages of LS is its simplicity and ease of implementation, making it resource-friendly and practical for industry applications [13]. LS does not require prior statistical knowledge of the channel, which is beneficial in scenarios where such information is unavailable [15]. Additionally, LS can be computationally efficient, especially when optimized to minimize the relative error between estimated and actual channel coefficients, leading to faster data processing [13].

However, LS has notable disadvantages, particularly its poor performance in low signal-to-noise ratio (SNR) environments, where it provides less accurate channel estimates compared to more sophisticated methods like Minimum Mean Square Error (MMSE) [16], [15]. LS can also introduce significant modeling errors when used to decouple pilot matrices, which can affect the accuracy of channel estimation [17]. Furthermore, LS is less effective in handling pilot contamination and interference, which can degrade the uplink rate in Massive MIMO systems [18]. Despite these limitations, LS remains a widely used method due to its straightforward implementation and low computational complexity, making it suitable for scenarios where computational resources are limited [19].

While LS is straightforward to use and efficient in terms of computation, it does have some drawbacks. Firstly, it is vulnerable to interference, particularly in noisy conditions. Secondly, LS ignores any previous statistical data about the channel, which can affect its precision. Thirdly, LS requires many pilot symbols to obtain reliable channel estimates. Although LS techniques are fundamental in estimating channels for MIMO systems, combining them with other methods and algorithms is essential for overcoming their shortcomings and improving performance in large-scale MIMO scenarios.

B. Minimum Mean Square Error (MMSE)

Besides being used as a signal detection technique [20], the Minimum Mean Square Error (MMSE) method addresses the limitations of LS by incorporating noise statistics. It aims to minimize the mean squared error between the estimated and actual channel response. This approach generally provides better accuracy than LS, especially in noisy conditions.

The Minimum Mean Square Error (MMSE) method is a prominent channel estimation technique in Massive MIMO systems, known for its high accuracy in acquiring channel state information (CSI) essential for optimal system performance. MMSE assumes that the channel is a random variable with known statistical properties. It estimates the channel coefficients by minimizing the expected squared error between the estimated and actual channel response. This minimization is achieved by using the channel's prior distribution and the noise statistics. MMSE estimators are particularly effective in environments with spatially correlated Rician fading, where

they can achieve improved normalized mean square error (NMSE) as the Rician K-factor decreases, indicating better performance under Rayleigh fading conditions [21]. However, the classical linear MMSE estimator is computationally intensive, especially in Massive MIMO contexts, prompting the development of alternative methods like the rank-1 subspace channel estimator, which offers lower complexity while maintaining high accuracy [22].

The MMSE method's reliance on accurate channel covariance matrices is a critical factor, as imperfections in these matrices can significantly affect estimation accuracy [23]. To address this, techniques such as the generalized eigenvalue decomposition (GEVD) have been proposed to estimate low-rank channel covariance matrices, enhancing MMSE performance in uplink cellular systems [24]. Additionally, model-based approaches and Bayesian estimators have been explored to reduce computational complexity while maintaining estimation quality [25], [23]. Despite these advancements, MMSE estimators still face challenges such as interference in multi-user environments, which can be mitigated by incorporating channel estimation errors into the MMSE detector [26].

One of the primary advantages of MMSE is its ability to provide accurate channel state information (CSI), which is crucial for the performance of Massive MIMO systems, especially in uplink scenarios where interference between user equipments (UEs) can be significant [24], [27]. MMSE estimators are effective in mitigating interference and improving spectral and energy efficiency, particularly when dealing with pilot contamination [28]. Additionally, MMSE can be adapted to various system configurations, such as those involving spatially correlated Rician fading channels, where it shows improved normalized mean square error (NMSE) as the Rician K-factor decreases [21].

However, MMSE channel estimation also has notable disadvantages, including its computational complexity, which can be a significant challenge in systems with large numbers of antennas or when one-bit quantization is used at the receiver [29]. To address this, techniques such as polynomial expansion have been proposed to reduce complexity while maintaining estimation performance [30]. Furthermore, MMSE requires accurate estimates of channel covariance matrices, which can be difficult to obtain, especially in low-rank scenarios [24], [27]. Despite these challenges, MMSE remains a popular choice due to its optimality in estimation theory and its ability to adapt to different channel conditions and system requirements [29], [31].

Both the Minimum Mean Square Error (MMSE) and Least Squares (LS) methods are widely used for channel estimation, each with distinct advantages and limitations. The MMSE estimator is known for its optimality in minimizing the mean square error, making it highly effective in scenarios with high interference and noise, as it requires knowledge of the channel covariance matrix to mitigate interference between user equipments (UEs) in neighboring cells [24], [27]. This method is particularly beneficial in spatially correlated Rician fading channels, where it achieves lower normalized mean square error (NMSE) as the Rician K-factor decreases [21]. However, MMSE's computational complexity and requirement for channel covariance information can be a drawback [29]. On the other hand, the LS method is simpler and does not require prior knowledge of the channel statistics, making it easier to implement [32]. It is effective in scenarios with high signal-to-noise ratio (SNR) and low interference, as demonstrated in MIMO-OFDM systems [32]. Despite its simplicity, LS can suffer from higher estimation errors compared to MMSE, especially in correlated channels [33]. In 5G systems, a modified entropy-based LS (MELS) has been proposed to enhance LS performance, outperforming both LS and MMSE at high SNR values [34].

While MMSE remains a robust method for channel estimation in Massive MIMO systems, ongoing research continues to refine its efficiency and accuracy, addressing computational and interference challenges [34], [35].

C. Compressed Sensing (CS)

Compressed Sensing (CS) leverages the sparse nature of wireless channels to reduce the number of required pilot symbols. CS is based on the principle that many signals, including wireless channels, can be represented as sparse vectors in a suitable basis. This means that only a small number of coefficients are non-zero. CS algorithms exploit this sparsity to recover the channel coefficients from a smaller number of measurements than would be required by traditional methods. This technique is particularly effective for Massive MIMO channels, especially in millimeter-wave (mmWave) communications. CS reconstructs sparse channels from fewer measurements, thereby reducing the pilot overhead.

In frequency division duplex (FDD) systems, the pilot overhead is particularly burdensome, and CS offers a solution by leveraging the sparsity of the channel. For instance, structured compressive sensing (SCS) schemes reduce pilot overhead by exploiting spatio-temporal common sparsity in delay-domain MIMO channels, using non-orthogonal pilots and adaptive algorithms like the adaptive structured subspace pursuit (ASSP) to enhance estimation accuracy [36] [37].

In MIMO-OTFS systems, radar sensing information is utilized to aid channel estimation by identifying strong angle-delay-Doppler taps, transforming the problem into a sparse recovery task [38]. Deep learning approaches, such as the two-step orthogonal matching pursuit (OMP) method, integrate CS with neural networks to improve channel state information (CSI) estimation in mmWave systems, even in low SNR conditions [39] [40]. Additionally, algorithms like the zebra optimization-based CoSaMP enhance estimation accuracy by optimizing the atomic matching process [41]. The generalized block adaptive matching pursuit (gBAMP) algorithm further refines channel estimation by optimizing index sets and using adaptive iterative stop conditions [42]. These methods collectively demonstrate that CS-based techniques can significantly reduce pilot and feedback overhead while maintaining high estimation accuracy, thereby enhancing the spectral and energy efficiency of massive MIMO systems [43] [44].

Compared to LS and MMSE, CS offers several advantages. First, it can significantly reduce the pilot overhead, which is especially important in scenarios with limited resources. Second, CS can provide accurate channel estimates even with a small number of measurements. However, CS also has some disadvantages. Its performance may degrade in non-sparse environments, and it can be computationally expensive for sparse recovery.

There are three types of CS methods stated in Table 1.

TABLE I
COMPRESSED SENSING METHODS

Method	Pros	Cons
Orthogonal Matching Pursuit (OMP)	Simple and computationally efficient, making it suitable for real-time applications. It is a greedy algorithm that iteratively selects the atom that has the highest correlation with the residual. This simplicity can be advantageous in scenarios where computational resources are limited. However, OMP can get stuck in local minima, especially when the signal is highly correlated. This can lead to suboptimal performance in some cases.	Can get stuck in local minima, especially when the signal is highly correlated. This can lead to suboptimal performance in some cases.
Basis Pursuit (BP)	Formulates the channel recovery problem as a convex optimization problem, which guarantees a global optimal solution. This makes BP more robust to noise and can provide better performance than OMP in some cases. However, BP can be computationally expensive, especially for large-scale problems.	Can be computationally expensive, especially for large-scale problems.
Compressive Sampling Matching Pursuit (CoSaMP)	Combines the strengths of OMP and BP. It is more robust to noise than OMP and can provide better performance than BP in some cases. However, CoSaMP is more complex than OMP and can be computationally expensive.	More complex than OMP and can be computationally expensive.

D. Kalman Filtering

Kalman filtering is a recursive estimation technique that updates the channel state based on prior knowledge and new measurements. It is well-suited for time-varying channels and is often used in scenarios involving high user mobility. Kalman filtering models the channel as a dynamic system with a state vector that evolves over time. The state vector contains the channel coefficients and their derivatives. Kalman filtering uses a prediction step to forecast the channel state based on the previous state and a measurement update step to correct the prediction based on new measurements.

Kalman filtering methods in channel estimation for massive MIMO systems are pivotal due to their ability to dynamically track and predict channel state information (CSI) in time-varying environments. The Multi-Stage Kalman Filter (MSKF) is a notable approach that leverages a reduced delay-line equalizer and Krylov-space based techniques to achieve fast convergence and reduced channel tracking errors, making it suitable for large-scale MMIMO systems [45]. The Vector Kalman Filter (VKF) is another method that utilizes autoregressive (AR) parameters from spatial channel models (SCM) to predict channels, offering a balance between computational complexity and prediction accuracy compared to machine learning-based methods [46] [47] [48]. In time-varying MIMO-OFDM systems, Kalman filters are used to track regularized zero-forcing (RZF) precoding coefficients, significantly reducing computational complexity by avoiding pseudo-inverse calculations [49].

Adaptive Kalman filters are advantageous in handling channel aging and varying user mobility, providing effective channel coefficient predictions for precoder construction [50]. Additionally, Kalman filters can estimate CSI based on received data without relying heavily on channel statistics, thus reducing the need for frequent pilot transmissions [51]. The use of Kalman filters in TDD massive MIMO systems allows for longer intervals between pilot transmissions, enhancing spectral efficiency by accommodating high Doppler spreads [52]. In STBC MIMO-OFDM systems, Kalman filters improve channel estimation accuracy by utilizing orthogonal pilot sequences and dynamic tracking properties, which are crucial in dynamic multipath environments [53]. Finally, the Kalman filter's ability to adaptively track time-domain changes in channels is enhanced by leveraging space-time reciprocity in antenna arrays, thus improving estimation accuracy in MIMO systems [54].

Compared to LS, MMSE, and CS, Kalman filtering offers several advantages. First, it is well-suited for time-varying channels and can provide accurate channel estimates even in dynamic environments. Second, Kalman filtering can be implemented in a recursive manner, which is efficient for real-time applications. However, Kalman filtering also has some disadvantages. It requires accurate initial state information and can be computationally intensive for large-scale systems.

III. MACHINE LEARNING-BASED CHANNEL ESTIMATION METHODS

Machine learning (ML) methods offer a promising alternative to traditional channel estimation techniques. By leveraging data-driven models, ML can learn complex channel characteristics from historical data, making it well-suited for Massive MIMO systems with their scale and dynamic nature.

Machine learning-based methods for channel estimation in Massive MIMO systems have emerged as powerful tools to address the challenges posed by the complexity and dynamic nature of wireless communication environments. Deep learning models, such as deep belief networks (DBNs) and convolutional neural networks (CNNs), have been effectively utilized to enhance channel estimation accuracy by learning spatial structures and channel statistics, as demonstrated by the DBN-BES technique, which achieves low root mean square error (RMSE) even in low signal-to-noise ratio (SNR) conditions [55].

In high-mobility scenarios, deep learning frameworks like the one proposed for MIMO-OTFS systems leverage CNNs to transform frequency-selective fading channels into quasi-time-invariant channels, significantly improving bit error rate (BER) and normalized mean squared error (NMSE) while reducing computational complexity by 80% [56]. Additionally, CNN-based models have been shown to outperform traditional methods like least squares (LS) and minimum mean square error (MMSE) in low SNR regimes, providing flexibility across various channel conditions without requiring prior statistical knowledge [15]. Other innovative approaches include the use of graph neural networks (GNNs), which incorporate system topology to improve generalization across different antenna configurations [57].

Furthermore, learning-based methods employing non-orthogonal pilots in grant-free multiple access scenarios have demonstrated promising performance in achieving low bit error rates in Massive MIMO systems [58]. Techniques such as the Spatial-Frequency UNet++ exploit spatial and frequency associations to enhance channel estimation accuracy [59].

These advancements highlight the potential of machine learning to not only improve estimation accuracy but also reduce computational complexity, making them suitable for real-world applications in 5G and beyond [60], [16].

A. Deep neural networks (DNNs)

Deep neural networks (DNNs) have been successfully applied to channel estimation tasks. DNNs can approximate the mapping between received pilot signals and the channel response, capturing non-linearities and complex relationships in wireless channels. This enables them to outperform traditional methods, especially in dynamic environments. However, DNNs require large training datasets and can be computationally expensive to train.

Implementing channel estimation in Massive MIMO systems using DNNs involves several innovative approaches that leverage the capabilities of deep learning to enhance performance and efficiency. One such method is a two-stage estimation process that combines pilot-aided and data-aided channel estimation. In the first stage, a two-layer neural network (TNN) and a deep neural network (DNN) are used to jointly design the pilot and the channel estimator, optimizing the pilot length relative to the number of transmit antennas. This is crucial because traditional methods assume the pilot length is equal to or larger than the number of antennas, which is not always feasible in Massive MIMO systems due to resource constraints [61]. The second stage involves further refining the channel estimation accuracy through iterative processes using another DNN, which minimizes the mean square error (MSE) of channel estimation. This iterative approach is shown to converge quickly, typically within five iterations, making it practical for real-time applications [61]. Additionally, deep learning-based methods can be categorized into data-driven and model-driven approaches. Data-driven methods use DNNs to directly map received signals to channel parameters, while model-driven methods, such as those using sparse Bayesian learning (SBL), unfold traditional algorithms into DNNs to capture complex channel sparsity structures effectively [62].

Another approach involves using a Channel State Information Network combined with a gated recurrent unit (CsiNet-GRU) to enhance the recovery quality and balance the trade-off between compression ratio and complexity in Massive MIMO systems. This method also employs dropout techniques to reduce overfitting during the learning process, resulting in significant performance improvements over existing techniques [63]. Furthermore, the use of reinforcement learning (RL) in semi-data-aided channel estimation can reduce communication latency by selecting reliable detected symbol vectors, thus optimizing the channel estimation process in Massive MIMO systems [64]. These methods collectively demonstrate the potential of DNNs to address the challenges of channel estimation in Massive MIMO systems, offering solutions that improve accuracy, reduce latency, and manage computational complexity effectively.

B. Convolutional neural networks (CNNs)

Convolutional neural networks (CNNs) are particularly effective for extracting spatial features from data. In Massive MIMO, CNNs can process channel state information (CSI) and predict the channel based on the spatial correlation between antennas. This approach offers improvements in accuracy and robustness, especially in multi-user MIMO systems. However, CNNs require careful tuning of the network architecture and can be computationally expensive during inference.

CNNs have been increasingly utilized for channel estimation in wireless communication systems due to their ability to handle complex, non-linear problems and their robustness to imperfect channel state information (CSI). CNNs are particularly effective in Massive MIMO systems, where they can refine coarse least squares (LS) estimations by exploiting channel correlations in both frequency and time domains, leading to improved performance and reduced overhead [62].

The CNN-based approach is advantageous because it can process imperfect CSI with strong robustness, which is crucial in practical scenarios where perfect CSI is often unattainable [65]. The architecture typically involves convolutional layers

that extract features from the input data, followed by fully connected layers that convert these features into the desired output dimensions, such as combiner weights or refined channel estimates [66]. This structure allows CNNs to learn the statistics of the channel model and acquire sparsity features in the angle domain, which are essential for accurate channel estimation [66]. Moreover, CNNs have been shown to significantly decrease computational complexity compared to traditional algorithms, making them a practical choice for real-time applications [65].

In some implementations, CNNs are combined with other neural network architectures, such as long short-term memory (LSTM) networks, to enhance their ability to handle fast time-varying channels and further improve estimation accuracy [66]. The use of CNNs in channel estimation is part of a broader trend of applying deep learning techniques to various physical layer problems in wireless communications, demonstrating their versatility and effectiveness in addressing the challenges posed by Massive MIMO systems [62].

CNNs offer a promising solution for channel estimation by providing a balance between performance, robustness, and computational efficiency, which are critical for the next generation of wireless communication systems. Moreover, the integration of attention mechanisms within these architectures can further refine the model's focus on relevant features, leading to even greater improvements in performance and adaptability in dynamic environments.

C. *Recurrent neural networks (RNNs)*

Recurrent neural networks (RNNs) and their variant LSTM are designed to handle sequential data and capture time dependencies in channel matrices, making them ideal for time-varying channel estimation in mobile environments. LSTMs can capture long-term dependencies in time-series data, which is particularly useful for tracking slow-fading channels in Massive MIMO systems. However, RNNs and LSTMs can be computationally expensive and require large-scale training.

The RNN-based approach leverages the inherent time and frequency correlations in wireless channels, which allows for more accurate channel estimation without the need for extensive channel-state-information (CSI) feedback or pilot assignment [67]. The LSTM networks are trained to predict the current channel matrix using a series of past channel matrices, optimizing the number of time steps considered to balance between capturing time correlation and avoiding excessive randomness [67]. This method is particularly beneficial in scenarios with long time coherence and channel hardening, where it outperforms traditional blind detection and least squares (LS) estimators [67].

Additionally, RNNs, including LSTM and Gated Recurrent Unit (GRU) architectures, have been proposed for doubly-selective channel estimation, addressing challenges posed by multi-path propagation and Doppler effects in dynamic environments. These RNN-based schemes demonstrate superior performance in terms of bit error rate and throughput across various mobility scenarios and modulation orders, while also reducing computational complexity and execution time compared to conventional methods [68].

Furthermore, the integration of deep learning techniques, such as combining LSTM with deep neural networks (DNNs), enhances the generalization capabilities of channel estimation models, allowing them to adapt more efficiently to non-stationary environments and reduce pilot overhead [69], [70]. These advancements highlight the potential of RNNs in improving the accuracy and efficiency of channel estimation in modern wireless communication systems, making them a promising tool for future developments in this field.

D. *Autoencoders*

Autoencoders are used for dimensionality reduction in high-dimensional systems like Massive MIMO. They compress the channel state information into a lower-dimensional space while maintaining key features for accurate channel reconstruction. This can reduce the computational complexity of channel estimation and improve efficiency.

Autoencoders, particularly variational autoencoders (VAEs) and convolutional neural network (CNN) autoencoders, have been explored for channel estimation in various contexts. The use of VAEs in channel estimation is highlighted in the context of underdetermined systems, where they are employed to parameterize an approximation to the mean squared error (MSE)-optimal estimator. This approach is advantageous as it does not require perfect channel state information (CSI) during the offline training phase, which is a significant improvement over other deep learning-based methods that typically demand such data [71].

Additionally, CNN autoencoders have been utilized in differential encoding networks for CSI estimation, where they have shown to outperform traditional compressed sensing approaches. These autoencoders are used to encode and feedback estimation errors, leveraging their ability to compress error terms effectively. This method combines unrolled optimization networks with autoencoders, demonstrating superior performance compared to previous autoencoder-based approaches [72].

Furthermore, the integration of autoencoders in channel estimation frameworks allows for the exploitation of sparsity in channel realizations, particularly in the angular domain, which enhances the network's ability to handle interference and improve estimation accuracy [73]. These applications underscore the versatility and effectiveness of autoencoders in addressing the challenges of channel estimation, such as pilot contamination and feedback compression, in modern wireless communication systems. The use of autoencoders, therefore, represents a promising direction for improving the efficiency and accuracy of channel estimation processes in various MIMO system configurations. However, autoencoders require large datasets and their performance is limited by the network design.

E. *Reinforcement learning (RL)*

Reinforcement learning (RL) is another promising approach to channel estimation. RL agents can learn policies that optimize the estimation process over time by interacting with the environment. This allows RL to adapt to different channel

conditions without explicit training data, making it robust to changing environments.

The use of RL for channel estimation is explored in the context of optimizing the selection of detected symbols for a semi-data-aided channel estimator. This approach involves formulating an optimization problem that adaptively selects these symbols, which is then solved using an efficient RL algorithm. The RL-based channel estimator demonstrates superior performance in terms of normalized mean square error (NMSE) and block error rate (BLER) compared to conventional pilot-aided methods by leveraging detected symbol vectors as additional pilot signals [64]. This method is particularly advantageous as it can be universally applied to any soft-output data detection method that computes log-likelihood ratios (LLRs) of transmitted data bits, thus enhancing its applicability across various data detection scenarios [64].

The RL algorithm's ability to utilize a priori probabilities (APPs) obtained from maximum a posteriori (MAP) data detection methods further underscores its versatility and effectiveness in channel estimation tasks. This approach not only improves the accuracy of channel estimation but also reduces the reliance on traditional pilot signals, thereby optimizing the overall communication system performance. The integration of RL in channel estimation represents a significant advancement in wireless communication, offering a robust framework for enhancing signal processing capabilities in complex and dynamic environments. However, RL requires exploration and may converge slowly.

TABLE II
ML-BASED METHODS

Method	Advantages	Disadvantages
DNNs	- Highly flexible models capable of learning complex non-linear relationships between received signals and channel coefficients. - Can capture both spatial and temporal correlations in the channel, making them suitable for a wide range of wireless environments. - Relatively easy to train and deploy compared to other ML methods.	- Require large training datasets, which can be challenging to obtain and label. - Computationally expensive, especially for large-scale models, requiring significant computational resources. - Sensitive to overfitting, which can lead to poor generalization performance.
CNNs	- Efficiently extract spatial features from CSI, which is crucial for channel estimation in Massive MIMO systems. - Computationally efficient compared to DNNs, making them suitable for real-time applications. - Relatively easy to train and deploy, especially when using pre-trained models.	- May struggle to capture temporal dependencies in time-varying channels, limiting their effectiveness in certain environments. - Require careful tuning of the network architecture and hyperparameters to achieve optimal performance.
RNNs/LSTMs	- Handle sequential data effectively, making them well-suited for time-varying channels in mobile environments. - Can capture long-term dependencies in the channel, which is important for predicting future channel states. - Relatively easy to train and deploy compared to other ML methods.	- Computationally expensive, especially for large-scale models and long sequences. - Can suffer from the vanishing gradient problem, which can make training difficult. - Sensitive to noise and outliers in the data.
Autoencoders	- Reduce the dimensionality of CSI, which can improve computational efficiency and reduce storage requirements. - Can be trained unsupervised, which can be advantageous when labeled data is limited. - Can capture important features of the channel while reducing noise and redundancy.	- May not capture all important features of the channel, leading to suboptimal performance. - Sensitive to noise and outliers in the data. - Difficult to train, especially for complex architectures.
RL	- Adapts to changing channel conditions without requiring explicit training data. - Can optimize channel estimation performance over time, improving accuracy and efficiency. - Can be used in environments with limited or no prior knowledge of the channel.	- Computationally expensive, especially for complex environments and large state spaces. - Can converge slowly, especially in challenging environments. - Difficult to tune and evaluate RL agents.

In addition to these methods, hybrid approaches that combine ML with traditional techniques have also been explored. For example, hybrid methods can use ML to learn the parameters of a traditional channel model or to improve the accuracy of traditional estimation algorithms.

One of the key challenges in ML-based channel estimation is the need for large training datasets. These datasets must be representative of the diverse channel conditions that the system will encounter in practice. Collecting and labeling such datasets can be time-consuming and expensive. To address this challenge, researchers have explored techniques such as data augmentation, transfer learning, and generative models.

Another challenge is the computational cost associated with training and deploying ML models. DNNs, CNNs, and RNNs can be computationally intensive, especially for large-scale models. To address this challenge, researchers have explored techniques such as model compression, quantization, and hardware acceleration.

Despite these challenges, ML-based channel estimation methods offer a promising avenue for addressing the challenges of channel estimation in Massive MIMO systems. By leveraging the power of data-driven models, ML can learn complex channel characteristics and provide accurate channel estimates in dynamic and challenging environments. However, further research is needed to address the computational and data requirements of ML-based methods and to explore new hybrid approaches that combine the strengths of ML and traditional techniques.

IV. DISCUSSION

Machine learning (ML) methods offer a promising alternative to traditional channel estimation techniques. By leveraging data-driven models, ML can learn complex channel characteristics from historical data, making it well-suited for Massive MIMO systems with their scale and dynamic nature.

TABLE III
COMPARISON OF ML AND TRADITIONAL BASED METHODS

Method	Complexity	Accuracy	Adaptability	Pilot Overhead	Data Requirements	Hardware Requirements
LS	$O(n^2)$	High MSE at low SNR	Static channels	Requires many pilot symbols	No prior channel statistics required	Can run on CPUs
MMSE	$O(n^3)$	Low MSE, especially at low SNR	Requires channel statistics	Requires many pilot symbols, but fewer than LS	Requires channel covariance matrix	May require GPUs for large-scale system
Compressed Sensing	Sparse recovery algorithms can be computationally intensive	Good for sparse channels, degrades in non-sparse environment	Works best in sparse environment	Significantly reduces pilot overhead	Requires sparse channel representation	May require GPUs for real-time applications
Kalman Filtering	Recursive updates, $O(n^2)$	Good for time-varying channels	Excellent for dynamic environments	Requires periodic pilot updates	Requires initial state information	Can run on CPUs, but GPUs may speed up processing
DNN	Training: $O(n^4)$ Inference: $O(n^3)$	Captures non-linearities, low MSE	Adapts well to dynamic environment	Learns channel structure, hence reducing pilot overhead	Requires large labeled datasets	Requires GPUs for training and inference
CNN	Training: $O(n^4)$ Inference: $O(n^3)$	captures spatial correlations, low MSE	Adapts well to multi-user MIMO	Learns spatial features, reduces pilot overhead	Requires large labeled datasets	Requires GPUs for training and inference
RNN/LSTM	Training: $O(n^4)$ Inference: $O(n^3)$	Captures temporal dependencies, low MSE	Excellent for time-varying and mobile environments	Learns temporal features, reduces pilot overhead	Requires large labeled datasets with temporal sequences	Requires GPUs for training and inference
Autoencoder	Training: $O(n^4)$ Inference: $O(n^3)$	Good for dimensionality reduction, moderate MSE	Works well for high-dimensional systems	Compresses CSI, reduces pilot overhead	Requires large datasets for training	Requires GPUs for training and inference
Reinforcement Learning	Exploration and policy optimization can be computationally intensive	Adapts to changing environments, low MSE	Excellent for dynamic and non-stationary environments	Reduces reliance on pilot signals	Requires interaction with the environment for training	Requires GPUs for training and inference

A. Accuracy

ML-based methods generally outperform traditional methods, especially in complex and dynamic environments. They can capture intricate relationships in the channel that traditional methods may overlook, such as non-linear dependencies, spatial correlations, and temporal variations. This is particularly beneficial in scenarios with rapidly changing channel conditions or when the channel is highly correlated. For example, in Massive MIMO systems with a large number of antennas, ML methods can effectively exploit the spatial correlation between antennas to improve channel estimation accuracy.

B. Complexity

While ML methods often require more computational resources due to their complexity and the need for large training datasets, their improved accuracy and adaptability can justify the increased computational cost. In many cases, the benefits of ML-based methods outweigh the additional computational overhead, especially in applications where high accuracy and adaptability are critical. For instance, in autonomous vehicles or critical infrastructure, accurate channel estimation is essential for reliable communication, and the increased computational cost of ML methods may be acceptable in exchange for improved performance.

C. Pilot Overhead

ML methods can significantly reduce pilot overhead by learning the structure of the channel more efficiently. This is particularly beneficial in scenarios with limited resources or bandwidth constraints, such as in mobile communication systems or IoT networks. By reducing the number of pilot symbols required for channel estimation, ML methods can improve spectral efficiency and increase data throughput. For example, in IoT networks where devices have limited power and bandwidth, ML-based channel estimation can help reduce the overhead associated with transmitting pilot symbols, enabling more efficient communication.

D. Adaptability

ML methods like LSTMs and RL excel in environments with high mobility or time-varying channels. They can adapt to changing conditions and provide accurate channel estimates in real-time, which is essential for applications like mobile communication, vehicular networks, and wireless sensor networks. For instance, in mobile communication systems where users are constantly moving and the channel conditions are changing rapidly, ML-based methods can continuously learn and adapt to the channel, ensuring reliable communication even in challenging environments.

E. Additional Considerations

To provide a clearer and more structured overview of the additional considerations in machine learning-based channel estimation, Table IV summarizes key aspects such as data quality, model selection, interpretability, privacy, hardware acceleration, and hybrid approaches.

TABLE IV
ADDITIONAL CONSIDERATIONS

Aspects	Descriptions
Data Quality and Quantity Considerations	
Data Collection	Collect diverse datasets covering various environments (indoor, outdoor, urban, rural), frequency bands, and propagation conditions.
Data Cleaning	Remove outliers, inconsistencies, and errors using techniques like outlier detection, imputation, and normalization.
Data Augmentation	Generate additional training data through noise injection, rotation, and scaling to improve model robustness.
Data Labeling	Accurately label channel estimates in training data to provide correct supervision. This task may require domain expertise.
Model Selection and Hyperparameter Tuning	
Model Architecture	Choose a model suitable for the task (e.g., DNNs for non-linear relationships, CNNs for spatial features).
Hyperparameter Tuning	Experiment with learning rate, batch size, number of layers, and activation functions using techniques like grid search or Bayesian optimization.
Interpretability and Explainability	
Visualize Decisions	Use feature importance plots or decision trees to understand how the model makes predictions.
Identify Biases	Detect and mitigate biases arising from training data or model architecture.
Explain Behavior	Provide human-understandable explanations for model decisions to build trust and improve transparency.
Privacy and Security	
Data Encryption	Protect data from unauthorized access during transmission and storage.
Data Anonymization	Remove or disguise personal information to protect user privacy.
Model Security	Protect models from adversarial examples (misleading inputs) and model theft (stealing parameters).
Hardware Acceleration	
GPU-Based Training	Use GPUs to accelerate training, especially for large-scale models requiring parallel computations.

Aspects	Descriptions
TPU-Based Inference	Use TPUs to accelerate inference, making ML models more suitable for real-time applications.
Hybrid Approaches	
ML-Enhanced Traditional Methods	Use ML to learn parameters of traditional models or improve their accuracy (e.g., predicting channel coefficients).
Hybrid Architectures	Combine traditional and ML components (e.g., traditional estimator for initial estimation, ML for fine-tuning).

V. FUTURE RESEARCH DIRECTIONS

Channel estimation in Massive MIMO systems is a rapidly evolving field with numerous opportunities for future research. One promising avenue is the development of hybrid methods that combine the strengths of traditional and machine learning (ML) techniques.

1) **Hybrid Methods**

Hybrid methods can leverage the complementary advantages of traditional and ML-based approaches. For example, traditional methods can be used for initial channel estimation, providing a baseline estimate that can be refined by ML models. This can reduce the computational cost of pure ML models while improving their performance in dynamic environments. Additionally, hybrid methods can incorporate domain knowledge into the ML models, enhancing their interpretability and robustness.

One potential hybrid approach is to use a traditional channel estimator to provide an initial estimate of the channel, and then use an ML model to refine the estimate based on additional information, such as the received signal or the channel statistics. This can help to improve the accuracy of the channel estimate, especially in challenging environments. Another approach is to use ML models to learn the parameters of a traditional channel model, making it more adaptable to different channel conditions.

2) **Federated Learning**

Federated learning is another promising area of research for channel estimation in Massive MIMO systems. This technique allows multiple devices to train a shared ML model without sharing their raw data, preserving privacy and reducing communication overhead. Federated learning can be particularly useful in distributed Massive MIMO systems where channel data is collected from a large number of devices.

By using federated learning, channel estimation can be performed in a decentralized manner, reducing the reliance on a central server and improving privacy. Additionally, federated learning can enable the training of ML models on large-scale datasets that would be difficult or impossible to collect and process centrally.

3) **Real-Time ML Inference**

While training ML models can be computationally expensive, future work could focus on optimizing inference time to make real-time deployment feasible in Massive MIMO systems. Techniques like model pruning and quantization can help reduce the model size and speed up the estimation process. Additionally, hardware acceleration using specialized hardware like GPUs or TPUs can further improve the inference performance.

By optimizing inference time, ML-based channel estimators can be deployed in real-time applications, such as mobile communication systems and autonomous vehicles. This will enable more accurate and responsive channel estimation, leading to improved system performance.

4) **Cross-Layer Optimization**

There is growing interest in cross-layer optimization, where channel estimation is integrated with higher-layer functions like resource allocation and power control. By jointly optimizing these processes using ML models, system performance could be significantly enhanced. For example, ML models could be used to predict the channel conditions and allocate resources accordingly, or to optimize power control to maximize data throughput while minimizing interference.

Cross-layer optimization can help to achieve more efficient and reliable wireless communication by taking into account the interactions between different layers of the system. By jointly optimizing these layers, it is possible to achieve better overall system performance than by optimizing each layer in isolation.

5) **Explainable Machine Learning**

As ML models become more complex, their interpretability decreases. Future research should focus on developing explainable ML methods for channel estimation to increase transparency and trust in ML-based wireless systems. Explainable ML techniques can help to understand how the model makes decisions, identify potential biases, and improve the model's reliability.

Explainable ML is particularly important in critical applications where it is essential to understand how the model works and why it makes certain decisions. By making ML models more explainable, we can increase trust in their predictions and ensure that they are not biased or unfair.

In conclusion, channel estimation in Massive MIMO systems is a rapidly evolving field with numerous opportunities for future research. By developing hybrid methods, leveraging federated learning, optimizing real-time inference, exploring cross-layer optimization, and improving the explainability of ML models, we can continue to advance the state of the art in this critical area of wireless communication.

V. CONCLUSION

Massive MIMO systems are pivotal for modern wireless communication, offering significant improvements in capacity, spectral efficiency, and energy efficiency. However, accurate channel estimation remains a critical challenge, especially in dynamic and complex environments. Traditional methods like Least Squares (LS) and Minimum Mean Square Error (MMSE) are widely used due to their simplicity and computational efficiency, but they struggle in low SNR and high-mobility scenarios. In contrast, machine learning (ML)-based methods, such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), have demonstrated superior performance by capturing complex spatial and temporal correlations in the channel. These methods reduce pilot overhead, improve accuracy, and adapt well to

dynamic environments, though they require large datasets and significant computational resources.

Future research should focus on hybrid approaches that combine the strengths of traditional and ML-based methods, leveraging the simplicity of traditional techniques for initial estimates and the adaptability of ML for refinement. Additionally, advancements in federated learning, real-time ML inference, and cross-layer optimization can further enhance the efficiency and robustness of channel estimation in Massive MIMO systems. By addressing challenges such as data requirements, computational complexity, and model interpretability, ML-based methods hold great promise for advancing wireless communication in the era of 5G and beyond.

REFERENCES

[1] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, Nov. 2010, DOI: 10.1109/TWC.2010.092810.091092.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, Feb. 2014, DOI: 10.1109/MCOM.2014.6736761.

[3] R. Hidayat, A. F. Isnawati, and B. Setiyanto, "Channel estimation in MIMO-OFDM spatial multiplexing using Least Square method," in *International Symposium on Intelligent Signal Processing and Communication Systems*, Dec. 2011. DOI: 10.1109/ISPACS.2011.6146157.

[4] Z. Luo and D. Huang, "General MMSE Channel Estimation for MIMO-OFDM Systems," in *Vehicular Technology Conference*, Oct. 2008. DOI: 10.1109/VETECF.2008.151.

[5] N. Dubey and A. Pandit, "A Comprehensive Review on Channel Estimation in OFDM System," Mar. 2019, DOI: 10.24113/IJOSCIENCE.V5I3.201.

[6] Z. Fang and J. Shi, "Least Square Channel Estimation for Two-Way Relay MIMO OFDM Systems," *Etri Journal*, Oct. 2011, DOI: 10.4218/ETRIJ.11.0210.0424.

[7] B. Zhang, "A LS channel estimation algorithm based on adaptive filtering for OFDM system," *Journal of Xi'an University of Posts and Telecommunications*, Jan. 2011.

[8] G. Zhu, J. Dai, C. Chang, and W. Xu, "Robust uplink channel estimation for massive MIMO systems with general array geometry," *Digital signal processing*, Oct. 2022, DOI: 10.1016/j.dsp.2022.103793.

[9] S. Li, J. Li, H. Dong, and Z. Li, "A Massive MIMO Channel Estimation Algorithm Design Combined the SVD Method with LS Signal Detection," Aug. 2022, DOI: 10.1109/IWS55252.2022.9977532.

[10] "A Massive MIMO Channel Estimation Algorithm Design Combined the SVD Method with LS Signal Detection," *2022 IEEE MTT-S International Wireless Symposium (IWS)*, Aug. 2022, DOI: 10.1109/iws55252.2022.9977532.

[11] S. S. Devi, B. S. Kumar, S. Narayanan, and G. Vimal, "Channel Estimation of OFDM – MIMO," Dec. 2022, DOI: 10.1109/ICMNWC56175.2022.10032004.

[12] "Channel Estimation of OFDM – MIMO," *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Dec. 2022, DOI: 10.1109/icmnwc56175.2022.10032004.

[13] A. Hasan, S. M. A. Motakabber, F. Anwar, M. H. Habaebi, and M. I. Ibrahimy, "A Computationally Efficient Least Squares Channel Estimation Method for MIMO-OFDM Systems," *International Conference on Computer and Communication Engineering*, Jun. 2021, DOI: 10.1109/ICCCE50029.2021.9467142.

[14] A. Riadi, M. Boulouird, and M. M. Hassani, "Least Squares Channel Estimation of an OFDM Massive MIMO System for 5G Wireless Communications," Dec. 2018, DOI: 10.1007/978-3-030-21009-0_43.

[15] L. Carro-Calvo, A. de la Fuente, A. Melgar, and E. Morgado, "Massive MIMO Channel Estimation With Convolutional Neural Network Structures," *IEEE Transactions on Cognitive Communications and Networking*, Jan. 2024, DOI: 10.1109/tccn.2024.3435478.

