Digital Predistorion of 5G Millimeter-Wave Active Phased Arrays using Artificial Neural Networks

Jalili, Feridoon

DOI (link to publication from Publisher):
10.54337/aau499829274

Publication date:
2022

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

DIGITAL PREDISTORTION OF 5G MILLIMETER-WAVE ACTIVE PHASED ARRAYS USING ARTIFICIAL NEURAL NETWORKS

BY
FERIDOON JALILI

DISSERTATION SUBMITTED 2022

AALBORG UNIVERSITY
DENMARK

Digital Predistortion of 5G Millimeter-Wave Active Phased Arrays using Artificial Neural Networks

Dissertation submitted 2022

Feridoon Jalili

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

Dissertation submitted: June 2022

PhD supervisor: Prof. Gert Frølund Pedersen,
Aalborg University

Ph.D. Co-supervisors: Assoc. Prof. Ming Shen,
Aalborg University
Felice Francesco Tafuri, industrial co-supervisor,
Keysight Technology Inc.

PhD committee: Associate Professor John-Josef Leth (chairman)
Aalborg University, Denmark

Professor Christian Fager
Chalmers University of Technology, Sweden

Professor Nuno Borges Carvalho
Universidade de Aveiro, Portugal

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-883-0

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, 2022

Biography

Feridoon Jalili

Feridoon Jalili received an M.Sc. degree in electrical engineering from Aalborg University, Denmark in 1992. He started his industrial work in the research department of Terma Electronic in Aarhus, Denmark, where he worked on the design of S-band power amplifiers for ballistic radars. His work in mobile communication started in 1996 as an RF designer for Maxon Cellular Systems and later on in Ericsson Mobile Communications, Aalborg. Starting in 2001 he was a system architect at Infineon Technologies responsible for advanced industrial research within mobile technologies. In 2011 he joined Intel Corporation, Aalborg, Denmark as a senior staff RF system architect and was in charge of RF front-ends design of mobile platforms for Intel's tier 1 customers. He holds three U.S. patents in this field. He is currently a Ph.D. fellow in the section of antennas, propagation and millimeter-wave systems (APMS), Department of Electronic Systems at Aalborg University, Denmark. His research area is efficiency improvement of millimeter-wave circuits and systems with a special focus on digital predistortion techniques for the 5G antenna-array transmitters.

Abstract

Transmitter power efficiency is of fundamental importance in mobile communication. This is particularly important in the emerging 5th Generation (5G) of mobile communications due to the even harder challenges posed by the Millimeter Wave (mmWave) frequency range with propagation losses. Furthermore, the high level of hardware integration requires trade-offs between cost, size, current consumption and transmitter power. The 5th generation of mobile communication uses massive multiple inputs multiple outputs (mMIMO) transceivers together with directional radio transmission using a narrow beam to achieve high transmission gain to a specific user. This architecture requires antenna arrays with many elements and active devices directly connected to each element. With this configuration, challenges arise such as a limited area for active devices, antenna crosstalk, mutual coupling between elements and high power consumption.

The 5G signal's modulation format is based on orthogonal frequency division multiplexing (OFDM) which has a high peak to average power ratio (PAPR). Furthermore, 5G mmWave supports bandwidths as large as 400 MHz, which may lead to severe memory effects in the power amplifiers (PAs). The high PAPR and large bandwidths together with nonlinearity can have a huge impact on the error vector magnitude (EVM) and the adjacent channel leakage ratio (ACLR) distortion.

The conventional way of mitigating distortion and correcting for nonlinearities in PAs is through digital pre-distortion (DPD), where the PA's nonlinearities are captured so that an inverse model can be applied in the digital baseband. The most common DPD architectures rely on nonlinear behavioral models based on the Volterra series, such as the memory polynomial model (MPM) or generalized memory polynomial model (GMPM). These linearization methods have adequate performance but for wide bandwidth and high nonlinearity cases, the computational complexity is extreme and needs a huge number of coefficients. On the other hand, the hardware complexity of the large antenna-array transmitters in the 5G system introduces additional limitations for conventional DPD approaches.

Artificial neural networks (ANN) are well known to be able to learn any arbitrary nonlinear function according to the universal approximation theorem. When comparing

the ANN approach with the MPM approach, the MPM has inherent local approximating properties in contrast to the global approximation capability of ANNs, when modeling strongly nonlinear systems. In addition, when compared to classical models, the ANN may adapt better to extrapolating beyond the zone exploited for parameter extraction.

This Ph.D. thesis mainly deals with defining specific design challenges for mmWave active phased array (APA) architectures and proposing solutions and algorithms for linearity enhancement of APAs based on ANN together with advanced over-the-air (OTA) measurement techniques for verification of the proposed methods.

Resumé

I mobilkommunikation har senderens effektivitet fundamental betydning. Dette er især vigtigt i den nye 5. generation (5G) af mobilkommunikation på grund af de endnu sværere udfordringer, som Millimeter Wave (mmWave) frekvensområdet med udbredelsestab udgør. Desuden kræver det høje niveau af hardware integration afvejning mellem omkostninger, størrelse, strømforbrug og sendeeffekt. Den 5. generation af mobilkommunikation bruger "massive multiple input multiple outputs" (mMIMO) sendere og modtagere sammen med retningsbestemt radiotransmission ved hjælp af en smal bølge for at opnå højere transmissionsforstærkning til en specifik bruger. Denne arkitektur kræver antenne arrays med mange elementer og aktive enheder direkte forbundet til hvert element. Med konfigurationen opstår udfordringer da der er et begrænset område til aktive enheder, lækage mellem antennerne, gensidig kobling mellem elementer samt højt strømforbrug.

Modulationsformatet af 5G signaler er baseret på "ortogonal frequency division multiplxing" (OFDM), som har et højt spids til gennemsnit effektforhold (PAPR). Desuden understøtter 5G mmWave høj båndbredde på op til 400 MHz, hvilket kan føre til såkaldte "memory effect" i effektforstærkerne (PA'er). Den høje PAPR og store båndbredde gør sammen med ulineariteten at det kan have markant indflydelse på forvrængning af det ønskede signal (EVM) og kan generere forstyrrende signaler i nabokanaler (ACLR).

Den konventionelle måde at begrænse forvrængning og korrigere for ulineariteten i PA'erne er gennem digital pre-distortion (DPD), hvor PA's ulinearitet læres og en invers model anvendes i det digitale base band. De mest almindelige DPD-arkitekturer er afhængige af ulineære modeller baseret på Volterra-serien, såsom memory polynomial model (MPM) eller generalized memory polynomial model (GMPM). Disse linearisering metoder giver tilstrækkelig forbedring, men i tilfælde af større båndbredde og høj ulinearitet er beregningskompleksiteten ekstrem og kræver nogle gange et stort antal koefficienter.

Kunstig intelligens neurale netværk (ANN) er velkendte for at kunne lære enhver vilkårlig ulineær funktion ifølge den universelle justeringsteori. Når man sammenligner ANN-tilgangen med MPM-tilgangen, har MPM naturlige lokal-minimum justerings egenskaber i modsætning til ANNs globale justerings evne, når man modellerer stærke

ulineære systemer. Derudover kan ANN sammenlignet med klassiske modeller tilpasse sig bedre til at ekstrapolere udover den zone, der bruges til parameterekstraktion.

Denne PhD afhandling drejer sig hovedsageligt om at definere specifikke designudfordringer for mmWave "active phased array" (APA) arkitekturer og foreslår løsninger og algoritmer til linearitetsforbedring af APA'er baseret på ANN sammen med avancerede over-the-air (OTA) måleteknikker til verifikation af de foreslåede metoder.

Contents

Thesis Details

Thesis Title:	Digital Predistorion of 5G Millimeter-Wave Active Phased Arrays using Artificial Neural Networks
Ph.D. Candidate:	Feridoon Jalili
Supervisors:	Prof. Gert Frølund Pedersen, Aalborg University Assoc. Prof. Ming Shen, Aalborg University Industrial co-supervisor Felice Francesco Tafuri, Keysight Technology Inc.

This thesis is submitted as partial 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 papers with the main part of the thesis being scientific papers published in, or submitted to, peer-reviewed journals and conferences. The thesis is a result of three years of research in the section of Antennas, Propagation and Millimetre-wave Systems (APMS), Department of Electronic Systems, Aalborg University, Denmark.

This work was supported in part by the "Innovations Fund Denmark", project of MARS2 (Modular Advanced Radio for Satellite Services).

The main body of this thesis consist of the following papers:

[A] Feridoon Jalili, Martin H. Nielsen, Ming Shen, Ole K. Jensen, Jan H. Mikkelsen and Gert F. Pedersen, "Linearization of Active Transmitter Arrays in Presence of Antenna Crosstalk for 5G Systems", *2019 IEEE Nordic Circuits and Systems Conference (NORCAS)*

[B] Feridoon Jalili, Felice F. Tafuri, Ole K. Jensen, Yunfeng Li, Ming Shen and Gert F. Pedersen, "Linearization Trade-offs in a 5G mmWave Active Phased Array OTA Setup", *IEEE Access Year: 2020 / Volume: 8 / Journal Article*

[C] Feridoon Jalili, Daniel E. Serup, Ondrej Franek, Ming Shen and Gert F. Pedersen, "Antenna Array Inter-Element Coupling impact on Linearization of Active Phased

Array", *International Symposium on Networks, Computers and Communications (ISNCC) Year: 2021 | Conference Paper | Publisher: IEEE*

[D] Feridoon Jalili, Felice F. Tafuri, Ole K. Jensen, Yunfeng Li, Qingyue Chen, Ming Shen and Gert F. Pedersen, "Tuning of Deep Neural Networks for Over-The-Air Linearization of Highly Nonlinear Wide-Band Active Phased Arrays", *International Symposium on Networks, Computers and Communications (ISNCC, Year: 2021 | Conference Paper | Publisher: IEEE*

[E] Feridoon Jalili, Yufeng Zhang, Markko Hintsala, Qingyue Chen, Ming Shen and Gert F. Pedersen, "Highly Nonlinear and Wide-Band mmWave Active Array OTA Linearization Using Neural Network", *IET Microwaves, Antennas & Propagation | Year 2021 | Journal Article*

[F] Feridoon Jalili, Felice F. Tafuri, Kasper B. Olesen, Lauge F. Dyring, Jakob G. Brask, Ming Shen and Gert F. Pedersen, "Complexity Analysis of Artificial Neural Networks Used for Active Phased Array Linearization", *2022 IEEE MTT-S International Wireless Symposium (IWS), Year: 2022 | Conference Paper | Publisher: IEEE*

[G] Feridoon Jalili, Felice F. Tafuri, Qingyue Chen, Ming Shen and Gert F. Pedersen, "Bandwidth scalable behavioral modeling using neural network based on transfer learning", *IEEE Access Year: 2022 | Submitted 2022 | Journal Article*

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 have been provided to the Ph.D. school for approval prior to the submission of the thesis. These co-author statements have also been presented to the Ph.D. committee and included as a part of 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 a part of 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] Emilio J. Martínez-Pérez, Feridoon Jalili, Ming Shen, Jan H. Mikkelsen, Ole K. Jensen, Gert F. Pedersen, "T-LINC Architecture with Digital Combination and Mismatch Correction in the Receiver", *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC) Year: 2019 | Conference Paper | Publisher: IEEE.*

[2] Qingyue Chen, Yufeng Zhang, Feridoon Jalili, Zhugang Wang, Yonghui Huang, Yubo Wang, Ying Liu, Gert Frølund Pedersen, Ming Shen, "Robust Digital Signal Recovery for LEO Satellite Communications Subject to High SNR Variation and

Transmitter Memory Effects", *IEEE Access, Year: 2021 | Volume: 9 | Journal Article | Publisher: IEEE.*

[3] Yunfeng Li, Yonghui Huang, Martin H. Nielsen, Feridoon Jalili, Wei Wei, Jian Ren, Yingzeng Yin, Ming Shen and Gert F. Pedersend, "A Cross-Mode Universal Digital Pre-Distortion Technology for Low-Sidelobe Active Antenna Arrays in 5G and Satellite Communications", *Journal Article | Published: Electronics, Published: Year 2021.*

[4] Jakob Gjedsted Brask, Kasper Bruun Olesen, Lauge Føns Dyring, Arun Yadav, Feridoon Jalili, Yufeng Zhang, Ming Shen, "Deep Digital Signal Recovery for LEO Satellite Communication in Presence of System Perturbations", *2021 IEEE MTT-S International Wireless Symposium (IWS), Year: 2021 | Conference Paper | Publisher: IEEE.*

[5] Yunfeng Li, Yonghui Huang, Qingyue Chen, Feridoon Jalili, Kasper B. Olesen, Jakob G. Brask, Lauge F. Dyring, Gert F. Pedersen and Ming Shen, "Hybrid Digital Pre-distortion for Active Phased Arrays Subject to Varied Power and Steering Angle", *Journal Article | Published: IEEE Microwave and Wireless Components Letters , Published: Year 2022 .*

Thesis Structure

The thesis consists of two parts. Part I is the introductory chapters and part II includes a collection of the scientific papers written by the author.

Part I includes seven chapters, where chapter 1 provides a short introduction and describes the research objectives included in this thesis. Chapter 2 introduces challenges in the linearization of the transmitters in the 5G system. The measurement techniques are described in chapter 3. The behavioral models for 5G transmitters are in chapter 4. Chapter 5 covers the digital predistortion using artificial neural networks. The contribution summary is in chapter 6 and finally, chapter 7 concludes the thesis.

Part II includes the contributions in the form of 7 scientific papers (A to G).

.

Preface

I have been working with mobile development from 1st generation to now 5th generation. My last position in the industry was at Intel mobile communication where I have been working on the architecture and design of RF hardware for tier-one mobile phones manufacturers. The complexity of the 5G mmWave architecture in terms of beam-forming, wide bandwidth and advanced modulation schemes, have introduced huge challenges to the power efficiency of the transmitters. Using the conventional methods from previous generations of mobile phones for the hardware architecture, test, verification and efficiency improvement of the new generation seems not to be feasible. On the other hand development of the technologies based on artificial intelligence (AI) for mobile communications is growing very fast and industry and academia see the potential of these technologies. There are still different opinions in the society, where some think AI is promising and others think its potential for communication systems is limited. The great potential of AI in solving the new challenging problems that emerged in 5G/6G (e.g. linearization of active beam steerable array), has motivated me to start this Ph.D. work for exploration of these challenges. It was a great opportunity to combine my experience in hardware design together with the development of new algorithms to get comprehensive results that are validated not only based on simulations but also on hardware realization and measurements in the laboratory.

I would like to thank my supervisor Prof. Gert Frølund Pedersen for offering me funds to pursue my Ph.D. project and for his great support throughout my studies. I want also to thank my co-supervisor Assoc. Prof. Ming Shen for guiding me through this Ph.D. and for his invaluable support. A very special thanks goes to Dr. Felice Francesco Tafuri from Keysight Technology, Barcellona, for his guidance and very helpful discussions. Further I would like to thank my colleagues and specially Assoc. Prof. Ole Kiel Jensen, and head of laboratory, Kim Nielsen who helped me during this period. Thanks for all the good conversations and the enjoyable, professional and relaxing working environment.

In the end I would like to thank my wife who has been a huge support for me during these sometimes challenging years.

Feridoon Jalili, Aalborg University, June 16, 2022

Part I

Introductory Chapters

Introduction

1 Problem Statement

The motivation and research goals of this thesis are explained in this chapter.

1.1 Motivation

The 5th generation (5G) of mobile communication has introduced new services to fulfill the requirements of the growing need for connectivity. The three main services are: enhanced Mobile Broadband (eMBB), massive Machine Type Communications (mMTC) and Ultra-Reliable Low Latency Communications (URLLC) [1]. eMBB is an extension of the 4G for providing high data throughput and enhancing coverage. mMTC introduce the Internet of Things (IoT) technology, where a massive amount of devices are connected to a network. These devices have very low power consumption and communicate small streams of data. URLLC is used for Industry 4.0 applications [2] and operations where low latency is required such as autonomous driving or remote surgeries [3, 4]. Fig. 1 illustrates a block diagram of these services.
This thesis tries to reveal the mechanism and solutions for nonlinear devices, which are beyond the capacity of conventional methods. Some of the main challenges that constitute the motivation for this work are:

- Hardware architecture
- High PAPR and ultra-wide bandwidth
- Over-the-air measurement and linearization

1.1.1 Hardware architecture

5G mobile communication has introduced frequency range 2 (FR2, five bands between 24.25 GHz and 43.5 GHz) among others to enhance the increasing demand for higher

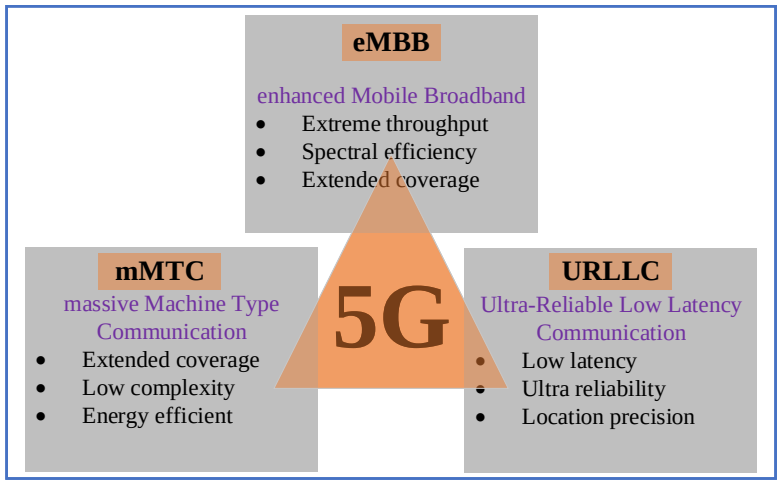


Fig. 1: Representation of hybrid beam-forming architecture.

data rate and coverage. In combination with the new frequency spectrum, new technologies such as massive multiple-input multiple-output (MIMO) have been introduced which can achieve spatial multiplexing by using beam-forming technology based on a large array of antennas [5]. The pure digital beam-forming architecture which was used for some applications in 4G can not be efficient in massive MIMO due to cost and size issues [6, 7]. Therefore hybrid beam-forming has been considered as an efficient solution. On the other hand, the implementation of conventional DPD requires a feedback loop for each branch of the antenna array which is practically not feasible in massive MIMO. Therefore, DPD applied on multiple transmitting branches which include one PA per branch is more practical. The nonlinear behavior of each PA in the branch can be different and it could be required to have different power levels for different branches. So there is a trade-off between the achievable linearization quality and complexity of the applied DPD. The beamforming structure arises additional new challenges such as antenna cross-talk and mutual coupling between elements. Therefore the cancellation of the linear crosstalk between the branches has been investigated heavily in academia and industry [8–10].

1.1.2 High PAPR and ultra-wide bandwidth

The 5G modulation scheme is based on orthogonal frequency division multiplexing (OFDM) which has a non-constant envelope and a high peak to average ratio (PAPR). This profile helps the system to utilize the frequency band efficiently while operating

at high power. One of the main challenges that need to be tackled while designing a transmitter with a non-constant envelope (spectrally efficient) is to maintain the requirements for linearity, e.g. below 12.5 % points EVM and below -28 dBc ACLR for downlink 64-QAM baseband signal [11]. These requirements for the in-band and out-of-band distortions need to be fulfilled while performing high efficiency in form of power consumption. On the other hand, the high power efficiency requires that active devices such as power amplifiers operate in the nonlinear region while maintaining the linearity requirements. Furthermore, 5G has introduced ultra-wide bandwidths, up to 400 MHz. The high bandwidth results in additional amplitude to amplitude (AMAM) and amplitude to phase (AMPM) gain distortions due to the memory effect. The conventional DPD techniques can not easily handle these new challenges without increasing the complexity.

