

A Deep-Unfolding Approach to RIS Phase Shift Optimization Via Transformer-Based Channel Prediction

Ishan Koralege , Arthur S. de Sena, Nurul H. Mahmood,
Farjam Karim, Dimuthu Lesthuruge, Samitha Gunarathne

*Centre for Wireless Communications, University of Oulu,
Oulu, Finland*

{ishan.koralege, arthur.sena, nurulhuda.mahmood, farjam.karim,
lesthuruge.silva, samitha.gunarathne}@oulu.fi

Abstract: Reconfigurable intelligent surfaces (RISs) have emerged as a promising solution that can provide dynamic control over the propagation of electromagnetic waves. The RIS technology is envisioned as a key enabler of sixth-generation networks by offering the ability to adaptively manipulate signal propagation through the smart configuration of its phase shift coefficients, thereby optimizing signal strength, coverage, and capacity. However, the realization of this technology's full potential hinges on the accurate acquisition of channel state information (CSI). In this paper, we propose an efficient CSI prediction framework for a RIS-assisted communication system based on the machine learning (ML) transformer architecture. Architectural modifications are introduced to the vanilla transformer for multivariate time series forecasting to achieve high prediction accuracy. The predicted channel coefficients are then used to optimize the RIS phase shifts. Simulation results present a comprehensive analysis of key performance metrics, including data rate and outage probability. Our results confirm the effectiveness of the proposed ML approach and demonstrate its superiority over other baseline ML-based CSI prediction schemes such as conventional deep neural networks and long short-term memory architectures, albeit at the cost of slightly increased complexity.

Keywords: Channel prediction, deep neural network, machine learning, reconfigurable intelligent surface, transformer.

1. INTRODUCTION

Sixth-generation (6G) wireless networks will have the ability to push beyond the limits of today's wireless systems with their groundbreaking expected features such as very low latency, ultra-high reliable connectivity, enhanced data security, and integrated intelligence that leverages machine learning (ML) capabilities (Chowdhury et al., 2020). Within this framework, reconfigurable intelligent surface (RIS) has emerged as a potential game-changer. RIS introduces a new layer of intelligence into the physical layer, allowing the stochastic wireless propagation environment to be somewhat controlled (Mahmood et al., 2023). This technology holds the promise of substantial improvements in signal strength, signal coverage, and network capacity by offering advanced control over the properties of electromagnetic waves (Basar et al., 2019). RISs are composed of passive reflecting elements that can be dynamically configured to manipulate the way wireless signals propagate toward receivers, enabling the achievement of diverse goals.

* The research leading to this paper was supported by the Research Council of Finland (former Academy of Finland) 6G Flagship program (Grant Number: 346208), and Business Finland's 6GBridge program through the projects Local 6G (Grant Number 8002/31/2022) and 6CORE (Grant Number 8410/31/2022).

Since RISs typically comprise only nearly passive components with no data processing capabilities, acquiring RIS-associated channel state information (CSI) is a fundamental challenge (Yuan et al., 2021). Classical methods of CSI estimation such as least squares and minimum mean-squared error, as in the work in (Ardah et al., 2021), rely on pilot signals, which incurs significant signaling overhead and channel acquisition delay, especially when the number of RIS elements is large. It is further exacerbated under dynamic wireless environments when the channel coherence time is short, thus requiring frequent CSI acquisition. These issues can be mitigated via learning-based approaches as they harness the power of ML to learn the optimal RIS phase shifts, thereby eliminating the requirement of complex mathematical modeling or overwhelming pilot training (Hashemi et al., 2023). However, such an approach suffers from slow convergence, large training overhead, and poor generalizability. A deep unfolding approach, where ML-based techniques are used to learn partial system blocks while adhering to conventional optimization approaches for the overall system design (Balatsoukas-Stimming and Studer, 2019) can be adopted to address these limitations. This results in improved performance at a much lower complexity. More specifically, we propose to replace the pilot-based CSI estimation procedure with an

ML-based CSI prediction method, which is then used to mathematically optimize the RIS phase shifts.

ML-assisted solutions have the potential to adapt and learn the dynamics of the CSI autonomously. By training ML models with large data sets, unforeseen channel characteristics can be captured, even in complex communications environments assisted by RISs. Conventional deep neural network (DNN) and recurrent neural network (RNN) architectures can provide, to some extent, satisfactory prediction results (Gao et al., 2021). However, both DNNs and RNNs have limitations such as the vanishing gradient phenomenon. This occurs when training DNNs with a large number of layers, including activation functions. In such cases, the gradients used to update the network become extremely small or even vanish as they are backpropagated. As a result, the convergence of the algorithm is substantially slowed. (Glorot and Bengio, 2010). Long short-term memory (LSTM) networks are an evolution of RNNs that have been proposed to prevent the vanishing gradient problem (Hochreiter and Schmidhuber, 1997), which enables the processing of longer data sequences. A recent architecture named *transformers* pushes the boundaries further. Transformers have excelled across various domains due to their superior characteristic performance (Tay et al., 2022), significantly outperforming most of the previous deep learning approaches.

