Neural Networks for Predicting the Optimal Beamforming Angles for Maximized Overall Wireless Network Capacity

1st *Parmida Geranmayeh*
Faculty of Mathematics and Natural Sciences
Department of Computer Science
Humboldt University of Berlin,
Berlin, Germany
Parmida.Geranmayeh@hu-berlin.de
0009-0007-9837-9395

2nd *Eckhard Grass*
Faculty of Mathematics and Natural Sciences
Department of Computer Science
Humboldt University of Berlin and
IHP – Leibniz-Institut für innovative Mikroelektronik
Frankfurt (Oder), Germany
grass@ihp-microelectronics.com
0000-0002-9718-4170

***Abstract*— Artificial Intelligence (AI) and Machine Learning (ML) are set to transform wireless communication systems. By using deep neural networks and extensive data, these technologies are improving wireless channel performance, making resource allocation more efficient, and enhancing signal detection. Their impact covers areas like source coding and channel estimation, introducing a new era of efficient and adaptable wireless communication. These advancements speed up networks, improve user experiences, and make networks more reliable. AI and ML are playing a crucial role in shaping the future of mobile and wireless technologies. In this paper, we explore how neural networks, through prediction, help to determine optimal beamforming angles to maximize the total channel capacity between transmitters and receivers.**

***Keywords*— Beamforming, Network Capacity, Neural Network**

I. Introduction

Neural networks play a crucial role in driving innovation in various fields, particularly excelling in wireless communication, network optimization, and the Internet of Things (IoT) [1]. The continuous evolution of wireless communication has brought era of data abundance, propelling advancements in big data analytics and machine learning [1]. In this dynamic environment, neural networks have become powerful tools for enabling the decoding, harnessing, and optimization of this wealth of information. Their capabilities enable us to meet the increasing demands of next-generation wireless networks, improve the efficiency of community wireless communication channels, and fine-tune network performance. Additionally, they adeptly handle the complexities of big data through advanced analytics [1]. In the realm of spatial-temporal analysis, particularly within cellular network traffic, neural networks exhibit their prowess. The development of a Time-Series Similarity-based Graph Attention Network (TSGAN) showcases their superiority over conventional prediction models, emphasizing the immense potential of neural networks in enhancing predictions across various time scales [2]. As we explore the realm of future communication systems, neural networks stand out, particularly in tasks such as predicting traffic patterns and managing networks effectively. Neural networks, with their ability to learn from training data, provide valuable capabilities for network optimization, even without explicit channel estimation. This adaptability amplifies their utility [3]. Furthermore, neural networks exhibit significant potential in intelligent wireless networks, especially through innovative adaptive learning [4]. This approach employs Transfer Learning (TL) techniques to efficiently allocate network resources while addressing critical challenges in the field [4]. Neural networks also play a pivotal role in the automatic discovery of channel innovation streamlines the optimization process, particularly benefiting massive Multiple Input Multiple Output (MIMO) systems by directly optimizing key system metrics, eliminating the need for explicit channel estimation [5]. Additionally, in the communication systems fields, machine learning, including neural-network-based reinforcement learning, contributes significantly. This specialized field maximizes data return, optimizes bandwidth utilization, and minimizes power consumption, underscoring the adaptability and efficiency introduced by neural networks in the communication systems domain [6]. In the dynamic world of the Internet of Things (IoT), Graph Neural Networks (GNNs) play a crucial role, proving to be powerful tools for analyzing IoT data. They offer deep insights into various IoT sensing environments, as evident in a comprehensive review [7].