[16] R. Thakur and V. N. Saxena, "A Survey on Learning-Based Channel Estimation Methods Used for 5G Massive Mimo System," Jul. 2023, DOI: 10.1109/icccnt56998.2023.10308216.

[17] G. Zhu, Y. Xia, and S. Li, "VBI-based Uplink Channel Estimation for Massive MIMO Systems," *IEEE International Conference on Electronic Information and Communication Technology*, Aug. 2022, DOI: 10.1109/ICEICT55736.2022.9909288.

[18] B. Sissokho, J.-P. Cances, and A. D. Kora, "Uplink Rate Based on Massive MIMO Channel Estimation Approach," *International Conference Frontiers Signal Processing*, Sep. 2018, DOI: 10.1109/ICFSP.2018.8552055.

[19] G. Sklivanitis, K. Tountas, D. A. Pados, and S. N. Batalama, "Small-Sample-Support Channel Estimation for Massive Mimo Systems," *International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2018, DOI: 10.1109/ICASSP.2018.8461599.

[20] Jyoti P. Patra, Bibhuti Bhusan Pradhan, and M. Rajendra Prasad, "An Ordered QR Decomposition based Signal Detection Technique for Uplink Massive MIMO System", *Infocommunications Journal*, Vol. XVI, No 1, March 2024, pp. 20–25., DOI: 10.36244/ICJ.2024.1.3

[21] J. F. Arellano, C. D. Altamirano, H. R. C. Mora, N. O. Garzón, and F. D. A. García, "On the Performance of MMSE Channel Estimation in Massive MIMO Systems over Spatially Correlated Rician Fading Channels," *Wireless Communications and Mobile Computing*, Apr. 2024, DOI: 10.1155/2024/5445725.

[22] B. Li *et al*., "Beyond MMSE: Rank-1 Subspace Channel Estimator for Massive MIMO Systems," Apr. 2024, DOI: 10.48550/arxiv.2404.13603.

[23] A. H. Shatti and E. A. Hussein, "Low-complex Bayesian estimator for imperfect channels in massive muti-input multi-output system," *International Journal of Electrical and Computer Engineering*, Dec. 2022, DOI: 10.11591/ijece.v12i6.pp6261-6271.

[24] "GEVD-based Low-Rank Channel Covariance Matrix Estimation and MMSE Channel Estimation for Uplink Cellular Massive MIMO Systems," *2022 30th European Signal Processing Conference (EUSIPCO)*, Aug. 2022, DOI: 10.23919/eusipco55093.2022.9909626.

[25] X. Li, Z. Song, L. Liu, X. Sha, and Y. Li, "Model-Based Structured Covariance-Aided Channel Estimation for Massive MIMO Systems," *International Conference on Speech Technology and Human-Computer Dialogue*, Nov. 2022, DOI: 10.1109/ICCT56141.2022.10072431.

[26] N. Zhao *et al*., "The MMSE MIMO Detector Under a New LMMSE Channel Estimation Error Analysis Method," Jul. 2023, DOI: 10.1109/iceict57916.2023.10245661.

[27] R. V. Rompaey and M. Moonen, "GEVD-based Low-Rank Channel Covariance Matrix Estimation and MMSE Channel Estimation for Uplink Cellular Massive MIMO Systems," arXiv: Signal Processing, Nov. 2021, DOI: 10.48550/arXiv.2111.11902

[28] E. Mukubwa and O. Sokoya, "Comparison of Improved MMSE and the Semi-Blind Channel Estimation Methods," *International Conference on Artificial Intelligence*, Aug. 2020, DOI: 10.1109/ICABCD49160.2020.9183807.

[29] M. Ding, I. Atzeni, A. Tolli, and A. L. Swindlehurst, "On the Optimal MMSE Channel Estimation for One-Bit Quantized MIMO Systems," arXiv.org, Apr. 2024, DOI: 10.48550/arxiv.2404.05536.

[30] Y. Wang and P. Fortier, "Polynomial Expansion-Based MMSE Channel Estimation for Massive MIMO-GFDM Systems," *Vehicular Technology Conference*, Nov. 2020, DOI: 10.1109/VTC2020-FALL49728.2020.9348717.

[31] J. Mirzaei, F. Sohrabi, R. S. Adve, and S. Shahbazpanahi, "MMSE-Based Channel Estimation for Hybrid Beamforming Massive MIMO with Correlated Channels," *International Conference on Acoustics, Speech, and Signal Processing*, May 2020, DOI: 10.1109/ICASSP40776.2020.9053980.

[32] A. S. Ahmed, M. M. Hamdi, M. S. Abood, A. M. Khaleel, M. Fathy, and S. H. Khaleefah, "Channel Estimation using LS and MMSE Channel Estimation Techniques for MIMO-OFDM Systems," *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Jun. 2022, DOI: 10.1109/hora55278.2022.9799887.

[33] M. H. Sadraei, M. S. Fazel, and A. M. Doost-Hoseini, "Ergodic spectral efficiency of massive MIMO with correlated Rician channel and MRC detection based on LS and MMSE channel estimation," *Iet Communications*, Oct. 2020, DOI: 10.1049/IET-COM.2019.0905.

[34] J. Kelvin, "Analytical Analysis on LS, MMSE and Modified Entropy-Based LS Channel Estimation Techniques for 5G Massive MIMO Systems," Jan. 2023, DOI: 10.1007/978-981-19-9512-5_40.

[35] Q. Liu, Y. Li, and J. Sun, "A Model-Driven Channel Estimation Method for Millimeter-Wave Massive MIMO Systems," *Sensors,* Feb. 2023, DOI: 10.3390/s23052638.

[36] Z. Gao, Y. Mei, and L. Qiao, "Compressive Sensing Sparse Channel Estimation in FDD Massive MIMO Systems," Oct. 2023, DOI: 10.1007/978-981-99-5394-3_3.

[37] "High Spectrum and Efficiency Improved Structured Compressive Sensing-Based Channel Estimation Scheme for Massive MIMO Systems," Jan. 2022, DOI: 10.1007/978-981-16-7610-9_19.

[38] S. Jiang and A. Alkhateeb, "Sensing Aided OTFS Massive MIMO Systems: Compressive Channel Estimation," May 2023, DOI: 10.1109/iccworkshops57953.2023.10283647.

[39] W. Tong, W. Xu, F. Wang, J. Shang, M. Pan, and J. Lin, "Deep Learning Compressed Sensing-Based Beamspace Channel Estimation in mmWave Massive MIMO Systems," *IEEE Wireless Communications Letters*, Sep. 2022, DOI: 10.1109/LWC.2022.3188530.

[40] "Deep Learning Compressed Sensing-Based Beamspace Channel Estimation in mmWave Massive MIMO Systems," *IEEE Wireless Communications Letters*, Sep. 2022, DOI: 10.1109/lwc.2022.3188530.

[41] S. Li and Y. Yang, "Improved compressed sensing channel estimation algorithm for MIMO-OFDM systems," May 2024, DOI: 10.1109/ccdc62350.2024.10587886.

[42] Y. Huang, Y. He, W. He, L. Shi, T. Cheng, and Y. Sui, "Channel Estimation in Massive MIMO Systems Based on Generalized Block Adaptive Matching Pursuit Algorithm," *IEEE Wireless Communications Letters,* Aug. 2020, DOI: 10.1109/LWC.2020.3013689.

[43] "A Compressive Sensing and Deep Learning-Based Time-Varying Channel Estimation for FDD Massive MIMO Systems," *IEEE Transactions on Vehicular Technology*, Aug. 2022, DOI: 10.1109/tvt.2022.3176290.

[44] N. Nouri, M. J. Azizipour, and K. Mohamed-pour, "A Compressed CSI Estimation Approach for FDD Massive MIMO Systems," *Iranian Conference on Electrical Engineering*, Aug. 2020, DOI: 10.1109/ICEE50131.2020.9260725.

[45] "Multi Stage Kalman Filter (MSKF) Based Time-Varying Sparse Channel Estimation With Fast Convergence," *IEEE open journal of signal processing*, Jan. 2022, DOI: 10.1109/ojsp.2021.3132583.

[46] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO Channel Prediction: Kalman Filtering Vs. Machine Learning," *IEEE Transactions on Communications*, Jan. 2021, DOI: 10.1109/TCOMM.2020.3027882.

[47] H. Kim, S. Kim, H. Lee, and J. Choi, "Massive MIMO Channel Prediction: Machine Learning Versus Kalman Filtering," *Global Communications Conference*, Dec. 2020, DOI: 10.1109/GCWKSHPS50303.2020.9367471.

[48] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO Channel Prediction: Kalman Filtering vs. Machine Learning," arXiv: Information Theory, Sep. 2020, DOI: 10.1109/TCOMM.2020.3027882

[49] S. Wu, X. Tao, D. Mishra, Y. Chen, and J. Xu, "Efficient Kalman Filter-Based Precoder Tracking for Time-Varying Massive MIMO-OFDM Systems," *IEEE Communications Letters*, Mar. 2020, DOI: 10.1109/LCOMM.2020.2981937.

[50] V. Arya and K. Appaiah, "Kalman Filter based Tracking for Channel Aging in Massive MIMO Systems," *International Conference on Signal Processing*, Jul. 2018, DOI: 10.1109/SPCOM.2018.8724442.

[51] A. Almamori and S. Mohan, "Estimation of channel state information for massive MIMO based on received data using Kalman filter," *IEEE Annual Computing and Communication Workshop and Conference*, Jan. 2018, DOI: 10.1109/CCWC.2018.8301698.

[52] S. Kashyap, C. Mollen, E. Björnson, and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with Kalman channel prediction," *International Conference on Acoustics, Speech, and Signal Processing*, Mar. 2017, DOI: 10.1109/ICASSP.2017.7952818.

[53] R. Tang, X. Zhou, and C. Wang, "Kalman Filter Channel Estimation in 2 × 2 and 4 × 4 STBC MIMO-OFDM Systems," *IEEE Access*, Sep. 2020, DOI: 10.1109/ACCESS.2020.3027377.

[54] R. Tang, X. Zhou and C. Wang, "Kalman Filter Channel Estimation in 2 × 2 and 4 × 4 STBC MIMO-OFDM Systems," in *IEEE Access*, vol. 8, pp. 189 089–189 105, 2020, DOI: 10.1109/ACCESS.2020.3027377.

[55] O. Pabbati and R. Joshi, "An optimized deep learning model for a highly accurate DOA and channel estimation for massive MIMO systems," *International Journal of Communication Systems*, Jul. 2024, DOI: 10.1002/dac.5902.

[56] M. Payami and S. D. Blostein, "Deep Learning-based Channel Estimation for Massive MIMO-OTFS Communication Systems," Apr. 2024, DOI: 10.1109/wts60164.2024.10536672.

[57] M. Ye, X. Liang, C. Pan, Y. Xu, M. Jiang, and C. Li, "Channel Estimation for mmWave Massive MIMO Systems Using Graph Neural Networks," Aug. 2023, DOI: 10.1109/iccc57788.2023.10233621.

[58] S. Park, Y. Kim, J. Jang, and H.-J. Yang, "Learning-Based Channel Estimation Method with Non-Orthogonal Pilots for Grant-Free Multiple Access in Massive MIMO Systems," *The Journal of Korean Institute of Communications and Information Sciences*, Nov. 2023, DOI: 10.7840/kics.2023.48.11.1464.

[59] K. Long, H. Liu, D. Qiu, and J. Yang, "A Channel Estimator of Millimeter-Wave Massive MIMO Systems Through Deep Learning," Nov. 2023, DOI: 10.1109/apcap59480.2023.10470212.

[60] J. Meng, Z. Wei, Y. Zhang, B. Li, and C. Zhao, "Machine learning based low-complexity channel state information estimation," *EURASIP Journal on Advances in Signal Processing*, Oct. 2023, DOI: 10.1186/s13634-023-00994-4.

[61] C. J. Chun, J. M. Kang, and I. M. Kim, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1228–1231, 2019. DOI: 10.1109/lwc.2019.2912378

[62] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi, "Deep Learning-Based Channel Estimation for Wideband Hybrid MmWave Massive MIMO," *IEEE Transactions on Communications*, vol. 71, no. 6, pp. 3679–3693, Jun. 2023, DOI: 10.1109/TCOMM.2023.3258484.

[63] H. M. N. Helmy, S. El Daysti, H. Shatila, and M. Aboul-Dahab, "Performance Enhancement of Massive MIMO Using Deep Learning-Based Channel Estimation," vol. 1051, no. 1, p. 012 029, Feb. 2021, DOI: 10.1088/1757-899X/1051/1/012029.

[64] T.-K. Kim, Y.-S. Jeon, J. Li, N. Tavangaran, and H. V. Poor, "Semi-Data- Aided Channel Estimation for MIMO Systems via Reinforcement Learning," *IEEE Transactions on Wireless Communications*, p. 1, Apr. 2022, DOI: 10.1109/twc.2022.3227312.

[65] X. Bao, W. Feng, J. Zheng, and J. Li, "Deep CNN and Equivalent Channel Based Hybrid Precoding for mmWave Massive MIMO Systems," *IEEE Access*, vol. 8, pp. 19 327–19 335, Jan. 2020, DOI: 10.1109/ACCESS.2020.2967402.

[66] H. Hirose, T. Ohtsuki and G. Gui, "Deep Learning-Based Channel Estimation for Massive MIMO Systems With Pilot Contamination," in *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 67–77, 2021, DOI: 10.1109/OJVT.2020.3045470.

[67] T. Faghani, A. Shojaeifard, K.-K. Wong, and A. Aghvami, "Recurrent Neural Network Channel Estimation Using Measured Massive MIMO Data" *34th IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2023*, Toronto, ON, Canada, September 5-8, 2023," Sep. 2023, DOI: 10.1109/PIMRC48278.2020.9217192

[68] A. K. Gizzini and M. Chafii, "RNN Based Channel Estimation in Doubly Selective Environments," *IEEE transactions on machine learning in communications and networking*, vol. 2, pp. 1–18, Jan. 2024, DOI: 10.1109/tmlcn.2023.3332021.

[69] J. Abouei et al., "RNN-FL-based Channel Estimation Approach in mmWave Massive MIMO Systems" *Authorea Preprints*, 2023.

[70] L. Ge, C. Shi, S. Niu, and G. Chen, "Mixed RNN-DNN based channel prediction for massive MIMO-OFDM systems," *Iet Communications*, Oct. 2023, DOI: 10.1049/cmu2.12685.

[71] M. Baur, N. Turan, B. Fesl, and W. Utschick, "Channel Estimation in Underdetermined Systems Utilizing Variational Autoencoders," Apr. 2024, DOI: 10.1109/icassp48485.2024.10447622.

[72] M. Rosario and Z. Ding, "Learning-Based MIMO Channel Estimation under Practical Pilot Sparsity and Feedback Compression", *IEEE Transactions on Wireless Communications* 22.2 (2022): 1161–1174, DOI: 10.1109/TWC.2022.3202750

[73] A. Kasibovic, B. Fesl, M. Baur, and W. Utschick, "Addressing Pilot Contamination in Channel Estimation with Variational Autoencoders", *arXiv preprint arXiv:*2409.07071, 2024, DOI: 10.48550/arXiv.2409.07071.

Amalia Eka Rakhmania received the Bachelor of Engineering degree from the University of Brawijaya, Malang, Indonesia, in 2012, Master of Science from National Central University, Taiwan, and Master of Engineering degree from the University of Brawijaya, Malang, Indonesia in 2015 all in electrical engineering, majoring in Telecommunication. She is currently a lecturer at the Department of Electrical Engineering, State Polytechnic of Malang, Indonesia. Her research interests include wireless and optical communication, interference mitigation, and the internet of things for smart cities.

Hudiono Hudiono received the Bachelor of Engineering and Master of Engineering. in electrical engineering, majoring in Telecommunication degree from the Tenth November Institute of Technology, Surabaya, Indonesia, in 1993 and 1998, respectively. He is currently a lecturer at the Department of Electrical Engineering, State Polytechnic of Malang, Indonesia. His research interests include radio communication, antenna propagation, wireless sensor network, and the internet of things for smart cities.

Umi Anis Ro'isatin received her Bachelor and Master of English Education from State University of Malang, Malang, Indonesia in 1988 and 2009, respectively. She is currently a lecturer at the Department of Mechanical Engineering, State Polytechnic of Malang, Indonesia. Her research interests include English for specific purpose, pedagogy, and English for second language.

Nurul Hidayati received the Bachelor of Engineering and Master of Engineering. in electrical engineering, majoring in Telecommunication degree from the Tenth November Institute of Technology, Surabaya, Indonesia, in 2016 and 2018, respectively. She is currently a lecturer at the Department of Electrical Engineering, State Polytechnic of Malang, Indonesia. Her research interests include internet of things, cooperative wireless network, and algorithms for wireless sensor network.

Horn Antenna Development at 80 GHz for Tank Level Probing Radar Applications

Lajos Nagy, *Member, IEEE*

***Abstract*—One important application and research area of radar technology is tank-level measurement and detection. Radar contactless level measurement is a safe solution even in extreme process conditions, such as significant overpressure, high temperatures, and the presence of corrosive vapors. The main categories of these principles are ultrasonic, and electromagnetic wave radars.**

We will now consider only radars using electromagnetic waves. The use of millimeter radio waves, which we use, is nowadays becoming more and more common also for automotive radars, human presence detection and human vital signs.

To meet electromagnetic requirements such as high gain, low spurious levels, and high bandwidth, special antennas are required.

The low sidelobe level and narrow main beam mainly reduce reflections from the side of the tank while the bandwidth determines the distance resolution of the measurement system. A further requirement is the small size and reliable manufacturability of the antenna. In the presence of corrosive vapors, antennas must be resistant to corrosion.

The article briefly summarizes the material parameter measurements required for the design of the radar, and the design of the main components. We analyzed the possible dielectric materials that can be used as random or dielectric lenses for such antennas. In the next part of the paper, we present a conical horn antenna design for the 80 GHz band and compare the parameters of an open horn antenna with those of a horn antenna with a dielectric lens. Finally, a tank-level radar designed with the Texas Instruments IWR1443 radar chip is presented.

***Index Terms*—Radar, antenna**

I. Introduction

Product quality control, production safety, and process economy can only be guaranteed can only be ensured by continuous measurements and the monitoring and intervention systems based on them. The main fields of application are the oil industry, petrochemical industry, chemical industry, food industry, pharmaceutical industry, transport, wastewater storage and treatment. Liquids, pastes, bulk solids, and liquefied gases are most stored in tanks, silos, or mobile containers.

Lajos Nagy, is with Budapest University of Technology and Economics, Budapest, Hungary, Department of Broadband Infocommunications and Electromagnetic Theory, Faculty of Electrical Engineering (E-mail: nagy.lajos@vik.bme.hu).

There are several classical and modern methods for measuring the product level in process and storage tanks. Such systems use microwave contactless radar, guided microwave radar, capacitive, magnetostrictive, and ultrasonic sensors.

Antennas for contactless microwave sensors are microstrip antenna systems, horn antennas, dielectric antennas, and slot antennas on waveguides.

Applications are in the chemical, petrochemical, pharmaceutical, water, and food industries, mobile tanks on vehicles and ships, and natural reservoirs such as seas, dams, lakes, and oceans. Typical tank heights for these applications are in the range of 0.5 m to 37 m.

In practical applications, two main measurement tasks can be distinguished:

- continuous level measurement, i.e., level indication,
- level detection, i.e., detection of an alarm limit to prevent overfilling.

Many level measurement devices are mounted on top of the tank and measure primarily the distance between their mounting position and the product's surface (Fig. 1).

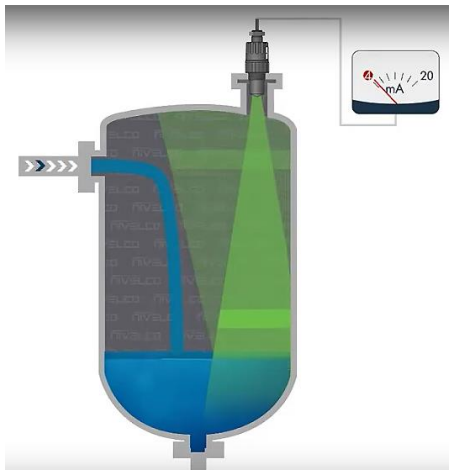


Fig. 1. Tank with liquid and non-contact sensors on the top of the tank

The sensor is placed in an opening on the top surface of the tank. As shown in the figure, the position of the sensor must be shifted laterally due to the fluid flowing through the inlet. This in turn causes reflections and consequently multiple reflections and makes level measurement difficult. (Fig. 1.)

DOI: 10.36244/ICJ.2025.1.4

Fig. 2. Horn antenna with house contains signal processing and communications circuits.

A typical horn antenna design is shown in Figure 2. The diameter of the antenna must be adapted to the size of the tank's opening to allow installation of the antenna.

For level measurement, a significant number of different principles measurement techniques are available [1,2], and it is advisable to select the optimum technique and sensor.

There is a recent trend for contactless radio sensors to operate in the increasingly higher microwave band.

Radio regulations distinguish between "tank level probing radar" inside closed metallic tanks or silos and "level probing radar" outside with more restrictions.

Tank Level Probing Radar (TLPR) applications are based on pulse RF, FMCW, or similar wideband techniques. TLPR radio equipment types can operate in all or part of the frequency bands as specified in Table I.

TABLE I
TANK LEVEL PROBING RADAR (TLPR) PERMITTED FREQUENCY BANDS [3]

	TLPR assigned frequency bands (GHz)
Transmit and receive	4,5 to 7
Transmit and receive	8,5 to 10,6
Transmit and receive	24,05 to 27
Transmit and receive	57 to 64
Transmit and receive	75 to 85

For aperture antennas, there is a clear relationship between the geometrical area of the aperture and the effective aperture area. This relationship is the aperture efficiency, which is typically between 0.4-0.7. The operating free-space wavelength gives the relationship between effective aperture area and antenna gain.

$$A_{geom}\eta_A = A_{eff} = \frac{\lambda^2}{4\pi}G \quad (1)$$

where

G the antenna gain,
A_{geom} the geometrical area of the aperture,
η_A aperture efficiency,
A_{eff} effective aperture area
λ operating free space wavelength.

For antennas considered lossless, the gain of the antennas can be expressed in terms of the main beam cone angle.

$$G \approx \frac{16}{\Theta^2} \quad (2)$$

where
Θ the antenna main beam cone angle.

It is easy to see from equations (1) and (2) that for the same aperture geometrical area, the main beam cone angle of the antenna decreases with increasing frequency.

$$\Theta \sim \lambda = \frac{c}{freq} \quad (3)$$

where
$freq$ operating frequency,
c the speed of light in a vacuum.

From the relationship in equation (3), it is clear, that for the same antenna aperture size, as the frequency increases, the beamwidth decreases, and the narrower antenna foot reduces the reflection from the tank sides. In addition, increasing the frequency has the further advantage of improving the radar distance measurement resolution and reducing the circuit dimensions. For these reasons we have chosen the 75-85 GHz frequency band from the allocated TLPR bands. [18, 19]

From relation (3), the main beam widths at 25 and 80 GHz can be compared. (Fig. 3)

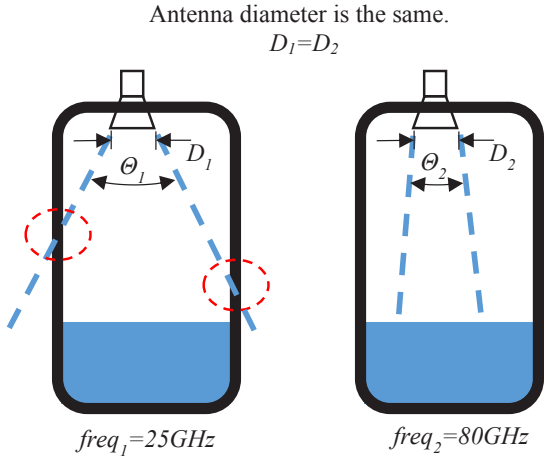


Fig. 3. Measurement in narrow tanks or silos [4]

In practice, the footprint size is often used to compare the size of the tank or silo and the main beam at each frequency. (Fig. 4.) The simulation parameters are aperture geometrical area 0.01m^2, aperture efficiency 0.6, distance 10m.

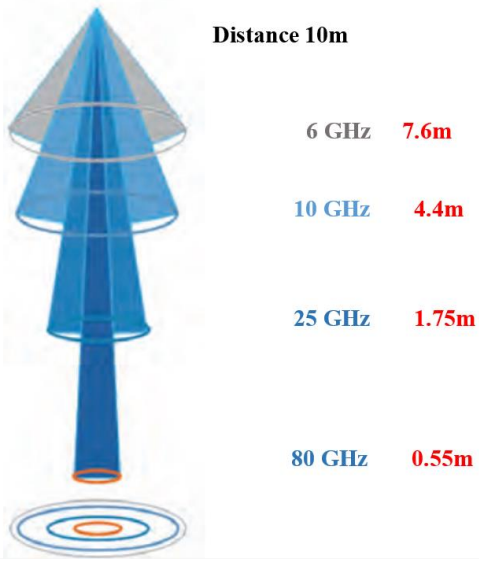


Fig. 4. Antenna footprint diameters for different frequencies

In summary, the antenna radiation field is inversely proportional to the aperture diameter of the antenna and the center frequency. Beam width decreases with increasing center frequency if the diameter of the antenna is kept constant. Furthermore, in the case of keeping the frequency constant, the beam width also decreases with the increasing diameter of the antenna. In conclusion, the beam width does not simply depend on one single parameter, but both parameters—center frequency and antenna diameter are degrees of freedom for determining the angular beam width. The choice of one specific antenna from a set of available antennas with different beam widths must be made dependent on the given application conditions.

II. Dielectric Materials and Permittivity Measurements

The most common plastic materials used as radomes and dielectric lenses for antennas are Polypropylene (PP), Polyvinylidene Fluoride (PVDF), and Polytetrafluoroethylene (PTFE). The electrical parameters, permittivity, and loss of these materials are examined below.

II.1 Dielectric materials

Polypropylene (PP) is one of the most widely used and low-cost thermoplastics with adequate physical properties, such as low density and high heat resistance [5]. PP is generally found as a homopolymer and copolymer. The first one consists of propylene monomers, and it has a high strength-to-weight ratio and good chemical resistance. The second one includes monomers in the PP backbone, and it is tougher and more flexible, with a lower melting point and high-impact resistance at low temperatures than PP homopolymer [6]. Because of all these advantages, polypropylene-based composites have been extensively used for automotive, construction, and packaging applications.

Polyvinylidene fluoride (PVDF) is a highly non-reactive and pure thermoplastic fluoropolymer material. Below 150 ºC, PVDF becomes ferroelectric. Thus, PVDF is an electroactive and semicrystalline polymer with pyro and piezoelectric properties at room temperature, which can be used for many applications [7]. It exhibits high mechanical strength, good chemical resistance, thermal stability, and excellent aging resistance [8]. Moreover, PVDF is an attractive polymer matrix for composite material with superior mechanical and electrical properties.

Polytetrafluoroethylene (FTFE) is a synthetic fluoropolymer of tetrafluoroethylene and is a PFAS that has numerous applications. The commonly known brand name of PTFE-based composition is Teflon. Polytetrafluoroethylene is a fluorocarbon solid, as it is a high-molecular-weight polymer consisting wholly of carbon and fluorine. PTFE is hydrophobic: neither water nor water-containing substances wet PTFE, as fluorocarbons exhibit only small London dispersion forces due to the low electric polarizability of fluorine. PTFE has one of the lowest coefficients of friction of any solid. PTFE is used as a non-stick coating for pans and other cookware. It is non-reactive, partly because of the strength of carbon–fluorine bonds, so it is often used in containers and pipework for reactive and corrosive chemicals. When used as a lubricant, PTFE reduces friction, wear, and energy consumption of machinery.

In the literature [9], the electrical properties of materials are generally known at lower frequencies (1 kHz, 1 MHz), and some materials are manufactured in several versions, so it is necessary to carry out measurements that give the electrical properties of materials in the millimeter waveband.

TABLE II
Electrical Properties of Plastic Materials [9]

Material	Dielectric constant @1MHz	Dissipation factor @ 1MHz	Volume resistivity Ohm/cm
PP	2.2-2.6	0.0003 - 0.0005	10^{16} -10^{18}
PVDF	8.4	0.06	10^{14}
PTFE	2.0-2.1	0.0003 - 0.0007	10^{18} -10^{19}

In the following, such a measurement procedure is presented and the results of electrical parameter measurements for the three materials are reported.

II.2 Material Parameter Measurements

In practice, microwave measurements of electrical material parameters are based on transmission and reflection measurements. (Fig. 5.) [10,11]

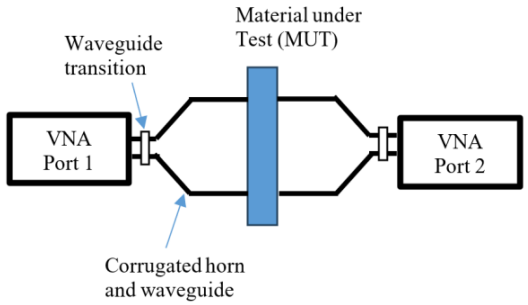


Fig. 5. Measurement set-up for transmission and reflection measurements of electrical material parameters.

According to the setup shown in Figure 5, the material parameter measurement is performed with 2 ports Vector Network Analyzer, and the sample is placed between two corrugated horn antennas for measurement. Properly sized corrugated horn antennas provide plane-wave excitation for the sample.

Measurements are preferably carried out using the Swissto12 Material Characterization Kit (MCK). (Fig. 6.) MCKs are used for measurement of the permittivity and loss tangent of planar specimens, foams, and multilayered materials at room temperature. The MCKs are available to cover the frequency range 50 GHz to 750 GHz.

Fig. 6. Measurement set-up with Swissto12 MCK and Vector Network Analyzer.

To model the measurements, a microwave signal flow graph network model of the measurement setup is presented. (Fig. 7.) [12-14]

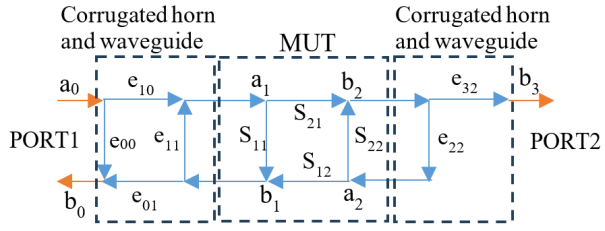


Fig. 7. Microwave signal flow graph of the measurement setup

The VNA measures the reflection and transmission between PORT1 and PORT2.

We introduce the evaluation of transmission parameters from signal flow graph of the measurement setup. The equations below can be used to express the transmission between PORT1 and PORT2.

$$e_{10}a_0 + e_{11}b_1 = a_1$$
$$S_{11}a_1 + S_{12}a_2 = b_1$$
$$S_{21}a_1 + S_{22}a_2 = b_2$$
$$b_3 = e_{32}b_2$$
$$e_{22}b_2 = a_2$$

The measured $S_{m,21}$ parameter $S_{m,21} = b_3/a_0$ can be expressed as.

$$\frac{b_3}{a_0} = \frac{S_{21}e_{10}e_{32}}{1 - S_{11}e_{11} - S_{22}e_{22} + e_{11}e_{22}(S_{11}S_{22} - S_{12}S_{21})}$$

The measured $S_{m,11}$ parameter $S_{m,11} = b_0/a_0$ can be similarly expressed.

Finally, the permittivity and loss factor of the sample are calculated from the S_{ij} scattering coefficients of the MUT.

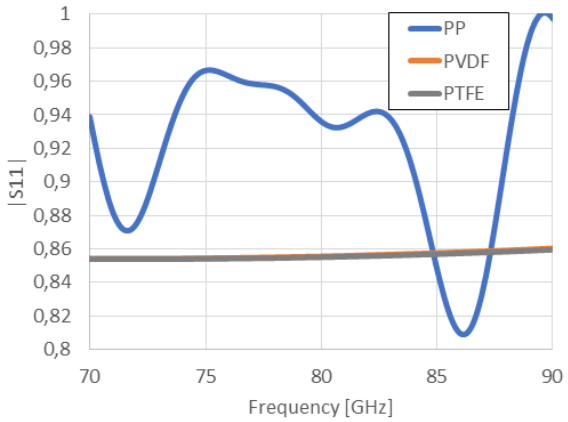


Fig. 8. Measured reflection parameters for PP, PVDF and PTFE samples.

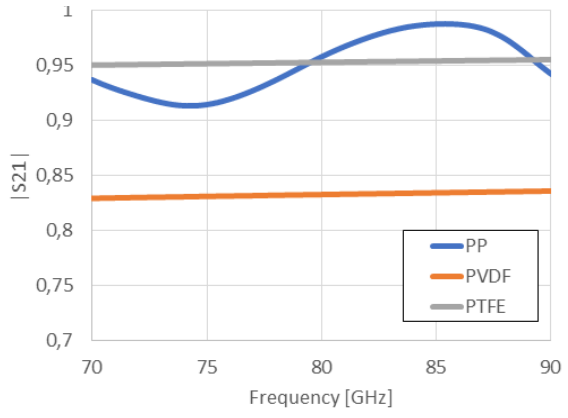


Fig. 9. Measured transmission parameters for PP, PVDF and PTFE samples.

The two material parameters (permittivity and loss tangent) for dielectric materials can be evaluated using S_{11} and S_{21}. These are introduced for PP, PVDF and PTFE in Figure 10 and 11.

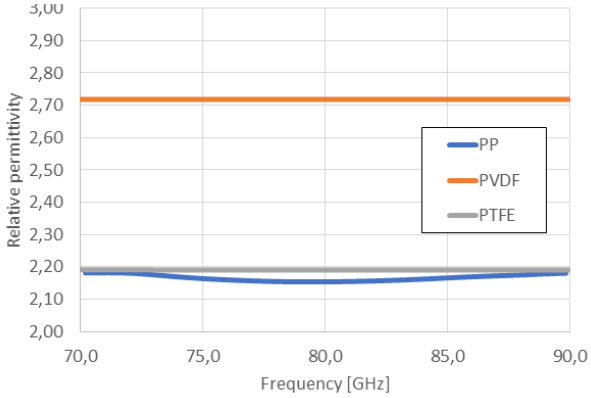


Fig. 10. Relative permittivity for PP, PVDF and PTFE samples.

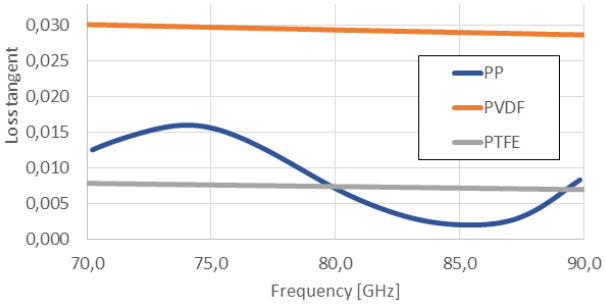


Fig. 11. Loss tangent for PP, PVDF and PTFE samples.

From the measurement results it can be concluded that the permittivity of PTFE can be considered constant over the entire frequency range investigated, and its loss factor is also essentially constant and can be the lowest.

For the horn antenna design with a dielectric lens, we use PTFE in the next sessions.

III. Single Chip Integrated FMCW Radar IWR1443

The IWR1443 device [15, 20] is an integrated single-chip millimeter wave (mmWave) sensor based on FMCW radar technology capable of operation in the 76- to 81 GHz band with up to 4 GHz continuous chirp. The device is built with TI's low-power 45-nm RFCMOS process, and this solution enables unprecedented levels of integration in a tiny form factor. The IWR1443 is an ideal solution for low-power, self-monitored, ultra-accurate radar systems in industrial applications such as building automation, factory automation, drones, material handling, traffic monitoring, and surveillance. The IWR1443 device is a self-contained, single-chip solution that simplifies the implementation of mmWave sensors in the 76 to 81 GHz band. The IWR1443 includes a monolithic implementation of a 3TX, 4RX system with built-in PLL and A2D converters. The device includes fully configurable hardware accelerator that supports complex FFT and CFAR detection. Additionally, the device includes two ARM R4F-based processor subsystems: one processor subsystem is for master control, and additional algorithms; a second processor subsystem is responsible for front-end configuration, control, and calibration. Simple programming model changes can enable a wide variety of sensor implementations with the possibility of dynamic reconfiguration for implementing a multimode sensor.

Automotive radar applications use generally all three Tx and four Rx channels for radar imaging but for our tank-level radar only one Tx and one Rx channels we apply.

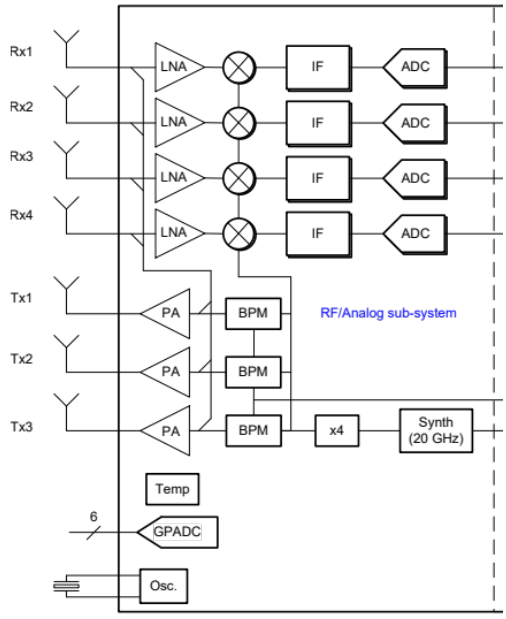


Fig. 12. RF sub-system functional block diagram of IWR1443 [15].

The Fig. 13. shows demo layout of IWR1443 chip with microstrip transmit and receive antennas.

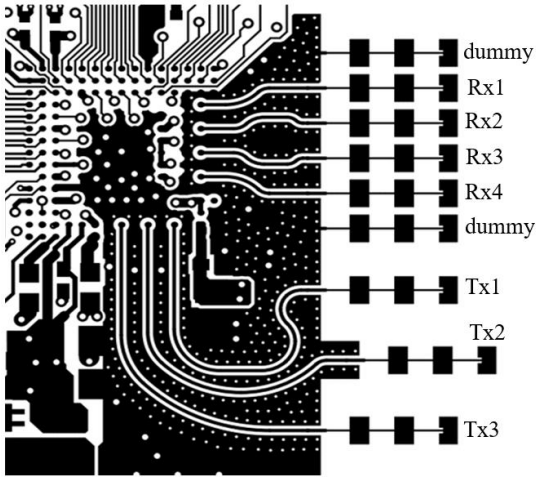


Fig. 13. FMCW radar layout (IWR1443BOOST) using IWR1443 [16].