Diverse ML-based strategies have been employed in the literature to predict RIS-associated CSI. In (Elbir and Coleri, 2022), a federated learning strategy with distributed convolution neural networks (CNNs) was used for channel prediction in a RIS-assisted multi-antenna system employing orthogonal frequency-division multiplexing. A distributed CNN framework for downlink channel prediction in a narrow-band multi-user system was proposed in (Dai and Wei, 2022). A real-time reinforcement learning-aided CSI measurement scheme for a RIS-assisted millimeter wave (mmWave) system was considered in (Zhang et al., 2022). The authors in (Zhang et al., 2021) developed schemes based on CNNs and fully connected DNNs for RIS channel extrapolation and beam searching in a single-antenna system. A deep denoising CNN was used for aiding RIS channel estimation in a mmWave system in (Liu et al., 2020). In (Nguyen et al., 2023), a strategy based on CNNs and LSTMs was proposed to predict the CSI of a RIS-assisted system employing non-orthogonal multiple access, and in (Xia et al., 2024), a transformer-aided scheme was proposed for predicting the CSI of an uplink RIS-assisted mmWave system.

To the best of our knowledge, only the work in (Xia et al., 2024) has exploited the transformer architecture for channel prediction in RIS-assisted networks. In this paper, we propose a novel transformer-based approach to predict the unknown CSI in a RIS-assisted communication system. The predicted CSI is then used to optimize the RIS phase shifts for downlink data transmission. We compare our proposed prediction scheme with other state-of-the-art learning-based approaches, evaluating the performance using metrics such as data rate and outage probability. Based on the ML architectures we developed, it is shown that the transformer outperforms LSTM and DNN architectures, though the architecture is slightly more complex. Furthermore, the DNN architecture shows the lowest perfor-

mance with the least complexity, while the LSTM method demonstrates higher performance and greater complexity than the DNN approach, yet lower performance and less complexity compared to the transformer architecture.

The remainder of the article is organized as follows. In Section 2, the system model is provided. Section 3 gives insights into the proposed ML-based CSI predictor. Section 4 provides the results and discussion. Finally, the conclusions are presented in Section 5.

2. SYSTEM MODEL

Consider a downlink RIS-aided single-input single-output (SISO) communication system as in Fig. 1, where the base station (BS) and the user equipment (UE) are each equipped with a single antenna, and the RIS comprises N reflecting elements. Let $f \in \mathbb{C}$, $\mathbf{h} \in \mathbb{C}^{N \times 1}$, $\mathbf{g} \in \mathbb{C}^{N \times 1}$ be the fading channels between the BS and the UE, the BS and the RIS, and the RIS and the UE, respectively. The channel vectors $\mathbf{h} = [h_1, \dots, h_i, \dots, h_N]^T$ and $\mathbf{g} = [g_1, \dots, g_i, \dots, g_N]^T$, where h_i (g_i) is the channel between the BS (UE) and i^{th} RIS element. We assume that the channel coefficients are Rayleigh distributed, i.e., $f \sim \mathcal{CN}(0, L_{\text{BS-UE}})$, $h_i \sim \mathcal{CN}(0, L_{\text{BS-RIS}})$, and $g_i \sim \mathcal{CN}(0, L_{\text{RIS-UE}})$, for $i = 1 \dots, N$, where L_a denotes the path loss coefficients in linear scale for the a^{th} link with $a \in \{\text{BS-UE}, \text{BS-RIS}, \text{RIS-UE}\}$. In this work, we consider the log-distance path loss model, which can be given in dB by

$$L_{a,\text{dB}} = L_{0,\text{dB}} + 10\eta_a \log_{10} \left(\frac{d_a}{d_0} \right), \quad (1)$$

where η_a and d_a represent the path loss exponent and the distance for the link a respectively, and $L_{0,\text{dB}}$ is the reference path loss in dB at a reference distance d_0 . Under this model, the received signal at the UE can be expressed as

$$y = \sqrt{P} (f + \mathbf{g}^T \Phi \mathbf{h}) x + n, \quad (2)$$

where $n \sim \mathcal{CN}(0, N_0)$ is the additive white Gaussian noise (AWGN), x is the transmitted symbol, P represents the transmit power and Φ is the matrix comprising the reflection coefficients of the RIS, which is a diagonal matrix defined as (Basar et al., 2019)

$$\Phi = \text{diag}(\phi), \quad (3)$$

where $\phi = [a_1 e^{j\phi_1}, a_2 e^{j\phi_2}, \dots, a_N e^{j\phi_N}]^T$ is the RIS phase shift vector, in which a_i is the amplitude and, ϕ_i is the adjustable phase induced by the i^{th} reflecting element. Due to the passive nature of the RIS, we can assume $a_i = 1, \forall i = 1, \dots, N$.

