An attention-based channel estimation algorithm for next-generation point to point communication systems

Kayode A. Olaniyi[1], Reolyn Heymann[1], Theo G. Swart[1,2]
[1]Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa
[2]Center for Telecommunications, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

Article Info

Article history:

Received Jun 26, 2023
Revised Apr 17, 2024
Accepted May 17, 2024

Keywords:

Attention mechanism
Channel estimation
Channel matrix
Encoder-decoder
Pilot densities

ABSTRACT

Accurate and robust estimation of channel parameters is essential in establishing reliable communication with characteristic optimal resource utilization in next-generation communication systems. Traditional techniques have limitations, such as the need for additional bandwidth and decreased spectral efficiency. Thus, there is a need for novel techniques that enhance the accuracy and robustness of channel parameter estimation in next-generation communication systems. To address this need, we propose in this paper a recurrent neural network (RNN)-based attention mechanism, to improve channel estimation accuracy and robustness in next-generation communication systems. The attention mechanism selectively focuses on the most relevant features while ignoring noise and interference. The attention network weights are initialized and are constantly updated in the course of network training. The weight values determine the significance of the features before passing them to the channel estimator. This allows the algorithm to adapt to varying channel conditions and improve its accuracy in challenging environments. The proposed attention-based algorithm performance is compared with three baseline techniques: learned denoising-based approximate message passing (LDAMP), Wasserstein generative adversarial networks (WGAN), and maximum likelihood (ML). The result evaluations indicate that the attention-based algorithm performs better than the existing artificial intelligence-based channel coding algorithms, in terms of robustness and accuracy.

Corresponding Author:

Theo G. Swart
Department of Electrical and Electronic Engineering Science
Faculty of Engineering and the Built Environment, University of Johannesburg
Johannesburg, South Africa
Email: tgswart@uj.ac.za

1. INTRODUCTION

Next-generation networks (NGNs) have the potential to provide communication speeds of up to 100 gigabits per second. These networks aim to support a wide range of services such as high-speed data, voice, video, and multimedia applications [1], [2]. However, the efficient and accurate delivery of these services heavily relies on the accurate representation and modeling of the channel parameters in real-time. Thus, improving channel coding and estimation techniques is critical in taking the full advantage of NGNs [3], [4]. Moreso, the ever-evolving state of communication technologies over the decades, has ensured channel estimation remains an open problem among researchers [5].

The channel estimation problem involves estimating the channel frequency response (CFR) or the channel impulse response (CIR) based on the received signal. The CIR is a time-domain characterization of the channel, while the CFR is a frequency-domain representation [6], [7]. Both representations contain information about the distortions introduced by the channel. Channel estimation remains a challenging problem especially because of variation in the characteristics of the channel under frequency and time. In addition, noise interference in the received signal can further complicate the estimation process. Traditional methods of channel estimation rely on pilot symbols, which are known symbols, that are inserted into transmitted signal to help estimate the channel. However, these methods can be inefficient and may not work well in some scenarios [8], [9].

Since the emergence of machine learning (ML) techniques, there has been growing interest in leveraging ML for communication applications, including end-to-end channel coding and estimation [1]. ML-based approaches have the potential to adaptively learn and optimize coding schemes based on data-driven insights, leading to improved performance and adaptability to varying channel characteristics [10]. Thus, ML-based channel approximation approaches offer an alternative solution to the challenges in the traditional methods of channel estimation. By training ML algorithms on a large dataset of known signals and their corresponding received signals, the algorithm can learn to detect characteristic patterns within the transmitted and the received information signal under various channel conditions. The trained artificial neural network (NN) model can then be utilized to predict channel estimation that effectively estimates channel characteristics from the received signals without the need for pilot symbols.

There is a conscientious effort by authors in schorlaly literature underscoring the applications of deep learning-based techniques to address the associated challenges of channel coding and estimation. Soltani *et al*. [11] presented a method called ChannelNet, which uses deep learning approach to estimate the wireless channel in communication systems. The study considered the evaluation of mean square error (MSE) metric of the proposed method over a range of signal-to-noise ratios (SNRs) and further compared the results with three cutting-edge algorithms: minimum mean square error (MMSE), estimated MMSE, and approximate linear MMSE. Results as presented by the authors show a significant improvement over these three cutting-edge algorithms that were compared. The study presented that while the MMSE algorithm performance may be competitive with ChannelNet, multiple ChannelNets are superior channel estimators.

Liao *et al*. [8] presented a simulation-based performance evaluation under frequency and time domain channel estimation techniques in a high-speed channel model with non-stationary and fast time-varying features. The result of the authors' simulation described that the proposed non-linear mapping estimation technique is more robust than other methods considering normalized mean squared error (NMSE) and bit-error-rate (BER), especially under high-speed scenarios. In their study, Bai *et al*. [12] conducted simulations to evaluate the performance of a NN application as a channel estimator operating under a time-varying, Rayleigh fading channel. The simulations utilized independent and identically distributed (IID) bit sequences and quadrature phase shift keying (QPSK) modulation to map the bits to symbols. The authors presented four distinct groups of simulation results for the NN estimator and compared its performance against traditional algorithms such as the MMSE estimator and the least squares (LS) estimator. The simulation results revealed that the sliding-block gated recurrent unit (SBGRU) estimator emerged as the optimal solution in the context of a time-varying channel, surpassing both the LS and "MMSE sim" estimators. Furthermore, the authors conducted a performance comparison between the multi-layer perceptron (MLP) and SBGRU estimators for different estimation block lengths. It was observed that the SBGRU estimator demonstrated effective tracking of the channel in most linear regions with minimal oscillation in non-linear regions. On the other hand, the LS and "MMSE sim" estimators exhibited significant vibration. The performance of the BGRU estimator exhibited a rapid decline with an increase in SNR due to the sliding operation introduction, which enabled the utilization of average channel information that is present within a specific time window.