Fig. 14 shows the main parameters of the antenna. The antenna gain, G=9.59dB, half power conical beam angle, Θ_{3dB}=25 degrees, sidelobe suppression ratio, SLSR=26dB.

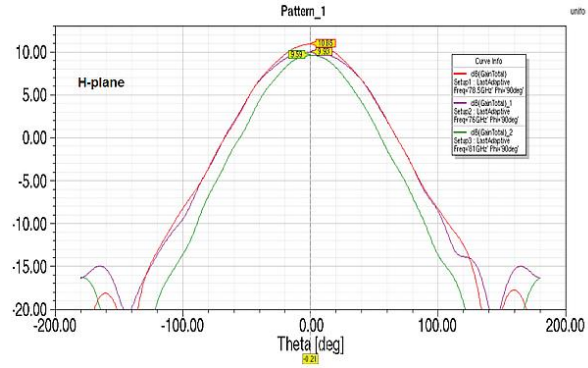


Fig. 14. Microstrip antenna radiation pattern for IWR1443BOOST Evaluation Module [16].

For the planned radar, the microstrip antennas of the IWR1443BOOST demo board cannot be used and new antenna should be designed for the next reasons.

The antenna gain should be increased, and conical beam angle should be decreased to suppress multiple reflections.

The antenna chosen for the design must also be resistant to pressure and corrosive humid media.

Due to these requirements, the antenna type chosen for the analysis was a conical horn antenna. Two types were investigated, the open horn antenna and the closed horn antenna with dielectric lens.

IV. Horn Antenna Design

Cross-sectional images of the analyzed horn antennas are shown in Figure 15 and 16. The material used for the lens was PTFE, which was found to be the best material based on a material parameter test in II. section.

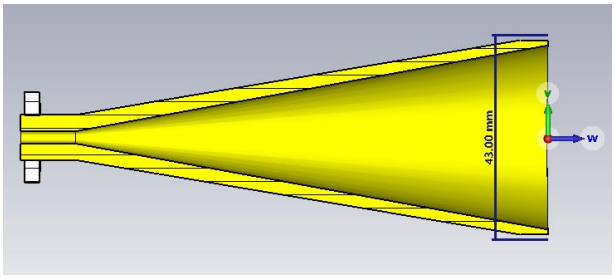


Fig. 15. Cross-sectional image of the analyzed open horn antenna.

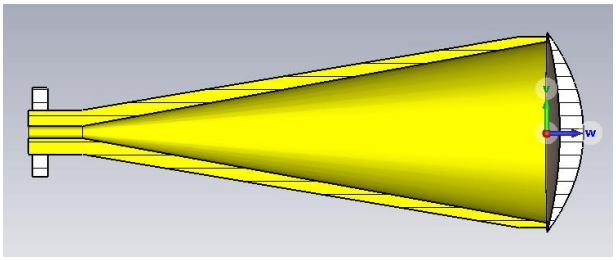


Fig. 16. Cross-sectional image of the analyzed closed horn antenna with dielectric lens.

The simulations were performed using the CST MWS electromagnetic field simulator and the results of the two main simulations (input reflection and radiation pattern) are shown in Figures 17 and 18.

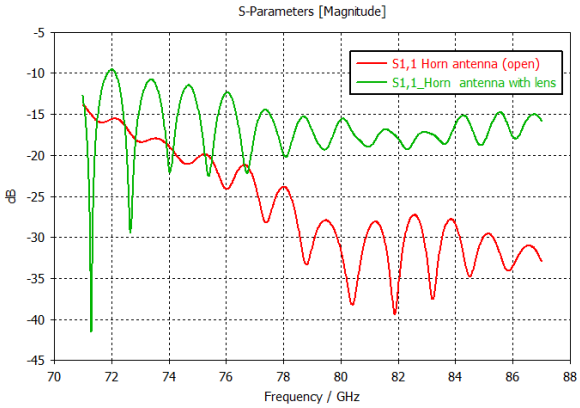


Fig. 17. Input reflection of the horn antennas.

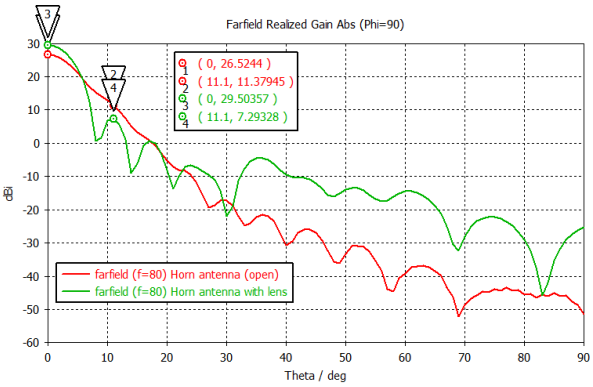


Fig. 18. Radiation pattern of the horn antennas.

The design requirement for input reflection generally below -10 dB, so each of the assumed antennas fulfill that. The antenna gain is optimized using lens on horn aperture and the gain is G=29.5dB. The half power conical beam angle, Θ_{3dB}=6 degrees.

V. Ring Hybrid Design

In radar technology, hybrid ring duplexers often are transmit/receive (TR) switch based on a simple rat-race coupler.

The rat-race coupler has four ports, each placed one-quarter wavelength away from each other around the top half of the ring. (Fig. 19.) A signal input on port 1 will be split between ports 2 and 4, and port 3 will be isolated. [17]

The full simulation has been performed for all gate input reflections and transmissions between gates. Of these, only the simulation results for the transmission between ports 2-3 and port 3 input reflections are reported for different ring radius r=0.5, 0.55 and 0.6 mm. (Fig. 20 and 21)

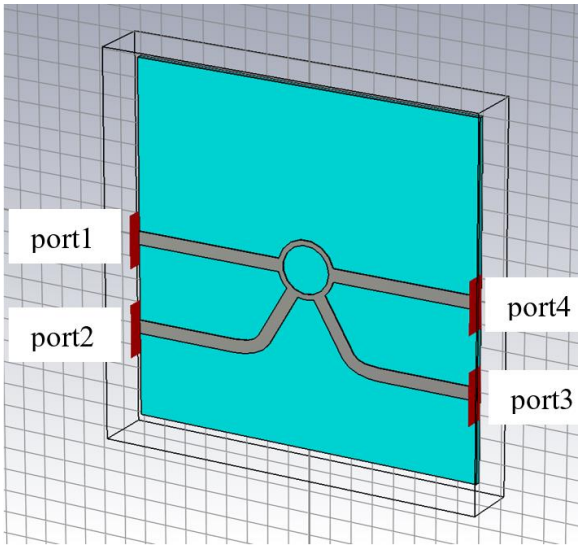


Fig. 19. Rat-race coupler CST simulation model.

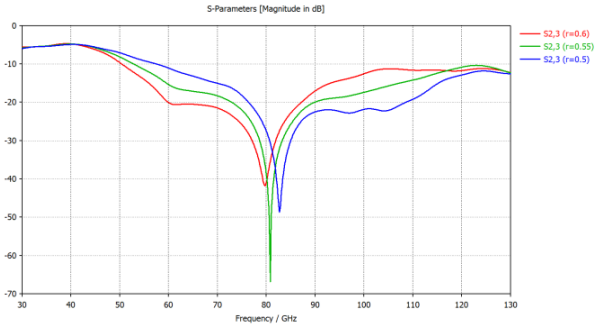


Fig. 20. Rat-race coupler isolation between port 2 and 3 (CST simulation).

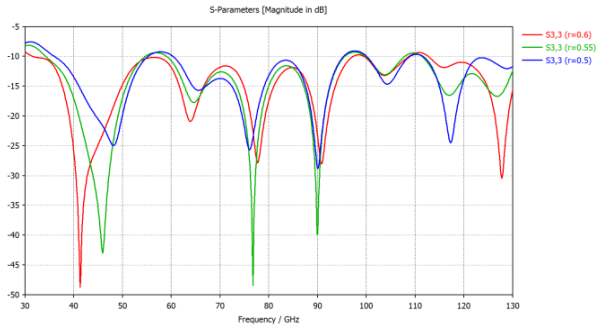


Fig. 21. Rat-race coupler input reflection for port 3 (CST simulation).

The RX1 port is used for reception, while the RX2 port is only used as a matched termination. The TX1 port of the chip's transmit channels is used. (Fig. 22 and 23.)

The radar sensor was tested with a flat target surface. The range-profile characteristics are shown in Fig. 24. for a target with distance of 22m.

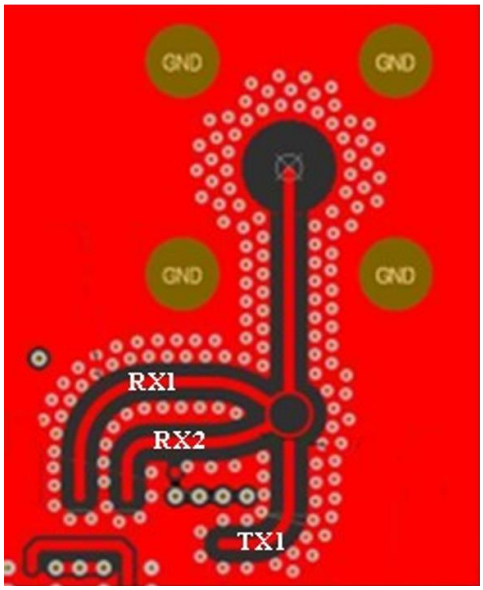


Fig. 22. Rat-race coupler in the radar layout.

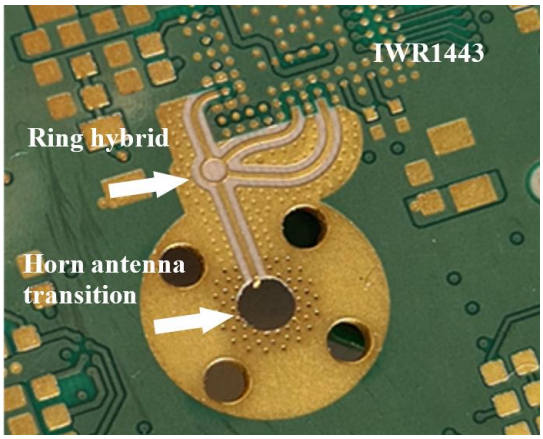


Fig. 23. Rat-race coupler and microstrip-conical waveguide transition.

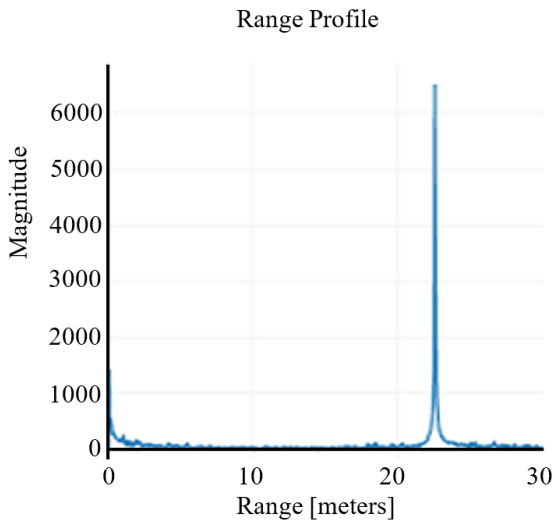


Fig. 24. Measured range-profile characteristics.

VI. SUMMARY

We presented a new systematic design of an 80 GHz radar sensor for contactless tank-level measurement. The reason for choosing this frequency band, mainly to reduce the main beamwidth of the antenna, has been presented. Detailed analysis of possible lens dielectric materials, horn antenna design results and ring hybrid for transmit-receive separation are introduced.

Layout of the antenna and the implemented tank level radar is introduced. Finally, field test was performed, and the measured range-profile test result is presented for the radar to demonstrate the correct functioning of the design.

As a continuation of this work, we plan to increase the range of radar measurements, both near and far.

ACKNOWLEDGMENT

The author very much appreciates the support from the NIVELCO Process Control Co. and the infrastructure of the Department of Broadband Infocommunication and Electromagnetic Theory at the Budapest University of Technology and Economics. The measurements were carried out in Karlsruhe at Karlsruhe Institute of Technology (KIT) at Höchstfrequenztechnik und Elektronik.

This research was funded by the Hungarian Fund 2020-1.1.2-PIACI-KFI-2021-00278.

REFERENCES

[1] Vogt, M., Michael, G.: Silo and tank vision: applications, challenges, and technical solutions for radar measurement of liquids and bulk solids in tanks and silos. *IEEE Microw. Mag*. 18(6), 38–51 (2017)

[2] Vass, Gergely. "The Principles of Level Measurement RF capacitance, conductance, hydrostatic tank gauging, radar, and ultrasonics are the leading sensor technologies in liquid level tank measurement and control operations." Sensors 17 (2000): 55-64.

[3] ETSI EN 302 372, ETSI EN 302 372 V2.1.1 (2016-12), Short Range Devices (SRD); Tank Level Probing Radar (TLPR) equipment operating in the frequency ranges 4,5 GHz to 7 GHz, 8,5 GHz to 10,6 GHz, 24,05 GHz to 27 GHz, 57 GHz to 64 GHz, 75 GHz to 85 GHz; Harmonised Standard covering the essential requirements of article 3.2 of the Directive 2014/53/EU, https://cdn.standards.iteh.ai/samples/46667/1fb304055d624fa1925191db920c0351/ETSI-EN-302-372-V2-1-1-2016-12-.pdf

[4] Dave Grumney, Understanding Radar Technology for Measuring Tank Level, InTech e-edition covering the fundamentals of automation, MARCH 2020, https://www.automation.com/getmedia/6d1ad741-4eb3-450f-9084-7833c2e46105/InTechFOCUS-march2020.pdf

[5] Gogoi R, Kumar N, Mireja S, Ravindranath SS, Manik G, Sinha S. Effect of hollow glass microspheres on the morphology, rheology and crystallinity of short Bamboo fiber-reinforced hybrid polypropylene composite. Journal of Metals. 2019;71(2):548-558. DOI: 10.1007/s11837-018-3268-3

[6] Gogoi R, Maurya AK, Manik G. A review on recent development in carbon fiber reinforced polyolefin composites. Composites Part C: Open Access. 2022;8:100279. DOI: 10.1016/j.jcomc.2022.100279

[7] Zhen Y, Arredondo J and Zhao G 2013 Open Journal of Organic Polymer Materials 3 99

[8] Kang G D and Cao Y M, Application and modification of poly(vinylidene fluoride) (PVDF) membranes – A review 2014 J. Membr. Sci. 463 145

[9] https://www.professionalplastics.com/professionalplastics/ElectricalPropertiesofPlastics.pdf

[10] James Baker Jarvis, Transmission/ Reflection and Short-Circuit Line Methods for Measuring Permittivity and Permeability, *NIST Technical Note* 1341, 1990. https://nvlpubs.nist.gov/nistpubs/Legacy/TN/nbstechnicalnote1341.pdf

[11] Alireza Kazemipour *et. al*., Analytical Uncertainty Evaluation of Material Parameter Measurements at THz Frequencies, *International Journal of Infrared and Millimeter Waves*, 24 July 2020. DOI: 10.1007/s10762-020-00723-0

[12] Alexandros I. Dimitriadis, Dielectric Measurements at mm-Wave Frequencies with the Material Characterization Kit (MCK), Towards Terahertz Technology for High Throughput Communications, EPFL, Lausanne, 5-7 February 2020

[13] A. Kazemipour *et al*., "Standard Load Method: A New Calibration Technique for Material Characterization at Terahertz Frequencies," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021, Art no. 1007310, DOI: 10.1109/TIM.2021.3077660.

[14] Y. Wang *et al*., "Material Measurements Using VNA-Based Material Characterization Kits Subject to Thru-Reflect-Line Calibration," in *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 5, pp. 466–473, Sept. 2020, DOI: 10.1109/TTHZ.2020.2999631.

[15] IWR1443 Single-Chip 76- to 81-GHz mmWave Sensor, SWRS211C – MAY 2017 – REVISED OCTOBER 2018 https://www.ti.com/lit/ds/symlink/iwr1443.pdf

[16] IWR1443BOOST Evaluation Module mmWave Sensing Solution, SWRU518D – May 2017 – Revised May 2020 https://www.ti.com/lit/ug/swru518d/swru518d.pdf

[17] Microstrip Hybrid Ring Coupler. Patent, Accession Number: ADD005347, 1978-06-06, https://apps.dtic.mil/sti/citations/ADD005347

[18] Tim Erich Wegner *et. al*., Fill level measurement of low-permittivity material using an M-sequence UWB radar, *International Journal of Microwave and Wireless Technologies* , Volume 15 , Special Issue 8: EuCAP 2022 Special Issue , October 2023, pp. 1299–1307

[19] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli and A. Yarovoy, "A Survey on Radar-Based Continuous Human Activity Recognition," in *IEEE Journal of Microwaves*, vol. 3, no. 3, pp. 938–950, July 2023

[20] Li, B.Y., Ding, S.Y., Ma, D., Wu, Y.X., Liao, H.J. and Hu, K.Y. (2024) LLMCount: Enhancing Stationary mmWave Detection with Multimodal-LLM. https://arxiv.org/abs/2409.16209

Lajos Nagy He received the Engineer option Communication) and PhD degrees, both from the Budapest University of Technology and economics (BME), Budapest, Hungary, in 1986 and 1995, respectively. He joined the department of Microwave Telecommunications (now Broadband Infocommunications and Electromagnetic Theory) in 1986, where he is currently an associate professor. He is a lecturer on graduate and postgraduate courses at BME on Antennas and radiowave propagation, Radio system design, Adaptive antenna systems and Computer Programming. His research interests include antenna analysis and computer aided design, electromagnetic theory, radiowave propagation, communication electronics, signal processing and digital antenna array beamforming, topics where he has produced more than 100 different book chapters and peer-reviewed journal and conference papers. Member of Hungarian Telecommunication Association, official Hungarian Member and Hungarian Committee Secretary of URSI, Chair of the IEEE Chapter AP/ComSoc/ED/MTT.

Physically Tenable Analysis and Control of Scattering from Reconfigurable Intelligent Surfaces

Botond Tamás Csathó, *Graduate Student Member, IEEE, Member, HTE,* and Bálint Péter Horváth

***Abstract*—Reconfigurable intelligent surface is a promising concept within the scope of smart radio environment, which is a key enabler of the future wireless networks. Efficient numerical modeling of such devices constitutes a fundamental and actively pursued research challenge. This study numerically analyzes the reflection properties of a particular reconfigurable intelligent surface with the aid of computational electromagnetics. A key advantage of utilizing full-wave simulations is that they capture all the physical phenomena within the structure, thus providing a physically stenable analysis method. An essential aspect of RIS modeling is the configuration pattern design of the surfaces. A standard objective function of the pattern design is the amount of energy reradiated toward the target direction. The utilization of full-wave simulations limits the applicable optimization methods. In this article, an intuition-based pattern search method is presented to design RIS configurations, with the radiation pattern of the RIS structure in free space as the objective function. The suggested method first identifies a set of configuration values, then exhaustively searches through their combinations, seeking for the highest anomalously reflected power. The first presented result is the demonstration of creating anomalous reflections, with the dominant reflection being electrically tunable. The second contribution is the aforementioned pattern search method, which enables the reradiation of the incident energy for numerous anomalous directions. The average scattering parameter amplitude for the scenario is 0.78. Finally, we also demonstrate the effect of the structure being finite in size. We conclude that the dominant radiation directions coincide with the modes of the infinite periodic counterpart.**

***Index Terms*—Reconfigurable Intelligent Surface (RIS), Intelligent Reflecting Surfaces (IRS), configurable MeTaSurfaces (MTS), periodic structures, full-wave simulation**

I. INTRODUCTION

THE smart radio environment (SRE) is a paradigm that has recently emerged to meet the unprecedented requirements of future wireless networks such as sixth generation (6G) [1]. The fundamental idea of SRE is to jointly optimize the wireless channel and the communicating endpoints [1]. The idea of reconfigurable intelligent surfaces (RISs) is a chief concept within the scope of SRE [2]. An RIS is a surface designed to configurably modify the scattered electromagnetic (EM) field [2].

An RIS consists of numerous elements (unit cells) whose EM properties can be electronically adjusted, for example, with tunable varactor diodes [3], switchable positive-intrinsic-negative diodes [4] or liquid crystal technology [5]. An RIS can be designed based on the well-established concept of reflectarrays or metasurface (MTS) technology [6]. The latter is superior due to the available EM field transformations [6], and utilized, e.g., to improve antenna structures [6]–[8]. It is important to note that RISs are primarily envisioned as nearly passive devices, meaning only their configuration requires energy. Numerous visions exist for scenarios where an RIS deployment is beneficial [9]. These include communicating with a user in blockage, creating additional signal paths, increasing the channel rank, and suppressing interference.

The interaction between the impinging EM field and the RIS can be described with macroscopic or microscopic models. Macroscopic models omit the particular structure of the unit cells and instead use some macroscopic parameters to account for the effect of the surface. Some frequently used macroscopic models are reviewed in [2], [10]. Numerous macroscopic models are employed to design control algorithms and large-scale performance evaluation. Some make simplifications that mask fundamental physical behavior, such as the interaction between neighboring RIS elements, e.g., [11]. In contrast, some macroscopic models are physically tenable; these are often used for the design of metasurfaces; one such model is the generalized sheet-transition condition [7]. Macroscopic models can also be included in ray-tracing simulators suitable for large-scale performance analysis [10].

On the contrary, microscopic models consider the physical implementation of unit cells. Conventionally, this requires full-wave numerical simulation of the structure. Therefore, these models are physically tenable and can capture fundamental phenomena. Full-wave simulations are commonly employed in the design phase [12], and to retrieve macroscopic parameters or validate macroscopic models [7]. Since the full-wave simulation of the complete propagating environment in 3D is resource-demanding, they are not feasible to conduct a large-scale performance evaluation. To circumvent the resource-demand issue, one can improve the efficiency of the numerical models [13]–[15], alternatively, the amalgamation of microscopic and macroscopic models can be employed [16]. The advantage of microscopic models lies in their applicability to almost any unit cell structure, particularly given the computational power available today. Hence, it is possible to analyze cases when homogenization-based macroscopic models are not suitable due to the physical dimensions of the unit cell relative to the wavelength.

Designing RIS control patterns constitutes a fundamental goal of RIS modeling. The choice of the RIS model is interconnected with the pattern design or beamforming methods. With simple macroscopic models, it is possible to optimize the RIS for a particular statistical channel realization [11], [17], [18]. With physically tenable macroscopic models, the typical goal is to tune the reflection pattern of the RIS to

The authors are with the Department of Broadband Infocommunications and Electromagnetic Theory, Budapest University of Technology and Economics, Budapest, Hungary. (E-mail: csatho.botond@edu.bme.hu, horvath.balint@vik.bme.hu)

DOI: 10.36244/ICJ.2025.1.5

achieve the desired pattern while omitting the effect of the environment [6], [12], [19]. From this perspective, microscopic models are similar; i.e., often the reflection pattern of the surface is optimized without considering the environment [13], [14]. In our work, a full-wave simulation-based approach is applied for RIS modeling, considering the size of the analyzed structure relative to the wavelength. An issue of significant importance with full-wave simulation-based RIS pattern design is that as the number of continuous control parameters increases, the optimization problem becomes more computationally demanding. The present article provides an intuition-based solution for this challenge. A more detailed comparison of different RIS pattern design methods is provided after discussing our contribution in Tab. III.

The considered design is based on a prototype described in the literature [3]. We assume that unit cells are configured periodically; consequently, only one period (so-called super-cell) is analyzed. As a first result, we present fundamental effects related to plane wave (PW) scattering from an RIS. Our main contribution is a method to generate RIS configuration patterns to achieve anomalous reflection. The presented method can be utilized to determine advantageous patterns if an RIS design is provided. We demonstrate that it is possible to direct power into configurable reflection directions. Our observations are consistent with the capabilities of antenna arrays and reflectarray technology; namely, by proper phase pattern design, it is possible to create anomalous reflection, i.e., turn the main beam in the desired direction [12]. Such anomalous reflections can be achieved, e.g., with phase-gradient metasurfaces designed based on the generalized Snell's law [12]. In a real-life RIS deployment, one should consider the limited physical size of an RIS. Therefore, as a last result, we compare the scattering properties of a finite-size RIS with the ones of an infinite periodic structure.

The remainder of the manuscript is organized as follows. Section II describes the core concept of EM field scattering from planar periodic structures. Subsequently, the utilized full-wave solver and the analyzed design are introduced in Sec. III. The results obtained from full-wave simulations and the description of the suggested pattern search method are presented in Sec. IV. Finally, conclusions are drawn in Sec. V.

The following mathematical notation is used throughout this paper. $\|\cdot\|$ denotes the 2-norm, $|\cdot|$ and $\arg(\cdot)$ are the absolute value and argument of a complex scalar, respectively. $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$ are sets of integer, real, and complex numbers, respectively.

II. Reflection from Periodic Structures

Let us start with an overview of the EM scattering from periodic structures, which is vital for analyzing the simulation results. We consider an infinite planar periodic structure consisting of rectangular lattices located at $z = 0$ with a PW incident on the surface. The period sizes and the angles describing the wave vectors of the incident and reflected PW, namely θ_i, ϕ_i, θ_r, and ϕ_r are shown in Fig. 1. Subscripts i and r indicate incident and reflected, respectively. The wave vectors

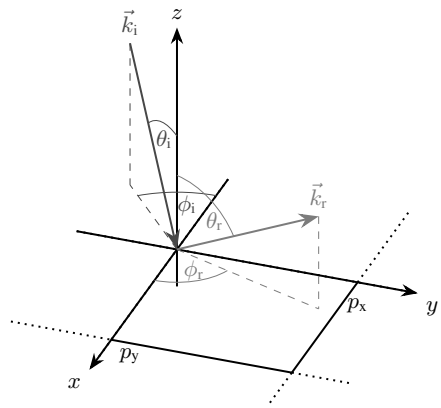


Fig. 1. Incident ($\vec{k}_\mathrm{i}$) and reflected ($\vec{k}_\mathrm{r}$) plane waves and and periodicity (p_x, p_y) of the structure

can be computed as

$$\begin{aligned} \left\|\vec{k}_\nu\right\| = k_\nu = \frac{2\pi}{\lambda} = \frac{2\pi f}{c} = 2\pi f\sqrt{\mu\varepsilon}, \\ \vec{k}_\nu = k_{\nu,\mathrm{x}}\hat{x} + k_{\nu,\mathrm{y}}\hat{y} + k_{\nu,\mathrm{z}}\hat{z}, \\ \text{with } \nu \in \{\mathrm{i},\mathrm{r}\}, \end{aligned} \tag{1}$$

where λ is the wavelength, f is the frequency, c is the speed of light in the medium, μ and ε are the permeability and permittivity of the medium in the $z \geq 0$ region, respectively. Furthermore, $\hat{x}$, $\hat{y}$, and $\hat{z}$ are the unit vectors in the X-, Y-, and Z-axes, respectively.

Applying the Floquet-Bloch theory, it can be shown that the scattered EM field of a periodic structure can be expressed as a linear combination of particular PW components [20]. It can be derived that the wave vector of these components is determine by the periodicity of the structure. According to Bhattacharyya [20], in the case of a rectangular grid, such as the one in Fig. 1, these wave vectors are

$$\begin{aligned} k_{\mathrm{r,x},m} = k_{\mathrm{i,x}} + \frac{2\pi m}{p_\mathrm{x}}, \\ k_{\mathrm{r,y},n} = k_{\mathrm{i,y}} + \frac{2\pi n}{p_\mathrm{y}}, \\ \{n,m\} \in \mathbb{Z}, \end{aligned} \tag{2}$$

where p_x and p_y are the length of the period of the structure along the X-axis and Y-axis, respectively, or in other terms, the size of the super-cell. Let us note here that equation (2) can be extended for general grids [20]. An m, n pair is also referred to as a mode. For each mode

$$k_\mathrm{i}^2 = k_\mathrm{r}^2 = k_{\mathrm{r},m,n}^2 = k_{\mathrm{r,x},m}^2 + k_{\mathrm{r,y},n}^2 + k_{\mathrm{r,z},m,n}^2 \tag{3}$$

must hold, which also defines $k_{\mathrm{r,z},m,n}$. Consequently,

$$\begin{aligned} k_{\mathrm{r,x},m}^2 + k_{\mathrm{r,y},n}^2 < k_\mathrm{r}^2 &\iff \text{propagating mode}, \\ k_{\mathrm{r,x},m}^2 + k_{\mathrm{r,y},n}^2 > k_\mathrm{r}^2 &\iff \text{evanescent mode}. \end{aligned} \tag{4}$$

Only propagating modes, i.e., when $k_{\mathrm{r,z},m,n} \in \mathbb{R}$ and $k_{\mathrm{r,z},m,n} > 0$ can deliver power to the far-field. The energy propagating in each mode depends on the boundary conditions (BCs) imposed by the structure.

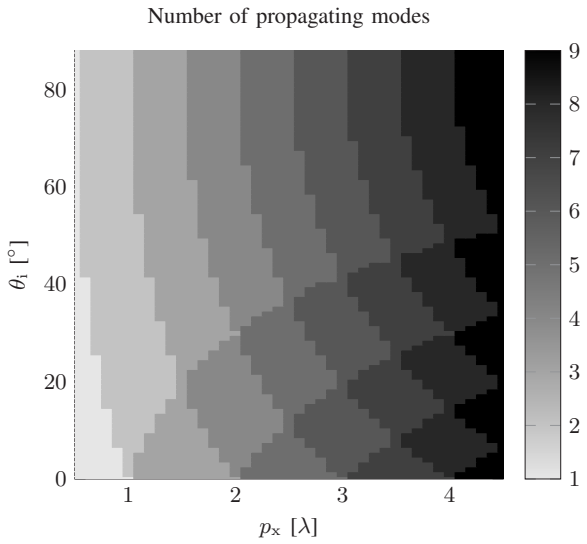


Fig. 2. Demonstrating the effect of varying the period size and incident angle on the number of propagating modes, assuming $\phi_\mathrm{i} = 0\,^\circ$, $p_\mathrm{y} = \lambda$

The number of propagating modes can be determined using (1), (2), (3), and (4). Figure 2 demonstrates the number of propagating modes for varying periodicity and incident angles. One can observe that as the period size increases with respect to the wavelength, the number of propagating modes also increases. Furthermore, the incident angle also affects the possible number of modes.

In conclusion, the period size and the angle of incidence govern the propagating modes By changing the structure, e.g., by modifying some geometrical parameter or tuning an electrically configurable component, it is possible to adjust the power corresponding to each propagating mode, as presented in Sec. IV.

III. Numerical Analysis

A. Simulation Environment

A commercial full-wave EM simulation software[1] is used to study scattering from an RIS in a physically consistent manner. The software solves Maxwell's equations with the finite element method. We assume that the RIS is configured periodically. Therefore, it can be treated as an infinite periodic structure, providing that the effect of the edges is neglected. Accordingly, analyzing only one period, i.e., a super-cell, is sufficient. PW excitation is used for the analysis, which does not restrict generality since an arbitrary EM field can be decomposed into the appropriately weighted sum of PWs with spatial Fourier transform [21]. HFSS can determine the possible propagating modes based on the size of the super-cell and the direction of the impinging PW. For each mode,

[1]Ansys® Academic Research High-Frequency Structure Simulator (HFSS) Release 2021 R2

TABLE I
Numerical data related to the RIS structure depicted in Fig. 3

Variable	Value	Variable	Value
p_x	31.4 mm	l_4	0.6 mm
p_y	22.6 mm	l_5	2.9 mm
l_1	20.3 mm	l_6	4.3 mm
l_2	4 mm	l_7	0.6 mm
l_3	8.3 mm	t_1	3 mm
ε_1	$2.65\varepsilon_0$	$\tan\delta_1$	0.005
μ_1	μ_0		

it defines two ports, one for transverse electric (TE) and one for transverse magnetic (TM) polarization[2].

We are interested in the scattering parameters (S-parameters) of the super-cell. These are initially defined as the ratio of incident and reflected voltages in an N-port network. HFSS can provide these metrics in terms of EM fields. According to the definition of S-parameters,

$$S_{j;k} = \left.\frac{\Psi_{\mathrm{r},j}}{\Psi_{\mathrm{i},k}}\right|_{\Psi_{\mathrm{i},k}=0 \text{ if } l\neq k}, \; j,k,l = 1,\ldots,\mathrm{N}, \tag{5}$$

where $S_{j;k} \in \mathbb{C}$, N is the number of ports defined by HFSS, and $\Psi_{\mathrm{r},j}$, $\Psi_{\mathrm{i},k}$ are unitless complex amplitudes of the reflected and incident fields corresponding to the j^{th} and k^{th} port, respectively [23]. Therefore,

$$\sum_{j=1}^{\mathrm{N}} \left|S_{j;k}\right|^2 \leq 1, \; \forall k, \tag{6}$$

must hold if the structure is passive because of the principle of conservation of energy. One needs to verify that the simulation results satisfy (6), since this is a straightforward indicator of a physically meaningful model.

As an example of the utilized notation, $S_{0,0,\mathrm{TE};1,0,\mathrm{TM}}$ is the scattering parameter corresponding to the $m = 0, n = 0$ mode in TE polarization, if the excitation is given in the $m = 1, n = 0$ mode with TM polarization.

B. Analyzed Unit Cell

Let us now introduce the studied RIS design. The unit cell arrangement is based on a manufactured and measured prototype [3]. We tune the original structure to operate around $f = 4.7\,\mathrm{GHz}$ (5G New Radio Frequency Range 1 n79 band). The sketch of the HFSS model is shown in Fig. 3, while the parameter values are summarized in Tab. I, where ε_0 and μ_0 denote the permittivity and the permeability of the vacuum, respectively. In the original design, there is a second substrate layer for the biasing lines, which is omitted in the HFSS model to save resources. Therefore, the biasing lines are connected to the ground.

In the HFSS model, periodicity is incorporated using `Lattice Pair` type BCs. A `Floquet Port` is the source of the PW excitation. The varactor diodes are modeled with `Lumped RLC` BCs, with zero resistance and inductance. The

[2]Let us note that the EM scattering from a boundary is usually described with TE and TM polarized components. These components are orthogonal to each other. Moreover, an arbitrary PW can be decomposed into TE and TM polarized components with respect to a plane [22].

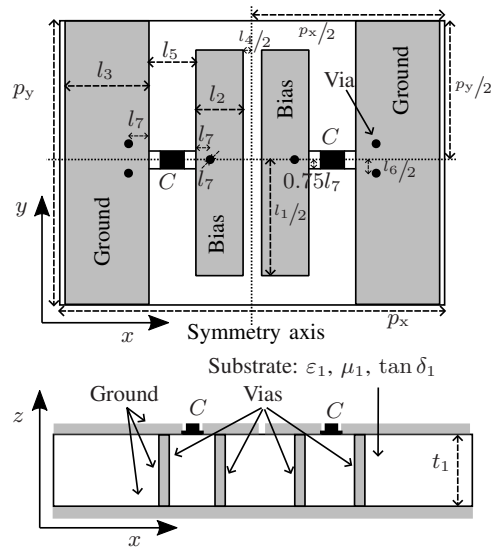


Fig. 3. Analyzed RIS structure. Parameter values are shown in Tab. I. $\tan\delta$ denotes the dielectric loss tangent.

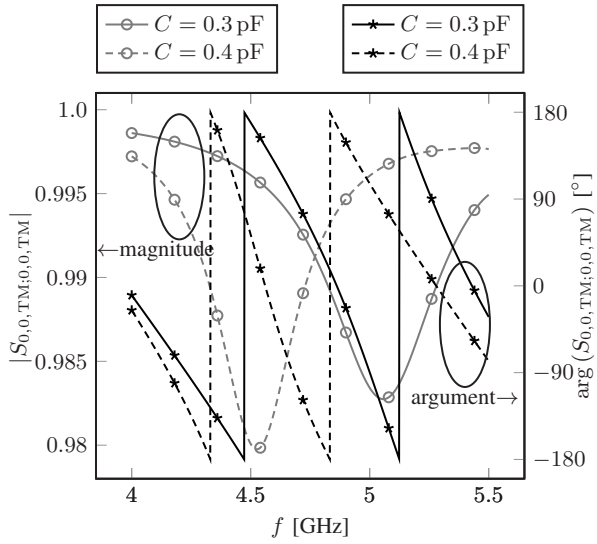


Fig. 4. Magnitude and argument of the complex-valued scattering parameter of a single unit cell $\theta_\mathrm{i} = \phi_\mathrm{i} = 0\,°$

capacitance (C) values are the control parameters of the RIS and equal within a unit cell. All chopper parts are considered with `Perfect E` BCs. The substrate is modeled as a user-defined dielectric, and there is a vacuum above the RIS.

The scattering parameters of the model are shown in Fig. 4. The curves are obtained with `Interpolating HFSS Frequency Sweep` based on the solutions corresponding to 10 different frequencies, with a mesh generated at $5.5\,\mathrm{GHz}$.

IV. Results and Discussion

A. Analyzing Scattering Properties

Having introduced the design, the first presented result is the angular behavior of a single unit cell. Assuming that the excitation is applied in TM mode, the scattering parameters are shown in Fig. 5. The results for the TE polarized excitation are very similar; thus, we omit the corresponding results for brevity. The S-parameters can exhibit extreme deviations for varying incident angles. Furthermore, one can observe that C also affects reflection properties. In the presented case, it can turn the mode conversion from TM to TE on and off.

In the case of a single unit cell, there is only one propagating mode with the considered physical parameters. Next, we add a second unit cell along the X-axes. Consequently, two other propagating modes appear due to the nature of periodic structures, as described in Sec. II. These new propagating modes can reflect power anomalously. Figure 6 presents the scattering parameters for varying incidence angle and control parameters. On the one hand, it is apparent that by varying the control parameters, the scattering parameter of a mode can be tuned. On the other hand, one can see that power can be directed in multiple directions simultaneously, leading to reradiation toward unintended directions. Some of these signal components might cause interference. Furthermore, results suggest that interference can be mitigated by appropriately designing the control patterns.