In the domain of computer vision-based IoT systems, Convolutional Neural Networks (CNNs) play a central role [7]. Known for their ability to identify patterns in images, CNNs (Convolutional Neural Networks) greatly improve the understanding and optimization of IoT data. This underscores the important role of neural networks in advancing modern technology [7]. In summary, neural networks occupy a prominent position in driving innovation across diverse domains, including wireless communication, network optimization, and IoT. Their adaptability, efficiency, and data-driven learning capabilities make them indispensable tools for addressing the multifaceted challenges posed by evolving technologies. The primary goal of this article is to present the efficacy and potential of neural networks to achieve the maximum total channel capacity in a wireless communication system. The utilization of neural networks in this context serves as a demonstrative example of their predictive capabilities. By using the power of neural networks, we aim to highlight their competence in making informed predictions, ultimately leading to the enhancement of channel capacity. This research emphasizes to underscores the practical and scientific vaslue of neural networks as a tool for optimization and prediction in the domain of wireless communication.

2. DESIGN PARAMETERS

In this section, we detail the fundamental aspects of our implementation. We conduct an exhaustive search involving a substantial number of combinations, specifically 390,625 in total. The computation time is 3 days with 72 CPUs and 2 threads per core. This extensive exploration is achieved by iterating through a range of -60 to 60 degrees with a 5-degree step for each of the four antennas. This results in a total of 25 values for each antenna, leading to the large number of 25×25×25×25 combinations. For each combination, the total channel capacity is calculated, and these values are stored by our algorithm. The combination that yields the maximum total channel capacity is sought as our primary goal. The environment is visualized in a specific area within Berlin, namely Northeast Berlin, as depicted in Figure 1. The weather condition is characterized by clear air with standard atmospheric parameters, and the buildings are modeled with concrete material. The wireless communication system operates at 28 GHz, with a fixed bandwidth of 2 GHz. A Uniform Rectangular Array (URA) is created using a rectangular patch antenna element, configured with 8 rows and 4 columns, and half-wavelength spacing between elements. The positions of the transmitter (Tx) and receiver (Rx) are defined as Tx1, Tx2, Rx1, and Rx2. Angles ranging from -60 to 60 degrees are considered in our calculations, with the goal of identifying the optimal configuration.

Tx1: (52.561075, 13.503874, 10)
Tx2:(52.560802, 13.504149, 10)
Rx1: (52.560834, 13.504560, 1)
Rx2: (52.561001, 13.503810,1)

The "ray-tracing" method is employed for channel modeling, utilizing the shooting and bouncing ray method (SBR) with a maximum one reflection. In this step, the channel capacity of the wireless communication system for different combinations of transmitter (Tx) and receiver (Rx) angles is evaluated. The goal is to ensure that the channel capacity is assessed. It should be noted that the setup is not specific to any unique city conditions; instead, a standard layout available in the MATLAB environment is utilized. This approach guarantees the robustness of our findings even with changes in city configurations or antenna positions. Within a nested loop structure, antenna patterns for various transmitter and receiver angles are calculated using the "beamsteer" function from MATLAB's Phased Array System Toolbox. This process entails creating a steering vector to represent the spatial response of the antenna array, defining scan angles for the main lobe beam, computing weights for beamforming based on these scan angles, and ultimately calculating the radiation pattern of the antenna array.

Fig. 1. (a) The city layout model. (b) The antennas placement in the model city [8].

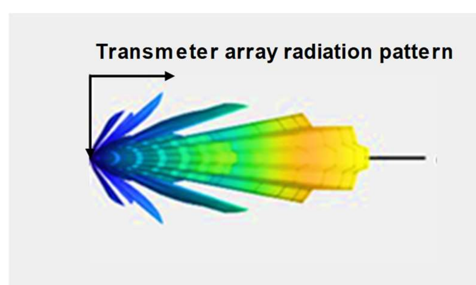

Fig. 2. The digitized antenna radiation pattern.

Different radiation patterns representing antenna directionality and power distribution in 3D space are generated by manipulating scan angles and weights.

Antenna elements, designed using classes such as "patchMicrostrip" and "phased.CosineAntennaElement," are integrated into array configurations to generate the desired radiation patterns, as depicted in Figure 2. Functionalities such as beam steering and beamforming within phased array systems are enabled by these elements. Following pattern generation, custom antenna elements and transmitter-receiver objects are created. The channel characteristics are calculated using the "ray-trace" function, and each ray is processed to extract channel parameters. The total power received and channel capacity for each transmitter-receiver pair are then calculated, and this information is stored in arrays. The iteration through combinations continues until the optimal combination, representing the highest total channel capacity achievable across all antennas, is found.