2.1 RIS Phase Shift Optimization

We wish to compute a phase shift matrix Φ that can maximize the signal power propagating through the RIS at the UE during downlink data transmission. To this end, we assume that the direct link is non-dominant. This can be the case when it is non-line of sight, while the BS-RIS and RIS-UE links are in line of sight. Under this assumption, we can recall the Cauchy-Schwartz inequality and achieved (Björnson et al., 2024)

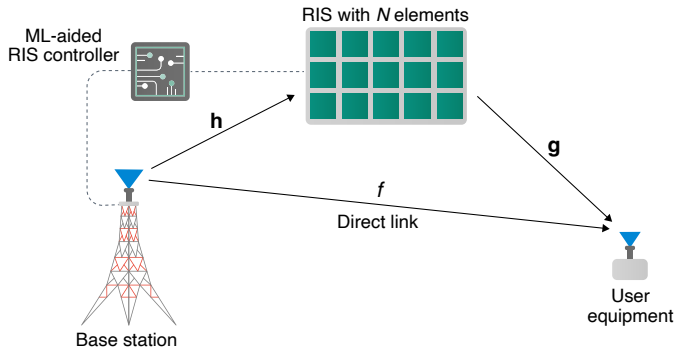


Fig. 1. System model. An RIS comprising N elements assists the communication between one single-antenna BS and one single-antenna UE.

$$|\mathbf{g}^T \Phi \mathbf{h}|^2 = |(\mathbf{h} \odot \mathbf{g})^T \boldsymbol{\phi}|^2 \leq \left(\sum_{i=1}^N |h_i g_i e^{j\phi_i}| \right)^2, \quad (4)$$

where \odot denotes the Hadamard product. The inequality in (4) provides an upper bound to the reflected power i.e., the maximum power that can be reflected by the RIS. The equality can be achieved if and only if $\phi_i = -\arg(h_i g_i)$. As a result, the desired optimal RIS phase shift vector can be given by

$$\boldsymbol{\phi}^* = e^{-j\arg(\mathbf{h} \odot \mathbf{g})}. \quad (5)$$

The optimum phase shift coefficients in (5) can be readily computed after predicting the CSI through ML, as discussed in the subsequent sections.

2.2 Performance Metrics

In this subsection, we present the performance metrics used to assess the effectiveness of the proposed ML-based CSI prediction framework. To this end, we first provide an expression for the signal-to-noise ratio (SNR). More specifically, by using (2) and (4), the instantaneous SNR observed at the UE, denoted as γ , can be computed by

$$\gamma = \frac{|(\mathbf{h} \odot \mathbf{g})^T \boldsymbol{\phi} + f|^2 P}{N_0}. \quad (6)$$

The outage probability, P_{out} , is the first important performance metric used to assess the performance of the UE. It is the probability of the instantaneous SNR falling below a defined threshold γ_{th} , i.e.,

$$P_{\text{out}} = P(\gamma < \gamma_{\text{th}}), \quad (7)$$

where P denotes probability.

Last, we consider the maximum data rate that can be achieved when the BS is transmitting at a fixed rate R_{th} , subject to a given outage probability P_{out} , which can be obtained as

$$R = (1 - P_{\text{out}})R_{\text{th}}, \quad (8)$$

where R_{th} represents the data rate threshold given by

$$R_{\text{th}} = \log_2(1 + \gamma_{\text{th}}). \quad (9)$$

3. CSI PREDICTION AND DATA PREPARATION

This section presents the proposed architecture of our transformer model which is used to make multivariate time

series predictions of the channel coefficients. Details on the generation of the data sets used to train, validate, and test the implemented ML models are also presented.

3.1 Transformer-Based CSI Prediction

In this subsection, we describe our proposed CSI prediction approach based on the ML transformer architecture, which we later exploit to optimize the RIS phase shifts. Transformers are a deep learning architecture initially introduced in (Vaswani et al., 2017) by a research group in Google. The key feature of this architecture is its state-of-the-art attention mechanism. The attention mechanism is a computational method that focuses on specific parts of the input data sequence, assigning varying degrees of importance to the different parts. This mechanism helps to inform the model where to pay attention when processing data. The vanilla transformer architecture consists of two parts, an encoder and a decoder, which consists of a sequence-to-sequence architecture. In this method, the model generates an output data sequence according to the given input sequence.

In this work, we aim to predict the CSI at the next time instance based on a given sequence of previous CSI samples. Therefore, the original sequence-to-sequence architecture is modified into a sequence-to-one architecture using only one encoder. More specifically, the encoder of our transformer model comprises an input embedding module, a positional encoding module, and a transformer encoder module, as shown in Fig. 2. The input embedding module transforms the dimension of the input data to the model dimension of the subsequent inner layer of the transformer. Since the transformer lacks a recurrent structure as in recurrent neural networks, it feeds the positional information to the output of the embedding layer separately. After that, using the acquired knowledge from the input sequence, an abstract representation is generated by the transformer encoder. The encoder consists of a multi-head attention block, layer normalization blocks, and a feedforward layer, as illustrated in Fig. 2. Then, the encoder output is sent through a fully connected (FC) linear layer and an activation function. Figure 2 shows the modified sequence-to-one transformer architecture used for CSI prediction, where we feed samples of real and imaginary channel data separately to forecast the next time instance of the channel coefficients as output.