Designing anomalously reflecting metasurfaces with perfect reflection is also described in the literature; see e.g. [12], [19]. It is noteworthy that it has also been shown that perfect anomalous reflection can be achieved, although with polarization conversion [24]. Therefore, this observation underpins the correctness of the analysis approach used in this article. It is important to emphasize that the full-wave simulation based analysis can be applied also to structures where homogenization is not feasible due to their physical dimensions relative to the wavelength. A comparison of different analysis methods is provided in Tab. III. The chosen approach also influences the potential pattern design or beamforming strategy. In the next subsection, we outline one possibility for maintaining manageable computational complexity when applying full-wave simulation.

B. RIS Pattern Search Method

In a nutshell, the fundamental difficulty arises from the number of continuous control parameters. Here, we propose an intuition-based solution to this problem by discretizing the continuous control parameters. The suggested pattern search method starts with identifying a capacitance set from which the C values are chosen. Then, we exhaustively search for optimal control parameter patterns of super-cells consisting of multiple unit cells.

To identify the set of capacitances, a single unit cell is considered. Assuming perpendicular incidence, C is tuned with the `Optimizer` available in the `Optimetrics` of HFSS to achieve a particular phase of the S-parameters at the carrier frequency ($f = 4.7\,\mathrm{GHz}$). The results are summarized in Tab. II. The authors of [3] also used a similar

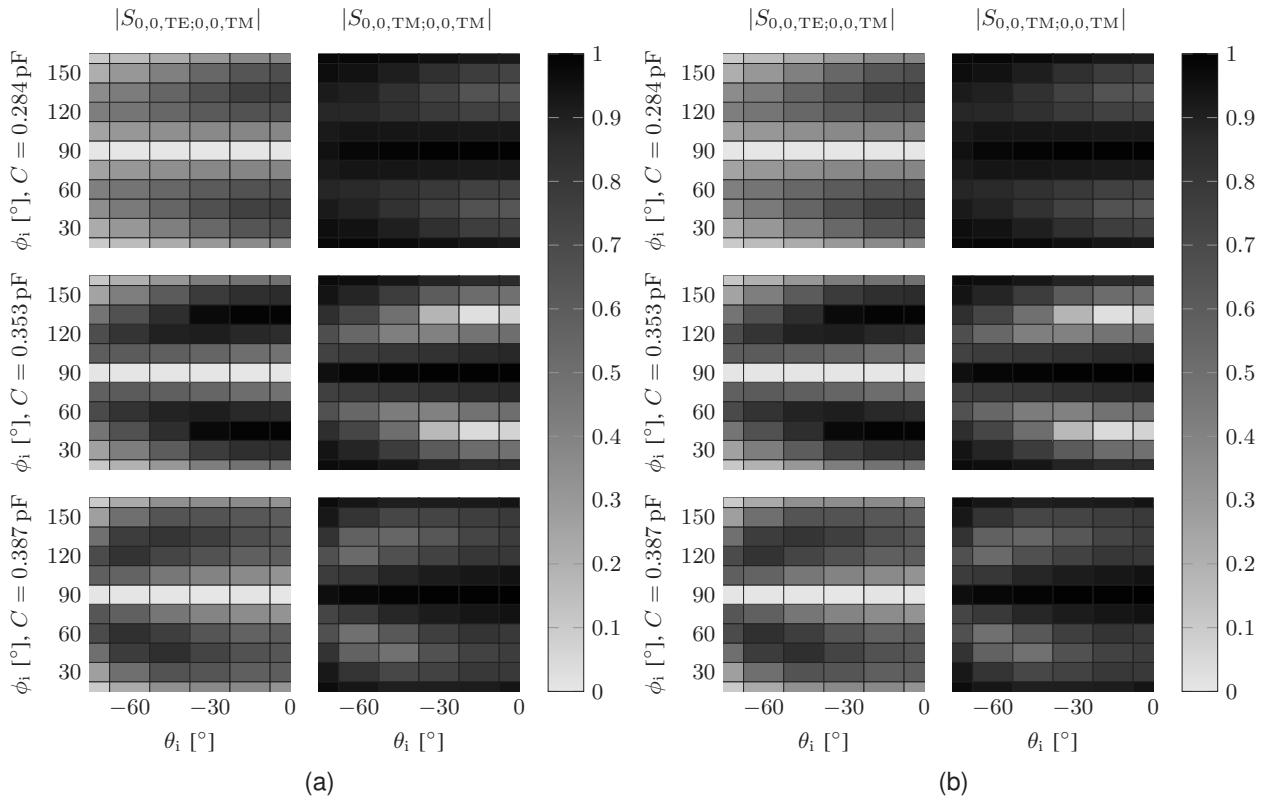


Fig. 5. Angular dependence of scattering parameters for a single unit cell, $f = 4.7\,\mathrm{GHz}$. (a) and (b) represents the magnitude and the phase of the S-parameters, respectively.

approach to determine the set of possible bias voltage values for each varactor diode. Subsequently, we choose control parameter values from Tab. II. For period sizes up to three unit cells, we use the capacitance values corresponding to $90°, 45°, 0°, -45°, -90°, -180°$. Whereas for larger period sizes, the values corresponding to $90°, 0°, -90°$ are employed to render the number of permutations manageable. Then, an exhaustive search is carried out for gradually increasing super-cell sizes. Simulations are evaluated for every possible control parameter permutation for each super-cell. First, we generate all possible permutations of the control parameters. Some permutations are then removed based on two aspects of the periodic BC. *(i)* the patterns which are cyclic shifts of an already evaluated one are ignored; e.g., only one of $C_1 = 0.284\,\mathrm{pF}, C_2 = 0.353\,\mathrm{pF}$ and $C_1 = 0.353\,\mathrm{pF}, C_2 = 0.284\,\mathrm{pF}$ is considered, where C_1 and C_2 denote the control parameter values of the two unit cells of the super-cell. *(ii)* a pattern is omitted if its periodicity is smaller than the size of the super-cell; e.g., $C_1 = C_2 = 0.353\,\mathrm{pF}$ is ignored for evaluating super-cells consisting of two unit cells since it is already considered for the case of a single unit cell. To enhance traceability, ϕ_i is set to $0\,°$. However, we would like to stress that the S-parameters depend on ϕ_i. An example of the obtained curves is shown in Fig. 6. As one can observe, the applied control pattern influences the amount of power directed into each mode.

TABLE II
CONTROL PARAMETER VALUES, $\theta_\mathrm{i} = \phi_\mathrm{i} = 0\,°$, $f = 4.7\,\mathrm{GHz}$

C [pF]	$\approx \arg\left(S_{0,0,\mathrm{TM};0,0,\mathrm{TM}}\right)$ [°]
0.051	135
0.284	90
0.330	45
0.353	0
0.374	−45
0.387	−90
0.423	−135
0.500	−180

The mode indices can be converted into reflected angles, given the incident angle and periodicity based on (1), (2), (3), and (4). Assuming $\phi_\mathrm{i} = 0\,°$, and $-90\,° < \theta_\mathrm{i} < 90\,°$, it follows that for propagating modes

$$\begin{aligned} k_{\mathrm{i,x}} &= k_0 \sin\theta_\mathrm{i}, k_{\mathrm{i,y}} = 0, \\ k_{\mathrm{r,x},m} &= k_{\mathrm{i,x}} + \frac{2\pi m}{p_\mathrm{x}}, \{m\} \in \mathbb{Z}, \\ \theta_\mathrm{r} &= \arcsin\frac{k_{\mathrm{r,x},m}}{k_0}. \end{aligned} \tag{7}$$

Let us note that (7) holds if modes corresponding to $n \neq 0$ are evanescent, which is valid for the considered period size, namely $p_\mathrm{y} = 22.6\,\mathrm{mm} \approx 0.35\,\lambda$.

The mode indices are converted to reflection angles using (7). Subsequently, we made a collection of possible θ_i-θ_r pairs.

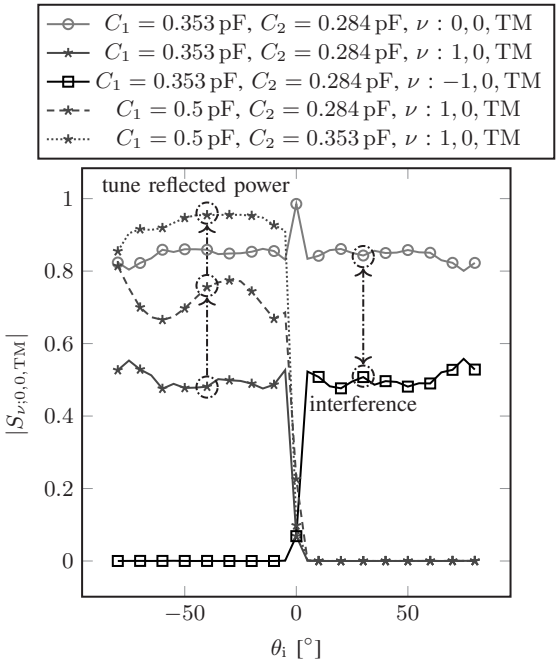


Fig. 6. Scattering parameters for a two unit cell size super-cell, $\phi_\mathrm{i} = 0\,°$, $f = 4.7\,\mathrm{GHz}$. TE modes are omitted due to their low power for these particular cases. Examples of power adjustment and interference is marked.

If a pattern for a θ_i-θ_r pair provides higher reflected power than the corresponding pattern in the database, the new pattern and its associated S-parameter are stored instead in the database for that particular θ_i-θ_r pair. Figure 7 illustrates the collection obtained for gradually increasing the maximum period size. The average of the amplitude of the scattering parameters, considering the largest period size is 0.78. Therefore, one can conclude that increasing the super-cell size makes it possible to direct power in more directions. Furthermore, one can tune the direction of reflection by choosing different patterns from the precalculated database. Continuing this approach for larger super-cell sizes and more control parameter values provides more options, presumably with higher reflection coefficients. Table III compares this approach with other strategies described in the literature relying on different RIS models. In conclusion, this intuition-based method is physically tenable and can handle structures where homogenization is not feasible, requiring higher but manageable complexity.

C. Effect of Finite Size

In practice, an RIS structure is finite in size. In turn, we also investigate the effect of this condition, which is studied analytically in the literature, see, e.g. [19]. To demonstrate this effect, we consider a six unit cell large structure, remove the periodic boundary conditions, and terminate the simulation domain with a perfectly matched layer. For the sake of simplicity, we select control parameters with a two unit cell period size and repeat it three times to obtain the patterns of the six unit cell large RIS. In this way, the pattern still has some periodicity, supporting the comparison with the periodic case. Afterward, the directivity [23] as the function of the reflection angle ($D\left(\phi_\mathrm{r}\right)$) is evaluated in the far-field.

Since a pattern consisting of two unit cells is repeated three times to obtain a six unit cell large structure, the reflection angles corresponding to each propagating mode are computed using (7) with $p_\mathrm{x} = 62.8\,\mathrm{mm}$, i.e., the size of two unit cells of the considered design along the X-axis, see Tab. I.

To compare the results of the infinite periodic structure with their finite counterparts, the directions corresponding to the propagating modes are plotted on the $D\left(\phi_\mathrm{r}\right)$ curves; see Fig. 8. For large incidence angles (θ_i), there is a notable difference between the peaks of $D\left(\phi_\mathrm{r}\right)$ and the directions of the propagating modes. However, as the angle of incidence decreases, the peaks approach the directions indicated by the propagating modes. The effect of varying the control parameters is also apparent. For uniform configurations, specular reflection dominates, i.e., the $0, 0$ mode. By appropriately changing the control pattern, anomalous reflections can be achieved; see the curves corresponding to $C_1 = 0.353\,\mathrm{pF}, C_2 = 0.5\,\mathrm{pF}$ and $C_1 = 0.5\,\mathrm{pF}, C_2 = 0.353\,\mathrm{pF}$ in Fig. 8. From the authors' perspective, these results indicate that it might be beneficial to consider infinite periodic structures at the pattern design phase because it simplifies the simulation. However, one should keep in mind that for a finite-size RIS the directions corresponding to the highest radiated power might deviate from the directions in the case of an infinite periodic structure.

V. CONCLUSION

We briefly introduce periodic structures essential for understanding the full-wave simulation results. Then, a comprehensive numerical analysis of the scattering from the utilized RIS design is presented. The analyzed structure is on a length scale, where homogenization-based modeling is not feasible. Full-wave simulations were employed to address this issue. Our main contribution is the approach that enables the identification of RIS configuration patterns creating anomalous reflection. In particular, the described method is physically meaningful due to the utilization of computational electromagnetics. With this approach, a database of configuration patterns can be obtained, each targeting a different reflection direction and a low level of parasitic reflection. An RIS control algorithm can be designed based on such a database, which chooses the best control pattern from the control patterns constructed offline. The primary limitation of this approach is its run-time and computation resource demand. As the number of control parameters grows, the number of simulations also increases. This issue can be partially addressed through parallelization, which is available in most commercial full-wave solvers. Another potential solution is to limit the period size of the pattern and repeat it along the surface. We also present the following effects described in the literature. First, reflection in unintended directions might lead to implementation challenges. Second, because of the finite size of the RIS, scattering from the surface can not be described with the superposition of Floquet modes; instead, for example, the directivity can be utilized. It is noteworthy that in the evaluated scenario, the directivity peaks divert from the directions corresponding

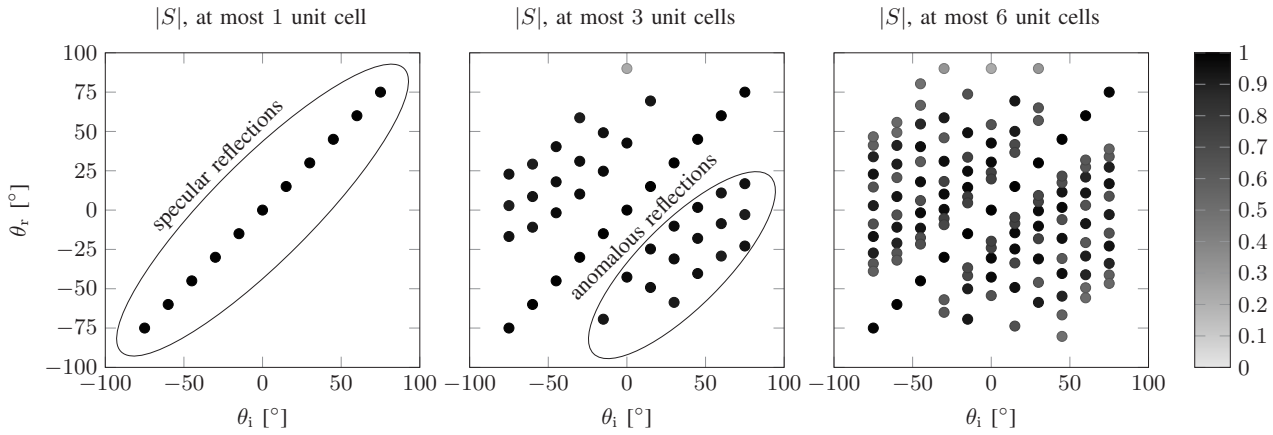


Fig. 7. Possible reflection directions, $\phi_\mathrm{i} = 0\,^\circ$, $f = 4.7\,\mathrm{GHz}$. The excitation is TM polarized. For the results, both reflected polarizations are considered.

TABLE III
COMPARING DIFFERENT RIS MODELING AND CORRESPONDING PATTERN DESIGN APPROACHES

Ref.	Methodology	Objective	Strength	Limitation
[11], [17]	The effect of scatterers is represented by complex numbers in a diagonal or dense matrix.	Typically, the goal is to maximize the throughput of the RIS-aided link.	This model integrates well into the stochastic models of unstructured rich scattering channels; thus, it is possible to consider the environment.	Such models often mask fundamental physical mechanisms.
[18]	The interaction of the RIS elements is accounted for using a multiport network theory model.	Synthesize a desired radiation pattern by tuning the impedance values describing the RIS elements.	It could be more accurate than [11] and it is still suitable for numerous evaluations, e.g., in optimization loops.	Network theory-based equivalent models are usually valid for a certain frequency range.
[6], [12], [19]	The RIS is replaced by an equivalent homogenized boundary condition.	Synthesize a desired radiation pattern by creating a spatially varying boundary condition.	This approach captures all the essential physical phenomena, provided that the conditions required for homogenization are fulfilled.	It can be cumbersome to establish the correspondence of the physical structure and the parameters of the homogenized BC.
[13], [14]	An integral equation-based approach is applied. The metallic pattern is considered with an impedance type BC.	Tune the impedance BC to achieve a target radiation pattern.	It accurately captures the interaction between each part of the RIS.	Adding other structures to the model, such as vias or circuit components might be challenging.
This work	Full-wave simulation is applied.	Design a control pattern dictionary for the angle of incident and reflection pairs.	This approach captures all the physical phenomena and can be applied to structures being large relative to the wavelength.	The simulation demands more computational resources than other approaches.

to the propagating modes of the infinite case only for large incidence angles. Therefore, the results of the periodic case can serve as a reliable guideline when designing RIS patterns. Building on the presented results, a possible future direction is to apply sensitivity analysis to identify the capacitance set from which the patterns are constructed.

REFERENCES

[1] M. D. Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020, DOI: 10.1109/JSCA.2020.3007211.

[2] M. D. Renzo, F. H. Danufane, and S. Tretyakov, "Communication models for reconfigurable intelligent surfaces: From surface electromagnetics to wireless networks optimization," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1164–1209, Sep. 2022, DOI: 10.1109/JPROC.2022.3195536.

[3] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627–8640, Dec. 2021, DOI: 10.1109/TCOMM.2021.3116151.

[4] L. Dai, B. Wang, M. Wang, X. Yang, J. Tan, S. Bi, S. Xu, F. Yang, Z. Chen, M. D. Renzo, C.-B. Chae, and L. Hanzo, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45 913–45 923, Mar. 2020, DOI: 10.1109/ACCESS.2020.2977772.

[5] O. H. Karabey, A. Gaebler, S. Strunck, and R. Jakoby, "A 2-D electronically steered phased-array antenna with 2×2 elements in LC display technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 5, pp. 1297–1306, May 2012, DOI: 10.1109/TMTT.2012.2187919.

[6] E. Martini and S. Maci, "Theory, analysis, and design of metasurfaces for smart radio environments," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1227–1243, Sep. 2022, DOI: 10.1109/JPROC.2022.3171921.

[7] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, and D. R. Smith, "An overview of the theory and applications of metasurfaces: The two-dimensional equivalents of metamaterials," *IEEE Antennas and Propagation Magazine*, vol. 54, no. 2, pp. 10–35, Apr. 2012, DOI: 10.1109/MAP.2012.6230714.

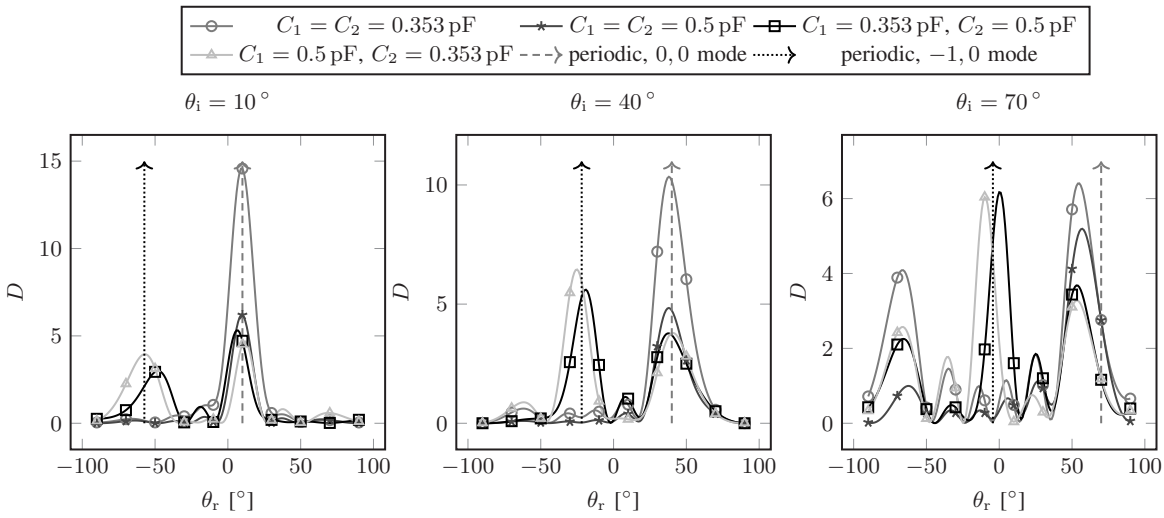


Fig. 8. Directivity of the finite RIS structure, $\phi_\mathrm{i} = 0\,^\circ$, $f = 4.7\,\mathrm{GHz}$. Arrows are normalized to match D.

[8] M. H. Jwair and T. A. Elwi, "Metasurface antenna circuitry for 5G communication networks," *Infocommunications Journal*, vol. 15, no. 2, pp. 2–7, 2023, DOI: 10.36244/ICJ.2023.2.1.

[10] V. Degli-Esposti, E. M. Vitucci, M. D. Renzo, and S. A. Tretyakov, "Reradiation and scattering from a reconfigurable intelligent surface: A general macroscopic model," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 10, pp. 8691–8706, Oct. 2022, DOI: 10.1109/TAP.2022.3149660.

[11] E. Björnson, H. Wymeersch, B. Matthiesen, P. Popovski, L. Sanguinetti, and E. de Carvalho, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 135–158, Mar. 2022, DOI: 10.1109/MSP.2021.3130549.

[12] A. Díaz-Rubio, V. S. Asadchy, A. Elsakka, and S. A. Tretyakov, "From the generalized reflection law to the realization of perfect anomalous reflectors," *Science Advances*, vol. 3, no. 8, Aug. 2017, DOI: 10.1126/SCIADV.1602714.

[13] G. Xu, V. G. Ataloglou, S. V. Hum, and G. V. Eleftheriades, "Extreme beam-forming with impedance metasurfaces featuring embedded sources and auxiliary surface wave optimization," *IEEE Access*, vol. 10, pp. 28 670–28 684, 2022, DOI: 10.1109/ACCESS.2022.3157291.

[14] V. G. Ataloglou, T. Qiu, and G. V. Eleftheriades, "Design of dual-polarization reflective impedance metasurfaces for 3D beam-shaping," in *2024 IEEE International Symposium on Antennas and Propagation and INC/USNC-URSI Radio Science Meeting*. Firenze, Italy: IEEE, Jul. 2024, pp. 1501–1502, DOI: 10.1109/AP-S/INC-USNC-URSI52054.2024.10686554.

[15] B. T. Csathó, Z. Badics, J. Pávó, and B. P. Horváth, "Accelerated 3-D analysis of metasurfaces for RIS applications by characteristic cell functions," *IEEE Transactions on Magnetics*, vol. 60, no. 12, pp. 1–4, Dec. 2024, DOI: 10.1109/TMAG.2024.3480354.

[16] Y. Liu and C. D. Sarris, "Efficient propagation modeling for communication channels with reconfigurable intelligent surfaces," *IEEE Antennas and Wireless Propagation Letters*, vol. 21, no. 10, pp. 2120–2124, Oct. 2022, DOI: 10.1109/LAWP.2022.3192269.

[17] Özlem Tuğfe Demir and E. Björnson, "Wideband channel capacity maximization with beyond diagonal RIS reflection matrices," *IEEE Wireless Communications Letters*, vol. 13, no. 10, pp. 2687–2691, Oct. 2024, DOI: 10.1109/LWC.2024.3439489.

[18] A. Abrardo, A. Toccafondi, and M. D. Renzo, "Design of reconfigurable intelligent surfaces by using S-parameter multiport network theory — Optimization and full-wave validation," *IEEE Transactions on Wireless Communications*, vol. 23, no. 11, pp. 17 084–17 102, Nov. 2024, DOI: 10.1109/TWC.2024.3450722.

[19] A. Diaz-Rubio and S. A. Tretyakov, "Macroscopic modeling of anomalously reflecting metasurfaces: Angular response and far-field scattering," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 10, pp. 6560–6571, Oct. 2021, DOI: 10.1109/TAP.2021.3076267.

[20] A. K. Bhattacharyya, *Phased Array Antennas: Floquet Analysis, Synthesis, BFNs, and Active Array Systems*. Hoboken, NJ, USA: Wiley, 2006.

[21] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*. Hoboken, NJ, USA: Wiley, 2003.

[22] C. A. Balanis, *Advanced Engineering Electromagnetics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2013.

[23] ANSYS, Inc., ANSYS® Academic Research High-Frequency Structure Simulator, Release 21.2, *HFSS Help*, Jul. 2021.

[24] C. Yepes, M. Faenzi, S. Maci, and E. Martini, "Perfect non-specular reflection with polarization control by using a locally passive metasurface sheet on a grounded dielectric slab," *Applied Physics Letters*, vol. 118, no. 23, Jun. 2021, DOI: 10.1063/5.0048970.

Botond Tamás Csathó is currently pursuing his Ph.D. at the Budapest University of Technology and Economics. He received his M.Sc. in 2021 from the same institute. His research interests include electromagnetic modeling of metasurfaces and signal processing for multiple antenna communications.

Bálint Péter Horváth received his M.Sc. 2013 and Ph.D. 2018 degrees in Electrical Engineering from the Budapest University of Technology and Economics, where he is currently a senior lecturer at the Department of Broadband Infocommunications and Electromagnetic Theory. His research interests include signal processing in communications systems, software defined radio, electromagnetic simulations and computational model validation of wireless devices.

Quantum Network Security: A Quantum Firewall Approach

Shahad A. Hussein[1], Suadad S. Mahdi[1] and Alharith A. Abdullah[1]

Abstract—**The increasing prominence of quantum networks has necessitated the exploration of their vulnerabilities and the development of effective countermeasures. This paper investigates the potential threats faced by quantum networks, particularly focusing on the exploitation of quantum TCP-three-way handshake connections. To mitigate these attacks, a novel approach involving the implementation of a quantum firewall is proposed. The paper emphasizes that the security of quantum networks is primarily reliant on pre-established agreements for creating quantum entanglement among devices, which inherently limits external attacks. However, it highlights the adverse impact of quantum assaults on network availability due to the consumption of quantum bits required for establishing connections. By leveraging unique node identification and coherence time of quantum memory, the proposed quantum firewall effectively mitigates the effects of attacks while ensuring network availability. Through this security strategy, the paper demonstrates the robustness of the quantum firewall in safeguarding the integrity and operation of quantum networks against potential threats.**

Index Terms—**Quantum Internet, Quantum Repeater, Quantum Attack, Quantum Firewall**

I. INTRODUCTION

The rapid evolution of quantum technology has led to the emergence of quantum networks as a vital component of modern communication systems [1]. As these networks become increasingly complex and critical, ensuring their security against potential threats has become of paramount importance. Consequently, recent research has dedicated significant attention to examining the security aspects of quantum networks, as evidenced by numerous articles in the field.

Quantum networks offer new avenues for secure and efficient information transfer by leveraging the principles of quantum mechanics [2]. However, these networks also introduce unique challenges and vulnerabilities that must be thoroughly understood and effectively addressed. Researchers have directed their efforts toward investigating the security implications of quantum networks, exploring potential threats, and devising strategies to mitigate them.

The exploration of security in quantum networks spans various dimensions, encompassing the protection of quantum communication protocols and the resilience of network infrastructure [3]. Researchers have developed novel techniques and frameworks to ensure the confidentiality, integrity, and availability of quantum information transmitted over these networks, aiming to establish a robust foundation for their secure operation.

In the context of this research, one notable paper [4] introduces a quantum intrusion detection system (QIDS) that combines conventional and quantum approaches to effectively protect systems against sophisticated attacks. The QIDS enhances accuracy, precision, and reduces false positives, with evaluations conducted using Distributed Denial of Service (DDoS) assaults generated by Mirai botnets.

Despite advancements in quantum technology, the practicality of a quantum internet is limited due to the constraint of point-to-point qubit transmission. Overcoming this limitation necessitates the implementation of quantum routers. Notably, studies described in [5][6] present quantum router designs that leverage teleportation and protocols for managing entangled pairs, with validation performed using quantum simulators.

Furthermore, [7] focuses on attacks targeting quantum repeaters, which are akin to traditional Internet routers, evaluating their vulnerabilities in terms of secrecy, integrity, and availability. The authors develop a framework for exploring network-wide vulnerabilities, emphasizing the role of classical computing and networking aspects in addressing the overall security concerns of quantum networks.

In [8], classical network theory and graph theory are employed to address security and key management challenges in quantum networks. The proposed communication architecture prioritizes high security by reducing the number of intermediary nodes. Additionally, key management and data scheduling algorithms enhance data transmission efficiency.

The paper outlined in [9] explores denial of service (DoS) attacks against actual quantum key distribution (QKD) equipment, where attackers deplete the QKD key reserves of the Key Management System (KMS). The authors propose safety measures to mitigate such attacks and underscore the significance of integrating QKD into standard telecommunications networks for ensuring communication security. They also emphasize the importance of treating QKD keys as valuable and scarce resources.

This paper delves into the security aspects of quantum networks, specifically focusing on the vulnerabilities associated with quantum TCP-three-way handshake connections

Submitted on 2024.11.22
[1] University of Babylon, Babil, Iraq;
(E-mail: {shahad.alshamary, suadadsafaa, alharith}@uobabylon.edu.iq)

DOI: 10.36244/ICJ.2025.1.6

[10][11][12]. It also elucidates a strategy for attacking quantum networks, shedding light on potential risks associated with the utilization of quantum communication protocols. The disruption caused by attackers exploiting quantum TCP-three-way handshake connections raises significant concerns. In response to this identified weakness, our research proposes and implements a novel solution: a quantum firewall.

Thus, the primary objectives of this paper are twofold: first, to uncover weaknesses in quantum networks that can be exploited through organized assaults, and second, to establish a quantum firewall as a proactive defense mechanism against such threats. Our findings highlight the importance of establishing quantum entanglement between devices and emphasize the vital role of pre-established agreements in safeguarding against external attacks [13].

By addressing these objectives, our paper contributes to the overall understanding and enhancement of quantum network security. We aim to uncover vulnerabilities, propose effective countermeasures, and provide insights for the development of secure and resilient quantum networks in an increasingly interconnected world.

II. THE PRELIMINARIES

A. The Quantum Network

Quantum networks represent an innovation in information processing and communication, utilizing quantum mechanics principles to provide capabilities previously unavailable to conventional networks [14]. Before delving into the details of this research and innovating a quantum firewall, it is significant to understand some essential concepts in the field of quantum networks which include the following:

a) Quantum bits, or qubits, differ considerably from classical bits, the qubits exist in multiple states at the same time due to the property of quantum superposition [15]. This characteristic significantly improves quantum networks' information processing capability.

Quantum entanglement is the basic idea behind quantum networks. Which is phenomenon shows the existence of a unique relationship between quantum particles. This principle was proposed by the scientist Einstein, who pointed out the existence of a shared state of two particles. Where the two particles are affected together, even if the action occurs on only one of the them. In addition, this type of quantum correlation holds regardless of the distance between the particles [1][13]. Moreover because of this quantum interconnection, measuring the state of one particle make it possible to obtain the state of the other also. Thus, leads to breaking the entangled system. Therefore, quantum entanglement has become a major resource for secure quantum communication [14].

Although the quantum internet depends on the entanglement swapping of entangled pairs of qubits (experiments showed that there are four pairs of entangled qubits, and each pair has two qubits called Bell states that can be represented in equations 1, 2, 3 and 4). The pairs resulting from entanglement swapping cannot be in independent states. In other words, the pairs resulting from this process are in an entangled state. In addition, the state of one pair cannot be separated into two independent states. Moreover, it depends on the measurement result. This is due to quantum mechanics, which is subject to the principle of probability. However, the state of each entangled pair is known after the completion of the quantum swap. At last, the behavior of quantum networks is affected by this concept, which is essential to several aspects such as the Quantum Key Distribution (QKD) and repeaters [16][17].

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (1)$$

$$|\beta_{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \quad (2)$$

$$|\beta_{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \quad (3)$$

$$|\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \quad (4)$$

c) Quantum information is sensitive as well as vulnerable to decoherence [18]. Quantum memories (QM), also known as quantum optical memories, are essential components for storing and retrieving quantum states. They are considered the main component responsible for storing a large amount of quantum information in quantum bits of both entangled and non-entangled types, which represent the core of communication in the quantum internet. Furthermore, the lifetime of quantum memories is very short as it depends on the coherence time, which represents the amount of time a quantum system maintains its precise superposition and is crucial for quantum networks to function properly [19], as well as the quantum measuring device such as Bell State Measurement (BSM), and entanglement generator are considered components of quantum memories. Quantum memories have different types, the most famous of which are nitrogen vacuum centers (NV) QM and trapped ionic QM.

d) Quantum nodes represent the heart of the quantum internet. Where, long-range communications between the sender and the receiver are carried out through these nodes. This is done by the entanglement swapping between the quantum memories present in these nodes. Although quantum nodes are subject to quantum laws quantum laws in the process of transferring information, they use the classical internet to exchange control messages between the nodes of the quantum network [12]. In addition, these quantum nodes communicate with each other through quantum and classical channels, like optical fibers or free space[1].

e) Quantum Key Distribution (QKD) referred to the methods employ by quantum network for secure communication. QKD uses quantum features to allow two parties to create a secret key that is secure against attacks like eavesdropping [20] [21].

f) Quantum teleportation a property distinguishing quantum networks that enables the transport of quantum information from one point to another without any physical movement of particles [22]. This mechanism influences the scalability and long-distance communication capacities of quantum networks [23].

All of these concepts establish a structure for understanding the details and constraints of using quantum physics to effectively and securely communicate information.

B. Quantum Transmission Control Protocol (QTCP)

QTCP is the quantum version of TCP, allowing nodes in quantum networks to communicate in a reliable and orderly

manner by using quantum three-way handshake process as shown Fig.1 [12]. QTCP uses qubits, allowing several states to exist at the same time by leveraging quantum superposition principles. This distinguishing feature enables more efficient and complicated transfer of data between quantum nodes, hence improving the overall performance of quantum communication systems.

Furthermore, QTCP uses quantum entanglement to build reliable links between nodes, using the inherent correlations between entangled particles to enable instantaneous and secure data transfer. Despite this, QTCP poses new security issues, such as the vulnerability to eavesdropping and the possibility of quantum state modification during transmission [12].

Additionally, to address these weaknesses and preserve the integrity and secrecy of quantum communication, some security solutions including are employed like intrusion prevention, detection systems and firewall. From point of the quantum view the firewall uses quantum mechanics concepts to create strong defensive mechanisms against possible attacks on Quantum TCP connections, hence protecting the integrity and security of quantum networks.

Understanding the complexities of Quantum TCP is critical for understanding the unique difficulties and solutions in quantum network security, emphasizing the need for more research and development in this quickly expanding sector.

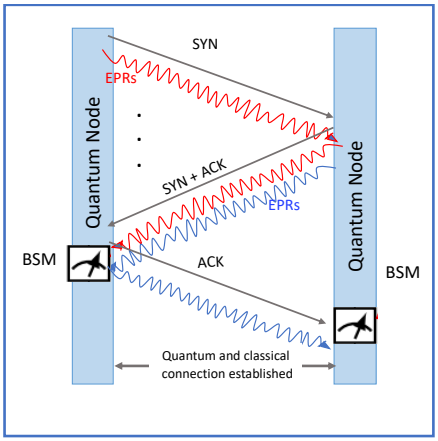


Fig. 1. The Three-way Handshek Process of QTCP [10]

C. *Distributed Denial of Service (DDoS) attacks*

In traditional networks, overloading a target with a flood of traffic from several sources is known as DDoS, which leaves services unreachable [19]. On the other hand, the introduction of Quantum DDoS in the quantum computing age has created a new danger scenario. Quantum DDoS takes advantage of quantum mechanics principles to impair the availability and operation of quantum networks, bringing the classic DDoS idea into the quantum domain. The new attack presents distinct obstacles that require detailed knowledge to design effective responses.

Quantum networks, although providing unparalleled benefits, can pose new vulnerabilities. These weaknesses are exploited by quantum DDoS attacks, which target the quantum channels and memory that contain the fundamental units of quantum information (qubits) as well as required for communication between quantum nodes, where the qubits and channels are attractive targets for attackers looking to disrupt network functioning.

Additionally, these attacks impair the coherence required to create and suffer quantum connections inside the network. Each attacking node adds to the decline of quantum channels, causing a cascade effect that compromises the overall availability and dependability of the quantum network.

Furthermore, unlike conventional networks, where external sources might launch DDoS attacks, quantum networks rely on previous agreements to achieve quantum entanglement between devices. Thus, Quantum DDoS attacks cannot come from external sources.

D. *Firewall Concept*

Before getting into the details of the proposed Quantum Firewall concept, it's important to have an understanding of the general idea of a firewall within the area of cyber security in traditional computing. A firewall is a key network security hardware or software that monitors, filters, and controls incoming and outgoing network traffic based on established security rules. It serves as a firewall between a trusted internal network and untrustworthy external networks, such as the Internet, thereby limiting unwanted access and possible cyber risks [24].

Firewalls are classified into various varieties, each with its own set of features and processes for detection and managing network traffic [23]. Packet-filtering firewalls, for example, Proxy firewalls, and another form, stateful inspection firewalls, also known as dynamic packet filtering, each of which is used depending on the security requirements of the network in which it is used. where the security policies that implement by each type of firewall specify how they should handle various forms of network traffic, enforce access controls and prevent unwanted activity. Furthermore, modern firewalls frequently include intrusion prevention and detection technologies to help identify and respond to possible security threats [24].

Understanding the fundamental ideas of firewalls in classical computing gives an ideal starting point for investigating novel methods of network security, including its use in protecting networks in emerging quantum computing such as the quantum firewall illustrated in Fig 2.