Gizzini and Chafii [6] provided a review of the deep learning approach to channel coding and estimation for doubly dispersive channels in dynamic environments. Traditional approaches use only a few pilot schemes for channel estimation which leads to degraded performance, especially considering a high mobility scenarios. Artificial intelligence-based schemes are low-complexity and robust, and the paper evaluates their performance in different scenarios such as mobility, modulation order, and frame length. Abdallah *et al*. [13] proposed a novel approach that estimates wideband cascaded channels in RIS-assisted multi-user millimeter-wave (mmWave) massive multiple-input-multiple-output (MIMO) systems. Unlike traditional methods, this technique explicitly considered the beam squint phenomenon and aims to minimize the training cost. The approach takes advantage of the inherent sparsity property shared among subcarriers and the double-structured sparsity property of users' angular cascaded channel matrices. By utilizing denoising NN algorithms to accurately predict and detect channel supports, this data-driven approach outperforms beam squint agnostic orthogonal matching pursuit (OMP) methods. It achieves a significantly lower NMSE of 5-6 dB and also reduces the lower bound gap to only 1 dB compared to the least-square benchmark.

Gao *et al*. [14] introduced a novel approach for channel estimation in hybrid analog-digital (HAD) architecture, which is generally adopted in practical mmWave massive MIMO systems. The HAD architecture offers cost and energy efficiency advantages but poses challenges for accurate channel estimation as a result of the limited number of radio frequency (RF) chains at transceivers. To address these challenges, the work proposed a deep learning-based channel approximation method that unfolds the sparse Bayesian learning (SBL) algorithm implementation in deep neural network (DNN). This approach leverages the power of deep learning NN to effectively capture complex channel sparsity structures across different domains. By jointly optimizing the measurement matrix, the proposed method mitigates practical effects such as beam squint and power leakage, which can lead to performance degradation in traditional compressive sensing (CS) algorithms. Through extensive evaluations, the authors demonstrate that their approach offers superior performance when compared with existing methods with regard to channel estimation accuracy while also maintaining reduced complexity. The tailored DNN architecture enables efficient representation and exploitation of intricate channel sparsity patterns, resulting in significant performance improvements. These promising results position the proposed DL-based approach as a highly viable and efficient solution for channel estimation in HAD systems.

Xu *et al*. [15] conducted performance evaluation of a novel dual-hop free-space optical (FSO) relay framework for deep-space communication systems. Addressing coronal turbulence, pointing errors, and plasma absorption, the study uses the Málaga fading distribution algorithm to model the coronal channel's scintillation index. The article presents comprehensive analytical formulations for crucial performance index metrics, including average bit error rate, ergodic capacity, and outage probability, considering various techniques including heterodyne detection, and intensity modulation with direct-detection. Through Monte Carlo simulations, the derived expressions are validated, demonstrating the superior performance of the relay-assisted system compared to a single FSO link. Additionally, the article provides analytical exploration into the impact of different parameters on the system performance, contributing valuable knowledge to the field of deep space communications.

Xu *et al*. [16] conducted the system performance evaluation of unmanned aerial vehicle (UAV)-assisted dual-hop FSO communication systems. Employing amplify-and-forward relaying and, intensity modulation/direct detection, the study addresses fading channels' impact on UAV-assisted systems. The analysis includes factors like atmospheric turbulence, attenuation loss, pointing errors and, angle-of-arrival fluctuations. Málaga distribution models the FSO link between the UAV and the destination. The study derives closed-form expressions for outage probability and average bit error rate, considering diverse system and channel parameters. Comparative analyses of modulation schemes are conducted, and simulation results validate the analytical findings.

A general problem with the reviewed methods is that they often rely on assuming a specific channel model and may not generalize well to other types of channels. Furthermore, traditional methods for channel estimation require known pilot symbols, which can be inefficient in high-dimensional, wideband systems [17]. Deep learning methods also often suffer from limitations such as poor performance in out-of-distribution settings and high computational complexity. Thus, to address these problems, we propose an attention-based channel estimation algorithm for next-generation wireless point-to-point communication systems. In communication systems, channel conditions can exhibit temporal variations, and recurrent neural networks (RNNs) are well-suited for modeling such dynamic patterns over time. The recurrent nature of these networks allows them to maintain a memory of past observations, enabling more effective learning and adaptation to changing channel states. Additionally, RNNs are suitable for scenarios where the relationship between input features and channel parameters is complex and non-linear, as they can capture intricate patterns in the data. Thus, we propose in this work, a novel approach for wireless channel coding and estimation using RNN and the attention mechanism.

2. METHOD

2.1. System model

We explore the case of a point-to-point MIMO communication system depicted in Figure 1 which comprises a transmitter K_T and receiver R_r. The primary objective of this system would be to establish reliable and efficient transmission of data between a transmitter and the receiver. The key factor in signal propagation within this system is the channel state information (CSI) matrix, denoted as H. This matrix represents the characteristics of the wireless channel between the transmitter and the receiver. The matrix H has dimensions $K_T \times R_r$, where K_T represents the transmitter's number of transmit antennas and, R_r represents the receiver's number of receive antennas. The CSI matrix characterizes the signal propagation in this system [2].

$$H \,\mathcal{E}\, \mathbb{C}^{K_T \times R_r} \tag{1}$$

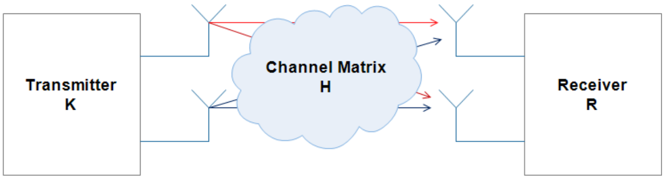

Figure 1. MIMO system model

The CSI matrix H is a complex-valued matrix, with each entry representing the channel gain between a specific transmit antenna and, a specific receive antenna. These channel gains capture the effects of multipath propagation, fading, and other environmental factors on the wireless signal as it travels from the transmitter point to the receiver point. We note that since the CSI matrix H is unknown at both the transmitter and the receiver, an accurate model estimation of the CSI is crucial for optimizing the performance of the MIMO communication system [18]. Once the CSI is estimated, it is used at the transmitter to design the transmission strategy, to maximize received information signal quality at the receiver. At the receiver, the estimated CSI is used for decoding the received encoded signal and recovering the original transmitted information.