The vanilla transformer architecture, originally introduced for natural language processing applications, contains two FC layers in the feedforward block and uses the rectified linear unit (ReLU) as the activation function (Vaswani et al., 2017). This setup is effective for capturing long-term variations in the multi-head attention output. However, given the relatively small window size, our problem requires observing and capturing short-term variations. Therefore, as shown in Fig. 3, we modified the original architecture by implementing two one-dimensional (1D) CNN layers instead of the two FC layers, as CNNs effectively capture local temporal patterns crucial for accurate forecasting. Additionally, we replaced the ReLU activation function with a hyperbolic tangent (Tanh) function. These modifications result in better performance in our time series forecasting problem.

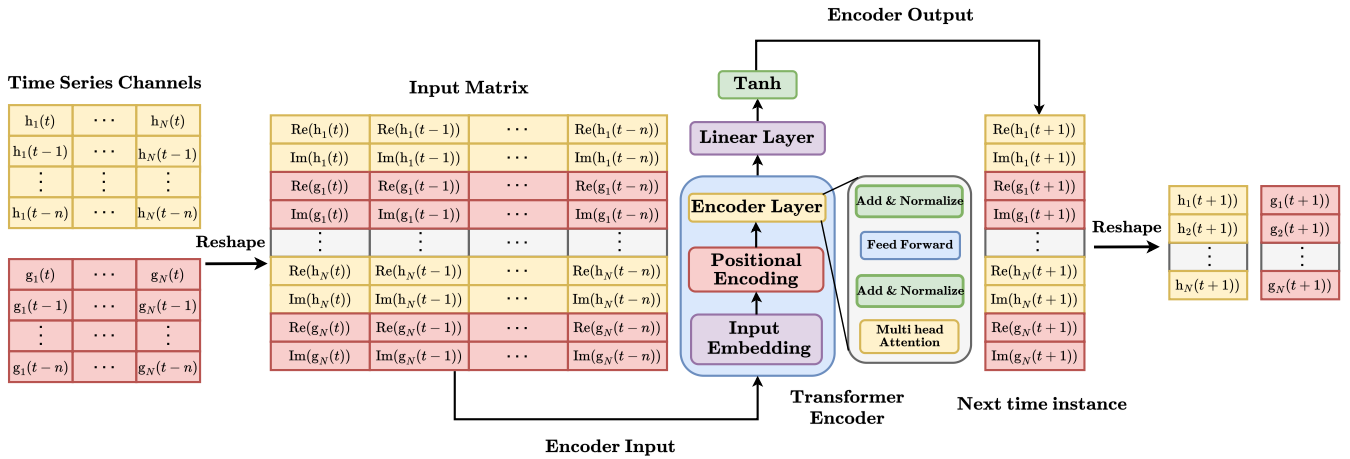


Fig. 2. Sequence-to-one transformer architecture with input and output.

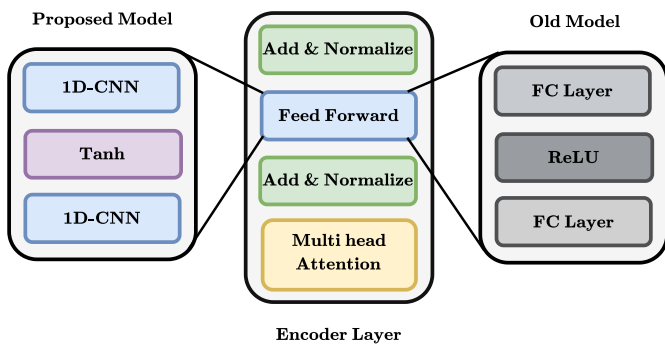


Fig. 3. Feedforward layer in the vanilla transformer architecture (right) and the proposed model (left).

3.2 Data Preparation

Most of the cited works on CSI prediction assume independent/uncorrelated fading, whereas real-world scenarios often exhibit correlation. We incorporate this important property in our work by considering time-correlated Rayleigh-distributed fading channel coefficients. We first generate uncorrelated Rayleigh-fading samples through simulation with zero mean and unit variance, which are then convoluted with a finite impulse response filter representing the correlation function. Before feeding time-correlated channel data to the ML models, the data is normalized for efficient convergence. Out of the 2550 time-correlated data samples, 80% is allocated for training, while the remaining 20% are allocated for validation. Another 50 samples are taken for testing per Monte Carlo iteration. In all the architectures, a moving window is deployed for iterative training, validation, and testing processes. To learn the temporal patterns, consecutive 10 samples are taken. Using this learned knowledge, 11th sample is predicted. A perfect CSI assumption is made for the training data. Though this is a strong assumption, it helps us to obtain useful information about the performance upper bound.