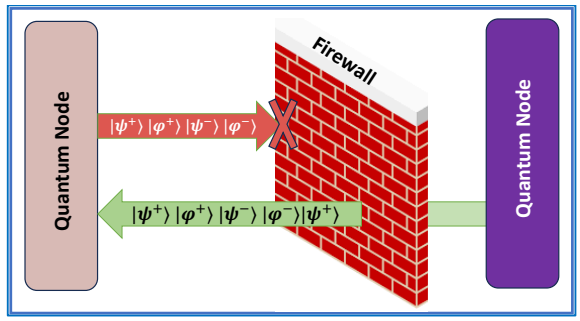


Fig. 2. The Quantum Firewall Concept

III. PROPOSED MECHANISM

The flow work of proposed mechanism outlines a comprehensive system designed to enhance efficiency and security as shown in Fig.3

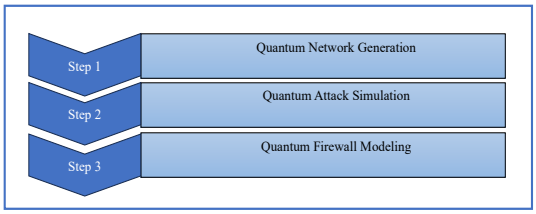


Fig. 3. The Flow Work of the Proposed System

A. Step1: Quantum Network Generation

The Python programming language v3.9, PyCharm Community 2021.3 environment, and a Mac operating system with the Apple M1 processor and 8GB of memory were used to construct the quantum network. The proposed quantum network comprises multiple nodes, each equipped with a quantum memory that has a limited capacity for storing quantum bits. The following factors were considered:

- The Q-network is composed of N nodes, Where the attack-nodes n are chosen according to the equation 5

$$n = round\ (\tfrac{1}{4}\ N) \qquad (5)$$

- The quantum memory has maximum capacity (Max_Cap) ranges from 3 to 9 qubits.
- The nodes are interconnected in a mesh network architecture.
- It is not possible for the quantum channels/links to share a single qubit. A visual representation of the network can be seen in Fig. 4.
- Each node has unique id as (1,2,3 .....n)

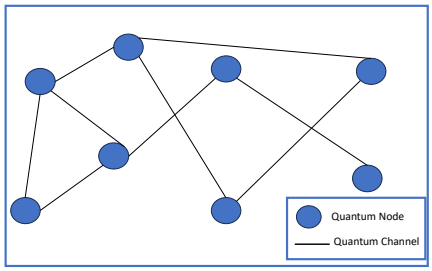


Fig. 4. The Proposed Netwok Architecture

B. Step2: Quantum Attack Simulation

The QTCP protocol's three-way handshake process is vulnerable to a quantum attack that exploits this weakness. The attack involves using EPR to establish Quantum SYN Flooding, as described in the following:

- Each fake node first calculates the number of connections with each neighbor, i.e., check the unbounded channels UC.
- The fake node then prepares a number of local entangled states EPRs in order to begin the quantum three-way handshake process (Connection establishment) according to equation 6

$$EPR_s = UC * Max\_Cap \qquad (6)$$

Moreover, the prepared states are formed as in equation 7 :

$$|\psi^+\rangle_{AiBi}, |\psi^-\rangle_{AiBi}, |\varphi^+\rangle_{AiBi}, |\varphi^-\rangle_{AiBi} \qquad (7)$$

Where the state is represented in two digits that are the node IDs A *and* B and i is an integer from 0 onwards.

- At this stage, the deceptive nodes send qTCP packets (quantum synchronization -qSYN- request) to all nearby nodes via all available channels, in order to occupy the entire channel. Additionally, consists of the sending node's identifier and the second qubit of the entangled bits generated, which represents a qubit stored in the quantum memory of the neighboring node. Nonetheless, the entanglement process remains unfinished as the deceptive node does not complete the three-way handshake process, resulting in half-open communications. the qSYN request is shown in Fig. 5.

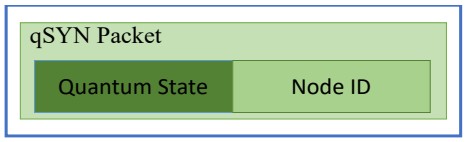


Fig. 5. The Quantum Synchronization Request Packet

- After the coherence time expires, the deceptive nodes resend the qSYN packet again (making qSYN flood attack) as it illustrated in Fig. 6, and it remains in this state repeatedly, consuming the quantum memory of neighboring nodes in addition to the channels between each pair of nodes.

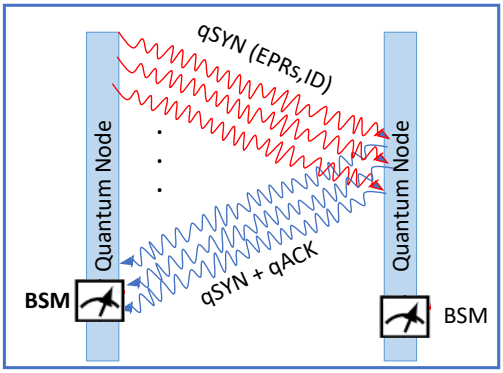


Fig. 6. The Quantum Synchronization Flooding

The flowchart of the quantum-TCP attack is displayed in Fig. 7.

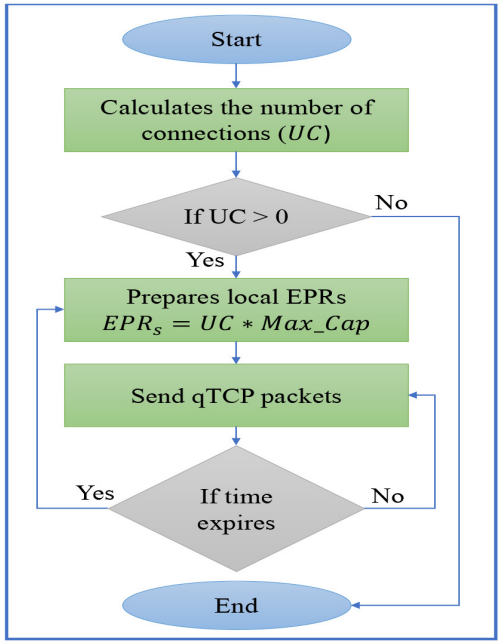


Fig. 7. The Quantum-TCP attack Flowchart

C. Step3: Quantum Firewall Modeling

The quantum firewall is software that runs on all quantum nodes in the network to protect them from quantum communications that may threaten the security and availability of the network. The proposed firewall works as follows:

- We assume that the firewall contains a database (DB) in which the received quantum packet information is recorded.
- The firewall records the time the quantum packet reaches the recipient.
- After that, it extracts the identifier from the quantum packet by performing the quantum measurement process, without sending back the measurement result during which the formation of the quantum link is completed.
- The extracted ID is compared with the IDs stored in the database. If there are no records for this packet, the firewall allows the quantum communication process to complete by performing the measurement process. However, if the number of requests (RC) is more than the threshold specified after the coherence time expires for the quantum memories (determined by comparing the information extracted from the quantum packet and the records belonging to the firewall), then the firewall blocks this node and all incoming requests as the flowchart in Fig. 8 illustrates.

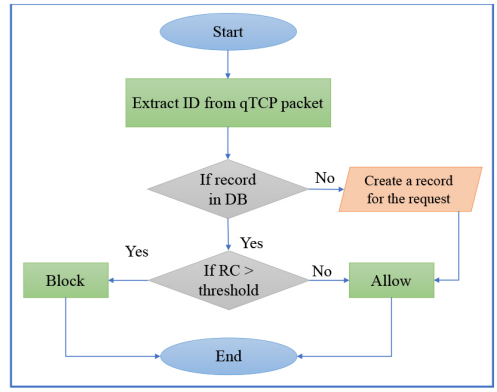


Fig. 8. The Quantum-Firewall Flowchart

IV. Result and Discussion

A. The First Case: All Nodes are Normal

In this case, all quantum nodes are functioning normally and in accordance with the settings prepared for the quantum devices. Every two neighboring nodes can perform quantum entanglement between their respective quantum memories, forming a quantum link between them based on the probability of generating quantum bits in each node as shown in Fig. 9. The entanglement process is completed through a quantum three-way handshake among the generated qubits with the highest probability.

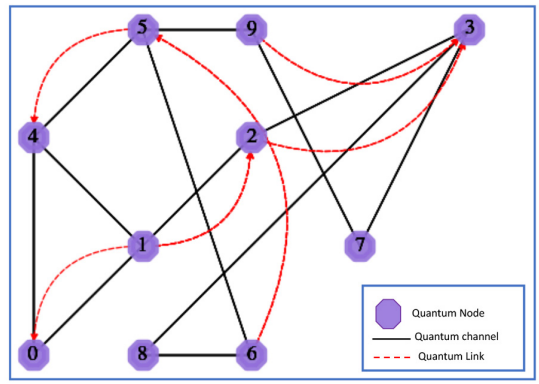


Fig. 9. Establishing Quantum Links Between the Quantum Nodes

B. The Second Case: There is an Attack on the Network in The Absence of a Firewall

The network in this case comprises of 10 quantum nodes, out of which three have been designated as deceptive nodes while the others are normal nodes. These three nodes are identified as 1, 5, and 7, and each node is connected to its neighboring nodes. Table 1 presents the results of the attack on the network, including the number of connections of each node and the number of pairs of entangled quantum bits (EPR) that are prepared at each attacking node. It has been observed that the number of these pairs is equal to or greater than the normal limit for each normal node.

TABLE I
THE COUNT OF CONNECTIONS AND EPRS GENERATED BY DECEITFUL NODES

Deceptive Nodes	Connections	EPRs
1	3	9
5	3	9
7	2	6

Additionally, Table 2 illustrates the encoding of the pairs of quantum bits (Entangled bits) that are prepared at each attacking node with the intention of sending one individual from each pair to the neighboring nodes to which it is connected.

For ease of work, the entanglement pair between two nodes was encoded with a symbol containing the node number and the qubit number. For example, there are 3 qubits generated between node 7 and node 3, and therefore the encoding of these qubits was as follows: |7030>, |7131>, |7232>. Where 7, and 3 represent the number of the two nodes, while 0, 1, 2 represent three different qubits, and so on for the rest of the symbols.

TABLE II
THE ENCODING OF THE PREPARED ENTANGLED BITS PAIRS

Deceptive Nodes	Neighboring Nodes	EPRs Encoding
1	0	|1000>, |1101>, |1202>
	2	|1320>, |1421>, |1522>
	4	|1640>, |1741>, |1842>
5	4	|5040>, |5141>, |5242>
	6	|5360>, |5461>, |5562>
	9	|5690>, |5791>, |5892>
7	3	|7030>, |7131>, |7232>
	9	|7390>, |7491>, |7592>

Moreover, Table 3 reflects the behavior of the quantum attack, listing the details of the connection establishment request packets when the scenario is run for five minutes. This table shows the stability of the packet type, represented by "q-SYN". Additionally, during each second, each deceptive node sends packets containing its identifier and one member of the pairs of the prepared entangled bits for all connected nodes simultaneously.

TABLE III
THE QUANTUM ATTACK BEHAVIOR

First coherent time			
Deceptive Nodes	1	5	7
Request Packets	(q-SYN,|00>,1) (q-SYN,|01>,1) (q-SYN,|02>,1) (q-SYN,|20>,1) (q-SYN,|21>,1) (q-SYN,|22>,1) (q-SYN,|40>,1) (q-SYN,|41>,1) (q-SYN,|42>,1)	(q-SYN,|40>,5) (q-SYN,|41>,5) (q-SYN,|42>,5) (q-SYN,|60>,5) (q-SYN,|61>,5) (q-SYN,|62>,5) (q-SYN,|90>,5) (q-SYN,|91>,5) (q-SYN,|92>,5)	(q-SYN,|30>,7) (q-SYN,|31>,7) (q-SYN,|32>,7) (q-SYN,|90>,7) (q-SYN,|91>,7) (q-SYN,|92>,7)
Second coherent time			
Deceptive Nodes	1	5	7
Request Packets	(q-SYN,|00>,1) (q-SYN,|01>,1) (q-SYN,|02>,1) (q-SYN,|20>,1) (q-SYN,|21>,1) (q-SYN,|22>,1) (q-SYN,|40>,1) (q-SYN,|41>,1) (q-SYN,|42>,1)	(q-SYN,|40>,5) (q-SYN,|41>,5) (q-SYN,|42>,5) (q-SYN,|60>,5) (q-SYN,|61>,5) (q-SYN,|62>,5) (q-SYN,|90>,5) (q-SYN,|91>,5) (q-SYN,|92>,5)	(q-SYN,|30>,7) (q-SYN,|31>,7) (q-SYN,|32>,7) (q-SYN,|90>,7) (q-SYN,|91>,7) (q-SYN,|92>,7)
Last coherent time			
Deceptive Nodes	1	5	7
Request Packets	(q-SYN,|00>,1) (q-SYN,|01>,1) (q-SYN,|02>,1) (q-SYN,|20>,1) (q-SYN,|21>,1) (q-SYN,|22>,1) (q-SYN,|40>,1) (q-SYN,|41>,1) (q-SYN,|42>,1)	(q-SYN,|40>,5) (q-SYN,|41>,5) (q-SYN,|42>,5) (q-SYN,|60>,5) (q-SYN,|61>,5) (q-SYN,|62>,5) (q-SYN,|90>,5) (q-SYN,|91>,5) (q-SYN,|92>,5)	(q-SYN,|30>,7) (q-SYN,|31>,7) (q-SYN,|32>,7) (q-SYN,|90>,7) (q-SYN,|91>,7) (q-SYN,|92>,7)

However, when the coherence time ends, the deceptive nodes continue to resend packets in the same way. This action occupies the communication channels of all connected nodes and consumes quantum memories by keeping the entanglement process incomplete. This incomplete half-quantum communication disrupts the measurement process and prevents then normal nodes from using quantum channels or memories for other quantum communications, ultimately causing network disruption.

Furthermore, Fig. 10 provides a clearer explanation of the requests that are transmitted from every malicious node to the nodes that are adjacent to it.

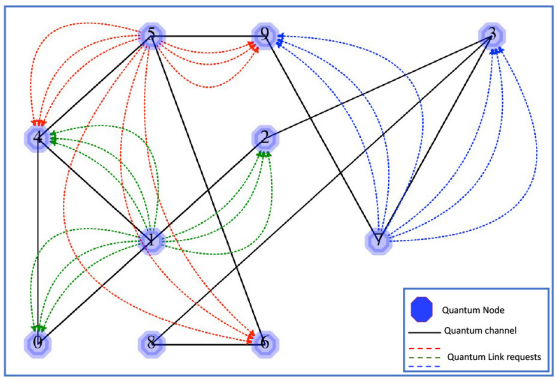


Fig. 10. q-SYN Requests Sent from Malicious Node to Its Neighboring Nodes

C. Third case: There is an attack on the network and the firewall has been activated.

The presence of the firewall in this case mitigates the impact of the DDoS attack, as it examines the received quantum packets (qTCP) and takes countermeasures against the attack. The results of activating the quantum firewall on quantum devices are shown in Table 4. It is clear from this table that the rules applied to the packets allow the quantum packets to be received and passed to the node to conduct quantum communication between the nodes only twice (as a threshold). Accordingly, the quantum firewall prevents nodes from receiving any packets belonging to a specific node when these

requests are repeated twice or more than that immediately after the coherence time expires. That is, when the coherence time ends and deceptive nodes re-send the same packets every time, the firewall in this case will prevent receiving any packets coming from these deceptive nodes.

TABLE IV
THE QUANTUM FIREWALL BEHAVIOR

Deceptive Nodes	1	5	7
Adjacent Nodes	0	4	3
Request Packets	(q-SYN,|00>,1) (q-SYN,|01>,1) (q-SYN,|02>,1)	(q-SYN,|40>,5) (q-SYN,|41>,5) (q-SYN,|42>,5)	(q-SYN,|30>,7) (q-SYN,|31>,7) (q-SYN,|32>,7)
Coherent Time	1	2	More than 2
Count Action	Allow	Allow	Deny

Finally, the firewall effectively decreased the number of requests being sent to the quantum nodes, resulting in maintain the availability of the quantum network.

V . CONCLUSION

This work has clarified two aspects of the quantum network. Firstly, it showed a method of attacking the network by employing the quantum TCP-three-way handshake connections. The other aspect was to defend against this attack or mitigate it by proposing and implementing a quantum firewall. From this work, we can conclude that according to quantum laws, the attack cannot be from outside the network due to the agreements prepared in advance in order to create quantum entanglement between quantum devices. Additionally, we can say that the quantum attack has an impact on the availability of the network, as each node performs an attack, it affects the quantum channels and memories in every node connected to it. This consumption of quantum bits prepared to create quantum links with other nodes causes the network to stop working as this action continues on all devices in the network.

Finally, when implementing a security method, it can be concluded that it can mitigate the impact of the attack and maintain availability. The firewall mainly depends on the ID of each quantum node as well as the coherence time of the quantum memories.

REFERENCES

[1] S. A. Hussein and A. A. Abdullah, "A comprehensive study of the basics of quantum networks," *2022 5th International Conference on Engineering Technology and Its Applications (IICETA)*, May 2022, DOI: 10.1109/IICETA54559.2022.9888324.

[2] L. Zhonghui, X. Kaiping, L. Jian, C. Lutong, L. Ruidong, W. Zhaoying, Y. Nenghai, W. David, S. Qibin, and L. Jun, "Entanglement-Assisted Quantum Networks: Mechanics, Enabling Technologies, Challenges, and Research Directions," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2133–2189, 2023, DOI: 10.1109/COMST.2023.3294240.

[3] K. Adarsh, B. Surbhi, K. Keshav, G. S. Manjula, D. S. Gayathri, D. J. Pacheco, A. Diego and M. Arwa, "Survey of Promising Technologies for Quantum Drones and Networks," in *IEEE Access*, vol. 9, pp. 125 868–125 911, 2021, DOI: 10.1109/ACCESS.2021.3109816

[4] P. Bhattacharyya, H. Sastry, V. Marriboyina and R. Sharma, *Smart and Innovative Trends in Next Generation Computing Technologies*. India: Springer, 2017, DOI: 10.1007/978-981-10-8657-1

[5] B. A. Huberman and B. Lund, "A quantum router for the entangled web," *Information Systems Frontiers*, vol. 22, no. 1, pp. 37–43, Dec. 2019, DOI: 10.1007/s10796-019-09955-5

[6] S. A. Hussein and A. A. Abdullah, "Hybrid routing protocol for quantum network based on classical and quantum routing metrics," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 1, p. 197, Jul. 2023, DOI: 10.11591/ijeecs.v31.i1.

[7] T. Satoh, S. Nagayama, S. Suzuki, T. Matsuo, M. Hajdušek, and R. Van Meter, "Attacking the quantum internet," *IEEE Transactions on Quantum Engineering*, vol. 2, pp. 1–17, Jan. 2021, DOI: 10.1109/TQE.2021.3094983

[8] H. Zhou, K. Lv, L. Huang, and X. Ma, "Quantum Network: security assessment and key management," *IEEE ACM Transactions on Networking*, vol. 30, no. 3, pp. 1328–1339, Jun. 2022, DOI: 10.1109/TNET.2021.3136943

[9] P. Burdiak *et al*., "Use-Case Denial of Service Attack on Actual Quantum Key Distribution Nodes," *In Proceedings of the 9th International Conference on Information Systems Security and Privacy (ICISSP 2023)*, Jan. 2023, DOI: 10.5220/0011672000003405

[10] F.-H. Hsu, Y. L. Hwang, C. Tsai, W. T. Cai, C.-H. Lee, and K. W. Chang, "TRAP: a Three-Way handshake server for TCP connection establishment," *Applied Sciences,* vol. 6, no. 11, p. 358, Nov. 2016, DOI: 10.3390/app6110358

[11] Q.-M. Ma, S. Liu, and X. Wen, "TCP Three-Way Handshake Protocol based on Quantum Entanglement," *Journal of Computers,* pp. 033–040, Oct. 2016, DOI: 10.3966/199115592016102703004

[12] N. Yu, C.-Y. Lai, and L. Zhou, "Protocols for packet Quantum network intercommunication," *IEEE Transactions on Quantum Engineering,* vol. 2, pp. 1–9, Jan. 2021, DOI: 10.1109/TQE.2021.3112594

[13] R. Ursin *et al*., "Entanglement-based quantum communication over 144 km," *Nature Physics*, vol. 3, no. 7, pp. 481–486, Jun. 2007, DOI: 10.1038/nphys629.

[14] Z. Li et al., "Entanglement-Assisted Quantum Networks: mechanics, enabling technologies, challenges, and research directions," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 4, pp. 2133–2189, Jan. 2023, DOI: 10.1109/comst.2023.3294240.

[15] J. Clarke and F. K. Wilhelm, "Superconducting quantum bits," *Nature,* vol. 453, no. 7198, pp. 1031–1042, Jun. 2008, DOI: 10.1038/nature07128.

[16] B. Nordén, "Quantum entanglement: facts and fiction – how wrong was Einstein after all?," *Quarterly Reviews of Biophysics*, vol. 49, Jan. 2016, DOI: 10.1017/s0033583516000111.

[17] R. Demkowicz-Dobrzański, A. Sen, U. Sen, and M. Lewenstein, "Entanglement enhances security in quantum communication," *Physical Review A*, vol. 80, no. 1, Jul. 2009, DOI: 10.1103/physreva.80.012311.

[18] A. Wallucks, I. Marinković, B. J. Hensen, R. Stockill, and S. Gröblacher, "A quantum memory at telecom wavelengths," *Nature Physics*, vol. 16, no. 7, pp. 772–777, May 2020, DOI: 10.1038/s41567-020-0891-z.

[19] H. M. Mohammad and A. A. Abdullah, "DDoS attack mitigation using entropy in SDN-IoT environment," *AIP Conference Proceedings*, Jan. 2023, DOI: 10.1063/5.0123465.

[20] Y. Jassem, and A. Abdulkareem. "Enhancement of quantum key distribution protocol for data security in cloud environment."*Icic International*, vol. 11, no. 3, pp. 279–288, 2020, DOI: 10.24507/icicelb.11.03.279

[21] S. S. Mahdi and A. A. Abdullah, "Enhanced Security of Software-defined Network and Network Slice Through Hybrid Quantum Key Distribution Protocol" *Infocommunications Journal*, vol. 14, no. 3, pp. 9–15, 2022, https://doi.org/10.36244/ICJ.2022.3.2

[22] A. Furusawa and P. Van Loock, *Quantum Teleportation and Entanglement: A hybrid approach to optical quantum information processing*. New York: John Wiley & Sons, 2011. DOI: 10.1002/9783527635283.ch8

[23] A. Singh, K. Dev, H. Šiljak, H. D. Joshi, and M. Magarini, "Quantum Internet—Applications, functionalities, enabling technologies, challenges, and research directions," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 4, pp. 2218–2247, Jan. 2021, DOI: 10.1109/comst.2021.3109944.

[24] J. M. Kizza, *Guide to computernetwork Security*, 6th ed. Berlin: Springer, 2024. DOI: 10.1007/978-3-031-47549-8.

Shahad A. Hussein is currently serving as an Assistant Lecturer at the University of Babylon in the College of Information Technology, located in Babylon, Iraq. She graduated with a Bachelor's degree in Information Technology with excellent grades, ranking first in her class from Babylon University in 2016. In 2017, she was awarded the first place at the national level for the Iraqi Science Day Award for her exceptional graduation research in her field of study. In 2022, she completed her M.Sc. degree from the College of Information Technology at Babylon University. Her current research field focuses on Quantum networks, Network Security, and various aspects related to the future internet.

Suadad S. Mahdi presently serves as a Lecturer at the University of Babylon, specifically within the College of Information Technology in Babylon, Iraq. Her academic journey includes earning a Bachelor's degree in Information Technology from Babylon University in 2016, followed by a Master's degree in Information Networks from the College of Information Technology (IT) in 2020. In 2024, she successfully completed her PhD in Information Networks from the same college. Currently, Suadad's research focuses on a wide range of topics, encompassing future internet, NFV, SDN, network security, cryptography, and quantum cryptography.

Alharith A. Abdullah received his B.S. degree in Electrical Engineering from Military Engineering College, Iraq, in 2000. MSc. degree in Computer Engineering from University of Technology, Iraq, in 2005, and his PhD. in Computer Engineering from Eastern Mediterranean University, Turkey, in 2015. His research interests include Security, Network Security, Cryptography, Quantum Computation and Quantum Cryptography.

Evaluation of traditional and eBPF-based packet processing in Kubernetes for network slicing

Ákos Leiter[†], Döme Matusovits[‡†], and László Bokor[‡§]

***Abstract*—In recent years, the proliferation of cloud-native technology enablers, such as microservice deployment and management with Kubernetes, have presented new challenges for telecommunications service providers. Strict data transmission requirements have emerged in various areas, such as immediate interventions in intelligent transportation, video conferencing, etc. With the advent of 5G networks, this demand can also be fulfilled thanks to an innovative technology called Network Slicing. In terms of its operation, we can separate networks into individual segments to continuously satisfy the desired service requirements. However, packet processing on top of Kubernetes may need to be changed to support the emerging number of microservices during slicing. This is where the Extended Berkeley Packet Filter (eBPF) comes into the picture to boost the capacity of data centers and keep service guarantees. This paper presents how eBPF can support network slicing through its performance evaluation in a Kubernetes environment.**

***Index Terms*—network slicing, eBPF, Kubernetes, 5G/6G mobile cellular architectures, cloud-native applications**

I. Introduction

Implementing end-to-end (E2E) network slicing is still a heavily researched problem in the telco industry. It is impossible to properly fulfill E2E network slicing requirements if one segment or domain of the network does not deal with service guarantees. For example, if the core network part of the E2E network slice instance has proper resource assignment and implementation, other network parts have to act similarly. Radio, transport, and data center networking must also be prepared for network slicing requirements. The end-to-end network slicing concept is depicted in Figure 1. All the network function elements are considered to run in the cloud environment.

In this work, we focus on data center networking, especially Kubernetes-based packet processing solutions, and whether or not they can ensure a particular service level that requires low latency and a guaranteed throughput during packet traversal. In the fifth-generation mobile network standards, communication with these constraints is called Ultra-reliable Low-latency Communication service (URLLC) [1]. We assume that the number of microservices (Kubernetes Services) will be continuously increasing (due to autoscaling, edge deployments, and further radio cloudification and decoupling [2] [3] [4] [5]), so we examine how this will affect network Key Performance Indicators (KPI) considering throughput and latency. We evaluate this on two test environments: the traditional Kubernetes packet processing method kube-proxy-based on the top of iptables and the extended Berkley Packet Filter (eBPF) [6] using the Cilium Container Networking Interface (CNI) [7] in the context of network slicing.

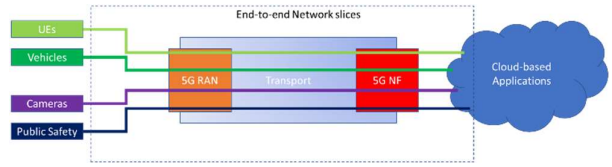


Figure 1 – End-to-end network slicing concept [8]

eBPF is powerful because it makes the Linux operating system programmable without modifying the kernel's source code or writing a new kernel module. Consequently, it also solves the complication of developing a monolithic Kernel, which saves much time when adding new features to the OS core. Furthermore, it provides an alternative to low-performed Netfilter-based packet processing. This is why Kubernetes has started to utilize eBPF in CNIs, which are responsible for Kubernetes' internal and external networking. This includes interface and IP address management and packet processing mechanisms. There are two CNIs publicly available that implement eBPF-based networking: Cilium and Calico [9]. Our test system relies on the Cilium-based solution.

This paper is organized as follows: Section II gives a technological introduction to iptables and eBPF. Then, Section III covers the most crucial technological background of Kubernetes and Cilium. The related work is presented in Section IV. The testbed details are explained in Section V, while measurement results are elaborated in Section VI. Conclusion and future work are drawn in Sections VII and VIII, respectively.

II. Background behind iptables and Extended Berkeley Packet Filter

Although in the Linux world, packet processing/filtering technologies are already available (such as nftables [10]), which can enhance the traditional Netfilter based approach, in Kubernetes, the iptables is still the most widely used option. It is developed under the Netfilter project [11], which consists of community-driven collaboratives. It is built up with

[†] Nokia Bell Labs Budapest, Bókay János u. 36-42, 1083 Hungary
[‡] Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[§] HUN-REN-BME Cloud Applications Research Group, Magyar Tudósok Körútja 2, H-1117 Budapest, Hungary
(E-mail: akos.leiter@nokia-bell-labs.com; dome.matusovits@nokia.com; bokorl@hit.bme.hu)

DOI: 10.36244/ICJ.2025.1.7

kernel modules linked to the kernel at runtime, extending the monolithic kernel's functionality. The iptables framework communicates (Figure 2) with different predefined hookpoints of the kernel's protocol stack. These are where the Nework Address Translation (NAT), Network Address and Port Translation (NAPT), packet filtering, and other packet manipulation procedures take place:

- NF_IP_PRE_ROUTING: This is where the incoming traffic directly enters the kernel stack. There isn't any routing processing at this point.
- NF_IP_LOCAL_IN: The routing procedures have already been done, and the packet has been forwarded to the local host.
- NF_IP_FORWARD: This is the same as the previous case, except that the packet is destined for a remote host.
- NF_IP_LOCAL_OUT: The traffic is locally generated and destined to a remote host
- NF_IP_POST_ROUTING: Packet processing procedures after routing

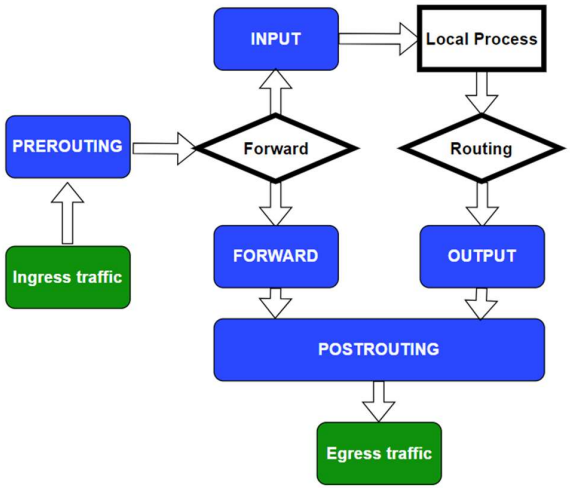


Figure 2 – The iptables packet processing in a nutshell

The iptables consists of rules, which contain targets. If a rule is evaluated, then the target is the action that needs to be executed (e.g., ACCEPT, DROP, RETURN, REJECT). Rules are part of chains that have two types: built-in (by the Linux OS) and custom (such as Kubernetes CNI-defined ones). The built-in chains are triggered by the abovementioned hookpoints respectively: PREROUTING (NF_IP_PRE_ROUTING), INPUT (NF_IP_LOCAL_IN), FORWARD (NF_IP_FORWARD), OUTPUT (NF_IP_LOCAL_OUT), POSTROUTING (NF_IP_POST_ROUTING).

The chains are located in tables. They are separated according to their appropriate functionality. For this reason, we can differentiate between Filter, NAT, Mangle, Raw, and Security tables. In the Filter table, the decision is made on whether the packet should enter or leave the network. The NAT rules can be found in the NAT table, as its name implies. The Mangle table contains packet manipulation rules. For configuration exemptions, where you do not want certain traffic to be tracked, you use the Raw table. It is designed to set a mark (NOTRACK) on a packet that has not wished to be tracked. For stricter access management, some Linux distributions include the security table. In order to achieve this, a mandatory access control (MAC) has been implemented.

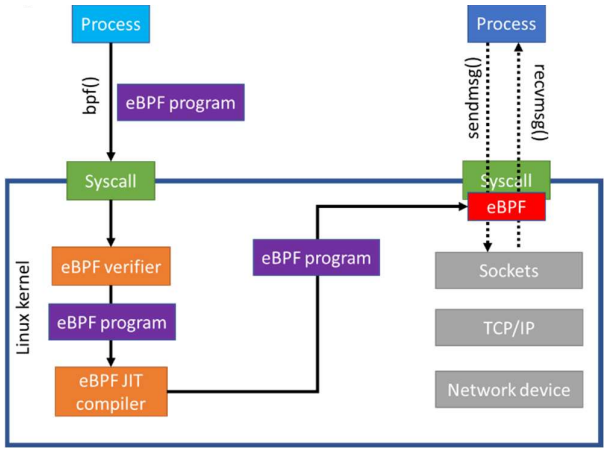


Figure 3 – eBPF execution flow

When the packet processing takes place, in the background the appropriate chain is selected within the table. Also, the desired rule should be applied in that chain. The problem occurs during the lookup phase. The table elements are not indexed; hence, the selection mechanism is sequential. At a small number of entries, it won't cause any problems. However, this number could be a significantly larger value in a production environment. In this case, we will experience substantial performance degradation in the packet processing. It could seriously harm SLA attributes, such as throughput and latency. That's why it is essential to develop a better solution that can enhance processing performance and help meet the requirements of URLLC communication.

Two well-known alternatives can boost packet processing: Vector Packet Processor (VPP) [12] and extended Berkeley Packet Filter (eBPF). The former solution implements a network stack, bypassing the Linux kernel. The essence here is that a new approach is being introduced to handling incoming traffic. The traditional Linux kernel-based solution is scalar processing, which typically processes one packet at a time. In contrast, at VPP, multiple packets are processed by their own network stack. These groups of packets are called vectors. In Kubernetes, the Calico CNI [9] is an example of its implementation. We do not go beyond this solution further, as our scope only focuses on eBPF.

In eBPF, we use eBPF programs to be loaded into the kernel from userspace (Figure 3). They are written in C language, but multiple development libraries can provide higher abstraction language levels, like bcc [13], libbpf [14], and eBPF Go[15]. Depending on the type of library, it uses a clang or Low Level Virtual Machine (LLVM) compiler to produce the so-called bytecodes from the source code. These are CPU-independent instruction sets translated by a Just-in- Time (JIT) compiler into machine-specific instruction sets. This way, we can optimize

the execution speed of the program using a natively compiled kernel code or a kernel module codebase. At a higher level abstraction, we can say that a virtual machine is practically embedded into the Linux kernel, where these eBPF programs run. They can be injected and executed on any level of the protocol stack. These are actually hookpoints, where certain events trigger the program execution. There can be many hookpoints, like kernel functions (kprobes) or user functions (uprobes) execution, system calls, or any tracepoints at the kernel. Also, eBPF programs can be attached to network interfaces or sockets as well (the latter two examples will be important in eBPF-based networking at Cilium). That's the key because, with this approach, we can extend the kernel's functionality without kernel source code modification or any usage of kernel modules. Additional functionalities, such as verification (depicted in Figure 3). improve and secure the execution compared to pure kernel modules, where there is no built-in and easy protection against, e.g., kernel panic. Before we get into the details of how eBPF can enhance packet processing in a Kubernetes environment, we introduce the related works that describe the most important eBPF use cases.

III. Relevant Parts of Kubernetes and Extended Berkeley Packet Filter

One of the most essential parts of our proposed testing environment is the Kubernetes integration of Cilium CNI, where our measurement results were gathered, depicted in Figure 4. It can leverage both iptables and eBPF-based networking to the whole cluster, where the Master and Worker nodes are located. The Master node is the control plane of Kubernetes. That is the component that involves the resource database (etcd) and the reconciliation loop mechanism (kube-controller manager), which controls the entire operation of the cluster. The API endpoint (apiserver) can also be found here, providing the cluster with reachability via HTTPS. What's more, the scheduling of workload resources (kube-scheduler) is also specified here. The Worker node provides the cluster's data plane. There, we could find the workloads that accomplish the desired services to be up and running.

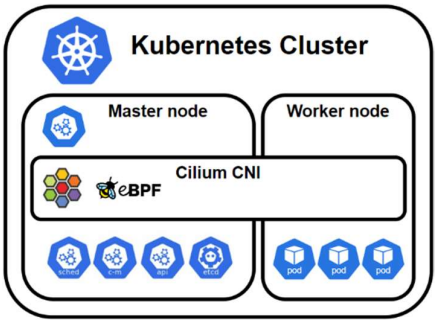


Figure 4 – High level architecture of Kubernetes with Cilium CNI

So far, all the cluster functionalities we mentioned have been implemented in Kubernetes' smallest unit, called the pod. Most of the time, a pod realizes a single container, but there can be a case when more than one container is embedded in a pod (e.g., a sidecar container, which receives the traffic, and there is another database container for information storage).

In both nodes, an entity should redirect control/data traffic to the desired endpoint (i.e., pod). That is where the kube- proxy comes into the picture. In most Kubernetes CNI solutions, iptables is used for packet processing. The kube-proxy's task is handling the appropriate chains, rules, and targets for traffic routing and manipulation. We aim to enhance packet processing performance by replacing iptables and hence, the kube-proxy.

eBPF also facilitates kernel programmability in Kubernetes [16]. Since there is only one kernel on a host, any application running in a container within a pod (in Kubernetes) must use the kernel whenever it requests access to hardware, manages files, or receives network messages. Regardless of the number of pods deployed on a machine, the kernel is always involved, whether we are talking about Bare Metal or a virtual machine. Containers do not have their own kernel; they use the existing kernel on the host machine. Thus, with proper eBPF instrumentation in the kernel, an agent can monitor all activities in the user space across all applications or cloud-native functions (micro-services). This enables complex eBPF tools to gain comprehensive observability across the entire node, providing deep insights into the cluster.

The two data paths that are associated with our experiment are shown in Figure 5. As a CNI, Cilium can deal with incoming traffic from the network interface of a Kubernetes Worker node or another Pod. Furthermore, the traffic destination can also be a Pod or the network interface of the Kubernetes Worker node. All the traffic goes through various iptables chains. The orange chains represent the default iptables chains; the blue ones are the Kubernetes-added ones. Cilium defines its own chains, depicted in purple.
The PREROUTING chain in Figure 5 is responsible for classifying whether traffic is local or must be forwarded. KUBE-SERVICES chains manage Kubernetes Services.

As shown in Figure 5, the many iptables chains on the data path can cause processing overhead and increased latency. This is where eBPF comes into the picture to circumvent issues with multiple iptables chains. An eBPF program can be loaded into the kernel to intercept traffic before the iptables-based processing starts. The hookpoint where the eBPF program is attached is called Traffic Control (TC). For incoming traffic, it is located before the PREROUTING, and for outgoing traffic, it can be found after the POSTROUTING chain. All of these mean that the eBPF- based solution intends to replace kube-proxy in Worker nodes that utilize iptables.

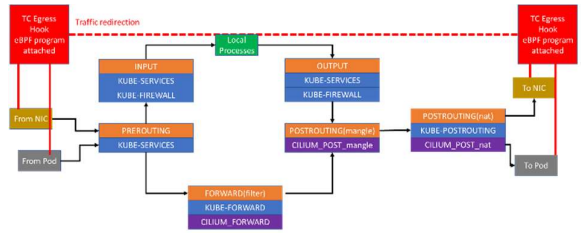


Figure 5 – iptables and eBPF-based Cilium data path [7]

IV. RELATED WORKS

eBPF can be used in many application fields related to security, observability, or performance enhancement scenarios. A summarization of the related papers is depicted in Figure 6.