2.2. Problem formulation

Assume we have a set of transmitted symbols $[x_1, x_2, \ldots, x_T]$ from a transmitter to a receiver through a wireless channel. Let $[y_1, y_2, \ldots, y_T]$ be the received symbols at the receiver end. There exist a relationship between the received symbols and the transmitted symbols represented by (2):

$$y_t = \sum_{i=1}^{T} h_{t,i} x_i + w_t \tag{2}$$

where $h_{t,i}$ represents the channel coefficient between the *i*-th transmitted symbol and, the *t*-th received symbol and, w_t is the additive white Gaussian noise with zero mean and variance. Thus, the goal is to estimate the channel coefficients $h_{t,i}$ for all values *i* and *t*.

2.3. Attention-based channel estimation

We present a deep learning technique that utilizes attention mechanisms to enhance channel coding and estimation accuracy in wireless communication. In this approach, we have formulated the channel estimation problem as a sequence-to-sequence learning problem, where the input information sequence is the received signal and the output sequence is the CIR. The attention mechanism is used to focus on the most representative features of the input signal while ignoring the noise, leading to better estimation performance. The next subsections describe the various stages of attention-based channel estimation.

2.3.1. Encoder-decoder design

The RNN with gated recurrent units (GRU) forms a major component in the encoder design. The input to the encoder is the received symbols sequence $[y_1, y_2, \ldots, y_T]$. At each iterative time step *t*, the RNN takes the received symbol y_t and the previous hidden state h_{t-1}, as input and generates the current hidden state h_t as output, i.e.,

$$h_t = f(h_{t-1}, y_t) \tag{3}$$

The GRU is a type of RNN that selectively chooses what aspect of information to keep and what part of the information to ignore in the hidden state. This is achieved through the use of gating mechanisms that monitor and control information flow into and out of the hidden state [19], [20]. The sequential data long-term dependencies are effectively captured with the GRU application. The output signal of the encoder is a sequence of hidden states $[h_1, h_2, \ldots, h_T]$. These hidden states capture the relevant information in the received symbols sequence and are used as input to the decoder for channel coefficient estimation. Algorithm 1 describes the encoder operation. In this algorithm, W_{yh}, W_{hh}, and b_h represent weights and biases of the RNN encoder, while tanh represents a non-linear hyperbolic tangent activation function. The algorithm takes as input the sequence of received symbols *y* and the initial hidden state h_0 of the encoder RNN, and produces as output the sequence of hidden states *H* produced by the encoder RNN. The algorithm processes the received symbols sequentially. It computes the current hidden state h_t using the previous hidden state h_{t-1} and the current received symbol y_t and, then appending current hidden state h_t to the sequence of hidden states *H*. The algorithm then returns the sequence of hidden states *H*. Figure 2 is a representation of the encoder-decoder architecture.

Algorithm 1. Encoder operation
Input: y // sequence of received symbols
h_0 // initial hidden state
W_{yh}, W_{hh}, b_h // weights and biases
Output: H: sequence of hidden states
1. Initialize $H = h_0$
2. $For\ t = 1\ to\ T$:
   a. Compute the current hidden state h_t using the previous hidden state h_{t-1} and the current received symbol y_t as
   $h_t = \tanh\left(W_{yh}\, y_t + W_{hh}\, h_{t-1} + b_h\right)$
   b. Append h_t to H
3. Return H

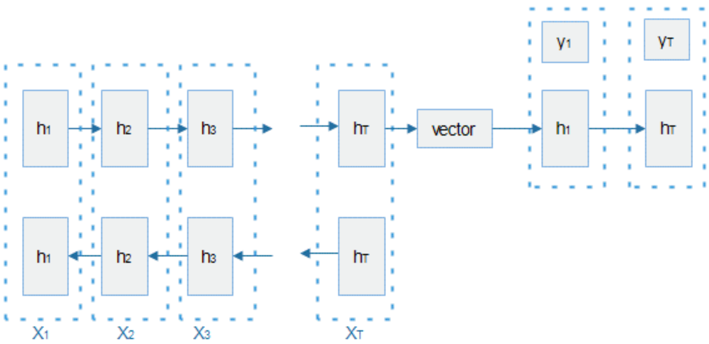

Figure 2. Encoder-decoder model

We implement the decoder as an RNN that processes the hidden states h_1, h_2, …, h_T produced by the encoder. At every iteration time step t, the decoder takes as input the previously estimated channel coefficient $\bar{h}_{t-1}$, the previous hidden state s_{t-1} and, a context vector c_t to generate as output, the currently estimated channel coefficient $\bar{h}_{t,i}$, i.e.:

$$\bar{h}_{t,i} = g(s_{t-1}, \bar{h}_{t-1,i,} c_t) \tag{4}$$

where g is a non-linear function feedforward NN. The context vector c_t represents the weighted sum of the hidden states h_1, h_2, …, h_T produced by the encoder, where the weights are learned during iterative training by an attention mechanism scheme [21]. Specifically, the context vector c_t is computed as (5):

$$c_t = \sum_{i=1}^{T} \alpha_{t,i} h_i \tag{5}$$

where $\alpha_{t,i}$ is the weight assigned to the i-th hidden state for predicting the t-th channel coefficient. The weights $\alpha_{t,i}$ are computed using a soft attention mechanism, which is the dot product between the decoder and the encoder's hidden states s_{t-1} , h_i respectively, i.e.,

