Deep Learning-Enabled Enhancements for Reliable and Efficient mmWave Communication.

Nielsen, Martin Hedegaard

DOI (link to publication from Publisher):
10.54337/aau695998374

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Nielsen, M. H. (2023). *Deep Learning-Enabled Enhancements for Reliable and Efficient mmWave Communication.* [PhD thesis, The Technical Faculty of IT and Design]. Aalborg Universitetsforlag. https://doi.org/10.54337/aau695998374

DEEP LEARNING-ENABLED ENHANCEMENTS FOR RELIABLE AND EFFICIENT MMWAVE COMMUNICATION

BY
MARTIN HEDEGAARD NIELSEN

DISSERTATION SUBMITTED 2023

AALBORG UNIVERSITY
DENMARK

Deep Learning-Enabled Enhancements for Reliable and Efficient mmWave Communication

Ph.D. Dissertation

Martin Hedegaard Nielsen

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7B
DK-9220 Aalborg

Dissertation submitted:	December 2023
PhD supervisor:	Associate Professor Ming Shen Aalborg University
PhD committee:	Associate Professor Gilberto Berardinelli (chairman) Aalborg University, Denmark Professor Nuno Borges Carvalho Universidade de Aveiro, Portugal Head of Department Alvaro Valcarce Nokia Bell Labs, France
PhD Series:	Technical Faculty of IT and Design, Aalborg University
Department:	Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-577-8

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2023

This thesis is written using LaTeX, and simulations have been run in either Matlab or Python for the entire thesis.

Abstract

As the next generation of wireless communication unfolds, 5G and the anticipated 6G networks are poised to revolutionize the industry by integrating with Low Earth Orbit (LEO) satellite systems. This integration, however, brings forward significant challenges, particularly in achieving power-efficient transmission within the constraints of satellite communication. This thesis delves into the intricacies of such systems, presenting innovative deep learning solutions to enhance their performance and reliability.

Drawing from the advancements in complex-valued neural networks, this work first introduces an autoencoder framework for power-efficient satellite communications, leveraging a Deep Complex Convolutional Network (DCCN) as a decoder. By employing a two-stage training scheme, the model adapts to varying channel conditions and front-end nonlinearities, achieving efficient transmission across different operational statuses without compromising bit error rates. This approach is substantiated through experimentation with a 28 GHz active phased array, demonstrating the potential of neural networks in handling non-linear OFDM waveforms.

Expanding on this, a novel hybrid model utilizing transfer learning is proposed to fortify end-to-end OFDM neural receivers against the complexities of multiple channel environments and nonlinear transmitters. The robustness of the DCCN is significantly enhanced, reducing training times and improving error rates, as validated through practical measurements on advanced hardware platforms.

In the critical domain of fault diagnosis within active phased arrays—vital for the operation of 5G and 6G radios—the thesis introduces a groundbreaking Deep Neural Network (DNN) approach. This method simplifies the traditionally complex diagnostic process, yielding high accuracy in detecting single and multiple element failures swiftly and effectively, even in challenging signal-to-noise scenarios. Such a development paves the way for practical, on-site deployment, reducing reliance on elaborate measurement setups.

Finally, addressing the pervasive issue of link blockages in mmWave communications, the thesis presents a cutting-edge application of Liquid Time Constant (LTC) networks. This biological ODE-inspired system demonstrates exceptional accuracy in predicting imminent blockages using only received signal power, marking a significant step towards

ensuring low-latency and reliable communication.

Collectively, these contributions underscore the transformative impact of deep learning in enhancing the robustness, efficiency, and diagnostic capabilities of next-generation wireless communication systems. This thesis not only offers a suite of novel methodologies but also bridges theoretical research with practical engineering applications, embodying the innovative spirit of a young engineer forging the path to a more connected future.

Resumé

Med udrulningen af næste generation af trådløs kommunikation står 5G og den forventede 6G-netværk klar til at revolutionere industrien gennem integration med lavbanesatellitsystemer (LEO). Denne integration medfører imidlertid væsentlige udfordringer, især når det gælder om at opnå energieffektiv transmission inden for rammerne af satellitkommunikation. Denne afhandling dykker ned i kompleksiteterne af sådanne systemer og præsenterer innovative dyblæringsløsninger for at forbedre deres præstationer og pålidelighed.

Med udgangspunkt i fremskridt inden for komplekse neurale netværk introducerer dette arbejde først en autoencoder-ramme for energieffektiv satellitkommunikation ved at benytte et Dybt Komplekst Konvolutionsnetværk (DCCN) som dekoder. Ved at anvende et todelt træningsskema tilpasser modellen sig til varierende kanalforhold og front-end ikke-lineariteter, hvilket opnår effektiv transmission på tværs af forskellige operationelle statusser uden at gå på kompromis med bitfejlratene. Denne tilgang er understøttet af eksperimenter med et 28 GHz aktivt fasestyret array og demonstrerer neurale netværks potentiale i håndteringen af ikke-lineære OFDM-bølgeformer.

Udvidelsen af dette omfatter en ny hybridmodel, der benytter transfer learning for at styrke end-til-end OFDM neurale modtagere mod kompleksiteterne af flere kanalomgivelser og ikke-lineære sendere. Robustheden af DCCN forbedres markant, hvilket reducerer træningstider og forbedrer fejlrate, som det er valideret gennem praktiske målinger på avancerede hardwareplatforme.

I det kritiske domæne for fejldiagnosticering inden for aktive fasestyrede arrays—vitalt for driften af 5G og 6G-radioer—introducerer afhandlingen en banebrydende Dyb Neuralt Netværk (DNN)-tilgang. Denne metode forenkler den traditionelt komplekse diagnostiske proces og giver høj nøjagtighed i opdagelsen af enkelte og flere elementfejl hurtigt og effektivt, selv under udfordrende signal-til-støj-scenarier. En sådan udvikling banebryder vejen for praktisk, on-site udrulning, hvilket reducerer afhængigheden af komplekse måleopsætninger.

Endelig, med hensyn til det udbredte problem med linkblokeringer i mmWave-kommunikation, præsenterer afhandlingen en banebrydende anvendelse af Liquid Time Constant (LTC)-netværk. Dette biologisk ODE-inspirerede system demonstrerer en-

estående nøjagtighed i forudsigelsen af forestående blokeringer ved kun at bruge modtaget signalkraft, hvilket markerer et væsentligt skridt mod at sikre lav-latens og pålidelig kommunikation.

Samlet set understreger disse bidrag den transformative indvirkning af dyblæring i forbedringen af robustheden, effektiviteten og diagnostiske kapaciteter af næste generations trådløse kommunikationssystemer. Denne afhandling tilbyder ikke blot en suite af nye metoder, men bygger også bro mellem teoretisk forskning og praktisk ingeniørarbejde, og indkapsler den innovative ånd hos en ung ingeniør, der baner vejen mod en mere forbundet fremtid.

Contents

Thesis Details

Thesis Title:	Deep Learning-Enabled Enhancements for Reliable and Efficient mmWave Communication.
Ph.D. Student:	Martin Hedegaard Nielsen
Supervisors:	Assoc. Prof. Ming Shen, Aalborg University Prof. Elisabeth De Carvalho, Aalborg University

This thesis is submitted in fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) from Aalborg University, Denmark. The thesis is structured as a collection of scientific papers, some of which are published in peer-reviewed journals and conferences. These papers contribute directly or indirectly to the extended summary of the thesis. The thesis represents three years of cross-sectional research in Antennas, Propagation, and Millimetre-wave Systems (APMS) and Connectivity at the Department of Electronic Systems, Aalborg University, Denmark.

This work was partly supported by the Department of Electronic Systems as part of the cross-section fellowship and partly by the "Innovations Fund Denmark" project of MARS2 (Modular Advanced Radio for Satellite Services).

The main body of this Ph.D. thesis consist of the following papers:

[A] Nielsen, Martin. H., De Carvalho, Elisabeth. & Shen, Ming., "A Two-stage Deep Learning Receiver for High Throughput Power Efficient LEO Satellite System with Varied Operation Status," *IEEE Access*, vol. 10, pp. 60904-60913, 2022.

[B] Nielsen, M. H., De Carvalho, E. & Shen, M., "Adapting to Nonlinear Transmitters with Hybrid Model Training for Neural Receivers," *IEEE Transactions on Cognitive Communications and Networking*, , vol. 9, no. 6, pp. 1657-1665, Dec. 2023.

[C] Nielsen, M. H., Zhang, Y., Xue, C., Ren, J., Yin, Y., Shen, M. & Pedersen, G. F., "Robust and Efficient Fault Diagnosis of mm-Wave Active Phased Arrays using Baseband Signal," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp. 5044-5053, July 2022.

[D] Martin H. Nielsen, Chia-Yi Ya, Ming Shen, Muriel Médard, "Blockage Prediction in Directional mmWave Links Using Liquid Time Constant Network," *2023 48th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, Montreal, QC, Canada, 2023, pp. 1-2.

According to the Ministerial Order no. 1039 of August 27, 2013, regarding the Ph.D. Degree, paragraph 12, article 4, a statement from each co-author about the Ph.D. students' contribution to the above-listed papers has been provided to the Ph.D. school for approval before the thesis submission. These co-author statements have also been presented to the Ph.D. committee and included in their assessment.

In addition to the main papers as the main content of this thesis, the following publications were also co-authored during the Ph.D. study. Since these papers are not part of the main body of this thesis, they have not been included in print. The reader is, therefore, kindly referred to their respective publishing channels as listed hereafter.

[1] Nielsen, M. H., Jespersen, M. H. & Shen, M., "Remote Diagnosis of Fault Element in Active Phased Arrays using Deep Neural Network," *2019 27th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 2019, pp. 1-4.

[2] Ojaroudi Parchin, N., A. Abd-Alhameed, R., Li, Y., Nielsen, M. H., & Shen, M. (2019). "Enhanced-Gain Dual-Polarized Slot Antenna with a Frequency Selective Surface." 2019 27th Telecommunications Forum (TELFOR) Belgrade, Serbia

[3] Ojaroudi Parchin, N., A. Abd-Alhameed, R., Li, Y., Nielsen, M. H., & Shen, M. (2019). "High-Performance Yagi-Uda Antenna Array for 28 GHz Mobile Communications. I 2019 27th Telecommunications Forum (TELFOR) (s. 1-4). IEEE.

[4] Feridoon Jalili, Martin. H. Nielsen, M. Shen, O. K. Jensen, J. H. Mikkelsen and G. F. Pedersen, "Linearization of Active Transmitter Arrays in Presence of Antenna Crosstalk for 5G Systems," 2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC), Helsinki, Finland, 2019, pp. 1-5, doi: 10.1109/NORCHIP.2019.8906927.

[5] Qingyue Chen, Yubo Wang, Arun Yadav, Patrick C. F. Eggers, Martin H. Nielsen, Yufeng Zhang, Ming Shen "Efficient Detection of Rare Beacon Events in GEO Satellite Communication Systems using Deep Learning" 2021 IEEE MTT-S International Wireless Symposium 10.1109/IWS52775.2021.9499607

[6] Yunfeng Li, Yonghui Huang, Martin H. Nielsen, Feridoon Jalili, Wei Wei, Jian Ren, Yingzeng Yin, Ming Shen, Gert F. Pedersen "A Cross-Mode Universal Digital Pre-

Distortion Technology for Low-Sidelobe Active Antenna Arrays in 5G and Satellite Communications" MDPI Electronics

[7] Qirui Hua, Martin H. Nielsen, et al. "Lab to Multi-Scene Generalization for Non-Line-of-Sight Identification with Small-Scale Dataset" IEEE Transactions on Artificial Intelligence doi: 10.1109/TAI.2023.3262763.

[8] Yuqing Xu, Zeliang An, Martin Hedegaard Nielsen, Ming Shen "Adversarial Attacks on Deep Learning based Identification of GaN Power Amplifiers" Transactions on Microwave Theory and Techniques

[9] Wei Wei, Martin H. Nielsen, Mathias H. Hannesbo, Ming Shen, & Gert F. Pedersen "Nonlinearity-Tracking Digital Predistortion of Active Phased Arrays in 5G Systems" submitted to IEEE Transaction on Industrial Electronics 2023

[10] Yunfeng Li, Yonghui Huang, Martin H. Nielsen, Feridoon Jalili, Wei Wei, and Ming Shen "A Cross-Mode Digital Pre-Distortion Technique for Linearization of RF Power Amplifiers" submitted to IEEE Transactions on Circuits and Systems II: Express Briefs 2023

Acknowledgement

When I embarked on my PhD journey, my aspiration was to make a significant contribution to the field of communication. Having just graduated with a Master's degree, I felt a thirst for research in engineering. Four years later, I am immensely grateful for the path I have taken and the multitude of places it has led me to. I wholeheartedly appreciate the opportunity I said yes to.

I am greatly indebted to my supervisor, Dr. Ming Shen, who introduced me to the intricacies of research, providing ample freedom and involving me in diverse projects. I am thankful to my co-supervisor, Prof. Elisabeth De Carvalho, for their patient guidance and engaging discussions, and to Prof. Gert Frølund Pedersen for funding my scholarly activities.

The camaraderie and collective efforts of my colleagues in the Antenna Propagation and Millimeter wave Systems (APMS) section have profoundly influenced my research journey. The invaluable feedback during paper drafting has helped sharpen my ideas. My deepest appreciation goes to them.

To Ben Krøyer, who welcomed me into his office for my initial years at AAU, our insightful discussions and laughs have been instrumental to my growth.

In addition, I would like to extend my heartfelt gratitude to Professor Muriel Médard and her team at Massachusetts Institute of Technology for hosting me. Muriel's wisdom, patient guidance, and constructive feedback have enriched my research perspective. The time spent with her team was filled with enlightening discussions, insightful feedback, and a deep sense of camaraderie.

I extend a special acknowledgment to my family, girlfriend and friends. Their unwavering support and encouragement have been my strength throughout this journey.

Lastly, I appreciate my assessment committee for their time, the Department of Electronic Systems for enabling my research, and the APMS section for their support.

Martin Hedegaard Nielsen
Aalborg University, December 11, 2023

Part I

Introductory Chapters

1 Introduction

In an era where the digitization of our world is accelerating, the imperative for faster and more reliable wireless communication systems intensifies. As we stand on the cusp of 6G networks and witness the expansion of satellite communication, the significance of millimeter-wave (mmWave) technologies has surged [1]. These technologies, with their high bandwidth and frequency capabilities, are becoming indispensable in leading the wireless communication revolution. Amidst the transformative wave of telecommunications, marked by relentless efforts to enhance data rates, minimize latency, and enrich user experiences, the power efficiency of power amplifiers has emerged as a crucial factor. These amplifiers, responsible for boosting signals from base stations to mobile devices, have seen their efficiency catapulted through advances in semiconductor materials, such as Gallium Nitride (GaN), and sophisticated linearization techniques like digital predistortion (DPD) [2, 3]. Remarkably, GaN-based power amplifiers have evolved from efficiencies near 50% in the early 2000s to surpassing 70% in recent times [4], addressing the heightened power demands of 5G and forthcoming 6G networks. Moreover, the foray of Artificial Intelligence (AI) into this domain promises further enhancements by compensating for non-linearities and dynamically optimizing amplifier operations for efficiency [5].

1.1 5G and Beyond

The advent of 5G, succeeding 4G LTE, has catalyzed a new telecommunications epoch by delivering faster, more robust networks capable of supporting a growing multitude of devices [6–8]. It introduces three pivotal services: Enhanced Mobile Broadband (eMBB), for heightened data throughput and coverage; Massive Machine Type Communications (mMTC), integrating IoT to link innumerable low-power devices; and Ultra-Reliable Low Latency Communications (URLLC), critical for latency-sensitive applications like autonomous vehicles and telesurgery [8]. The deployment of Frequency Range 2 (FR2), spanning 24.25 to 43.5 GHz, along with massive MIMO technologies, emboldens 5G to achieve elevated data rates and expansive coverage, albeit presenting efficiency challenges, particularly given cost and size constraints [8]. Addressing these efficiency concerns, especially in power consumption, is vital as the requisites for managing in-band and out-of-band distortions become increasingly stringent. This underscores the pivotal role of active devices, such as power amplifiers, which must operate within nonlinear regimes to attain optimal power efficiency. However, the emergence of ultra-wide bandwidths, extending to 400 MHz, introduces complex AMAM and AMPM gain distortions due to memory effects, challenging conventional DPD techniques and prompting the need for more sophisticated solutions [9–12].

1.2 Motivation

Denmark's strategic investments in the space sector, AI, and telecommunications aim to spur economic growth and societal advancement [13–15]. These initiatives are poised to foster a robust AI ecosystem, guarantee access to high-speed internet, upgrade telecommunication services, and leverage space data. Nevertheless, the quest for higher data transmission speeds and extended mobile device battery life underscores the significance of energy efficiency in wireless communication [16–19]. At the hardware level, this translates into a delicate interplay between the nonlinear behavior and overall energy efficiency of communication devices. The intrinsic trade-off between these elements has been a longstanding challenge [9–12], with nonlinear transmission leading to signal distortion, intermodulation distortion, and erroneous demodulation, thus impairing the integrity of wireless communication systems. To mitigate these challenges, the integration of deep learning techniques into mmWave communication systems is proposed.

However, the promise of mmWave technologies is tempered by significant reliability and efficiency challenges [16, 17, 20]. These include blockage susceptibility, channel management complexity, transmitter non-linearities, and the difficulty in diagnosing faults within active phased arrays (APAs) [16, 18, 21]. Overcoming these obstacles necessitates innovative solutions, and deep learning, as a potent subfield of AI, offers considerable promise in addressing complex problems [22–24]. Its capability for pattern recognition, environmental adaptation, and predictive accuracy can be pivotal in surmounting mmWave system challenges [5, 17, 20, 21]. Moreover, these advances, when coupled with the increasing availability of computational resources and data, could lead to breakthroughs that would dramatically shift the current paradigm of wireless communication.

1.3 Thesis Objectives and Methodology

This thesis sets out to apply and explore deep learning techniques to enhance the reliability and efficiency of mmWave front-end systems. The central research queries guiding this work are as follows:

1. How can complex-valued deep learning approaches enhance power-efficient transmission in 6G and satellite systems?

2. What strategies can be developed to handle multiple channels and nonlinear transmitters in mmWave communication effectively?

3. How can deep neural networks be utilized for quick and precise fault diagnosis in active phased arrays?

4. How can ordinary differential equation neural networks, specifically Long Short-Term Memory (LSTM)-based networks, be applied for precise blockage prediction in mmWave front-ends?

These research questions are addressed through a collection of scientific papers, each tackling specific aspects of the overarching objectives. Since different communication environments inherently require different adaptations, our approach aims to develop deep learning techniques and models that exhibit versatility across diverse scenarios.

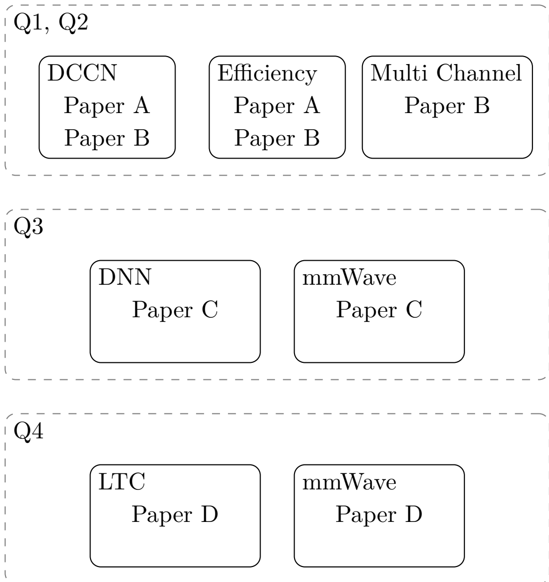


Fig. 1: Relationship between research questions, methods, and papers included in the thesis.

The research questions are addressed by employing various deep-learning techniques and models, which are discussed in detail throughout this thesis. For a more comprehensive understanding of these questions and how they are tackled in various scenarios, please refer to Figure 1.

1.4 Thesis Outline

The thesis is divided into two main parts. Part I, the remainder of which is structured as follows: Part I begins with Chapter 2, offering a detailed examination of the essential theoretical foundations. This chapter delves into the complexities of terrestrial and satellite communication systems. Chapter 3 emphasizing the application of deep learning for various purposes. These include channel estimation, addressing nonlinear transmitter challenges, predicting blockage status, and diagnosing faults specifically in mmWave communication. Chapter 4 presents the measurement setup and measurements conducted to have real world data to both simulate from and test the different neural network applications.

Chapter 5 presents detailed summaries of the individual research papers that form the core of the thesis. Each paper tackles different aspects of the primary research questions, contributing significantly to a more profound understanding of the topic.

Chapter 6 concludes Part I, drawing insights from the research and suggesting future research directions in this dynamic and advancing area.

Part II of the thesis, which constitutes its primary contributions, is a collection of scientific papers. These papers are methodically arranged to correspond with different elements of mmWave communication. Papers A and B explore complex-valued deep learning to enhance system efficiency and adaptability. Papers C and D, on the other hand, are centered on fault diagnosis and blockage prediction in mmWave systems. This structure allows for a systematic examination of how deep learning can be innovatively applied in the field of mmWave communication.

2 Wireless communication for Next-Generation Networks

In the realm of modern telecommunication, the development of new wireless technologies has experienced a rapid expansion. Among these technologies, 5G and 6G networks, along with satellite communication, have emerged as significant advancements with the potential to revolutionize connectivity and necessitate enhanced physical layer architectures. Concurrently, the applications of artificial intelligence, particularly in neural networks, have seen a substantial increase, offering innovative solutions to enhance the physical layer. This chapter aims to provide a comprehensive overview of these developments.

2.1 OFDM Modulation in 5G NR

Orthogonal Frequency Division Multiplexing (OFDM) is a critical modulation scheme adopted by the 5G New Radio (5G NR) standard [8]. As a multicarrier modulation technique, OFDM divides the available bandwidth into multiple orthogonal subcarriers, with each subcarrier transmitting a portion of the data.

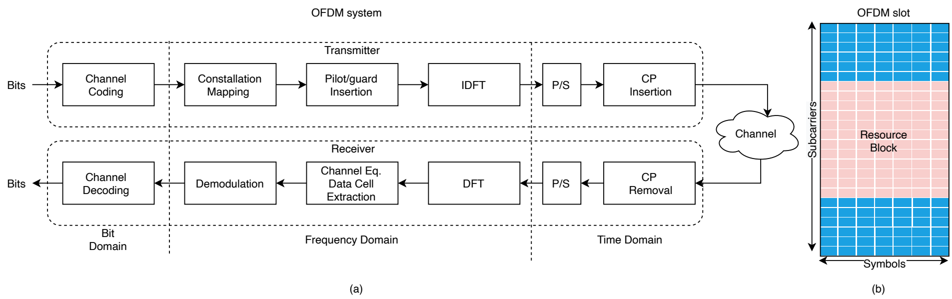


Fig. 2: a) Block diagram of a conventional communication system [25]. b) OFDM symbol grid, with the x-axis representing symbols and the y-axis representing subcarriers [25].

OFDM offers several advantages within the context of 5G NR:

- **Efficient Spectrum Utilization:** OFDM minimizes interference through the orthogonality of its subcarriers, enabling efficient spectrum use.
- **Flexibility:** The support for variable subcarrier spacing in 5G NR allows for the accommodation of various use cases and deployment scenarios, striking a balance between coverage and capacity.

- **Robustness Against Multipath Fading:** The inherent structure of OFDM provides resilience against multipath fading, which is prevalent in wireless communication.
- **Support for MIMO:** OFDM integrates seamlessly with Multiple Input Multiple Output (MIMO) technologies in 5G NR, facilitating increased data rates and link reliability.

The cyclic prefix (CP) is a critical element in OFDM, crucial for maintaining subcarrier orthogonality in multipath environments. 5G NR supports both normal and extended CP lengths to accommodate diverse propagation conditions. A collection of OFDM symbols forms a resource element (RE), and a collection of REs constitutes a Resource block, as illustrated in Figure 2. A communication slot comprises F consecutive OFDM symbols, allocating P pilots and D REs for data transmission. The length of a time-domain OFDM symbol is $S = N + N_{\text{cp}}$, where N_{cp} represents the CP length. Under m-ary modulation, an IQ sample carries m bits, with the constellation size being 2^m.

Conventional Demodulation

In 5G NR systems, demodulation is a critical process for accurate data transmission and reception. A conventional algorithm used for this purpose is the Linear Minimum Mean Square Error (LMMSE) estimator [26]. The LMMSE estimator for a received signal r can be represented as :

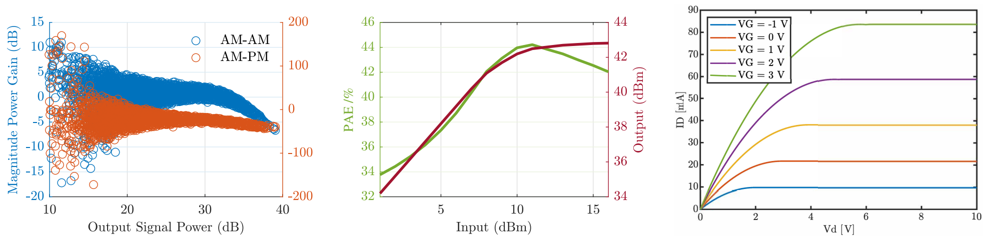
$$\hat{x} = \mathbf{R}_{\text{rx}}\mathbf{R}_{\text{xx}}^{-1}r, \tag{1}$$

where $\mathbf{R}_{\text{rx}}$ is the cross-correlation matrix between the received and the original signal, $\mathbf{R}_{\text{xx}}$ is the auto-correlation matrix of the original signal, and $\hat{x}$ is the estimated transmitted signal [26]. The LMMSE algorithm aims to minimize the mean square error between the estimated and actual transmitted signals. This method offers a robust demodulation mechanism within 5G NR systems, as illustrated in Figure 2. With the advent of neural receivers, there is potential for optimization in the demodulation process, leveraging machine learning to enhance performance beyond traditional algorithms [5].

2.2 Enhancing Power Efficiency in Communication Systems

Power amplifiers face the challenge of maintaining a delicate balance between linearity and efficiency. The 3GPP standardization mandates linear operation, which inevitably results in reduced efficiency [8].

The 5G modulation scheme, based on OFDM, optimizes frequency band usage while operating at high power. A significant challenge in designing a transmitter with a non-constant envelope is maintaining linearity while adhering to standardization requirements. For a 5G downlink 64-QAM baseband signal, the Error Vector Magnitude (EVM) is defined to be below 12.5% points and the Adjacent Channel Leakage Ratio (ACLR) below -28 dBc [27, 28].



(a) (b) (c)

Fig. 3: Nonlinear effect on a) Amplification in the AiP [25, 29], b) Power added efficiency of the AiP [25, 29], c) operation points in the GaN HEMT PA [25, 29].

To address this, various linearization techniques have been employed to enable the power amplifier to operate in nonlinear zones while adhering to standardization requirements and maintaining satisfactory efficiency. The industry has long utilized various effective methods to achieve these objectives, such as the Doherty Power Amplifier, Envelope Tracking (ET), and Digital Predistortion (DPD) [9, 12].

The scope of linearization extends beyond a single power amplifier with the advent of more recent technological advancements like MIMO and beamforming, critical components in 5G and 6G. It now involves linearizing an entire array of power amplifiers, each connected directly to the smaller patches of an active phased array antenna. This advanced structure necessitates the application of additional linearization methods, the specifics of which depend on the arrangement of these systems. The introduction of

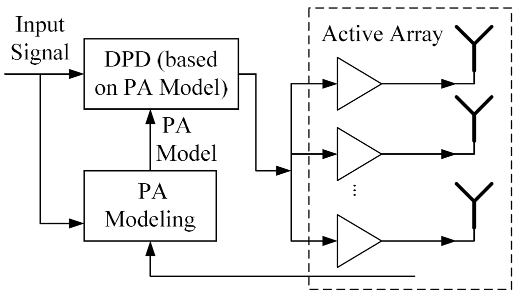
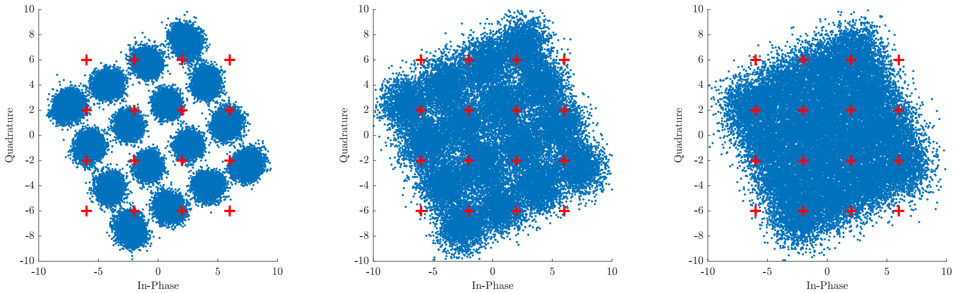


Fig. 4: Conventional Digital Predistortion in APA.

neural networks offers the possibility of re-engineering the conventional communication system using AI to solve some of the challenges in 5G communication systems.



(a) (b) (c)

Fig. 5: AiP nonlinear distortion effect as expressed in the received constellation. a) Nonlinear distorted signal effect on symbols [25]. b) Nonlinear distortion in AWGN channel [25]. c) Nonlinear distortion in a flat fading channel [25].