1.1.3 Over-the-air measurement and linearization

The 5G mmWave transmitter structure based on beam steering uses a highly integrated RF front-end which makes the conductive measurement and linearization techniques non-feasible. Furthermore, the nonlinear distortion is beamformed in the same direction as the intended signal [10, 12]. Assuming the same adjacent channel power ratio (ACPR) in all directions, the maximum ACP will be in the direction where the most power is sent which is the direction where the user is. Therefore in 5G mmWave, over-the-air (OTA) measurement of nonlinear distortion in different spatial sectors is required. It means that with a well-designed wireless network the linearity requirements for a multi-beam configuration can be reduced compared to the single antenna concept [13]. Therefore OTA measurement and linearization techniques is necessary for the 5G mmWave system and providing the appropriate OTA setup for the design, test and verification of the devices is an important topic.

1.2 Research Goals

The evolution of mobile technology toward enhanced 5G and 6G has imposed new challenges i.e. ultra-wide bandwidth, high PAPR, active phased array with a large number of elements and OTA linearization and verification. Industry and academia are trying to suggest solutions for these challenges together with architectures that involve improvements in form of cost, size and power efficiency. In this context the architecture of the transmitter has a major impact since it is one of the most power-consuming parts of the system. This Ph.D. project mainly deals with defining specific design challenges for mmWave active phased array architectures and proposing solutions and algorithms for linearity enhancement of active phased arrays together with advanced OTA techniques for verification of the proposed methods. The scientific contributions described in the following chapters include:

1. Identification of the challenges and proposing solutions for OTA linearization of mmWave active phased array (APA) transmitters.
2. A complexity analysis of different methods for linearization of the APA which includes a comparison between the artificial neural networks (ANN) method and the traditional memory polynomial (MPM) method.
3. An experimental comparison of state-of-the-art ANN-based DPD and MPM-based DPD.
4. Assessment of the robustness of ANN-based DPD with respect to operating condition such as the beam angle, signal bandwidth and power levels.
5. An investigation of the reduction of computational complexity using transfer learning neural network (TLNN) for systems with dynamically changing bandwidth.

2 Challenges in 5G Transmitter Linearization

Traditionally power amplifiers face a never-ending linearity-efficiency trade-off. Fulfilling the specifications required in 3GPP standardization requires linear operation while linear operation implies efficiency degradation. Linearization techniques allow the PA to operate in the nonlinear regions while fulfilling standardization requirements and providing sufficient efficiency. Several efficient techniques for achieving these goals have been used by the industry for years, such as Doherty PA, envelope tracking (ET) and digital predistortion (DPD). By introducing new technologies such as MIMO and beamforming, implemented in 4G and 5G, the linearization technique not only involves a single PA but linearization of an array of PAs connected directly to the small patches of an active phased array antenna. This new structure requires additional linearization techniques dependent on the topology of these systems. This chapter gives an overview and explanation of these challenges together with a description of the topologies considered for the modern transmitter system.

2.1 Beam-forming Architectures used in the Active Phased Arrays

5G mobile communication system in the FR2 frequency range uses a beam-forming architecture together with antenna arrays which have several elements to fulfill the requirements of high gain. The passive antenna gain, as used for BS type 1-O, is specified to 17 dBi [11]. These multi-element antenna arrays are often designed as active phased arrays (APA), where the transmitter power amplifiers or the receiver low noise amplifiers are directly connected to the antennas. Generally, there are three types of beam-forming architectures for the APA:

2.1.1 Full digital beam-forming architecture

In this architecture each antenna element has its own RF chain, digital to analog converter (DAC) and dedicated baseband section, illustrated in Fig. 2a. All the digital processing including the adjustment of phase and amplitude of each antenna is done in the baseband module and therefore there is no need for analog phase and amplitude shifter circuits in RF domain. There are two main constraints in dedicating an RF chain and a data converter to each antenna:

1. Space limitation: All devices in each chain (baseband and RF transceiver chips, PA, LNA and filters) should be packed behind each antenna element and since the elements are placed close to each other then there is a space limitation.

2. Power consumption: PAs and ADCs are power-hungry devices. A digital conversion per stage leads to a large demand for digital signal processing (many parallel

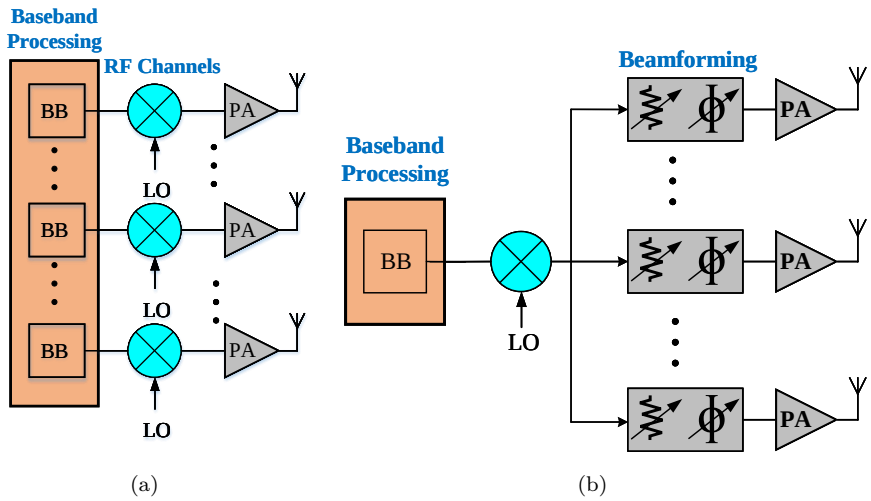


Fig. 2: (a): Full digital beam-forming architecture; (b): Analog beam-forming architecture

Giga samples per second data stream to be processed) which leads to excessive power consumption.

2.1.2 Analog beam-forming architecture

Analog beam-forming, Fig. 2b, is the simplest MIMO approach and it is dedicated to single-stream transmission. Analog beam-forming is implemented using a network of digitally controlled phase shifters. In Fig. 2b the phase and amplitude shifters are placed before the PAs but they could be placed in other positions in the RF chain. The phase shifter weights are adaptively adjusted using digital signal processing with a specific strategy to steer the beam and meet a given objective, for example, maximize transmitted signal power to specific user. Phase shifters can be active or passive. Active ones have problems with noise and linearity whereas passive ones have problems with large size and large insertion loss [14].

2.1.3 Hybrid beam-forming

In the hybrid approach which is shown in Fig. 3, the digital precoder/combiner can correct for lack of precision in the analog beam-forming, for example to cancel residual multi-stream interference. This allows hybrid precoding to approach the execution of the unlimited solutions [6, 7]. The optimization process is divided between analog and digital domains. In scenarios with multiple users, for each user, the individual baseband

modules can be allocated in order to increase the gain. Using more baseband modules increases the complexity but at the same time, the steering capability of the system will be advanced. So in hybrid beam-forming, there is a trade-off between complexity and capacity [14].

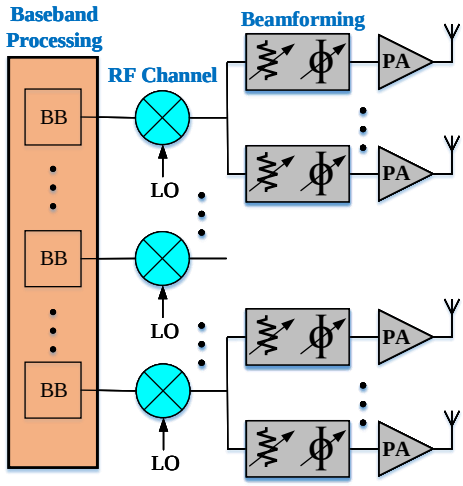


Fig. 3: Representation of hybrid beam-forming architecture.

2.2 Cross-talk in Active Phased Array

Crosstalk is used as a term that describes the unwanted coupling between signals in each PA-antenna chain (branch) of the array. Crosstalk can be categorized into two types: before PA (nonlinear crosstalk) and after PA (linear crosstalk) [15]. Generally, the power amplifiers of the active phased array need to operate close to the saturation region to provide high efficiency. The output impedance of the PA is not necessarily 50 ohm and may need to be matched to an optimum load which is a compromise among output power, linearity and efficiency requirements [16]. Since each PA is directly connected to the antenna without any isolator in-between, then the antenna mismatch and the mutual coupling between the antennas have a direct impact on the load impedance of the PA [17]. Part of the radiated waves from the PA output will be reflected back and also radiated to the other antenna elements. A block diagram of a generic active antenna array is shown in Fig. 4. It includes L antenna elements where for the kth branch, b_{2k} and a_{2k} represent the output and reflected signals from the antenna array, and a_{1k} and b_{1k} are the incoming and the reflected signals from amplifier's input.

The power injected into one antenna element of the array will be coupled back to

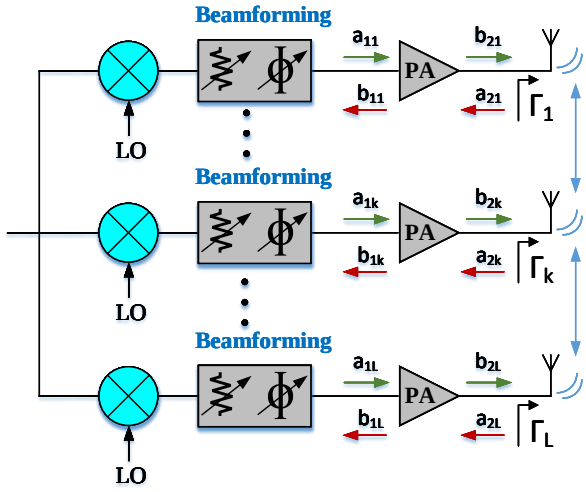


Fig. 4: Block diagram of a generic active antenna array including L antenna elements.

other antenna elements and as a result to the outputs of other PAs because of the electromagnetic coupling between the elements of the arrays. The steering angle and the radiated power level define the magnitude of the coupling. As illustrated in figure 4, $a_{2,k}$ can be described as following multiplication of s-parameter vectors [18]:

$$a_{2k} = \begin{bmatrix} s_{k1} & \dots & s_{kk} & \dots & s_{kL} \end{bmatrix} \cdot \begin{bmatrix} b_{21} & \dots & b_{2k} & \dots & b_{2L} \end{bmatrix}^T . \tag{1}$$

Reflection coefficient, Γ_k, for the kth element of the array is described as:

$$\Gamma_k = \frac{a_{2k}}{b_{2k}} = \frac{\sum_{i=1}^{L} S_{ki} b_{2i}}{b_{2k}}, \tag{2}$$

where the coefficients in the scattering matrix, S_{ki}, which are set by the characteristics of the antenna array, define the coupling between the k'th and i'th antennas. The elements in b_{2i} vector are complex coefficients for the input to the antenna. The reflection coefficient Γ_k, is a sum of the reflection from the k'th element and the coupling from all other elements. This inter-element coupling among antennas is called load modulation and has a major influence on the behavior of the active devices connected to the antenna elements. The impact of this load modulation on the digital pre-distortion of the active phased array is described in this thesis (papers A, B, C and E) through simulation and verification.

2.3 Feedback Receivers used for Training of DPD in Active Phased Array

For digital predistortion there is a need for a feedback receiver (FBR) to measure the nonlinear characteristic of the transmitter and use it for the training of the DPD. For an active phased array where the transmitter consists of several parallel branches, defining this feedback strategy is not straightforward. There are generally two methods for FBR construction; conductive or over-the-air (OTA). For the conductive methods, there have been several approaches proposed in recent years. Three major topologies proposed as conductive approaches, shown in Fig. 5 are:

- Partial feedback receiver
- Beam-oriented feedback receiver
- Time-shared feedback receiver

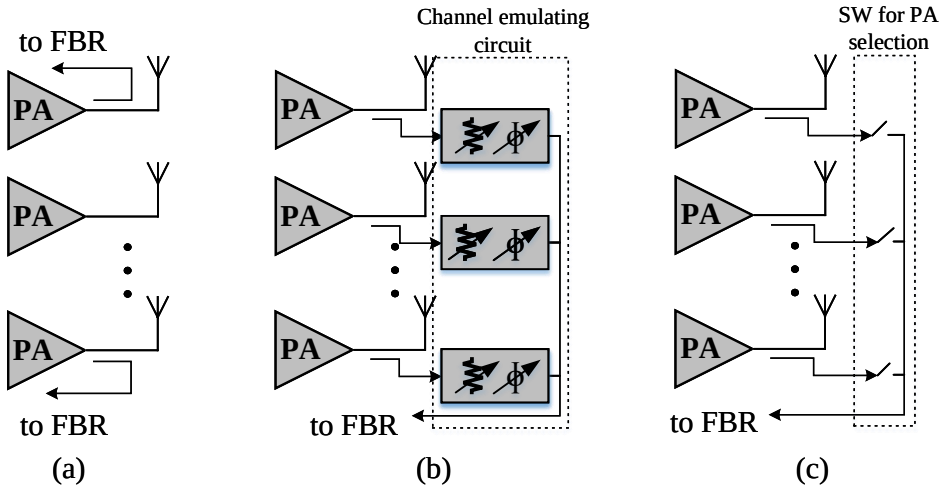


Fig. 5: Three major topologies proposed as conductive approaches; a: partial FBR, b: beam-oriented FBR, c: time-shared FBR.

2.3.1 Partial feedback receiver

In this approach, illustrated in Fig. 5a, the response of a few branches are used for the linearization of the whole array [19]. In this approach, the PAs in the array are assumed to behave similarly.

2.3.2 Beam-oriented feedback receivers

The beam-oriented approach, Fig. 5b is using a "virtual" main beam signal for linearization [20–22] and suggests a combined and sampled output of individual branches. Practically this approach emulates the channel by recovering its phase and amplitude. The draw-back is that besides the need for a symmetrically combining network, the FBR needs a separate path including phase shifter and control circuits which increase the complexity. This complexity could be relaxed by using the receiver path in time-division duplexing (TDD) system which is considered for the 5G FR2 system.

2.3.3 Time-shared feedback receiver

In this approach, Fig. 5c, instead of measuring the response of all PAs simultaneously, the measurement of each PA is time-shared. This approach has been proposed in [21, 23] and provides a model for each PA also while the response eventually is changed by the steering angle. Again the requirement for an additional receiver can be relaxed in TDD mode. This approach shows good linearity improvement but the need for the bulky feedback circuits makes it impractical to be used in the highly integrated active phased arrays structure.

2.3.4 Over-the-air single input single output feedback receiver

Due to issues with additional insertion losses and the limited possibility to access each PA's output in the array using conductive approaches, the OTA single input single output (SISO) FBR model, Fig. 6, has been proposed in [10, 24–28].

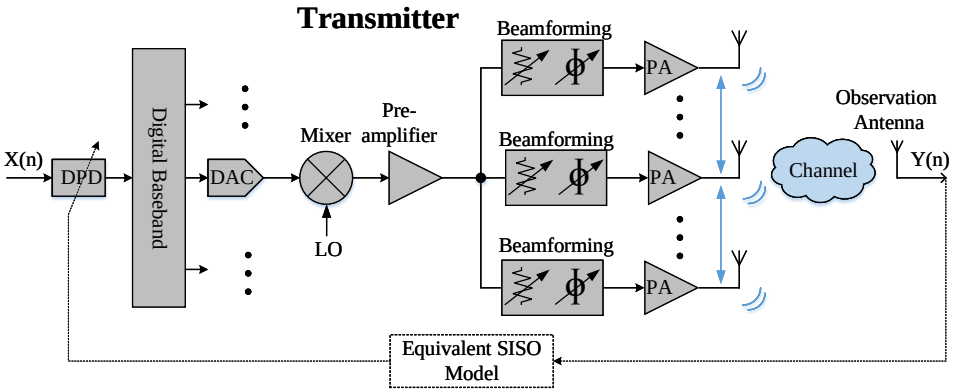


Fig. 6: Block diagram of equivalent SISO model for digital pre-distortion of hybrid beam-forming.

Generally OTA measurement and linearization has attracted a lot of interest in

academia and industry and mostly driven by 5G system. Despite remarkable progress, there are still challenges to be addressed such as antenna crosstalk, mutual coupling among antennas and power level variation in RF chains. In [29] the authors propose compensating for the potential differences among PAs. It allows linearization in all directions with a single DPD, in contrast to linearizing the main beam only. However, compensating for the mismatch requires analog circuits which introduce complexity and delay for large arrays and the potential changes in the PAs' behaviors due to crosstalk. In [18], a SISO OTA model has been introduced and a state-of-the-art analysis regarding linearization trade-offs in the OTA setup has been presented, which addresses most of the above-mentioned challenges. In [9] the effect of crosstalk between antennas has been analyzed by detailed simulation and measurements on a 4×4 active phased antenna array. A major practical challenge is the placement of the observation antenna. This could be the other user's device or as proposed for a heterogeneous network for an enhanced 5G system [30]. Besides the impact of multi-path and interference, since the reference antenna is placed at a certain steering angle and the actual beam is steered to a specific user then these two beams can interfere with each other. The observation antenna can be implemented as part of the receiver section of the same device (i.e. diversity receiver) which has been presented by [26]. This method uses an iterative procedure to eliminate the uncorrelated components from the feedback signal. There are some challenges related to this approach as it is practically a near-field measurement and the coupling between the patches of the diversity part which are not close to the transmitter patch could be very low. However, if the power level is high, the problem of low coupling can be ignored. These kinds of approaches using some "unused" part of the antenna construction during transmission is a promising approach and seem to be investigated in more detail by industry and academia in the coming years. An overview of the OTA linearization methods for 5G MIMO systems is introduced in [31]. In the present Ph.D. work, the SISO setup, shown in Fig. 6, has been used.

3 Measurement Techniques

This chapter explains the measurement setups and techniques used for the research topics in this thesis.

3.1 Over-The-Air Measurement Setup

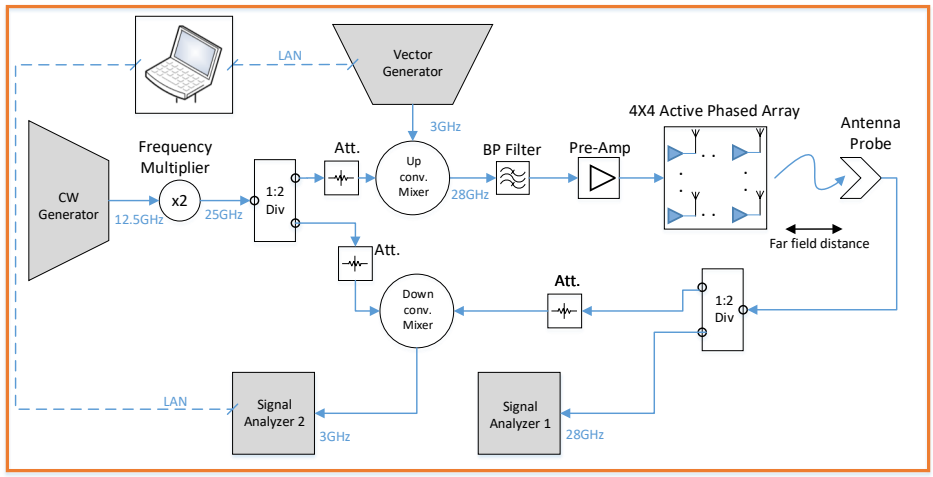


Fig. 7: The block diagram of the measurement setup.