4. RESULTS AND DISCUSSIONS

This section presents comprehensive simulation results to verify the effectiveness of our proposed transformer-

based CSI prediction framework. The number of parameters and other architectural details used by different ML strategies are also put into perspective. To demonstrate the performance advantages of our approach, we consider two baseline ML architectures, namely DNN and LSTM, where metrics such as data rate and outage probability are compared.

We consider the path loss exponent for the BS-RIS and the RIS-UE links to be 2.2, whereas the exponent for the direct BS-UE link is assumed to be 4.2. The reference path loss value and noise power are taken as -30 dB and -100 dB, respectively. The distances from the transmitter to the RIS, the RIS to the receiver, and the transmitter to the receiver are set to 38 m, 5 m, and 40 m. Here, the SNR threshold $\gamma_{th} = 1$, correspondingly $R_{th} = 1$ too. Unless stated otherwise, an RIS with eight elements is considered and the transmit power is set to 0dB.

4.1 Comparison with ML Baseline Architectures

In this section, the performance of the proposed transformer approach is compared with the conventional DNN and the LSTM baseline ML architectures. All ML models are trained for 100 epochs employed with the Adam optimizer alongside the root mean square error (RMSE) as the loss function. ReLU is used as the activation function for DNN and LSTM architectures while Tanh is used for the transformers. Architectures in all the models used a window size of 10. In each encoder layer, four attention heads and 20 feedforward layer dimensions are used across all transformer models.

Variation of Transmit Power with Fixed RIS Elements

Let us consider the system model as mentioned in Fig. 1 with the number of RIS elements set to $N = 8$. There are 16 channels associated with the RIS model, eight each for the BS-RIS and RIS-UE link, respectively. Since we are predicting real and imaginary channel values separately, 32 features are required to predict at once. Since the number of RIS elements remains constant in this scenario, we can employ a single model for each architecture to make predictions, as the input to each model remains unchanged.

Table 1. Optimized hyperparameters of DNN, LSTM, and Transformer models.

Hyperparameters	Values														
	DNN					LSTM					Transformer				
Number of RIS elements	4	8	12	16	20	4	8	12	16	20	4	8	12	16	20
Input and Output features	16	32	48	64	80	16	32	48	64	80	16	32	48	64	80
Neurons in LSTM layer	-	-	-	-	-	22	44	60	76	100	-	-	-	-	-
Neurons in FC layer 1	20	36	52	68	88	18	36	56	68	90	16	32	48	64	80
Neurons in FC layer 2	24	40	56	72	92	-	-	-	-	-	-	-	-	-	-
Neurons in FC layer 3	24	44	56	68	94	-	-	-	-	-	-	-	-	-	-
Neurons in FC layer 4	20	36	52	68	88	-	-	-	-	-	-	-	-	-	-
Transformer model dimensions	-	-	-	-	-	-	-	-	-	-	24	48	60	80	120

Table 2. Summary of metrics associated with DNN, LSTM, and Transformer models.

RIS Elements	Architecture	Train RMSE	Test RMSE	P
4	DNN	0.0335	0.0271	2, 280
	LSTM	0.0149	0.0172	3, 796
	Transformer	0.0134	0.0158	6, 838
8	DNN	0.0364	0.023	6, 968
	LSTM	0.0282	0.0191	16, 532
	Transformer	0.0131	0.0142	26, 702
12	DNN	0.0413	0.0259	14, 216
	LSTM	0.0329	0.0219	32, 552
	Transformer	0.0159	0.0157	46, 818
16	DNN	0.0535	0.0317	24, 024
	LSTM	0.0425	0.0266	52, 820
	Transformer	0.0177	0.0165	82, 894
20	DNN	0.0555	0.0322	39, 538
	LSTM	0.036	0.0228	89, 170
	Transformer	0.0182	0.0166	164, 630

The transmit power is varied from 0 dBm to 50 dBm by keeping other initial model parameters of the RIS the same. Optimized hyperparameters of the ML models only used for this scenario are tabulated under the columns where the number of RIS elements is eight for each approach as in Table 1. To measure the effectiveness of the architectures, a performance evaluation should be carried out. In this study, we use RMSE as the prediction evaluation criterion. It measures the RMSE between the predicted feature sequence and the actual sequence values in the test and training data sets separately. Apart from that, we have obtained the number of model parameters P associated with models which are calculated using weights and biases. Metrics obtained through the above model simulations are organized in Table 2. Moreover, when N is fixed at 8 it is evident that the proposed transformer-based approach significantly outperforms LSTM and DNN architectures, thereby reducing the training RMSE approximately by 115% and 177% and test RMSE approximately by 34% and 62%, respectively.