Transistor Variation

Transistors used in power amplifiers exhibit variations due to the manufacturing process, which directly impacts their non-linearities. Factors such as channel doping, gate oxide thickness, and gate line roughness directly impact the behavior of the transistor [30] [31]. This is evident in the voltage threshold behavior of a transistor, expressed as:

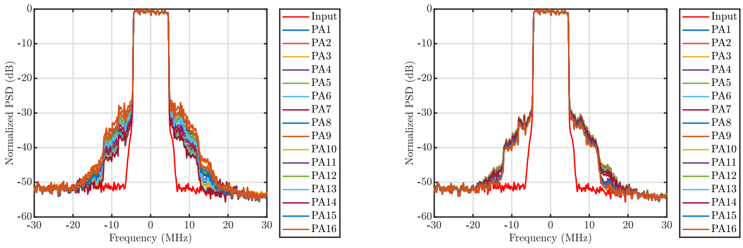
$$\sigma_{V_{Th}} = 3.19 \cdot 10^{-8} \frac{T_{ox} \cdot N_A^{0.4}}{\sqrt{L_{eff} \cdot W_{eff}}} \tag{2}$$

where N_A is the channel doping, T_{ox} is the oxide thickness, and L_{eff}, W_{eff} are the effective length and width of the channel respectively [30].

The drain current is a function of the relationship between gate voltage and voltage threshold, as expressed in the following equations:

$$I_D = \mu_n C_{ox} \frac{W}{L} [(V_{gs} - V_{Th}) V_{DS} - \frac{{V_{DS}}^2}{2}] \tag{3}$$

$$I_D = \mu_n C_{ox} \frac{W}{2L} (V_{gs} - V_{Th})^2 \tag{4}$$



(a) (b)

Fig. 6: Power spectrum of 16 PA's with a) $V_{gs} = -2.7$ V and B) $I_{DQ} = 100$mA.

where C_{ox} is oxide capacitance, μ_n is the electron mobility of the material, V_{gs} is the gate voltage, and V_{DS} is the voltage from drain to source.

The oxide capacitance, C_{ox}, can be calculated using the permittivity of the oxide and the thickness of the oxide:

$$C_{ox} = \frac{\epsilon_{ox}}{T_{ox}} \tag{5}$$

Equations (2) and (4) can be plotted for different V_{gs} and V_{DS} values, as shown in Figure 3(c).

Figure 3(c) illustrates that when the transistor enters saturation, it is linear, which is not entirely accurate. The equations are a simplified model of a transistor, and the lines are therefore linear when, in practice, they are increasing slightly in the saturated area.

Crucially, equations (2) and (4) demonstrate that only if C_{ox} is equal for each transistor will these equations yield the exact same value. Since C_{ox} varies due to manufacturing processes, it will cause a difference between transistors that results in different drain currents when the same gate voltage is applied to the transistors. This variation underscores the different non-linearities exhibited by transistors due to manufacturing variations as shown in Figure 6.

2.3 Satellite Communication

Satellite communication is going to plays a pivotal role in global telecommunication, offering reliable connectivity, particularly in remote areas where traditional networks fall short. As 6G technology unfolds, the focus has shifted toward integrating terrestrial and non-terrestrial networks, including satellite systems, to bolster future global connectivity [21, 32, 33].

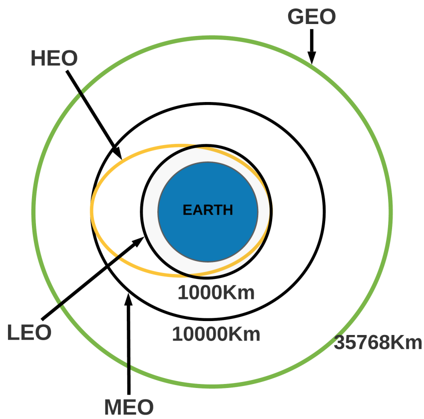


Fig. 7: Illustration of the orbits and altitudes of different types of communication satellites.

Fundamentally, satellite communication involves transmitting signals over vast distances between ground stations and onboard communication systems. This transmission process is characterized by the sending and receiving of signals at fixed intervals, dictated by the satellite's speed and orbit position.

There exist three primary categories of communication satellites: Geostationary Earth Orbit (GEO), Low Earth Orbit (LEO), and Medium Earth Orbit (MEO). GEO satellites, located approximately 36,000 kilometers above Earth, provide continuous coverage of specific regions and are ideal for applications like GPS. LEO satellites, positioned closer to Earth (around 1,200 kilometers above the surface), offer lower latency and can handle data-intensive tasks, such as 6G communication [6, 25, 34, 35]. However, achieving continuous coverage necessitates a network of LEO satellites due to their rapid movement. MEO satellites, positioned between 8,000 to 20,000 kilometers above Earth, strike a balance between GEO's coverage and LEO's low latency.

Ground Stations

Ground stations are integral to satellite communication systems, serving as the primary nodes for transmitting and receiving signals to and from satellites. These stations are equipped with a high-gain antenna and the necessary equipment for signal processing [33].

A ground station's task is tracking the satellites it communicates with. This process is relatively straightforward for Geostationary Orbit (GEO) satellites, but it becomes more complex for Low Earth Orbit (LEO) and Medium Earth Orbit (MEO) satellites due to their movement relative to the Earth's surface. For our study, we focus on downlink communication, which involves the transmission of signals from a satellite to a ground station on Earth.

The link budget can be calculated as:

$$P_{\text{received}} = P_{\text{transmitted}} + G_{\text{transmitter}} - L_{\text{path}} + G_{\text{receiver}} \tag{6}$$

This makes it possible to calculate the link budget of a satellite communication system.

Table 1: Link Budget example for Satellite Communication

Parameter	Value
$P_{\text{transmitted}}$	20 dBW
$G_{\text{transmitter}}$	30 dBi
L_{path}	150 dB
G_{receiver}	45 dBi

During a satellite overpass, the downlink changes. The direct line of sight path varies in distance and can potentially be obstructed [25]. This phenomenon is depicted in Figure 8.

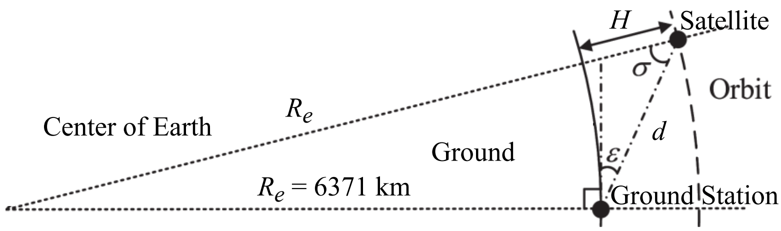


Fig. 8: Geometrical illustration of the satellite to ground data transmission link [25].

The distance between the satellite and the ground station fluctuates throughout the overpass, leading to changes in the communication link over time [25]. This variation can be mathematically described using the following equations:

$$(H + R_e)^2 = R_e^2 + d^2 - 2R_e d \cos(\frac{\pi}{2} + \epsilon_0), \tag{7}$$

where ϵ denotes the elevation angle, d represents the distance of data transmission, H is the orbital altitude of the LEO satellite, and R_e is the Earth's radius.

The distance can then be expressed as a function of ϵ:

$$d(\epsilon) = R_e[\sqrt{(\frac{H + R_e}{R_e})^2 - \cos^2(\epsilon) - \sin(\epsilon)}]. \tag{8}$$

Subsequently, the free space path loss can be formulated as:

$$L(\epsilon_0) = (\frac{4\pi d(\epsilon)}{\lambda})^2 = (\frac{4\pi f d(\epsilon)}{c})^2, \tag{9}$$

Given that the noise is fixed, the variation in Signal-to-Noise Ratio (SNR) regardless of distance can be expressed as:

$$\Delta\frac{S}{N}(dB/Hz) = (\frac{S}{N}_{5^\circ}) - (\frac{S}{N}_{90^\circ}) = \Delta L(\epsilon). \tag{10}$$

For a LEO satellite, the variation is approximately 12 dB in the link budget in the transmission window from 5° to 90° [36].

To optimize the efficiency of the satellite downlink, the transmitter should be able to adjust the transmission power such that the power received remains constant. Achieving this without non-linear operation is challenging. Therefore, driving the amplifier non-linearly can help ease the link budget.

Historically, the industry has leaned on pre-distortion techniques to enhance power efficiency and address non-linear amplifier responses. These techniques span a wide spectrum, from memory polynomial models leveraging lookup tables to innovative deep learning methodologies [7, 9, 10, 12, 36–38]. Memory polynomial models, in particular, employ an equation-based approach to describe non-linear behavior. They rely on a feedback loop between the power amplifier's input and output, using inverse polynomials to modify the input signal, thereby achieving the desired efficiency and linearity [9, 10, 12].

However, implementing these techniques in satellites presents unique challenges. The intricacies of establishing feedback loops in highly integrated active phased arrays, coupled with the potential changes in PAPR due to pre-distortion, make the optimization process complex and resource-intensive [37].

Uplink for Terrestrial Communication System

Terrestrial communication systems, particularly in uplink communication, involve the transmission of signals from a device to a base station. The link budget for the terrestrial system was calculated using a similar expression as the satellite system, but with considerations for terrestrial interference and obstacles.

Both the downlink for satellite communication and the uplink for terrestrial communication systems share the fundamental feature of transmitting signals from a device (satellite or mobile handset) to a receiving station (ground station or base station). While the principles of communication remain consistent, the challenges and parameters differ. Satellite communication in LEO has a reduced path loss due to the closer proximity of the satellite to the ground station. In contrast, terrestrial communication

Table 2: Link Budget for Terrestrial Communication

Parameter	Value
$P_{\text{transmitted}}$	0 dBW
$G_{\text{transmitter}}$	3 dBi
L_{path}	100 dB
G_{receiver}	15 dBi

systems often face challenges related to interference from buildings, trees, and other obstacles and the established link or channel will change rapidly. However, the mathematical representation, especially the link budget calculation, remains strikingly similar for both systems. This highlights the universality of communication principles, regardless of the medium or distance of transmission.

3 Communication Systems in the 6G Era

The advent of 5G NR advanced and the transition towards 6G represent a paradigm shift in communication systems, characterized by a significant increase in speed, reduced latency, and greater capacity [20]. These advancements are made possible by leveraging a range of technologies, including enhanced modulation schemes, massive MIMO (Multiple Input Multiple Output) systems, and the adoption of higher frequency bands, such as millimeter waves (mmWave) [39]. The integration of deep learning models such as DNNs and CNNs is also playing a pivotal role in improving various aspects of these complex networks [5].

3.1 Deep Learning in Communication Systems

The integration of deep learning techniques, such as Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs), into communication systems, has opened new frontiers for optimizing various aspects of wireless networks [16, 20]. DNNs, composed of multiple processing layers, are adept at modeling complex relationships in data by learning high-level abstractions. CNNs, a specialized kind of DNN, are particularly effective in handling data with a grid-like topology, such as images, making them suitable for tasks that involve spatial hierarchies in the data.

In the context of communication systems, DNNs can be trained to perform tasks such as signal detection, error correction, and channel estimation, while CNNs can be applied to analyze spectrum data, detect signal modulation types, and improve spatial filtering for MIMO systems. These deep learning models can be integrated as neural receivers to enhance the performance and reliability of 5G and 6G wireless communication networks.

Deep Neural Networks (DNNs)

DNNs are multilayered artificial neural networks with the capability to learn non-linear and complex representations of data [22]. They consist of an input layer, several hidden layers, and an output layer. Each layer comprises neurons that apply non-linear transformations to the input data, allowing the network to learn from a vast amount of unstructured data. DNNs have been instrumental in advancing areas such as computer vision, natural language processing, and autonomous systems.

Convolutional Neural Networks (CNNs)

CNNs are a category of deep neural networks that utilize convolutional layers to process data with a known grid-like structure [22, 24]. The convolutional layers apply a convolution operation to the input, passing the result to the next layer. This allows CNNs to take advantage of the spatial structure of the input data, making them particularly effective for image and video recognition tasks. In wireless communications, CNNs can analyze the spatial features of incoming signals for enhanced processing and interpretation.

The subsequent sections of this chapter will delve into the application of these advanced deep learning techniques to optimize and revolutionize the physical layer of wireless communication systems.

3.2 Neural Receivers

With the dawn of deep learning in communication systems has ushered in innovative solutions tailored for the intricacies of 6G communication [20]. The allure of deep learning lies in its ability to adapt to dynamically changing channels, as evidenced by the success of autoencoders in revamping communication systems, transmitters, or receivers [5, 40, 41]. Specifically, autoencoders have been championed for their prowess in facilitating energy-efficient communication systems [40]. To further enhance the energy efficiency of front-end transmitters, strategies such as Peak-to-average power ratio reduction and pre-distortion have been employed [42].

In the context of wireless communication systems, the term "neural receivers" is used to describe the application of AI as a receiver in a communication system.

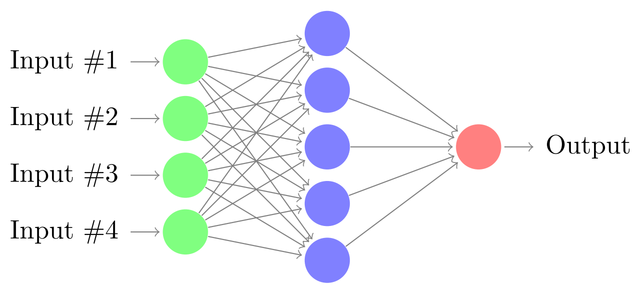


Fig. 9: Neural Network architecture, hidden layer (Blue) will change mathematically depending on DNN or CNN.

Here two different neural network structeres are important. Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs). DNNs consist of multiple layers of interconnected neurons, including a visible layer for data input and several hidden layers for complex computations and feature extraction. These networks excel at deciphering complex patterns within vast and intricate datasets, thanks to their deep architectures and numerous trainable parameters.

CNNs, a specialized kind of DNN, are tailored for processing data that has a grid-like topology, such as an image, which can be considered as a grid of pixels. The defining feature of CNNs is their convolutional layers, which apply a convolution operation to the input, passing the result to the next layer. This operation allows the network to be less sensitive to the location and orientation of features in the input, making it especially useful for tasks like object detection in images, where the object can appear in different places.

While neural network techniques have showcased superior performance [5, 34, 40–46], a prevailing assumption is the static nature of front-end nonlinearity at the receiver side [47, 48]. This assumption becomes a bottleneck when deploying neural receivers, given the dynamic nature of both the front-end and the environment post-training. The work in [49] offers a glimpse into the potential of deep learning-based receivers, particularly the 1D-TRNet, which exhibits resilience to nonlinear clipping distortion in OFDM systems. However, its limitations in handling multiple channels and nonlinear transmitters cannot be overlooked.

Neural receivers are trained using end-to-end methods, leveraging a conventional baseband description for the received signal [5]. The end-to-end system is mathematically represented as

$$\mathbf{Y} = \mathbf{H} * \mathbf{X} + \mathbf{W} \tag{11}$$

where $\mathbf{Y}$ is the received signal, $\mathbf{H}$ is the channel matrix, $\mathbf{X}$ is the transmitted symbol matrix, and $\mathbf{W}$ is the additive white Gaussian noise with variance σ^2 per element, and $*$ is the operation of convolution [29]. The channel model can be represented as

Fig. 10: Neural receiver example

a multipath fading channel, which can be mathematically modeled as a linear finite impulse response (FIR) filter. The front-end model is based on the popular memory polynomial model, which can describe the nonlinear distortion introduced by a nonlinear system.

The proposed architecture uses the input signal, $\mathbf{y}_{cp}$, and can the neural receiver can then utilise different neural network layer to remove the CP, do DFT like C-Conv layer that converts the signal $\mathbf{y}$ to the frequency domain $\mathbf{Y}$. This makes the neural receiver able to do both the channel estimation and compensate for the AiP nonlinearity [29].

Rationale for Neural Receiver Architecture

The architecture of the neural receivers deployed in this thesis was meticulously crafted to counteract the inherent nonlinearity of mmWave communication systems. By integrating DNNs and CNNs, the receivers capitalize on their respective strengths. DNNs for their pattern recognition capabilities across complex data structures and CNNs for their efficiency in processing grid-like data inputs, such as those found in signal processing applications.

The decision to employ a memory polynomial model for the front-end was influenced by its established efficacy in capturing nonlinear distortions prevalent in mmWave systems. This model serves as the foundation for the neural receiver's ability to accommodate the dynamic nature of mmWave channels and front-end hardware.

3.3 Training Neural Receivers

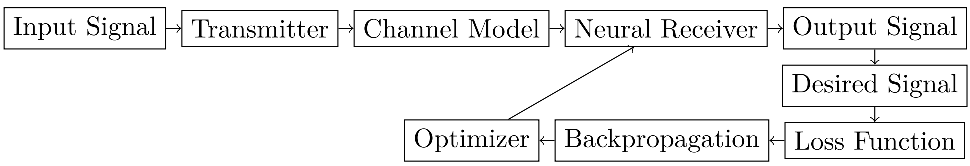


Fig. 11: Training diagram of End-to-end neural network with optimizer

Training a neural receiver involves using an end-to-end system that leverages a conventional baseband description for the received signal. The system encapsulates the effects of channel fading distortions and nonlinearity on the signal.

The amplification coefficients, which have a significant effect on the output, are computed via reverse modeling, depending on the peak-to-average power ratio and power levels of the baseband signals. The variations in channel fading distortions and nonlinearity effects on the IQ signal are illustrated in Figure 5.

In mathematical terms, the system is designed identical to Equation (10). Thus the transmission scheme defines which elements in $\mathbf{X}$ are used for pilots. The symbols, carrier-modulated based on the constellation (for example, QAM, QPSK), are assumed to have an average energy of one prior to the nonlinear front end [5]. Utilising a loss function it is possible to train the neural receiver. The total loss function is defined as the cross-entropy loss function,

$$\ell(\mathbf{b}, \tilde{\mathbf{b}}, \Phi) = -\ln\left(\frac{e^{\mathbf{b}_k}}{\sum_i e^{\tilde{\mathbf{b}}_i}}\right) + \epsilon \ell_{reg}(\Phi), \tag{12}$$

where k represents the index of the target bit, and $\epsilon < 1$ is a small constant.

3.4 Fault Detection in 5G and 6G Radios

The reliable operation of wireless communication systems hinges on their ability to detect and mitigate faults promptly. In the context of 5G and future 6G systems, maintaining seamless connectivity is paramount, making fault detection a critical aspect of system design.

Traditional fault detection methods, such as threshold-based techniques, have limitations when it comes to real-time detection and adaptability to dynamic network conditions. As wireless networks evolve and diversify, these methods may fail to provide robust fault detection solutions.

One approach is the Relative Entropy Variance (REV) method [**?**], which leverages statistical techniques to detect faults in wireless communication systems. This method focuses on monitoring statistical parameters and identifying deviations from expected values. By doing so, it can detect anomalies and potential faults in real-time, enhancing the fault detection capabilities of 5G and 6G radios.

The mathematical expressions involved in the REV method typically involve vector rotations and transformations. The electric field vector $\vec{E}$ can be represented in Cartesian coordinates as $\vec{E} = E_x\hat{\imath} + E_y\hat{\jmath} + E_z\hat{k}$, where E_x, E_y, E_z are the components of the electric field along the x, y, and z axes, respectively.

During the calibration process, the electric field vector is rotated by a known angle θ about a specific axis. The rotation matrix R can transform the electric field vector from its original orientation to its new orientation after rotation. For example, a rotation about the z-axis can be represented by the rotation matrix:

$$R_z(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{13}$$

The rotated electric field vector $\vec{E'}$ can then be obtained by multiplying the original electric field vector $\vec{E}$ by the rotation matrix R:

$$\vec{E'} = R\vec{E} \tag{14}$$

This rotation and transformation of the electric field vector are essential steps in the REV method for calibrating the APAs and accurately determining the electric field components.

The calibration of APAs using the REV method is crucial for ensuring accurate and reliable electric field measurements. However, the presence of faults in the active phased antenna arrays can lead to erroneous measurements and degrade the performance of the system. Traditional fault detection methods may not be effective in identifying subtle faults or faults that occur in complex patterns. Therefore, there is a need for a more advanced and robust fault detection technique.

3.5 Link Blockage Prediction for mmWave 5G Systems

mmWave frequencies are a key component of 5G communication systems, as they offer higher bandwidth and enable higher data rates compared to sub-6 GHz frequencies. However, mmWave signals are highly susceptible to blockage by obstacles such as buildings, vehicles, and even the human body [39, 50–53]. This can lead to link blockage, where the communication link between the transmitter and receiver is temporarily lost or degraded. Link blockage is a significant challenge for mmWave 5G systems, as it can lead to increased latency, reduced throughput, and degraded quality of service.

Predicting link blockage is crucial for optimizing the performance of mmWave 5G systems [16, 39]. Accurate blockage prediction can enable proactive handover to another base station or frequency before the blockage occurs, thereby minimizing the impact on the communication link. Various approaches have been proposed for link blockage prediction in mmWave 5G systems, including machine learning-based methods and statistical models [50–53].

Statistical models use statistical analysis of past blockage events to predict the likelihood of blockage in a given environment [52, 53]. Statistical models can be based on empirical data or theoretical models of mmWave propagation and blockage. Statistical models are computationally efficient and can be used for real-time blockage prediction. However, the accuracy of statistical models may be limited by the availability and quality of empirical data or the assumptions made in the theoretical models.

Machine learning-based methods use historical data and machine learning algorithms to predict link blockage [50–53]. For example, a machine learning model can be trained on a dataset of past blockage events and the corresponding environmental conditions, such as the location of the transmitter and receiver, the presence of obstacles, and the movement of people and vehicles. Once trained, the model can predict the likelihood of blockage for a given set of environmental conditions. Machine learning-based methods can provide accurate blockage predictions, but they require a large amount of training data and may not perform well in environments that are significantly different from the training data.

Justification of Predictive Methods

To address the challenge of link blockage in mmWave 5G systems, we implemented both statistical and machine learning-based predictive methods. Statistical models were utilized for their computational efficiency, which is paramount for real-time applications. These models are grounded in empirical data and theoretical models of mmWave propagation, providing a solid basis for prediction in stable environments with consistent blockage patterns. However, their performance may degrade in dynamic settings due to the limited scope of empirical data and the rigidity of their theoretical assumptions.

On the other hand, machine learning-based methods were selected for their adaptability and potential accuracy improvements. These models can uncover complex patterns within the historical data, allowing for more nuanced predictions that consider a multitude of environmental factors. Despite their higher computational demands and substantial data requirements, these models are crucial for maintaining system performance in highly variable environments. The limitations of machine learning models, particularly when applied to scenarios not represented in the training data, were mitigated by incorporating a diverse range of blockage scenarios during the training phase.

4 Measurement Techniques

This Section comprehensive overview of the experimental setup and data collection procedures employed throughout the research. The section begins with a detailed description of the measurement setup used for various investigations within the thesis. Subsequently, the experiments and data collection processes are discussed, followed by an exploration of fault scenarios.

4.1 Measurement Setup

The measurement setup, illustrated in Figure C.5, played a pivotal role in various investigations within this thesis. A 3GPP downlink OFDM modulated waveform was generated, featuring specific characteristics such as 64-QAM sub-carrier modulation, a sub-carrier spacing of 60 kHz, 1584 active sub-carriers, and a peak-to-average power ratio (PAPR) of 11.6 dB. This waveform was created using the R&S SMBV100B Vector Signal Generator, centered at 3 GHz. The signal bandwidths ranged up to 100 MHz, with a maximum sampling rate of 600 MHz [25, 29].

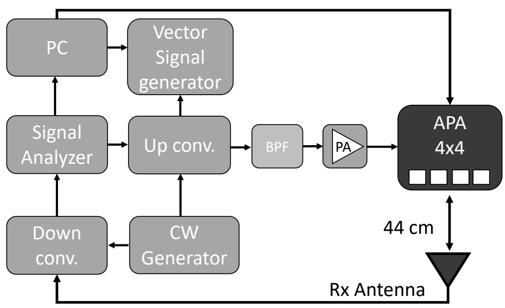


Fig. 12: Block diagram of the measurement setup [25].

In this experiment setup, a continuous wave signal was generated using an Agilent E3247C, set to emit at 12.5 GHz. This frequency was then doubled to 25 GHz via the MITEQ-MAX2M200400. The generated signal functioned as the local oscillator for both up-converting and down-converting operations, which were executed using two distinct active mixers, KTX321840 and KRX321840. These mixers were key in transitioning signals between 28 GHz and 3 GHz [25, 29].

To minimize local oscillator leakage, a band-pass filter was utilized. Additionally, a pre-amplifier was employed to deliver adequate power to the Anokiwave AWMF-0158 AiP. This AiP featured a complex arrangement of 16 branches, each including attenuators, phase shifters, and power amplifiers, as well as a 4x4 patch antenna array [25, 29].

The modulated signal was then transmitted over the air and received by an observation antenna positioned at a considerable distance. Post reception, the signal, down-converted to 3 GHz, was processed to baseband using a signal analyzer [25, 29].

To control the baseband I and Q samples for signal generation and capture, a host PC running MATLAB was utilized. The main beam of the array was controlled through a code-book and software tools, ensuring that all devices operated within their highly linear regions. The primary source of nonlinearity in the setup was associated with the APA as the device-under-test. Figure 13 provides a visual representation of the measurement setup in the lab.

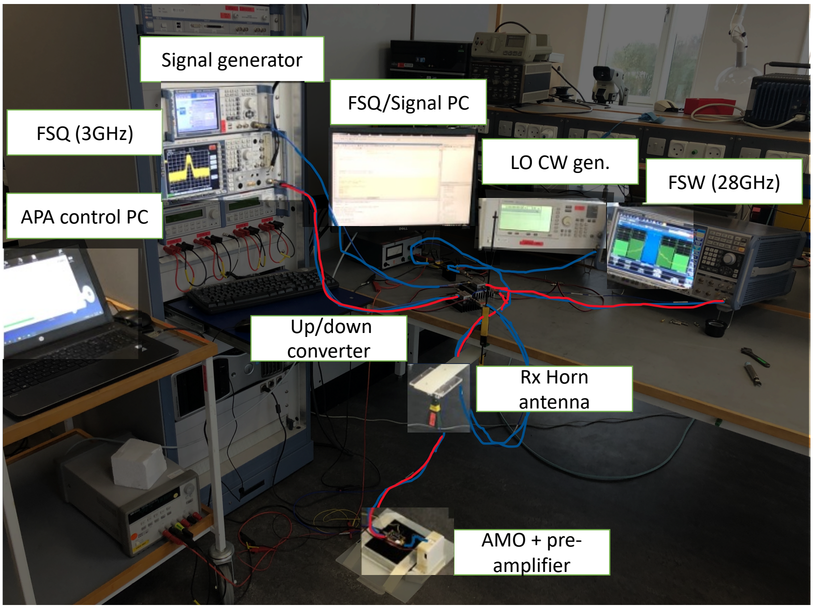


Fig. 13: Laboratory measurement setup [29].

4.2 Experiments and Data Collection

In our experiments, we focused on 16 different transistors, specifically the CREE GaN HEMT CGH40006P transistor. A 10 MHz LTE signal at 3.5 GHz, featuring a peak-to-average power ratio of 9.81 dB, was used as the input signal. All power amplifiers were biased with a drain-source voltage of 28 V and either a drain current of 100 mA or a gate voltage of -2.7 V before applying the signal [25, 29]. The measurements were conducted using the experimental setup depicted in Figure 12.

Fault Scenarios

Our research explored various fault scenarios, each corresponding to a different component within the active phased array (APA) system, including the antenna element, power amplifier, and phase shifter. These fault scenarios were controlled using software provided by AMOTECH [54].

Antenna Element Malfunction

One category of fault scenarios involved the malfunction of individual antenna elements. Specifically, we examined the impact of turning off a single antenna element while the entire APA was transmitting. Due to the combinatorial explosion of possibilities, multi-element failures were not practically feasible to collect and classify individually. Therefore, we opted for multi-label fault detection, where the system could identify multiple fault categories simultaneously. The goal was to train our Deep Neural Network (DNN) to correctly identify the faulty antenna element and distinguish the corresponding IQ waveforms.

Magnitude Attenuation Issue

Another set of experiments aimed to detect magnitude imbalances caused by undesired attenuation. Signal attenuation was controlled using AMOTECH software with a resolution of 0.5 dB (the lowest resolution available). The test involved attenuating one signal path by 0.5 dB using the software and then measuring the system's response 10 times, as outlined in our measurement procedure.

Phase Fault Detection

In a third category of experiments, we focused on the DNN's capabilities for detecting phase faults. Due to limitations in the AMOTECH 0404 phase controller, we introduced the minimum phase error possible—1-bit phase difference, equivalent to 5 degrees of phase shift. Each of the 16 antenna elements was shifted 5 degrees out of phase one by one to simulate the absolute minimum phase fault that could occur. This revised section maintains all essential information while eliminating redundancy, resulting in a more concise and organized presentation of your research methodology and data collection procedures.

5 Contribution Summary

This section presents the contributions of the papers included in Part II of this dissertation, offering a brief summary of motivations, contents, and main results.

5.1 Paper A

Two-stage Deep Learning Receiver for Power-Efficient LEO Satellite Communications

Motivation

The paper focuses on addressing the challenge of high throughput and power efficiency in Low Earth Orbit (LEO) satellite communication systems. With the rapid advancement of 5G and 6G networks, integrating these technologies into satellite communication is essential. One major challenge in this integration is achieving power-efficient transmissions, especially when dealing with nonlinearities introduced by power amplifiers in the transmitter.