The block diagram in Fig. 7 illustrates the measurement setup used for several investigations in this thesis. A 3GPP downlink OFDM modulated waveform with 64-QAM sub-carrier modulation, sub-carrier spacing of 60 kHz, 1584 active sub-carriers and peak to average power ratio (PAPR) of 11.6 dB has been generated using the R&S SMBV100B Vector Signal Generator and centered at 3 GHz. Bandwidths are up to 100 MHz and with the sampling rate of up to 600 MHz. A continued wave signal of 12.5 GHz generated by the Agilent E3247C, frequency-doubled to 25 GHz by the MITEQ-MAX2M200400, and used as the local oscillator (LO). Two active mixers, KTX321840 and KRX321840, are used for up-conversion to 28 GHz and down-conversion to 3 GHz. The LO leakage is attenuated using a band-pass filter and a pre-amplifier is used for delivering sufficient input power to the Anokiwave AWMF-0158 [32] active phased array (APA). The APA integrates 16 branches of attenuators, phase shifters and PAs together with 16 patch antennas in a 4×4 configuration. The over-the-air (OTA) modulated signal is captured using an observation antenna placed in the far-field. The received signal after down-conversion to 3 GHz, is converted to baseband using the signal analyzer.

For uploading the baseband I and Q samples to generator and capturing them by the R&S FSW signal and spectrum analyzer, a host PC running in Matlab is used. The main beam of the array is controlled by using the appropriate code-book and software tools. The system is calibrated and all devices are in their highly linear region. The only source of nonlinearity is related to the APA as the device-under-test. Fig. 8 shows the measurement setup in laboratory.

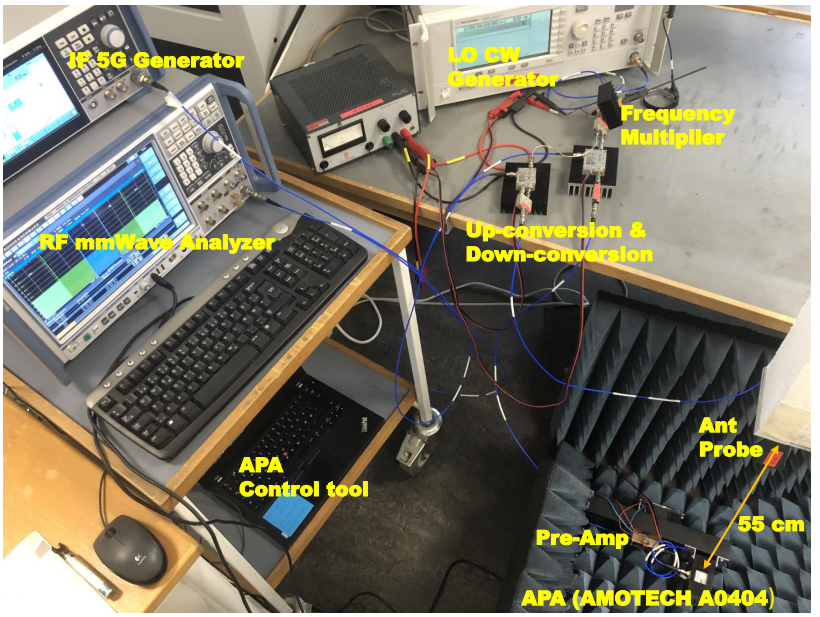


Fig. 8: The measurement setup.

3.2 OTA Measurement using Compact Antenna Test Range

For the 5G testing at mmWave frequencies, the compact antenna test range (CATR) OTA measurement method has been identified as the best approach. The idea is that during the radio transmission and reception measurements, the waves illuminated at the receiver antenna are desired to be uniform plane waves. CARTs can generate nearly planar waves at a shorter distance compared to the far-field requirements of the wave. Fig. 9 illustrates such CART chamber [33]. The performance of the measurement in CATR is limited by several parameters such as aperture blockage, direct radiation from the source to the test antenna, diffractions from the edges of the reflector and feed support, depolarization coupling between the two antennas and wall reflections [34]. However, the actual designed CATR chamber used for this work is specified for high

accuracy measurements in the frequency range from 6 GHz to 40 GHz [33]. The block diagram of the OTA measurement using the CATR chamber is shown in Fig. 10.

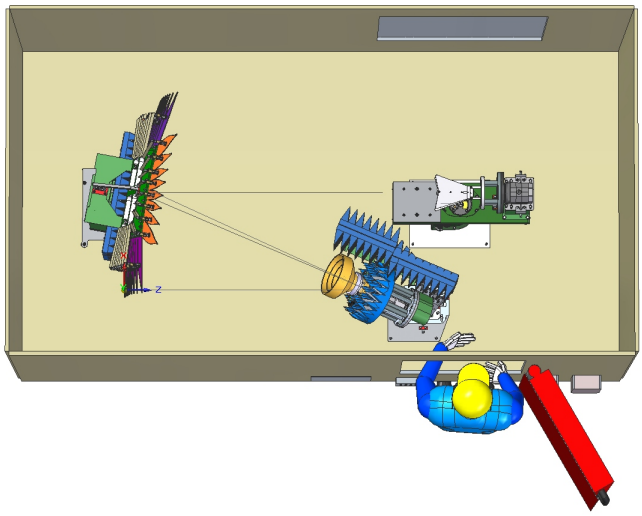


Fig. 9: Illustration of a CATR measurement chamber [33].

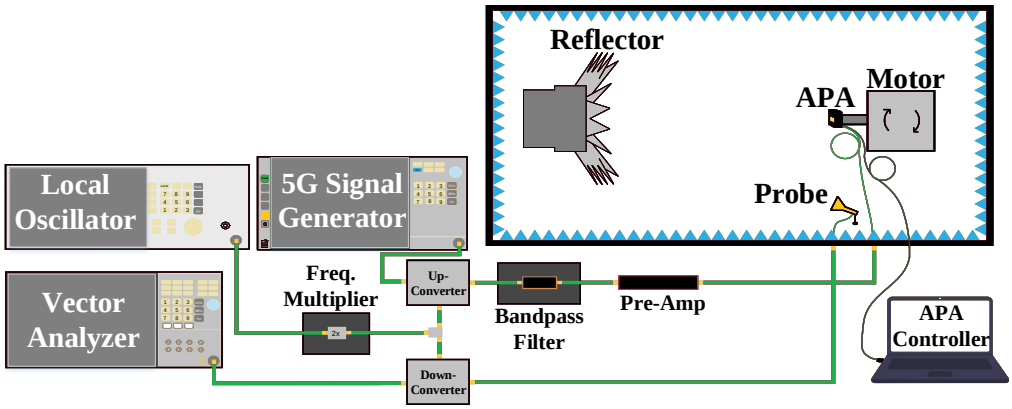


Fig. 10: The block diagram of the OTA measurement using CATR chamber.

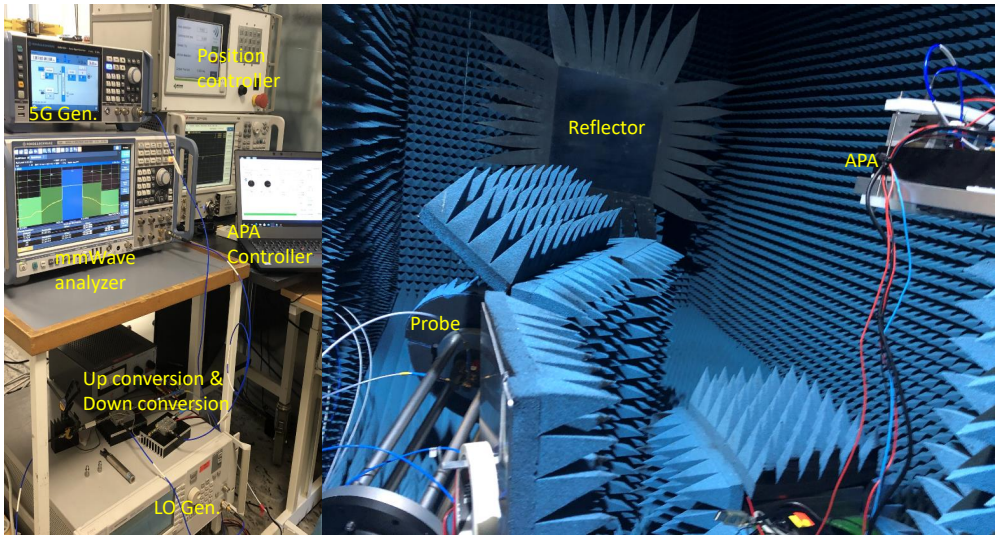


Fig. 11: The OTA measurement setup using CATR chamber.

3.3 Total Radiated Power Measurement

In the 5th generation frequency rang 2 transmitter specifications, 3GPP has defined the adjacent channel leakage ratio (ACLR) specifications based on OTA measurement of the total radiated power (TRP) of the transmitter. This means that the system does not depend on the spatial direction of the ACLR but integrates the transmitted power over the angular domains. For a beam steered transmitted signal, the worst case of nonlinear distortion is in the directions where the intended users are positioned [13]. Therefore quantitative measurements in this work have been done to evaluate if using main beam ACLR is a valid method for characterizing the linearization performance of a beam steerable array. For doing this investigation TRP ACLR measurements have been performed and compared with main-beam-only measurements. In [35] the estimated TRP for a discrete set of measured directions is described as:

$$\mathrm{TRP} = \frac{1}{4\pi} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \mathrm{EIRP}(\phi, \theta) \cdot \sin\theta \, d\theta \, d\phi \tag{3}$$

For $\Delta\theta = \frac{\pi}{N_e}$ and $\Delta\phi = \frac{2\pi}{M_a}$ then:

$$\frac{1}{4\pi} \cdot \frac{\pi}{N_e} \cdot \frac{2\pi}{M_a} = \frac{\pi}{2N_e M_a}, \tag{4}$$

where the number of azimuth angles (ϕ_n) and elevation angles (θ_n) are defined as N_a and M_e respectively. The estimated TRP can be calculated as:

$$\mathrm{TRP}_{\mathrm{Estimate}} = \frac{\pi}{2N_eM_a} \sum_{n=0}^{N_e-1} \sum_{m=0}^{M_a-1} \mathrm{EIRP}(\phi_n, \theta_m) \cdot \sin\,\theta_m \tag{5}$$

The radiated power in each angular case as a sum over both linear polarizations is defined as $\mathrm{EIRP}(\phi_n, \theta_m)$. The TRP-ACLR in linear scale is calculated as the total radiated power of the adjacent channel divided by the total radiated power of the in-band channel:

$$\begin{aligned} \text{TRP-ACLR} &= \frac{\frac{\pi}{2N_eM_a} \sum_{n=0}^{N_e-1} \sum_{m=0}^{M_a-1} \mathrm{P}_{adj.ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m}{\frac{\pi}{2N_eM_a} \sum_{n=0}^{N_e-1} \sum_{m=0}^{M_a-1} \mathrm{P}_{ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m} \\ &= \frac{\sum_{n=0}^{N_e-1} \sum_{m=0}^{M_a-1} \mathrm{P}_{adj.ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m}{\sum_{n=0}^{N_e-1} \sum_{m=0}^{M_a-1} \mathrm{P}_{ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m} \end{aligned} \tag{6}$$

The block diagram and lab setup for the measurements are shown in Fig. 10 and Fig. 11. The following procedure is applied for all specific azimuth angles, ϕ, and elevation angels, θ:

- The APA is placed at the positioner and the coordinate system is aligned.
- The main beam of the APA is adjusted to the direction of the observation antenna (the reflector).
- The main and adjacent channels' power are measured.
- The measurement is repeated for all ϕ and θ angles.
- The TRP-ACLR is now calculated according to (6).

For the applied TRP-ACLR measurement in this work, the steering angle of the APA is set to 0 degrees. The position of APA has been adjusted by θ from 0 to 180 degrees in steps of 10 degrees and for each step, the ϕ angle is changed from -90 degrees to 90 degrees in 20 degrees steps. For each angular setting, the TRP-ACLR is measured using Eq. (6). This procedure determines the radiated power in the front hemisphere and it is assumed that the radiated power in the back hemisphere is negligible.

4 Behavioral Models for 5G Transmitters

State of the art power amplifier (PA) modeling can generally be divided into two major groups, physical models and behavioral models [36]. They can be distinguished by the type of data needed for their extraction. Making a physical model requires a nonlinear model of the active device, a nonlinear model of the passive components and the nonlinear equations relating to terminal voltages and currents. These models are then applied to the circuit-level simulation and the accuracy of the results is limited by the quality of the used device models and the accuracy of the simulator.

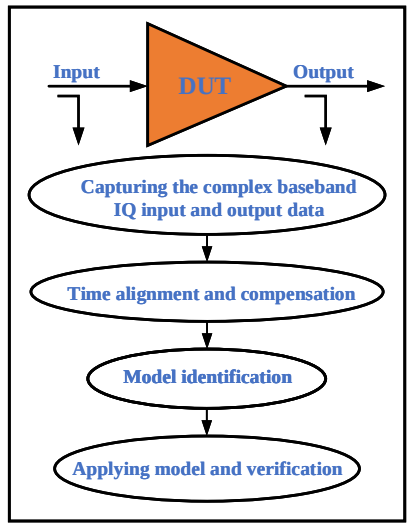


Fig. 12: Procedure for behavioral model extraction; from measurements to model validation.

4.1 Behavioral Modeling Theory

The behavioral model considers PA as a black box and there is no need for a-priori knowledge about the PA's physical model, nor the circuit topology. The behavioral model is based on experimentally measured input and output of the PA and with a lower computational cost compared to the physical model. This approach is a appropriate choice for system-level simulation and the design of an algorithm for linearization and optimization. It is especially suitable for the cases where the nonlinear part is not a single nonlinear device but a set of nonlinear devices integrated into the system such as an active phased array which is the scope of work in this thesis. The behavioral model extraction procedure is illustrated in Fig. 12 where the key steps from measurements

to model validation are shown.

The relation between the input and output signals in discrete time can be modelled by the classical approach of modelling. There are several structures to model and emulate the memory effect of RF power amplifiers. Some of the most used models are Hammerstein model [37], Wiener model [38], memory polynomial model, generalized memory polynomial model, Volterra series model and artificial neural networks model.

4.2 Polynomial Models

4.2.1 Volterra series model

The block diagram of the Volterra series model is shown in Fig. 13.

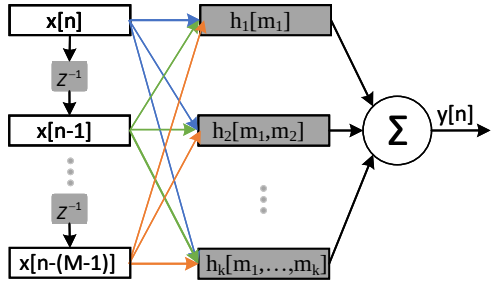


Fig. 13: The block diagram of the Volterra model.

This model is the classical approach of modelling the full behaviour of a nonlinear device [39–41] and is represented by Eq. (7).

$$
\begin{aligned}
y[n] &= \sum_{k=1}^{K} \sum_{m_1=0}^{M-1} \cdots \sum_{m_k=0}^{M-1} h_k[m_1,\ldots,m_k] \prod_{j=1}^{k} x[n-m_j] \\
&= \sum_{m_1=0}^{M-1} h_1[m_1]x[n-m_1] \\
&+ \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} h_2[m_1,m_2]x[n-m_1]x[n-m_2] \\
&+ \ldots \\
&+ \sum_{m_1=0}^{M-1} \cdots \sum_{m_K=0}^{M-1} h_K[m_1,\ldots,m_K] \prod_{j=1}^{K} x[n-m_j], \qquad (7)
\end{aligned}
$$

where K and M are the order of the nonlinearity and the memory depth respectively and $h_k(m_1, \ldots, m_k)$ are the parameters of the model, which are often referred to as the "Volterra kernels" in literature. At each of the k^{th} Volterra kernel, the n^{th} sample of the input signal $x[n]$ is mixed with the $M-1$ preceding samples. It means that the k^{th} kernel includes all possible combinations of k time shifts of the input signal, which includes all types of memory effects. Therefore, the Volterra series is considered the most complete model. The draw-back is that the computational complexity of the model is very high [42].

4.2.2 Memory polynomial model

Memory polynomial model (MPM) has been widely used in the last decade for PA modeling due to the easiness in parameters extraction [43–47]. The system representation is illustrated in Fig. 14 and represented by following equation:

$$y[n] = \sum_{m=0}^{M} \sum_{k=1}^{K} a_{mk} x[n-m] \cdot |x[n-m]|^{k-1}, \tag{8}$$

where a_{mk} are the coefficients that modelings the PA.

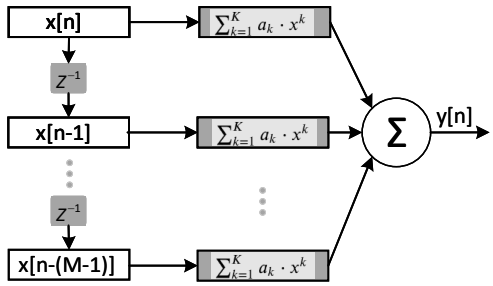


Fig. 14: The block diagram of the memory polynomial model.

As can be immediately seen from Fig. 14 the memory polynomial contains quite fewer terms than the Volterra series, that is because the memory polynomial excludes cross-terms.

4.2.3 Generalized memory polynomial model

For the cases with deep memory effect, the generalized memory polynomial model (GMPM) is more suitable [48], block diagram of this model is shown in Fig. 15. The model consists of three components: the regular MP, a backward MP term and a forward MP term. The idea is that the last two additional terms can enhance the model

with additional degrees of freedom and capabilities, defining a new structure able to model memory effects situated around time constants that are located in a number of past and future locations:

$$\begin{aligned} y[n] = & \sum_{m=0}^{M_a} \sum_{k=1}^{K_a} a_{mk} x[n-m] \cdot |x[n-m]|^{k-1} \\ & + \sum_{m=0}^{M_b} \sum_{k=2}^{K_b} \sum_{p=0}^{P} a_{mkp} x[n-m] \cdot |x[n-m-p]|^{k-1} \\ & + \sum_{m=0}^{M_c} \sum_{k=2}^{K_c} \sum_{q=0}^{Q} a_{mkq} x[n-m] \cdot |x[n-m+q]|^{k-1}, \end{aligned} \tag{9}$$

where M_a, K_a are the number of coefficients for the regular MP and M_b, K_b, P are the number of coefficients related to the backward MP and M_c, K_c, Q are the number of coefficients related to the forward MP.

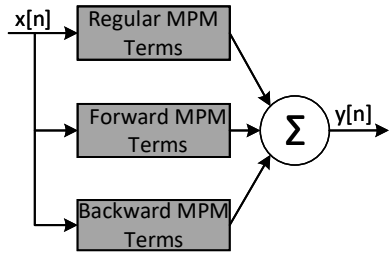


Fig. 15: The block diagram of the generalized memory polynomial model.

4.3 Artificial Neural Network Model

Challenges such as high bandwidth and high nonlinearity lead to huge complexity and the explosion of MPM-based algorithms. With the increased complexity and consequently increase of the number of unknown kernel coefficients in the model, the MPM approach is less practical [49]. Artificial neural networks (ANNs) are well known to be able to learn any arbitrary nonlinear function according to the universal approximation theorem [50]. Inspiration of ANN origins from the behavior of neurons in the human brain. By training these neurons the structure of data can be recognized or a complex nonlinear function can be modeled. Fig. 16 illustrates the common block diagram of such a model.