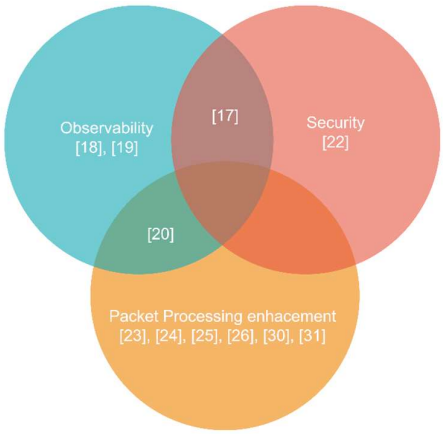


Figure 6 – eBPF use cases by summarized literatures

Regarding observability, eBPF can be used to monitor certain events. Attaching the written eBPF code to the appropriate hookpoints of the Linux kernel can trigger these eBPF programs to collect analytical traffic stream data. *David Soldani et al*. [17] used this approach to estimate cloud-native functions' energy consumption and derive performance counters and gauges for transport networks, 5G applications, and non-access stratum protocols. Furthermore, *Abderaouf Khichane et al*. [18] [19] have found a more profound way of measuring the behavior of a network function or protocol. Also, they could identify potential bottlenecks and SLA violations more accurately. *Carmine Scarpitta et al*. [20] describe a high-performance solution for end-to-end delay monitoring for SRv6-based networks. It leverages the Simple Two-way Active Measurement Protocol (STAMP) [21] to monitor the delay between two nodes called STAMP Session-Sender and Session-Reflector. The monitoring is implemented with eBPF programs.

Packet filtering mechanisms could be achieved more efficiently, like in the abovementioned paper by *David Soldani et al*. [17], where they detected and responded to unauthorized access to cloud-native resources in real time using eBPF. *Dominik Scholz et al*. [22] give a brief overview of analyzing the performance of eXpress Data Path (XDP), the lowest level before the network stack. They used for installing application-specific packet filtering configurations acting on the socket level. It is implemented with eBPF programs that are attached to the XDP hookpoint. It is well applicable for DoS prevention. Their case studies focus on performance aspects. Their packet filtering approach with eBPF doesn't have as much engineering cost. The performance losses are below 20%, while security is improved through better isolation between applications.

Attaching eBPF programs to the Linux kernel's protocol stack could also enhance the packet processing performance. It could be more efficient than the traditional netfiler [8] approach. That's the key point, as our goal was to evaluate performance using eBPF technology. *Matteo Bertrone et al*. [23] describe how the acceleration of packet processing can be achieved by emulating the iptables filtering semantic with eBPF, using Traffic Control (TC) or XDP. Depending on their use cases, such as delivering local traffic directly to the output port or connection tracking, they configured the data path respectively. This firewall solution is called bpf-iptables.

The paper by *Sebastiano Miano et al*. [24] extends the above scope by diving deep into the overall architecture of bpf-iptables, mentioning additional enhancements that make this technology perform better. Nftables [11] is also considered a relevant firewall alternative in these measurement scenarios. These are similar measurements in that they consider the TC hookpoint as we did. However, they only measure throughput, and the testbed is not in a cloud- native environment. We will also measure the latency and evaluate performance using a virtualized network infrastructure. Besides the observability aspect, in this previously mentioned approach by *Carmine Scarpitta et al*. [20], they managed to build the monitoring system where the eBPF implementation outperforms their examined solutions with negligible impact on the forwarding capability of the router. It uses XDP hookpoint, which differs from our scenario. Also, they only considered the throughput, similarly to paper [24].

Jung-Bok Lee et al. [25] implement an eBPF-based load-balancer. They also compare the performance of their eBPF-based solution to normal iptables, as we did in this paper. They developed a containerized high-performance load balancer that uses eBPF with the Linux kernel to distribute traffic, which can be easily managed via Kubernetes. They conducted tests simulating real-world traffic patterns using Internet Mix (IMIX) traffic streams. Their experimental results show that the proposed load balancer significantly outperforms the Destination Network Address Translation-based iptables solution, with the performance gap widening as packet size decreases. The measurements were conducted in a cloud environment, but their scope was only throughput performance scenarios, as in the previous papers. Also, they used XDP, instead of TC.

Federico Parola et al. show [26] a case study for Multi-access Edge Computing (MEC) technology, which is relevant in implementing the User Plane Function (UPF) deployed near the Radio Access Network (RAN), enabling telcos to provide services at close proximity to mobile users. In this scenario, high-performance data plane technologies, such as Data Plane Development Kit (DPDK) [27], may not be appropriate because they require dedicated resources like CPU cores and network interfaces. Furthermore, its proprietary drivers make it challenging to maintain and integrate DPDK. For this reason, they came up with a new idea to implement some of the functionalities of Mobile Gateway with eBPF/XDP, such as GPRS Tunneling Protocol Handling, QoS Management, Traffic Classifying, and Routing. They evaluated this approach with different Mobile Gateway data plane technologies like BESS [28], OpenvSwitch-DPDK (OvS-DPDK), and OvS-kernel [29].

The results show that eBPF competes with traditional kernel-bypass technologies. Although some performance degradation can be seen in some cases, it is still worth it because of higher integration with the kernel and more flexible resource usage. They used XDP hook as opposed to our case. Likewise, latency wasn't taken into account in these performance evaluations.

Dushyant Behl et al. [30] present a paper about the feasibility of eBPF for efficient implementation of network functions. They propose an eBPF-based framework to make the usage of eBPF CNI-agnostic. Their approach allows for replacing existing network functions with independent, eBPF-based modules. They were using multiple hookpoints: TC, XDP, and socket. We used only the TC hookpoint in our testbed. Also, they focused on enhancing the packet processing on the socket level by examining the throughput, where they could achieve a consistent 50% increase per scenario. There weren't any other attributes considered in their approach.

Code reusability is also an issue in the field of eBPF. *Federico Parola et al.* [31] address this problem by using PolyCube [32] [33]. PolyCube facilitates the development of efficient, modular, and dynamically reconfigurable network functions that run within the Linux kernel. This solution significantly improves Pod-to-Pod, Pod-to-Service, and Internet-to-Service throughput even in multi-node clusters compared to Flannel [21], Calico, and Cilium. This is the closest approach to our measurement use cases: it is based on Kubernetes, the traffic flow path is similar (Pod-to-Service scenario at least), and the eBPF hookpoint is the same (TC). They were even replacing the kube-proxy control plane element with eBPF programs as we did (they also achieved that with Cilium CNI in one of their test cases). However, they were scaling the associated pods to the Kubernetes Services not the number of Services itself. This is because they were curious about the load-balancing performance attributes when using eBPF. Also, as we can see in the previous papers, they only examined the throughput as a KPI.

V. The Implemented Test Environments

Based on Section III, we have created two test environments (Figure 8 and Figure 9) to study the performance of both solutions. The testbeds are built on OpenStack; the high-level design can be seen in Figure 7.

A. General principles of the test environments

The blue line represents the incoming, and the purple line shows the outgoing direction of the traffic (Figure 8, Figure 9). All the traffic originates from the *ITGSend* module of the Distributed Internet Traffic Generator (D-ITG) [34] on the client. The traffic is received by *ITGRecv* module, which is embedded in a Kubernetes Pod. There is a dedicated signaling port for connection establishment. To expose our D-ITG pod outside of the cluster, we need Kubernetes services (actually, ClusterIP is an exception because it makes the pod accessible only within the cluster). We can choose between ClusterIP, NodePort, LoadBalancer, and ExternalName. The latter option isn't remarkable for us since it only applies to mapping a service to a DNS name. Since the ITGSend module remains in the client network, the pod will be accessible through the Worker Node's interface with a private IP address. That means the NodePort service will be enough as it opens a port on the Worker node's interface and redirects the traffic to the pod. Note that LoadBalancer is preferred in production. We can use more protocols that can flow through it. Moreover, since it exposes the pod by acquiring an IP address for the desired service, we can make it accessible on the Internet (with a public IP address). However, for simplicity, we used NodePorts instead. Furthermore, the allocation of IPv4/v6 addresses for many services would harden the building of the testbed. The data ports were randomized. The maximum number of NodePorts is 2767, so when all of the ports were reserved, the remaining services were replaced with the type of ClusterIP during the service number increase, detailed in Section VI. To preserve the client's source IP, we use an annotation in the service definition file called *externalTrafficPolicy=Local*. With this annotation, the kube-proxy/eBPF program only proxies requests to local endpoints, which means we can avoid SNAT translation to node IP during any considered traffic flow.

B. High-level design

The client and the router VMs were placed in the client network created by OpenStack. The Kubernetes cluster – including the master and the worker node – was in the data center network, which was also created by OpenStack. All the elements in the test system (Client, Router, Master, and Worker Node) are OpenStack-instantiated virtual machines. We installed Ubuntu OS with version 20.04 (5.4.0 Linux kernel version) for the VMs. Also, we reserved 20Gb virtual memory with 1VCPU (1 core) and 2Gb RAM for the Client and the Router. Regarding the Master and Worker node instances, the setup was 40Gb memory with 2VCPUs (2 cores) and 4Gb RAM. The CPU clock rate was configured with 1500 MHz for each setup. There is also a management node to examine the system behavior without affecting the measurements, represented by orange lines. The black lines show the actual traffic path to be measured.

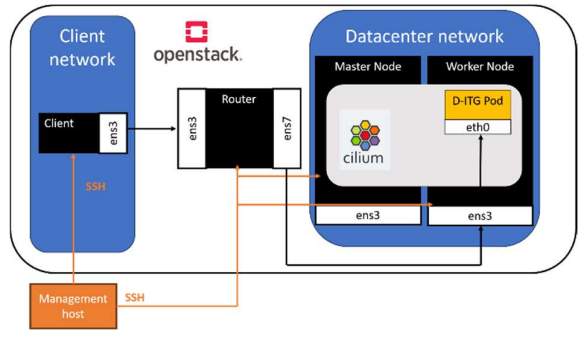


Figure 7 – The high-level testbed design with implementation details

C. Test environment for kube-proxy (iptables)

The kube-proxy-based test environment is shown in Figure 8. The red rectangles represent the relevant iptables chains through which the traffic goes.

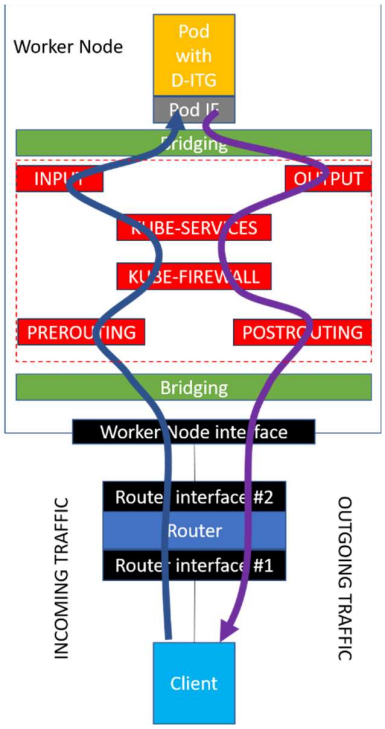


Figure 8 – Test environment #1: iptables-based forwarding

D. Test environment for eBPF

The eBPF-based test environment is depicted in Figure 9. The green rectangle shows the hook points where the eBPF program is attached. This means that after the packet arrives at the node interface, the eBPF program is triggered, and the packet processing and forwarding continue without iptables interaction.

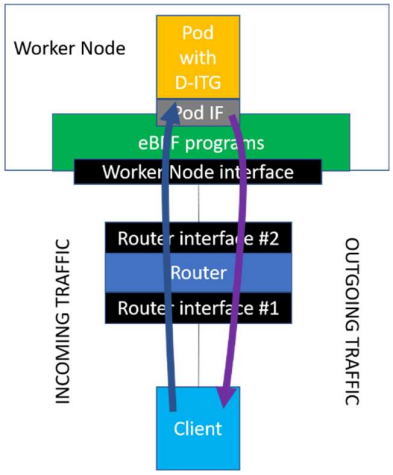


Figure 9 – Test environment #2: eBPF-based forwarding

VI. MEASUREMENTS DESCRIPTIONS AND RESULTS

We have examined several aspects of packet processing for our evaluation purposes. As we mentioned earlier, every measurement uses Kubernetes Services with NodePorts. A port number is associated with the node's IP address and will be translated to the Pod's IP and port number where you can reach the server. This Kubernetes object is responsible for routing traffic from the worker node's interface to the Pod handled by the kube-proxy/eBPF program. The traffic distribution between NodePorts is random. D-ITG is used for traffic generation. The test environments introduced in Section IV are applied.

We have continuously increased the number of Kubernetes Services to conclude the related bottlenecks of Kubernetes. Meanwhile, we evaluated two packet processing methods: normal kube-proxy-based (iptables) and eBPF- based.

E. TCPthroughputmeasurements

We used a relative scale as it is tough to determine maximum throughput in a virtualized environment. All the virtual links have been limited to 500 Mbps. One hundred measurements have been executed in every scenario – with 30-second-long TCP streams – where the number of Kubernetes Services is increased by 1000 (except from 1 to 1000). Altogether, 1100 measurements were evaluated overall.

Goal: Concluding the difference between kube-proxy (iptables) and eBPF-based packet processing in the case of IPv4 and IPv6 and within the context of throughput behavior.

Measurement results: From the data point of view, we highlight the standard deviation (Table 1) as there is no significant difference between the minimum, maximum, average, and median values. These values are also represented in Figure 10 and Figure 12, showcasing a different perspective on the measurement data.

TABLE I
eBPF-BASED THROUGHPUT STANDARD DEVIATION RATIOS COMPARED TO IPTABLES-BASED IN THE CASE OF IPv4 AND IPv6

Number of Kubernetes Services	Standard deviation ratio (IPv4)	Standard deviation ratio (IPv6)
1	0.20	0.85
1000	1.01	0.78
2000	1.07	0.95
3000	1.41	1.22
4000	0.93	0.56
5000	3.05	1.33
6000	0.92	1.20
7000	0.68	0.74
8000	0.91	0.56
9000	0.67	1.07
10000	0.67	0.88

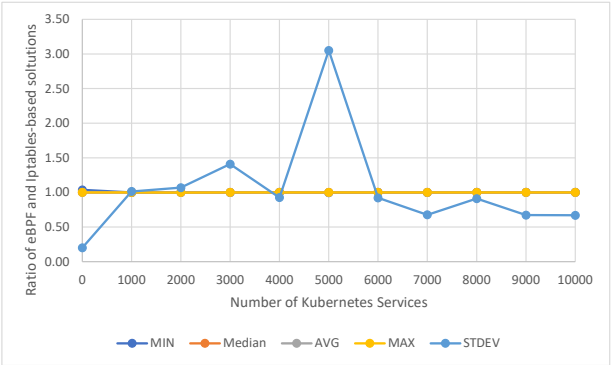


Figure 10 – Summarized diagram of throughput measurements between iptables and eBPF with IPv4

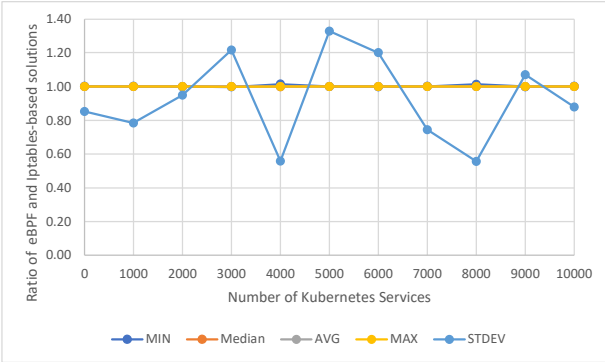


Figure 11 – Summarized diagram of throughput measurements between iptables and eBPF with IPv6

Conclusion: In the case of a high number of Kubernetes Services, the standard deviation of throughput is lower when eBPF is used in most cases. IPv6-based throughput values are more stable compared to IPv4 in both approaches, as the standard deviation of the eBPF to iptables ratio is lower.

F. Latency measurements

We also use a relative scale, just as we did in the case of throughput measurements. UDP traffic originated 100 times in every scenario with a 30-second-long flow, with the same service scaling as in the throughput measurements. This means 1100 measurements were in summary.

Goal: Concluding the difference between kube-proxy (iptables) and eBPF-based packet processing in the case of IPv4 and IPv6 within the context of latency behavior.

Measurement results: From the data point of view, we highlight the maximum and standard deviation (Table 2, Table 3) as there is no significant difference between minimum, average, and median values. The data is also represented in the graphs of Figure 12 and Figure 13.

TABLE II
eBPF-BASED LATENCY MAXIMUM AND STANDARD DEVIATION RATIO COMPARED TO IPTABLES-BASED IN THE CASE OF IPv4

Number of Kubernetes Services	Maximum	Standard deviation ratio
1	1.10	1.82
1000	1.21	1.91
2000	1.17	1.15
3000	1.18	1.49
4000	0.86	0.62
5000	0.90	0.47
6000	1.13	1.22
7000	0.80	0.50
8000	1.04	1.08
9000	0.91	0.54
10000	0.84	0.40

TABLE III
eBPF-BASED LATENCY MAXIMUM AND STANDARD DEVIATION RATIO COMPARED TO IPTABLES-BASED IN THE CASE OF IPv6

Number of Kubernetes Services	Maximum	Standard deviation ratio
1	1.08	1.01
1000	0.92	0.93
2000	0.95	0.93
3000	0.91	0.94
4000	0.86	0.87
5000	1.16	1.57
6000	1.05	1.00
7000	1.18	0.98
8000	0.65	0.74
9000	1.09	1.26
10000	0.87	0.82

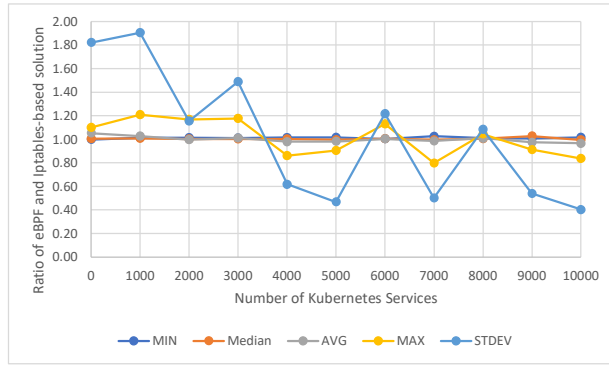


Figure 12 – Summarized diagram of delay measurements between iptables and eBPF with IPv4

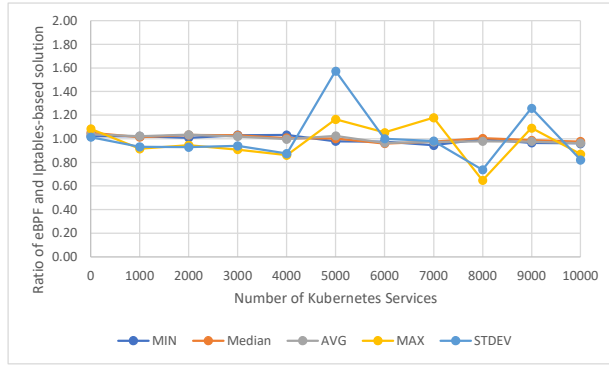


Figure 13 – Summarized diagram of delay measurements between iptables and eBPF with IPv6

Conclusion: The maximum latency values are higher for fewer Kubernetes services when using eBPF over IPv4. However, for IPv6 traffic, eBPF performs better in the case of fewer Kubernetes Services.

The standard deviation of latency for eBFP over IPv4 is lower in the case of 5 out of 11 scenarios. However, the greater the number of Kubernetes services used, the lower the standard deviation trend-wise in the case of eBPF. For IPv6, the standard deviation tends to be lower for eBPF with fewer Kubernetes services. However, this is not so significant compared to the IPv4 cases. Overall, IPv6 latency is more stable than IPv4 from the standard deviation point of view as the fluctuation of the values is lower.

G. Measurements conclusion

Generally, we can say that the eBPF-based solutions are more "stable" as the standard deviation is lower.

Generally, we can say that, based on the number of Kubernetes services, it is worth considering which type of Kubernetes packet processing is used because appropriate solutions can be suggested for different cases.

This experience can be utilized in URLLC network slices. Slicing SLAs always specify how reliably a particular parameter has to be kept (e.g., 99.999% of the time). These SLA requirements may be maintained better with lower latency fluctuation in certain eBPF cases. This can also contribute to telecommunication systems' overall software availability, as Varga et al. detailed in [35] [36].

Voice over IP services can also consider the results as lower jitter can be reached concerning the number of Kubernetes services in the telco cloud hardware.

H. Lessons learned

There were several difficulties during the creation of the test environment. Firstly, it is essential to differentiate the architecture of the measurement tools. As for the iPerf [37], there is a client and a server entity, and the connection establishment happens at the same port through which the data traffic flows. This means that the observed traffic is influenced by the signaling messages. In the case of D-ITG, there are dedicated ports for signaling and traffic generation, respectively. The desired data port we want to use is sent over the control plane as a plain-text message. Therefore, the NodePorts to pod's port translation won't happen. This means we cannot send traffic to the pod. Our solution was to choose the same port for the pod's port and the NodePort, so we had to define the NodePort to achieve that manually. Furthermore, there were some cases when the server was shut down randomly. So, we had to handle this and consider it inside the automatized shell script, which we used for measurements. Moreover, D-ITG components are more separated by their functionalities than iPerf. There are several entities present: ITGSend (at the client, it establishes the connection and generates the traffic), ITGRecv (at the server, it receives the traffic), ITGDec (at the client, it decodes the measurement result saved in a config file). Beyond the scope of our testbed, other entities can still be used for different scenarios, like ITGLog and ITGManager. So, we can see that the overall architecture of D-ITG is more complex than that of iPerf, which uses a simple point-to-point client-server model. Even though it is hard to implement D-ITG measurement in a cloud-native environment, this is still a valid solution as it is very flexible and has accurate traffic generation [38] [39]. It is also important to mention that the most unstable test scenarios were the IPv6-based traffic generations, where we used eBPF programs for packet processing. There were some cases where the traffic generator crashed. Not to mention that the higher the throughput was, the more time it took for ITGDec to decode the config file. Unfortunately, it is a limitation of the D-ITG software, which caused a massive impact on the measurement time. Delay measurement test cases took about 8-9 hours, and for the throughput analysis, it was 16-18 hours. All in all, generalizing the applied scripts required continuous development and spared much time to be usable. In a cloud environment – especially in public clouds – it is impossible to fully isolate a particular workload. Background traffic and other workloads may affect the measurement system and the performance of network functions. This might add an additional deviation in the results.

VII. CONCLUSION

In this paper, we have shown that there is room for eBPF to improve network performance in several use cases of Kubernetes-based telco cloud infrastructures. With the help of our results, operators can choose the packet processing methods that are the most suitable for their usage. With a significant service number, the throughput and latency values are more stable with IPv4 and eBPF. In IPv6-based measurements, the use of eBPF gives more stable results in most of the cases.

eBPF is not just about performance improvements; it is a complete framework supporting more straightforward and secure software development. Telecommunication networks can also benefit from better observability of network functions, which supports a variety of fields to be measured, such as energy consumption, SLA violation, logging, protocol analysis, security, etc. We believe this holistic approach will affect the whole telecommunication landscape. Even though there are some cases where eBPF does not outperform iptables, the application fields and feature sets mentioned above are worth the cost.

eBPF can support network slicing itself due to the increased observability, which leads to more control over particular data paths. Thus, it is easier to fulfill slice availability and performance requirements.

VIII. FUTURE WORK

The scalability of Kubernetes is crucial, and it is not just with Kubernetes services. It is worth examining how many Worker Nodes, Ingress controllers, etc., can be safely used by telecommunication applications or even in a regular IT cloud. As an example, Calico Typha deals also with scalability [40]. This pertains to, e.g., regulatory requirements where advanced logging is needed, which must also be scalable. In a Kubernetes-based telco environment, it is worth examining how eBPF can solve issues related to networking itself, such as assigning multiple interfaces to a pod or eliminating Network Address Translation (NAT) by Kubernetes services [41]. Furthermore, additional measurement points can be added to identify which packet processing section can cause increased volatility.

ACKNOWLEDGMENT

The authors thank Nandor Galambosi and Jozsef Varga for their constructive criticism and valuable comments while preparing the manuscript.

REFERENCES

[1] *IMT-2020 requirements*. Accessed: Aug. 21, 2024. [Online]. Available: https://www.3gpp.org/technologies/3gpp-meets-imt-2020

[2] G. Blinowski, A. Ojdowska, and A. Przybyłek, 'Monolithic vs. Microservice Architecture: A Performance and Scalability Evaluation', *IEEE Access*, vol. 10, pp. 20 357–20 374, 2022, DOI: 10.1109/ACCESS.2022.3152803.

[3] O. Al-Debagy and P. Martinek, 'A Comparative Review of Microservices and Monolithic Architectures', in *2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2018, pp. 000 149–000 154. DOI: 10.1109/CINTI.2018.8928192.

[4] H. T. Nguyen, T. Van Do, and C. Rotter, 'Scaling UPF Instances in 5G/6G Core With Deep Reinforcement Learning', *IEEE Access*, vol. 9, pp. 165 892–165 906, 2021, DOI: 10.1109/ACCESS.2021.3135315.

[5] P. Mach and Z. Becvar, 'Mobile Edge Computing: A Survey on Architecture and Computation Offloading', *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017, DOI: 10.1109/COMST.2017.2682318.

[6] 'Extended Berkeley Packet Filter (eBPF)'. [Online]. Available: https://ebpf.io/

[7] 'Cilium Kubernetes CNI'. Accessed: Nov. 05, 2023. [Online]. Available: https://cilium.io/

[8] 'Nokia official documentation of Network Service Platform'. Accessed: Jun. 26, 2024. [Online]. Available: https://documentation.nokia.com/nsp/24-4/Transport_Slice_Controller/Overview.html

[9] 'Calico Kubernetes CNI'. Accessed: Nov. 05, 2023. [Online]. Available: https://docs.projectcalico.org/getting-started/kubernetes/

[10] 'Nftables'. Accessed: Jun. 28, 2024. [Online]. Available: https://netfilter.org/projects/nftables/

[11] *Netfilter project*. Accessed: Jul. 10, 2024. [Online]. Available: https://www.netfilter.org/

[12] 'Vector Packet Processing'. Accessed: Feb. 04, 2025. [Online]. Available: https://fdio-vpp.readthedocs.io/en/latest/overview/whatisvpp/what-is-vector-packet-processing.html

[13] 'BCC – Toolkit and library for efficient BPF-based kernel tracing'. Accessed: May 10, 2024. [Online]. Available: https://ebpf.io/applications/

[14] 'libbpf – C-based library'. Accessed: Mar. 22, 2024. [Online]. Available: https://docs.kernel.org/bpf/libbpf/libbpf_overview.html

[15] 'The eBPF Library for Go'. Accessed: Mar. 14, 2024. [Online]. Available: https://ebpf-go.dev/

[16] L. Rise, *Learning eBPF: Programming the Linux Kernel for Enhanced Observability Networking and Security, pp. 218, 2023*. [Online]. Available: https://github.com/lizrice/learning-ebpf

[17] D. Soldani *et al.*, 'eBPF: A New Approach to Cloud-Native Observability, Networking and Security for Current (5G) and Future Mobile Networks (6G and Beyond)', *IEEE Access*, vol. 11, pp. 57 174– 57 202, 2023, DOI: 10.1109/ACCESS.2023.3281480.

[18] A. Khichane, I. Fajjari, N. Aitsaadi, and M. Gueroui, '5GC-Observer: a Non-intrusive Observability Framework for Cloud Native 5G System', in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–10. DOI: 10.1109/NOMS56928.2023.10154433.

[19] A. Khichane, I. Fajjari, N. Aitsaadi, and M. Gueroui, '5GC-Observer Demonstrator: a Non-intrusive Observability Prototype for Cloud Native 5G System', in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–3. DOI: 10.1109/NOMS56928.2023.10154369.

[20] C. Scarpitta, G. Sidoretti, A. Mayer, S. Salsano, A. Abdelsalam, and C. Filsfils, 'High Performance Delay Monitoring for SRv6-Based SD-WANs', *IEEE Transactions on Network and Service Management*, vol. 21, no. 1, pp. 1067–1081, 2024, DOI: 10.1109/TNSM.2023.3300151.

[21] G. Mirsky, G. Jun, H. Nydell, and R. Foote, 'Simple Two-Way Active Measurement Protocol'. in Internet Request for Comments, no. 8762. RFC Editor, Fremont, CA, USA, Mar. 2020. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8762.txt

[22] D. Scholz, D. Raumer, P. Emmerich, A. Kurtz, K. Lesiak, and G. Carle, 'Performance Implications of Packet Filtering with Linux eBPF', in *2018 30th International Teletraffic Congress (ITC 30)*, 2018, pp. 209–217. DOI: 10.1109/ITC30.2018.00039.

[23] M. Bertrone, S. Miano, F. Risso, and M. Tumolo, 'Accelerating Linux Security with eBPF iptables', in *Proceedings of the ACM SIGCOMM 2018 Conference on Posters and Demos*, in SIGCOMM '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 108–110. DOI: 10.1145/3234200.3234228.

[24] S. Miano, M. Bertrone, F. Risso, M. V. Bernal, Y. Lu, and J. Pi, 'Securing Linux with a faster and scalable iptables', *SIGCOMM Comput. Commun. Rev.*, vol. 49, no. 3, pp. 2–17, Nov. 2019, DOI: 10.1145/3371927.3371929.

[25] J.-B. Lee, T.-H. Yoo, E.-H. Lee, B.-H. Hwang, S.-W. Ahn, and C.-H. Cho, 'High-Performance Software Load Balancer for Cloud-Native Architecture', *IEEE Access*, vol. 9, pp. 123704–123716, 2021, DOI: 10.1109/ACCESS.2021.3108801.

[26] F. Parola, F. Risso, and S. Miano, 'Providing Telco-oriented Network Services with eBPF: the Case for a 5G Mobile Gateway', in *2021 IEEE 7th International Conference on Network Softwarization (NetSoft)*, 2021, pp. 221–225. DOI: 10.1109/NetSoft51509.2021.9492571.

[27] 'Data Plane Development Kit (DPDK)'. Accessed: Jul. 01, 2024. [Online]. Available: https://www.dpdk.org/

[28] S. Han, K. Jang, A. Panda, S. Palkar, D. Han, and S. Ratnasamy, 'SoftNIC: A software NIC to augment hardware', *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2015-155*, 2015.

[29] B. Pfaff *et al.*, 'The design and implementation of open {vSwitch}', in *12th USENIX symposium on networked systems design and implementation (NSDI 15)*, 2015, pp. 117–130.

[30] D. Behl, H. Huang, P. Kodeswaran, and S. Sen, 'On eBPF extensions to Kubernetes CNI datapath', in *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 2023, pp. 207–209. DOI: 10.1109/COMSNETS56262.2023.10041357.

[31] F. Parola, L. D. Giovanna, G. Ognibene, and F. Risso, 'Creating Disaggregated Network Services with eBPF: the Kubernetes Network Provider Use Case', in *2022 IEEE 8th International Conference on Network Softwarization (NetSoft)*, 2022, pp. 254–258. DOI: 10.1109/NetSoft54395.2022.9844062.

[32] S. Miano, F. Risso, M. V. Bernal, M. Bertrone, and Y. Lu, 'A Framework for eBPF-Based Network Functions in an Era of Microservices', *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 133–151, 2021, DOI: 10.1109/TNSM.2021.3055676.

[33] S. Miano, M. Bertrone, F. Risso, M. Tumolo, and M. V. Bernal, 'Creating Complex Network Services with eBPF: Experience and Lessons Learned', in *2018 IEEE 19th International Conference on High Performance Switching and Routing (HPSR)*, 2018, pp. 1–8. DOI: 10.1109/HPSR.2018.8850758.

[34] S. Avallone, S. Guadagno, D. Emma, A. Pescape, and G. Ventre, 'D-ITG distributed Internet traffic generator', in *First International Conference on the Quantitative Evaluation of Systems, 2004. QEST 2004. Proceedings.*, 2004, pp. 316–317. DOI: 10.1109/QEST.2004.1348045.

[35] J. Varga, A. Hilt, J. Bíró, C. Rotter, and G. Jaro, 'Reducing operational costs of ultra-reliable low latency services in 5G', *Infocommunications Journal*, vol. X, pp. 37–45, 2018, DOI: 10.36244/ICJ.2018.4.6.

[36] J. Varga, A. Hilt, C. Rotter, and G. Járó, 'Providing Ultra-Reliable Low Latency Services for 5G with Unattended Datacenters', in *2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, 2018, pp. 1–4. DOI: 10.1109/CSNDSP.2018.8471756.

[37] 'Iperf3 – traffic generator'. Accessed: Nov. 05, 2023. [Online]. Available: https://iperf.fr/iperf-download.php

[38] G. Aceto, C. Guida, A. Montieri, V. Persico, and A. Pescapè, 'A First Look at Accurate Network Traffic Generation in Virtual Environments', in *2022 IEEE Symposium on Computers and Communications (ISCC)*, 2022, pp. 1–6. DOI: 10.1109/ISCC55528.2022.9913058.

[39] D. Perepelkin and M. Ivanchikova, 'Problem of Network Traffic Classification in Multiprovider Cloud Infrastructures Based on Machine Learning Methods', in *2021 10th Mediterranean Conference on Embedded Computing (MECO)*, 2021, pp. 1–5. DOI: 10.1109/MECO52532.2021.9460171.

[40] 'Calico Typha'. [Online]. Available: https://docs.tigera.io/calico/latest/reference/typha/

[41] Ákos Leiter *et al*., 'Cloud-Native IP-Based Mobility Management: A MIPv6 Home Agent Standalone Microservice Design', presented at the CSNDSP 2022

Ákos Leiter graduated as a Computer Engineer MSc at the Department of Networked Systems and Services (HIT), Budapest University of Technology and Economics (BME) in 2015, specializing in Computer Networks. His thesis was about proposing an operator-centric, dynamic flow mobility protocol with IP in the Evolved Packet Core. He is a PhD candidate at HIT's Multimedia Networks and Services Laboratory (MEDIANETS) and a research engineer at Nokia Bell Labs. His main research field is Network Function Virtualization and Software Defined Networking, including Orchestration and Network Automation. His work-in-progress PhD thesis is about the cloudification of the Mobile IPv6 protocol family on top of Kubernetes.

Döme Matusovits graduated from the Budapest University of Technology with a BSc in Electrical Engineering in 2024. He is currently pursuing his MSc degree at the Department of Networked Systems and Services (HIT), specializing in computer and mobile networks. His ongoing MSc thesis focuses on the Orchestration of Network Slices and Inter-Slice Handover. His main research fields include 5G, Network Slicing, and Software Defined Networking within the scope of Network Automation. The research is conducted in close collaboration with Nokia Bell Labs and HIT. Since 2023, he has been working at Nokia in the Cloud and Network Services, developing 5G products.

László Bokor received his Ph.D. degree in computer engineering from Budapest University of Technology and Economics (BME) in 2014. He is currently an associate professor at the Department of Networked Systems and Services (HIT), where he leads the Vehicular Communications Research Group founded within strong industry-academic cooperation. He is a member of the HTE (Scientific Association for Infocommunications Hungary), the Hungarian Standards Institution's Technical Committee for Intelligent Transport Systems (MSZT/MB 911), the TPEGoverC-ITS Task Force within the TPEG Application Working Group of TISA, the ITS Hungary Association (the Hungarian organization of ERTICO's Network of National ITS Associations), and the BME's Multimedia Networks and Services Laboratory, where he participates in different R&D projects. His research interests include IPv6 mobility, SDN/NFV-based mobile networks, network simulation, mobile healthcare infrastructures, and V2X communication in cooperative intelligent transportation systems.

Support Vector Machines: Theory, Algorithms, and Applications

Mohammed Jabardi

***Abstract*—Support Vector Machines, or SVMs, are a strong group of supervised learning models that are commonly used for tasks like regression and classification. SVMs are based on the theory of statistical learning and try to find the best hyperplane that maximizes the gap between different classes. This makes it easier to apply to new data. Since kernel functions are used with SVMs, they are more flexible and can handle both linear and nonlinear situations well. Even though they have a strong theoretical base, they still face problems in the real world, like being hard to code and difficult to tune parameters, especially for big datasets. Recent improvements, like scalable solvers and estimated kernel methods, have made them a lot more useful. This essay talks about SVM theory, its main algorithms, and how it is used in the real world. It shows how it is used in bioinformatics, banking, and image processing, among other areas.**

***Index Terms*—Support Vector Machine, Classification, Regression, Machine Learning, Hyperplanes, Kernel Functions.**

I. INTRODUCTION

Support Vector Machine (SV), is one of the simplest and most refined classification methods in machine learning. Unlike neural networks, SVMs can work with very small datasets and are not inclined to overfitting. The SVM is used to classify each object by representing points in an N-dimensional space and the coordinates of these points, which are usually called features. [2]

SVM perform the classification procedure by drawing a hyperplane that is a line in 2 or 3 dimensional in a plane such a way that all points of one category are on one side of the hyperplane and all points of other categories are on the other side. If there are multiple hyperplanes, SVM try to find the one that best separate the two categories, in the sense that maximizes the distance to points in either category [3-4]. This distance called the Margin and all the points fall exactly on the margin are called the Supporting Vectors. To find the hyperplane in the first place the SVM requires for training set or set of points that already labeled with the correct category, this is why SVM is said to be supervised learning algorithm. In the background SVM solve a convex optimization problem that maximize this margin and where constraints say that points for each category should be fall in the correct side of the hyperplane. [5-6]

While it's mainly used for binary classification, SVM can also handle multiclass problems by using strategies like one-vs-all (comparing one class against all others) or one-vs-one (building a classifier for each pair of classes [7]. Originally presented by Vapnik, SVMs are well-known for their kernel-based approach to classification and regression tasks [8–10]. In data mining, pattern recognition, and machine learning, their remarkable generalization capacity, optimal solutions, and discriminative power have attracted plenty of attention. Originally presented by Vapnik, SVM is well-known for its kernel-based method of handling regression and classification problems [8–10]. The data mining, pattern recognition, and machine learning groups in recent years have shown great interest in its exceptional generalization capacities, optimal solutions, and discriminative capability [11]. To maximize the separation margin in a high-dimensional feature space, SVMs optimize decision functions directly from training data [12–18]. This strategy not only minimizes training data errors but also improves generalization abilities. The support vector machine algorithm or SVM it looks at the extremes of the data sets and draws a decision boundary also known as a hyperplane near the extreme points in the data set so essentially the support vector machine algorithm is a frontier which best segregates the two classes. [19-20].