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{\sum_{j=1}^{T} \exp(e_{t,j})} \tag{6}$$

where $e_{t,i}$ is the attention score between the decoder and encoder hidden states s_{t-1}, h_i, respectively, and is calculated as (7):

$$e_{t,i} = s_{t-1}.h_i \tag{7}$$

where "·" denotes the dot product. During training, we minimize the mean squared error (MSE) between the estimated channel matrix H and, the true channel matrix H^* using backpropagation and stochastic gradient descent. During testing, the trained attention mechanism based RNN computes efficient channel matrix H estimation from the received information signal y and, the known input sequence x. The channel estimation loss function is represented as a MSE between the estimated channel coefficients $\bar{h}_i$ and, the true channel coefficients h_i, - given by (8):

$$L = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{T}(h_{t,i} - \bar{h}_{t,i})^2 \quad (8)$$

where T is the length of the transmitted symbol sequence. The goal is to achieve a minimal loss function during training so that the model can learn to accurately estimate the channel coefficients.

2.3.2. Encoder network architecture

The encoder scheme used is a RNN architecture with a GRU that efficiently captures long-term dependencies in sequential data. The GRU is composed of a set of recurrently connected units, which update their internal state based on the input information signal at each time step and the previous internal state. Each unit has two gates comprising a reset gate and an update gate. The reset gate is a sigmoidal activation function effectively determines how much of the previous hidden state h_{t-1} information should be forgotten or reset., while the update gate controls the amount of new information being added to the current hidden state. The equations for the GRU at time step *t* are [22], [23]:

Reset gate: $r_t = sigmoid\ (W_r\,.\,[h_{t-1}, x_t] +\ b_r)$ (9)

Update gate: $z_t = sigmoid\ (W_z\,.\,[h_{t-1}, x_t\ ] +\ b_z\ )$ (10)

Candidate hidden state: $h_{t'} = \tanh\ (W_h\,.\,[r_t *\ h_{t-1}\ , x_t\ ] +\ b_h)$ (11)

Hidden state: $h_t =\ z_t *\ h_{t-1} + (1 -\ z_t\ ) *\ h_{t'}$ (12)

where x_t is the input at time step *t*, h_{t-1} is the previous hidden state, $b_r, b_z,$ and b_h are bias terms, "*" denotes element-wise multiplication, z_t is the update gate at time step *t*, $W_r, W_z,$ and W_h are weight matrices that are continuously learned and updated during training. The output of the GRU at time step *t* is the hidden state h_t, which is used as input to the next time step. The initial hidden state h_0 is set to a small random vector.

2.3.3. Decoder architecture

The decoder is an attention mechanism based RNN architecture. It takes as input the hidden states produced by the encoder and, generates the estimated channel coefficients. The decoder RNN is initialized with a trainable vector s_0. At each time step *t*, the previously estimated channel coefficient $\bar{h}_{t-1,i,}$ is embedded using a trainable embedding matrix *E*, which maps each channel coefficient to a fixed-dimensional vector. The decoder uses an attention mechanism to generate a context vector c_t as a weighted sum of the encoder hidden states h_1, h_2,…, h_T. The attention weights $\alpha_{t,i}$ are computed using a soft attention mechanism, based on the dot product between the decoder and encoder's hidden states s_{t-1}, h_i, as (6). The RNN-based decoder input is the embedded previous estimated channel coefficient $\bar{h}_{t-1,i,}$, the context vector c_t and the previous hidden state s_{t-1} and produces the current hidden state s_t. The currently estimated channel coefficient $\bar{h}_{t,i}$ is generated from the current hidden state s_t and the context vector c_t using a feedforward NN, as shown in (4). The decoder RNN terminates after generating the final estimated channel coefficient $\bar{h}_{T,i}$. The decoder operation is summarized in Algorithm 2.

Algorithm 2. Decoder operation
1. Initialize the decoder hidden state s_0 and the previously estimated channel coefficient $\bar{h}_{0,i}$ as zero vectors.
2. For *t* = 1 to *T*, do the following:
   a. Compute the context vector c_t using Equation (5).
   b. Compute the attention weights $\alpha_{t,i}$ using Equation (6).
   c. Compute the attention scores $e_{t,i}$ using Equation (7).
   d. Update the decoder hidden state s_t.
   e. Compute the estimated channel coefficients $\bar{h}_{t,i}$ using Equation (4).
   f. Update the previously estimated channel coefficient $\bar{h}_{t-1,i}$ as $\bar{h}_{t,i}$.
3. Compute the loss function *L* using Equation (14).
4. Update the decoder parameters by backpropagating the gradients of the loss function through the network.
5. Repeat steps 2 to 4 for several epochs or until convergence is reached.