Paper content

The authors propose a two-stage deep learning receiver architecture for OFDM (Orthogonal Frequency Division Multiplexing) systems to address this issue. The architecture uses a Deep Complex Convolution Neural Network (DCCN) trained in two stages. The first stage involves training with an AWGN (Additive White Gaussian Noise) channel and a fixed nonlinear front-end, while the second stage utilizes transfer learning to adapt to flat fading channels. This training approach enables the receiver to efficiently handle varied operation statuses, such as different power levels and steering angles, without compromising the bit error rate. The paper elaborates on the DCCN architecture, training process, and the use of an end-to-end simulator for both training and testing.

Main results

The DCCN receiver's performance was rigorously evaluated using several metrics and scenarios. When tested over an Average White Gaussian Noise (AWGN) channel, the DCCN demonstrated the ability to handle AWGN and non-linear distortion simultaneously for different modulation forms. However, it was noted that if the APA model is not known a priori, the DCCN struggles to handle non-linear distortion effectively.

In tests involving a flat fading channel, the DCCN was compared to legacy receivers like LMMSE and ALMMSE. The DCCN was found capable of compensating for nonlinearity and performed comparably to these legacy systems in flat fading conditions. It was observed that both LMMSE and DCCN could not effectively compensate for non-linearity caused by the AiP without prior awareness of it. Interestingly, the DCCN performed slightly better in conditions of low Signal-to-Noise Ratio (SNR), where high noise levels blocked the non-linearity from impacting the DCCN's performance. When trained with the AiP model, the DCCN's performance improved significantly, achieving results only marginally worse than conventional approaches like ALMMSE and on par with LMMSE.

Further, the DCCN's ability to adapt to different power levels and steering angles was assessed. It showed that the DCCN could compensate for variations in power levels without model alteration. In scenarios where the non-linear frontend model changed to a different AiP, the DCCN required retraining for the new front end, but only in the second step of the training process. This feature greatly reduces the cost and time of reapplying the model to different APA models and channel environments.

These findings demonstrate the DCCN receiver's robustness and adaptability in various operational conditions, making it a promising solution for high throughput and power-efficient communication in LEO satellite systems.

5.2 Paper B

Adaptive Neural Receivers for Nonlinear Transmitters and Multiple Channels

Motivation

This paper addresses the challenge of adapting neural receivers to nonlinear transmitters in Orthogonal Frequency Division Multiplexing (OFDM) systems. Traditional approaches often struggle with nonlinearity induced by power amplifiers, resulting in signal distortion and degradation in wireless communication systems. The motivation for this study is to develop a robust neural receiver using a deep complex convolutional network (DCCN) that can effectively manage multiple channels and nonlinear transmitters without the need for frequent retraining.

Paper content

We propose a novel hybrid model transfer learning approach for training OFDM neural receivers. This method incorporates a two-step training process using a mixed model of different channels and front-end sampled models. The approach aims to expedite the development of a more robust neural receiver capable of managing variations in nonlinear operation regions and biasing points. The paper details the methodology, including the architecture of the DCCN neural network, training parameters, and performance evaluation metrics. The training approach is designed to handle obscured or noisy front-end transmissions, and the network architecture is optimized to adapt to different fading channels and non-ideal nonlinear samples.

Main results

The proposed hybrid model training approach showed significant improvements in bit error rate (BER) performance and training efficiency compared to traditional methods.

The DCCN neural receiver achieved a substantial improvement in BER. Specifically, there was a 35% enhancement in BER performance compared to traditional methods. This improvement was consistent across various modulation schemes and channel conditions, including BPSK, QPSK, 8QAM, and 16QAM, in EPA, EVA, and ETU mobile channels. The training time was reduced by 19%, which highlights the method's efficiency. The quicker training process was attributed to the use of a distribution of randomly drawn Rayleigh fading models, making the training process more agile and suitable for adapting to random unknown channels.

The hybrid-trained DCCN showed adaptability to different nonlinearities induced by various operational points, bias voltages, and input power level variances. This adaptability was evident in its ability to maintain BER performance even in the presence of nonlinear front-end distortions and in highly mobile channel conditions.The DCCN receiver could adapt to most expected scenarios, including imperfect AiP models and noisy power amplifier models. It was capable of accurately demodulating the received signal and compensating for differences in noisy models, maintaining good performance despite such challenges.

This research indicates that the hybrid model training approach can significantly enhance the performance and adaptability of neural receivers in complex wireless communication scenarios, particularly in the context of 6G technologies.

5.3 Paper C

Efficient Fault Diagnosis in mm-Wave Active Phased Arrays

Motivation

This paper addresses the critical need for efficient and timely fault diagnosis in millimeter-wave (mmWave) active phased arrays (APAs), which are key components in 5G and 6G communication systems. Traditional fault diagnosis methods are costly, time-consuming, and complex, making them impractical for on-site deployment. The paper aims to develop a method that can quickly and accurately diagnose faults in APAs using a simpler and more cost-effective approach.

Paper content

The authors propose a novel fault diagnosis method using a Deep Neural Network (DNN) that exploits the baseband in-phase and quadrature signals for classifying different faults in APAs. This approach requires only a single probe and one measurement point, making it fast and efficient. The DNN is designed to detect faults in antenna elements, power amplifiers, and phase shifters, addressing a broader range of potential issues than traditional methods. The paper details the architecture of the DNN, the fault scenarios considered, the data collection and measurement setup, and the training process of the DNN.

Main results

The proposed method demonstrated high accuracy in diagnosing faults in a 28 GHz Active Phased Array (APA). It achieved an accuracy of 99% for single-element failure detection and 80% for multi-element failure detection. This high level of accuracy was consistent across various test scenarios, including on-off antenna elements, phase variations, and magnitude attenuation variations.

The method is effective even in low signal-to-noise ratio (SNR) environments, maintaining above 90% fault detection accuracy at a 4 dB SNR. This capability is crucial for ensuring reliable fault diagnosis in challenging communication environments.

The DNN was able to correctly identify the class of the fault with an accuracy ranging between 88% and 99%, depending on the specific class. Considering the DNN had to distinguish among 49 different fault classes, these results are highly significant. The various fault classes included no faults, antenna off, attenuation with 0.5 dB, and a phase shift of 5 degrees, each distinguishable by the DNN.

The DNN was capable of predicting every failure in the test dataset in 1.1 seconds, and it took only 0.006 seconds to predict a single measurement. This efficiency in fault detection makes the method highly suitable for on-site deployment, especially since the required I and Q signals are easily acquired using a simple receiver front end. This approach significantly simplifies the measurement setup and data acquisition process compared to conventional methods.

These results collectively indicate that the DNN-based approach for fault diagnosis in mmWave APAs is not only highly accurate and efficient but also robust against various environmental factors, making it a promising solution for real-world applications in 5G and 6G communication systems.

5.4 Paper D

Blockage Prediction in Directional mmWave Links Using LTC Network

Motivation

This paper addresses the challenge of blockage prediction in millimeter-wave (mmWave) and terahertz (THz) communication, crucial for achieving high data rates in 6G and beyond. The highly directional transmission in these bands is vulnerable to blockage, causing sudden communication interruptions. Existing methods for blockage prediction use machine learning approaches that require data from multiple sources or prior statistical observations, making them costly and less generalizable.

Paper content

We propose a novel approach using a Liquid Time Constant (LTC) network, inspired by biological systems, for predicting future blockage status of mmWave links using only received signal power. The LTC network, based on ordinary differential equations (ODEs), shows strong expressiveness, stability, and performance in handling time series. It is capable of making predictions without prior knowledge of the outdoor scenario or retraining, making it highly generalizable. The method involves training the LTC network with an indoor dataset containing pre-blockage power signature under controlled blockage movement and then applying the trained model to outdoor datasets with uncontrolled blockage events.

Main results

The LTC network demonstrated an impressive accuracy of over 97.85% in predicting blockage for the immediate future timeslot (t+1) across all outdoor scenarios. This high level of accuracy highlights the model's effectiveness in real-time prediction and its potential to enhance the reliability of mmWave communication systems.

Compared to the baseline models, the LTC network showed a substantial improvement in accuracy for predicting blockage in the near future (t + 5 and t + 10 timeslots). The accuracy improvement ranged from 12% to 39%, indicating the model's superior predictive capabilities over a longer horizon, which is crucial for proactive network management and planning.

The proposed LTC model's sparse network structure and its high generalization capabilities make it an attractive option for pre-blockage prediction in high-frequency communication systems. This generalization ability is particularly important as it implies the model can be effectively applied in various outdoor scenarios without the need for extensive retraining or tuning for each specific environment.

These results demonstrate the LTC network's potential to significantly enhance the prediction and management of blockage events in mmWave communication systems, thereby contributing to the overall reliability and efficiency of 6G and beyond wireless networks.

6 Conclusion & Future work

In this thesis, the focus has been on exploring emerging aspects of mmWave 5G and 6G communications systems, specifically addressing four key research areas that are crucial for the advancement of this field. The work has resulted in developments that contribute to the understanding of power-efficient transmission, channel and transmitter management, fault diagnosis using deep neural networks, and blockage prediction with deep learning techniques.

It is acknowledged that the work presented here has certain limitations. The two-stage deep learning receiver, designed for LEO satellite systems, faces challenges around power consumption and processing capabilities, which are critical for real-world applications. The hybrid model for channel and transmitter adaptation, while effective under current test conditions, may need further refinement to address scalability and adaptability in evolving communication scenarios. The neural network-based fault diagnosis system, although showing potential, requires additional testing and validation under diverse environmental conditions. Lastly, the blockage prediction methods using deep learning, are still in their nascent stage and need further development to handle the dynamic nature of mmWave communication environments effectively.

6.1 Future Work

Future research directions include enhancing the receiver architecture to better balance power efficiency with performance, particularly focusing on its application in edge devices where power resources are limited.

Further development of the hybrid model is necessary, especially in integrating more advanced machine learning techniques that can more effectively address channel fading and equipment degradation, while also aiming to reduce computational demands.

Improving the fault diagnosis systems in active phased arrays is another area for future work. Efforts should be directed towards creating more robust machine learning models that can perform reliably in various operational conditions typical of mmWave communication systems.

In blockage prediction, incorporating the prediction with communication systems will be an important step towards enhancing their impact on real-time adaptability and reliability in mmWave communications.

Overall, this thesis contributes to the body of knowledge in next-generation communication systems and suggests pathways for further research. The integration of machine learning with communication systems presents an opportunity for significant advancements in network efficiency and reliability. The findings of this thesis have practical implications for the telecommunications sector, suggesting applications that extend beyond the academic domain.

Moving forward, a multidisciplinary approach that blends the theoretical aspects of communication theory with practical machine learning insights is recommended. Such a direction is expected to facilitate the development of intelligent communication systems that are efficient, robust, and adaptable, meeting the demands of an increasingly connected and technologically evolving world.

References

[1] S. Tripathi, N. V. Sabu, A. K. Gupta, and H. S. Dhillon, "Millimeter-wave and terahertz spectrum for 6g wireless," *CoRR*, vol. abs/2102.10267, 2021. [Online]. Available: https://arxiv.org/abs/2102.10267

[2] T. Kaneko, K. Shiikuma, and K. Kunihiro, "Gan hemt high efficiency power amplifiers for 4g/5g mobile communication base stations," pp. 994–997, 2014.

[3] P. Varahram, S. Jamuar, S. Mohammady, M. Hamidon, and S. Khatun, "Power amplifiers linearization based on digital predistortion with memory effects used in cdma applications," pp. 488–491, 2007.

[4] Y. Komatsuzaki, K. Nakatani, S. Shinjo, S. Miwa, R. Ma, and K. Yamanaka, "3.0–3.6 ghz wideband, over 46amplifier with frequency dependency compensating circuits," pp. 22–24, 2017.

[5] O'Shea, T. J. and Hoydis, J., "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[6] K. Rahil and Frisco, TX, US, "System and method for satellite-long term evolution (S-LTE) air interface," U.S Patent 20 100 068 993, March, 2010.

[7] L. Siyang, Q. Fei, G. Zhen, Z. Yuan, and H. Yizhou, "LTE-satellite: Chinese proposal for satellite component of IMT-advanced system," *China Communications*, vol. 10, no. 10, pp. 47–64, Oct 2013.

[8] 3GPP, "Technical specification group radio access network; nr; physical channels and modulation (release 16)," ETSI, Tech. Rep., 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/16.02.00_60/ts_138211v160200p.pdf

[9] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[10] A. Molina, K. Rajamani, and K. Azadet, "Digital Predistortion Using Lookup Tables With Linear Interpolation and Extrapolation: Direct Least Squares Coefficient Adaptation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980–987, March 2017.

[11] D. Cox, "Linear amplification with nonlinear components," *IEEE Transaction on Communication*, vol. 22, no. 12, pp. 1942–1945, 1974.

[12] Y. Guo, C. Yu, and A. Zhu, "Power adaptive digital predistortion for wideband rf power amplifiers with dynamic power transmission," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 11, pp. 3595–3607, 2015.

[13] Danish space sector. [Online]. Available: https://www.openaccessgovernment.org/danish-space-sector-strategy-future/33249/

[14] D. M. of Finance. (2018) Danish ai strategy. [Online]. Available: https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf

[15] T. D. Government. (2018) Strategy for denmark's digital growth. [Online]. Available: https://eng.em.dk/media/10500/digital-growth-strategy-uk-web-2.pdf

[16] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, “Ten challenges in advancing machine learning technologies toward 6g,” *IEEE Wireless Communications*, vol. 27, no. 3, pp. 96–103, 2020.

[17] A. Jagannath, J. Jagannath, and T. Melodia, “Redefining wireless communication for 6g: Signal processing meets deep learning with deep unfolding,” *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, pp. 528–536, 2021.

[18] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5g,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[19] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang, and D. Zhang, “A Survey on Green 6G Network: Architecture and Technologies,” *IEEE Access*, vol. 7, pp. 175 758–175 768, 2019.

[20] J. Hoydis, F. Ait Aoudia, A. Valcarce, and H. Viswanathan, “Toward a 6g ai-native air interface,” *IEEE Communications Magazine*, vol. 59, no. 5, pp. 76–81, 2021.

[21] N. Alliance, “6g requirements and design considerations,” https://www.ngmn.org/publications/6g-requirements-and-design-considerations.html.

[22] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[23] Y. Goldberg, “A primer on neural network models for natural language processing,” *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, 2016.

[24] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[25] M. H. Nielsen, E. de Carvalho, and M. Shen, “A two-stage deep learning receiver for high throughput power efficient leo satellite system with varied operation status,” *IEEE Access*, 2022.

[26] W. Zhou and W. H. Lam, “A fast lmmse channel estimation method for ofdm systems,” *EURASIP J. Wireless Commun. Netw.*, vol. 2009, no. 1, pp. 1–9, Dec. 2009.

[27] “Etsi ts 138 211,” 3GPP, Technical Specefication 38.211, 2020, release 16. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/16.02.00_60/ts_138211v160200p.pdf

[28] “Etsi ts 138 214,” 3GPP, Technical Standard 38.214, 2020, release 16. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138214/16.02.00_60/ts_138214v160200p.pdf

[29] M. et al., “Adapting to nonlinear transmitters with hybridmodel training for neural receivers,” *Journal name*, vol. Volume number, p. Page numbers, Year.

[30] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, “Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale mosfets,” *IEEE Transactions on Electron Devices*, vol. 50, no. 9, pp. 1873–1852, 2003.

[31] N. Seoane, A. Garcia-Loureiro, K. Kalna, and A. Asenov, “Analysis of the impact of intrinsic parameter fluctuations in a 50 nm inp hemt,” in *2007 Spanish Conference on Electron Devices*, 2007, pp. 92–95.

[32] Huawei 6G Research Team, "6g: The next horizon," https://www.huawei.com/en/huaweitech/future-technologies/6g-the-next-horizon, Nov. 2021.

[33] M. Sweeting, "Modern small satellites-changing the economics of space," *Proceedings of the IEEE*, vol. 106, no. 3, pp. 343 – 361, 2018.

[34] Y. Zhang, Z. Wang, Y. Huang, W. Wei, G. F. Pedersen, and M. Shen, "A digital signal recovery technique using dnns for leo satellite communication systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6141–6151, 2021.

[35] J. Wang, L. Li, and M. Zhou, "Topological dynamics characterization for LEO satellite networks," *Computer Networks*, vol. 51, no. 1, pp. 43–53, 2007.

[36] Q. Chen, Y. Zhang, F. Jalili, Z. Wang, Y. Huang, Y. Wang, Y. Liu, G. F. Pedersen, and M. Shen, "Robust Digital Signal Recovery for LEO Satellite Communications Subject to High SNR Variation and Transmitter Memory Effects," *IEEE Access*, vol. 9, pp. 135 803–135 815, 2021.

[37] V. Karunanithi, C. Verhoeven, and W. Lubbers, "High efficiency transmitter architecture compatible with CCSDS and ECSS standards for nano-satellite missions," in *32nd AIAA International Communications Satellite Systems Conference*, 2014, p. 4252.

[38] Y. Zhang, Z. Wang, Y. Huang, W. Wei, G. F. Pedersen, and M. Shen, "A Digital Signal Recovery Technique Using DNNs for LEO Satellite Communication Systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6141–6151, 2021.

[39] I. Akyildiz *et al.*, "Terahertz band communication: An old problem revisited and research directions for the next decade," 2021.

[40] O'Shea, T. J. and Karra, K. and Clancy, T. C., "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *Proc. IEEE Int. Symp. on Signal Processing and Information Technology (ISSPIT)*, 2016, pp. 223–228.

[41] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "Ofdm-autoencoder for end-to-end learning of communications systems," in *Proc. IEEE Int. Workshop Signal Proc. Adv. Wireless Commun. (SPAWC)*, 2018.

[42] M. Kim, W. Lee, and D. Cho, "A novel papr reduction scheme for ofdm system based on deep learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, 2018.

[43] T. Wada, T. Toma, M. Dawodi, and J. Baktash, "A denoising autoencoder based wireless channel transfer function estimator for ofdm communication system," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 530–533.

[44] B. Zhu, J. Wang, L. He, and J. Song, "Joint transceiver optimization for wireless communication phy using neural network," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1364–1373, 2019.

[45] Z. Zhao, M. C. Vuran, F. Guo, and S. D. Scott, "Deep-waveform: A learned ofdm receiver based on deep complex-valued convolutional networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2407–2420, 2021.

[46] F. Ait Aoudia and J. Hoydis, "End-to-end learning for ofdm: From neural receivers to pilotless communication," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1049–1063, 2022.

[47] T. J. O'Shea, T. Roy, N. West, and B. C. Hilburn, "Physical layer communications system design over-the-air using adversarial networks," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 529–532.

[48] V. Raj and S. Kalyani, "Backpropagating through the air: Deep learning at physical layer without channel models," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2278–2281, 2018.

[49] Y. Xie, X. Liu, K. C. Teh, and Y. L. Guan, "Robust deep learning-based end-to-end receiver for ofdm system with nonlinear distortion," *IEEE Communications Letters*, vol. 26, no. 2, pp. 340–344, 2022.

[50] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmwave beam and blockage prediction using sub-6 ghz channels," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5504–5518, 2020.

[51] A. Alkhateeb and I. Beltagy, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," 2018.

[52] A. E. Kalør, O. Simeone, and P. Popovski, "Prediction of mmwave/thz link blockages through meta-learning and recurrent neural networks," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2815–2819, 2021.

[53] S. Wu *et al.*, "Blockage prediction using wireless signatures: Deep learning enables real-world demonstration," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 776–796, 2022.

[54] AMOTECH, "Datasheet on amo aip aaipk428gc-a0404-xxx," https://www.amo-tech.com/datasheet/AAiP28G-A0404-EVB-R05_EVB-R05_190212_b.pdf, Feb. 2019.

Part II

Papers

Paper A

A Two-stage Deep Learning Receiver for High Throughput Power Efficient LEO Satellite System with Varied Operation Status

Nielsen, Martin. Hedegaard., De Carvalho, Elisabeth. & Shen, Ming.,

The paper has been published in the
IEEE Access Vol. 10, pp. 60904-60913, 2022,

Abstract

Low Earth orbit satellites are expected to be one of the biggest suppliers of wireless communication within the coming years. For this to happen 5G and 6G networks, are crucial to be implemented in satellite communication. This comes with the problem of power-efficient transmissions. This paper exploits recent advances in complex-valued deep learning to cope with this challenge. The proposed approach is based on the autoencoder structure, where a legacy orthogonal frequency division modulation (OFDM) transmitter is used as an encoder and a deep complex convoluted network (DCCN) is used as decoder/receiver. Different from other state-of-the-art receiver architectures based on one-stage trained neural networks, our proposed DCCN adopts a two-stage training scheme, where the first stage is trained using AWGN channel and a fixed non-linear front end. The second stage uses a transfer learning to adapt to the flat fading channels and the front end model can be changed to compensate for different front ends, significantly reducing training time.This allows for power-efficient transmission at different operation statuses (e.g. radiated power levels and steering angles) without compromising the bit error rate in both average white Gaussian noise (AWGN) and flat fading channels. A K-band (28 GHz) active phased array in package (AiP) transmitting a 5G NR OFDM signal with a bandwidth of 100 MHz was used as the main front end test vehicle for validating the proposed DCCN. Satisfactory bit error rates were achieved while the AiP was driven into saturation with high power efficiency at different power levels and steering angles. This work demonstrates, for the first time, the promising capability of deep neural networks in processing varied operation staged non-linear OFDM waveforms in the form of an auto-decoder receiver.

1 Introduction

Satellite communication has been a topic of interest for many years, but recently it has been developing rapidly with the advancements in low Earth orbit (LEO) satellites that are usually deployed at an altitude of 500–1500 km to ensure low latency in established communication links [1]. Dependent on the altitude deployed the satellite will over the course of a transmission window different distances to the ground making the communication channel dynamic and varying free space path loss. This is conventionally overcome at the transmitter side [2].

Currently, the satellite sector is changing its service focus from TV and maritime applications towards broadband Internet services using the LEO satellites and 5G signals [3–5]. With higher data rates needed for Internet access, highly bandwidth-efficient modulation schemes, such as multi-sub-carrier orthogonal frequency division modulation (OFDM), are wanted. Multi-subcarrier OFDM signals can provide the necessary bandwidths for future space communication and are already being used in some satellite systems [6–8]. However, this comes with the cost of a high peak-to-average power ratio (PAPR) and low efficiency in the transmitter front end, since to ensure linearity back off is typically used.

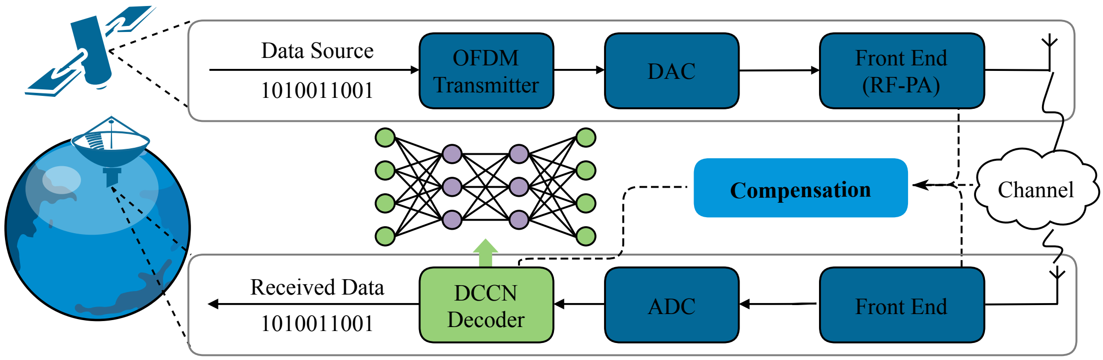


Fig. A.1: Concept figure of the proposed system architecture where a DCCN receiver is used to compensate for the front end.

For many years the industry and research practice for increasing power efficiency is use pre-distortion to deal with non-linear amplifier responses. Pre-distortion can be implemented in a plethora of ways all from look up tables using memory polynomial models to different deep learning techniques [6, 8–23]. Memory polynomial unlike deep learning uses an equation to describe non-linear behavior. It is based on a feedback loop between the input and output of the power amplifier (PA) to achieve pre-distortion. It uses the inverse polynomial to change the input signal such that the wanted efficiency/linearity of the system is achieved [9–11].

Implementing pre-distortion in satellites is not as easy, since the optimization can be costly for the satellites [12]. When a highly integrated active phased array is used, it is very difficult to establish the feedback loop. Further, the PAPR change in the input signal caused by pre-distortion can in some cases not be realized [12]. Commonly the deep learning approach is based on autoencoders, where both the transmitter and receiver are changed to neural networks. Some research reduces the PAPR [14, 20], while some gaining better bit error rate (BER) performance [13, 15, 16]. However to our knowledge no autoencoders have gained both high BER performance and are able to combat high non-linearity without changing the transmitter completely [13–20]. However, using deep learning comes with challenges [4, 5, 24]. With the wanted integration of satellites into the 5G network, multiple satellites can and will have to connect to the same ground station in these new mega-constellations [4, 24]. This can cause issues having to deal with multiple slightly different front ends. Another inconvenience is that the satellite has limited power on board and using pre-distortion or an autoencoder can be a become resource problem.

Instead of relying on changing the on-board communication system, this paper proposes to train a receiver that can compensate for the different non-linear behavior present in the received signal. Then only the power output from the transmitter can be increased and the satellite can achieve the performance increases that conventional pre-distortion gives a front-end transmitter by only changing the receiver side. This allows for multiple satellites to be serviced by the same ground station without having to change the satellite architecture. The idea was initially proposed in [21]. It was shown possible to recover the received signal and compensate for the non-linear distortion before the actual receiver and BER performance was later shown in [22]. However, it lacked more realistic channels and results for how it compensates for different amplifiers, steering angles, and power levels without retraining the model. A fading channel, unlike AWGN, distort and corrupts the signal. Hence, using pre-processing for OFDM signals is both time-consuming and inconvenient. Instead using complex-valued neural networks as receivers has the potential to solve these issues faster and as reliable as legacy receivers [23]. Complex valued networks are similar to deep neural networks but they can process and understand complex numbers without preprocessing. This paper proposes to combine the use of a complex-valued neural network [23] with the idea of recovering signals [21, 22] to provide a new receiver structure that can both decode and compensate for the non-linearity in a single complex-valued neural network for an OFDM system. This paper proposes a two-stage Deep Complex Convolution Neural Network (DCCN) which is trained in two steps to achieve the wanted performance. The special two-stage DCCN comes with the benefit that it is very robust to variations in the non-linear signal. Further, it can quickly adapt to different front ends with only one retraining of step two. This makes quick adaptation to different satellites very fast at the ground station. Different operation modes for the same front end are possibly compensated for. This is possible without retraining for operation modes such as different power saturation, different bandwidths, and different steering angles, these are all compensated for due to the two-stage DCCN.

To train the neural network this paper uses an online end-to-end numeric simulator that uses a legacy OFDM transmitter, together with real experimental data gathered from the front end to train and test the DCCN.

The paper is structured as follows. In Sec. II the problem and communication system are described, Sec. II-A the legacy OFDM transmitter is described, Sec. II-B the end-to-end simulator is detailed and Sec. II-C describes the channel models used. Sec. III describes the acquisition and results of the measured front-end models together with the results of the end-to-end simulator. In Sec. IV the deep learning receiver architecture is elaborated and the training process is described. The results of the deep learning receiver are the presentation in Sec. VI and discussed in Sec. VII. The paper is concluded in Sec. VIII.

2 Communication system & Problem formulation

The satellite communication system starts when the satellite establishes a communication link with the ground station [2, 21, 22, 25–28]. The window for an LEO satellite starts when the elevation angle ϵ is greater than 5°. The satellite will then travel with a constant velocity towards the ground station where the elevation angle is 90°. The satellite will then travel from the ground station until the elevation angle is 5°. The distance to the ground station will not be constant and can be geometrically described as shown in Fig. A.2.

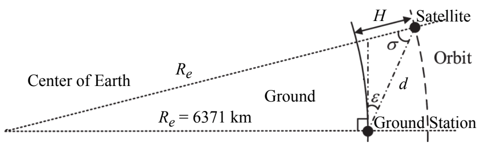


Fig. A.2: Geometrical illustrating of the satellite to ground data transmission link.

The differentiation in the distance from the ground station causes the transmission distance to vary. The link will therefore change over time. This can be described mathematically.

$$(H + R_e)^2 = R_e^2 + d^2 - 2R_e d \cos(\frac{\pi}{2} + \epsilon_0), \tag{A.1}$$

where ϵ represents the elevation angle, d is the distance of data transmission, H denotes the orbital altitude of LEO satellite, R_e is the earths radius. This makes it possible to denote the distance as a function of ϵ

$$d(\epsilon) = R_e[\sqrt{(\frac{H + R_e}{R_e})^2 - \cos^2(\epsilon) - \sin(\epsilon)}]. \tag{A.2}$$

Then the free space path loss can be expressed as

$$L(\epsilon_0) = (\frac{4\pi d(\epsilon)}{\lambda})^2 = (\frac{4\pi f d(\epsilon)}{c})^2, \tag{A.3}$$

since the noise can is fixed the variation in SNR regardless of distance is

$$\Delta\frac{S}{N}(dB/Hz) = (\frac{S}{N}_{5^\circ}) - (\frac{S}{N}_{90^\circ}) = \Delta L(\epsilon). \tag{A.4}$$

Based on the equation the variation for a satellite in LEO at an altitude of 800 km is approximately 12 dB in the link budget between 5° and 90° [22]. To have the possibility of an efficient as possible satellite downlink, the transmitter should be able to vary transmission power such that the power received is constant. However to keep the transmitter efficient the transmitter should be driven non-linearly.