3. IMPLEMENTATION

In this phase, we will develop four distinct neural network models. Among these models, two employ the sigmoid activation function, while the other two utilize the hyperbolic tangent activation function. As depicted in Figure 3, these activation functions are utilized in both one-layer (a) and two-layer (b) neural network architectures. For the one-layer models, we have configured them to comprise 14 neurons. In contrast, the two-layer models consist of 14 neurons in the first layer and 11 neurons in the second layer.

Additionally, the Identity function, designated as the output layer, has been integrated into these models. To facilitate the training and validation process, 70% of the available data has been allocated. The remaining 30% is reserved for evaluating the model's performance and assessing its accuracy.

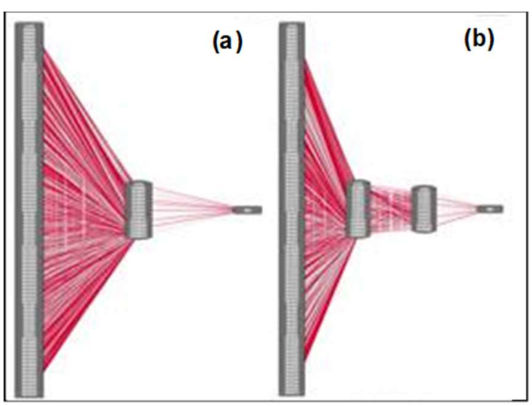

Fig. 3. 1-Layer and 2-Layer Neural Network Architectures.

It's important to emphasize that all the data has been standardized, a process involving the rescaling of scale-dependent variables, to ensure consistency in the model's training and testing processes. To enhance the model's predictive accuracy and minimize errors, we've employed the gradient descent method that has been determined as an optimization algorithm [9]. Batch training was employed for our extensive dataset because it is particularly beneficial for large data volumes. Moreover, batch training offers the advantage of shorter training time. Furthermore, the learning rate coefficient, which influences the size of parameter updates in each iteration, has been set at 0.1. Finally, we've defined the number of training epochs as 2000, ensuring an adequate number of iterations for the model to converge and capture patterns within the data.

As depicted in Figure 4, the 2-layer sigmoid model exhibits the highest average error, amounting to 1.7%. Following closely is the 1-layer sigmoid model with an average error of 1.6%. Both of these sigmoid models underperform when compared to their hyperbolic counterparts. Notably, the single-layer hyperbolic tangent model stands out with an impressive performance, yielding a mere 1.1% average error. However, it's the two-layer hyperbolic tangent model that truly excels, boasting a remarkable 0.2% average error. This model has proven itself as a highly accurate predictor of channel capacities. It's worth emphasizing that, as evident from the results presented in Figure 4, increasing the number of layers and perceptrons in a neural network doesn't invariably lead to performance improvements. It's vital to carefully consider the most efficient and optimal number of layers and perceptrons for the model. This is because the relationship between network complexity and performance is nuanced and calls for a thoughtful, data-driven approach.

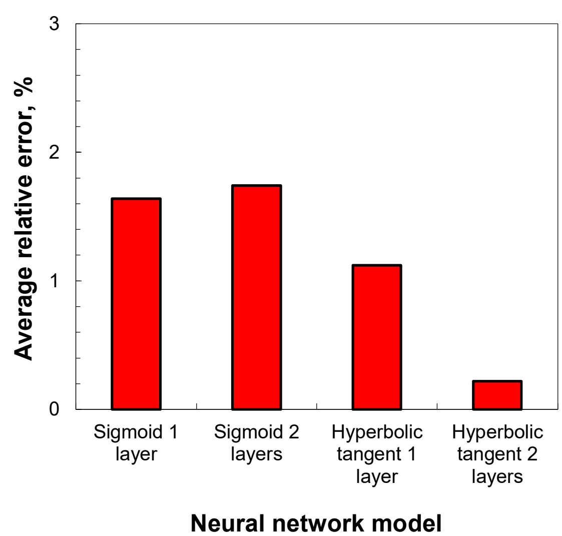

Fig. 4. Performance Analysis of Neural Network Models.