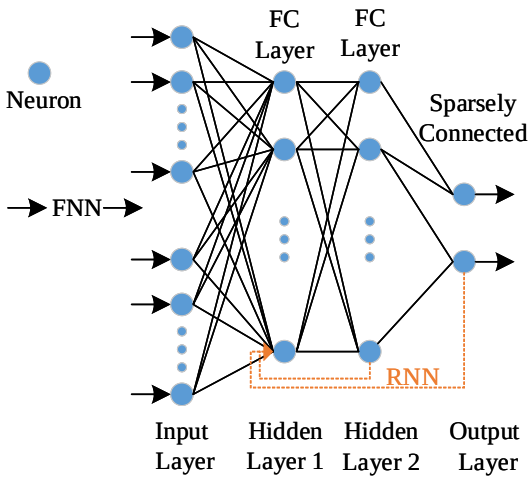


Fig. 16: Block diagram of a general neural network model.

ANN has a layered sequential computation structure that by training can recognize patterns in data and predict outcomes. Between the input and output layers of an ANN are the hidden layers. The first part of a hidden layer is the fully connected (FC) layer, also known as a dense layer. An FC structure in a densely populated ANN may increase requirements for hardware resources, but in many applications, the weight of some interconnections can be set to zero without loss of accuracy, which results in sparsely connected layers [51]. The sparse structure is out of the scope of this work. As illustrated in Fig. 16, there are two common strictures of the network, Feedforward networks (FNNs) and recurrent (or feedback) networks (RNNs). These networks can be distinguished by the interconnection pattern or architecture [51]. The feedforward network is considered for this work since it is the most used ANN and according to the universal approximation theorem, it can approximate any nonlinear function with any desired error [52].

4.3.1 Single-layer perceptron model

The fundamental building block of ANN is based on the single-layer perceptron (SLP) model. Fig. 17 shows the block diagram of an SLP model. A parallel combination of several SLP makes the architecture of a multi-layer perceptron (MLP) which presents the general structure of the ANN. In SLP, x_i is the input to the j^{th} neuron from layer k, and each x_i is then multiplied by the weight w_{ji} and the results are added together. A bias weight, b_j, is added and the output of j^{th} neuron is calculated by multiplying an activation function, $\sigma(.)$, as shown in Eq. (10).

$$y_j{}^{(k)}(n) = \sigma^{(k)} \left(\sum_{n=1}^{N} w_{ji} x_i{}^{(k-1)} + b_j{}^{(k)} \right), \tag{10}$$

where y_i is the output from the j^{th} neuron k^{th} layer.

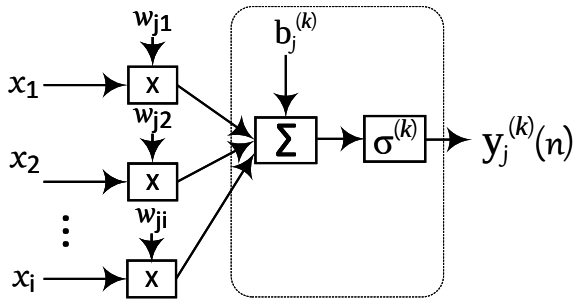


Fig. 17: Block diagram of single layer preceptron (j^{th} neuron at k^{th} layer).

For linearization of a specific nonlinear circuit, different configurations for ANN architecture, learning algorithm, or activation function could be used and there is not a universal recipe. Mainly trial and error are often used for finding the optimum configuration. However, knowing the physical behavior of the circuit and its impairments will help in defining and optimizing the appropriate model for the circuit.

4.3.2 Architecture of artificial neural network

Generally an ANN is constructed by mapping an input function to an output function and by choosing the proper structure and parameters of the ANN, then it is possible to fit any arbitrary nonlinear functions using regression. For enabling the dynamic nonlinear system identification, and solving the time series prediction, a time-delayed neural network (TDNN) which includes tapped delay line has been introduced in [51, 53]. For extracting the amplitude and phase information of the modulated signal, the input IQ data should be represented either in complex-valued (CV) form or in real-valued (RV). In the case of RV, a pair of the input and a pair of the output represent the real and imaginary parts, and a weight and activation function is separately defined for the real and imaginary parts. Several state-of-the-art linearization techniques based on RVTDNNs have recently been introduced in the literature. In [54, 55], the authors propose a solution for performance imperfections such as crosstalk, power amplifier nonlinearities along with modulator imperfections like in-phase and quadrature (I/Q)

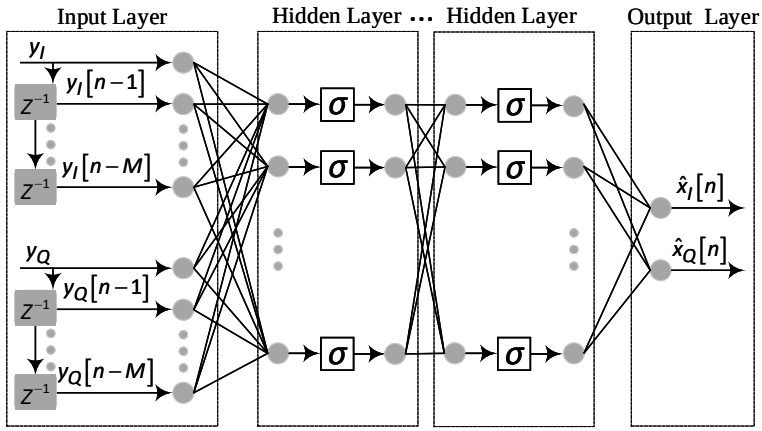


Fig. 18: Multilayer RVTDNN neural network architecture.

imbalance and DC-offset for a wide bandwidth direct-conversion transmitter. Another similar approach, where only the magnitude of the input signal encounters a nonlinear operation, and the phase information is recovered with a linear weighting operation has been introduced in [56]. The RVTDNN structure offers superior performance and easy baseband implementation when used for inverse modeling of PAs with strong nonlinearities and memory effects [57]. Fig. 18 represents such a structure and has been used for the investigations in this thesis. The structure of the ANN is based on real-valued while the training data are complex-values, therefore the number of neurons at input and output of the architecture are multiplied by two. The input and output data are separated as $y_I[n-M]$, $y_Q[n-M]$, $\hat{x}_I[n]$ and $\hat{x}_Q[n]$ where n is the number of the I and Q data used in the training. The memory effects are modeled by the delayed replica up to the memory depth of M. The mathematical way of describing the function of the dense layers is that each output of the k^{th} layer, $y^{(k)}(n)$ is a weighted sum of all inputs, $x^{(k-1)}$, multiplied by activation function:

$$\mathbf{y}^{(k)}(n) = \sigma^{(k)}\left(\mathbf{W}\mathbf{x}^{(k-1)} + \mathbf{B}\right), \tag{11}$$

where $\sigma^{(k)}(\cdot)$ is the activation function, $\mathbf{y}^{(k)} \in \mathbb{R}^{L\times 1}$ are the output values of the M'th neuron, $\mathbf{W} \in \mathbb{R}^{L\times M}$ are trainable coefficients, $\mathbf{x}^{(k-1)} \in \mathbb{R}^{M\times 1}$ are the outputs of the previous layers and $\mathbf{B} \in \mathbb{R}^{L\times 1}$ are trainable biases. Thus, M is the number of outputs of the previous layer, and L is the number of inputs to the next layer. Eq. (11) describes the operations in each dense layer. The optimization algorithm used in this work is the adaptive moment estimator, Adam [58]. This algorithm is based on a gradient descent

which gets more computationally efficient by using momentum and randomized batches to avoid local minima. A batch is defined as the number of training samples used for estimating the error gradient. The weights and biases of each FC layer are distinctive and are optimized by back-propagation for improving the accuracy of predictions.

4.3.3 Activation function

The activation function makes the ANN able to fit arbitrary nonlinear functions. The activation function used in this thesis is the rectified linear units (ReLU) [59]. This function is considered to be less computationally expensive than hyperbolic tangent (Tanh) [60] and Sigmoid [61] and involves simpler mathematical operations [62, 63]. The ReLU activation function is defined by:

$$\sigma_{\mathrm{ReLU}}(x) = \max(0, x) \tag{12}$$

According to Eq. (12), ReLU introduces nonlinearity by setting negative inputs to 0, which also adds sparsity to the ANN and simplifies the computations [62].

4.3.4 Loss function

The difference between the predicted values and reference values is evaluated by using the loss function. Different type of loss function has been used during the investigation in this thesis such as Huber loss and mean square error (MSE) loss. Huber loss function is defined as:

$$\mathrm{loss}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{B} \sum_{i=1}^{B} \varepsilon_i, \tag{13}$$

where B is the batch size and ε_i as a combined set of square and absolute errors:

$$\varepsilon_i = \begin{cases} 0.5(x_i - \hat{x}_i)^2, & \text{if} \quad |x_i - \hat{x}_i| < \delta, \\ \delta|x_i - \hat{x}_i| - \frac{1}{2}\delta^2, & \text{otherwise.} \end{cases} \tag{14}$$

where Huber loss function is defined by setting δ to 1. MSE loss function is defined as:

$$\mathrm{loss}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{B} \sum_{i=1}^{B} (x_i - \hat{x}_i)^2. \tag{15}$$

Huber loss function uses a kind of smooth loss function for regression tasks and therefor it is robust against large errors [64].

5 Digital Predistortion using Artificial Neural Network

Digital predistortion (DPD) is a system-level linearizer used for extending the linear range of PAs. However, the resulting power efficiency achieved with linearization techniques applied to PAs operating close to the nonlinear region for the systems with high peak to average ratio (PAPR) is limited. For improving the efficiency, the operation condition of the PA can be forced to follow the signal's envelope by using envelope tracking (ET) [65, 66]. Another approach is using switch-mode amplifying classes in Doherty PAs [67]. High efficient switch-mode PAs are also used in linearization techniques using nonlinear components (LINC). The idea of LINC is to convert the amplitude modulation of the input signal into phase modulation so the efficient switch-mode PAs can be used [68]. Considering the 5G system with wide bandwidth, beam steering and carrier aggregation, in highly efficient techniques such as ET, Doherty and LINC, DPD facing challenges to compensate for both static nonlinearities and dynamic memory effects because of the increased number of parameters required in the DPD model. As a consequence, the model identification and adaption process invoke high computational complexity which may result to overfitting and unreliability [69]. ANN-based DPD can offer robust solutions compared to conventional approaches due to their powerful nonlinear mapping. Some comparison to the conventional approach based on use cases such as highly nonlinear systems, beam-dependent load modulation and reduction of computational complexity for bandwidth scaleable DPD has been introduced in this work.

5.1 Digital Predistortion Theory

Digital predistortion (DPD) consists of identifying the behavioral model of the PA and applying an inverse model of that to the input with a result of canceling distortion at the PA's output. The concept is illustrated in Fig. 19. The linearized output signal is obtained by multiplying the predistorter to the input signal of the PA. It means that once the relationship between input and output is known, then the coefficients used for predistortion can be found using the behavioral model. Theoretically, these coefficients which describe the characteristic of the PA need be calculated only once. But the PA behavior will change by parameters such as temperature drift, aging, power supply variations, etc. [70]. Therefore for an adaptive system, the coefficients for the predistorter need to be updated frequently at a defined time interval.

5.2 DPD Identification Techniques

DPD identification is based on a set of input and output baseband measurements of the device-under-test (DUT), which is performed by using a measurement setup based on

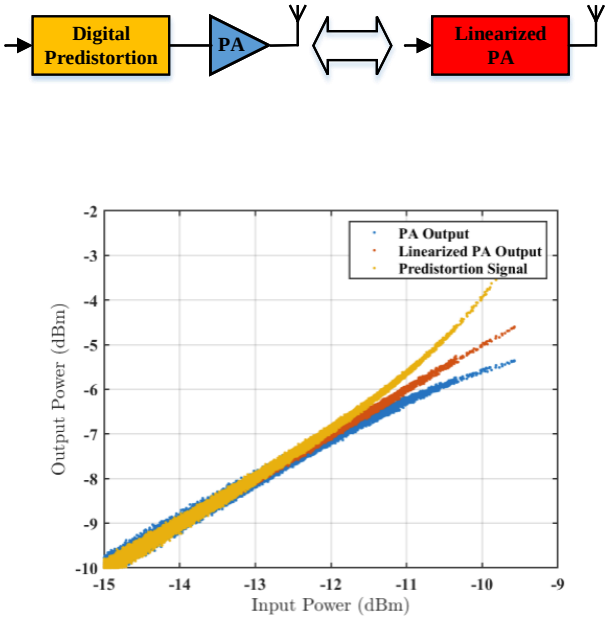


Fig. 19: Representation of digital predistortion concept.

a network analyzer or signal generator and signal analyzer. The procedure is based on criteria of minimization of an error defined by the previously chosen goal function. DPD is implemented by continuously updating the model coefficients of the predistorted signal in the feedback loop of the model. The model coefficients can be estimated either offline, by using a set of input and output samples of the PA, or online in an adaptive real-time measurement of the PA's input and output [71, 72]. Generally, there are two learning architectures for estimating the model coefficients, indirect learning architecture (ILA) and direct learning architecture (DLA).

5.2.1 Indirect learning architecture

The most popular identification algorithm in DPD is the indirect learning architecture (ILA) which has been implemented in many state-of-the-art applications [73–76]. Fig. 20 shows the block diagram of this architecture. In this architecture, the inverse behavior of the PA is directly calculated and used as a predistorter and applied to the PA. This is also called post-inverse, and can be performed either in an offline mode using least squares (LS) optimizer [47] or can be performed in online mode in an adaptive way using a least mean square (LMS) algorithm [77]. The most important advantages of IDL are the simplicity and a low computational complexity [78]. However, there are two major

problems with the IDL architecture. Firstly, in case of the noise in the measured PA output, the predistorter may converge to a biased solution. Secondly, in case of high saturation, the efficiency of the ILA is limited [79].

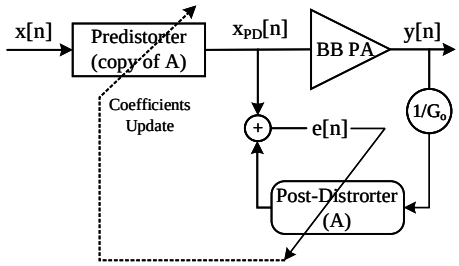


Fig. 20: Block diagram representing indirect learning architecture.

5.2.2 Direct learning architecture

In the direct learning architecture (DLA), Fig. 21, a direct model of the PA is calculated, inverted and used as a predistorter. For the online adaptive mode scenarios, the error produced in the PA output is minimized by using adaptive algorithms such as LMS [72]. An issue with this approach is that the DPD identification, a direct model of the PA, requires more iterations and consequently a higher computational complexity. But compared to the ILA, the DLA can achieve higher accuracy [73]. In offline mode, a set of the input and output of the PA together with the LS algorithm is used to identify the predistorter. This approach used as a look-up table (LUT) based linearization, can significantly reduce the computational time. However, for a system with high complexity, this approach requires a large lookup table [80].

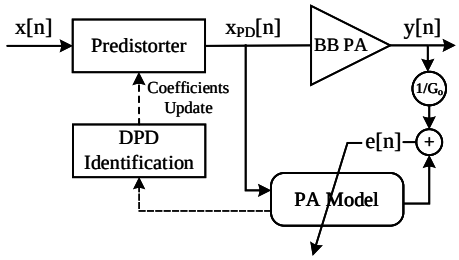


Fig. 21: Block diagram representing direct learning architecture.

5.3 Digital Predistortion using Artificial Neural Networks

Due to their strong adaptive nature and approximation capability, ANNs are very attractive for the behavioral modeling of PAs. Through the work in this thesis the real-value time-delay neural networks (RVTDNN) model has been used due to its superior performance and easy baseband implementation [54]. Fig. 22 illustrates the system-level ANN-based DPD arcitecture.

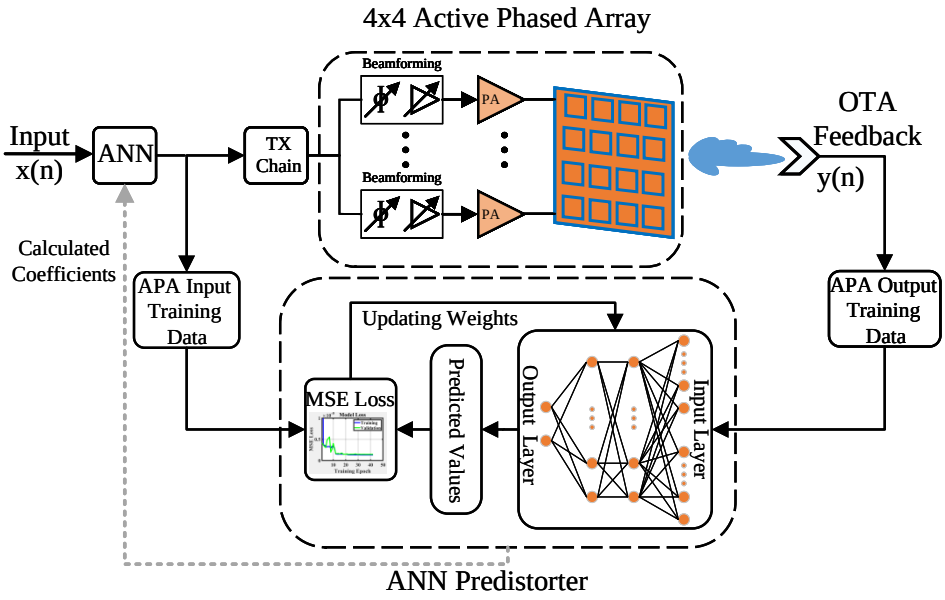


Fig. 22: The system-level ANN-based DPD concept.

ANN parameters are updated gradually for reducing the losses between the predicted values and the reference input. The ANN learns the features of the training data as a regression model which is generally used for the scenarios where the ANN needs to learn the characteristic nonlinear distortions.
Python 3.8.4 on Visual Studio Code is used for obtaining all operations. Keras 2.3.0-tf, version 2.2.0 is used for building and training the ANN.

5.3.1 Training and optimization procedure

For each power level, ANN is trained to reach the minimum mean square error (MSE) level defined by the cost function while varying the number of the time-delays and with a fixed number of neurons. Then the number of neurons is tuned for the best linearization

performance. Finally, the number of the hidden layers has been tuned keeping the other parameters unchanged. There is a trade-off between the number of delay lines, the number of neurons and the number of hidden layers dependent on the target complexity and linearization quality [81]. Fig. 23 shows a typical optimization result using four hidden layers, where the optimal choice is a trade-off between the ACLR, the EVM and the number of multiplications.