2.1 OFDM transmitter

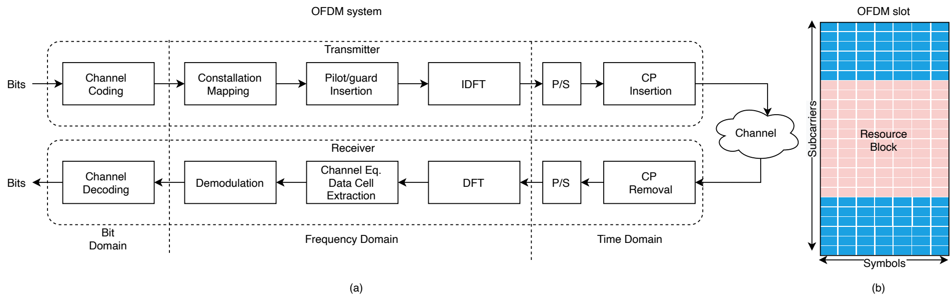


Fig. A.3: a) Block diagram of a PHY layer OFDM communication system. b) OFDM slot illustration, subcarriers on y axis and symbols on x axis. A resource block is shown to be 12 subcarriers and 7 symbols.

5G NR is an evolution of LTE signal that is proposed to become the future communication signal for satellites [6, 7]. OFDM is the basis for 5G and LTE signals therefore, this paper only considers the OFDM signal since any channel coding/precoding associated with LTE or 5G NR happens before OFDM transformation, see Fig. A.3 a. OFDM is based on a frame structure where a communication slot is composed of multiple OFDM symbols. The notations related to the OFDM slot are as follows: an OFDM symbol contains N subcarriers, where N is the size of DFT/IDFT. Among the N subcarriers, a total of G nullified guard subcarriers are placed at the center (DC guard band) and the edge guard band. A subcarrier in an OFDM symbol is referred to as a resource element (RE). A collection of REs is referred to as a Resource block as shown in Fig. A.3. A communication slot contains F consecutive OFDM symbols, in which P and D REs are allocated to pilot and data, respectively. The length of a time-domain full OFDM symbol is $S = N + N_{cp}$ where N_{cp} is the length of CP. Under m-ary modulation, an IQ sample carries m bits, and the size of the constellation is 2^m.

2.2 End-to-end simulator

The end-to-end simulator is built based on Fig. A.3. It uses a random bit generator to produce the bit sequence $\mathbf{b}$ that is encoded such that $\mathbf{b} \in \pm 1$. The encoded bits are then converted to complex-valued in-phase and quadrature (IQ) data by mapping to a constellation on the IQ plane. From the IQ data, an OFDM frequency-domain symbol $\mathbf{X}$ is is created by inserting pilot signals and guard bands into the IQ data, and then $\mathbf{X}$ is transformed to a time-domain OFDM symbol, $\mathbf{x}$, via an N-point IDFT and a subsequent parallel to serial (P/S) conversion.

Next, the Cyclic Prefix (CP) is prepended to $\mathbf{x}$ to create a time-domain full OFDM symbol, $\mathbf{x}_{cp}$. The baseband signal $\mathbf{x}_{cp}$ is then upconverted to RF and sent to the

RF front end model. The RF front end model is created based on the popular Memory Polynomial model that can describe the non-linear distortion given the input and output of a non-linear system. It is given as

$$y(t) = \sum_{n=1}^{N} \sum_{m=0}^{M} a_{nm} x(t - \tau_m) |x(t - \tau_m)|^{n-1}, \tag{A.5}$$

where N is the polynomial order while M is the memory effect, i.e, the number of previous samples that have effect on the current output. To determine the amplifier coefficients a_{nm} the input and output of the device need to be captured, and reverse modeling is used to find them. The coefficients are dependent on the signal power levels and how high the PAPR of the signal was at the input and output baseband signals.

The now distorted signal is propagated over a channel model and the received IQ samples are then represented as OFDM-time domain samples $\mathbf{y}_{cp}$. Then, the CP is removed from $\mathbf{y}_{cp}$ and the rest of the IQ samples, $\mathbf{y}$, are transformed to the frequency-domain OFDM symbol, $\mathbf{Y}$, via DFT. Based on $\mathbf{Y}$, a channel equalizer outputs the estimated transmit frequency-domain IQ data $\hat{\mathbf{X}}$ which is then demodulated to soft bits (log-likelihood) $\tilde{\mathbf{b}}$ which are converted to binary outputs bits via hard decision to become $\hat{\mathbf{b}}$. The end-to-end simulator can then calculate the Bit Error Rate depending on how much difference there is between $\hat{\mathbf{b}}$ and the ground truth $\mathbf{b}$.

2.3 LEO channel

According to [29, 30], the LEO communication channel cannot be treated as an AWGN channel. It is stated that it can be seen as a fading environment. We therefore model our channel after the well described flat fading model [31] since it is a good baseline model:

$$\mathbf{y} = \mathbf{x} * \mathbf{h} + \mathbf{n}, \mathbf{Y} = \mathbf{X} \odot \mathbf{H} + \mathbf{N}, \tag{A.6}$$

where $\mathbf{x}, \mathbf{y}, \mathbf{n}$ are the time-domain transmitted and received signals and white noise, respectively and $\mathbf{n}$ is the channel impulse response. $\mathbf{Y}, \mathbf{X}, \mathbf{H}, \mathbf{N}$ are the frequency domain transforms. $*$ is the convolution and $\odot$ is the elementwise product.

3 Measurement of Non-linear Hardware

To have a practical RF front end model in the end-to-end simulator, an active phased array in package (AiP) is used for measuring the non-linear behavior of the front end. The AMOTECH A0404 AiP uses 4 Anokiwave AWMF-0158 beam forming devices and a 4 by 4 patch antenna array for transmitting [32]. In this paper, the changes happening in the transmission window i.e. different input power is given and a variation in the steering angle are measured and investigated.

To capture the data for use in the hardware models the AiP is measured over the air using a 28 GHz 5G NR signal with a bandwidth of 100 MHz, compliant with the 3GPP downlink specification for 5G NR OFDM modulated with a peak to average power ratio of 10.6 dB generated by the R&S SMBV100B signal generator. The 5G NR signal from the generator is up-converted from 3 GHz. A continuous-wave (CW) signal has been multiplied by two into 25 GHz for up-conversion and down-conversion as shown in Fig. C.6. A pre-amplifier is used to push the AiP into compression, the pre-amplifier is in backoff to ensure linearity. This setup is shown in Fig. C.6.

The data is captured by the observation horn antenna placed 42 wavelengths away (44 cm) and aligned with the main beam which is connected to and analyzed in the spectrum analyzer (R&S FSW 67 GHz). The procedure for taking measurements is as follows:

1. I and Q waveform for the 5G NR signal is uploaded using the R&S VISA tool from the PC to the vector signal generator.
2. The APA is driven into the non-linear region.
3. The I and Q at the receiver are then captured from the signal analyzer using the R&S VISA tool.
4. The signal is then time aligned with the input signals samples post-process to ensure each sample corresponds to the correct time sample of the previously recorded input signal.

3.1 Measured results

The measurement results are shown in Fig. A.5. Fig. A.5a the power spectrum density is shown. The 38 dB output power signal shows the AiPs non-linear behavior and the high adjacent channel power present. Fig. A.5b shows the amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) behavior of the AiP. The non-linear saturation and memory effects can be seen both in the spread of points and the steep drop after 30 dB for AM/AM and the rotation in phase for symbols in the AM/PM. Looking at the power added efficiently (PAE) curve, Fig. A.5c it is clear that the AiP is not very efficient. To have the AiP be as efficient as possible it is better to have an output power level of around 38 dBm at all times.

3.2 Simulation results on measured AiP model

To determine the non-linearity effect on the communication system the simulator is tested in three different scenarios, AiP distortion only, with AWGN and AiP distortion, and with fading and AiP distortion. The noisy baseband signal results from the three tests are shown in Fig. B.3. The resulting constellation patterns show that without

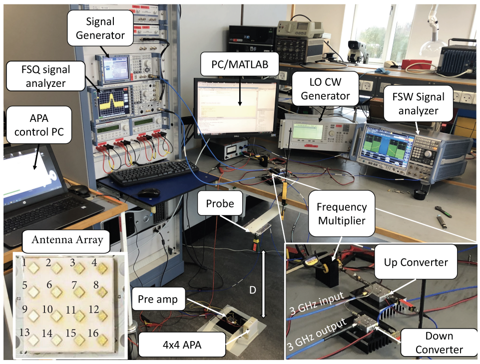


Fig. A.4: Measurement setup of the AiP with all instruments used labeled.

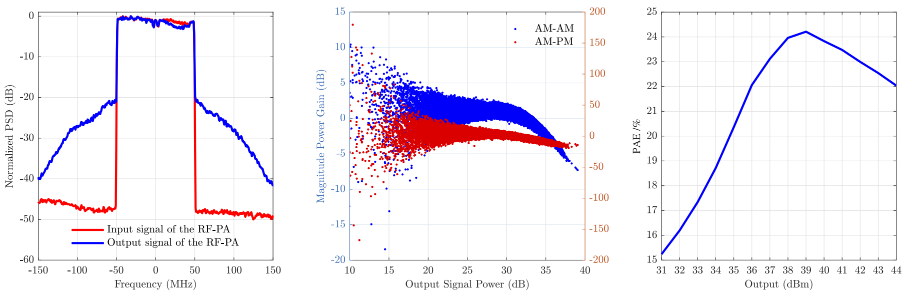


Fig. A.5: Measurement of the AiP. a) Measured power density response of the 100MHz 5G NR response. b) Measured AM/AM AM/PM responds. c) AiP power added efficiently vs input power.

proper modifications to understand the non-linearity and fading it is not possible to decode the symbols given.

4 DCCN architecture

The neural network-based OFDM receiver uses similar signal processing modules as legacy OFDM receivers with layers functioning differently. The proposed architecture is shown in Fig. A.7 which includes recommendations founds in [23].

The proposed architecture uses the input signal, $\mathbf{y}_{cp}$, and uses the first complex dense layer to remove the CP. The DFT like C-Conv layer converts the signal $\mathbf{y}$ to the frequency domain $\mathbf{Y}$. The channel estimator uses four dense layers to both compensate for the AiP and estimate the channel response, $\hat{\mathbf{H}}$. After the channel estimator, it then

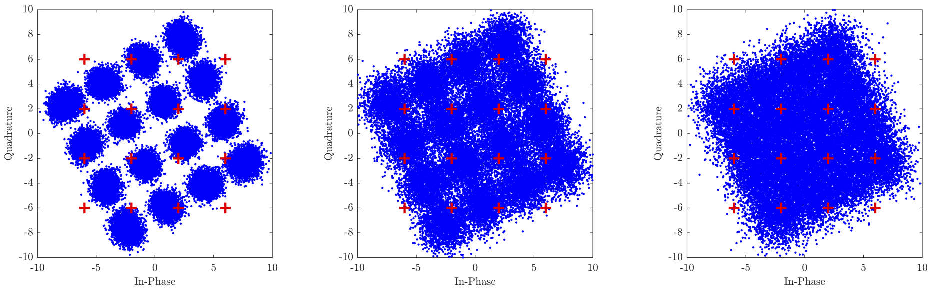


Fig. A.6: Effects of the non-linear distortion from the AiP and AWGN and fading on the constellation with SNR at 30 dB. a) Constellation of the non-linear distorted signal without the channel. b) Constellation of the non-linear distorted signal with the AWGN. c) Constellation of the non-linear distorted signal with flat fading.

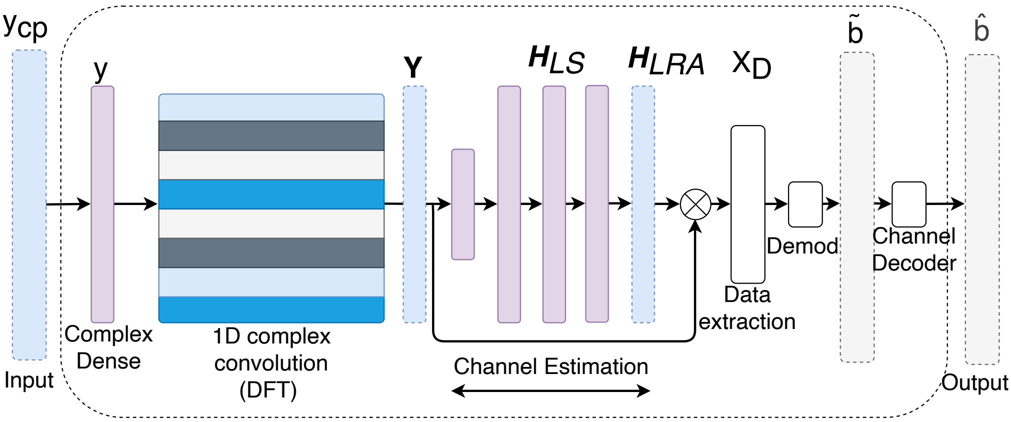


Fig. A.7: Neural network receiver architecture to deal with non-linear amplifier and different fading channels.

does element-wise complex division to the equalization so that it can estimate $\hat{\mathbf{X}}$. To do the estimation the first dense layer of $\mathbb{C}_{FN\ddot{O}P}$, where F is OFDM symbols per coherence slot, N is number subcarriers which are proportional to the DFT/ODFT size, and P number of pilots. The dense layer is designed to locate pilots and estimate channel coefficients on pilots $\hat{H}^{P}_{LS}$, where LS is least-square similar to LMSSE. The operation can be shown as equal to that of LMSSE which matematically can be described as,

$$\hat{\mathbf{X}} = \frac{\mathbf{Y}}{\hat{H}_{LS}}, \hat{H}_{LS} = \mathcal{L}(\frac{\mathbf{Y}_P}{\mathbf{X}_P}), \tag{A.7}$$

where $\hat{\mathbf{X}}$ is the estimated signal, $\hat{H}_{LS}$ contains the LS channel estimates, $\mathbf{X}_P$ and $\mathbf{Y}_P$ are transmitted and received pilots, and $\mathcal{L}()$ is the interpolation operation. To obtain $\hat{H}$ an interpolation of $\hat{H}^P_{LS}$ to the entire slot and channel estimation is done in the next three dense layers and a 2D filter of size (F, N). The 2D filter can be mathematically described as follows

$$\hat{\mathbf{H}}_{LRA} = \mathbf{U}\mathbf{D}_p\mathbf{U}^H\hat{H}_{LS}, \tag{A.8}$$

where $\mathbf{D}_p$ is a diagonal matrix with trainable parameters $\delta_k = \frac{\lambda_k}{\lambda_k + \frac{\beta}{\alpha}}$, $\mathbf{U}$ is a unitary matrix containing the singular vectors of the frequency domain covariance matrix of channel realizations denoted $\mathbf{R}_{HH} = E\{\mathbf{H}, \mathbf{H}^H\}$. Instead of setting the parameters explicitly they are learned directly from data. As the non-linearity is treated at the receiver side to be part of the channel a separate layer is not needed. However, through testing, it has been shown that having three dense instead of only two has better generalization capabilities than using two [23].

To gain the time domain signal the network uses a complex dense to take out the Resource Elements (RE) so we get $\hat{\mathbf{X}}_D$ and then using an IDFT like C-Conv layer it is possible to transform $\hat{\mathbf{X}}$ to time-domain $\hat{\mathbf{x}}$. The rest of the forward network converts IQ samples to soft bits, where an input IQ sample $\mathbb{C}$ is treated as a vector of 2 real numbers, $\mathbb{R}^2$. The extracted IQ vector and its non-linear (Leaky ReLU) activation, A0, are concatenated to a tensor of shape [B, D, 4], where B is the number of slots in a batch of input signals. A0 is fed to a small dense layer of $\mathbb{R}_{4\text{Ö}2^m}$ followed by another Leaky ReLU activation, A1, of which the output tensor is reshaped to [B, D,m, 2] and then activated by a softmax function along its last dimension to produce a soft bit–a vector of likelihoods of ± 1.

4.1 Training

The training of the DNN receiver as illustrated in Fig. B.6. The training is done online using the end-to-end simulator which generates a random binary stream $\mathbf{b}$. The stream is turned into time-domain OFDM symbols, $\mathbf{x}_{cp}$ by the OFDM transmitter. The transmitted signal $\mathbf{y}_{cp}$ is then distorted by a non-linear AiP model together with a channel model that adds noise and fading to $\mathbf{x}_{cp}$. Thus the received signal, $\mathbf{y}_{cp}$ is the training data, and $\mathbf{b}$ is the labels. The output of the receiver is given as soft bits, $\tilde{\mathbf{b}}$ and after the hard decision the $\hat{\mathbf{b}}$. The loss function is a weighted sum of the cross entropy

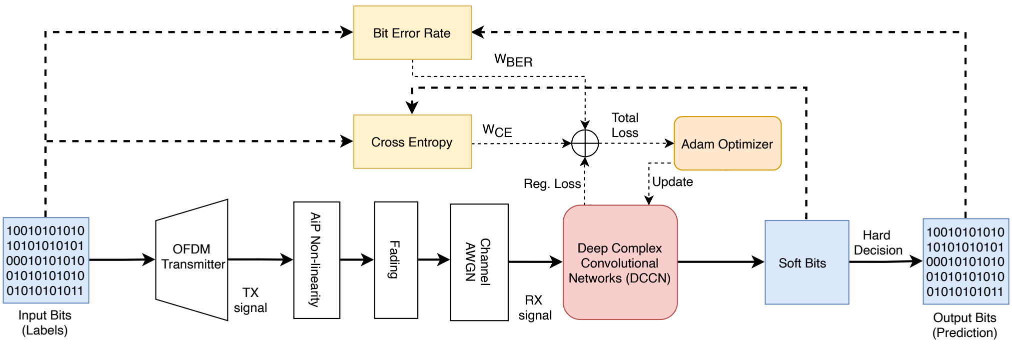


Fig. A.8: Training diagram of the proposed DCCN receiver.

loss and the regularization loss

$$\ell(\mathbf{b}, \tilde{\mathbf{b}}, \Phi) = -\ln\left(\frac{e^{\mathbf{b}_k}}{\sum_i e^{\tilde{\mathbf{b}}_i}}\right) + \epsilon \ell_{reg}(\Phi), \tag{A.9}$$

where k is the index of the target bit, and $\epsilon < 1$ is a small constant. The Adam optimizer that is doing the back-propagation is randomly initializing Φ of DCCN receivers and updates the loss function accordingly.

It is difficult to train the DCCN for fading channels due to the severe distortion that will be implemented on top of the already non-linear distortion introduced by the front-end model. Therefore a two-stage training method is implemented. In stage 1 the neural network is trained with the APA model and AWGN channel. In stage 2, the DCCN is retrained with flat fading and the AiP model for better BER in flat fading channels. This is done by transfer learning. The flow graph of the TensorFlow session freezes the CP remover, data extraction, demodulator, and channel decoder and only keeps training the 1D complex convolution and the channel estimator. Thus the already pre-existing weights in a new session can be used to further enhance the neural network for flat fading and non-linearity. The loss function is the same throughout stages 1 and 2. The graph editing technique enables back-propagation when the second half of the forward network is frozen. This two-stage training approach can increase the data efficiency by reusing the same pre-trained receiver in stage 2 for different fading settings and different active phased array (APA) models.

This increased training efficiency since multiple different channels can be trained with the same initial model. This also allows for more fine-tuning training to compensate for the APA and given channel. To optimize training all operations are vectorized and all data are fed as tensors in the code. Loops are avoided and the learning rate is decayed exponentially for fine-tuning as the training proceeds. For training data, all data is generated online. Hence iterations are used instead of epochs. Early stopping is implemented to make sure that when the BER has not improved for 20 iterations the DCNN model is saved for testing.

Determining SNR during training is not straightforward, however, [23] provides a recommendation of using fixed SNR when only considering AWGN channels and a mix of high and low SNR during flat fading. We choose to use an SNR of 5 dB during AWGN in stage 1 and a variational SNR during flat fading training, stage 2. In stage two different SNRs are used with a mix of both high and low SNR with about a 30/70 split, this is because the channel estimator relies more heavily on the channel statistics and a variation of SNR is therefore needed. This variation is not needed for AWGN since it is a simpler channel model and the DCCN can better generalize for the AWGN using a low SNR of around 3-5 dB.

5 Results

The proposed DCCN receiver is evaluated using numeric results. For comparison, legacy OFDM receivers like LMMSE, ALMMSE, etc. are used. Both Average White Gaussian Noise channel (AWGN) and a Flat fading channel are used with and without training for the non-linear APA model. The selected APA model used for training and evaluation is the AiP at 38 dBm output power 100MHz with steering angle 0. It is highly non-linear as shown in Fig. B.3. APA models for different output powers (31-39 dBm) are used. To further evaluate the DCCN an APA model based on [21] is also used to determine how well the DCCN handles never before seen APA models. Fig. A.9 shows that the DCCN is capable of handling AWGN and non-linear distortion at the same time for different modulation forms. The blue solid line is for the DCCN with no APA model during training. It shows that if the APA model is not known a-priory the DCCN cannot handle the non-linear distortion very well. The blue dotted line is legacy receivers with non-linear distortion. It shows it is not possible to handle non-linear distortion on its own. The striped lines are all LMMSE cases for AWGN with no implementation of an APA model. Hence, if the non-linear distortion is handled at the transmitter side the baseline for legacy receivers is slightly better in higher modulation orders as shown. It is possible to drive the AiP at the highest PAE of 24% while maintaining the BER performance of legacy receivers in different modulations using the DCCN receiver. For other AiPs, APAs, and amplifiers this potential benefit can be higher.

5.1 Soft bit results

Fig. A.10 shows some spread is still present in the DCCN predicted soft bits, but compared to Fig. B.3c there is a big difference. The previously shown phase shift is now removed. Further small variations are shown but have no impact on the hard decision used on the soft bit prediction. The small variations are due to the non-linearity and noise not being fully eliminated.

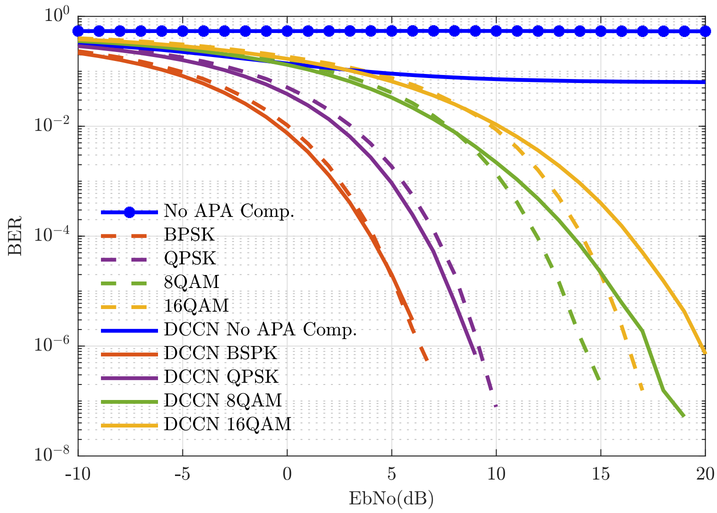


Fig. A.9: Bit error rate plot with the trained decoder over AWGN channel(solid) and LMMSE BER results for comparison(dotted).

5.2 Flat fading

To test the DCCN in flat fading with the AiP it is compared to different legacy receivers, like LMSSE, ALMSSE, Spline, etc. [33]. The transmitter is once again without an APA model. Fig. A.11 shows that the DCCN can compensate for the non-linearity while simultaneous deals just as well as legacy receivers with the flat fading channel. Both LMMSE and DCCN can not compensate for the non-linearity caused by the AiP if they are not aware of it, similar to the AWGN performance. The DCCN does slightly better only in low SNR. This is due to high noise blocking the non-linearity from the DCCN. When trained with the AiP model the DCCN works very well. It is only a few BER worse than conventional approaches like ALMMSE and the DCCN is equal to LMMSE in performance.

5.3 Output Power, and Steering Angle Difference

It is evaluated how well the DCCN compensates for different non-linear models from both the same AiP and different front ends.

The DCCN shows that it is capable of compensating for different power levels without changing the model. If the non-linear frontend model is changed to a different AiP or APA the DCCN cannot any longer function as shown in Fig. A.12. The dotted line represents the changed front-end model and the DCCN does not function prop-

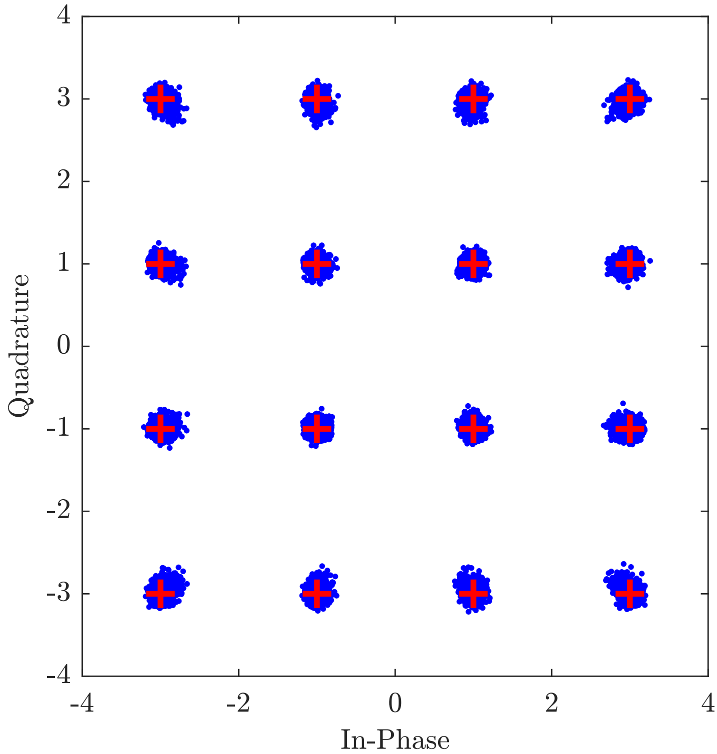


Fig. A.10: Constallation of what the soft bit output of the DCCN receiver. Red crosses is a normal 16QAM decoder scheme. SNR = 20 dB smaller SNR will cause more spread of blue dots.

erly. However, since a two-stage training method is used the front end model can be changed in step 2 which significantly eliminates this problem. This makes training time significantly less costly for reapplying this model to different APA models and channel environments since training time is reduced by one step.

6 Discussion

The method given in this paper shows that the receiver can handle different power levels without having to change the front end model. Further, it shows that the given DCCN receiver can be retrained for any given front end model while only changing the model used in step two of the training process. This is possible due to the equalizer

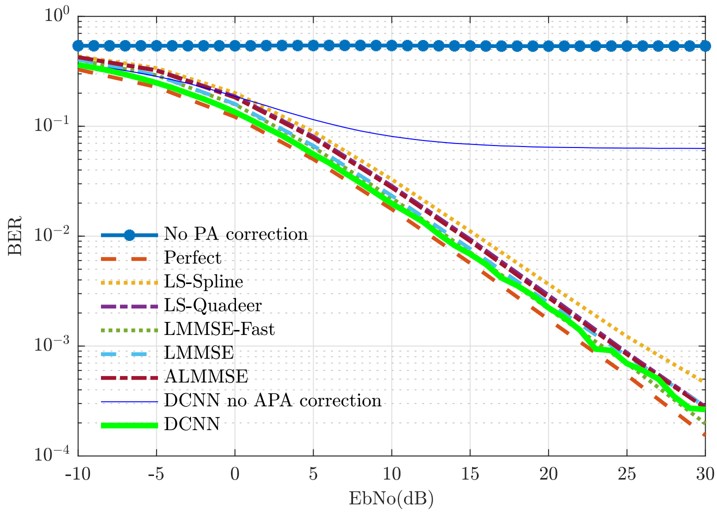


Fig. A.11: Bit error rate plot with the trained decoder over flat fading channel vs conventional OFDM receiver codes.