2.3.4. Attention mechanism

The attention mechanism scheme in our channel estimation NN model is responsible for selectively focusing on the most informative parts of the hidden states when predicting the channel coefficients in the decoder. The attention mechanism computes a context vector c_t as a weighted sum of the encoder's hidden states, where the weights are iteratively updated and learned by the model during training. The attention mechanism is a soft attention mechanism based on the dot product between the decoder and encoder's hidden states [24]. Specifically, we compute the attention scores $e_{t,i}$ between the decoder's hidden state s_{t-1} and the i-th encoder hidden state h_1 as (13):

$$e_{t,i}=s_{t-1}^{T}.Wh_i \tag{13}$$

where Wh_i is the weight matrix that transforms the encoder's hidden state h_i to the same dimension as the decoder's hidden state s_{t-1}. The transpose operation of the decoder hidden state vector ensures that the dimensions of the vectors match the dot product operation. The resulting attention score $e_{t,i}$ indicates the magnitude of attention that should be given to the i-th hidden state when predicting the t-th channel coefficient. A softmax activation function is applied to the attention scores to derive the equivalent attention weights $\alpha_{t,i}$. The context vector c_t is subsequently computed as a weighted sum of the encoder hidden states using the attention weights as described in (5). The context vector is concatenated with the decoder's hidden state s_{t-1} and used in the decoder scheme to predict the channel coefficients $h_{t,i}$ at time t. The encoder-decoder model with the attention mechanism is shown in Figure 3.

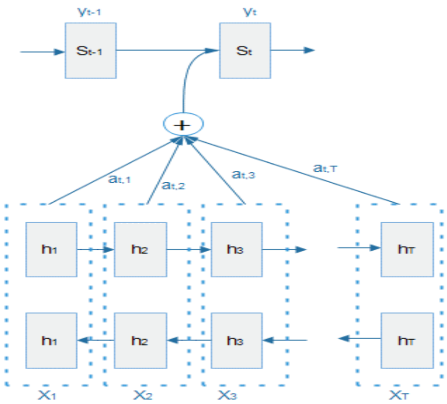

Figure 3. Encoder-decoder model with attention

2.3.5. Loss function

The loss function is a measure of the difference between the predicted output channel coefficients $\bar{H}$ and the true channel coefficients H. We compute our loss function as the MSE, which is defined as (14):

$$L=\frac{1}{T}\sum_{i,t=1}^{T}(h_{t,i}-\bar{h}_{t,i})^2 \tag{14}$$

where T represents the length of the sequence.

The MSE loss measures the average squared difference between the predicted and the true channel coefficients over the entire sequence. The goal of training the RNN is to find minimal MSE loss over the training data. The loss is iteratively minimized by adjusting the RNN parameters, and employing a backpropagation through time (BPTT) algorithm process.

3. EVALUATION PROCEDURE

3.1. Data description

To evaluate our proposed attention-based mechanism scheme, we develop the CDL-A and Rician fading channel models by defining the parameters as shown in Table 1 for the CDL-A channel model. They include the center frequency, bandwidth, time duration, delay spread, maximum Doppler shift, and spatial correlation. We select these parameters to accurately represent real-world wireless channel characteristics. Using this model, we then generate channel data for both the line-of-sight (LOS) and non-line-of-sight (NLOS) components. The LOS component accounts for the direct path between the transmitter and receiver, while the NLOS component represents the reflected and scattered paths. This combination creates a challenging scenario for channel estimation. Secondly, we incorporate the Rician fading channel model, by adding random fluctuations to the channel characteristics. The Rician factor was selected to control the strength of the LOS component relative to the NLOS component. This factor influences the severity of the fading effect, allowing us to evaluate the channel estimation method's performance under different fading conditions. Next, we generate a dataset consisting of 50,000 wireless channel samples for the two models. Each data sample represents a snapshot of the channel at a specific time instant, including the channel gains and phases for the LOS and NLOS components.

Table 1. Parameters used for the CDL-A and Rician fading channel

Channel parameters	Value
Center frequency C_f	2.4 GHz
Bandwidth	20 MHz
Time duration	1 ms
Delay spread	10 ns
Maximum Doppler shift	100 Hz
Spatial correlation	0.8
Rician factor	5

To ensure consistency within the dataset, we normalized the channel samples based on the average channel power observed in the training set. This normalization process eliminates variations in channel power levels and guarantees a fair comparison during the evaluation phase. The generated dataset has been partitioned into three sets: training, test, and validation using the 70%-15%-15% ratio. The training set was used to train the attention-based model, allowing it to learn the underlying patterns and relationships between the channel inputs and received signals. The validation set was employed to fine-tune the model's hyperparameters and monitor its performance during the training process. To prevent overfitting, we use the dropout regularization technique to randomly drop units during training, preventing co-adaptation of hidden units. We further use early the stopping mechanism to stop training when the performance on the validation set stops improving. For the hyperparameters settings, weight decay is disabled (set to 0.000), and the Adam optimizer is chosen with initial learning rate of 0.0001, a beta1 value of 0.9, and AMSGrad set to false. For the training phase, a batch size of 32 and four workers are specified, with a total of 400 training epochs. Additionally, an annealing power of 2 is applied during training, and the logging of all sigmas is set to false.

Finally, the test set served as an independent dataset to evaluate the model's performance and compare it with existing algorithms. This model adopts an encoder-decoder with attention architecture. The encoder network extracts relevant feature patterns from the observed received signal information and produces the parameters for the attention mechanism. The decoder network generates channel estimates cognizant of the outputs from the attention mechanism system and the encoder's hidden state.

3.2. Performance metrics

To quantitatively assess the performance, we compute the NMSE metric which objectively measures the quality of the estimated channel distributions as compared to the ground truth channel inputs. The NMSE of the proposed method is then compared with three baseline approaches, i.e., Wasserstein generative adversarial networks (WGAN) [25], learned denoising-based approximate message passing (LDAMP) [26], and maximum likelihood (ML) [27]. The normalized MSE is defined in (15):

$$NMSE\ [dB] = 10 \log_{10} \frac{\|H_{est} - H\|_F^2}{\|H\|_F^2} \tag{15}$$

The average SNR is calculated as ${N_t}/{\sigma_{pilot}^2}$ [28], where N_t represents the number of channels and σ_{pilot}^2 denotes the pilot size.