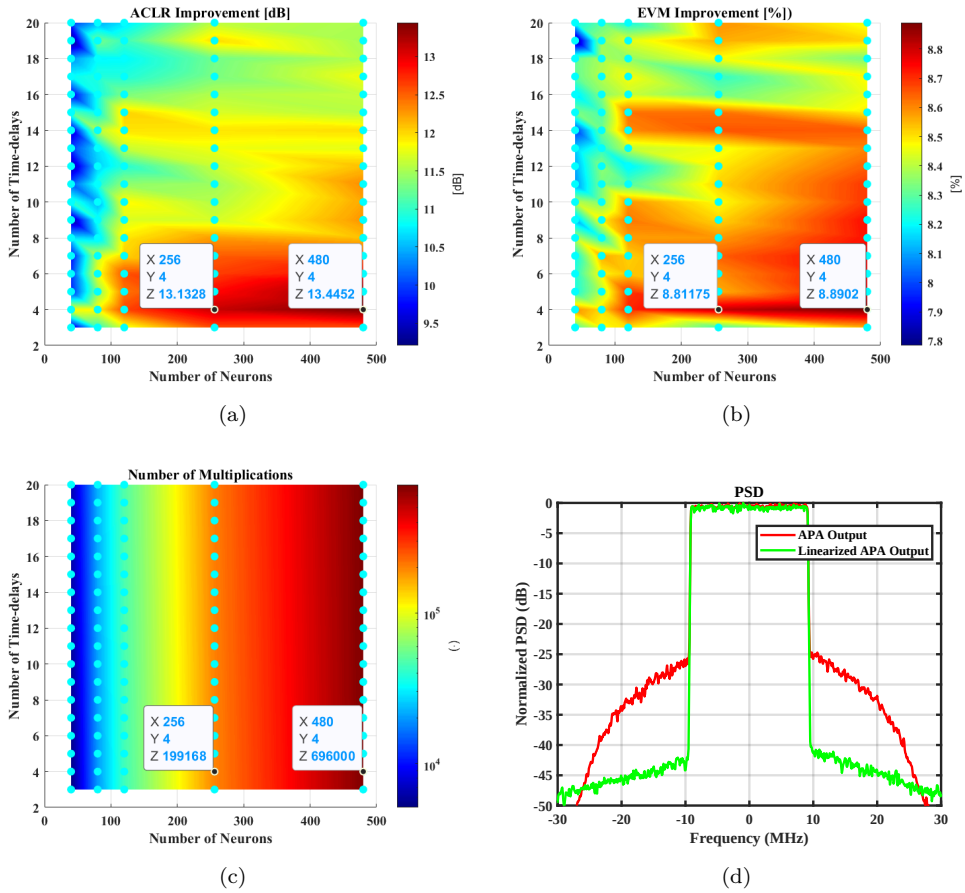


Fig. 23: Optimization parameters for a general model; (a): The ACLR values; (b): The EVM values; (c): The number of multiplications required to apply the corresponding ANN DPD model; (d): The PSD result.

The measurements results of this experiment show that by keeping the number of

time-delays to 4 and the number of neurons to 256, it is possible to achieve an ACLR improvement of 13.1 dB and EVM improvement of 8.8 % while keeping the number of multiplications as low as possible, i.e. 199 k. Fig. 23a, 23b and 23c show the out-of-band distortion in form of ACLR, in-band distortion in form of EVM and complexity of the experiment in form of the number of multiplications, respectively, for the optimized model. Increasing the number of neurons to a higher number e.g. 480 will give negligible improvement of ACLR and EVM while the number of the multiplications which set the complexity level, will increase drastically to 696 k. The power spectral density (PSD) plot in Fig. 23d shows the achieved out-of-band improvement in form of ACLR. The detailed results can be found in "paper G" attached to this thesis.

5.4 Measurement Results

This section presents some measurement results that summarize some of the most important results obtained within this Ph.D. project. It gives a robustness assessment of ANN-based DPD concerning beam angle, highly compressed active device and signal bandwidth. For more detailed explanations please refer to the research papers attached at the end of this thesis.

5.4.1 Beam-dependent load modulation

Due to mutual coupling and cross-talk between antennas of the APA, there could be a distortion in form of beam-dependent load modulation at PAs outputs. The experimental results show that re-using the trained coefficients of one angle for other angles will degrade the linearization performance. Fig. 24 shows the experimental results of the impact of load modulation on DPD based on MPM and ANN linearization approaches. In this experiment the trained model for 0 degrees has been used for the entire beam steering range, i.e. -78 degrees to +78 degrees. ANN-based linearization shows only slightly better performance compared to the MPM-based. For both models the best performance is achieved when the steering angle is close to the training angle. The step size of the angle for the re-training depends on the target linearity requirements and the actual set-up. The detailed results can be found in papers B, C and E attached to this thesis. In [82] a combined ANN digital predistortion method where the input and output data of several steering angles are combined and used for training of the predistorter has been proposed. The proposed ANN DPD provides consistently good linearization for the entire range of steering angles considered. This approach does not provide the same good performance compared to the case where a single-beam training model is used at the trained angle. However, it demonstrates the strength of ANN compared to MPM and it is worth investigating it further in future work.

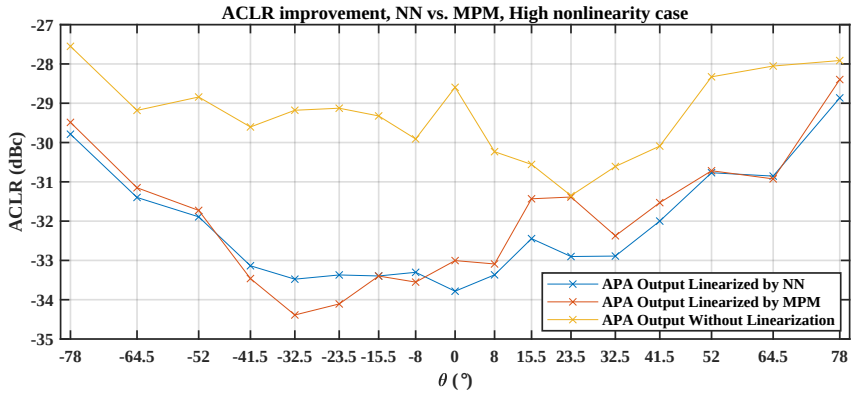


Fig. 24: Experimental result of the impact of load modulation (trained at 0 degree) [83].

5.4.2 Comparison between ANN and MPM digital predistortion for highly nonlinear active phased array

A 100 MHz wide orthogonal frequency-division multiplexing (OFDM) signal with a peak-to-average power ratio (PAPR) of 11.6 dB has been used as input to the APA in this experiment. For the ANN parameters, one hidden layer, five time-delays and 100 neurons have been chosen. For MPM a nonlinearity order of 5 and memory depth of 8 has been chosen as optimal parameters. Both linearization models get almost the same level of complexity in terms of the number of multiplications (1200 for the ANN and 828 for the MPM) by using these configurations. The APA is driven into two different nonlinear regions with the main-beam powers at 34 dBm and 31 dBm. Fig. 25 shows the experimental results for both models. The ANN approach gives a much better linearization result compared to the MPM approach for the high nonlinearity case whereas in the case of low nonlinearity they perform almost equally. A further investigation based on the predistorter signal's peak to average power ratio (PAPR) for both models, shows that for the MPM case the PAPR of the pre-distorted signal reaches extremely high values. On the contrary, the proposed ANN model does not contain inherently such an "explosion" effect for high non-linearities. The polynomial model has inherent local approximating properties where the ANN can behave as global approximation which is a potential advantage when modeling strong nonlinear systems. These results confirms that ANN principally works better than the classical MPM approach to extrapolate beyond the utilized zone of parameter extraction [84]. The detailed results of this experiment are in paper E attached to this thesis.

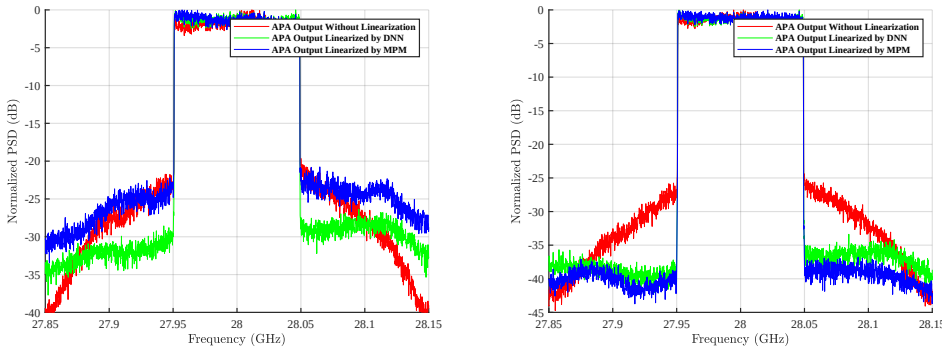


Fig. 25: linearization results while the APA is driven in two different nonlinearity cases; (a): Main beam power = +34 dBm; (b): Main beam power = +31 dBm [83].

5.4.3 Bandwidth-scalable DPD using transfer learning

5G mobile communication has introduced a very broad bandwidth, up to 400 MHz, with a dynamic bandwidth change during transmission and reception. By increasing the bandwidth the digital predistortion algorithm needs to increase the number of coefficients used in the identification algorithm. For the ANN case, it means an increase in the input dimensions of the model and number of layers and an increase in the number of neurons [85]. Existing works show that an increase in the number of the tapped element (memory depth) improves the performance [81]. Also in the same reference, it has been shown that keeping the number of neurons unchanged and then increasing the number of the hidden layers from one to three will also improve the linearity performance. Increasing the ANN parameters results in an increased training time and complexity. The transfer learning neural network (TLNN) approach can apply knowledge learned for linearization of the previous operating bandwidth into the new operating bandwidth and reduce the training time and complexity. The study of TL is motivated by the fact that one can intelligently apply knowledge learned previously to solve new problems faster or with better solutions [86, 87]. In this concept, the knowledge obtained from a trained model with a large dataset is used in a second model to obtain a faster learning time while maintaining high performance. Fig. 26 shows a block diagram illustrating the concept. The transfer learning approach has been used on bandwidth-scalable DPD in one of the experiments of this work. Part of the trained model of 20 MHz bandwidth has been frozen and reused for linearization of 100 MHz BW signal. Fig. 27 shows power density measurement results before and after linearization using regular ANN and TL-ANN. The detailed linearization results in terms of EVM and ACLR are listed

in Table 1. The proposed TL-ANN can achieve similar linearization performance with much lower computational complexity. By re-using the frozen sub-model it is possible to reduce the number of hidden layers from 4 to one and the number of neurons from 256 to 16. This will result in the relaxation of the computational complexity in terms of multiplications from 69120 to 224. The detailed results of this experiment are in paper G attached to this thesis.

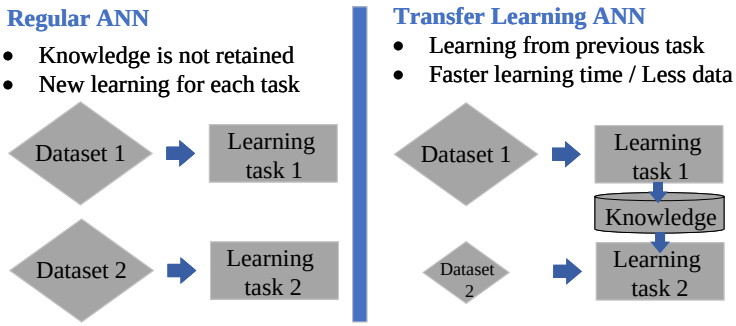


Fig. 26: Block diagram illustrating the transfer learning concept.

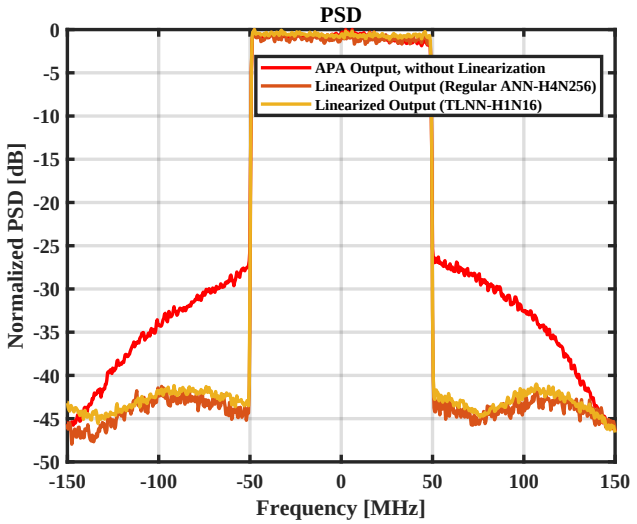


Fig. 27: Power spectral density results for linearization of 100 MHz BW signal; regular ANN vs. TLANN (TLNN with 1 hidden layer and 16 neurons is denoted as "TLNN-H1N16").

Table 1: Number of multiplications comparison between regular ANN and proposed TL ANN for a 100 MHz bandwidth signal. *)The average of the left and the right sides ACLR.

	Number of Multiplications	**EVM (without/with DPD), (Improv.)**	**ACLR *) (without/with DPD), (Improv.)**
Regular ANN (**256 Neurons, 4 hidden layers**)	69120	(10.3 % / 1.6 %), (8.7 %)	(34.7 dBc / 43.7 dBc), (9 dB)
TL ANN (**16 Neurons, 1 hidden layer**)	224	(10.3 % / 1.7 %), (8.6%)	(34.7 dBc / 43.7 dBc), (9 dB)

6 Contribution Summary

This section presents the research contribution of the papers applied to the Part II of this dissertation with a brief summary of motivations, contents and main results.

6.1 Paper A

Linearization of Active Transmitter Arrays in Presence of Antenna Crosstalk for 5G Systems
Feridoon Jalili, Martin H. Nielsen, Ming Shen, Ole K. Jensen, Jan H. Mikkelsen and Gert F. Pedersen
Published in the *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*

6.1.1 Motivation

The active phased arrays are constructed using several antennas together with a power amplifier connected to each antenna. The distance between antennas in the array is relatively short due to the limited area. This configuration makes linearization of each single PA unrealistic since we do need space for adding a coupler to the output of each PA. Furthermore, the individual PA's behavior is influenced by antenna crosstalk and the mutual coupling between antenna. A solution to this problem could be treating the amplifiers and antennas as one system and linearizing the main beam signal at the receiver rather than on each single power amplifier.

6.1.2 Paper content

For the measurements, an array of 4 antennas with one PA connected to each and the receiver antenna spaced one meter from the transmitter has been used. The power amplifiers are from CREE CGH400006P and as input signal, an LTE10 OFDM base station downlink signal with a center frequency at 3.5 GHz has been used. For evaluating the impact of mutual coupling a four elements antenna array having element spacing of d = [0.1 0.2 0.3 0.4 0.5 0.6] wavelength, λ, has been carried out.

6.1.3 Main results

The measurements show that the best results for EVM and ACLR are achieved with 0.5λ spacing. For investigating the coupling among antennas, the s-parameters for the array with 4 antennas have been measured which show that the strongest coupling is approximately -12 dB which explains why the variation in ACLR and EVM is not so high. The complexity of pre-distortion used in this work is the time alignment. The accurate time delay needed between input and output of the PA has been calculated by a simulation. An 8 dB improvement of ACLR has been achieved by using the system level linearization and without doing intensive s-parameter measurements needed in existing work.

6.2 Paper B

Linearization Trade-offs in a 5G mmWave Active Phased Array OTA Setup
Feridoon Jalili, Felice F. Tafuri, Ole K. Jensen, Yunfeng Li, Ming Shen and Gert F. Pedersen
Published in the *IEEE Access Year: 2020 | Volume: 8 | Journal Article*

6.2.1 Motivation

The 5G mmWave active phased array has normally several analog transmitter circuits connected directly to the antennas. Understanding the key factors contributing to the total nonlinearity of the whole array in such a complex system is not straightforward. Firstly 5G systems are using the so-called hybrid beam-forming where the number of analog RF chains is higher than the number of the baseband controllers. So a digital control unit may not directly linearize the individual PA at each antenna branch. Secondly, due to the high level of integration, the placement of the feedback circuits at the output of each PA is challenging. Thirdly, the feedback receiver needs to have very high bandwidth, i.e. up to 5*400 MHz as well as a high-speed analog to digital converter. With this complexity, having many feedback circuits is infeasible. Furthermore, using multiple antennas introduces cross-talk which increases the complexity of the algorithm for mitigating the impact. Finally, the typical implementation topology for mmWave is to up-convert the existing sub-6 GHz system into mmWave frequency. The linearity of these sub-6 GHz circuits has an impact on the overall system linearization.

6.2.2 Paper content

There is a trade-off between the optimal performance and the accepted complexity in terms of cost and size. In this work, we analyze these trade-offs by doing the measurement on state-of-the-art hardware devices together with the latest defined signal constellation in 5G transmitters. We consider the whole system as a single input single

output (SISO) system and use the conventional memory polynomial together with over-the-air (OTA) measurements. Following topics on linearization of the active phased array have been investigated:

i. impact of up-conversion from sub-6 GHz into mmWave
ii. impact of beam angle on linearization
iii. reusing trained coefficients in the backed-off region

6.2.3 Main results

Through simulation and measurement the following trade-offs have been identified:

i. a highly efficient active phased array needs to deliver the required maximum power while it is in compression. Since the efficiency of the last chain in the transmitter has high priority, so for getting the best efficiency for the whole system it is required to either run the pre-stages in the linear region or use a separate linearization loop for those circuits.

ii. due to load modulation which is a consequence of beamforming and crosstalk, the trained beam needs to be updated with a new set of training after some degree shift of the beam. For the actual setup, a new training after approximately ±15 degree shift of the main beam is required.

iii. The linearized system is out-performing the backed-off system above the trained output level, at the trained level and for the actual setup, 2 dB below the trained level. For the highly backed-off scenarios using a trained coefficient is not useful.

6.3 Paper C

Antenna Array Inter-Element Coupling impact on Linearization of Active Phased Array

Feridoon Jalili, Daniel E. Serup, Ondrej Franek, Ming Shen and Gert F. Pedersen

Published in the *International Symposium on Networks, Computers and Communications (ISNCC) Year: 2021 | Conference Paper | Publisher: IEEE*

6.3.1 Motivation

Due to the high level of integration in active phased array transmitters, the output stages are directly coupled to antennas without any isolator in between. In this configuration, the load impedance of these stages is affected by the mismatch of the antenna and crosstalk. Furthermore, the impedance at each antenna is not only related to the reflections from its own antenna but also related to coupling from other antennas in the array due to electromagnetic coupling among antennas. This coupling may vary during beam steering. This kind of linear distortion is called load modulation and needs to be considered during digital predistortion. Since a new training of the coefficients is costly, it is desired to keep the trained coefficients as long as possible during beam steering.

Load modulation limits this reuse of the coefficients. In this paper, the impact of the load modulation on the linearization of the active phased array is investigated through 3D simulation of the radiation patterns and measurement.

6.3.2 Paper content

The antenna array used in measurement has been designed as a 4×4 planar patch antenna on a low-loss Rogers RO4403 substrate. For simulation, the time domain solver in CST MICROWAVE STUDIO® (MWS) was used with a hexahedral mesh and a perfectly matched layer (PML) boundary condition. For investigation of the impact of load modulation, the 4x4 patch antenna has been exited in both sequential mode, i.e. all ports except the exited port are terminated, and simultaneous mode, i.e. all ports are exited. By applying appropriate phase shifting, the beam is shifted at various angles while exiting the ports. The combined polar and 3D far-field together with the reflection for each port are simulated. The input and output data of a corresponding 4x4 active phased array running at 28 GHz has been measured by placing an observation receiver in the far-field and located at the main beam while varying the beam direction in both horizontal and vertical steering angles. The DPD trained at $\theta = 0$ degree has been used for all steering angles to quantify the variation of ACLR and EVM as a result of inter-element coupling among antennas. A 4x4 active phased array based on Anokiwave AWMF-0158 hardware has been used as the device-under-test. For the experiment, a 3 GHz LTE10 signal, compliant with the 3GPP downlink orthogonal frequency-division multiplexing (OFDM) modulation with a peak to average power ratio of 10.6 dB has been up-converted to 28 GHz and used as input to the active array.

6.3.3 Main results

The reflections for specific ports during sequential excitation have been compared with simultaneous mode excitation. The results show that the reflection coefficient gets 5.5 dB worse in simultaneous excitation when the steering angle is changed by 9 degrees at 28 GHz. This change in impedance results in degradation of the linearization performance. Measurement results on the active array indicate this impact. For allowing only 0.5 dB degradation, a new training after approximately 5 degrees shift of the main beam is required which is a hard limit and requires several additional pieces of training during a wide range of beam steering.