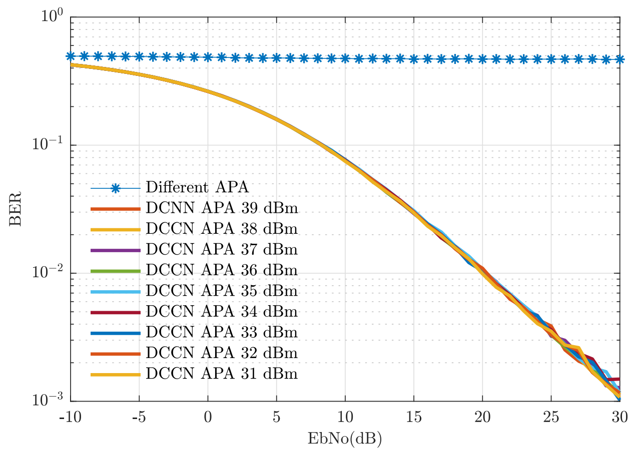


Fig. A.12: Bit error rate plot with the trained decoder over flat fading channel with different power outputs from the APA.

used for channel selective fading and is the most important aspect to achieve high BER performance with non-linear amplifiers. Further, this method makes it possible to not

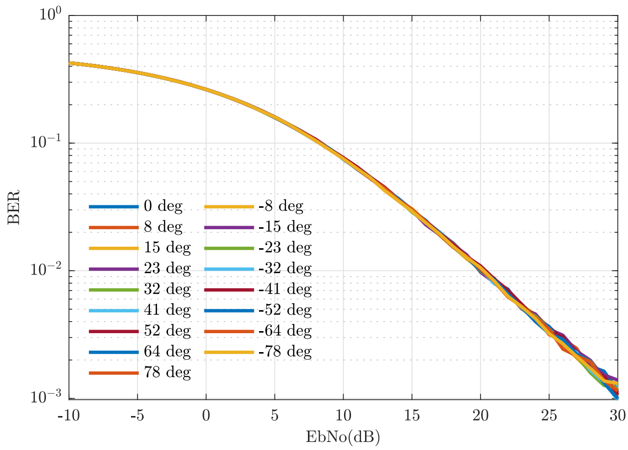


Fig. A.13: Bit error rate plot with the trained decoder over flat fading channel with steering angle changing.

change our transmitter architecture. Thus this can be readily implemented in satellite to ground transmissions. And since only the DCCN needs to be changed at ground station it can service a lot of satellites. For the physical implementation of the system, the model can be pre-trained before deployment and then uploaded to the ground station. If the BER drifts significantly the model can be retrained for the new channel and front end situation. Since the model only needs a synthetic model of the transmitter front end any changes to the whole system can be directly updated and retrained off the actual deployed system.

The impact of steering angle for the AiP can change the non-linearity behavior slightly, but as shown in A.13 this variation is negligible and the DCCN performs as expected. The complex convolution layer that is doing the DFT-like operation together with the equalization layers is why it is possible to adapt to the change in the non-linearity. Since the AiP at different power outputs does not dramatically change the non-linear behavior the DCCN is capable of equalizing for it without re-training. This is due to the strong generalization capabilities of first the DFT-like convoluted layer and second the equalizer. The convoluted layer makes it possible for the DCCN to understand IQ symbols. Any change and variation due to the front end can be estimated by the channel equalizer since the front end disturbances are seen to be part of the channel matrix $\mathbf{H}$. This comes with the cost of generalization for different front ends. However for the same AiP the generalization is very good. Both output power variation and steering angle variaiton is possible to be compensated for at the bit prediction side meaning a good BER is maintained even with high distortion in the signal.

With these findings it is capable that a single ground station can have multiple DCCN models trained for different satellites and compensate for them all. If a multiple satellites are to be serviced by the same ground station and they use similar front ends no retraining is needed for the DCCN model. Only if the front end changes completely a retrain is needed.

Since a legacy transmitter is used, the DCCN is a very robust receiver that can use legacy OFDM transmitters with non-linearity. To use this receiver a single measurement of the given front end with a known signal is needed. Hence for mobile communication systems, the proposed system could be difficult to use due to multiple users all with different APAs. For future work, it is wanted to investigate possible solutions for adapting to different APAs. Hence, the limitation of the proposed DCCN is how many APA models it can compensate for without retraining. It should be noted that the proposed DCCN does not suppress the adjacent channel power at the transmitter side and it could be a problem for terrestrial radio systems. But this is not an issue as band-pass filtering is usually already included before feeding the signal to antenna in most satellite transmitters to fulfill the interference regulations of the International Telecommunication Union (ITU).

To help fast-track evaluation of different APAs, using the over-the-air measurement setup shown in Fig.C.6, it is wanted to implement the trained receiver into a software-defined radio setup. This is left for future work.

7 Conclusion

In this paper, we presented a fully trainable two-stage deep learning receiver that addressed particular non-linearity issues in the satellite transmitter. The deep learning receiver can achieve equal BER to that of conventional approaches and compensate for non-linearity in a dynamically changing transmission link without the need for pre-distortion techniques in the transmitter. This is due to the unique two-stage training scheme proposed that integrates the estimation for the channel and non-linear front end into one. Using an online end-to-end simulator both for training and testing it is possible to numerically validate the proposed method.

The DCCN can deal with different steering angles and output power levels from the same AiP without retraining. It is only needed to include a new AiP in step two, significantly reducing training time for different transmitters to the same ground station. Numerical validation results also show that it is possible to almost double the power efficiency of the chosen AiP without sacrificing BER. This is different from conventional OFDM receivers like ALMMSE and LMMSE since no architecture change at the transmitter side is needed.

References

[1] J. Wang, L. Li, and M. Zhou, "Topological dynamics characterization for LEO satellite networks," *Computer Networks*, vol. 51, no. 1, pp. 43–53, 2007.

[2] S. Egami, "Closed-loop transmitting power control system for k-band satellite communications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 4, pp. 577–584, 1983.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[4] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. F. M. Montoya, J. C. M. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff, J. Querol, L. Lei, T. X. Vu, and G. Goussetis, "Satellite Communications in the New Space Era: A Survey and Future Challenges," *IEEE Communications Surveys Tutorials*, vol. 23, no. 1, pp. 70–109, 2021.

[5] F. Fourati and M.-S. Alouini, "Artificial Intelligence for Satellite Communication: A Review," *Intelligent and Converged Networks*, vol. 2, no. 3, pp. 213–243, 2021.

[6] L. Siyang, Q. Fei, G. Zhen, Z. Yuan, and H. Yizhou, "LTE-satellite: Chinese proposal for satellite component of IMT-advanced system," *China Communications*, vol. 10, no. 10, pp. 47–64, Oct 2013.

[7] K. Rahil and Frisco, TX, US, "System and method for satellite-long term evolution (S-LTE) air interface," U.S Patent 20 100 068 993, March, 2010.

[8] M. Sweeting, "Modern small satellites-changing the economics of space," *Proceedings of the IEEE*, vol. 106, no. 3, pp. 343 – 361, 2018.

[9] Y. Guo, C. Yu, and A. Zhu, "Power Adaptive Digital Predistortion for Wideband RF Power Amplifiers With Dynamic Power Transmission," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 11, pp. 3595–3607, 2015.

[10] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[11] A. Molina, K. Rajamani, and K. Azadet, "Digital Predistortion Using Lookup Tables With Linear Interpolation and Extrapolation: Direct Least Squares Coefficient Adaptation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980–987, March 2017.

[12] V. Karunanithi, C. Verhoeven, and W. Lubbers, "High efficiency transmitter architecture compatible with CCSDS and ECSS standards for nano-satellite missions," in *32nd AIAA International Communications Satellite Systems Conference*, 2014, p. 4252.

[13] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "Ofdm-autoencoder for end-to-end learning of communications systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[14] M. Kim, W. Lee, and D. Cho, "A novel PAPR reduction scheme for OFDM system based on deep learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, 2018.

[15] J. Kim, B. Lee, H. Lee, Y. Kim, and J. Lee, "Deep learning-assisted multi-dimensional modulation and resource mapping for advanced OFDM systems," in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–6.

[16] T. Wada, T. Toma, M. Dawodi, and J. Baktash, "A denoising autoencoder based wireless channel transfer function estimator for OFDM communication system," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 530–533.

[17] T. Van Luong, Y. Ko, N. A. Vien, M. Matthaiou, and H. Q. Ngo, "Deep energy autoencoder for noncoherent multicarrier MU-SIMO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 3952–3962, 2020.

[18] T. J. O'Shea, T. Roy, N. West, and B. C. Hilburn, "Physical Layer Communications System Design Over-the-Air Using Adversarial Networks," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 529–532.

[19] V. Raj and S. Kalyani, "Backpropagating Through the Air: Deep Learning at Physical Layer Without Channel Models," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2278–2281, 2018.

[20] B. Zhu, J. Wang, L. He, and J. Song, "Joint Transceiver Optimization for Wireless Communication PHY Using Neural Network," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1364–1373, 2019.

[21] Y. Zhang, Z. Wang, Y. Huang, W. Wei, G. F. Pedersen, and M. Shen, "A Digital Signal Recovery Technique Using DNNs for LEO Satellite Communication Systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6141–6151, 2021.

[22] Q. Chen, Y. Zhang, F. Jalili, Z. Wang, Y. Huang, Y. Wang, Y. Liu, G. F. Pedersen, and M. Shen, “Robust Digital Signal Recovery for LEO Satellite Communications Subject to High SNR Variation and Transmitter Memory Effects,” *IEEE Access*, vol. 9, pp. 135 803–135 815, 2021.

[23] Z. Zhao, M. C. Vuran, F. Guo, and S. D. Scott, “Deep-Waveform: A Learned OFDM Receiver Based on Deep Complex-Valued Convolutional Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2407–2420, 2021.

[24] P. A. Oche, G. A. Ewa, and N. Ibekwe, “Applications and Challenges of Artificial Intelligence in Space Missions,” *IEEE Access*, pp. 1–1, 2021.

[25] F. Alagoz and G. Gur, “Energy efficiency and satellite networking: A holistic overview,” *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1954–1979, 2011.

[26] M. Ibnkahla, Q. Rahman, A. Sulyman, H. Al-Asady, J. Yuan, and A. Safwat, “High-speed satellite mobile communications: technologies and challenges,” *Proceedings of the IEEE*, vol. 92, no. 2, pp. 312–339, 2004.

[27] A. Sridhar and A. Ephremides, “Comparison of satellite and cellular architectures for downlink broadcast data transmission,” *Int. J. Satell. Commun. Netw.*, vol. 25, no. 1, p. 27–43, jan 2007. [Online]. Available: https://doi.org/10.1002/sat.844

[28] ——, “Sat04-1: Satellite vs. cellular broadcasting : A multi-criteria comparison,” in *IEEE Globecom 2006*, 2006, pp. 1–5.

[29] M. Rahman, T. Walingo, and F. Takawira, “Impact of Varying Wireless Channel on the Performance of LEO Satellite Communication System,” in *AFRICON 2015*, 2015, pp. 1–6.

[30] M. Yang, F. Meng, S. Shi, and Q. Guo, “Markov Chain Based Two-State Satellite Mobile Channel Model,” in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–5.

[31] P. Pradhan, O. Faust, S. Patra, and B. Chua, “Channel Estimation Algorithms for OFDM Systems,” in *Int. J. of Signal and Imaging Systems Engineering*, vol. 5, 01 2011.

[32] *Datasheet on AMO AiP AAiPK428GC-A0404-XXX*, AMOTECH, 2 2019, eVB-R05 190212 b.

[33] O. Edfors, M. Sandell, J.-J. van de Beek, S. Wilson, and P. Borjesson, “OFDM Channel Estimation by Singular Value Decomposition,” *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 931–939, 1998.

Paper B

Adapting to Nonlinear Transmitters with Hybrid Model Training for Neural Receivers

Nielsen, M. H., De Carvalho, E. & Shen, M

The paper has been published in the
IEEE Transactions on Cognitive Communications and Networking , Vol. 9, no. 6, pp. 1657-1665, Dec. 2023, doi: 10.1109/TCCN.2023.3307948..

Abstract

This paper proposes a novel hybrid model transfer learning approach designed for end-to-end OFDM neural receivers that effectively manage multiple channels and nonlinear transmitters. The hybrid model transfer learning method uses mixed Rayleigh channels and other obscured front-end models. This two-step process compensates for nonlinear front-end realizations and different channels, training a robust neural receiver. The neural receiver used is a deep complex convoluted network (DCCN), which replaces the conventional communication blocks with trainable layers that can correct the transmitter's nonlinear performance and other imperfections in the physical layer. This training approach improves the DCCN by 35% for bit error rate (BER), and training time can be reduced by 19% compared to other training approaches for the same tasks while adapting to different fading channels and being robust to noise in power amplifier models. Measurements on both a 28 GHz active phased array in package (AiP) and a GaN Hemt PA show that the trained DCCN can adapt to nonlinear behavior without sacrificing BER. This work demonstrates how training for multiple device operation states and channels helps develop a robust deep neural network capable of demodulating OFDM symbols subject to nonlinear distortions in multiple channel environments without retraining.

1 Introduction

The growing demand for higher-speed data transmissions and extended battery life in mobile devices has made energy efficiency a paramount concern in the field of wireless communication [1–4]. At the hardware level, energy efficiency involves a tradeoff between the nonlinear behavior and the energy efficiency of front-end hardware devices. This tradeoff poses a significant challenge that has been studied for many years [5–8]. Nonlinear transmission results in severe signal distortion, intermodulation distortion, and erroneous demodulation, all of which degrade the performance of wireless communication systems. We propose deep learning as a solution to mitigate this degradation.

1.1 Related Work & Motivation

Recent advancements in deep learning communication systems show significant potential in tackling the challenges posed by 6G communication. Deep learning end-to-end trained systems have been proposed as excellent solutions to the problems faced in 6G communication, such as adaptation to dynamically changing channels [9–11]. These systems employ autoencoders to overhaul the communication system, transmitter, or receiver, thereby enhancing the performance of the communication system.

[9] proposes the use of autoencoders to manage energy-efficient communication systems. For front-end transmitters to be energy efficient, they must use either Peak-

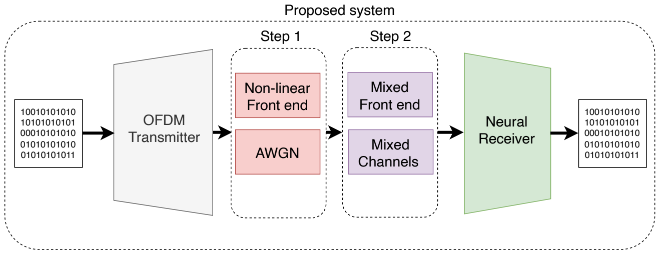


Fig. B.1: Our proposed training approach of the neural receiver. Inspired by [16]. We use the same two-step training approach but expand it with the ability to adapt nonlinearity using a mixed model training approach.

to-average power ratio reduction or pre-distortion to achieve better energy efficiency in the transmitter [12].

Neural networks techniques have achieved greater performance [9–17]. However, they assume that the front-end nonlinearity is well known at the receiver side and cannot change during operation [18, 19]. This can pose a problem when deploying neural receivers, as both the front-end and the environment can change post-training. In [20], a deep learning-based receiver using a one-dimensional transmit and recovery net (1D-TRNet) is robust to nonlinear clipping distortion for OFDM systems. The 1D-convolution kernel described in the paper can handle nonlinear distortion and achieve a better BER than conventional deep learning techniques. However, while it performs similarly to traditional techniques like LMMSE, it also significantly increases the system's complexity. Furthermore, 1D-TRNet cannot manage multiple channels and nonlinear transmitters.

We build upon the groundwork established by [16], which provided a standard training scheme for developing neural receivers with complex convolution for OFDM in fading channels. Utilizing the DCCN neural receiver, we seek to mitigate the nonlinear behavior of different front-end operation stages caused by power adaptability and biasing point differences in the transistor without changing the transmitter side. For individual front-ends, it has been shown to be possible to train the neural receiver to adapt to the nonlinearity [21–23].

To train robust and adaptable deep learning models that can adapt quickly to multiple scenarios without retraining, a deep learning training scheme is needed that can handle obscured or noisy front-end transmissions as the trained nonlinearity. Therefore, optimizing the trainability of the deep learning receiver is crucial to the success of these systems, considering that retraining the model can be both time-consuming and costly.

We propose to use initial training conditions, a mixed front-end, and a channel model approach based on real devices to create a neural receiver capable of adapting to various operational scenarios. Our goal is to demonstrate that the proposed trained DCCN neural receiver can effectively mitigate nonlinear distortion without compromising system performance, and enhance energy efficiency in the front-end of 6G communication systems, all without altering the transmitter side.

1.2 Main Contributions

This work presents a novel hybrid model transfer learning approach designed to effectively reduce the bit error rate (BER) when interacting with nonlinear front-ends operating at operation points. We utilize a two-step training method that incorporates a random mix of different channels and front-end sampled models (as depicted in Fig. B.1. In contrast to previous approaches [16], we augment training rules to expedite the development of a more robust neural receiver capable of managing variations in nonlinear operation regions and biasing points. Additionally, we address the challenge posed by limited knowledge of the front-end device, a problem that arises from noisy samples in the deployment environment.

The contribution of this paper can be summarized as follows:

- We propose a novel hybrid model transfer learning approach for end-to-end OFDM neural receivers to handle transmitters with nonlinear front-ends. This approach would make transmitters more power efficient without inband nonlinearity constraints.
- We provide an analysis on determining the initial conditions for training and methods for training end-to-end receivers.
- We utilize gate voltage and drain current data for different biasing points of the transistor to show the DCCN can handle multiple operation points.
- Transfer learning facilitates robustness in multiple channel environments and with different transistors. In this paper, we only have samples of two completely different transistors. Sixteen transistors had different biasing points. The transistor trained on where not in the data set.
- The obscurity of the front-end samples enables the achievement of these results. In other words, it's crucial to train on non-ideal, nonlinear samples to allow the AI to develop more robustness through noisy training.

The remaining parts of this paper are structured as follows. Section II introduces our method, the DCCN, and discusses models and measurements. Section III outlines our training approach. Section IV presents a comparison of the bit error rate across various channels and different nonlinear power amplifier models. Section V summarises the paper and discusses potential extensions.

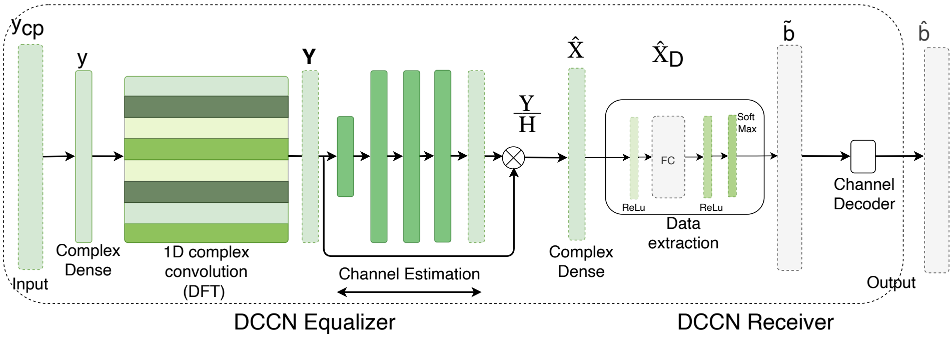


Fig. B.2: Neural network architecture and building blocks of the proposed network. Each neural receiver layer replaces a function from conventional receivers. Proposed architecture similar to [16, 22].

2 Method

A DCCN receiver is used as the base for the neural receiver. The DCCN neural network utilizes signal processing components similar to those in traditional OFDM receivers, with each layer performing distinct functions. Fig. B.2 illustrates the proposed architecture that incorporates similar architecture as [16]. The receiver uses the initial complex dense layer to remove the cyclic prefix from the input signal, y_{cp}. The signal is then converted from the time domain to the frequency domain using the DFT-like C-Conv layer, producing Y. The channel estimator, composed of four dense layers, compensates for amplitude imbalance and estimates the channel response, H. Subsequently, the equalization process involves element-wise complex division, enabling the estimation of $\hat{X}$. Then a data extraction procedure using a complex dense layer, relu activation functions and softmax for output facilitates the conversion to a soft bits likelihood vector, $\tilde{b}$. Tab. B.1 provides a detailed overview of the implemented architecture. A similar architecture without our data extraction layer was used with great success in [16].

We opt for complex convolution over regular convolution for each layer. Our experiments with layer simplification showed that it would not converge and train correctly, a similar finding observed in [16]. The complex convolution allows the neural network to process the IQ signal as a whole, enabling the neural receiver to learn more about the received signal compared to when the IQ signal is split into I and Q components. Furthermore, we refrain from comparing our DCCN receiver with other state-of-the-art methods. Through our own experiments, DeepRX struggles to handle nonlinear signals with unknown variations. This makes it challenging to obtain meaningful results from training DeepRX to handle nonlinear power amplifiers in channel environments other than CDL channels. We believe this is because, unlike DCCN, DeepRX does not incorporate a DFT-like operation in the network.

Table B.1: DCCN network architecture layer specification.

Layer Type	Shape
Input	(Batch, OFDM symbols (N), FFT size)
Complex Dense	(Batch,N,FFT size)
Complex Convolution	(N,N,1) Kernel (3, 3)
Batch Normalization	(Batch,N,FFT size)
Dense	(Batch,N,FFT size)
Dense	(Batch,N,FFT size)
Dense	(Batch,N,FFT size)
Complex Dense	(Batch,N,FFT size)
Max Pooling	(Batch,N,FFT size)
Complex Dense	(Batch,N,FFT size)
Relu	(Batch,N,FFT size)
Dense	(Batch,N,FFT size)
Relu	(Batch,N,FFT size)
Soft max Output	(Batch,OFDM symbols (N),FFT size)

2.1 End-to-end system

The end-to-end system is built around the conventional description of a baseband received signal,

$$\mathbf{Y} = \mathbf{H} * \mathbf{X} + \mathbf{W} \tag{B.1}$$

where $\mathbf{Y}$ is the received signal, $\mathbf{H}$ is the channel matrix, $\mathbf{X}$ is the transmitted symbol matrix, and $\mathbf{W}$ is the additive white Gaussian noise with variance σ^2 per element, and $*$ is the operation of convolution. Some elements in $\mathbf{X}$ are used for pilots dependent on the transmission scheme. The symbols are carrier-modulated based on the constellation (e.g., QAM, QPSK) and are assumed to have an average energy of one before the nonlinear front end. The channel model can be represented as a multipath fading channel, which can be further characterized by a linear finite impulse response (FIR) filter [24],

$$y_t = \sum_{l=0}^{L-1} x_{t-l} \sum_{k=1}^{K} \sqrt{\Omega_k} z_k \operatorname{sinc}\left(\frac{\tau_k}{T_s} - l\right) \tag{B.2}$$

where z_k is a complex-valued random variable, ω, τ_k represents the tap delay profile of the fading process, and T_s is the sampling period of the discrete signal. The length of the filter L is correlated with the second term, which we denote z_l. L is chosen so $|z_l|$ is small when $l < 0$ and $l \geq L$. For Rayleigh fading, the real and imaginary parts of z_k

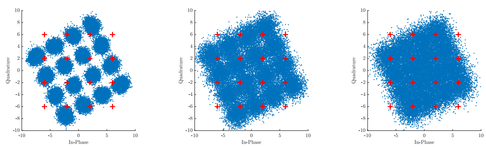
are i.i.d. Gaussian random variables. Thus, $|z_k|^2$ follows a Rayleigh distribution. K is the number of paths in a multipath fading channel. For flat fading channels $K = 1$, i.e., all channel coefficients on all subcarriers are identical. For a multipath fading channel $K > 1$ is a frequency-selective channel, and the channel coefficient varies by subcarrier.

The front-end model is based on the popular memory polynomial model, which can describe the nonlinear distortion introduced by a nonlinear system. The model is given by an equation that specifies how the system's output at a given time depends on the input and previous inputs [5].

$$y(t) = \sum_{n=1}^{N} \sum_{m=0}^{M} a_{nm} x(t - \tau_m) |x(t - \tau_m)|^{n-1}, \tag{B.3}$$

where N is the polynomial order, M is the memory effect, i.e., the number of previous samples that affect the current output. The amplifier coefficients, a_{nm}, are determined using reverse modeling, which involves capturing the input and output of the system and using this data to estimate the coefficients. These coefficients depend on the input signal's peak-to-average power ratio and the baseband signals' power levels [5].

The effects of different channel fading distortions and nonlinearity effects on the IQ signal are shown in Fig. B.3.



(a) (b) (c)

Fig. B.3: AiP nonlinear distortion effect as expressed in the received constellation. a) Nonlinear distorted signal effect on symbols. b) Nonlinear distortion in AWGN channel. c) Nonlinear distortion in a flat fading channel.

2.2 Measurement setup

We identify two distinct challenges related to measurements. The first involves nonlinearity at varying input power levels, while the second pertains to nonlinearity induced by different operation points. To address the first issue, we select the AMOTECH A0404 active phased array in package (AiP) as the front-end device for our model. This device comprises four Anokiwave AWMF-0158 beamforming devices and a 4×4 patch antenna array for transmission [25]. To achieve energy efficiency, the front end must be driven nonlinearly by operating the AiP at different operation points. To assess our

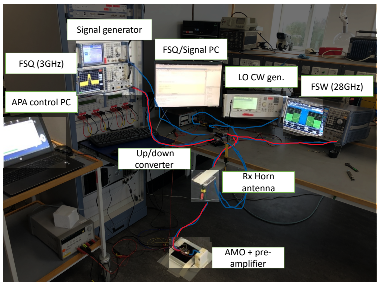


Fig. B.4: Measurement setup of the AiP with all instruments labeled.

DCCN's performance in such a scenario, we conducted various front-end measurements at different input powers.

The AiP nonlinearity is captured using an over-the-air measurement setup, shown in Fig. C.6. The input signal to the system is a 3 GHz 5G NR signal with a bandwidth of 100 MHz, with a peak-to-average power ratio of 10.6 dB is generated by the R&S SMBV100B signal generator. The signal is up-converted to 28 GHz for the AiP. A linear pre-amplifier drives the AiP into saturation. The setup is shown in Fig. C.6. The main beam is aligned with the receiver, a horn antenna placed 44 cm away (42 wavelengths), and connected directly to the spectrum analyzer (R&S FSW 67 GHz) for capturing the data. The measurement procedure is as follows:

1. The 5G NR input signal is uploaded from the PC to the signal generator using the R&S VISA tool.
2. The IQ signal is received and captured using the R&S VISA tool for the FSW.
3. Time alignment with the input and output signal is done post-process to ensure each sample corresponds correctly with the transmitted signal.

We achieve the obscurity of the AiP measurement in post-processing by using artificial noise. We opt for this approach as there is no advantage in capturing the obscured signal over creating it artificially.

For the different biasing points, we measure various CREE GaN HEMT CGH40006P transistors. Due to limitations in the biasing point of the AiP, we cannot utilize it. However, we can employ the same 5GNR mmWave setup at a lower frequency. The measurement is identical except for the up-down converter setup. All PAs are biased with a drain-source voltage of 28 V, and we investigate how a gate voltage of -1 to 3 V impacts the PA's nonlinear behavior.

Fig. B.5 depicts the nonlinear behavior of the AiP and the GaN HEMT PA. Subfigure (a) demonstrates the nonlinear effect on the AiP's amplification. The amplification

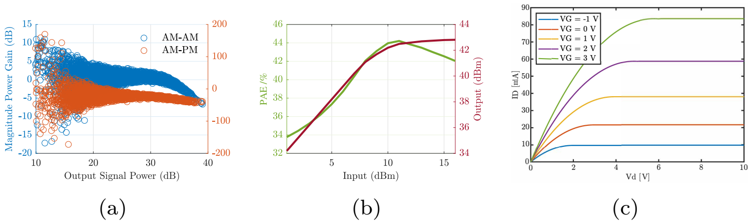


Fig. B.5: Nonlinear effect on a) Amplification in the AiP, b) Power added efficiency of the AiP, c) operation points in the GaN HEMT PA

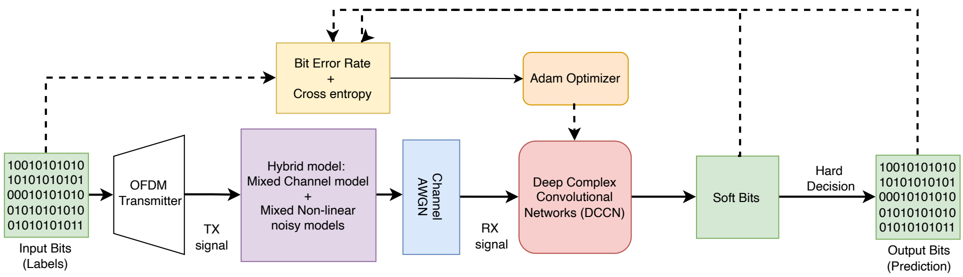


Fig. B.6: Proposed end-to-end autoencoder approach to train the DCCN receiver based on [16]. Mixed fading is not used in the first step, and only one front-end model is used. In the second step, the mixed fading and obscured front-ends are used to retrain the network and get better performance.

changes with variations in the amplitude or phase of the input signal, which can significantly impact the quality of the transmitted signal. Subfigure (b) shows the AiP's power-added efficiency (PAE), defined as the ratio of the output power minus the input power to the consumed DC power. The graph shows the PAE's variation with different gate voltages (V_G), providing an indication of the AiP's efficiency. Subfigure (c) displays the operation points in the GaN HEMT PA, determined by the DC current and voltage. The graph shows the operation points' change with different gate voltages (V_G), which can affect the PA's linearity and efficiency.

3 Training Neural receiver using two-step transfer learning

The neural receiver is trained using a two-step transfer learning approach that utilizes a mixed model for nonlinear power amplifiers and channel models. The nonlinearity model of the power amplifier (PA) is derived from a combination of four distinct PA models.

Our training approach is based on [16], expands to incorporate randomly generated channel models and a randomly selected PA model drawn from the measurements between 31-39 dBm output power. The power amplifier models are made noisy to increase the model's adaptability, with 5,10, and 15 dB SNR.