It's important to note that over-fitting has not happened in these models. Firstly, over-fitting typically occurs when the dataset is limited in size, but in our case, we have a substantial amount of data. Secondly, over-fitting tends to happen when a neural network has a large number of layers. However, we've employed either one or two layers, keeping our model relatively simple. Additionally, we've imposed limits on the number of training epochs, further preventing over-fitting.

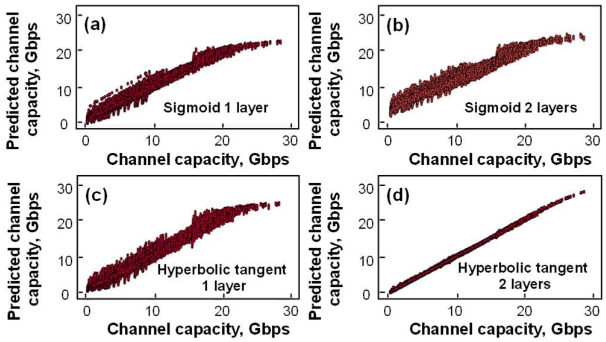

Fig. 5. Models Comparison for Channel Capacity Predictions.

Figure 5 displays the numerical predictions of channel capacity for each data point, categorized by the respective models. Notably, the two-layer hyperbolic activation function outperforms the single-layer hyperbolic function, demonstrating the superior accuracy of the hyperbolic family in channel capacity estimation. In these plots, the horizontal axis represents the actual data values, while the vertical axis represents the predicted values per Gbps. The visual representation highlights the effectiveness of the two-layer hyperbolic tangent model in closely approximating real values, forming a curve that closely resembles the identity line (y=x). Following the two-layer hyperbolic model, the one-layer hyperbolic and one-layer sigmoid models also exhibit curves with minimal deviations from the ideal. In contrast, the two-layer sigmoid model demonstrates lower estimation accuracy when compared to the other models.

Figure 6 illustrates the relationship between predicted channel capacity and residuals. On the X-axis, you'll find the predicted channel capacity values, each corresponding to a specific data point. The Y-axis represents residuals, which are the disparities between observed channel capacity and the predicted values. This plot serves as a tool for evaluating the accuracy of the models' predictions. When the residuals are close to zero ($y = 0$), it signifies that the model's predictions closely align with the actual channel capacity. Positive residuals indicate overestimation, while negative residuals suggest underestimation. These plots reveal that the two-layer hyperbolic activation function model exhibits the highest accuracy. This is evident by the concentration of residuals close to zero, indicating minimal prediction deviations. In contrast, one-layer and two-layer sigmoid models demonstrate weaker performance, as their residuals are more scattered, suggesting varying degrees of prediction accuracy. Totally, these plots help us discern the accuracy and reliability of the different models, with the two-layer hyperbolic model emerging as the most accurate choice.

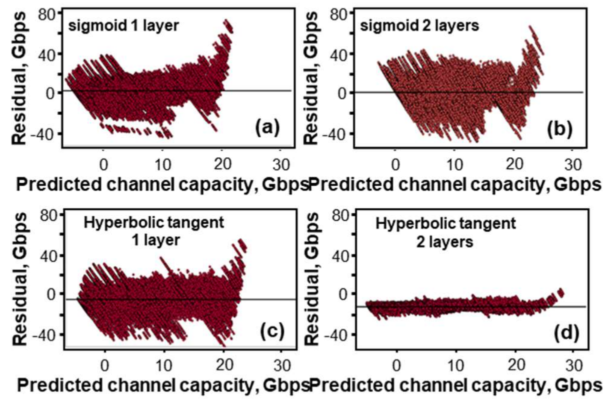

Fig. 6. Models Accuracy Assessment via Residual Analysis.