Variation of RIS Elements with Fixed Transmit Power

In this scenario, number of RIS elements is the variable. Let us consider instances where the number of RIS elements is 4, 8, 12, 16, and 20 by keeping other initial model parameters of the RIS the same. Therefore, the number of channels associated with the RIS are 8, 16, 24, 32, and 40, respectively. Then, according to our model configuration (considering real and imaginary values separately as fea-

tures), the number of features we need to handle would be 16, 32, 48, 64, and 80, respectively. For each case, we need to optimize the ML models separately since the number of input features is changing. The optimized hyperparameters for each architecture are summarized in Table 1. It shows the number of neurons available in each layer of the respective architecture, the number of neurons in the LSTM layer, and the transformer model dimensions.

Table 2 presents the summary of metrics associated with all three ML architectures such as the train RMSE, the test RMSE, and the number of model parameters. According to the observations, it is clear that the transformer architecture has outperformed the DNN and LSTM architectures in terms of performance when compared with both the train and test RMSE values. This shows that the transformer architecture can significantly outperform the state-of-the-art ML architectures, albeit at the cost of higher complexity. Hence, the transformer architecture is preferable in scenarios where the prediction accuracy is to be prioritized over computational complexity. However, the increased computational complexity can be easily handled by using optimized hardware such as tensor processing units.

4.2 RIS Phase Optimization with Predicted CSI

Using the prediction values obtained from the ML models, we can calculate the optimal phase shift vector of the RIS model as mentioned in (5). Then plugging the test data (actual) channel values and the calculated RIS optimal phase shift vector to (6), we can obtain the maximized SNR values. From there onwards, (7) and (8) can be obtained with respect to the maximized SNR values. Therefore, we have shown that RIS could be optimized by predicting the CSI in a RIS-assisted communication system. Let us focus on the two major scenarios of our research output. The generated results shown in Fig. 4 and Fig. 5 compare key scenarios, including the outputs with optimized-phase RIS, fixed-phase RIS, without the RIS, and the integrated outputs predicted by the transformer, LSTM, and DNN models. Additionally, in the fixed-phase RIS scenario, Φ is set as an identity matrix.

The system performance for a fixed number of RIS elements with varying transmit power is illustrated in Fig. 4. Specifically, Figure 4(a) demonstrates the variation in outage probability, while Figure 4(b) shows the variation in average rate. As the transmit power increases, the outage probability decreases, and the average rate increases. In Fig. 4(a), the lowest outage probability is observed when the RIS is optimized. It is the optimum scenario when

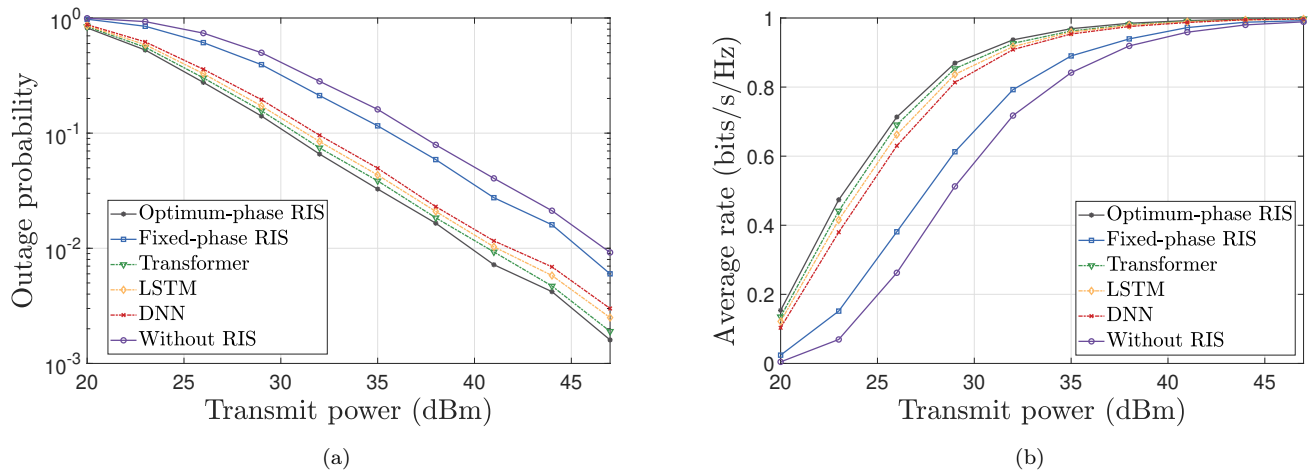


Fig. 4. System performance for a fixed number of RIS reflecting elements ($N = 8$) in terms of (a) outage probability, and (b) average rate, when the transmit power is varied.