Similar to [16], the first step involves training the neural receiver on an AWGN channel. However, we train the entire network, equalizer, and receiver and implement a single nonlinear model in the first step. In the second step, we employ a transfer learning scheme, freezing the output layer weights inspired by [26]. In this step, we specifically train the neural network equalization layers to replicate the robustness of linear minimum mean squared error (LMMSE) to tap delay mismatch. This is done since the final layers are used to demodulate, and the initial layers are used for channel estimation. The network is then retrained using our hybrid model composed of different channel models and nonlinear transmitters. The channel models are drawn from a distribution function that can generate different tap delay models with a Doppler spread from 0Hz to 300Hz. To overcome local minima associated with richer multipath, shorter delay spread fading models are used to smooth the overall loss.

The nonlinear model employs the mixed channel approach outlined in [16], generating four distinct models based on empirical AiP measurements. The nonlinearity model, given by (B.3), encapsulates the inherent nonlinear characteristics of the AiP, as presented in Fig.B.5. The model is then modified synthetically using Gaussian noise perturbations to create variations. We use three different noise levels at 5, 10, and 15 SNR and a noise-free model, with one model being swapped for another as the channel model changes during training.

We observed that if we train the DCCN using a freeze layer transfer learning method and only retrain the final layers, we are unable to decode. We obtain a BER of 0.4 or higher for all noise levels, and the neural network does not converge properly, especially when dealing with nonlinear responses. We believe this is due to the complex convolution, which enables the learning of nonlinear behavior in the convolution, thereby propagating the relationship between nonlinear effects on the IQ signal throughout the network.

3.1 Determining Training Parameters

We determine the training parameters using a combination of manual tuning and grid search. We experiment with various values for the learning rate, early stopping, batch size, number of epochs, and with and without cyclic prefixes, considering both long and short versions. All parameters are illustrated in Fig. B.7.

The sweep reveals that using a minimum SNR value of -2.0 during the initial training step ensures a good validation BER for the DCCN, even in step 2. As in [16], we achieve the best results when we include the cyclic prefix and set it to long.

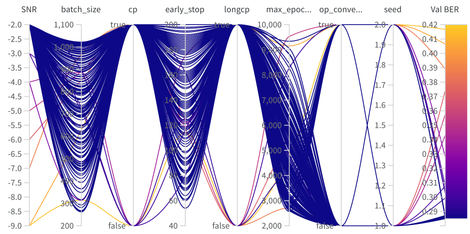


Fig. B.7: WandB parameter sweep of all possible tunable parameters showing that SNR is most important to gain good BER performance. It intelligently finds the best parameter in the first column and tries to reduce the goal of low BER even further.

Table B.2: Best training parameters for the DCCN receiver.

Setting	Value
Maximum iterations	5000
Early stop	50 iterations of no BER change or if BER = 0
Initial learning rate	0.001
Learning rate decay	2% pr 500 iterations
SNR (dB)	-2 to 20 dB
Batch size	512
Optimizer	Adam

Tab. B.2 summarizes the training parameters used for performance evaluation. These parameters, determined from the sweep in Fig. B.7, are the best for training. With the chosen parameters, we can reduce training time efficiently, as we don't need to train as many values as in [16], and we can stop earlier.

3.2 Training Loss

We train our DCCN receiver using binary cross-entropy as the loss function. This loss function is commonly used for binary classification tasks and measures the difference between the predicted and actual outputs. In our case, the predicted output is a probability distribution over the input symbols. The actual output is a binary vector indicating whether a symbol was transmitted.

We also evaluate the performance of our trained network using the BER as a metric. BER , measures the number of bit errors per total number of transmitted bits, and provides a quantitative measure of the network's ability to recover the transmitted symbols. Thus, the total loss function can be defined as the cross-entropy loss function,

$$\ell(\mathbf{b}, \tilde{\mathbf{b}}, \Phi) = -\ln\left(\frac{e^{\mathbf{b}_k}}{\sum_i e^{\tilde{\mathbf{b}}_i}}\right) + \epsilon \ell_{reg}(\Phi), \tag{B.4}$$

where k represents the index of the target bit, and $\epsilon < 1$ is a small constant. The Adam optimizer is initialized randomly. Updates are performed per coherence slot of transmission within a batch size. Throughout the training, we monitor both the binary cross-entropy loss and the BER. Our goal is to minimize both metrics by adjusting the network's hyperparameters and training it on a diverse set of channel models.

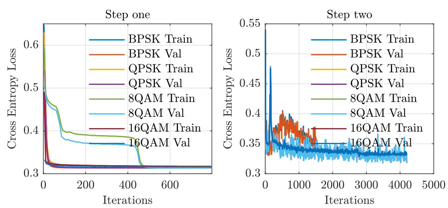


Fig. B.8: Crossentropy Loss, a) is loss pr iteration for initial training b) is loss pr iteration for the second training stage of the proposed DCCN architecture.

Fig. B.8 illustrates a low loss for the first step, contrasted by a high oscillation in the second step. The oscillation of the loss occurs due to the potential for the channel to change from one instance to another, complicating the achievement of a smooth loss curve akin to the training curve in the first step. However, this is not an issue as the aim is to manage multiple different channel realizations. The declining and settling loss curve indicates that our network is not overfitting.

4 Performance Evaluation

A mixed numeric simulator is used for performance evaluation. The simulation uses (B.3) as the nonlinearity model based on real-world experimental data as the input and output of the model. The evaluated modulation schemes include BPSK, QPSK, 8QAM, and 16QAM. The channel models used for evaluation are the Extended Pedestrian A model (EPA), Extended Vehicular A model (EVA), and Extended Typical Urban model (ETU), mobile channels found in the 3GPP standard. Different models are contained in training and withheld for the performance evaluation to evaluate the robustness of the proposed neural receiver to different power amplifiers' nonlinearity due to operation point bias voltage difference and input power level variance.

The parameters of the different channel models are as follows To evaluate the performance of different training methods, we compare our approach, which uses mixed

Table B.3: Key parameters of 3GPP EPA, EVA, and ETU channel models

	EPA	EVA	ETU
Number of Paths	7	9	9
Max Doppler Shift (Hz)	5	70	300
Relative Power (dB)	0.0 to -20.8	-1.5 to -20.9	-1.0 to -16.9

Rayleigh channel models drawn from a distribution and different obscured front-end models, against a single-channel method. The evaluation initially tests the effectiveness of training different modulation schemes using mixed Rayleigh training versus static channel training, evaluated in a flat fading channel. Fig. B.9 shows the results of this comparison, where our new hybrid training approach (denoted as "hybrid" in the figure) outperforms the static channel approach.

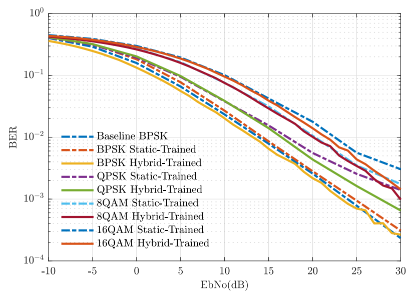


Fig. B.9: Comparison of BER for different modulations using Hybrid Model Training Approach and Static Channel Training Approach, evaluated in a flat fading Rayleigh channel and a fully known AiP model at 32 dBm with no perturbation.

Our approach demonstrates similar performance across different modulation types, allowing the DCCN receiver to maintain performance while compensating for the non-linear front end.

Fig. B.10 demonstrates that the new training method enables the DCCN receiver to improve BER and compensate for the channels simultaneously without significant loss. This is achieved while the power amplifier is in saturation, maintaining a PAE at 24%. We observe that the flat fading implementation, using a mixed Rayleigh channel, performs significantly better at high SNR.

Our method enables the DCCN adapt to multiple channels, performing better than the static training method. As Fig. B.10 shows, the previous training method struggled in highly mobile channels. In contrast, our approach significantly improves decoding in the presence of high channel mobility.

Tab. B.4 shows that our training approach, which uses a distribution of randomly

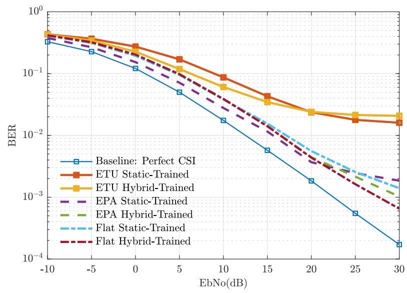


Fig. B.10: Comparison of BER using our Hybrid training scheme compared to single static channel approach BER. Test achieved at 8QAM 32 dBm AiP nonlinearity with different 3GPP channel models.

Table B.4: Comparison of different training time processes relative to the GPU in an unknown tap delay channel at 30 SNR.

Training Scheme	Time	Val. BER	Modulation
Single channel [22]	**7455.1**	0.052618	16 QAM
	5758	0.040873	8 QAM
	6783.2	0.025806	QPSK
	3864.7	0.049292	BPSK
Two channels	8720.1	0.07524	16 QAM
	6688.4	0.049406	8 QAM
	8517.3	0.016422	QPSK
	7790	0.0099228	BPSK
Four channel mix [16]	12265	0.04705	16 QAM
	9688.6	0.034533	8 QAM
	10471	0.01743	QPSK
	8730.7	0.022586	BPSK
Our approach	9927.3	**0.045018**	16 QAM
	8574.6	**0.032042**	8 QAM
	7709.8	**0.0096985**	QPSK
	7378	**0.0070007**	BPSK

drawn Rayleigh fading models, makes the training process quicker than training with a single-channel or a two-channel approach. This demonstrates that our approach significantly enhances performance in random unknown channels. Thanks to its agility, our approach enables faster training. We have reduced the training process by utilizing earlier stops compared to [16]

4.1 Non-perfect sampled Front-end model

The hybrid model transfer learning method uses mixed Rayleigh channels and AiP-impaired models to enable a robust and efficient DCCN receiver that can adapt to most expected scenarios, as shown in Fig. B.10. If the BER changes dramatically, as shown in Fig. B.11, the front-end model can be sampled directly from the environment and the DCCN can be retrained. This provides the significant advantage of enabling the DCCN receiver to compensate for non-ideal deployment scenarios in 6G communication.

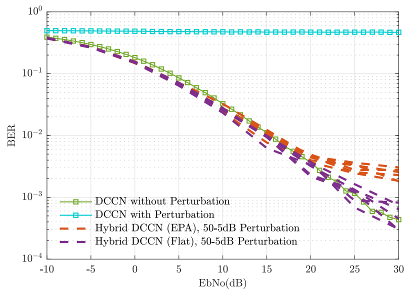


Fig. B.11: Comparison of BER evaluation in a Flat fading channel and an EPA 3GPP channel using static channel training vs Hybrid approach. Pertubation from 50 dB to 5 dB SNR of a 32 dB AiP model.

Fig. B.11 shows that the DCCN cannot compensate for noisy models without using our hybrid model training. Using our proposed hybrid model training approach, we demonstrate that the DCCN can learn and adapt to multiple noisy models without the need for retraining. The DCCN receiver can be trained for each specified sampled noise level of 5, 10, and 15 dB SNR, and it can still correctly demodulate the received signal while compensating for the differences in noisy power amplifier models. The impact of an imperfect AiP model where the nonlinear behavior is changed or obscured can be handled as shown in Fig. B.11. Even when the power amplifier model is noisy, our approach maintains good performance. The DCCN receiver can adapt to each noise level, correctly demodulate the received signal, and compensate for the differences in noisy power amplifier models.

4.2 Testing of different nonlinearity due to average power level changes

We evaluate the adaptability of the neural receiver to changes in the transmitter's power level, which induce varying nonlinearities. Our goal is to determine whether our hybrid method can train the DCCN to accommodate differences in output power nonlinearity. Therefore, the input power given to the AiP will differ, which causes nonlinear behavior at different output power levels from 31 to 39 dB.

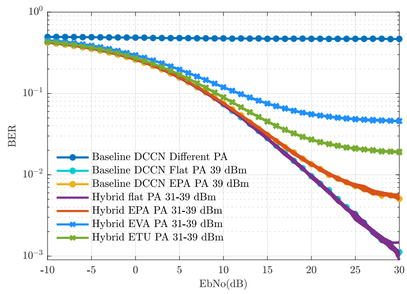


Fig. B.12: Different AiP input power levels effect on received nonlinearity. Output power variation is from 31-39dB. If DCCN is trained using the regular method, it can adapt to a single nonlinearity but not other nonlinearities. Our hybrid model is tested in all 3 channels while we only show the DCCN in 2 channels however, if retrained for ETU and EVA, it will follow the same curves.

The hybrid DCCN training method, which uses a mixed Rayleigh channel and front-end model, proves to be a robust estimator, performing well even when the nonlinearity changes due to variations in the transmitter's power. As shown, it is possible to adapt to different 3GPP channels and nonlinearities without retraining. Similarly, it achieves the same BER as regular training DCCN, maintaining the ability to adapt to multiple power level nonlinearities.

4.3 Testing of different power amplifier operation points

We examine the impact of various operation points and their associated nonlinearities on the neural network receiver. The evaluation is conducted using a hybrid-trained neural network, leveraging AiP mixed model training. While training the hybrid DCCN never sees the GaN Hemt PA nonlinear response. We test in all four different 3GPP channels and check for the robustness of the BER's stability over the different SNR ranges. Fig. B.13 shows that using our training approach of mixed channels and noisy nonlinear models, it is possible to train the neural network receiver for the nonlinearity of different operation points without seeing these nonlinearities in training. However, upon closer inspection of the nonlinear characteristics of the different operation points, we observe that they share some similarities with the AiP, which could explain their adaptability.

5 Conclusion

This paper proposed a hybrid model transfer learning method for training OFDM neural receivers, enabling them to manage nonlinear transmitters in an end-to-end envi-

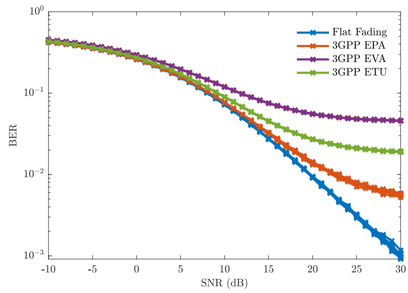


Fig. B.13: Comparison of BER performance using the DCCN on never seen different GaN Hemt PAs using different Gate Voltage and Drain Current to bias the operation point.

ronment. The transfer learning approach enables the DCCN receiver to comprehend imperfect information about front-end nonlinearity and adapt to various nonlinearities induced by different working points and input levels, all without the need for retraining. During training, we use a combination of different Rayleigh channels and measured models of nonlinear front-ends to enhance the robustness of the neural receiver.

To predict bits amidst transmitter nonlinearity, we replace the DFT/IDFT and soft decoder with deep learning layers. These layers process the IQ samples using multiplication rules in complex fields and classify the bits according to the selected modulation scheme.

By using complex-valued neural networks, DCCN can integrate the functions of conventional receivers, such as LMMSE, and concurrently equalize the nonlinear symbols for decoding by the neural receiver.

Our results indicate that our hybrid model training approach yields a 19% faster training time compared to other methods, primarily due to the implementation of early stopping. However, it is slower than training dedicated to a single channel. With DCCN, it is feasible to achieve maximum power efficiency through AiP, even without comprehensive knowledge of the nonlinear behavior, without compromising the BER. The hybrid model-trained DCCN can also adapt to various working points, as validated by the biasing of transistor voltages.

Moreover, our hybrid model training approach can improve BER by up to 50%, significantly outperforming other training methods.

In future work, we aim to extend the benefits of the transfer learning scheme demonstrated by the DCCN neural receiver to other neural receivers.

Furthermore, we have observed that without incorporating highly nonlinear models into the training scheme, adaptation becomes impossible. However, a single model within the training set can adapt to multiple models due to the noise effect applied to the front-end models. Therefore, we aim to explore the potential of using advanced deep learning techniques to enhance the robustness of the neural receiver against out-of-distribution nonlinearities. To achieve this, it will be necessary to measure different mmWave front-end devices and restructure both the training process and the deep learning layers.

References

[1] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten challenges in advancing machine learning technologies toward 6g," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 96–103, 2020.

[2] A. Jagannath, J. Jagannath, and T. Melodia, "Redefining wireless communication for 6g: Signal processing meets deep learning with deep unfolding," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, pp. 528–536, 2021.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[4] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang, and D. Zhang, "A Survey on Green 6G Network: Architecture and Technologies," *IEEE Access*, vol. 7, pp. 175 758–175 768, 2019.

[5] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[6] A. Molina, K. Rajamani, and K. Azadet, "Digital Predistortion Using Lookup Tables With Linear Interpolation and Extrapolation: Direct Least Squares Coefficient Adaptation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980–987, March 2017.

[7] D. Cox, "Linear amplification with nonlinear components," *IEEE Transaction on Communication*, vol. 22, no. 12, pp. 1942–1945, 1974.

[8] Y. Guo, C. Yu, and A. Zhu, "Power adaptive digital predistortion for wideband rf power amplifiers with dynamic power transmission," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 11, pp. 3595–3607, 2015.

[9] O'Shea, T. J. and Karra, K. and Clancy, T. C., "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *Proc. IEEE Int. Symp. on Signal Processing and Information Technology (ISSPIT)*, 2016, pp. 223–228.

[10] O'Shea, T. J. and Hoydis, J., "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[11] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "Ofdm-autoencoder for end-to-end learning of communications systems," in *Proc. IEEE Int. Workshop Signal Proc. Adv. Wireless Commun. (SPAWC)*, 2018.

[12] M. Kim, W. Lee, and D. Cho, "A novel papr reduction scheme for ofdm system based on deep learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, 2018.

[13] T. Wada, T. Toma, M. Dawodi, and J. Baktash, "A denoising autoencoder based wireless channel transfer function estimator for ofdm communication system," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 530–533.

[14] B. Zhu, J. Wang, L. He, and J. Song, "Joint transceiver optimization for wireless communication phy using neural network," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1364–1373, 2019.

[15] Y. Zhang, Z. Wang, Y. Huang, W. Wei, G. F. Pedersen, and M. Shen, "A digital signal recovery technique using dnns for leo satellite communication systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6141–6151, 2021.

[16] Z. Zhao, M. C. Vuran, F. Guo, and S. D. Scott, "Deep-waveform: A learned ofdm receiver based on deep complex-valued convolutional networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2407–2420, 2021.

[17] F. Ait Aoudia and J. Hoydis, "End-to-end learning for ofdm: From neural receivers to pilotless communication," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1049–1063, 2022.

[18] T. J. O'Shea, T. Roy, N. West, and B. C. Hilburn, "Physical layer communications system design over-the-air using adversarial networks," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 529–532.

[19] V. Raj and S. Kalyani, "Backpropagating through the air: Deep learning at physical layer without channel models," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2278–2281, 2018.

[20] Y. Xie, X. Liu, K. C. Teh, and Y. L. Guan, "Robust deep learning-based end-to-end receiver for ofdm system with nonlinear distortion," *IEEE Communications Letters*, vol. 26, no. 2, pp. 340–344, 2022.

[21] Y. Zhang, Z. Wang, Y. Huang, W. Wei, G. F. Pedersen, and M. Shen, “A Digital Signal Recovery Technique Using DNNs for LEO Satellite Communication Systems,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6141–6151, 2021.

[22] M. H. Nielsen, E. de Carvalho, and M. Shen, “A two-stage deep learning receiver for high throughput power efficient leo satellite system with varied operation status,” *IEEE Access*, 2022.

[23] Q. Chen, Y. Zhang, F. Jalili, Z. Wang, Y. Huang, Y. Wang, Y. Liu, G. F. Pedersen, and M. Shen, “Robust Digital Signal Recovery for LEO Satellite Communications Subject to High SNR Variation and Transmitter Memory Effects,” *IEEE Access*, vol. 9, pp. 135 803–135 815, 2021.

[24] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling Methodology and Techniques*. New York, NY, USA: Springer, 2006.

[25] AMOTECH, “Datasheet on amo aip aaipk428gc-a0404-xxx,” https://www.amo-tech.com/datasheet/AAiP28G-A0404-EVB-R05_EVB-R05_190212_b.pdf, Feb. 2019.

[26] A. Brock, T. Lim, J. Ritchie, and N. Weston, “Freezeout: Accelerate training by progressively freezing layers. arxiv 2017,” *arXiv preprint arXiv:1706.04983*.

Paper C

Robust and Efficient Fault Diagnosis of mm-Wave Active Phased Arrays using Baseband Signal

Nielsen, M. H., Zhang, Y., Xue, C., Ren, J., Yin, Y., Shen, M. & Pedersen, G. F.

The paper has been published in the
IEEE Transactions on Antennas and Propagation Vol. 70, No. 7, pp. 5044-5053, July 2022.

Abstract

One key communication block in 5G and 6G radios is the active phased array (APA). To ensure reliable operation, efficient and timely fault diagnosis of APAs on-site is crucial. To date, fault diagnosis has relied on measurement of frequency domain radiation patterns using costly equipment and multiple strictly controlled measurement probes, which are time-consuming, complex, and therefore infeasible for on-site deployment. This paper proposes a novel method exploiting a Deep Neural Network (DNN) tailored to extract the features hidden in the baseband in-phase and quadrature signals for classifying the different faults. It requires only a single probe in one measurement point for fast and accurate diagnosis of the faulty elements and components in APAs.

Validation of the proposed method is done using a commercial 28 GHz APA. Accuracies of 99% and 80% have been demonstrated for single- and multi-element failure detection, respectively. Three different test scenarios are investigated: on-off antenna elements, phase variations, and magnitude attenuation variations. In a low signal to noise ratio of 4 dB, stable fault detection accuracy above 90% is maintained. This is all achieved with a detection time of milliseconds (e.g 6 ms), showing a high potential for on-site deployment.

1 Introduction

The commercial availability of 5G communication systems is estimated to triple in the next coming years. The primary enabler of 5G and the next evolution 6G is the use of mmWave active phased arrays (APAs) as the primary communication front end system [1]. Quality of Service (QoS) is one of the prime motivators for the evolution of communication systems, fault diagnosis of communications systems has never been more important to meet the QoS user expectation. It is critical to identify the failures in the APA efficiently, or even on-site, to make repairs easy and keep the downtime of devices low. Furthermore, failures are not only limited to antenna elements as conventional fault diagnosis techniques have focused on [2–19]. Faults can also happen in the front end circuits such as power amplifiers (PAs) or phase shifters, which makes the fault diagnosis a challenging multi-dimensional problem as illustrated in Fig. C.1.

Fault detection in antenna arrays has been a research topic for many years [2–19]. The techniques for determining faults vary in the mathematic formulation and use different measurement methods. However, they all have in common that they use the electric field component of the antenna array. The measurements also need to be in either near field or far-field and use either 1 probe with a lot of measurements needed or using a plane of probes with fewer measurements needed.

The most prevalent method used in the literature has been the rotating element electric field vector (REV) method [20, 21], hence it is used as the base comparison.

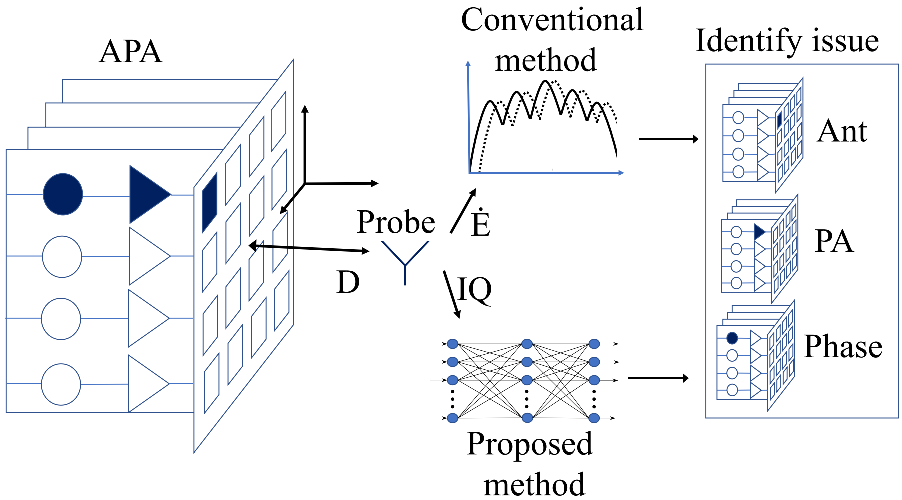


Fig. C.1: APA diagnosis concept, where the conventional approach is using a math estimator based on measured electric field denoted $\dot{E}$ and antenna pattern, and the novel approach proposed in this paper is using DNN and acquired IQ baseband signal at the receiver. The potential faulty components are highlighted with color. The three potential errors are phase shifters represented using circles, power amplifiers represented using triangles, and antenna elements represented by squares, respectively.

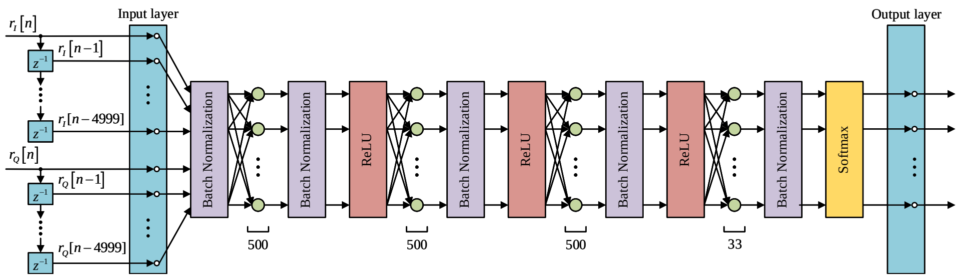


Fig. C.2: The neural network tailored for APA fault diagnosis is a simple 4-layered feedforward architecture, where each hidden layer has a size of 500 nodes.

The special thing with REV is that compared to other approaches for diagnosis of APAs which mostly limits malfunctions to passive antenna elements and phase errors [2–19, 22, 23]. The REV method does not look at passive elements but can detect amplitude, phase, and element faults with very accurate estimations hence why it is also used for calibration arrays before deployment.

Another fault diagnosis method is the inverse scatter problem (ISP) or more specifically electromagnetic inverse problem (EIP) where the antenna array is seen as the scatter. Instead of relying on the antenna pattern, it uses the time-domain of the measurement to make a tap-delay model of the received signal. This measurement can then be reshaped using mathematical equations to produce an answer for what shape the scatter has. This has been a thoroughly researched topic [24–28]. However, it needs multiple input sources/measurement probes to be used. Therefore array diagnosis efficiencies of conventional methods have been significantly limited and not suited for on-site deployment.

Deep Neural Network (DNN) has shown promises in many fields of fault diagnosis [29–32]. Recent research in planar array diagnosis has shown that DNN can identify failures if given the full RF domain antenna radiation pattern for each failure type [33–36].

The multidimensional problem of identifying multiple types of faulty components is tricky and is only solvable using full radiation patterns from anechoic chamber measurements. 5G measurement setups are very sensitive to the positioning of probes due to the scaling down of signal wavelengths at mm-wave frequencies. RF domain-based synthetic method is therefore not suitable for making measurement setup simpler. [2–19]. Newer research has shown that using a comparable method to REV using DNNs it is possible to reduce calculation complexity. But the methods do not reduce measurement complexity [36–39].

This paper proposes a new APA diagnosis technology using baseband IQ signals shown in Fig. C.1. In-phase and Quadrature (IQ) baseband signals are used in many modern digital systems, and the baseband signal characteristics are unique to the APA. Baseband signals typically have a frequency range of up to several hundreds of MHz and are much easier to sample and process. Research has also shown that IQ signals are easier to handle than amplitudes and phases [40]. Hence, to simplify the measurement setup and data acquisition this uniqueness of the baseband signal is exploited in this paper.

The challenge for conventional synthetic methods to use baseband IQ signals is that it is not a trivial task to model the detailed dependence of the IQ signals on the APA components. According to the theory of deep learning, however, a DNN with sufficient parameters can model any mathematical equations [41]. The DNN used in this work consists of 4 layers, and 5549147 trainable parameters, providing compelling characterization capability. Therefore the DNN can capture the "hidden" features in the IQ signals, which cannot be characterized by conventional techniques. The novelty of this work is the proposed method's capability to detect not only the failures in the antenna array but also the failures of active components in the front end. This is new and different from our previous work done at sub 3 GHz with planar arrays [42]. Moreover, the proposed method eliminates the need for multiple measurements at restricted locations opening up new avenues for measurement and test setups, which greatly increases its potential for on-site deployment. To our best knowledge, work that can provide similar outcomes has not been reported yet.

This paper is organized as follows. Section II describes the new proposed method. In this Section the network, faults, measurement setup and training details are described. In section III the experimental results and test of the model are presented. Section IV is a discussion of the proposed method and a comparison with the conventional methods given earlier. Finally, the paper is concluded in Section V.

2 Proposed Method

The exploitation of the baseband signal starts with understanding the basic model of the communication system. The communication system can be seen as a dynamic nonlinear system with memory as the basic model description. This leads to the following distortions constellation, intersymbol interference and adjacent channel interference. Any small imperfection in the front end system will have impact on these and will be part of the baseband signal. Being a dynamic nonlinear system it can be described by Volterra theory where complex baseband input signal x(n) and output y(n) are related as

$$y(n) = \sum_{k=1}^{\infty} \sum_{n_1} \dots \sum_{n_k} h_k(n_1, n_2, ...n_k) \prod_{r=1}^{p} x(n - r_r) \prod_{q=p+1}^{k} x(n - n_q)^* \quad \text{(C.1)}$$

where $h_k(n_1, n_2, ..., n_k)$ are the Volterra kernels. Since any imperfection in the transmitter will have an impact on the output due to the model basis for the Volterra series this can be exploited for the classification and detection of failures in the APA. Using the IQ baseband over a channel that is assumed to be static while transmitting it is possible to create a machine learning algorithm for classification that is better suited for monitoring.