Figure 7 illustrates the maximum total channel capacity predicted by our models. Our main goal is to design neural network models that identify optimal beamforming angles for maximizing total channel capacity, rather than seeking the maximum value within the dataset. Nonetheless, comparing these models can be important for finding the maximum total channel capacity. In this figure, the hyperbolic tangent models with one layer predicts 29.1 Gbps and two layers predicts 29.4 Gbps as the maximum channel capacity. Subsequently, the one-layer sigmoid model forecasts 28.9 Gbps, while the two-layer sigmoid model predicts 29.0 Gbps. Notably, the maximum total channel capacity is 29.5 Gbps. This reiterates the point that increasing the number of neural network layers, despite its associated costs and complexity, doesn't guarantee enhanced efficiency and predictive performance.

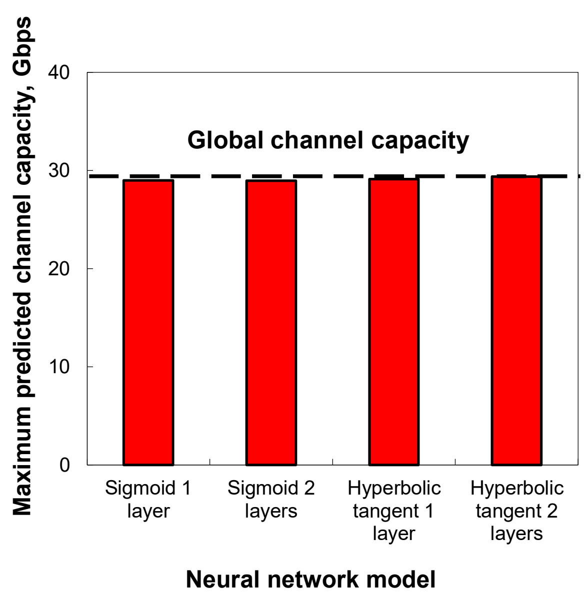

Figure. 7. Maximum total channel capacity comparison among neural network models.

In the next phase, we decided to change the weather condition. Given the superior precision and increased anticipated channel capacity demonstrated by the 2-layer hyperbolic tangent model, we opted to utilize this model for predicting the total channel capacity during rainy weather. To accomplish this, we randomly selected 100 sets of angles for 2 transmitters and 2 receivers, inputting them into the model. By considering the weights and the number of perceptrons, we derived an output indicating the optimal angles for both transmitters and receivers in this scenario. As shown in Figure 8, this process was repeated 20 times, resulting in an impressive 90% accuracy. In rainy conditions, the maximum total channel capacity is 29.3 Gbps, and our model successfully recommended antenna angles that achieved a near-perfect total capacity with average of 26.5 Gbps. It should be noted that the duration of each prediction and execution of each run has been reduced to less than 2 minutes.

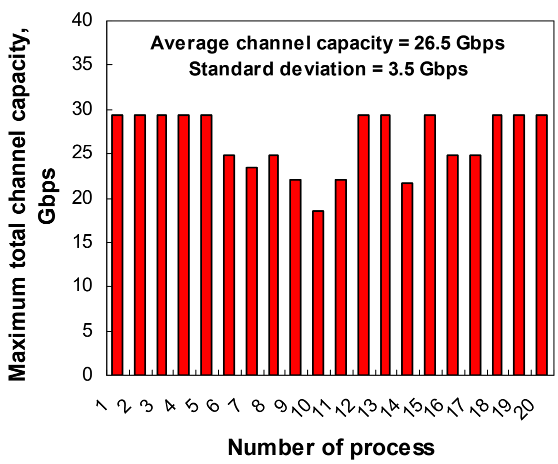

Fig. 8. Prediction of total channel capacity in rainy weather by using a 2-layer hyperbolic tangent model.