M. Jabardi is with Department of Software, College of Education, University of Kufa, Najaf, Iraq. (E-mail: mohammed.naji@uokufa.edu.iq)

DOI: 10.36244/ICJ.2025.1.8

A. SVM Historical Perspective and their Evolution.

Vladimir Vapnik and Alexey Chervonenkis's landmark paper, "A Theory of Learning from General Examples" (1964), laid the foundation for statistical learning theory. It emphasized minimizing generalization error rather than training error, a principle central to SVMs [21]. Vapnik and Boser further developed SVMs in their 1992 paper, "Pattern Recognition Using an Insensitive Loss Function," where they detailed classical SVM algorithms for binary classification and introduced the concept of support vectors [22]. The development of kernel methods by Christopher J.C. Burges in "A Tutorial on Support Vector Machines for Pattern Recognition" (1998) enabled SVMs to effectively handle non-linearly separable data [23].

B. Application of SVN

The late 1990s and early 2000s saw the development of robust SVM software libraries like LIBSVM and SMO, making them readily accessible to practitioners. This accessibility sparked a surge in SVM applications across various fields, including [24-25]:

- Text classification: Spam filtering, sentiment analysis, topic modeling.
- Image classification and object detection: Handwritten digit recognition, object detection, image segmentation.
- Bioinformatics and computational biology: Gene classification, protein analysis, disease prediction.
- Financial forecasting: Stock market prediction, credit risk assessment.

- Healthcare and medical diagnosis
- Cybersecurity and intrusion detection
- Environmental science and climate modeling.

C. SVMs Advantages and Disadvantages [26-28]:

Advantages:

- Particularly successful in high-dimensional environments.
- Memory effective: Makes only use of a subset of the training points.
- flexible: the kernel method helps here.
- Scalability and efficiency: advances in large-scale SVMs and mass dataset handling optimizing techniques.
- Multi-class and multi-label classification stretches SVMs beyond binary classification to address intricate data structures.
- Designs new kernel functions and investigates adaptive kernels automatically that learn from data automatically in kernel learning and adaptation.
- This method minimizes generalization error and improves performance with fresh data by concentrating on margin maximizing and capacity control.
- Robustness: Greater robustness results from larger margins lowering susceptibility to data changes or noise. learning models.

Disadvantages:

- Not appropriate for big datasets: Long training times can be problematic.
- Sensitive to kernel and hyperparameter decisions: These can greatly affect results.
- Difficult Interpretability: Complicating knowledge are complex elements including high-dimensional decision limits and kernel modifications.
- Handling big datasets might result in major computational expenses and increase training times, hence computationally demanding.
- Correct parameter settings define performance; inadequate calibration can produce less than ideal results.
- Lack probabilistic outcomes: procedures like Platt scaling are required for probabilities; SVMs mostly produce binary classifications without direct probability estimations.
- Understanding complex models is challenging: Particularly with nonlinear kernels, intricate decision boundaries complicate models for interpretation.
- Scalability problems: Memory and computing restrictions can make training on very big datasets unworkable.

II. RECENT REVIEW ARTICLES AND SURVEYS ON SVMS

TABLE I
THE SUMMARY OF THE LITERATURE REVIEW

Reference	Description
[25-28]	Introduces machine learning, focusing on supervised learning and SVMs. Explores SVM capabilities, applications, and future prospects.
[29]	SVM for two-class classification, kernels, and penalty functions, furthermore covering multiclass methods, one-class SVDD, and Support Vector Regression for handling outliers and non-linear data.
[30-31]	Provides an overview of SVM applications, challenges, and emerging trends, highlighting their utility in various fields.
[32]	Comprehensive review of SVMs, covering fundamental concepts, kernel methods, optimization algorithms, and applications.
[33]	Surveys SVM applications in bioinformatics, healthcare, finance, image processing, and natural language processing.
[34]	Focuses on interpretable SVMs, covering techniques like rule extraction, feature importance analysis, and model-agnostic methods.
[35]	Reviews challenges and solutions for large-scale SVM training, including stochastic gradient descent and distributed computing.
[36]	Explores hybrid models combining SVMs with deep learning to improve performance and address individual limitations.
[37]	Highlights emerging SVM applications in bioinformatics, healthcare, finance, and natural language processing.
[38]	Develops an automated facial expression recognition system using SVM, MLP, and KNN classifiers with HOG and PCA for feature extraction.
[39]	Examines linear SVM classification, focusing on solvers, improvements, empirical findings, and future research directions.
[40]	Demonstrates SVM's effectiveness in predicting Alzheimer's disease using MRI data, emphasizing its potential in medical diagnosis.
[41]	Proposes a bagged ensemble SVM technique for speech emotion recognition, contributing to Human-Computer Interaction research.
[42]	Introduces an SVM-based intrusion detection framework with naive Bayes feature embedding, improving network security.

[43] Presents a U-Net-based method for melanoma classification in dermoscopy images, using segmentation, feature extraction, and SVM.
[44] Combines deep neural networks (DNN) and multiclass SVMs for classification, using K-means clustering for feature extraction.
[45] Enhances SVM classification capabilities by incorporating dynamic graph learning and self-paced learning.
[46] Highlights SVM's role in interpreting neuroimaging data for brain disorder research and precision psychiatry.
[47] Proposes a CNN-SVM hybrid model for diagnosing faults in rotating machinery, improving early-stage fault detection.
[48] Combines deep learning and SVM for identifying and predicting rice leaf diseases.
[49] Introduces a method for detecting malaria parasites using deep neural networks and SVM with transfer learning.
[50] Explores SVM's role in image classification, discussing its evolution, variants, and applications.
[51] Uses PSO, GA, and Grid Search to optimize SVM parameters for risk assessment in railway transportation systems.
[52] Addresses factors affecting SVM performance in classifying nonlinearly separable problems, providing insights for future research.
[53] Proposes a deep learning method for breast cancer detection using mammography, combining DNN and multiclass SVM.
[54] Introduces DeepSVM-fold, a computational predictor for protein fold recognition, offering improved accuracy over existing methods.

III. METHODOLOGY

One of the most well-known supervised learning methods is Support Vector Machine (SVM). Its main job is to sort things into groups, but it can also help with error problems in machine learning [24]. In n-dimensional space, the SVM algorithm tries to find the best line or decision boundary that splits it into classes. This makes it easy to put new data points into the right category. A hyperplane is the name for this best border. SVM finds the most important extreme points or vectors for defining this hyperplane [25–26].

The Core Concepts [27-30]

Support Vector Machines (SVMs) aim to identify a hyperplane with the largest possible margin, resulting in a robust classification model. The mathematical method maximizes the squared norm of the weight vector under constraints guaranteeing class separation. Lagrange multipliers help to simplify this optimization issue. Crucially important data points defining the decision-making range are support vectors. Then, depending on their feature vectors, a decision function groups newly occurring data points.

Important notes:

Hyperplanes are data point classification boundary. Points on opposing sides of the hyperplane fall into several categories. The number of features determines the hyperplane's dimension; for instance, a hyperplane with two features is a line and with three it becomes a plane.

Margins: The margin is the distance between the hyperplane and the closest data points, known as support vectors. This distance can be mathematically expressed as: $2/(\|w\|)$. The Euclidean norm of the weight vector w is denoted as $\|w\|$.

Maximizing Margins: SVMs strive to find the hyperplane with the widest margin, enhancing the classifier's generalization capabilities.

Regularization: This technique helps prevent overfitting in SVMs by introducing a penalty term in the objective function, which encourages the model to prefer simpler decision boundaries over complex ones that perfectly fit the training data.

Support Vectors: These are data points close to the hyperplane that significantly influence its position and orientation and are essential for constructing the SVM

Support Vectors: These are data points close to the hyperplane that significantly influence its position and orientation and are essential for constructing the SVM. Altering these points would shift the hyperplane, as

$\vec{x}.\vec{w} = c$ (the point lies on the decision boundary)
$\vec{x}.\vec{w} > c$ (positive samples)
$\vec{x}.\vec{w} < c$ (negative samples)

illustrated in Figure 1.

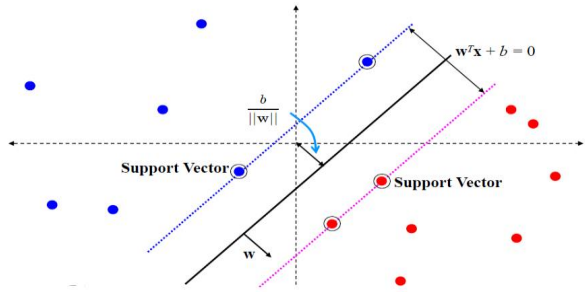


Fig. 1. Support vectors

Kernel Functions: These functions transform the input data into a higher-dimensional feature space, making it easier for a linear classifier to separate the data. Kernels help capture complex, nonlinear patterns like curves and circles. Common types include linear, polynomial, radial basis function (RBF), and sigmoid. There are two different categories that are classified using a decision boundary or hyperplane as shown in Figure. 2.

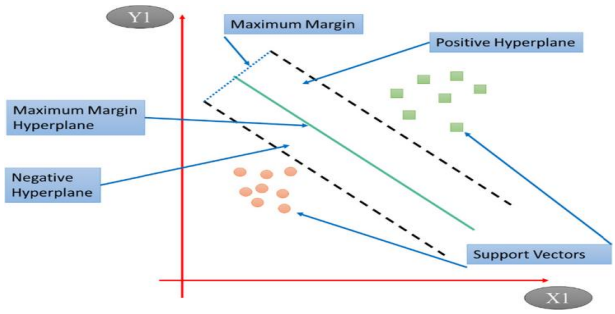


Fig. 2. The hyperplanes used to classify data points [30]

Consider a random point X, and determine whether it is above or below the hyperplane, or on it, as shown in Figure 3. First, represent X as a vector. Then, construct a vector (w) perpendicular to the hyperplane. Suppose C is the distance from the origin to the decision boundary along (w). Project X onto (w) through a dot product. If the dot product exceeds C, X is above the plane; if less, below; if equal, on the decision boundary [31].

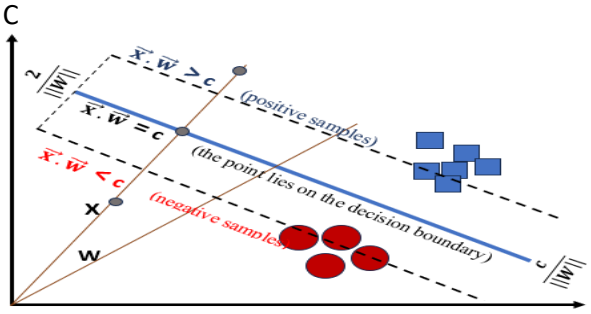


Fig. 3. The core concept of the SVM algorithm

A. Types of Support Vector Machine

Based on the nature of the decision boundary, Support Vector Machines (SVM) can be divided into two main parts as shown in Figure 1:

Support Vector Machines (SVM) can be categorized into two main types based on the nature of the decision boundary

1. Linear SVM:
   This type is applicable when the data is perfectly linearly separable. This means that the data points can be divided into two classes using a single straight line in a two-dimensional space, as shown in Figure 4.A.
2. Non-Linear SVM:
   When the data is not linearly separable, a Non-Linear SVM is used. This occurs when the data points cannot be separated into two classes by a straight line. In such cases, advanced techniques like the kernel trick are employed to classify the data. Most real-world applications involve non-linearly separable data, hence the kernel trick is commonly used, as depicted in Figure 4.B.

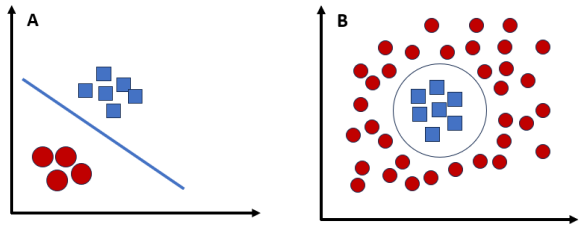


Fig.4. **A:** Linearly Separable Data **B:** Non-Linearly Separable Data

B. Mathematical Formulation

1. **Linear SVM**:
   This type is applicable when the data is perfectly linearly separable. This means that the data points can be divided into two classes using a single straight line in a two-dimensional space, as shown in Figure 4.A.

   1.1. Data: a dataset of points ({Xi}, {Yi}), where Xi is an n-dimensional vector representing the features of a data point and Yi is the class label (+1 or -1).

   1.2. Hyperplane Equation: The hyperplane is defined as:
   $W_T \times X + b = 0$ ..... (1)
   Where W is a weight vector normal (perpendicular) to the hyperplane X is an input vector b is the bias term

   1.3. Constraints: For a data point to be correctly classified, we need:
   $Yi\ (W_T \times X_i + b) \geq 1$ for $Y_i = +1$ ...(2)
   $Yi\ (W_T \times X_i + b) \leq -1$ for $Y_i = -1$ ...(3)

   1.4. Optimization Problem: Maximizing the margin is equivalent to minimizing the following objective function: $W = \frac{||W^2||}{2}$ (subject to the constraints above). This is a quadratic optimization problem, usually solved using techniques like Lagrange multipliers.

2. **Non-Linear SVMs (The Kernel functions)**:
   The popular kernel types that we can use transform the data into high dimensional feature space are polynomial kernel, radial basis function RPF or RBF kernel and sigmoid kernel [32]. Choosing the correct kernel is a non- trivial task and may depend on specific task at hand no matter which kernel we choose, just we need to tune the kernel parameters to get good performance from a classifier. A popular parameter tuning technique includes k-fold cross-validation. Some of the most common kernel functions for support vector machines include:

   2.1. The Linear Kernel:
   The linear kernel, or dot product kernel, is the simplest function. It calculates the dot product of input feature vectors in the original input space. Mathematically, it is expressed as in equation (4).
   $K(xi, xj) = xiT \times xj$ ..... (4)

Advantages:

- Efficiency: The linear kernel excels in computational efficiency, involving only a simple dot product operation, making it suitable for high-dimensional data where other kernels might become computationally expensive.
- Interpretability: It offers the highest level of interpretability among kernel functions. The decision boundary learned by the SVM is a hyperplane in the original feature space, and the weights assigned to each feature reveal their relative importance for classification.
- No Hyperparameter Tuning: Unlike most other kernels, the linear kernel requires no hyperparameter tuning, making it easier to use and reducing the risk of overfitting due to poorly chosen hyperparameter.

Limitations:

- Limited to Linearly Separable Data: Its main limitation is that it can only handle data that is already linearly separable in the original feature space. If the data exhibits complex, non-linear relationships, the linear kernel will not effectively learn a separation boundary.
- Less Flexible: Due to its simplicity, the linear kernel is less flexible in modeling complex non-linear patterns compared to kernels like polynomial or RBF.

2.2. Polynomial Kernel:

The polynomial kernel calculates the similarity between two vectors by raising the dot product of the original vectors to a given power d, adding non-linearity to the decision boundary. Mathematically, it is expressed as in equation (5).

$$K(xi, xj) = (xiT \times xj + 1)d \quad \text{..... (5)}$$

2.3. Radial Basis Function (RBF) Kernel:

A lot of users work with the RBF kernel, which is also known as the Gaussian kernel. The Gaussian distribution is used to measure the distance between two vectors in the feature space to find out how close they are. This is helpful when there isn't a clear line between the input data. It can be written mathematically as shown in equation (6). If you change the hyperparameter gamma, it changes how wide the Gaussian distribution is. Radial Basis Function (RBF) kernel is a popular and flexible choice for SVMs that work with data that is not linear.

$$K(x_i, x_j) = e^{(-\text{gamma} \,\|\, K(Xi, Xj) \,\|2)} \quad \text{……. (6)}$$

Advantages:

- Useful for Non-Linear Data: The RBF kernel is great at turning data into a higher-dimensional feature space by using a Gaussian function. This change makes it possible for SVMs to see complicated, non-linear connections between traits that weren't possible in the original space.
- Stability: The RBF kernel is less likely to suffer from the curse of dimensionality than the polynomial kernel. It works well with data that has a lot of dimensions and doesn't have the same risk of overfitting.
- Fewer Hyperparameters: The RBF kernel only has one hyperparameter, called gamma. However, it is not as sensitive to setting hyperparameters as the degree parameter of the polynomial kernel. This can make the process of choosing a model easier.

Limitations:

- Interpretability: Like most non-linear kernels, the RBF kernel sacrifices some interpretability compared to the linear kernel. The decision boundary becomes less intuitive in the original feature space.
- Computational Cost: Although generally more efficient than the polynomial kernel for high dimensions, the RBF kernel can still be computationally expensive, especially for very large datasets.
- Hyperparameter Tuning: While less sensitive than the polynomial kernel, the RBF kernel's performance still depends on finding the optimal gamma value. Careful hyperparameter tuning is essential.

2.4. Optimal Kernel Selection:

An Empirical Methodology The effectiveness of SVMs relies on carefully choosing the right kernel function that suits the specific challenge. Here's an analysis of typical options and their practical applications:

1- The linear kernel:

Situation: Consider classifying emails as either spam or legitimate. Possible features include word frequency and the presence of spam keywords. The linear kernel is a good initial choice because the relationship between these features is likely linear (a higher frequency of spam keywords often indicates spam). This model is computationally efficient and interpretable, with the decision boundary being a straight line in the original feature space.

2- The polynomial kernel:

This is used when assessing handwritten digits for recognition. Pixels can serve as features, but defining a clear decision boundary to separate different digits (such as 6 and 8) with a linear plane is often challenging. A low-degree polynomial kernel, such as quadratic, can introduce non-linearity, enabling the SVM to more accurately capture the curved features of some digits. However, using a high-degree polynomial kernel may lead to overfitting, underscoring the need for careful parameter tuning.

2.5. The Radial Basis Function (RBF)

Situation: Categorize photographs featuring various species of animals, such as cats, dogs, and birds. Pixel intensities and local features are used. However, it should be noted that the relationships between them exhibit a high degree of nonlinearity. The RBF kernel performs exceptionally well in this context. It turns input points into a space with an unlimited number of dimensions. This lets the SVM find complex spatial patterns, like the edges and textures that are unique to each species. The parameter (σ) controls how smooth the decision limit is. A higher π number makes the transition smoother, which could group animals that look alike, like all cats, even if they are in different positions. A lower number of σ allows for more complex differences, which could make it easier to tell the difference between different breeds.

2.6. SVMs Constraints and Difficulties [25-27]:

- It can be hard to work with very large datasets because SVMs can require a lot of processing power and memory, especially when working with very large datasets that are high dimensional.
- Choosing the Kernel Function: Picking the correct kernel function and its values has a big impact on how well SVM works. Finding the best kernel and tweaking its settings, on the other hand, can be hard and needs specialized knowledge.
- Noise Sensitivity: SVMs have great dataset outlier and noise sensitivity, overfitting can result from this sensitivity upsetting the decision limit. Sometimes preprocessing methods like eliminating outliers and lowering noise are required.
- Binary Focus: SVMs are mostly meant for binary classification tasks, so multiclassification becomes difficult. Usually, they are adapted for multiclassification utilizing one-vs- one or one-vs- all approaches. These techniques, however, can have problems with scalability and lower performance, hence stressing SVMs' shortcomings.
- Lack of Interpretability: SVMs generate a black-box model that makes it challenging to grasp the learned decision bounds and the fundamental data linkages. In sectors such banking and healthcare, interpretability is absolutely vital.
- Unbalanced Datasets: SVMs may perform badly in imbalanced datasets, in which case cases across classes vary significantly. To solve this, one could need different evaluation measures, resampling, or class weighting.

Despite these challenges, SVMs remain a robust and widely used approach in machine learning for tasks like classification, regression, and anomaly detection. Researchers continue to explore ways to overcome these obstacles and improve the efficiency and scalability of SVMs in various fields.

2.7. Why SVMs Are Computationally Expensive

The following causes make the SVMs computationally expensive:

1. **Kernel Computations:** SVMs calculate the kernel matrix when employing non-linear kernels; this structure has a size of (n × n), so it is not viable for high n.
2. **Dense Data Visualizations:** The computational cost rises significantly for extremely dimensional or sparse data such text data.
3. **Memory Requirement:** Large datasets are blocked by storing the intermediate optimization variables and kernel matrix using a lot of RAM.
4. **Time Complexity:** The time complexity of solving the quadratic optimization problem in SVMs is typically within range $O(n^2)$ and $O(n^3)$, where n is the number of training samples.

2.8. SVM and large dataset.

Using SVM with large datasets can be challenging due to computational complexity and memory requirements. The SGD-based SVMs, GPU acceleration, parallelization, Dimensionality reduction, and feature selection methods are utilized to overcome the scalability of SVMs for large datasets. The summary of these methods is summarized in Table II.

TABLE II
METHODS FOR OVERCOMING THE SCALABILITY OF SVMS FOR LARGE DATASETS

Methods	Approach	Advantages	Trade-offs
SGD-Based SVMs	Incremental optimization using small batches.	scalability and efficiency.	Noisy updates, suboptimal solutions
Parallelization	Allocate training across multi-processors	Operates large datasets efficiently	Demands distributed computing resources
GPU Acceleration	Use GPUs to parallelize computations.	Substantial speed-up for huge datasets	Needs GPUs and technical libraries.
Dimensionality reduction and feature selection	Decrease feature number.	Deflates computational cost.	Loss of information

2.9. Types of SVM Based on Functionality

SVM techniques can be functionally categorized based on their type and purpose. Below are the main types of SVM functions as shown in Table III.

TABLE III
TYPES OF SVM BASED ON FUNCTIONALITY

Function	Objective	Approach	Key Idea	Use Case Example
Binary Classification	Separate two classes using a hyperplane that maximizes the margin.	Optimize hyperplane $w \cdot x + b = 0$ to maximize margin. Use kernel functions (e.g., linear, polynomial, RBF) for nonlinear boundaries. Minimize $\frac{1}{2}\|\boldsymbol{w}\|^2$ subject to constraints.	Maximize margin between two classes.	Spam detection, medical diagnosis.
Multiclass Classification	Classify data into more than two classes.	**OneVsOne**: Train $\boldsymbol{k} * \frac{\boldsymbol{k-1}}{\boldsymbol{2}}$ binary classifiers for pairwise comparisons. **OneVsRest**: Train K binary classifiers, one per class. **Native Multiclass SVM**: Directly optimize multiclass objective.	Extend binary classification to multiple classes.	Handwritten digit recognition, image classification.
Multilabel Classification	Assign multiple labels to each data point.	Binary Relevance: Train a binary classifier for each label. Classifier Chains: Use predictions from previous classifiers as features. Adapted Algorithms: Modify loss function for multilabel tasks.	Assign multiple labels to a single instance.	Stock price prediction, housing price estimation.
Regression (SVR)	Predict continuous values .	Find function $f(x) = w \cdot x + b$ that deviates from true values by at most ϵ. Use slack variables for errors outside ϵ - tube. Apply kernel functions for nonlinear relationships.	Predict continuous values with a margin of tolerance.	Large-scale datasets, real-time applications.

IV. SVM AND SOME COMMON MACHINE LEARNING TECHNIQUES

Support Vector Machines (SVMs) provide a robust and interpretable classification technique that excels in processing high-dimensional data. Nevertheless, it is worth noting that alternative algorithms such as Logistic Regression, Decision Trees, Random Forests, and Neural Networks may be more appropriate for addressing the particular problem and data attributes, as indicated in Table IV. Assessing various algorithms on a dataset is crucial for making a well-informed conclusion regarding the most efficient performance of a particular activity.

SVMs are renowned for their effectiveness in high-dimensional domains and their ability to mitigate some issues caused by the curse of dimensionality.

TABLE IV
SVM AND SOME COMMON MACHINE LEARNING TECHNIQUES

	Strength	Limitation
SVM	**Non-linearity:** This can be efficiently addressed by employing kernel functions to handle non-linear data. **High dimensionality**: Effective in feature spaces with a large number of dimensions. **Resistant to extreme values**: Logistic regression is more vulnerable to outliers in the data than other statistical methods. **Enhanced computational efficiency**: SVMs exhibit slower training times, particularly when dealing with extensive datasets.	**The interpretability** of the model is challenging due to the intricate decision boundaries in a high-dimensional space, particularly when employing kernels. **Cost of computation**: Training SVMs can be computationally demanding, particularly when dealing with extensive datasets using kernel approaches. **Parameter tuning** poses a significant challenge when selecting the appropriate kernel function and its corresponding hyperparameter.
Logistic Regression	Its computing efficiency during training and its ability to effectively handle datasets are characterized by a high number of features and a relatively small number of data points. Offers interpretability by providing insights into the relationship between features and the target variable through the model coefficients. The calibration process allows for the straightforward determination of probability estimates pertaining to class membership.	Restricted to linear data: Optimal for data that can be separated linearly. May exhibit suboptimal performance when dealing with intricate, non-linear associations. The model's performance can be considerably affected by outliers present in the data.
Random forest	The ensemble nature of random forests generally leads to good accuracy in a wide range of classification and regression applications. By integrating numerous decision trees, variation can be decreased and generalization enhanced. Random forests have shown success in handling datasets with many different qualities. Using a random feature selection approach at every decision tree split help to prevent overfitting. Though not as clearly understandable as single decision trees, random forests allow one to obtain feature relevance scores to identify the variables most likely to contribute to the predictions of the model. Interpretability: Although not as easily comprehensible as individual decision trees, it is possible to get feature importance scores from random forests in order to ascertain the attributes that provide the greatest contributions to the model's predictions.	The interpretability of random forests may be comparatively lower than that of simpler models such as logistic regression, mostly due to the intricate ensemble of decision trees involved. Training random forests can have a significant computational cost, especially considering large datasets. The operation involves training several decision trees. Blackbox nature: While feature importance provides some understanding, grasping the complex internal mechanics of the ensemble can prove challenging.
Artificial Neural Networks	(ANNs) have the potential to learn intricate patterns and correlations from extensive datasets, rendering them well-suited for jobs that challenge conventional algorithms. Non-linearity: In contrast to less complex models such as logistic regression, artificial neural networks (ANNs) can accurately capture non-linear associations between features and the target variable. Feature extraction: ANNs can autonomously acquire pertinent features from unprocessed data, preventing manual feature engineering in some scenarios. Well-trained ANNs are effective in real-world prediction challenges because they can generalize well to unknown data.	**Black box nature**: Artificial neural networks' (ANNs') inner mechanisms can be complex and difficult to understand. This can have a disadvantage when interpretability is of great relevance. **Computational cost**: Training ANNs—especially large and deep ones—may be time-consuming and expensive. This work calls for large datasets and significant computational resources. ANNs are highly dependent on both the quality and quantity of data available. Insufficient, noisy, or biased training data can lead to poor performance. **Overfitting** occurs when artificial neural networks (ANNs) are not adequately regularized, leading to worse performance on unseen data.

V. CONCLUSION

Support Vector Machine sometimes known as SVM is an example of a typical form of supervised learning algorithm that was designed expressly for classification problem. The basic objective is to locate the hyperplane that most effectively divides data points into those belonging to distinct classes while simultaneously increasing the margin. The margin is the distance that separates the hyperplane from the data points that are closest to it, which are referred to as the support vectors. SVM is one of the most fundamental approaches to machine learning, which are renowned for their robust theoretical underpinning and their capacity to generate appropriate decision boundaries for categorization. Because they make use of kernel functions, they are able to effectively manage complex and non-linear data, which enables them to be adapted to a wide variety of applications, including text classification and picture recognition.

REFERENCES

[1] Abdullah DM, Abdulazeez AM. Machine learning applications based on SVM classification a review. Qubahan Academic Journal. 2021 Apr 28;1(2):81-90. DOI: 10.48161/qaj.v1n2a50

[2] Hussain SM, Jilani MN, Haq MU, Shaikh MS. A Machine Learning Approach to Arabic Phoneme Classification through Ensemble Techniques. In 2024 5th International Conference on Advancements in Computational Sciences (ICACS) 2024 Feb 19 (pp. 1–7). IEEE. Artificial Neural Networks Random forest Logistic Regression

[3] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). DOI: 10.1007/s42979-021-00592-x

[4] Prasad SC, Anagha P, Balasundaram S. Robust pinball twin bounded support vector machine for data classification. Neural Processing Letters. 2023 Apr; 55(2):1131-53.

[5] Otchere DA, Ganat TO, Gholami R, Ridha S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. Journal of Petroleum Science and Engineering. 2021 May 1; 200:108182.

[6] Zulfiqar M, Kamran M, Rasheed MB, Alquthami T, Milyani AH. Hyperparameter optimization of support vector machine using adaptive differential evolution for electricity load forecasting. Energy Reports. 2022 Nov 1; 8:13333-52.

[7] Alwahedi F, Aldhaheri A, Ferrag MA, Battah A, Tihanyi N. Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models. Internet of Things and Cyber-Physical Systems. 2024 Jan 3.

[8] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer, 1998.

[9] Roushangar K, Ghasempour R. Supporting vector machines. In Handbook of Hydro informatics 2023 Jan 1 (pp. 411–422). Elsevier.

[10] Roy A, Chakraborty S. Support vector machine in structural reliability analysis: A review. Reliability Engineering & System Safety. 2023 May 1; 233:109126.

[13] B.-Y. Sun, D.-S. Huang, H.-T. Fang, Lidar signal denoising using least-squares support vector machine, IEEE Signal Process. Lett. 12 (Feb 2005) 101–104.

[14] P. Chen, B. Wang, H.-S. Wong, D.-S. Huang, Prediction of protein b-factors using multi-class bounded SVM, Protein Peptide Lett. 14 (Feb 2007) 185–190.

[15] X. Liang, L. Zhu, D.-S. Huang, Multi-task ranking SVM for image segmentation, Neurocomputing 247 (Jul 2017) 126–136.

[16] J. Cervantes, F. García Lamont, A. López-Chau, L. Rodríguez Mazahua, J. Sergio Ruíz, Data selection based on decision tree for SVM classification on large data sets, Appl. Soft Comput. J. (2015).

[17] V. A. Naik, A. A. Desai, online handwritten gujarati character recognition using svm, mlp, and k-nn, in: 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017, pp. 1–6.

[18] J. L. Raheja, A. Mishra, A. Chaudhary, Indian sign language recognition using svm, Pattern Recognin. Image Anal. 26 (Apr 2016) 434–441.

[19] Hamel LH. Knowledge discovery with support vector machines. John Wiley & Sons; 2011 Sep 20.

[20] Brereton RG, Lloyd GR. Support vector machines for classification and regression. Analyst. 2010; 135(2):230-67.

[21] Vapnik, V. N., & Chervonenkis, A. Y. (1964). A theory of learning from general examples. Proceedings of the 3rd Annual ACM Conference on Theory of Computing (pp. 151–160). ACM.)

[22] Vapnik VN. An overview of statistical learning theory. IEEE transactions on neural networks. 1999 Sep;10(5):988-99.

[23] Burges CJ. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 1998 Jun;2(2):121-67.

[24] Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020 Sep 30; 408:189-215.

[25] Soman KP, Loganathan R, Ajay V. Machine learning with SVM and other kernel methods. PHI Learning Pvt. Ltd.; 2009 Feb 2.

[26] Ghosh S, Dasgupta A, Swetapadma A. A study on support vector machine based linear and non-linear pattern classification. In2019 International Conference on Intelligent Sustainable Systems (ICISS) 2019 Feb 21 (pp. 24–28). IEEE.

[27] Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020 Sep 30; 408:189-215. DOI: 10.1016/j.neucom.2019.10.118

[28] Steinwart I, Christmann A. Support vector machines. Springer Science & Business Media; 2008 Sep 15.

[29] Hubert K, Elisha B. Support Vector Machines (SVM): Explaining SVM and its application in regression tasks for sales forecasting. 2023

[30] Ishfaq K, Sana M, Ashraf WM. Artificial intelligence – built analysis framework for the manufacturing sector: performance optimization of wire electric discharge machining system. The International Journal of Advanced Manufacturing Technology. 2023 Oct; 128(11-12):5025-39.

[31] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995). DOI: 10.1007/BF00994018

[32] Hofmann T, Schölkopf B, Smola AJ. A review of kernel methods in machine learning. Mac-Planck-Institute Technical Report. 2006 Dec 14; 156.

[33] radhan A. Support vector machine-a survey. International Journal of Emerging Technology and Advanced Engineering. 2012 Aug; 2(8):82-5.

[34] Martin-Barragan B, Lillo R, Romo J. Interpretable support vector machines for functional data. European Journal of Operational Research. 2014 Jan 1; 232(1):146-55.

[35] Menon AK. Large-scale support vector machines: algorithms and theory. Research Exam, University of California, San Diego. 2009 Feb 27; 117.

[36] Gamal M, Abbas H, Sadek R. Hybrid approach for improving intrusion detection based on deep learning and machine learning techniques. InProceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020) 2020 (pp. 225–236). Springer International Publishing.

[37] Byvatov E, Schneider G. Support vector machine applications in bioinformatics. Applied bioinformatics. 2003 Jan 1;2(2):67-77.

[38] Dino HI, Abdulrazzaq MB. Facial expression classification based on SVM, KNN and MLP classifiers. In2019 International Conference on Advanced Science and Engineering (ICOASE) 2019 Apr 2 (pp. 70–75). IEEE.

[39] Chauhan VK, Dahiya K, Sharma A. Problem formulations and solvers in linear SVM: a review. Artificial Intelligence Review. 2019 Aug 15; 52(2):803- 55.

[40] Battineni G, Chintalapudi N, Amenta F. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Informatics in Medicine Unlocked. 2019 Jan 1; 16:100200.

[41] Bhavan A, Chauhan P, Shah RR. Bagged support vector machines for emotion recognition from speech. Knowledge-Based Systems. 2019 Nov 15; 184:104886.

[42] Gu J, Lu S. An effective intrusion detection approach using SVM with naïve Bayes feature embedding. Computers & Security. 2021 Apr 1; 103:102158.

[43] Seeja RD, Suresh A. Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). Asian Pacific journal of cancer prevention: APJCP. 2019; 20(5):1555.

[44] Patil RS, Biradar N. Automated mammogram breast cancer detection using the optimized combination of convolutional and recurrent neural network. Evolutionary intelligence. 2021 Dec; 14:1459-74.

[45] Hu R, Zhu X, Zhu Y, Gan J. Robust SVM with adaptive graph learning. World Wide Web. 2020 May; 23:1945-68.

[46] Pisner DA, Schnyer DM. Support vector machine. In Machine learning 2020 Jan 1 (pp. 101-121). Academic Press.

[47] Gong W, Chen H, Zhang Z, Zhang M, Wang R, Guan C, Wang Q. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. Sensors. 2019 Apr 9; 19(7):1693.

[48] Jiang F, Lu Y, Chen Y, Cai D, Li G. Image recognition of four rice leaf diseases based on deep learning and support vector machine. Computers and Electronics in Agriculture. 2020 Dec 1; 179:105824.

[49] Vijayalakshmi A. Deep learning approach to detect malaria from microscopic images. Multimedia Tools and Applications. 2020 Jun; 79:15297-317.

[50] Chandra MA, Bedi SS. Survey on SVM and their application in image classification. International Journal of Information Technology. 2021 Oct; 13:1-1.

[51] Huang W, Liu H, Zhang Y, Mi R, Tong C, Xiao W, Shuai B. Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. Applied Soft Computing. 2021 Sep 1; 109:107541.

[52] Roman I, Santana R, Mendiburu A, Lozano JA. In-depth analysis of SVM kernel learning and its components. Neural Computing and Applications. 2021 Jun; 33(12):6575-94.

[53] Kaur P, Singh G, Kaur P. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. Informatics in Medicine Unlocked. 2019 Jan 1; 16:100151.

[54] Liu B, Li CC, and Yan K. DeepSVM-fold: protein fold recognition by combining support vector machines and pairwise sequence similarity scores generated by deep learning networks. Briefings in bioinformatics. 2020 Sep; 21(5):1733-41.

Mohammed Jabardi received the B.Sc. degree in Computer Science from the University of Technology, Baghdad, Iraq, in 1990, the M.Sc. degree in Computer Science from Jamia Hamdard University, New Delhi, India, in 2014, and the Ph.D. degree in Information Technology from the University of Babylon, Babylon, Iraq, in 2020.

He is currently an Assistant Professor with the Department of Software, College of Education, University of Kufa, Najaf, Iraq. His research interests include data mining, machine learning, and artificial intelligence applications. He has over six years of experience in scientific and academic research at the Information Technology Research Center, University of Kufa. Since 2017, he has been teaching courses on AI applications, programming, compiler construction, data mining, and machine learning for both undergraduate and master's students. He has also been supervising master's students since 2021.

Dr. Jabardi has served as a scientific reviewer for numerous Scopus-indexed journals, including AIJ, IJCSM, IJEECS, BEEI, and IJECE, and has contributed to the organization of academic conferences such as ICECCME 2023.

A Siamese-based Approach to Improve Parkinson's Disease Detection and Severity Prediction from Speech Using X-Vector Embedding

Attila Zoltán Jenei[1], Réka Ágoston[1], and István Valálik[2]

***Abstract*—Parkinson's disease is incurable and is considered one of the most common neurological diseases. It is a progressive disease, which highlights the importance of early detection. Machine learning-based diagnostic support is desirable since the diagnosis is based on history, visual inspection, and drug tests. Speech is presumed to be one of the promising biomarkers that can predict the state of the disease. Combining speech data with deep learning feature extraction in Siamese-based architecture may improve the detection compared with direct regression with acoustic and prosodic features. Read text-based speech samples were acquired from 98 patients with Parkinson's disease and 107 healthy participants. Feature vectors were extracted with pre- trained x-vector embedding and were used directly with a support vector regressor in a nested cross-validation setup (baseline approach). Furthermore, pairs were allocated, and difference vectors were calculated. These difference vectors were then used to train support vector regressor models in nested cross-validation (Siamese-based approach). Severity predictions and classification were performed with the outcomes. The Siamese-based setup outperformed the baseline approach both in regression and classification metrics. The relative improvement in root mean square error is 14.4%, and the Pearson correlation is 12.5% at best. After the classification, the relative improvement is 6.0% in sensitivity, 3.0% in specificity, and 4.5% in accuracy. Furthermore, comparing the test sample to not only one but multiple others decreases the average standard deviation of the predicted severity by 16.5% in relative value. Changing only the architecture of the traditional examination setup to a Siamese-based approach may increase the performance of the models.**

***Index Terms*—Classification, Deep-learning, Parkinson's Disease, Siamese Network**

I. Introduction and Literature Study

Parkinson's disease (PD) is one of the most common neurological disorders, which also manifests in movement disabilities. PD affects mainly the aging population and has a prevalence of 1% after 60 years [1]. The importance of detecting the disease in an early stage is its progressive nature and because it is incurable, according to recent clinical knowledge. The development of the disease is characterized by the loss of dopamine-producing cells with the appearance of Lewy protein aggregates [2].