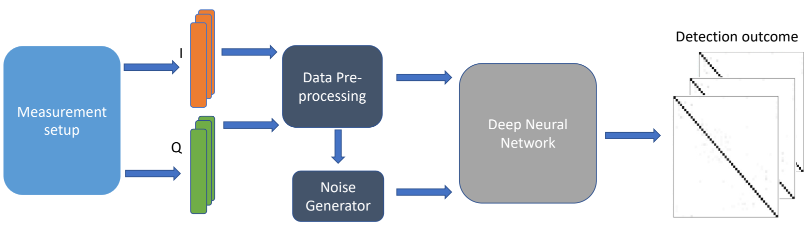


Fig. C.3: The training process of the DNN, where the IQ signal is acquired from measurement setup and pre-processed for the DNN to handle. Here noise can be added for further robustness tests. After going through the DNN different detection outcomes (e.g. confusion matrix) can be given.

The DNN used to identify the different faults is trained using supervised learning to map the different classes to the faults. DNN has been used instead of more complicated networks like CNN or RNN. These other networks can potentially decrease the number of neurons needed in the network but are more complex to feed and pre-process the data. Since pre-processing and labeling the data is the most costly task it has been decided to keep the network simple. For simplifying the training each instance of training data is grouped in batches were 200 training data samples are grouped and processed through the network. Going through all the training data available is one epoch. The programming language used for the DNN is Python and the machine learning library used is PyTorch v 1.4.

2.1 Architecture of DNN

The DNN is composed of four different types of layers, including Fully Connected Layer (FC Layer), Batch Normalization Layer (BN Layer), activation function layer and hidden layers. The full network architecture can be seen in Fig. C.2. In the FC Layer, weights and biases are expressed by $\mathbf{W}_i$ and $\mathbf{B}_i$ where i denotes the ith layer. The output of the FC ith layer can be expressed as

$$\mathbf{y}_i = \mathbf{W}_i\mathbf{x}_i + \mathbf{B}_i, \tag{C.2}$$

where x_i is the input to the ith layer. The number of neurons in the layer is determined by iterative tests to achieve the best performance.

The BN Layer normalizes the mean and variance of the input data to 0 and 1 respectively and then gives the input data a new mean and variance corresponding to the new dimension of the data. This reduces the time cost of the entire training significantly [43]. The output of a BN Layer can be written as

$$\hat{y}_i = \gamma \frac{y_i - E[y_i]}{\sqrt{\mathrm{var}[y_i] + \epsilon} + \beta}, \tag{C.3}$$

where γ and β are the scaling factors for the mean and variance in the new dimensional space. Further, γ and β are learnable parameters. ϵ is a constant parameter that prevents denominator from being zero. Generally, it is set to 0.001.

The output of the BN Layer is fed into the activation function layer which is chosen to be the Rectified Linear Unit (ReLU) function. The ReLU activation function outputs the positive part of its input and everything else as zero.

$$f(x) = x^+ = \max(0, x), \tag{C.4}$$

where x is the input to a neuron. ReLu is chosen over Leaky Relu, since the IQ data can be represented by the magnitude and phase, and rectifying the sine wave does not have a significant impact on these two key factors and the sine wave are horizontally symmetric.

Instead, using half period of the signal reduces the chance for the model to capture too detailed features (e.g. noise) to avoid overfitting. Testing also confirms this since we haven't seen an increase in accuracy due to using leaky ReLu. This procedure transfers using forwards propagation the raw data from the first layer to the last layer. The useful features of the raw data are then extracted layer by layer. After each batch, the weights in the DNN are updated according to the loss which is evaluated by cross-entropy loss in combination with a softmax function

$$\ell(k, \boldsymbol{\pi}) = -\ln\left(\frac{\mathrm{e}^{\pi_k}}{\sum_i \mathrm{e}^{\pi_i}}\right), \tag{C.5}$$

where k is the index of the target class and $\boldsymbol{\pi}$ are the unnormalized posterior class probabilities which are the output of the last layer of the network. To optimize the network weights and biases, the popular Stochastic Gradient Descent (SGD) is chosen in combination with an adaptive learning rate that decreases by a factor of e^{-1} each time the accuracy score of the system stops and stagnates after two epochs. This is to ensure that the gradient descent keeps decreasing to a global minimum without introducing overfitting issues in the network. The SGD optimizer is given as

$$Q(w) = \frac{1}{n}\sum_{i=1}^{n} Q_i(w), \tag{C.6}$$

where w, is the parameter that minimizes $Q(w)$ and is found by solving the following equation

$$w - \eta\nabla Q(w) = w - \eta\sum_{i=1}^{n}\frac{\nabla Q_i(w)}{n}, \tag{C.7}$$

where η is the learning rate of the DNN. The final layer in the DNN is the output layer which handles the classes that are to be determined. This layer can change size depending on how many fault scenarios are to be classified and can be expanded to fit new fault scenarios.

2.2 Fault Scenarios

The different fault scenarios investigated in this work include three major components used in the APA: antenna element, power amplifier, and phase shifter. The different fault scenarios are established by controlling the APA using a software program provided by AMOTECH [44].

Antenna Element Malfunction

Single antenna element malfunction is one antenna element being turned off using AMOTECH software controller while the whole APA is transmitting. Multi-element

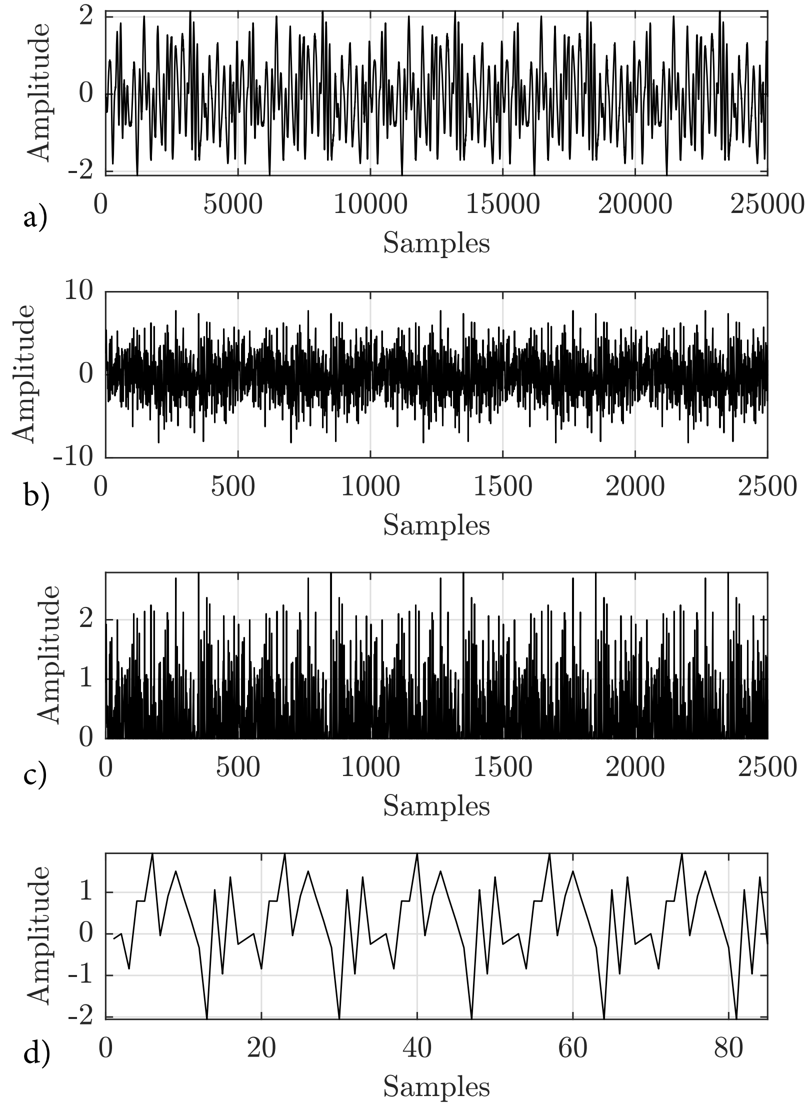


Fig. C.4: Outputs of different layers showing the influence of each layer of the DNN. a) is the Batch Layer, b) is the Linear layer, c) is the ReLU layer and d) is the output layer.

failure is difficult to show since with 16 elements there are a total of 120 combinations for 2 element failure, for 3 element failure there are 560, for four 1820, etc. This becomes a very complex measurement issue and too many classes for multi-class classification to solve. Instead, multi-label fault detection can be used. With multi-label classification, the measurement issue persists but the classification issue is no longer growing too large but can be kept to be 16 classes for element detection since multi classes can be predicted. If all 16 elements can fail the combination of potential failures is 65535, which is very impractical to collect. Instead, this paper limits failures to be grouped into failure categories of faults as 1,2,..,6. Errors of over 6 faulty elements are too many element failures for the power to not have dropped significantly enough that the array is deemed not working. The goal of the test is for the DNN to determine the correct antenna element being turned off and distinguish the different IQ waveforms that are being received.

Magnitude Attenuation Issue

The goal of this test is to determine if the DNN can detect magnitude imbalance caused by undesired attenuation. In the test, the signal attenuation was controlled by the AMO software with a resolution of 0.5 dB(the lowest resolution available). The test is carried out by attenuating one signal path with 0.5 dB using the AMO software and then measuring 10 times as outlined in the measurement procedure.

Phase Fault Detection

The goal of this test is to determine the DNN's capabilities for handling phase faults. Due to limitations of the AMOTECH 0404 phase controller, 1bit phase difference is the lowest phase error possible to introduce while not breaking the APA. 1bit phase difference for the AMOTECH A0404 is 5° of phase shift. Each of the 16 antenna elements is then shifted 5° out of phase from the rest one by one to mimic the absolute minimum phase fault that can happen.

2.3 Data Collection and Measurement Setup

Validation of the DNN was done using a 3 GHz LTE signal, compliant with the 3GPP downlink OFDM modulated also to be used in 5G and satellite high throughput communication [45, 46]. The signal has a peak to average power ratio of 10.6 dB generated by the R&S SMBV100A signal generator.

The 3 GHz OFDM signal from the generator is converted to 28 GHz using an up-converter. A continuous-wave (CW) signal has been multiplied by two into 25 GHz for up-conversion and down-conversion as illustrated in Fig. C.5. The leakage from the local oscillator in the up-converted 28 GHz modulated signal is filtered out using a bandpass filter. To avoid any nonlinearity in multiplier and up-converter, the signal levels in these stages are kept in the linear region of these devices according to their specifications. The 28 GHz signal is then fed to AMOTECH A0404 [44] that includes

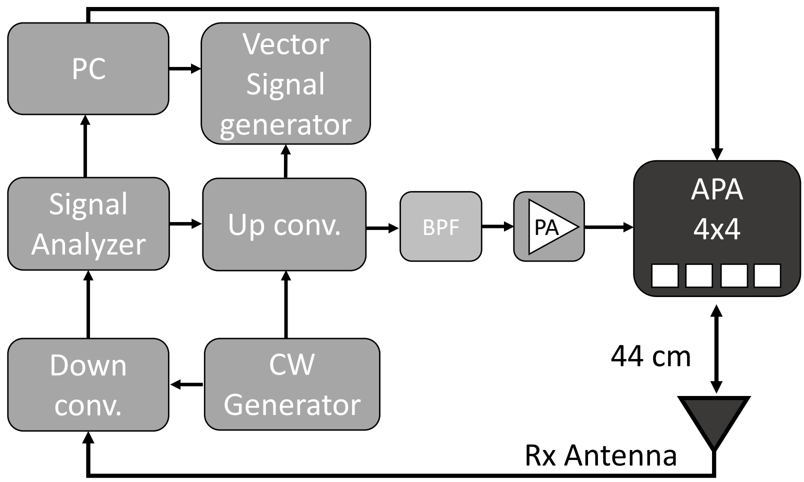


Fig. C.5: Block diagram of the measurement setup.

4 Anokiwave AWMF-0158 beam forming devices and a 4 by 4 patch antenna array as shown in Fig. C.6. The data is captured by the observation horn antenna placed 42 wavelengths away (44 cm) and aligned with the main beam at 0 deg.

The measurement setup for the 4 by 4 AMO APA is shown in Fig. C.6. The captured signal is split into two signal paths. One is analyzed in the spectrum analyzer (R&S FSW 67GHz) for monitoring the actual ACPR. The other is down-converted to a 3 GHz signal and captured by an (R&S FSQ) spectrum analyzer for getting access to I and Q data. It is chosen to only conduct experiments on single failure detection. This means that only one antenna element can be off, have the wrong phase shift, or radiate less power. The signal is captured 10 times with a small random time interval. The procedure for taking measurements is as follows:

1. I and Q waveform for the LTE signal is uploaded by MATLAB from the PC to the vector signal generator.

2. The AMO is manipulated to have a fault using custom software to control the different circuits and antennas.

3. The I and Q at the receiver are then captured from the signal analyzer 10 times with a random time interval between each measurement.

4. Another antenna element or circuit is chosen to be faulty and the measurement is repeated for all 16 antennas.

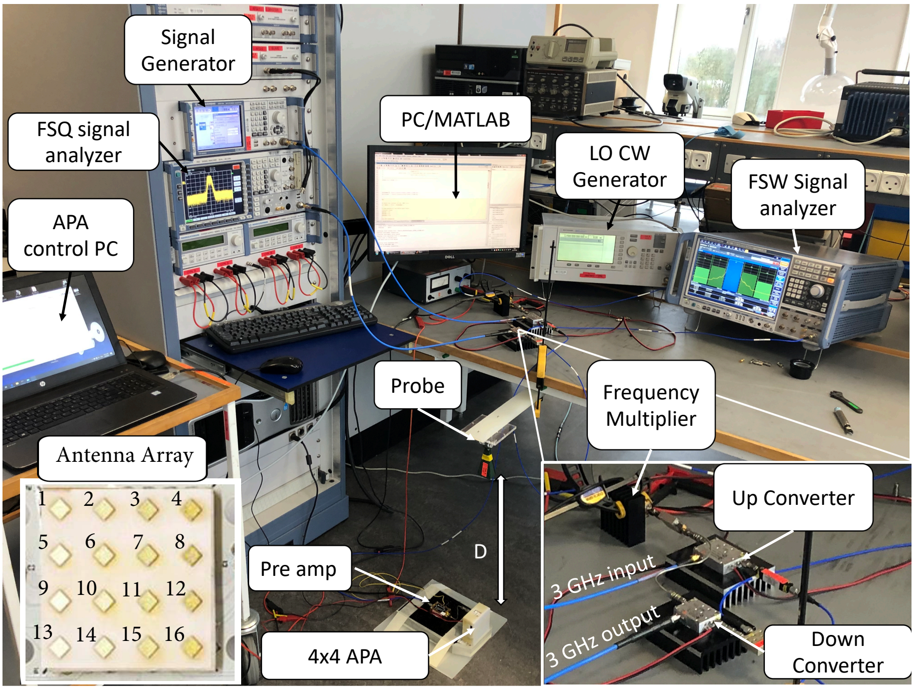


Fig. C.6: The data acquisition measurement setup.

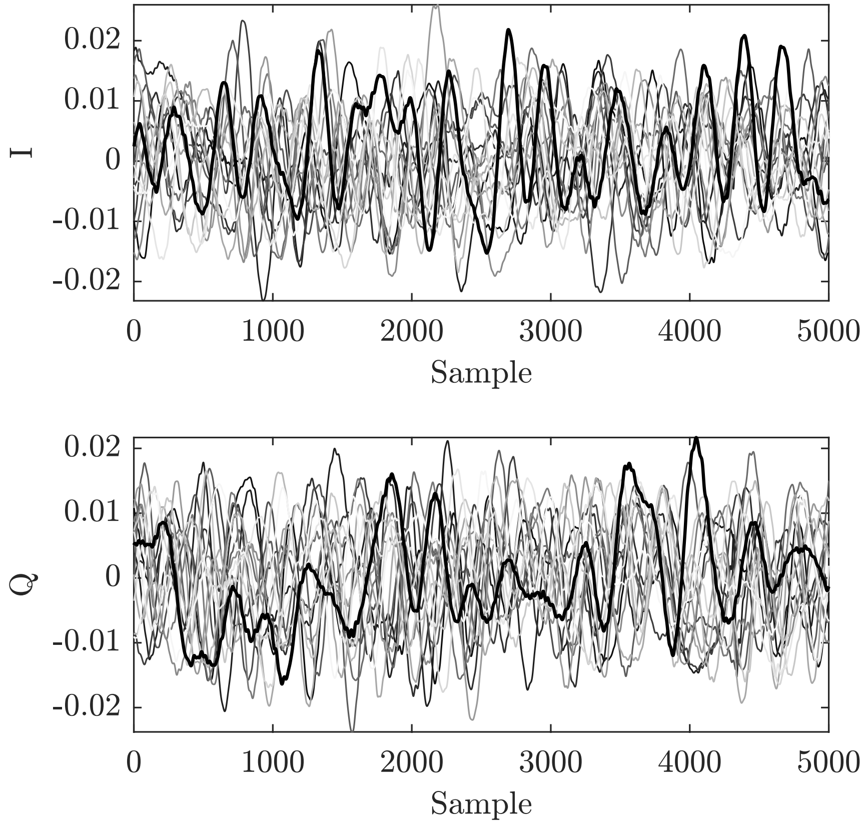


Fig. C.7: Minibatch of measured IQ signals (two examples highlighted with thicker line width) for each of the 17 different magnitude faults.

2.4 Training of the DNN

For each fault scenario, 10 measurements of $5 \cdot 10^6$ samples at a sample rate of 10 kHz are captured using the spectrum analyzer. The 10 measurements are done at random times. This is to ensure the training data is intolerant to time shifts. For training data each class is given as 10 measurements of 5 million samples. To make it easier to handle the data in memory on the GPU it is split into training samples of 5000 I & 5000 Q samples which are interpreted as one data sample of size 10000 when training the DNN. This expands the number of trainable data samples the DNN can re-iterate over. For testing, a new measurement of 5 million IQ samples per class is done. It is also split into smaller data samples of the same size. Fig. C.7 shows different IQ signals features and illustrates

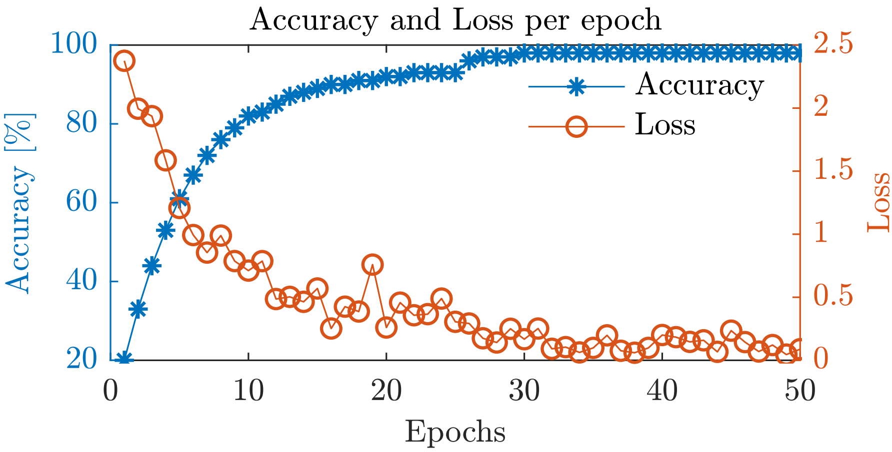


Fig. C.8: The accuracy and loss versus epochs during training.

that the human eyes or synthetic approaches with close form equations characterized simply characterize the features. But as can be seen in the results section, DNN can classify the IQ signals for different faulty cases very well. Pytorch does not support complex numbers. Therefore, the input signal is split into the real and imaginary parts of the IQ sequence. The neural network will find a concatenation between the two in training. The measured IQ data are split randomly into 70% and 30% to form training and testing sets, respectively. The IQ data is put as the first 5000 samples if I and then the next 5000 samples of Q as one vector of size (1,10000) for the input of the neural network. The training data is passed through the network over multiple epochs to learn all the features of the network. After each epoch, the network gets an accuracy score. To keep improving the DNN and make sure no overfitting happens adaptive learning rate is implemented. If the accuracy score of the network does not improve for three epochs, the learning rate is lowered by a factor of ten. This is illustrated in Fig. C.8.

The loss of the network logarithmically decreases while the accuracy exponentially increases as the training data goes through more epochs. It is evident that even though the cross-entropy plateaus in the last training steps, the accuracy keeps improving epoch by epoch until it nearly reaches 100 %. The DNN has a final accuracy of 99 % and a loss very close to 0.

3 Experimental Validation

To evaluate how well the DNN performs in terms of accuracy, the test data is passed through the trained DNN. The results are represented in a confusion matrix shown in

Fig. C.9. From the test data, the accuracy of predicting the correct class is between 88 % and 99 % dependent on which class is predicted. These are very good results considering the DNN has to determine which of the 49 fault classes it is. Hence, the DNN will with a minimum of 88 % confidence correctly identify the failure in the AMO APA. The different labels in Fig. C.9 are denoted by 1 for no faults, by 2-17 for antenna off, by 17-33 for attenuation with 0.5 dB, and by 34-49 for a phase shift of 5 degrees. Each failure is then distinguishable as a label for the DNN to output and the failure label can then be mapped to the component malfunction. In Fig. C.9 the resulting accuracy of each class can be seen. The difference between leakyReLu and ReLu has been investigated together with a comparison to existing methods which can be seen in Tab. C.1. The confusion matrix shows a diagonal matrix where the prediction is between 90% and 98% meaning the DNN can correctly recognize the true class from the test data. Any wrong prediction will be represented as a shade that is not on the diagonal. Further testing has been done for multi-element testing, here faults are grouped in clusters for easier prediction, there are 1-3 element failures, full-chip failure meaning a full chip has failed and 2 PA and 2 phase shifter faults. To predict every failure in the test data set takes 1.1 s while it takes 0.006 s for a single measurement to be predicted on an NVIDIA TITAN RTX GPU. This makes the on-site deployment of the proposed method promising. And this is especially true as the needed I and Q signals are easy to acquire using a simple receiver front end.

The resulting test is shown in Fig. C.10, here it is possible to show the clear separation between different fault cases similar to the performance shown for single element detection.

3.1 Noise's Effect on Diagnosis

To determine how robust the DNN is, a Signal to Noise Ratio (SNR) test is done. To demonstrate this, the test data is polluted with Average white Gaussian noise (AWGN), with a Signal to Noise Ratio (SNR) ranging from -5 dB to 9 dB. The resulting performance of the DNN is displayed in Fig. C.11.

In Fig. C.11, when SNR is (-5 - -2) one class is predicted for all cases(black straight line down). Fig. C.11 shows that when the SNR level is between -1 to 3, the addition of noise does not impede the prediction, and from SNR levels of above 4 dB, the DNN model predicts consistently the correct class though with less accuracy.

3.2 Statistical Analysis of Performance in SNR

To test the Statistical performance of the DNN different distributions of the 70/30 split data are used. It is possible to see how well the DNN is trained and what worst-case and best-case scenarios for data are. By varying the random split of data 10 times, the batches are different from each other with enough variation to test the network's robustness to variations in test data. During this test, all three types of failures, antenna

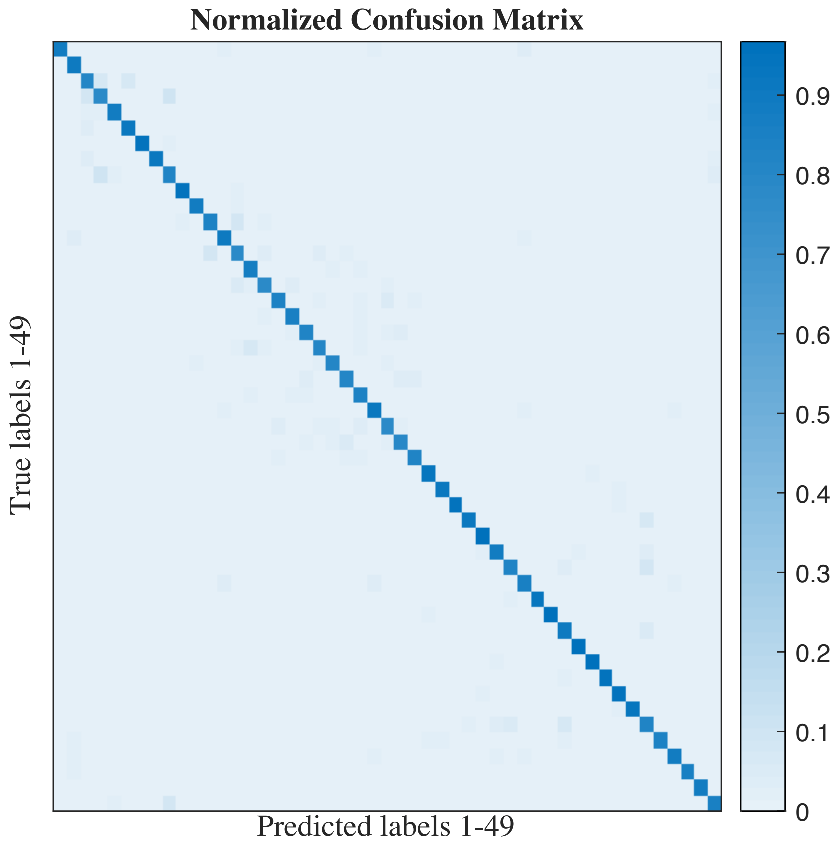


Fig. C.9: Normalized confusion matrix showing all 50 fault testing cases.

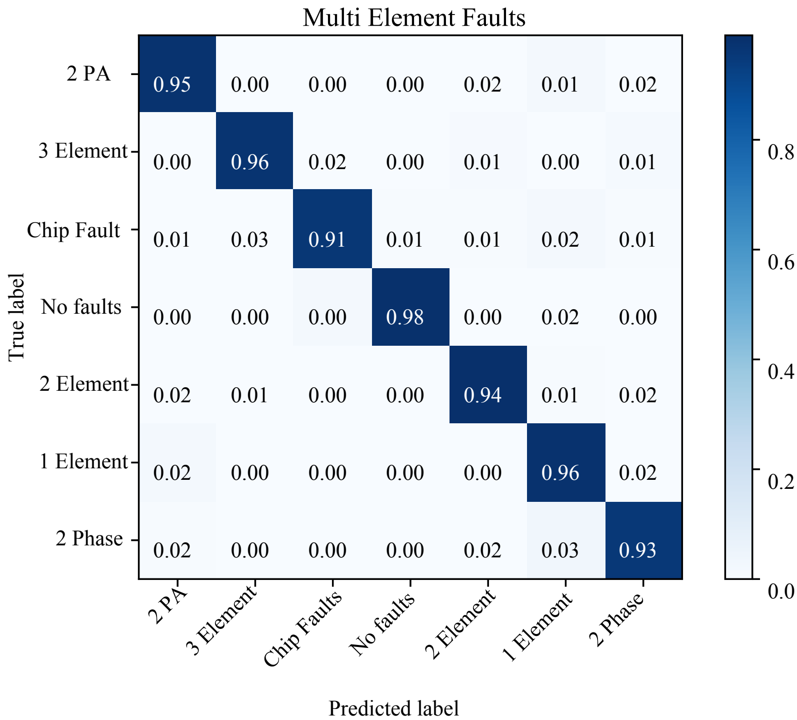


Fig. C.10: Normalized confusion matrix for multi-element fault detection for up to 6 faults.