6.4 Paper D

Tuning of Deep Neural Networks for Over-The-Air Linearization of Highly Nonlinear Wide-Band Active Phased Arrays
Feridoon Jalili, Felice F. Tafuri, Ole K. Jensen, Yunfeng Li, Qingyue Chen, Ming Shen and Gert F. Pedersen

Published in the *International Symposium on Networks, Computers and Communications (ISNCC) Year: 2021 | Conference Paper | Publisher: IEEE*

6.4.1 Motivation

Many state-of-the-art digital predistortions are successfully using artificial neural networks (ANN) for linearization of power amplifiers. For a system with wide bandwidth and high nonlinearity, ANN may need multiple hidden layers and a significant number of neurons to tackle the additional complexity raised by an active phased array transmitter in 5G. However, an analysis of the complexity vs. performance trade-off intrinsic to the usage of a high-complexity ANN model for linearization of the actual active phased array devices is an interesting investigation. This work gives a guideline for minimizing the computational cost of ANN implementation by tuning its major parameters.

6.4.2 Paper content

A 5G modulated signal at 28 GHz together with a 4x4 active phased array was used for investigations. The bandwidth is 100 MHz and the sampling frequency is 600 MHz. 100 k IQ samples of input and output of the active array have been captured and 70% of the data has randomly been used for training and the remaining 30% for testing. For training, the proposed ANN is implemented using TensorFlow 1.14 through the Keras API in Python 3.8.4. The number of delay lines at the input of the ANN, the number of neurons in each layer and the number of hidden layers are set appropriately to avoid underfitting and overfitting. For this purpose, the memory depth has been chosen to a low number, and then the other parameters have been initialized to get the best linearization parameters in terms of ACLR and EVM. The optimization is continued by tuning the memory depth and keeping other parameters at their optimum level.

6.4.3 Main results

For each power level, the ANN is trained to reach a mean square error (MSE) of less than 1E-6. The best performance is achieved by setting the memory depth to 14. With this fixed number of delay lines, the number of neurons is tuned for best linearization which shows that there is only a minor impact from the number of neurons if it is at least 150. In the end, the number of hidden layers has been tuned keeping the other parameters unchanged. The results show that the system reaches the optimum linearization performance using 3 hidden layers.

6.5 Paper E

Highly Nonlinear and Wide-Band mmWave Active Array OTA Linearization Using Neural Network

Feridoon Jalili, Yufeng Zhang, Markko Hintsala, Qingyue Chen, Ming Shen and Gert F. Pedersen
Published in the *IET Microwaves, Antennas & Propagation | Year 2021 | Journal Article*

6.5.1 Motivation

Active phased array (APA) transmitters together with multi-input multi-output (MIMO) systems have been chosen for 5G mobile communication sytem operating at mmWave in order to improve the system capacity and data rates. The traditional linearization techniques based on the Volterra series model face new challenges such as:
i. highly integrated front-ends and a large number of PA's do not allow placement of feedback circuits for each branch.
ii. wide bandwidth.
iii. need for higher efficiency together with enhanced modulation scheme results in driving the power amplifier into very high nonlinearity.
The challenges can in the worst case lead to huge complexity and explosion of the Volterra-series model approach, which is commonly used in MPM. A DPD based on MPM for such a highly complex system needs to increase the number of unknown kernel coefficients in the model which is practical. In recent years several ANN architectures have been proposed and applied to linearize highly nonlinear multi-PA devices-under-test (DUTs) such as the active phased array. A comparison between ANN and MPM for such a highly nonlinear and wide bandwidth scenario tested on state-of-the-art hardware and OTA verification is an interesting topic.

6.5.2 Paper content

A SISO OTA modeling where the entire transmitter has been considered as a two-port system together with an observation receiver in far-field is used in this work and the linearization results of both ANN-based and MPM-based DPD techniques for linearization of the antenna array have been given. For mitigating the memory effect due to the wide bandwidth, two dynamic neural structures have been proposed in the literature; recurrent neural networks (RNNs) which utilize feed-forward and feedback signal processing and real-valued time-delay neural network (RVTDNN). Due to superior performance and easy baseband implementation, the RVTDNN has been chosen as the ANN architecture in this work. For the high nonlinearity cases, the model needs a low learning rate during training at the cost of the training time. In the present paper, a so-called batch normalization (BN) layer is added to each hidden layer to increase the learning rate. Furthermore, the proposed RVTDNN uses the rectified linear units (ReLU) activation function, which is less computationally expensive than hyperbolic tangent (Tanh) and Sigmoid because it involves simpler mathematical operations. To make it comparable with conventional MPM, an ANN including only one hidden layer together with a minimum number of neurons has been implemented. A RVTDNN using

the ReLU activation function together with a total radiated power (TRP) measurement of a real 5G test bench is the unique content of this work.

6.5.3 Main results

When using MPM-based models for linearization of a deeply saturated PA, the peaks of the predistorted signal reach extremely high values and hence leading to a huge increase in the predistorted signal's peak to average power ratio (PAPR). The nonlinear kernel of the NN does not contain inherently such an "explosion" effect for high non-linearities. The proposed RVTDNN structure in this work is based on supervised learning which means that while we train the ANN we are using low envelope fluctuations, i.e. the desired I and Q at the output layer. This allows the ANN to learn what the characteristics of a signal with low envelope fluctuations are.

6.6 Paper F

Complexity Analysis of Artificial Neural Networks Used for Active Phased Array Linearization
Feridoon Jalili, Felice F. Tafuri, Kasper B. Olesen, Lauge F. Dyring, Jakob G. Brask, Ming Shen and Gert F. Pedersen
Published in the *2022 IEEE MTT-S International Wireless Symposium (IWS) Year: 2022 | Conference Paper | Publisher: IEEE*

6.6.1 Motivation

The new challenges introduced by the 5G mmWave active phased array have increased the complexity of the conventional techniques such as the memory polynomial model (MPM) used for digital predistortion. Rich experimental results are showing that the artificial neural network (ANN) based DPD approach can be a good alternative for addressing the new challenges. The complexity of digital preditortion can be defined as the number of multiplications during the training of the predistorter. For the recent proposed 5G multiple-input multiple-output (MIMO) transmitters architecture, the capability of ANN in comparison to MPM in terms of multiplication is still an open and ongoing discussion.

6.6.2 Paper content

A predistorted signal modeling (PSM) has been designed for optimizing the number of time delays and neurons for achieving the target linearity performance and afterward the number of multiplications has been bench-marked. The real-valued time-delay neural network (RVTDNN) has been used as the ANN structure. The applied MPM, which is a deviation of the Hammerstein model and has been proven effective for removing

nonlinearity and memory effects has been used for bench-marking. The comparison is based on verification results of a highly nonlinear and wide-band 5G testbed measured over-the-air (OTA) in a compact antenna test range (CATR) chamber. The APA device is based on Amotech AAiPK428GC-A0404 which includes four Anokiwave AWMF-0158 transceivers. It integrates 16 branches of attenuators, phase shifters and PAs and 16 patch antennas in a 4×4 active phased array. A host PC is used for capturing and uploading the IQ samples.

6.6.3 Main results

Although the DPD complexity-performance trade-off will naturally be correlated to the device-under-test and to the MIMO architecture of the array, the methodology can be used to identify the optimal complexity-performance trade-off when choosing the ANN parameters. Increasing the number of time-delays or neurons will not necessarily result in better linearity and there is not a direct relation between the number of multiplication and optimum results for each test case. We are providing a method that allows system engineers to choose an ANN complexity level per a given wanted linearity.

6.7 Paper G

Bandwidth Scalable Behavioral Modeling using Neural Network Based on Transfer Learning
Feridoon Jalili, Felice F. Tafuri, Qingyue Chen, Ming Shen and Gert F. Pedersen
Submitted to the *IEEE Access Year: 2022 | Volume: 8 | Journal Article*

6.7.1 Motivation

One of the major differences between the introduced 5G mobile communication with generations behind is the dynamic behavior of the parameters in the system, such as sub-carrier spacing (SCS), modulation scheme, bandwidth, etc. The existing digital predistortion algorithms need to be capable to cope with such dynamic behavior. Furthermore, this dynamic behavior is together with the increasing complexity of the system, for example increasing signal bandwidth which leads to strong nonlinearity and memory effects and as a result to the increasing complexity and adaption time. For such a dynamic system, the existing DPD methods need to update a huge amount of coefficients rapidly which may make the method infeasible. Artificial neural network (ANN) has been widely used in the modeling of nonlinear devices. In this work, a cost-effective real-valued time-delayed neural network (RVTDNN) method has been used in several scenarios. However, by increasing the bandwidth and nonlinearity, the RVTDNN method needs to increase the input dimensions of the model and number of layers, and a larger number of neurons which may make the model slow for a dynamic system. Transfer learning ANN approach which can intelligently apply knowledge

learned previously to solve new problems faster appears to be a good solution for the linearization of systems with dynamic bandwidth. This method allows the domains, tasks, and distributions used in training and testing to be different.

6.7.2 Paper content

A single input single output (SISO) model where the entire transmitter has been considered as a two-port system has been used. The proposed TL-RVTDNN method consists of three phases; offline learning using the source dataset, using a 20 MHz bandwidth signal, identifying the transfer learning algorithm, and online application based on the target dataset ie. the 100 MHz bandwidth signal. The first few layers of the network, which are used for extracting the nonlinear characteristics of the active phased array in 20 MHz bandwidth, are frozen after executing offline training. For training, the input and output datasets are divided as 70 % for training and 30 % for validation. The weights and biases of the ANN are using the Huber loss function and Adam is used as the optimization algorithm.

6.7.3 Main results

The optimum model of 20 MHz bandwidth is modified by removing the last fully connected hidden layer. This model is frozen as a transferred pre-design model and used to linearize the 100 MHz bandwidth signal. The frozen pre-defined model from 20 MHz training and one fully connected fine-tuning hidden layer is used For TL-RVTDNN. Four different sets of neurons are used for fine-tuning. The TL-RVTDNN linearization can provide the same linearization result as the regular ANN by using only one hidden layer and 16 Neurons. This means that the frozen layers based on 20 MHz bandwidth already contain enough information to compensate for the nonlinearities of the 100 MHz bandwidth signal. Since the time for calculating the coefficients of the incremental layers is reduced, this approach grants the ability to an adaptive re-identification of the ANN model.

7 Conclusion

The new generation of wireless communication systems operating at mmWaves, use a new architecture based on the active phased array (APA) at the transceiver's RF front-end which requires integration of several power-hungry devices directly connected to the antennas. This new architecture together with a high peak to average ratio (PAPR) and wide bandwidth of the signal has introduced new challenges for enhancing the power efficiency while mitigating distortions. Digital predistortion (DPD) has been considered an effective technique to cope with the unwanted nonlinear distortion in the past generations of mobile platforms. However, the number of parameters required for predicting the nonlinearities in the DPD technique based on the existing Volterra-series-based models such as the memory polynomial model (MPM) is growing exponentially which increases the computational complexity and cost. In recent years artificial neural networks (ANNs) are strongly showing their advantages in terms of reduction of computational complexity and enhancing the performance of the emerging 5G and 6G mobile communication systems. This dissertation presents an overview of major linearization challenges in 5G mmWave transmitters and proposes possible solutions based on OTA measurements and ANN-based predistortion.

The implementation of a feedback receiver for training of the behavioral model of the nonlinear devices for DPD is not straightforward and many types of research and investigations are ongoing for finding a proper hardware implementation. Defining the mutual coupling among antennas using an S-parameter matrix and using it together with the DPD algorithm is not feasible since an accurate S-parameter measurement on a highly integrated architecture is not possible. Therefore an equivalent single-input single-output (SISO) model has been used for measurements.

The performance and complexity of the proposed ANN architecture have been compared with the existing MPM techniques together with detailed analyses, simulations and measurements by using state-of-the-art 5G active phased array. OTA measurements for capturing data and validation together with the required time alignment have been defined step-by-step in several works. Since the distortion is also beam-formed in a steerable transmission configuration, TRP measurement has been done by using a compact antenna test range (CATR). In the case of transmitting with a highly nonlinear ANN, the ANN-based linearization shows superior performance due to lower PAPR in the predistorter.

Detailed analysis and definition for the proposed architecture of ANN including a proposed input-output configuration have been given. The benefit of using different optimization algorithms, choice of activation functions, loss functions and the influence of hyper-parameters used in the model have been analyzed and verified on several setups and presented in the published papers.

Furthermore, the enhanced ANN architecture using an algorithm based on transfer learning has been analyzed and the advantage of that on the dynamical bandwidth

changing 5G system has been investigated.
On the hardware configuration, the setup includes a state-of-the-art 4×4 active phased array from Anokiewave together with 16 integrated patch antennas. The hardware setup is using an up-converter and a down-converter between sub-6 GHz and mmWave which is close to what is implemented in the latest available 5G mobile phones and gives an indication of the issues and proposals related to this configuration.
The presented work is far from being exhaustive for all the aspects and mostly focuses on research topics where the active phased array is studied from a system-level point of view. All aspects of the pros and cons of using the SISO model for DPD are far from adequately covered. A large privilege has been given to the modeling and characterization concepts using artificial neural networks and to the practical implementation of the whole active phased array transmitter using a measurement setup in the compact range chamber.

For future work
For most of the measurements in this work an intently controlled environment has been used where the impact from hardware imperfections of the elements in the APA branches and channel impact are kept ideal. A future work may include investigations of these impacts.
The robustness of ANN-based DPD concerning input power level variation is also one of the future directions for this research.
Due to the APA's spectral regrowth, the sampling rate of the feedback receiver needs to be several times the bandwidth of the transmitter signal for DPD identification. Taking the signal bandwidth growth in 5G and 6G, leads to very wide bandwidth requirements for the feedback receiver and increased hardware complexity. Using the transfer learning ANN approach together with implementing DPD on a small portion of the transmitter's bandwidth and re-using that for the whole bandwidth may relax the feedback receiver's complexity.
The on-line adaptive DPD design is also a major investigation topic. There is a need for a far-filed observation antenna for providing the OTA feedback signal for adaptive on-line DPD. This observation antenna might be available from other user's devices as introduced for a heterogeneous networks in 5G and beyond. The implementation of such a system is an interesting topic. The feedback signal could also be obtained from the receiver antenna of the same phone or base station but the proper implementation techniques are still under discussion. One promising proposal is to use the diversity antenna. However, there may be an issue with low power-coupling between the transmitter and the diversity antennas. Furthermore this approaches may need digital signal processing for near-field to far-field transformation in antenna measurements and yet there is no adequate information about the feasibility of this approach.

References

[1] Petar Popovski, Kasper Fløe Trillingsgaard, Osvaldo Simeone, and Giuseppe Durisi. 5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view. *Ieee Access*, 6:55765–55779, 2018.

[2] Rahul Rai, Manoj Kumar Tiwari, Dmitry Ivanov, and Alexandre Dolgui. Machine learning in manufacturing and industry 4.0 applications, 2021.

[3] Salvador V Balkus, Honggang Wang, Brian D Cornet, Chinmay Mahabal, Hieu Ngo, and Hua Fang. A survey of collaborative machine learning using 5g vehicular communications. *IEEE Communications Surveys & Tutorials*, 2022.

[4] Leandra Börner Valdez, Rabi R Datta, Benjamin Babic, Dolores T Müller, Christiane J Bruns, and Hans F Fuchs. 5g mobile communication applications for surgery: An overview of the latest literature. *Artificial Intelligence in Gastrointestinal Endoscopy*, 2(1):1–11, 2021.

[5] Erik G Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L Marzetta. Massive mimo for next generation wireless systems. *IEEE communications magazine*, 52(2):186–195, 2014.

[6] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave mimo systems. *IEEE journal of selected topics in signal processing*, 10(3):436–453, 2016.

[7] Andreas F Molisch, Vishnu V Ratnam, Shengqian Han, Zheda Li, Sinh Le Hong Nguyen, Linsheng Li, and Katsuyuki Haneda. Hybrid beamforming for massive mimo: A survey. *IEEE Communications magazine*, 55(9):134–141, 2017.

[8] Feridoon Jalili, Martin H Nielsen, Ming Shen, Ole K Jensen, Jan H Mikkelsen, and Gert F Pedersen. Linearization of active transmitter arrays in presence of antenna crosstalk for 5g systems. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, pages 1–5. IEEE, 2019.

[9] Feridoon Jalili, Daniel E Serup, Ondrej Franek, Ming Shen, and Gert F Pedersen. Antenna array inter-element coupling impact on linearization of active phased array. In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, pages 1–5. IEEE, 2021.

[10] Alberto Brihuega, Mahmoud Abdelaziz, Lauri Anttila, Matias Turunen, Markus Allén, Thomas Eriksson, and Mikko Valkama. Piecewise digital predistortion for

mmwave active antenna arrays: Algorithms and measurements. *IEEE Transactions on Microwave Theory and Techniques*, 68(9):4000–4017, 2020.

[11] 3GPP 3rd Generation Partnership Project. Base station radio transmission and reception, 3GPP ts 38.104. http://www.3gpp.org/ftp/Specs/archive/38_series/38.104/, 2018.

[12] Lauri Anttila, Alberto Brihuega, and Mikko Valkama. On antenna array out-of-band emissions. *IEEE Wireless Communications Letters*, 8(6):1653–1656, 2019.

[13] Christian Fager, Thomas Eriksson, Filipe Barradas, Katharina Hausmair, Telmo Cunha, and Jose Carlos Pedro. Linearity and efficiency in 5g transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5g active antenna transmitter context. *IEEE Microwave Magazine*, 20(5):35–49, 2019.

[14] Ada SY Poon and Mazhareddin Taghivand. Supporting and enabling circuits for antenna arrays in wireless communications. *Proceedings of the IEEE*, 100(7):2207–2218, 2012.

[15] S. A. Bassam, M. Helaoui, and F. M. Ghannouchi. Crossover digital predistorter for the compensation of crosstalk and nonlinearity in mimo transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 57(5):1119–1128, 2009.

[16] S Crips. RF power amplifiers for wireless communications, artech house, 1999.

[17] Zahidul Islam Shahin, et. al. Efficient dpd coefficient extraction for compensating antenna crosstalk and mismatch effects in advanced antenna system. Department of Electrical and Information Technology LTH, Lund University SE-221 00 Lund, Sweden, 2017.

[18] Feridoon Jalili, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Ming Shen, and Gert F Pedersen. Linearization trade-offs in a 5g mmwave active phased array ota setup. *Ieee Access*, 8:110669–110677, 2020.

[19] Hongmin Li, Gang Li, Yikang Zhang, Wen Qiao, and Falin Liu. Forward modeling assisted digital predistortion method for hybrid beamforming transmitters with a single pa feedback. In *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, pages 179–182. IEEE, 2018.

[20] Xin Liu, Qian Zhang, Wenhua Chen, Haigang Feng, Long Chen, Fadhel M Ghannouchi, and Zhenghe Feng. Beam-oriented digital predistortion for 5g massive mimo hybrid beamforming transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(7):3419–3432, 2018.

[21] Nuutti Tervo, Janne Aikio, Tommi Tuovinen, Timo Rahkonen, and Aarno Parssinen. Digital predistortion of amplitude varying phased array utilising over-the-air combining. In *2017 IEEE MTT-S International Microwave Symposium (IMS)*, pages 1165–1168. IEEE, 2017.