4. RESULTS AND DISCUSSION

The results of our experiments are presented and we provide a detailed discussion on the performance evaluation of the proposed technique compared to three cutting-edge algorithms: LDAMP [26], WGAN [29], and ML. The performance metric used in our evaluation is the NMSE between the estimated and true channel realizations. Figure 4 displays the observed MSE for the CDL-A channel model across different ranges of SNR. The results of our proposed technique show it outperforms all the comparison methods in terms of the NMSE. This indicates that our technique can estimate the channel parameters more accurately, leading to a closer match between the estimated and actual channel realizations.

Among the comparison methods, ML shows the best performance, exhibiting lower MSE values compared to WGAN and LDAMP. LDAMP utilizes a powerful data-driven algorithm with deep unrolling and end-to-end learning, which allows it to effectively capture the underlying channel characteristics. However, it is important to note that our proposed technique achieves competitive performance with ML and surpasses WGAN and LDAMP in terms of the NMSE. The competitive performance of our technique can be attributed to its ability to leverage the specific characteristics of the CDL-A channel model.

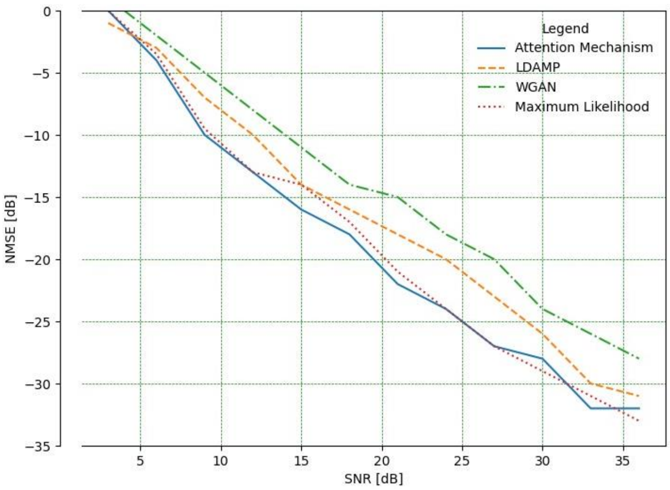

Figure 4. Channel estimation performance using methods trained on CDL-A

Furthermore, the performance of the proposed method is evaluated in the Rician fading channel model. The Rician fading channel is commonly used to model wireless communication in scenarios where there is a dominant LOS component along with multipath fading. In terms of channel estimation, the Rician fading channel presents challenges due to the presence of both deterministic and random components. Figure 5 demonstrates the performance of the different techniques across a range of SNR values. We observe that the attention mechanism-based method achieved the lowest MSE across all SNR levels, indicating its effectiveness in capturing the channel characteristics and mitigating the effects of fading. The comparison methods showed competitive performance, especially in scenarios with lower and higher SNRs. This result highlights the potential of the proposed method in capturing the intricate nature of the Rician fading channel and improving channel estimation accuracy.

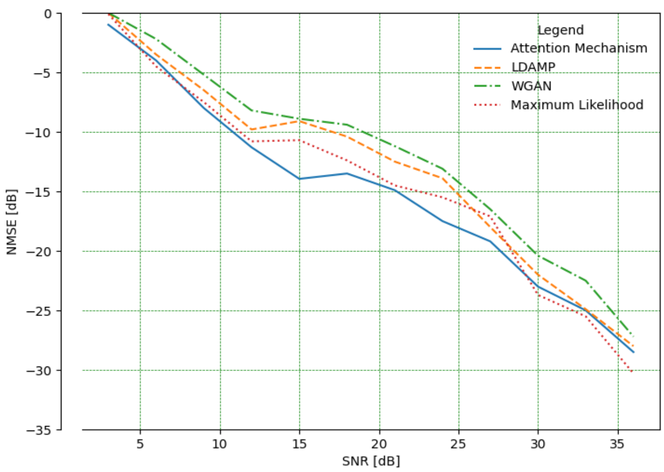

Figure 5. Channel estimation performance using methods trained on Rician fading channel

We further investigated how the variation in the number of pilots impacts the performance of the proposed technique. Figure 6 illustrates the results of our simulations, depicting the NMSE for different pilot densities. As expected, we observed a trend where decreasing the pilot density resulted in increased MSE for all the evaluated techniques. This is because a lower number of pilots provide less information about the channel, leading to less accurate estimation. In the proposed scheme, we observe that as the pilot density decreased, the NMSE increased gradually. However, the increase in NMSE was relatively moderate, indicating that even with a reduced number of pilots, this method could still provide reasonably accurate channel estimates.

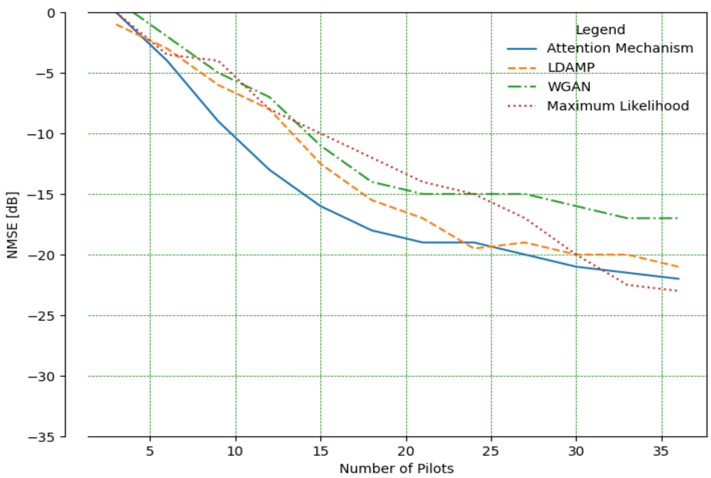

Figure 6. Channel estimation performance at different pilot densities

In contrast, the comparison methods showed more significant sensitivity to the number of pilots. As the pilot density decreased, the MSE increased more rapidly compared to the proposed method. This suggests that these techniques rely more heavily on the availability of a sufficient number of pilots to accurately capture the channel characteristics. These findings provide valuable insights for designing efficient channel estimation strategies, considering the compromise between the number of pilots and estimation accuracy. The result highlights the need to carefully choose the pilot density based on the specific requirements of the communication system and the capabilities of the estimation techniques employed. Figure 7 reveals the BER performance across all SNR conditions. These results suggest that the attention mechanism is more robust in terms of error performance, i.e., fewer errors occur during data transmission, resulting in more reliable communication. It further indicates the ability to handle a wider range of channel conditions. In real-world scenarios, channel conditions can vary significantly, and a communication system that performs well across these variations is highly desirable.