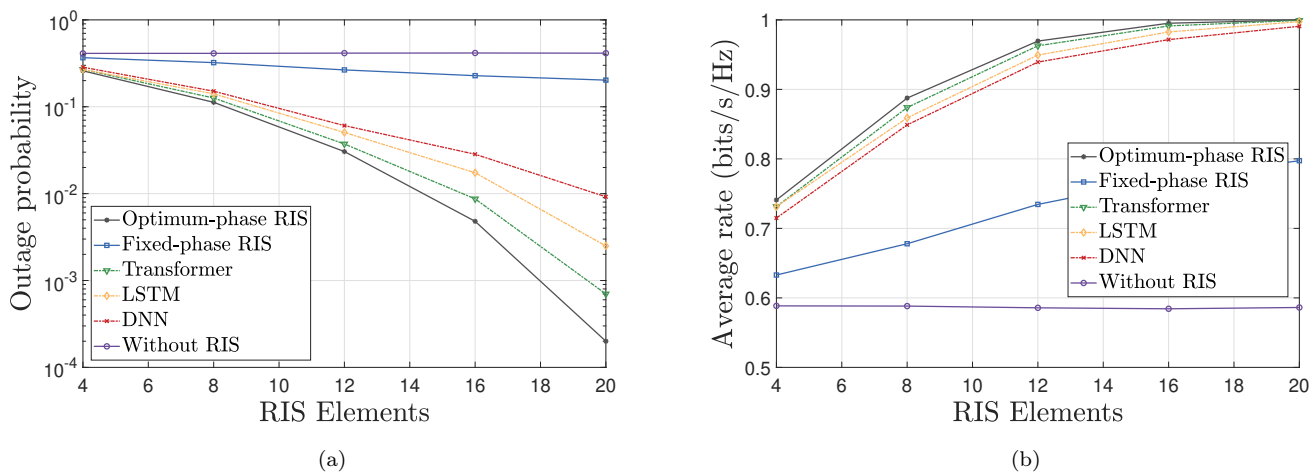


Fig. 5. System performance for fixed transmit power (30 dBm) in terms of (a) outage probability, and (b) average rate when the number of RIS reflecting elements is varied.

the CSI is accurately known at the transmitter. This is the performance upper/lower bound and is not possible in practice. The transformer model closely approaches the performance of the optimized RIS, outperforming both the LSTM and DNN models. Furthermore, to achieve an outage probability of 0.01, the optimized RIS scenario requires a transmit power of 39.79 dBm, compared to 46.66 dBm for the scenario without RIS. In addition, to achieve the same outage probability, the required transmit powers for the transformer, LSTM, and DNN models are 40.67 dBm, 41.13 dBm, and 41.85 dBm, respectively. In Fig. 4(b), the average rate is maximized when the RIS is optimized and the transformer model gets very close to the upper bound established by the optimized RIS. Moreover, it can be observed that when the transmit power is 25 dBm, the performance gap between the scenarios where the transformer model and the system without RIS is approximately 0.4 bits/s/Hz. The LSTM and DNN models also demonstrate notable performance improvements, although they fall short of the transformer model's performance.

The system performance for a fixed transmit power, while varying the number of RIS elements, is presented in Fig.

5. Figure 5(a) illustrates the variation in outage probability, whereas Figure 5(b) depicts the variation in average rate. Generally, as the number of RIS elements increases, the outage probability decreases, and the average rate increases for scenarios incorporating RIS. Figure 5(a) shows that the outage probability is minimized when the RIS is optimized. The transformer architecture yields the closest results to the optimized RIS, followed by the LSTM and DNN architectures. As per the observations, for a 12-element RIS system, the optimized RIS scenario achieves an outage probability of 0.03, while the transformer, LSTM, and DNN scenarios result in outage probabilities of 0.037, 0.05, and 0.06, respectively. In Fig. 5(b), the average rate is maximized when the RIS is optimized and the transformer architecture again demonstrates the closest prediction to the average rate achieved with the optimized RIS. When the number of RIS elements reaches eight, the performance gap between the scenario with the transformer model and the scenario without RIS is approximately 0.3 bits/s/Hz.

From the results, it is evident that the application of RIS has increased the overall system performance noticeably

and the best results are given when the phase is optimized. In every graph, prediction curves lie between the optimum phase and the system without RIS. Out of the three prediction curves, the transformer prediction curve goes very closely with the optimum phase RIS output showing that the transformers have provided the most convincing results compared to both the DNN and LSTM architectures.

5. CONCLUSIONS

Harnessing the full potential of RIS technology hinges on accurate CSI estimation/prediction since the optimum RIS phase shifts are a function of the corresponding composite channel's CSI. In this study, we proposed a novel sequence-to-one transformer architecture to predict the RIS-associated CSI, enabling the efficient and accurate optimization of the RIS phase shifts. For the CSI prediction task, three architectures namely DNN, LSTM, and transformers were utilized. For the time series channel sample data set created, the transformer architecture provided the lowest RMSE value outperforming DNN and LSTM methods in the scenarios discussed above. After that, the predicted CSI from ML models was fed into the RIS model for phase optimization. According to the results obtained, the transformer was the better multivariate time series prediction architecture out of the three architectures in terms of performance but at the cost of higher complexity. Optimizing the RIS phase shift based on transformer-predicted CSI was found to perform very close to the optimum case when the CSI was assumed to be accurately known. Conversely, the DNN architecture yielded the lowest performance and was also the least complex architecture. The LSTM architecture was positioned in between the other two architectures, offering a middle ground in terms of both performance and complexity. Finally, this study concluded that the proposed sequence-to-one transformer architecture provided promising results for channel prediction regarding RIS phase optimization.