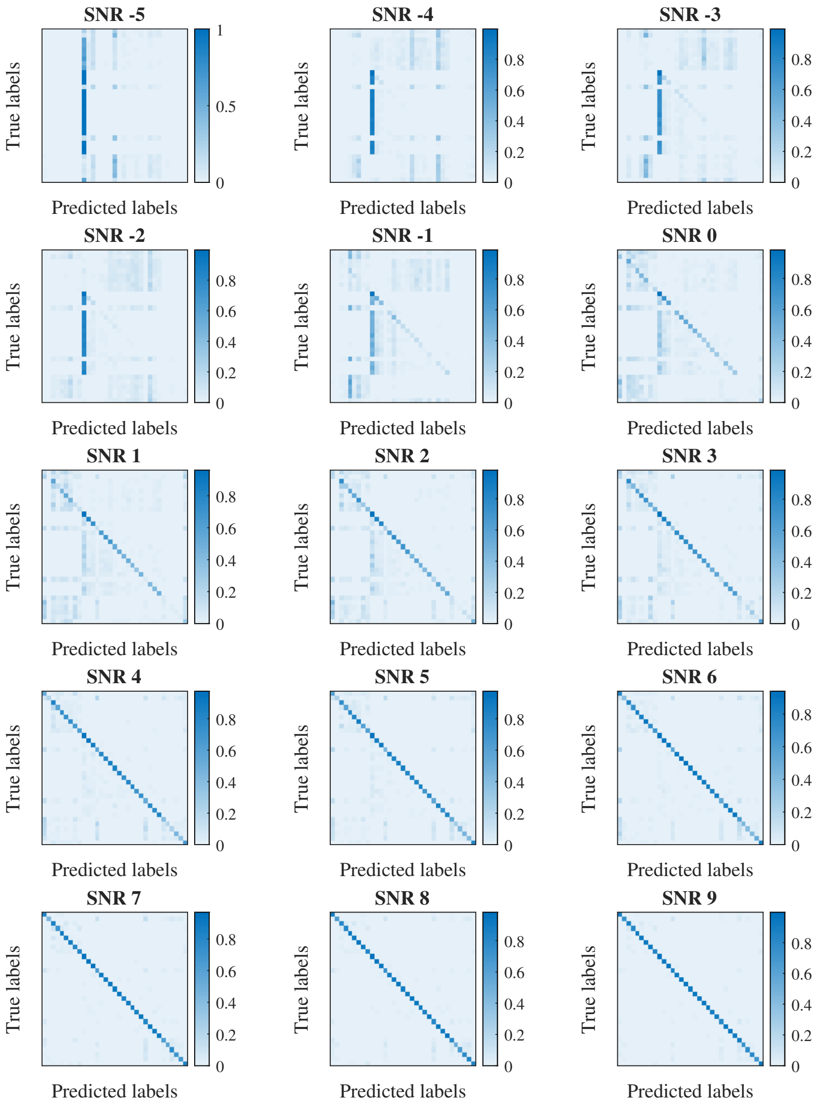


Fig. C.11: Multiple confusion matrixes with SNR -5-9 dB.

elements, PAs, and phase shifters, have been included. Box plots of accuracy versus SNR are used to illustrate the difference of runs over SNR as shown in Fig. C.12 to illustrate the statistical variance over 10 runs. By further looking into the distributions of the different SNR values, it is very clear how the distribution of predictions behaves in

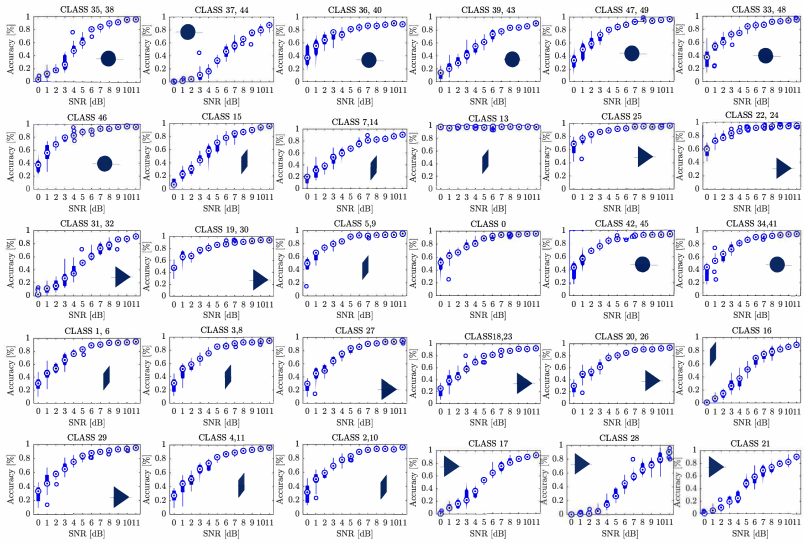


Fig. C.12: Statistical accuracy versus SNR box-plots for three typs of failures: antenna element, PA, and phase shifter. Class 0 is no failures, classes 1-17 are antenna on off detection, 18-33 is amplifier faults and 34-49 is phase shifter faults. Since some classes have same distribution of detection accuracy they have been plotted together to reduce space.

Table C.1: Comparison with Previous Works.

Reference	Method	Accuracy	Robustness Estimate
[20]	REV	100.0%	0%
[22]	Fast only	90.0%	N/A
[23]	Complex signal	100%	N/A
[21]	OTA calibration	100%	50%
[33]	DNN	80.0%	Yes
This work	**DNN**	**99.9%**	**90%**
	DNN LeakyRelu	**93.8%**	**86.7%**

different SNR values. For high SNR values, the predictions cluster around 80-100 % while for low SNR values the distribution gets spread out across the x-axis.

4 Discussion

The DNN can distinguish the 48 faulty classes and 1 not faulty class from each other with an accuracy of between 90 % and 99 % while the faults are of the smallest possible deviation to be introduced in the commercial APA of 0.5 dB magnitude in a single PA and 5° phase shift per antenna element. The DNN's biggest advantage compared to radiation pattern approaches is the ability to use the baseband signal single measurement to predict the faulty components' location and failure type. Conventional and other deep learning approaches can guarantee the correct prediction of antenna elements however it requires multiple antenna probes or measurements points. Big measurement setups are not needed after training as the model only requires simple IQ baseband signals. Further, since the trained DNN model only takes up 22.3 MB of space, it can be used in many devices with limited storage space.

Multi-element classification is shown to be possible, with the difficulty of determining which element is faulty is was not possible to be solved with the method proposed in this paper alone. Multi-label classification is a possible solution to help solve this, however it comes with significant challenges. Simply using IQ data is of too small dimension to directly use multi-label classification. The dimensionality of the data makes multilabel classification only work as a coin toss and is therefore not presented in this paper. Further investigation is needed to solve this problem with more advanced machine learning techniques and measurement setup.

EIP/ISP are related problems but also very different from the proposed solution. This work uses 5G OFDM baseband signal of IQ at a normal sample size to determine faults in a transmitter at a normal frame size.EIP/ISP needs tap delay measurements and this requires a different measurement setup/different pre-processing of the data to get the needed results. This works baseband approach is a new approach to the same problem and could be beneficial in using for over the air (OTA) in situ fault diagnosis.

The DNN is adaptable for different deployment scenarios using new data and transfers learning schemes to update the DNN quickly. This can make it possible to expand the methods for larger antenna arrays.

Instead of having the APA work sub-optimal or not at all, a simple reconfiguration of the APA steering vector can be done to compensate for malfunctioning hardware as long as errors are kept under 2 faulty elements, otherwise, the power drop is significant for this small 16 element array. Moreover, the DNN shows that even at SNR above 4 dB the accuracy of the predictions is maintained above 90 %. However, at SNR below 4 dB, the DNN has difficulties and one class (class 13) is predicted with above 98 %. However, it always gets predicted even when SNR is below 4 dB. It can therefore be concluded that it is possible to trust the DNN when SNR is above 4 dB but when below 4 dB the results should not be given 100 % trust. A reason for this behavior is because the model has found a strong weight for class 13, and when noise is polluting the IQ signal so much that the signal is no longer recognizable from what the DNN has previously encountered this class 13 will be predicted. To potentially eliminate this and be able to handle much larger noise in the signal the DNN should be trained with high noise signals and is future work for this method. But it is very impressive to achieve the robustness of the model without noise addition during training. More details have been provided, but out of range is a bit more difficult. As known, a deep learning classifier does not understand out of range and it will be trying to put it into the learned classes. Hence it has not been included as it is seen as arbitrary that out of range does not work for deep learning systems which is one of the drawbacks of the proposed system.

Other possible future work for this method includes the distance relationship between probe position for training data and testing data, the impact multiple failures will have on the detection accuracy of the array, and the impact of noise in testing and training data.

5 Conclusion

This paper presents a new deep learning-based method for fault diagnosis of active phased arrays (APAs) widely used in 5G and LEO satellite communication systems. The main contribution of this work is the development of the DNN for multiple fault detection while eliminating the need for a big number of measurement points, time-consuming procedures, and costly chambers paving the way for on-site deployment of the fault diagnosis technology. The proposed fault diagnosis technique was tested both for accuracy and robustness using a 28 GHz commercial APA with 48 experimental fault scenarios. The trained deep neural network can distinguish component failures between single path magnitude attenuation in the gain down to 0.5 dB, phase variation of 5°, and array element failure. Prediction accuracy of up to 99 % for scenarios without the added noise and above 90% in presence of significant noises (e.g. SNR of 6 dB) has been achieved. The concept proposed in this work has the potential to be extended to on-site fault diagnosis for more system blocks (e.g. switches, filters, etc.) in advanced communication systems such as 5G and 6G communication systems.

References

[1] F. Boccardi, R. W. H. J. au2, A. Lozano, T. L. Marzetta, and P. Popovski, "Five Disruptive Technology Directions for 5G," 2013.

[2] R. Palmeri, A. F. Morabito, and T. Isernia, "Planar Arrays Diagnosis Through Phaseless Measurements: A Compressive-Sensing-Inspired Approach," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*, March 2019, pp. 1–4.

[3] W. Li, W. Deng, Q. Yang, and M. D. Migliore, "Fast non-convex compressed sensing approach for diagnosis of defective array elements using planar near-field measurements," *IET Microwaves, Antennas Propagation*, vol. 13, no. 11, pp. 1940–1947, 2019.

[4] J. J. Lee, E. M. Ferren, D. P. Woollen, and K. M. Lee, "Near-field probe used as a diagnostic tool to locate defective elements in an array antenna," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 6, pp. 884–889, June 1988.

[5] A. Chakraborty, B. N. Das, and A. Bhattacharya, "Detection of Localised Array Fault from Near Field Data," in *Antennas and Propagation Society Symposium 1991 Digest*, vol. 3, Jun. 1991, pp. 1408–1411.

[6] O. M. Bucci, M. D. Migliore, G. Panariello, and P. Sgambato, "Accurate diagnosis of conformal arrays from near-field data using the matrix method," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 3, pp. 1114–1120, March 2005.

[7] M. D. Migliore, "A Compressed Sensing Approach for Array Diagnosis From a Small Set of Near-Field Measurements," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 6, pp. 2127–2133, June 2011.

[8] T. Takahashi, Y. Konishi, S. Makino, H. Ohmine, and H. Nakaguro, "Fast Measurement Technique for Phased Array Calibration," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 7, pp. 1888–1899, July 2008.

[9] J. Ronen and R. Clarke, "Monitoring techniques for phased-array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 33, no. 12, pp. 1313–1327, December 1985.

[10] M. Salucci, A. Gelmini, G. Oliveri, and A. Massa, "Planar Array Diagnosis by Means of an Advanced Bayesian Compressive Processing," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 11, pp. 5892–5906, Nov 2018.

[11] Y. Chen and I. Tsai, "Detection and Correction of Element Failures Using a Cumulative Sum Scheme for Active Phased Arrays," *IEEE Access*, vol. 6, pp. 8797–8809, 2018.

[12] M. E. Eltayeb, T. Y. Al-Naffouri, and R. W. Heath, "Compressive Sensing for Millimeter Wave Antenna Array Diagnosis," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2708–2721, June 2018.

[13] C. Xiong, G. Xiao, Y. Hou, and M. Hameed, "A Compressed Sensing-Based Element Failure Diagnosis Method for Phased Array Antenna During Beam Steering," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 9, pp. 1756–1760, Sep. 2019.

[14] R. Iglesias, F. Ares, M. Fernandez-Delgado, J. A. Rodriguez, J. Bregains, and S. Barro, "Element Failure Detection in Linear Antenna Arrays Using Case-based Reasoning," *IEEE Antennas and Propagation Magazine*, vol. 50, no. 4, pp. 198–204, Aug. 2008.

[15] M. Kando, "Fault Detection Method with Antennas [Communication Wires]," in *Proceedings of 5th International Conference on Properties and Applications of Dielectric Materials*, vol. 2, May. 1997, pp. 1140–1143.

[16] K. M. Lee, R. S. Chu, and S. C. Liu, "A Built-in Performance Monitoring/fault Isolation and Correction (PM/FIC) System for Active Phased Array Antennas," in *Proceedings of IEEE Antennas and Propagation Society International Symposium*, vol. 1, Jun. 1993, pp. 206–209.

[17] G. Oliveri, P. Rocca, and A. Massa, "Reliable Diagnosis of Large Linear Arrays—A Bayesian Compressive Sensing Approach," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 10, pp. 4627–4636, 2012.

[18] A. Buonanno and M. D'Urso, "A Novel Strategy for the Diagnosis of Arbitrary Geometries Large Arrays," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 880–885, 2012.

[19] Kuan-Min Lee, Ruey-Shi Chu, and Sien-Chang Liu, "A built-in performance-monitoring/fault isolation and correction (PM/FIC) system for active phased-array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 11, pp. 1530–1540, Nov 1993.

[20] S. Mano and T. Katagi, "A Method for Measuring Amplitude and Phase of each Radiating Element of a Phased Array Antenna," *Electronics and Communications in Japan (Part I: Communications)*, vol. 65, no. 5, pp. 58–64, 1982. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecja.4410650508

[21] Z. Wang, F. Zhang, H. Gao, O. Franek, G. F. Pedersen, and W. Fan, "Over-the-air Array Calibration of mmWave Phased Array in Beam-steering Mode Based on Measured Complex Signals," *IEEE Transactions on Antennas and Propagation*, pp. 1–1, 2021.

[22] R. Long, J. Ouyang, F. Yang, W. Han, and L. Zhou, "Fast Amplitude-Only Measurement Method for Phased Array Calibration," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 4, pp. 1815–1822, 2017.

[23] ——, "Multi-Element Phased Array Calibration Method by Solving Linear Equations," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 6, pp. 2931–2939, 2017.

[24] R. Guo, X. Song, M. Li, F. Yang, S. Xu, and A. Abubakar, "Supervised Descent Learning Technique for 2-D Microwave Imaging," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 5, pp. 3550–3554, 2019.

[25] Z. Lin, R. Guo, M. Li, A. Abubakar, T. Zhao, F. Yang, and S. Xu, "Low-Frequency Data Prediction With Iterative Learning for Highly Nonlinear Inverse Scattering Problems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 10, pp. 4366–4376, 2021.

[26] Z. Wei and X. Chen, "Deep-Learning Schemes for Full-Wave Nonlinear Inverse Scattering Problems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 1849–1860, 2019.

[27] ——, "Physics-Inspired Convolutional Neural Network for Solving Full-Wave Inverse Scattering Problems," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 9, pp. 6138–6148, 2019.

[28] G. Chen, P. Shah, J. Stang, and M. Moghaddam, "Learning-Assisted Multimodality Dielectric Imaging," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 3, pp. 2356–2369, 2020.

[29] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3768–3774, 2015.

[30] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, “An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[31] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, “Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, 2019.

[32] L. Wen, X. Li, L. Gao, and Y. Zhang, “A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[33] D. Vakula and N. Sarma, “Using Neural Networks for Fault Detection in Planar Antenna Arrays,” *Progress In Electromagnetics Research Letters*, vol. 14, Jan. 2009.

[34] A. Patnaik and C. Christodoulou, “Finding Failed Element Positions in Linear Antenna Arrays Using Neural Networks,” in *2006 IEEE Antennas and Propagation Society International Symposium*, Jul. 2006, pp. 1675–1678.

[35] A. Patnaik, B. Choudhury, P. Pradhan, R. Mishra, and C. Christodoulou, “An ANN Application for Fault Finding in Antenna Arrays,” *Antennas and Propagation, IEEE Transactions on*, vol. 55, pp. 775 – 777, Apr. 2007.

[36] K. Chen, W. Wang, X. Chen, and H. Yin, “Deep learning based antenna array fault detection,” in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–5.

[37] X. Wang, K. Konno, and Q. Chen, “Diagnosis of Array Antennas Based on Phaseless Near-Field Data Using Artificial Neural Network,” *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 7, pp. 3840–3848, 2021.

[38] G. Álvarez Narciandi, J. Laviada, Y. Álvarez López, G. Ducournau, C. Luxey, C. Belem-Goncalves, F. Gianesello, N. Nachabe, C. D. Rio, and F. Las-Heras, “Freehand System for Antenna Diagnosis Based on Amplitude-Only Data,” *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 8, pp. 4988–4998, 2021.

[39] G. Zheng, Q. Zhang, and S. Li, “Failure Diagnosis of Antenna Array Using Residual Convolutional Neural Network,” in *2021 International Applied Computational Electromagnetics Society (ACES-China) Symposium*, 2021, pp. 1–2.

[40] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, “Deep Learning-Based Signal Modulation Identification in OFDM Systems,” *IEEE Access*, vol. 7, pp. 114 631–114 638, 2019.

[41] M. A. Nielsen, *Neural Networks and Deep Learning.* Determination Press, 2015.

[42] M. H. Nielsen, M. H. Jespersen, and M. Shen, "Remote Diagnosis of Fault Element in Active Phased Arrays using Deep Neural Network," pp. 1–4, 2019. [Online]. Available: https://sfx.aub.aau.dk/sfxaub?sid=pureportal&doi=10.1109/TELFOR48224.2019.8971339

[43] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 448–456. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html

[44] *Datasheet on AMO AiP AAiPK428GC-A0404-XXX*, AMOTECH, 2 2019, eVB-R05 190212 b.

[45] K. Rahil and Frisco, TX, US, "System and method for satellite-long term evolution (s-lte) air interface," Patent 20 100 068 993, March, 2010. [Online]. Available: http://www.freepatentsonline.com/y2010/0068993.html

[46] L. Siyang, Q. Fei, G. Zhen, Z. Yuan, and H. Yizhou, "LTE-satellite: Chinese proposal for satellite component of IMT-Advanced system," *China Communications*, vol. 10, no. 10, pp. 47–64, Oct 2013.

Paper D

Blockage Prediction in Directional mmWave Links Using Liquid Time Constant Network

Martin Hedegaard Nielsen, Chia-Yi Ya, Ming Shen, & Muriel Médard

The paper has been published in the
2023 48th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), Montreal, QC, Canada, 2023, pp. 1-2, doi: 10.1109/IRMMW-THz57677.2023.10299092.

Abstract

We propose to use a liquid time constant (LTC) network to predict the future blockage status of a millimeter wave (mmWave) link using only the received signal power as the input to the system. The LTC network is based on an ordinary differential equation (ODE) system inspired by biology and specialized for near-future prediction for time sequence observation as the input. Using an experimental dataset at 60 GHz, we show that our proposed use of LTC can reliably predict the occurrence of blockage and the length of the blockage without the need for scenario-specific data. The results show that the proposed LTC can predict with upwards of 97.85% accuracy without prior knowledge of the outdoor scenario or retraining/tuning. These results highlight the promising gains of using LTC networks to predict time series-dependent signals, which can lead to more reliable and low-latency communication.

1 Introduction

Millimeter (mmWave) and terahertz (THz) communication is a promising technology for achieving high data rates in 6G and beyond. However, highly directional transmission in these bands makes the link vulnerable to blockage, causing sudden interruptions to a communication link. Consequently, reliable mmWave and THz communication requires predicting the occurrence of blockage [1].

Blockage prediction using machine learning based approaches has been investigated in the literature for both in-band and out-of-band solutions. While the state-of-the-art methods rely on information from multiple sources, either from other frequency spectrums [2] or a camera feed [3], it is more advantageous if we can predict without these additional resources or hardware. Prior work also proposed to use prior statistical observations for meta-learning [4], however, it requires prior information about the surrounding and thus is costly to generalize. In [5], an in-band proactive predictor based on a pre-blockage power signature was proposed using deep learning to enable the prediction of the wireless link status. However, the prediction accuracy decreases significantly when predicting even just slightly further into the future.

In this work, we propose to use Liquid Time Constant (LTC) network [6] for blockage prediction. Since the LTC network is based on ordinary differential equations (ODEs), it has strong expressiveness, stability, and performance in handling time series. Using the publicly available dataset on mmWave blockage scenarios [5], we demonstrate the superior generalization of LTC networks by training using the indoor dataset, which contains pre-blockage power signature under controlled blockage movement and apply the trained model to the outdoor dataset with uncontrolled blockage events. We show accuracy in blockage prediction above 97.85% for all outdoor scenarios for the immediate future timeslot $(t + 1)$. In addition, we demonstrate a 12% to 39% accuracy

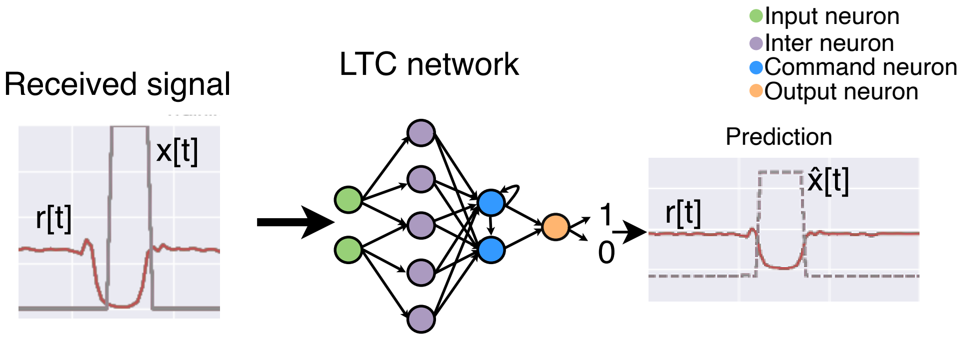


Fig. D.1: Proposed architecture and system model for THz blockage prediction $x[t]$ using LTC network. The input data is the received signal as a function of time samples $r[t]$. The LTC is trained to recognize the pre-blockage signature contained in the data.

improvement in predicting blockage in the near future ($t+5$ and $t+10$ timeslots) compared to the baseline [5]. The proposed model's sparse network and high generalization capabilities make it an attractive option for pre-blockage prediction in high-frequency communication systems.

2 Problem formulation

We address the problem of proactively identifying directional link blockage status using received mmWave signal power. We consider a system with one transmitter and one receiver, each equipped with a directional beam. After the link is established, the link can experience short blockages at unknown times.

We formulate the blockage prediction as a discrete-time problem where $t \in \mathbb{Z}$ is the index of the time samples. The received signal power is represented by $r[t]$, and the link blockage status is defined as $x[t] \in \{0,1\}$ where 1 indicates blockage and 0 indicates no blockage. At time t, given the received signal power samples S_{ob} with an observation duration of T_{ob}, we predict whether the link is blocked in the future T_k timeslots. That is,

$$S_{ob} = \{r[t-T_{ob}+1], \cdots, r[t-1], r[t]\}, \tag{D.1}$$

and the prediction is represented by

$$\hat{x}[t+K], \forall K \in \{1, \cdots, T_k\} \tag{D.2}$$

Our goal is to maximize the blockage prediction success probability:

$$\max \mathbb{P}(\hat{x}[t+K] = x[t+K] | S_{ob}), \; \forall K \in \{1, \cdots, T_k\}. \tag{D.3}$$

In this paper, we propose to solve the blockage prediction problem using a LTC network, as illustrated in Fig. D.1.

3 Proposed Architecture with Liquid Time Constant Network

Liquid Time Constant (LTC) network was first proposed in [6], where LTC is introduced as variations of continuous time models loosely inspired by biological signals. Since LTC networks are based on a system of ordinary differential equations (ODEs), they have strong expressiveness, stability, and performance in modeling time series over a short to medium time. For time-series prediction tasks, the LTC network has been shown to outperform other modern RNNs, and long short-term memory networks on most metrics [6]. Moreover, the LTC network requires fewer neurons and generalizes easier than conventional RNN networks [6]. This behavior of the LTC networks makes it a perfect fit for handling the time series blockage prediction problem.

Neural Circuit Policy (NCP), which creates a sparse network according to [7], is jointly employed to implement an LTC network with fewer resources. Further, the NCPs allow the LTC network to be adaptable to previously unseen data and provide robustness to nonideal data samples, which enables offline training using pre-blockage signature. Once the LTC network is trained, it is deployed to make blockage prediction at each time slot t for the future T_k time slots.

4 Evaluation

To evaluate the performance of an LTC network on blockage prediction, we use a publicly available dataset collected with a directional transmitter and a directional receiver at 60 GHz, consisting of indoor and outdoor scenarios [2, 5]. For the indoor scenario, a controlled blockage event happened during data collection. In comparison, for the outdoor scenarios, uncontrolled blockage events were caused by passing vehicles. The dataset includes the power readings and the ground truth labels for blockage for all time instances. We use the pre-blockage signature under a controlled blockage [5] from the indoor dataset to train the neural network. We then use the trained LTC network to predict blockage events on the outdoor dataset.

We construct an 8-neuron LTC network consisting of two input neurons, four inter neurons, two control neurons, and one output neuron as described in [6] and illustrated in Fig. D.1. We then use the NCPs defined in [7] to construct the neural network, which helps create a very sparse and efficient network for training. We define the input as a series of time data with 64 beam indexes, with power values over time. The output is then determined by the output neuron, either 0 or 1. We then train the neural network for 40 epochs over the complete indoor data set using an Adam optimizer, a cross-entropy loss function, and a learning rate of 0.02. After the training, the LTC network is evaluated on the outdoor scenarios. Since not all 64 beams receive a strong enough signal, we exclude the predictions where the received power is lower than 0.4 normalized received power.

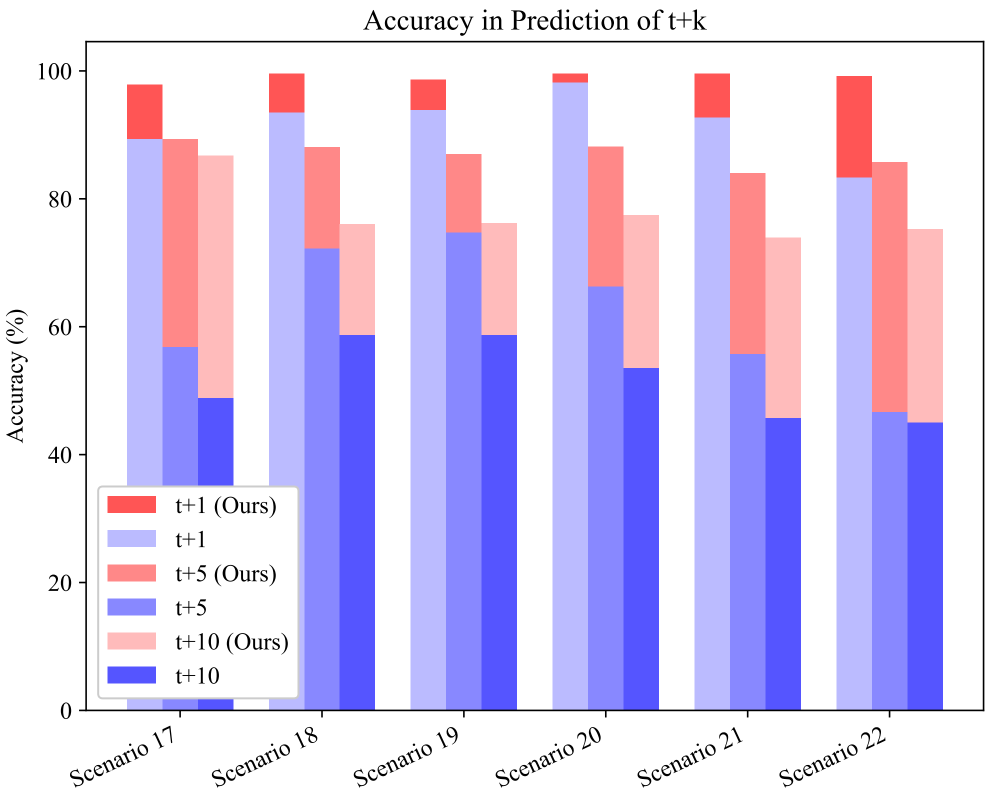


Fig. D.2: Blckage prediction accuracy of different outdoor scenarios, using our approach (red) and the baseline approach [2] (blue).

Fig. D.2 presents the accuracy of blockage prediction in different outdoor scenarios with the scenario numbers defined in the dataset. Fig. D.2 shows that our proposed model outperforms the baseline approach given in [2]. The LTC network accuracy scores range from 73.95% to 99.6%, with the lowest score being significantly better than the state-of-the-art. Our approach shows, in some scenarios, an above 80% accuracy for $t+5$ and above 70% for $t + 10$, indicating our model's advantage in predicting further time in the future. The proposed model's sparse network and high generalization capabilities make it an attractive option for pre-blockage prediction in sub-terahertz communication systems.

Further, the neural network can be quickly deployed and does not need to be trained for individual environments but shows high generalization capabilities for pre-blockage prediction. For future work, we will explore how this pre-blockage prediction can help improve reliability and latency in THz communication systems.

References

[1] Akyildiz, Ian, et al. "TeraHertz Band Communication: An Old Problem Revisited and Research Directions for the Next Decade." arXiv, 2021, https://doi.org/10.48550/arXiv.2112.13187.

[2] M. Alrabeiah and A. Alkhateeb, "Deep Learning for mmWave Beam and Blockage Prediction Using Sub-6 GHz Channels," in IEEE Transactions on Communications, vol. 68, no. 9, pp. 5504-5518, Sept. 2020, doi: 10.1109/TCOMM.2020.3003670.

[3] Alkhateeb, Ahmed, and Beltagy, Iz. "Machine Learning for Reliable mmWave Systems: Blockage Prediction and Proactive Handoff." arXiv, 2018, https://doi.org/10.48550/arXiv.1807.02723.

[4] A. E. Kalør, O. Simeone and P. Popovski, "Prediction of mmWave/THz Link Blockages Through Meta-Learning and Recurrent Neural Networks," in IEEE Wireless Communications Letters, vol. 10, no. 12, pp. 2815-2819, Dec. 2021, doi: 10.1109/LWC.2021.3118269.

[5] S. Wu, et al., "Blockage Prediction Using Wireless Signatures: Deep Learning Enables Real-World Demonstration," in IEEE Open Journal of the Communications Society, vol. 3, pp. 776-796, 2022, doi: 10.1109/OJCOMS.2022.3162591.

[6] Hasani, Ramin, et al. "Liquid Time-constant Networks." arXiv, 2020, https://doi.org/10.48550/arXiv.2006.04439.

[7] Lechner, Mathias, et al. “Neural circuit policies enabling auditable autonomy.” Nature Machine Intelligence 2 (2020): 642-652.

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-577-8

AALBORG UNIVERSITY PRESS