The diagnosis of PD is based on the patient's history and examination. The cardinal symptoms are resting tremor, bradykinesia, and rigidity that started asymptotically [3]. The patient may have small handwriting, a masked face, and a soft voice. Vocal disorders are prominent symptoms as they manifest in 90% of PD cases [4]. These motor symptoms are less notable by visual observation at the early stage, which stresses the need for such a diagnostic support system.

The speech affected by PD may be monotonic, have less intensity, and can include sudden stops and starts. Tremors also can appear in the phonation. Speech perceptive analysis is part of the Unified Parkinson's Rating Scale (UPDRS), part III: Motor Examination, where the irregularity can be rated between 0 and 4 [5]. Next to the perceptual analysis, objective methods using descriptive features from the speech can also facilitate the diagnosis in clinical settings. Especially, speech can be acquired non-invasively and analyzed even on a mobile device [6].

According to [7], speech-related studies can be categorized into the following four aspects: a) phonatory, b) articulatory, c) prosodic, and d) cognitive-linguistic studies. Generally, all of them can be used to build a diagnostic support system; however, b) and c) are the most commonly applied in this area. The most common features are jitter, shimmer, noise ratio, pitch, intensity, articulation rate, or pauses. With these descriptors, the machine learning algorithms (such as support vector machine (SVM) or k-nearest neighbor (k-NN)) can reach up to 97% accuracy [8], [9], [10].

In addition to the manual features, deep learning-based feature extraction is also applied to maximize the disease representation in the features. As medical data is hard to acquire in large amounts, transfer learning from another domain is a form of use. These models are initially trained on a large dataset for general purposes, such as speaker recognition. The x-vector and e-capa time-delay neural networks are based on speaker recognition and are still applied in PD detection [11], [12]. The transfer learning-based deep learning models can improve the detection and make the process robust (avoiding overfitting). In [13], Sztahó and his colleagues classified 85 PD and 85 healthy individuals using an x-vector approach with probabilistic linear

[1] Department of Telecommunications and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary (E-mail: jenei.attila.zoltan@vik.bme.hu, orcid id: 0000-0003-1007-9907, reka10007@gmail.com).

[2] Department of Neurosurgery, St. John's Hospital, Budapest, Hungary (E-mail: valalik@parkinson.hu).

DOI: 10.36244/ICJ.2025.1.9

discriminant analysis (PLDA) and achieved 84.1% accuracy. The severity of the disease was measured with the Hoehn and Yahr (H&Y) scale [14]. This test assigns a number to the patient between 0 and 5, where 0 means no deviation from normal functioning while 5 means the patient is in bed or a wheelchair due to the disease. This scale is not linear, so an H&Y score of 2 does not mean two times as severe symptoms as the H&Y score of 1. Using this scale, the patients had a mean score of 2.7 with a standard deviation of 1.1. The participants read the "The North Wind and the Sun" tale.

Another approach to overcoming limited medical data is to use distance-based solutions where the pairs of samples could be allocated with higher freedom. One of its applications is in the siamese networks. Bhati and his colleagues [15] used Long Short-Term Memory (LSTM)-based siamese networks to learn better PD representation. Then, they trained classifiers to detect PD with the features resulting from the siamese part. Shalaby and Belal [16] used Siamese networks to enhance the data clustering before classification. They improved the detection by 9.5% relative accuracy.

In the [17] study conducted in 2017, the authors used speech samples from 51 PD patients and 27 healthy participants in Hungarian language. The severity was measured with the H&Y scale. The mean severity was 2.58, with a standard deviation of 0.9. The speech samples included the read version of "The North Wind and the Sun" tale. Prosodic features were used with several regression and classification algorithms. The authors achieved 77.8% sensitivity and 83.6% accuracy with classification, 1.052 RMSE, and 0.73 Pearson correlation with SVM and support vector regression (SVR).

The authors repeated the study [18] two years later. They used samples from 55 PD and 33 healthy participants. The mean severity was almost the same as in the previous study. The prosodic features were extended with various acoustic ones. They reached 84.8% sensitivity, 81.8% specificity, and 83.5% accuracy with SVM. Regression was conducted with SVR, 1.071 RMSE, and 0.72 R^2 (the square of the Pearson correlation coefficient), which also resulted from that study.

We introduced sample pairs and extracted features with x-vector technology based on these studies. Using the difference in the feature vectors, we examined PD's classification and severity regression in the Hungarian language. Our goal is to examine how PD can be detected when we switch from individual samples (and feature vectors) to sample pairs (and feature vector differences) in the same process.

After the *Introduction*, we will present the applied materials and the examination methods in the *Methodology*. We will present the regression and classification results in the *Results* section. Finally, we will discuss the findings and conclude the work in the *Discussion* and *Conclusion* sections.

II. METHODOLOGY

The setup of the experiments can be seen in Fig. 1, where two approaches are detailed. The Baseline approach includes the x-vector feature extraction from the Hungarian Parkinson Speech Dataset (HPSD) and the severity prediction with regressor algorithms. The Siamese-inspired approach takes two samples from the dataset, extracts x-vector features, and calculates the vector difference. Then, this dissimilarity is used to predict severity. In the following subsections, we will detail all components of the approaches.

A. Hungarian Parkinson Speech Dataset

PD patients were recorded at the Semmelweis and Virányos Clinic, Budapest, alongside healthy participants as a control group. 42 females and 56 males were in the PD class with a 65.4 average age and 8.4 standard deviation. 70 females and 37 males were in the healthy control (HC) class with 45.8 average age and 17.7 standard deviation. The sex distribution and age between the two classes appeared significant, with a p-value lower than 0.05 (sex by chi-square test, age by Mann-Whitney U test). These may affect the results when compared to others in the literature. However, we propose a baseline to highlight our findings so the new technique will be compared with a traditional one with the same dataset.

The diagnosis and the severity estimation were done by the neurosurgeon doctor using the H&Y scale. These were made using history, visual examination, and drug tests. The mean severity is 2.8 H&Y with a 0.9 standard deviation. The distribution of the

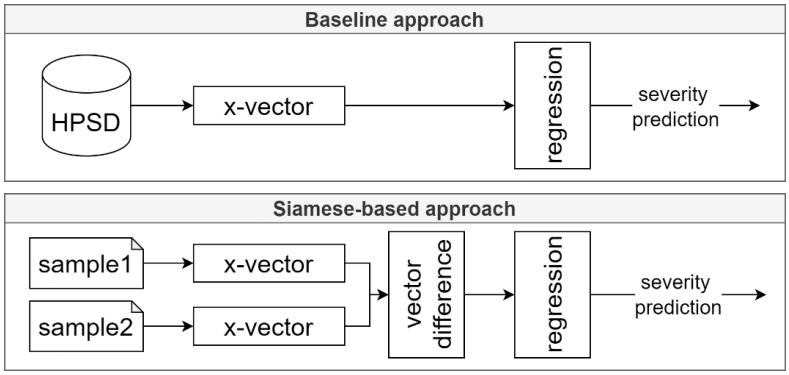


Fig. 1. The Baseline and Siamese-based examination process. HPSD is the dataset, sample1 and sample2 are paired subjects from the dataset.

H&Y severity in the dataset can be seen in Table 1. The HC samples were in the 0 stages of the H&Y scale.

TABLE I
DISTRIBUTION OF THE PATIENTS ACCORDING TO THEIR H&Y SEVERITY SCORE.

	1	1.5	2	2.5	3	4
PD	10	1	18	12	31	26

The speech task was to read the "*The North Wind and the Sun*" tale, which resulted in around one-minute-long recording per participant. All participants were native Hungarian speakers, and the recordings were done in Hungarian language. The samples were acquired with a clip-on microphone on 16 kHz sampling frequency and 16-bit quantization.

All participants were informed in advance about the research and the use of the samples and metadata. Data acquisition did not mean any harm or risk to the participants. Informed consent was collected from the people involved, and they had (and have) the option to change their minds anytime.

B. Preprocessing and Feature Extraction

The speech samples were normalized to peak value before feature extraction. After that, feature extraction was done with x-vector embedding originating from the speaker recognition applications [19]. The x-vector is a time-delay neural network (its architecture is in Fig. 2) that observes not only the speech frame at time point *t* but its surroundings in the frame level layers. The frame refers to the (sliding) time window, where features like filterbanks can be defined. These layers learn the local and speaker-specific patterns.

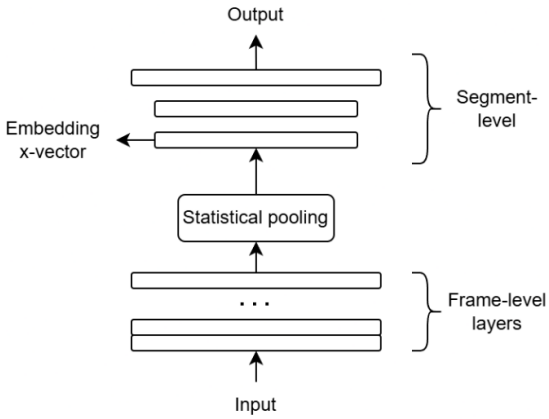


Fig. 2. The x-vector embedding architecture, which parts from the input layer are the frame level layers, statistical pooling, and segment level layers before the output.

Following the frame-level layers, the statistical pooling calculates the mean and standard deviation after aggregating the information by frames. This creates a fixed-length representation of the entire utterance Then, the speaker discriminative features are learned in the segment level from where the x-vector can be extracted with a dimension of 512.

The embedding was implemented through the SpeechBrain toolkit [20], which stores the pre-trained models on the Huggingface website. We used this pre-trained version of the model without further training. The initial training was done on the training data of Voxceleb1 and Voxceleb2 datasets (in English). From the resulting feature vector, the top 65 features were selected with the Random Forest algorithm.

Next, pairs were allocated for the Siamese-based approach to use the features directly with the predictive algorithms (Baseline approach). HC-HC (similar) and HC-PD (dissimilar) pairs were created randomly without repeating already allocated pairs. As a result, 205 pairs for the similar category and 205 for the dissimilar category were examined. 200 pairs were of the same sex, while 210 pairs were of the opposite sex. In this approach, age was not considered during the pair allocation.

C. Severity Prediction and Classification

To estimate the severity of PD in the H&Y scale, we employed a regression method using the Support Vector Regression (SVR) algorithm [21]. The decision was based on the study [18] where the SVR was the prominent algorithm to predict PD severity in the Hungarian language.

The *C*, *gamma*, and *epsilon* parameters were optimized for nested cross-validation. The optimization includes an outer 10-fold cross-validation setup where 10% (one fold) was separated as an independent test set, and the rest was used in the inner 5-fold cross-validation loop. In this inner loop, 80% of the remaining data was used for training and 20% for optimization.

The target severity was normalized between 0 and 1 before inputting them into the predictive models. As a result, the output was generated between 0 and 1, which was up-scaled to the original H&Y scale. With these outputs and the original scores, a linear regression was performed, measuring the mean absolute error (*eq. 1*) and the root mean square error (*eq. 2*). In these metrics, y_i is the original severity of the *i*-th sample, $\hat{y}_i$ is the predicted score of the *i*-th sample, and *n* is the number of samples.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{2}$$

The R^2 value shows how linearly the estimated score fits the original severity score (*eq. 3*). Its value is between 0 and 1, within which a value close to 1 shows a better fit than a value close to 0. The $\overline{y}_i$ in the denominator is the mean of the predicted scores.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{3}$$

We also created classifications with the output score by drawing a decision threshold at H&Y stage 1. Below the score 1, the samples were classified as HC and above as PD. The classification is correct in True Positive (TP) and True Negative (TN) cases and not correct in False Positive (FP) and False Negative (FN) cases.

From these, metrics like sensitivity (*sens – eq. 4*), specificity (*spec – eq. 5*), and accuracy (*acc – eq. 6*) were derived.

$$sens = \frac{TP}{FN + TP} \tag{4}$$

$$spec = \frac{TN}{TN + FP} \tag{5}$$

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

D. Experimental Cases

In this study, we created a Baseline approach, where x-vector embeddings were extracted from the speech samples, and then in a nested cross-validation setup, SVR models were trained and tested. With the output, regression, and classification were performed.

After that, a Siamese-based approach was made by pairing the samples. The x-vector embeddings were extracted from both samples in the pairs similarly, and then the difference between the two vectors was calculated. This vector was the input to the SVR models for regression and classification. Within this approach, we had two cases: 1) create only ONE pair with the selected test sample, and 2) create TEN pairs with the selected test sample and average the predicted scores. Since the train/test split was done after the pair allocation, one pair was only in the train, validation, or test sets. One speaker could be in more sets but not in the same pair.

The *Results* will be divided according to these three cases: a) baseline, b) Siamese approach with one pair, and with ten pairs.

III. RESULTS

A. Baseline approach

The results of the predicted and the original severity scores are presented in the left plot (a part) of Fig. 3 with blue dots for the Baseline approach. The red line represents the perfect mapping of the original scores. Table 2 includes the metrics presented in the next section for all three cases.

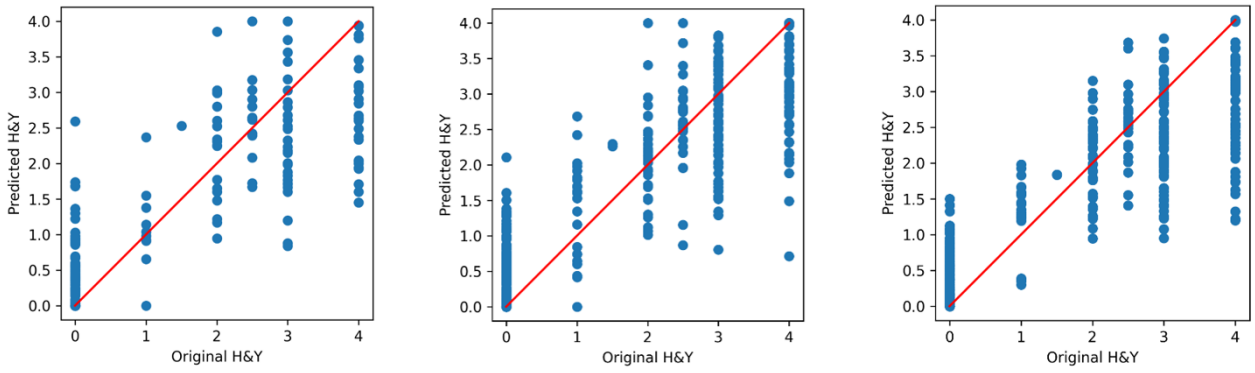
The MAE and RMSE resulted from the predicted and original scores of 0.59 and 0.85. The R^2, which describes the fit of the predicted data to the original, is 0.69. That means that the original score accounts for 69% of the change in the estimated score. Checking the standard deviation of the predicted scores at different H&Y stages, stage 0 had the lowest value at 0.46, and stage 3 had the highest at 0.79. The standard deviation for the other stages is 0.58 (stage 1), 0.76 (stage 2), and 0.73 (stage 4). This deviation can be seen in the figure where the points at stage 3 are scattered farther from the red line. The model predicts stages 1, 2, and 2.5 symmetrically while stage 0 is over-, and stage 4 is underestimated. Also, stage 3 has a slight bias toward lower scores.

After transforming these scores to labels (performing the classification), 91.6% sensitivity, 92.4% specificity, and 92.0% accuracy scores were achieved. The ratio between the positive and negative classes also remains in the predictive scores, as sensitivity and specificity have almost the same values.

B. Siamese-based approach

The results of the Siamese-based approach can be seen in Fig. 3, where the middle plot (b part) presents the one-pair case, and the right side (c part) presents the average of the ten-pair case.

Comparing the test sample with only one other sample (using one pair) resulted in 0.51 MAE, 0.73 RMSE, and 0.78 R^2 metrics. The standard deviations are 0.41, 0.70, 0.62, 0.69, and 0.70 at the stages of the severity scale from 0 to 4, respectively. Symmetrical distribution can be seen at stages 1, 2, and 2.5.



a) Baseline approach b) Siamese approach with ONE pair c) Siamese approach with TEN pairs

Fig. 3. Original and predicted severity scores
with the Baseline approach (a) and with Siamese approach (b – where one pair, c – where ten pairs were allocated.).

Stage 0 is over-, and stages 3 and 4 are underestimated by the models. After calculating classification metrics, 95.1% sensitivity, 91.7% specificity, and 93.4% accuracy were achieved.

When we compared one test sample to ten other test samples and averaged the predicted scores, 0.55 MAE, 0.77 RMSE, and 0.75 R^2 were the results. The standard deviations were 0.31, 0.50, 0.55, 0.69, and 0.72 according to the stages from 0 to 4. Here, stages 2 and 2.5 seem symmetrical, while stages 0,1 are over-, and 3 and 4 are underestimated. After the classification, 97.1% sensitivity, 95.1% specificity, and 96.1% accuracy were observed.

Comparing the three experiment cases, *no significant difference* can be noted between the predicted severity scores or the classification results with Mann-Whitney tests next to 5% significant level. Furthermore, the difference was calculated between the predicted and the original severity score separately for the male-male/female-female and the female-male pairs. The mean difference was 0.04 for the pairs from the same gender and 0.009 for the pairs from the opposite gender. The standard deviation of the differences was around the same.

IV. Discussion

The results of the Baseline and Siamese-based approaches are summarized in Table 2, along with the metrics described above. Baseline refers to the Baseline approach, ONE pair to the one-pair version, and TEN pairs to the ten-pair version of the Siamese approach. Bold style marks the best values for each metric.

Comparing the regression metrics to the different cases, it can be seen that the Siamese-based approach decreased the error in the predictions as MAE and RMSE decreased. Both one-pair and ten-pair cases were better in metrics than the Baseline approach. However, the ten-pair version of the Siamese-based approach did not yield better than the one-pair version examining the regression metrics. A similar tendency can be seen with the R^2 as it is higher with the Siamese-based approach than the Baseline. However, the ten-pair version is not better than the one-pair version.

Comparing the classification results, the improvement is continuous. Sensitivity increased from 91.6% to 97.1%, while the specificity was slightly lower with one pair but increased with ten pairs. The accuracy improved from 92.0% to 96.1%.

TABLE II
SUMMARY TABLE OF THE BASELINE AND SIAMESE-BASED (ONE AND TEN-PAIR VERSIONS) APPROACHES.

	MAE	RMSE	R2	sens	spec	acc
Baseline	0.59	0.85	0.69	91.6%	92.4%	92.0%
ONE pair	**0.51**	**0.73**	**0.78**	95.1%	91.7%	93.4%
TEN pair	0.55	0.77	0.75	**97.1%**	**95.1%**	**96.1%**

Comparing our results with similar studies [17], [18], it can be seen that the results achieved in this study outperformed them. Even the Baseline method reached lower error and higher classification metrics. This could be due to the extended database and the nested cross-validation setup for the optimization.

The results could improve compared to the Baseline method by facilitating the Siamese-inspired approach. This result highlights the possible benefit of increasing the number of input data by creating pairs. However, comparing someone to only one other person is not always reliable. We compared one test sample to ten others to overcome this limitation and averaged the predicted scores. This technique seemingly did not improve the regression metrics but did improve the classification. This is because the average standard deviation decreased from 0.62 to 0.55 by averaging the pair's predictions. With this, the samples that were misclassified with only one pair were classified correctly by the other 9 pairs. 188 TN and 195 TP classification were done with the one-pair case, while 195 TN and 199 TP with the ten-pair case.

Comparing our results to the existing solutions in the international literature is cumbersome due to the different datasets with different languages. However, we believe that our results fit into this international level. Furthermore, the x-vector was used here without further training; it was initially trained with English samples. The usage of such model is promising since we could use it in the Hungarian language without the language's specifications.

The present study concentrated on the the x-vector algorithm. Nevertheless, the employment of alternative algorithms in speech technology (like transformers) has the potential to enhance the findings of this study.

VI. Conclusion

PD is one of the most common neurological disorders that is not curable according to recent clinical knowledge. Early detection is important to maintain the quality of life and slow disease progression. Speech is one of the promising biomarkers that can capture the pre-motor symptoms of the disease. Many studies used acoustic and prosodic features to describe the speech and used them with statistical methods or machine learning algorithms to distinguish between PD and other groups. Next to the manual features deep learning-based feature extraction and detection became prevalent due to the more detailed representation. However, these solutions require a huge amount of data.

In this study, we explored the possibilities of the Siamese-based architecture with PD patients and HC participants using read-text-based speech. The results indicated that the Siamese-based approach outperforms the Baseline in regression and classification. Pairing the sample with more than one other sample decreases the standard deviation of predictions and improves the classification further. The results fit the literature since they outperform studies with similar datasets and are also comparable with similar studies that use different languages or databases.

It should be noted that this study used the x-vector trained on English utterances while the samples in scope were in Hungarian. However, the results show high performance without specifying the embedding to Hungarian. This is also promising for language-independent solutions. Additional

analysis may be required to discover possible influencing effects like age or sex (even though the results showed no significant influence of these effects).

Acknowledgments

The work was funded by the National Research, Development and Innovation Office – NKFIH, project K143075.

REFERENCES

[1] O.-B. Tysnes and A. Storstein, 'Epidemiology of Parkinson's disease', *J Neural Transm*, vol. 124, no. 8, pp. 901–905, Aug. 2017, DOI: 10.1007/s00702-017-1686-y.

[2] E. Tolosa, A. Garrido, S. W. Scholz, and W. Poewe, 'Challenges in the diagnosis of Parkinson's disease', *The Lancet Neurology*, vol. 20, no. 5, pp. 385–397, May 2021, DOI: 10.1016/S1474-4422(21)00030-2.

[3] P. Rizek, N. Kumar, and M. S. Jog, 'An update on the diagnosis and treatment of Parkinson disease', *CMAJ*, vol. 188, no. 16, pp. 1157–1165, Nov. 2016, DOI: 10.1503/cmaj.151179.

[4] G. Solana-Lavalle and R. Rosas-Romero, 'Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation', *Biomedical Signal Processing and Control*, vol. 66, p. 102 415, Apr. 2021, DOI: 10.1016/j.bspc.2021.102415.

[5] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, 'Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation', *J Neural Transm*, vol. 124, no. 3, pp. 303–334, Mar. 2017, DOI: 10.1007/s00702-017-1676-0.

[6] T. J. Wroge, Y. Ozkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, 'Parkinson's Disease Diagnosis Using Machine Learning and Voice', in *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Philadelphia, PA: IEEE, Dec. 2018, pp. 1–7. DOI: 10.1109/SPMB.2018.8615607.

[7] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, 'Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects', *Biomedical Signal Processing and Control*, vol. 66, p. 102 418, Apr. 2021, DOI: 10.1016/j.bspc.2021.102418.

[8] O. Yaman, F. Ertam, and T. Tuncer, 'Automated Parkinson's disease recognition based on statistical pooling method using acoustic features', *Medical Hypotheses*, vol. 135, p. 109 483, Feb. 2020, DOI: 10.1016/j.mehy.2019.109483.

[9] A. Benba, A. Jilbab, A. Hammouch, and S. Sandabad, 'Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease', in *2015 International Conference on Electrical and Information Technologies (ICEIT)*, Marrakech, Morocco: IEEE, Mar. 2015, pp. 300–304. DOI: 10.1109/EITech.2015.7163000.

[10] K. Wu, D. Zhang, G. Lu, and Z. Guo, 'Learning acoustic features to detect Parkinson's disease', *Neurocomputing*, vol. 318, pp. 102–108, Nov. 2018, DOI: 10.1016/j.neucom.2018.08.036.

[11] L. Moro-Velazquez, J. Villalba, and N. Dehak, 'Using X-Vectors to Automatically Detect Parkinson's Disease from Speech', in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain: IEEE, May 2020, pp. 1155–1159. DOI: 10.1109/ICASSP40776.2020.9053770.

[12] L. Jeancolas *et al.*, 'X-Vectors: New Quantitative Biomarkers for Early Parkinson's Disease Detection From Speech', *Front. Neuroinform.*, vol. 15, p. 578 369, Feb. 2021, DOI: 10.3389/fninf.2021.578369.

[13] D. Sztahó, A. Z. Jenei, I. Valálik, and K. Vicsi, 'The Effect of Speech Fragmentation and Audio Encodings on Automatic Parkinson's Disease Recognition', *JBiSE*, vol. 15, no. 01, pp. 6–25, 2022, DOI: 10.4236/jbise.2022.151002.

[14] R. Bhidayasiri and D. Tarsy, 'Parkinson's Disease: Hoehn and Yahr Scale', in *Movement Disorders: A Video Atlas*, in Current Clinical Neurology., Totowa, NJ: Humana Press, 2012, pp. 4–5. DOI: 10.1007/978-1-60327-426-5_2.

[15] S. Bhati, L. M. Velazquez, J. Villalba, and N. Dehak, 'LSTM Siamese Network for Parkinson's Disease Detection from Speech', in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Ottawa, ON, Canada: IEEE, Nov. 2019, pp. 1–5. DOI: 10.1109/GlobalSIP45357.2019.8969430.

[16] M. Shalaby, N. A. Belal, and Y. Omar, 'Data Clustering Improves Siamese Neural Networks Classification of Parkinson's Disease', *Complexity*, vol. 2021, pp. 1–9, Jun. 2021, DOI: 10.1155/2021/3112771.

[17] D. Sztaho, M. G. Tulics, K. Vicsi, and I. Valalik, 'Automatic estimation of severity of Parkinson's disease based on speech rhythm related features', in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen: IEEE, Sep. 2017, pp. 000 011–000 016. DOI: 10.1109/CogInfoCom.2017.8268208.

[18] D. Sztaho, I. Valalik, and K. Vicsi, 'Parkinson's Disease Severity Estimation on Hungarian Speech Using Various Speech Tasks', in *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Timisoara, Romania: IEEE, Oct. 2019, pp. 1–6. DOI: 10.1109/SPED.2019.8906277.

[19] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, 'Deep Neural Network Embeddings for Text-Independent Speaker Verification', *Interspeech 2017*, Aug. 2017, DOI: 10.21437/interspeech.2017-620

[20] M. Ravanelli *et al.*, 'SpeechBrain: A General-Purpose Speech Toolkit'. arXiv, Jun. 08, 2021. Accessed: Jun. 24, 2024. [Online]. Available: http://arxiv.org/abs/2106.04624

[21] M. Awad and R. Khanna, 'Support Vector Regression', in *Efficient Learning Machines*, Berkeley, CA: Apress, 2015, pp. 67–80. DOI: 10.1007/978-1-4302-5990-9_4.

Attila Zoltán Jenei was born in Debrecen, Hungary in 1995. He graduated from the Budapest University of Technology and Economics as a Biomedical Engineer (Master's Degree, 2020). Since January 2020, he has been a department engineer and Ph.D. student at the Laboratory of Speech Acoustics, Department of Telecommunications and Artificial Intelligence, Faculty of Electrical Engineering and Informatics. His research focuses on diagnostic support for Parkinson's disease with non-invasive medical data. He participated in the Student Research Societies of Budapest University of Technology and Economics and was awarded in 2017 and 2019. From 2021, he was the Vice President, and from 2023 to 2024, he was the President of the Department of Engineering Sciences in the National Association of Doctoral Students.

Réka Ágoston was born in Dunaújváros in 1999. She earned a bachelor's degree in biochemical engineering from Pannon University. She continued her studies at the Budapest University of Technology and Economics (BME), where she pursued a master's degree in biomedical engineering. For her master's thesis, she used machine learning to analyze speech data for Parkinson's disease severity detection, combining her passion for biomedical data and machine learning.

István Valálik MD., Ph.D., MSc. neurosurgeon and head physician of the Department of Neurosurgery at St. John's Hospital, Budapest, honorary associate professor at the University of Debrecen. In 2011 he defended PhD thesis "CT-guided stereotactic thermolesion and deep brain stimulation in the treatment of Parkinson's disease", in 2015 MSc in Health Services Management. His scientific interest focused on movement disorders, MR-tractography-based surgical planning, psychiatric surgery, acoustic and motion analysis. He developed a planning software for stereotactic brain surgery and portable neuro-navigation system. Since 2010 he is acting in the Executive Committee of the European Society for Stereotactic and Functional Neurosurgery (www.essfn.org). In 2013 he was awarded by the Hungarian Academy of Sciences for the book "Stereotactic and Functional Neurosurgery". In 2019 he participated in mission of successful surgical separation of Bangladeshi craniopagus twins.

21st International Conference on Network and Service Management

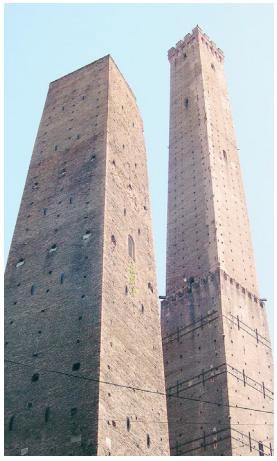
AI and Sustainability in the Future of Network and Service Management

27 - 31 October, 2025 - Bologna, Italy

CALL FOR PAPERS

The 21st International Conference on Network and Service Management (CNSM) is inviting authors to submit original contributions to network and service management research. CNSM is a selective single-track conference that covers all aspects of network and service management, pervasive systems, enterprises, and cloud computing environments. In particular, CNSM 2025 will focus on AI and Sustainability in the Future of Network and Service Management.

Papers accepted and presented at CNSM 2025 will be published as open access on the conference website and will be submitted for possible publication in IEEE Xplore. Authors of selected papers accepted for publication in the CNSM 2025 proceedings will be invited to submit an extended version of their papers to the IEEE Transactions on Network and Service Management journal.



Important Dates

Paper Submission:
16 June 2025
Rebuttal Period:
20-22 August 2025
Acceptance Notification:
27 August 2025
Camera Ready due:
14 September 2025

Technical Program Co-Chairs

Daphné Tuncer,
Institut Polytechnique de Paris, France
Alberto Egon Schaeffer-Filho,
Federal University of Rio Grande do Sul, Brazil
Davide Borsatti,
University of Bologna, Italy

General Co-Chairs

Walter Cerroni,
University of Bologna, Italy
Mauro Tortonesi,
University of Ferrara, Italy

Topics of Interest (but not limited to)

Technologies
- Communication Protocols
- Middleware
- Overlay Networks
- Peer-to-Peer Networks
- Technologies for Computing Continuum
- 5G/6G Networks
- Federated and Distributed Learning
- Generative AI and Large Language Models
- Green and Sustainable Networking
- Information Visualization
- Software-Defined Networking
- Monitoring and Measurements
- Multi-Access Edge Computing
- Network Function Virtualization
- Orchestration
- Operations and Business Support Systems
- Control and Data Plane Programmability
- Distributed Ledger Technology
- Digital Twins for Networks and Services
- Reinforcement Learning
- Secure and Dependable Networking

Service Management
- Multimedia Services
- Content Delivery Services
- Cloud/Edge Computing Services
- Data Services
- Internet Connectivity and Internet Access Services
- Internet of Things Services
- Security Services
- Context-Aware Services
- Information Technology Services
- Service Assurance

Functional Areas
- Fault Management
- Configuration Management
- Accounting Management
- Performance Management
- Security Management

Management Paradigms
- Centralized Management
- Hierarchical Management
- Distributed Management
- Federated Management
- Autonomic and Cognitive Management
- Policy- and Intent-Based Management
- Model-Driven Management
- Pro-active Management
- Energy-aware Management
- QoE-Centric Management

Methods
- Artificial Intelligence and Machine Learning
- Mathematical Logic and Automated Reasoning
- Optimization Theories
- Control Theory
- Probability Theory, Stochastic Processes, and Queuing Theory
- Evolutionary Algorithms
- Economic Theory and Game Theory
- Data Mining and (Big) Data Analysis
- Computer Simulation Experiments
- Testbed Experimentation and Field Trials
- Software Engineering Methodologies

Paper Submission

Authors are invited to submit original contributions that have not been published or submitted for publication elsewhere. Papers should be prepared using the IEEE 2-column conference style and are limited to 9 pages including references (full papers) or 5 pages including references (short papers). Papers must be submitted electronically in PDF format through the EDAS system (the submission link will be available soon).

Papers exceeding page limits, multiple submissions, and self-plagiarized papers will be rejected without further review. All other papers will get a thorough single-blind review process, followed by a rebuttal phase.

For further information, please check http://www.cnsm-conf.org/2025/.

ETFA 2025 Porto Portugal 9-12 September

30th IEEE International Conference on Emerging Technologies and Factory Automation

CALL FOR PAPERS

General Co-Chairs
Marina Indri
Luís Almeida
Technical Program Co-Chairs
Antonio Visioli
Mario de Sousa
Ex-officio (TCFA Chair)
Moris Behnam
Finances Co-Chairs
Paulo Portugal
TBD
Publications Co-Chairs
Pedro Santos
Mohammad Ashjaei
Special Sessions Co-Chairs:
Andrea Bonci
Thilo Sauter
Marco Porta
Work-in-Progress Co-Chairs
Svetlana Girs
Ramon Vilanova
Luca Leonardi
Workshops Co-Chairs
Frank Golatowski
Lucia Lo Bello
Diversity and Inclusion Co-Chairs
Lucia Lo Bello
Regina Roos
Teresa Delgado
Industry Forum Co-Chair:
Gil Gonçalves
Bruon Iafelice
Publicity and Web Co-Chairs:
Gowhar Javanmardi
Zenepe Satka
Alberto Morato

Welcome to ETFA 2025. This conference is the flagship event of the IEEE IES Technical Committee on Factory Automation (TCFA) and a premier event sponsored by the IEEE Industrial Electronics Society. It aims at bringing together the international community to present the latest research results, share new ideas and engineering breakthroughs, and discuss state-of-the-art challenges and future directions in technology and innovation in the broad domain of Automation with a focus on Industrial and Factory Automation.

The organizing committee cordially invites high-quality papers representing original work, including but not limited to the following topics:

- Adaptive and Intelligent Control Systems
- Advanced Motion Control
- Autonomous Robotic Systems, Artificial Intelligence, and Machine Learning
- Automotive Control and Transportation Systems
- Biomechatronics and Bioengineering Systems
- Compliant and Soft Robotics
- Haptics and Human-Robot Interaction
- Industry Applications, Information Technology, and Advanced Manufacturing
- Micro-Electro-Mechanical Systems (MEMS) and Nanotechnologies
- Network-based Control Systems and Applications
- Sensors and Actuators
- Smart Materials and Structures
- Visual Servo Systems, Machine Vision, and Image Processing

ETFA 2025 will also seek Special Sessions and Workshops covering subjects or cross-subjects belonging to the topics of interest, or novel related topics.

Up to date information will be available in the following website. Stay tuned!:

etfa2025.ieee-ies.org

Timeline

Special Session proposals deadline..... February 13, 2025..... Notifications -- February 21, 2025
Workshop proposals deadline..... February 28, 2025..... Notifications -- March 14, 2025
Regular and SS submissions deadline..... April 18, 2025..... Notifications -- May 23, 2025
Work-in-Progress submission deadline..... May 30, 2025..... Notifications -- June 20, 2025
Final versions deadline..... July 4, 2025

(These dates may be updated. **Always check the website**)

IEEE Advancing Technology for Humanity
IES
U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO
CISTER - Research Centre in Real-Time & Embedded Computing Systems

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on

https://journals.ieeeauthorcenter.ieee.org/
Then click: "IEEE Author Tools for Journals"
- "Article Templates"
- "Templates for Transactions".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *Index Terms (Keywords)*. For the final version of accepted papers, please send the short cvs and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information. When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue
g) Document Object Identifier (DOI)

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984. DOI: 10.1109/TEI.1984.298778

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to submit their papers electronically via the following portal address:

https://www.ojs.hte.hu/infocommunications_journal/about/submissions

If you have any question about the journal or the submission process, please do not hesitate to contact us via e-mail:

Editor-in-Chief: Pál Varga – pvarga@tmit.bme.hu

Associate Editor-in-Chief:

József Bíró – biro@tmit.bme.hu

László Bacsárdi – bacsardi@hit.bme.hu

33rd International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems

Call for Papers

The MASCOTS 2025 conference encourages original submissions describing state-of-the-art research in the areas of the performance evaluation of computer systems and networks as well as in related areas. Papers describing results of theoretical and/or practical significance are welcome. Experimental, modeling, and simulation studies are all in the scope of the conference. Papers focusing on novel performance evaluation methods or providing insights on design and runtime management tradeoffs are particularly encouraged.

The submission deadline is **May 18th, 2025 AoE Sunday.**

Topics of interest include (but are not limited to):

> Big data and advanced machine learning techniques for system design, optimization, and cybersecurity
> Cloud/edge/fog technologies
> Combining quality of service and cybersecurity in system performance
> Computer architectures, multi-core processors, accelerators (e.g., GPUs), and memory systems
> Computer networks, protocols, and algorithms
> Databases and big data systems and technologies
> Internet-of-Things
> Mobile systems
> Multimedia systems
> Operating systems and virtualization technologies
> Security in computer and communication systems
> Smart grids and cyber-physical systems
> Social networks
> Storage and file systems
> Sustainable (or energy-efficient) computing
> Web systems, enterprise applications, and web services
> Wireless, mobile, ad-hoc, and sensor networks

Some of the best submitted papers that fall below the acceptance threshold for the main conference program, will be invited to the MASCOTS'25 Workshop. If they are orally live presented at the workshop by one registered author, they will then be published in the MASCOTS'25 proceedings as 4-page short papers.

All questions about submissions should be emailed to mascots2025 at easychair dot org.

Important Dates

Paper submission deadline: May 18th, 2025 AoE
Notifications: August 8th, 2025
Author registration deadline: September 5th, 2025
Camera-ready: September 12th, 2025

Technically Co-Sponsored by the IEEE Computer Society (approval pending)

IEEE COMPUTER SOCIETY

Sponsors

SORBONNE UNIVERSITÉ
IITIS
LIP6
slices

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its 1000 individual members, the Scientific Association for Infocommunications (in Hungarian: Híradástechnikai és Informatikai Tudományos Egyesület, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society.

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **FERENC VÁGUJHELYI** • *elnok@hte.hu*

Secretary-General: **GÁBOR KOLLÁTH** • *kollath.gabor@hte.hu*

Operations Director: **PÉTER NAGY** • *nagy.peter@hte.hu*

Address: H-1051 Budapest, Bajcsy-Zsilinszky str. 12, HUNGARY, Room: 502

Phone: +36 1 353 1027

E-mail: *info@hte.hu*, Web: *www.hte.hu*