REFERENCES

- Ardah, K., Gherekhloo, S., de Almeida, A.L., and Haardt, M. (2021). TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems. *IEEE Signal Process. Lett.*, 28, 513–517.
- Balatsoukas-Stimming, A. and Studer, C. (2019). Deep unfolding for communications systems: A survey and some new directions. In *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, 266–271. doi:10.1109/SiPS47522.2019.9020494.
- Basar, E., Di Renzo, M., De Rosny, J., Debbah, M., Alouini, M.S., and Zhang, R. (2019). Wireless communications through reconfigurable intelligent surfaces. *IEEE access*, 7, 116753–116773.
- Björnson, E., Demir, Ö.T., et al. (2024). *Introduction to Multiple Antenna Communications and Reconfigurable Surfaces*. Now Publishers, Inc.
- Chowdhury, M.Z., Shahjalal, M., Ahmed, S., and Jang, Y.M. (2020). 6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions. *IEEE open j. Commun. Soc.*, 1, 957–975. doi:10.1109/OJCOMS.2020.3010270.
- Dai, L. and Wei, X. (2022). Distributed machine learning based downlink channel estimation for RIS assisted wireless communications. *IEEE Trans. Commun.*, 70(7), 4900–4909. doi:10.1109/TCOMM.2022.3175175.
- Elbir, A.M. and Coleri, S. (2022). Federated learning for channel estimation in conventional and RIS-assisted massive MIMO. *IEEE Trans. Wireless Commun.*, 21(6), 4255–4268. doi:10.1109/TWC.2021.3128392.
- Gao, S., Dong, P., Pan, Z., and Li, G.Y. (2021). Deep multi-stage CSI acquisition for reconfigurable intelligent surface aided MIMO systems. *IEEE Commun. Lett.*, 25(6), 2024–2028. doi:10.1109/LCOMM.2021.3063464.
- Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249–256. JMLR Workshop and Conference Proceedings.
- Hashemi, R., Ali, S., Mahmood, N.H., and Latva-Aho, M. (2023). Deep Reinforcement Learning for Practical Phase-Shift Optimization in RIS-Aided MISO URLLC Systems. *IEEE IoT J.*, 10(10), 8931–8943. doi:10.1109/JIOT.2022.3232962.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Liu, S., Gao, Z., Zhang, J., Renzo, M.D., and Alouini, M.S. (2020). Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces. *IEEE Trans. Veh. Technol.*, 69(8), 9223–9228. doi:10.1109/TVT.2020.3005402.
- Mahmood, N.H., Berardinelli, G., Khatib, E.J., Hashemi, R., De Lima, C., and Latva-aho, M. (2023). A functional architecture for 6G special-purpose industrial IoT networks. *IEEE Trans. Ind. Inf.*, 19(3), 2530–2540. doi:10.1109/TII.2022.3182988.
- Nguyen, C., Hoang, T.M., and Cheema, A.A. (2023). Channel estimation using CNN-LSTM in RIS-NOMA assisted 6G network. *IEEE trans. mach. learn. commun. netw.*, 1, 43–60. doi:10.1109/TMLCN.2023.3278232.
- Tay, Y., Dehghani, M., Bahri, D., and Metzler, D. (2022). Efficient transformers: A survey. *ACM Comput. Surv.*, 55(6), 1–28.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Xia, G., Liu, H., and Long, K. (2024). Transformer-empowered parallel channel prediction for fast-paced and dynamic RIS-aided wireless communication systems. *IEEE Commun. Lett.*, 1–1. doi:10.1109/LCOMM.2024.3383427.
- Yuan, X., Zhang, Y.J.A., Shi, Y., Yan, W., and Liu, H. (2021). Reconfigurable-intelligent-surface empowered wireless communications: Challenges and opportunities. *IEEE Wireless Commun.*, 28(2), 136–143.
- Zhang, Q., Saad, W., and Bennis, M. (2022). Millimeter wave communications with an intelligent reflector: Performance optimization and distributional reinforcement learning. *IEEE Trans. Wireless Commun.*, 21(3), 1836–1850. doi:10.1109/TWC.2021.3107520.
- Zhang, S., Zhang, S., Gao, F., Ma, J., and Dobre, O.A. (2021). Deep learning optimized sparse antenna activation for reconfigurable intelligent surface assisted communication. *IEEE Trans. Commun.*, 69(10), 6691–6705. doi:10.1109/TCOMM.2021.3097726.