[22] Mahmoud Abdelaziz, Lauri Anttila, Alberto Brihuega, Fredrik Tufvesson, and Mikko Valkama. Digital predistortion for hybrid mimo transmitters. *IEEE Journal of Selected Topics in Signal Processing*, 12(3):445–454, 2018.

[23] Sara Hesami, Sina Rezaei Aghdam, Christian Fager, Thomas Eriksson, Ronan Farrell, and John Dooley. Single digital predistortion technique for phased array linearization. In *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2019.

[24] Eric Ng, Yehia Beltagy, Patrick Mitran, and Slim Boumaiza. Single-input single-output digital predistortion of power amplifier arrays in millimeter wave rf beamforming transmitters. In *2018 IEEE/MTT-S International Microwave Symposium-IMS*, pages 481–484. IEEE, 2018.

[25] Eric Ng, Ahmed Ben Ayed, Patrick Mitran, and Slim Boumaiza. Single-input single-output digital predistortion of multi-user rf beamforming arrays. In *2019 IEEE MTT-S International Microwave Symposium (IMS)*, pages 472–475. IEEE, 2019.

[26] Xin Liu, Wenhua Chen, Long Chen, Fadhel M Ghannouchi, and Zhenghe Feng. Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks. *IEEE Transactions on Microwave Theory and Techniques*, 67(12):5235–5248, 2019.

[27] Xiaoyu Wang, Yue Li, Chao Yu, Wei Hong, and Anding Zhu. Digital predistortion of 5g massive mimo wireless transmitters based on indirect identification of power amplifier behavior with ota tests. *IEEE Transactions on Microwave Theory and Techniques*, 68(1):316–328, 2019.

[28] Xiaoyu Wang, Chao Yu, Yue Li, Wei Hong, and Anding Zhu. Real-time single channel over-the-air data acquisition for digital predistortion of 5g massive mimo wireless transmitters. In *2019 IEEE MTT-S International Wireless Symposium (IWS)*, pages 1–3. IEEE, 2019.

[29] Chao Yu, Jianxin Jing, Han Shao, Zhi Hao Jiang, Pinpin Yan, Xiao-Wei Zhu, Wei Hong, and Anding Zhu. Full-angle digital predistortion of 5G millimeter-wave massive mimo transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 67(7):2847–2860, 2019.

[30] Petar Popovski. *Wireless Connectivity: An Intuitive and Fundamental Guide*. John Wiley & Sons, 2020.

[31] Marina Jordão, Rafael Caldeirinha, Arnaldo SR Oliveira, and Nuno Borges Carvalho. A survey on over-the-air linearization methods for mimo systems. *Energies*, 14(8):2225, 2021.

[32] Anokiwave. *AWMF-0158 28 GHz Silicon 5G Tx/Rx Quad Core IC*, 2019.

[33] Antenna Systems Solutions S.L. (ASYSOL). Compact antenna test range. https://asysol.com/portfolio/mini-compact-antenna-test-range/, 2020.

[34] Constantine A Balanis. *Antenna theory: analysis and design*. John wiley & sons, 2015.

[35] Evolved Universal Terrestrial Radio Access. Base station radio transmission and reception. volume 36, pages 96–97, 2018.

[36] Jose Carlos Pedro and Nuno Borges Carvalho. *Intermodulation distortion in microwave and wireless circuits*. Artech House, 2003.

[37] YH Ku. The volterra & wiener theories of nonlinear systems: Martin schetzen. 531 pages, diagrams, illustr., John Wiley, New York, 1980., 1982.

[38] Martin Schetzen. Nonlinear system modeling based on the wiener theory. *Proceedings of the IEEE*, 69(12):1557–1573, 1981.

[39] Jacob Engberg and Torben Larsen. *Noise theory of linear and nonlinear circuits*. Wiley, 1995.

[40] N Norholm, C Iversen, and Torben Larsen. Gaas mesfet large-signal modelling for multiport volterra series analysis. *IEE Proceedings-Circuits, Devices and Systems*, 144(1):40–44, 1997.

[41] Anding Zhu and Thomas J Brazil. An overview of volterra series based behavioral modeling of rf/microwave power amplifiers. In *2006 IEEE annual wireless and microwave technology conference*, pages 1–5. IEEE, 2006.

[42] Fadhel M Ghannouchi, Oualid Hammi, and Mohamed Helaoui. *Behavioral modeling and predistortion of wideband wireless transmitters*. John Wiley & Sons, 2015.

[43] J. Kim and K. Konstantinou. Digital predistortion of wideband signals based on power amplifier model with memory. *Electronics Letters*, 37(23):1–2, 11 2001.

[44] Magnus Isaksson and David Wisell. Extension of the hammerstein model for power amplifier applications. In *ARFTG 63rd Conference, Spring 2004*, pages 131–137. IEEE, 2004.

[45] Chokri Jebali, Ali Gharsallah, Noureddine Boulejfen, and Fadhel M Ghannouchi. Performance assessment of rf power amplifier memory polynomial models under different signal statistics. In *2009 16th IEEE International Conference on Electronics, Circuits and Systems-(ICECS 2009)*, pages 383–386. IEEE, 2009.

[46] Lei Ding, Guo Tong Zhou, Dennis R Morgan, Zhengxiang Ma, J Stevenson Kenney, Jaehyeong Kim, and Charles R Giardina. Memory polynomial predistorter based on the indirect learning architecture. In *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, volume 1, pages 967–971. IEEE, 2002.

[47] Lei Ding, Guo Tong Zhou, Dennis R Morgan, Zhengxiang Ma, J Stevenson Kenney, Jaehyeong Kim, and Charles R Giardina. A robust digital baseband predistorter constructed using memory polynomials. *IEEE Transactions on communications*, 52(1):159–165, 2004.

[48] Dennis R Morgan, Zhengxiang Ma, Jaehyeong Kim, Michael G Zierdt, and John Pastalan. A generalized memory polynomial model for digital predistortion of rf power amplifiers. *IEEE Transactions on signal processing*, 54(10):3852–3860, 2006.

[49] Changsoo Eun and Edward J Powers. A new volterra predistorter based on the indirect learning architecture. *IEEE Transactions on Signal Processing*, 45(1):223–227, 1997.

[50] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[51] Pere L Gilabert, David López-Bueno, Thi Quynh Anh Pham, and Gabriel Montoro. Machine learning for digital front-end: a comprehensive overview. *Machine Learning for Future Wireless Communications*, pages 327–381, 2020.

[52] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[53] Qingyue Chen, Yufeng Zhang, Feridoon Jalili, Zhugang Wang, Yonghui Huang, Yubo Wang, Ying Liu, Gert Frølund Pedersen, and Ming Shen. Robust digital signal recovery for leo satellite communications subject to high snr variation and transmitter memory effects. *IEEE Access*, 2021.

[54] Dongming Wang, Mohsin Aziz, Mohamed Helaoui, and Fadhel M Ghannouchi. Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters. *IEEE transactions on neural networks and learning systems*, 30(1):242–254, 2018.

[55] Praveen Jaraut, Meenakshi Rawat, and Fadhel M Ghannouchi. Composite neural network digital predistortion model for joint mitigation of crosstalk, I/Q imbalance,

nonlinearity in MIMO transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(11):5011–5020, 2018.

[56] Yikang Zhang, Yue Li, Falin Liu, and Anding Zhu. Vector decomposition based time-delay neural network behavioral model for digital predistortion of rf power amplifiers. *IEEE Access*, 7:91559–91568, 2019.

[57] Meenakshi Rawat, Karun Rawat, and Fadhel M Ghannouchi. Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 58(1):95–104, 2009.

[58] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[59] Xavier Glorot, Antoine Bordes, and Y. Bengio. Deep sparse rectifier neural networks. *Journal of Machine Learning Research*, 15, 2010.

[60] Abdalla E Abdelrahman, Oualid Hammi, Andrew K Kwan, Azzedine Zerguine, and Fadhel M Ghannouchi. A novel weighted memory polynomial for behavioral modeling and digital predistortion of nonlinear wireless transmitters. *IEEE Transactions on Industrial Electronics*, 63(3):1745–1753, 2015.

[61] Reina Hongyo, Yoshimasa Egashira, Thomas M Hone, and Keiichi Yamaguchi. Deep neural network-based digital predistorter for doherty power amplifiers. *IEEE Microwave and Wireless Components Letters*, 29(2):146–148, 2019.

[62] X Glorot, A Bordes, and Y Bengio. Deep sparse rectifier neural networks in journal of machine 964 learning research. *Universite de Technologie de Compiègne, Compiegne, France*, page 965.

[63] Changliu Liu, Tomer Arnon, Christopher Lazarus, Christopher Strong, Clark Barrett, and Mykel J Kochenderfer. Algorithms for verifying deep neural networks. *arXiv preprint arXiv:1903.06758*, 2019.

[64] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[65] Zhancang Wang. Demystifying envelope tracking: Use for high-efficiency power amplifiers for 4g and beyond. *IEEE Microwave Magazine*, 16(3):106–129, 2015.

[66] Quynh Anh Pham, David López-Bueno, Teng Wang, Gabriel Montoro, and Pere L Gilabert. Multi-dimensional lut-based digital predistorter for concurrent dual-band envelope tracking power amplifier linearization. In *2018 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, pages 47–50. IEEE, 2018.

[67] Raymond Pengelly, Christian Fager, and Mustafa Ozen. Doherty's legacy: a history of the doherty power amplifier from 1936 to the present day. *IEEE Microwave Magazine*, 17(2):41–58, 2016.

[68] Emilio J Martínez-Pérez, Feridoon Jalili, Ming Shen, Jan H Mikkelsen, Ole K Jensen, and Gert F Pedersen. T-linc architecture with digital combination and mismatch correction in the receiver. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, pages 1–5. IEEE, 2019.

[69] Jessica Chani-Cahuana, Mustafa Özen, Christian Fager, and Thomas Eriksson. Digital predistortion parameter identification for rf power amplifiers using real-valued output data. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 64(10):1227–1231, 2017.

[70] F. M. Ghannouchi, O. Hammi, and M. Helaoui. *Behavioural Modelling and Predistortion of Wideband Wireless Transmitters*. John Wiley & Sons, 1 edition, 7 2015.

[71] Dayong Zhou and Victor E DeBrunner. Novel adaptive nonlinear predistorters based on the direct learning algorithm. *IEEE transactions on signal processing*, 55(1):120–133, 2006.

[72] Mazen Abi Hussein, Vivek Ashok Bohara, and Olivier Venard. On the system level convergence of ila and dla for digital predistortion. In *2012 International Symposium on Wireless Communication Systems (ISWCS)*, pages 870–874. IEEE, 2012.

[73] Henna Paaso and Aarne Mammela. Comparison of direct learning and indirect learning predistortion architectures. In *2008 IEEE International Symposium on Wireless Communication Systems*, pages 309–313. IEEE, 2008.

[74] Leticia Aladren, Paloma Garcia-Ducar, Jesus de Mingo, Cesar Sanchez-Perez, and Pedro Luis Carro. Behavioral power amplifier modeling and digital predistorter design with a chirp excitation signal. In *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2011.

[75] Felice Francesco Tafuri, Cataldo Guaragnella, Marco Fiore, and Torben Larsen. Linearization of rf power amplifiers using an enhanced memory polynomial predistorter. In *NORCHIP 2012*, pages 1–4. IEEE, 2012.

[76] Lei Ding, Raviv Raich, and G Tong Zhou. A hammerstein predistortion linearization design based on the indirect learning architecture. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages III–2689. IEEE, 2002.

[77] Padmanabhan Madampu Suryasarman and Andreas Springer. A comparative analysis of adaptive digital predistortion algorithms for multiple antenna transmitters. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 62(5):1412–1420, 2015.

[78] Fa-Long Luo. *Digital front-end in wireless communications and broadcasting: circuits and signal processing*. Cambridge University Press, 2011.

[79] Jessica Chani-Cahuana, Per Niklas Landin, Christian Fager, and Thomas Eriksson. Iterative learning control for rf power amplifier linearization. *IEEE Transactions on Microwave Theory and Techniques*, 64(9):2778–2789, 2016.

[80] Xiaowen Feng, Bruno Feuvrie, Anne-Sophie Descamps, and Yide Wang. A digital predistortion method based on nonuniform memory polynomial model using interpolated lut. In *2015 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR)*, pages 1–3. IEEE, 2015.

[81] Feridoon Jalili, Yufeng Zhang, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Qingyue Chen, Ming Shen, and Gert F Pedersen. Tuning of deep neural networks for over-the-air linearization of highly nonlinear wide-band active phased arrays. In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, pages 1–4. IEEE, 2021.

[82] Alberto Brihuega, Lauri Anttila, and Mikko Valkama. Neural-network-based digital predistortion for active antenna arrays under load modulation. *IEEE Microwave and Wireless Components Letters*, 30(8):843–846, 2020.

[83] Feridoon Jalili, Yufeng Zhang, Markku Hintsala, Ole Kiel Jensen, Qingyue Chen, Ming Shen, and Gert Frølund Pedersen. Highly non-linear and wide-band mmwave active array ota linearisation using neural network. *IET Microwaves, Antennas & Propagation*, 2022.

[84] Pere Lluís Gilabert Pinal. *Multi look-up table digital predistortion for RF power amplifier linearization*. Universitat Politècnica de Catalunya, 2008.

[85] Zhijun Liu, Xin Hu, Ting Liu, Xiuhua Li, Weidong Wang, and Fadhel M Ghannouchi. Attention-based deep neural network behavioral model for wideband wireless power amplifiers. *IEEE Microwave and Wireless Components Letters*, 30(1):82–85, 2019.

[86] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[87] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

Part II

Papers

Paper A

Linearization of Active Transmitter Arrays in Presence of Antenna Crosstalk for 5G Systems

Feridoon Jalili, Martin H. Nielsen, Ming Shen, Ole K. Jensen, Jan H. Mikkelsen and Gert F. Pedersen

The paper has been published in the
2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC) .

Abstract

By increasing the number of antennas and power amplifiers connected to each in 5G system, the linearization methods like digital pre-distortion (DPD) of each power amplifier is inefficient due to antenna imperfections such as crosstalk. As a result of limited area the distance between antennas in array can vary which leads to unwanted coupling between antennas. A solution to this problem could be treating the amplifiers and antennas as one system and linearizing the main beam signal at the receiver rather than on each single power amplifier. In the work described in this paper, the whole system including amplifiers, antennas and the receiver is treated as a 2-ports system and the impacts of the above mentioned constraints are investigated.

1 Introduction

Due to growing demand for higher data rate, which needs extended bandwidth, new mm-Wave frequency bands are introduced as enabling technology in 5G mobile system. Together with the requirements of massive MIMO, highly integrated beam-steerable active arrays consisting of a high number of PAs and antennas are considered as an efficient solution [1]. The active array topology where the PAs are placed after the phase shifters, and just before the antennas, gives several benefits such as reducing the power handling of phase shifters, reducing the output power requirements for each element and allowing small integrated devices to be used while connected to the antenna. On the other hand active arrays place many challenges for the traditional linearization methods used for PA efficiency:

- Antenna crosstalk
- Coupling vs distance between antennas
- Varying power levels at antenna branches due to side-lobe control requirements

In this paper, the linearization of power amplifiers in the presence of cross talk is investigated. The task includes the characterization of the crosstalk and its impact on the system linearity and DPD of PAs under effect of antenna crosstalk. The applied DPD technique captures the combined non-linearity of the whole array including the PAs and antenna elements. Specifically the particular focus is to reduce the system complexity, while maintaining the linearization performance. The investigations in this paper are done by combining measurements together with analyses in Matlab.

The paper is organized as follows: Section 1 is the introduction. Section 2 discusses system theory and background. Measurement set-up of proposed system DPD and results have been provided in section 3. Finally, the conclusions of this work are presented in section 7.

2 System Theory and Background

2.1 Antenna crosstalk in arrays

A common beam-forming structure, the so-called "hybrid beam forming", is shown in Fig. A.1. Each sub-array includes a DAC and a modulator. The number of antennas in each sub-array is higher than the number of modulators in order to minimize the power consumption since the DAC, the modulator and the preceding digital part are the most power consuming elements [2]. Crosstalk as coupling from one branch to another of transmitter in an active antenna array can be categorized in two types: before PA and after PA.

2.1.1 Crosstalk before PA

This is mainly due to RF leakage through the common LO and coupling between different transmit paths because of the electromagnetic coupling [3].

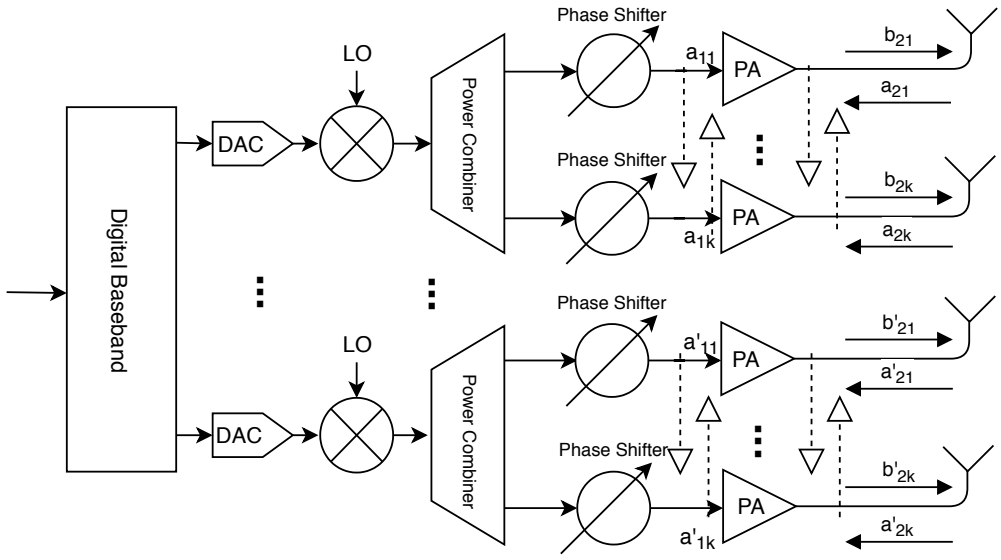


Fig. A.1: Model of crosstalk in Hybrid beam-forming

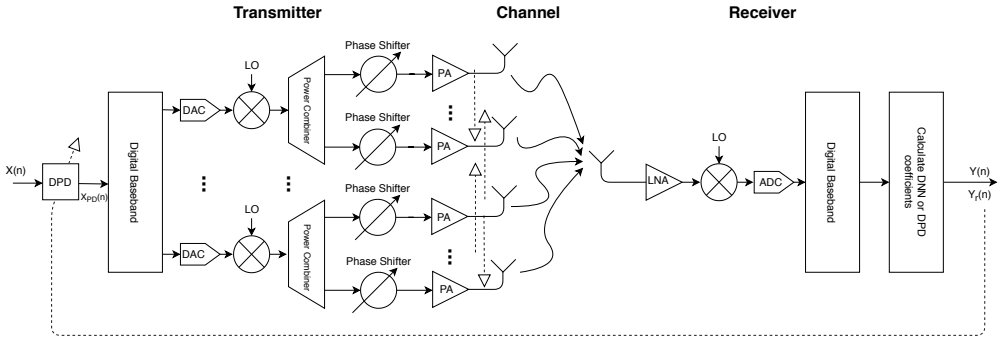


Fig. A.2: Estimation of DPD coefficient at receiver side

2.1.2 Crosstalk after PA

When no isolators are present at the PAs' outputs, there is mutual coupling between them via the antennas [4]. A sketch of this is shown in Fig. A.1 where a_{1k} is the incoming signal to the amplifier, b_{2k} is the output from the amplifier and a_{2k} is the reflected signal form the antenna array at the kth branch.