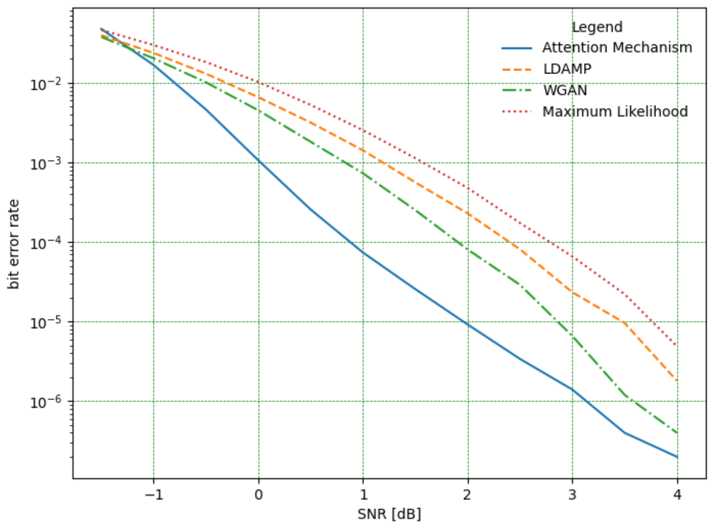

Figure 7. BER performance at different SNRs

5. CONCLUSION

This paper presents an attention-based channel estimation technique for point-to-point MIMO communication systems. The proposed technique utilizes RNN with GRUs, and attention mechanisms to further improve the performance accuracy of channel coefficient estimation. The encoder-decoder architecture captures the relevant information in the received symbols sequence and generates estimated channel coefficients. The encoder processes the received information symbols sequentially and generates a sequence of hidden states that capture only the relevant features. The decoder uses an attention mechanism to focus only on the most informative parts of the encoder's hidden states when predicting the channel coefficients. The attention scores are computed based on the dot product between the decoder's hidden state and the encoder's hidden state. The model has been trained and the loss function monitored using the MSE metric, which measures the difference between the predicted channel coefficients and the true channel coefficients. The goal of training is to minimize this loss function and consequently improve the accuracy of channel coefficient estimation.

The proposed technique offers several advantages. It can handle the effects of multipath propagation, fading, and other environmental factors on the wireless signal. The attention mechanism helps to selectively target the most important features of the received signal, leading to improved estimation performance. The use of deep learning techniques allows the model to learn complex relationships and capture long-term dependencies in the sequential data. Future work will explore variations of attention mechanisms, such as self-attention or transformer-based architectures, which have shown promising results in other domains.

REFERENCES

[1] O. D. Okey *et al.*, "Quantum Key Distribution Protocol Selector Based on Machine Learning for Next-Generation Networks," *Sustainability (Switzerland)*, vol. 14, no. 23, p. 15901, Nov. 2022, doi: 10.3390/su142315901.

[2] J. M. Kang, C. J. Chun, and I. M. Kim, "Deep Learning Based Channel Estimation for MIMO Systems with Received SNR Feedback," *IEEE Access*, vol. 8, pp. 121162–121181, 2020, doi: 10.1109/ACCESS.2020.3006518.

[3] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward Next Generation Open Radio Access Networks: What O-RAN Can and Cannot Do!," *IEEE Network*, vol. 36, no. 6, pp. 206–213, Nov. 2022, doi: 10.1109/MNET.108.2100659.

[4] S. S. Gill *et al.*, "AI for next generation computing: Emerging trends and future directions," *Internet of Things (Netherlands)*, vol. 19, p. 100514, Aug. 2022, doi: 10.1016/j.iot.2022.100514.

[5] Z. Wang, L. Liu, and S. Cui, "Channel Estimation for Intelligent Reflecting Surface Assisted Multiuser Communications: Framework, Algorithms, and Analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020, doi: 10.1109/TWC.2020.3004330.

[6] A. K. Gizzini and M. Chafii, "A Survey on Deep Learning Based Channel Estimation in Doubly Dispersive Environments," *IEEE Access*, vol. 10, pp. 70595–70619, 2022, doi: 10.1109/ACCESS.2022.3188111.

[7] J. M. Kang, C. J. Chun, and I. M. Kim, "Deep-learning-based channel estimation for wireless energy transfer," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2310–2313, Nov. 2018, doi: 10.1109/LCOMM.2018.2871442.

[8] Y. Liao, Y. Hua, X. Dai, H. Yao, and X. Yang, "ChanEstNet: A Deep Learning Based Channel Estimation for High-Speed Scenarios," in *IEEE International Conference on Communications*, vol. 2019, pp. 1–6, 2019, doi: 10.1109/ICC.2019.8761312.

[9] M. N. Drakshayini and M. R. Kounte, "A review of wireless channel estimation techniques: challenges and solutions," *International Journal of Wireless and Mobile Computing*, vol. 23, no. 2, pp. 193–203, 2022, doi: 10.1504/ijwmc.2022.126390.