4. CONCLUSION AND FUTURE WORK

This paper investigates the role of neural networks in predicting the optimal beamforming angles to maximize channel capacity between transmitters and receivers. At first, four neural network models are developed, with two using sigmoid activation and two employing hyperbolic tangent activation. These models are configured with one or two layers, and the data is standardized for consistency. The gradient descent method and a learning rate coefficient of 0.1 are employed to enhance predictive accuracy. Results show that the two-layer hyperbolic tangent model outperforms other models in prediction with a 0.3% average error, emphasizing that increasing the number of layers and perceptrons doesn't always lead to improved performance. Subsequently, in the second phase, we applied our trained and designed model to predict the total channel capacity during rainy weather. Our findings revealed that our model accurately forecasted channel capacity under rainy conditions with an impressive 90% accuracy. Hence, the article explores the important role of neural networks in facilitating the prediction of channel capacity and emphasizes their significance in achieving the result. For future research, employing reinforcement learning for all antennas is a potential approach. This approach enables antennas to autonomously learn optimal angles for transmission and reception with zero knowledge of the environment, and they effort to maximize the total channel capacity through autonomous learning.

ACKNOWLEDGMENT

The authors thank the Deutsche Forschungsgemeinschaft (DFG), Project number 457407152, for providing financial support of the research project entitled “5GEnabled Real Time Communications for "Tactile Internet" (5G-Remote)”.

REFERENCES

[1] P. Rohini, S. Tripathi, C.M. Preeti, A. Renuka, J.l.A. Gonzales. D. Gang, “A study on the adoption of wireless communication in big data analytics using neural networks and deep learning”, 2nd Int. Conf. on Adv. Comput. and Innovat. Tech. in Eng. (ICACITE), India, April 2022, DOI: 10.1109/ICACITE53722.2022.9823439.

[2] Z. Wang, J. Hu, G. Min, Z. Zhao, Z. Chang, Z. Wang, “Spatial-temporal cellular traffic prediction for 5G and beyond: A graph neural networks-based approach”, IEEE Trans. Ind. Inf., vol. 19, pp. 5722-5731, April 2023.

[3] G. Dong, M. Tang, Z. Wang, J. Gao, S. Guo, L. Cai, R. Gutierrez, B. Campbell, L.E. Barnes, M. Boukhechba, “Graph neural networks in IoT: A survey”, March 2022, DOI: 10.48550/arXiv.2203.15935.

[4] A. Thantharate, C. Beard, “Adaptive6G: Adaptive resource management for network slicing architectures in current 5G and future 6G systems”, J. Netw. Syst. Manag., vol. 31, October 2022, DOI: 10.1007/s10922-022-09693-1.

[5] W. Yu, F. Sohrabi, T. Jiang, “Role of deep learning in wireless communications”, IEEE BITS Inf. Theory Mag., vol. 2, pp. 56-72, DOI: October 2022, DOI: 10.1109/MBITS.2022.3212978.

[6] P. Ferreira, R. Paffenroth, A.M. Wyglinski, T.M. Hackett, S.G. Bilén, D. J. Mortensen, “Reinforcement learning for satellite communications: from LEO to deep space operations”, IEEE Commun. Mag., vol. 57,pp. 70-75, May 2019.

[7] J. Suárez-Varela, P. Almasan, M. Ferriol-Galmés, K. Rusek, F. Geyer, X. Cheng, X. Shi, Sh. Xiao, F. Scarselli, A. Cabellos-Aparicio, P. Barlet-Ros, “Graph neural networks for communication networks: context, use cases and opportunities”, IEEE Netw., vol. 37,pp. 70-75, May/June 2023, DOI: 10.1109/MNET.123.2100773.

[8] Terrain source: GMTED2010, Data available from the U.S. geological survey, source: Esri, Maxar, Earthstar Geographics, and the GIS user community, (2010).

[9] M.J. Kochenderfer, T.A. Wheeler, “Algorithm for optimization”, MIT press, ISBN: 978-0262039420, (2019).