[10] Y. Liao, Y. Hua, and Y. Cai, "Deep Learning Based Channel Estimation Algorithm for Fast Time-Varying MIMO-OFDM Systems," *IEEE Communications Letters*, vol. 24, no. 3, pp. 572–576, Mar. 2020, doi: 10.1109/LCOMM.2019.2960242.

[11] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep Learning-Based Channel Estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, Apr. 2019, doi: 10.1109/LCOMM.2019.2898944.

[12] Q. Bai, J. Wang, Y. Zhang, and J. Song, "Deep Learning-Based Channel Estimation Algorithm over Time Selective Fading Channels," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 125–134, Mar. 2020, doi: 10.1109/TCCN.2019.2943455.

[13] A. Abdallah, A. Celik, M. M. Mansour, and A. M. Eltawil, "Deep-Learning Based Channel Estimation for RIS-Aided mmWave Systems with Beam Squint," in *IEEE International Conference on Communications*, vol. 2022-May, pp. 1269–1275, May 2022, doi: 10.1109/ICC45855.2022.9839142.

[14] 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.

[15] G. Xu, Q. Zhang, Z. Song, and B. Ai, "Relay-Assisted Deep Space Optical Communication System over Coronal Fading Channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 6, pp. 8297–8312, Dec. 2023, doi: 10.1109/TAES.2023.3301463.

[16] G. Xu, N. Zhang, M. Xu, Z. Xu, Q. Zhang, and Z. Song, "Outage Probability and Average BER of UAV-Assisted Dual-Hop FSO Communication With Amplify-and-Forward Relaying," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 8287–8302, Jul. 2023, doi: 10.1109/TVT.2023.3252822.

[17] A. K. Gizzini, M. Chafii, A. Nimr, and G. Fettweis, "Deep Learning Based Channel Estimation Schemes for IEEE 802.11p Standard," *IEEE Access*, vol. 8, pp. 113751–113765, 2020, doi: 10.1109/ACCESS.2020.3003286.

[18] W. M. Salama, M. H. Aly, and E. S. Amer, "Enhanced deep learning based channel estimation for indoor VLC systems," *Optical and Quantum Electronics*, vol. 54, no. 9, p. 535, Sep. 2022, doi: 10.1007/s11082-022-03904-4.

[19] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[20] M. Li, Y. Wang, Z. Wang, and H. Zheng, "A deep learning method based on an attention mechanism for wireless network traffic prediction," *Ad Hoc Networks*, vol. 107, p. 102258, Oct. 2020, doi: 10.1016/j.adhoc.2020.102258.
[21] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, "CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances," *Information Sciences*, vol. 546, pp. 835–857, 2021, doi: 10.1016/j.ins.2020.09.003.
[22] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/neco_a_01199.
[23] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks," *arXiv preprint arXiv:1909.09586*, 2019.
[24] G. Brauwers and F. Frasincar, "A General Survey on Attention Mechanisms in Deep Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3279–3298, Apr. 2023, doi: 10.1109/TKDE.2021.3126456.
[25] E. Balevi, A. Doshi, A. Jalal, A. Dimakis, and J. G. Andrews, "High Dimensional Channel Estimation Using Deep Generative Networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 18–30, Jan. 2021, doi: 10.1109/JSAC.2020.3036947.
[26] C. A. Metzler, A. Mousavi, and R. G. Baraniuk, "Learned D-AMP: Principled! Neural network based compressive image recovery," *Advances in Neural Information Processing Systems*, vol. 2017-December, pp. 1773–1784, 2017, doi: 10.5555/3294771.3294940.
[27] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, Jan. 2018, doi: 10.1109/TSP.2017.2771725.
[28] M. Arvinte and J. I. Tamir, "MIMO Channel Estimation Using Score-Based Generative Models," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 3698–3713, Jun. 2023, doi: 10.1109/TWC.2022.3220784.
[29] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," *34th International Conference on Machine Learning, ICML 2017*, vol. 1, pp. 298–321, 2017, doi: 10.5555/3305381.3305404.

BIOGRPHIES OF AUTHORS

Kayode A. Olaniyi received a B.Sc. in Physical Sciences 1998 and M.Sc. degrees in Electric and Electronic Engineering from the University of Ibadan, Nigeria, in 2004. He is a researcher in the Department of Electrical and Electronic Engineering and also a member of the Institute for Intelligent Systems of the University of Johannesburg. He can be contacted at email: 217097911@student.uj.ac.za.

Reolyn Heymann received her B.Ing. Electrical and Electronic Engineering degree (cumlaude) and B.Sc. Information Technology degree at the Rand Afrikaans University in 2004. She received her M.Ing. Electrical and Electronic Engineering degree (cumlaude) from the University of Johannesburg in 2007 and in 2016 her doctorate in digital communications, specifically coding theory. She was previously the Director of the Centre for Collaborative Digital Networks in the Faculty of Engineering and the Built Environment at the University of Johannesburg, South Africa. She is the chair of the IEEE South Africa Information Theory Society and the faculty representative of the UJ IEEE-HKN honors student chapter. She can be contacted at email: reolyn.heymann@gmail.com.

Theo G. Swart received the B.Ing. and M.Ing. degrees in Electric and Electronic Engineering from the Rand Afrikaans University, South Africa, in 1999 and 2001, respectively (both cumlaude), and the D. Ing. degree from the University of Johannesburg, South Africa, in 2006. He is an associate professor in the Department of Electrical and Electronic Engineering Science and the Director of the UJ Center for Telecommunications. He is a senior member of the IEEE and was previously the chair of the IEEE South Africa Chapter on Information Theory. He was also a specialist editor in Communications and Signal Processing for the SAIEE Africa Research Journal. He can be contacted at email: tgswart@uj.ac.za.