The relationship between a_{2k} and the output signal b_{2k} is determined by the characteristics of the antenna array's s-parameters. The system model of the multi-antenna transmitter could therefore be split into a crosstalk and mismatch model. This model could also be used together with the DPD that holds the model for the PA, to linearize it [5].

A major problem with this approach is that the s-parameters for the antennas must be known and that the method needs feedback from each amplifier which is a problem if several amplifiers are used.

2.2 On-line system linearization

A novel solution to the above mentioned concerns could be to treat amplifiers and antennas as a single system and then do the DPD estimation for the main beam at the receiver side as depicted in Fig. A.2. In order to make this configuration to work, the transmitter has to send a pilot sequence which is known to own receiver. Then the receiver can obtain the coefficients for the DPD algorithm and feed those back to the transmitter. Calculation of the linearization coefficients can either be done as traditional memory polynomial or Deep Neural Network (DNN). On-line system linearization can also be used in satellite communication as post distortion analysis for power efficiency [6, 7] done in ground-station receiver. The benefit with the last approach is that it includes also the channel response.

3 Measurements

3.1 Measurement setup

Figure A.3 shows the measurement setup in lab which is consisting of an array of 4 antennas with one PA connected to each and the receiver antenna spaced one meter from the transmitter. Power amplifier of type CREE CGH400006P [8] is used for the measurements.

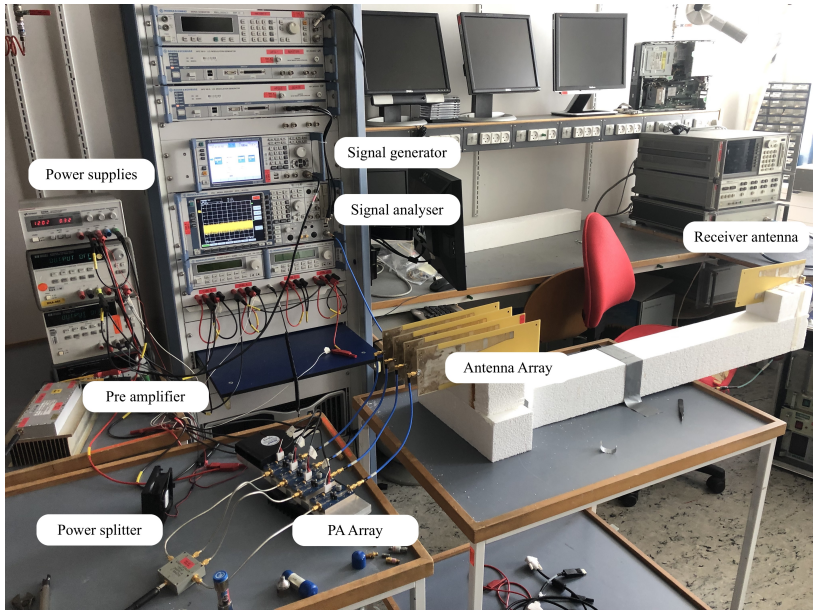


Fig. A.3: Measurement setup.

The gate voltage for the amplifier is adjusted to Vg = - 2.798 V resulting in a quiescent drain current of Id = 100 mA. Fig. A.4 shows the gain response of the PA with 1 dB input compression point at 22 dBm and gain of about 8 dB. The signal generator generates an input signal that is a 10 MHz LTE signal with a center frequency at 3.5 GHz and peak to average power of 9.8 dB. A pre-amplifier at the output of the generator provides the necessary gain in order to get an average power of 22 dBm at each PA input.

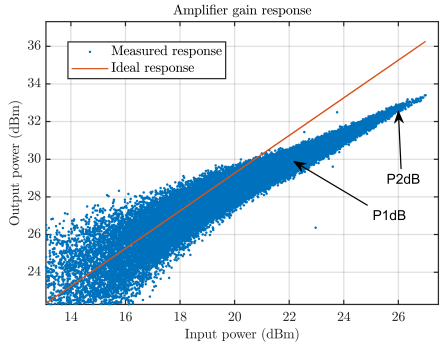


Fig. A.4: Amplifier gain response.

3.2 Single PA linearization vs system level linearization

This section demonstrates how much mutual coupling and crosstalk between antennas influence the non-linear distortion, which can not be compensated accurately by conventional single PA linearization.

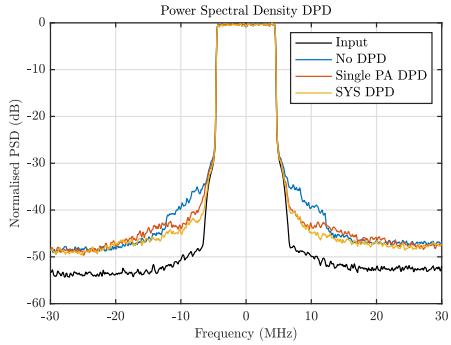


Fig. A.5: PDS for single PA DPD vs system DPD.

Measurement results are shown in Fig. A.5. In the first set of measurements each PA is linearized using a standard DPD on a single PA without connection to antenna. Then these standard DPD's coefficients are applied to the system including antenna and the measurement is done at receiver antenna. The result is called "Single PA DPD". In the second set of measurement the linearization is done based on the combined beam-form signals of all four power amplifiers at receiver antenna. The result is called "SYS DPD". As shown in Fig. A.5, the normalized Power Spectral Density (PSD) of adjacent channel gets approximately 2 dB lower with the proposed model in Fig. A.2 compare to "single

PA" model.

3.3 Impact of coupling between antennas

For investigating the impact of correlation between antennas on DPD, a set of measurements with varying space between the four transmitter antennas, d = [0.1 0.2 0.3 0.4 0.5 0.6] times wavelength (λ), has been carried out. The same quiescent drain current (100 mA) has been applied to each amplifier. Measurements with and without DPD for each spacing has been carried out and for each set of measurements the Amplitude to Amplitude (AM/AM) distortion, normalized PSD and Adjacent Channel Power Ratio (ACPR) are measured. Fig. A.6 shows that as a result of coupling between antenna the gain of array is reduced by 4 dB comparing 0.5λ distance between antennas with 0.1λ. Figures A.7 and A.8 showing an improvement of PSD and an ACPR of approximately 6 dB by doing the proposed system DPD.

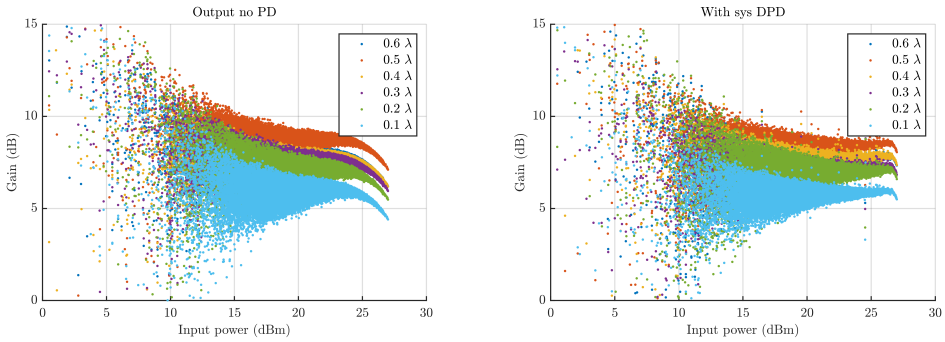


Fig. A.6: AM/AM with same drain current but different spacing between antennas.

However, the ACPR difference when going from 0.1λ to 0.5λ is only app. 1 dB. An explanation could be that in the worst case spacing (0.1λ) the coupling between antennas is not so high. For investigating the coupling rate between antennas, the s-parameters for the array with 4 antenna has been measured. The strongest coupling is approximately -12dB as it is shown in Fig. A.9. This means that in the actual setup the worst case coupling is still not significant and could not have major influence in linearization. Further investigation with stronger coupling between antennas needs to be carried out.

3.4 Impact of power variation of each branch on DPD

First set of measurements are conducted by keeping the quiescent current at all four amplifiers constant at 100 mA. This measurement is called "Same Current". A second

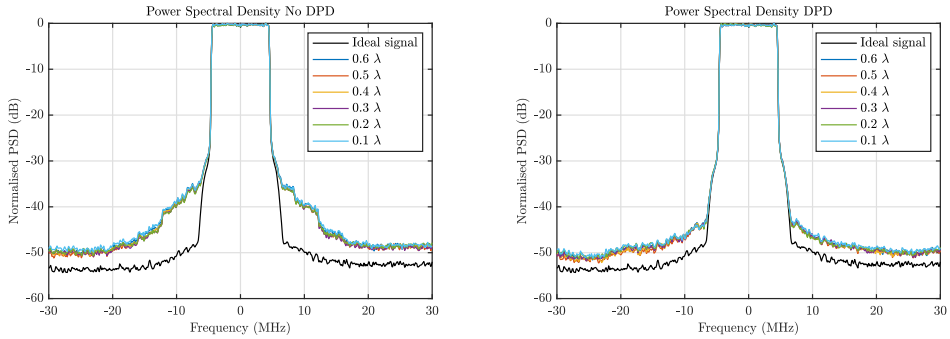


Fig. A.7: PSD with same Drain current but different distance between antennas, using system DPD.

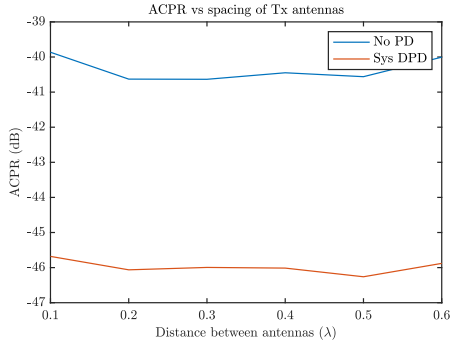


Fig. A.8: ACPR with same drain current versus distance between antennas.

set of measurements, called "Same Volt", are applied by keeping the quiescent gate voltage of amplifiers constant at -2.7 V which results in different drain current in PAs and consequently different output powers. For each set of measurements the AM/AM and normalized PSD are measured. Measurement results are shown in Fig. A.10 and Fig. A.11.

The results show that if the amplifiers with antennas are treated as a whole, then the DPD algorithm is able to improve the normalized PSD of the adjacent channel on the received beam regardless of power variation on each branch.

4 Conclusion

In this paper a DPD technique for linearizing of the antenna array in presence of crosstalk has been presented. By using the actual technique on system level it is pos-

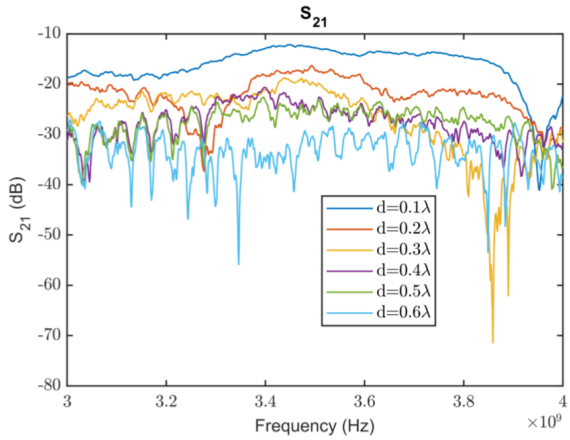


Fig. A.9: Measured S21 with four antennas.

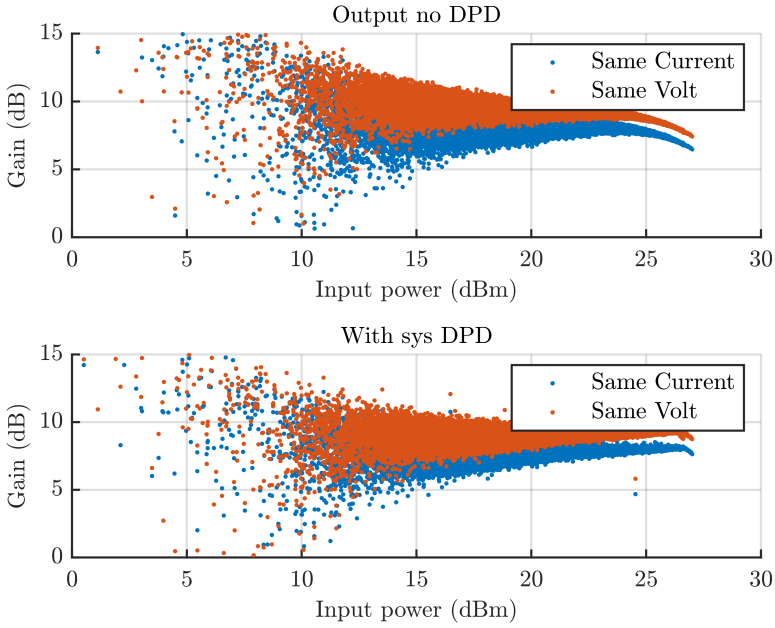


Fig. A.10: AM/AM with same drain current vs same gate voltage.

sible to reduce the adjacent channel power by about 8dB without doing any intensive

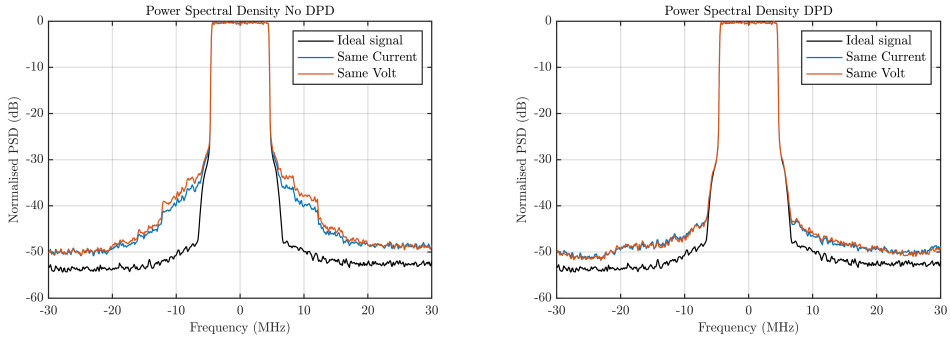


Fig. A.11: Power Spectral Density with same drain current vs same gate voltage.

S-parameter measurements needed in existing work. Additionally it has been shown that with the system level DPD it is possible to improve the overall power spectral density although the array's chains have different power levels. So treating the amplifiers and antennas as one system and linearizing the main beam signal at the receiver rather than on each single power amplifier works and individual DPD is not needed. The complexity of pre-distortion used in this work is the time alignment. To determine the time alignment needed between input and output of PA, a simulation must be done. The time alignment is then made into a phase shift between the two input signals. The difference in the phase from input signal to output signal needs be calculated and used for the rest of the simulations. The presented system level linearization technique needs be proven in other scenarios such as multi path environment and varying distance between transmitter and receiver antennas. In order to use the presented system level DPD in practice, it is required a reverse link to pass the DPD coefficients, and synchronization of the training sequence. Main difficulty is that the receiver bandwidth must be large enough to cover the adjacent channels.

References

[1] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results. *IEEE communications magazine*, 52(2):106–113, 2014.

[2] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave mimo systems. *IEEE journal of selected topics in signal processing*, 10(3):436–453, 2016.

[3] Abubaker Abdelhafiz, Laleh Behjat, Fadhel M Ghannouchi, Mohamed Helaoui, and Oualid Hammi. A high-performance complexity reduced behavioral model and digital predistorter for mimo systems with crosstalk. *IEEE Transactions on Communications*, 64(5):1996–2004, 2016.

[4] Zahidul Islam Shahin, et. al. Efficient dpd coefficient extraction for compensating antenna crosstalk and mismatch effects in advanced antenna system. Department of Electrical and Information Technology LTH, Lund University SE-221 00 Lund, Sweden, 2017.

[5] Katharina Hausmair, et. al. Digital predistortion for multi-antenna transmitters affected by antenna crosstalk. http://www.ieee.org/publications_standards/publications/rights/index.html, 2017.

[6] Mouna Ben Mabrouk, Guillaume Ferré, Eric Grivel, and Nathalie Deltimple. A novel digital post-distortion and detection technique for rf power amplifiers in cognitive radio systems. In *2015 IEEE International Wireless Symposium (IWS 2015)*, pages 1–4. IEEE, 2015.

[7] Ziv Alina and Ofer Amrani. On digital post-distortion techniques. *IEEE Transactions on Signal Processing*, 64(3):603–614, 2015.

[8] Cree. Cgh40006p datasheet. https://www.mouser.com/ds/2/90/gh40006p-876367.pdf, 2019.

Paper B

Linearization Trade-offs in a 5G mmWave Active Phased Array OTA Setup

Feridoon Jalili, Felice F. Tafuri, Ole k. Jensen, Yunfeng Li, Ming Shen, Gert F. Pedersen

The paper has been published in the
IEEE Access Year: 2020 | Volume: 8 | Journal Article.

Abstract

The new generation of 5G mobile communication systems is using millimeter wave (mmWave) active phased arrays (APA) which have up to hundreds of individual analog transmitter and receiver chains and antennas. For these highly integrated systems linearization of each analog path is very challenging. Therefore a single input single output (SISO) system in combination with over the air (OTA) measurement is considered as an efficient approach for linearization. However, the knowledge about the dependency of the total SISO nonlinearity on the contributions from different blocks in the antenna array, as well as the linearization trade-offs is still missing.

In this paper, an overview of the possible linearization trade-offs in an OTA setup with a mmWave APA is provided. The linearization technique is applied to a 4x4 active phased array containing up-conversion of a sub 6 GHz LTE10 signal to an RF frequency of 28 GHz. Through measurements, the effects on adjacent channel power ratio (ACPR) and error vector magnitude (EVM) have been investigated for the following scenarios: i. impact from the up-converter, ii. impact of the steering angle due to antenna crosstalk and iii. a linearity comparison between a linearized and a backed-off system.

1 Introduction

For modern communication systems, high power efficiency is required while maintaining linear operation to meet stringent spectral requirements. For the 5th generation of mobile communication and inter-satellite communication highly integrated beam-steerable active arrays consisting of a large number of PAs and antennas are considered as an efficient solution to fulfill the new requirements [1].

Microwave power amplifiers can achieve a higher efficiency in terms of transmitted power vs. supplied power if driven as close as possible to the saturation point [2]. Unfortunately the more the PA approaches saturation, the more it behaves nonlinearly and it does not fulfill the linearity requirements dictated by the mobile communication standard. Therefore a linearity and efficiency trade-off arises which RF engineers have to deal with if they want to have the PA working with a reasonable efficiency. Back-off strategies can be used to avoid nonlinear effects: the dynamic range of the amplifier's input signal is shifted down to a lower power level so that the amplifier's output is not severely distorted by the nonlinear behavior. A solution to avoid the drastically low power efficiency in PAs using back-off is the recourse to a linearization technique. Different existing linearization methods are able to reduce the nonlinear distortions while keeping the PA as efficient as possible [3]. Digital pre-distortion (DPD) has been widely used for improvement of transmitter efficiency but generally on a single power amplifier and single antenna [4]. The new generation of active arrays considered for 5G mobile communication, is using a set of highly integrated active arrays, as illustrated in

Fig.B.1. The active array topology where the PAs are placed after the phase shifters, and just before the antennas, gives several benefits such as power dissipation in and required power handling capability of the phase shifters, reducing the output power requirements for each element and allowing small integrated devices to be used while connected to the antennas. However, the increased complexity in the active phased array also makes it difficult to have a comprehensive understanding of the key factors contributing to the total nonlinearity of the whole array. Lacking this knowledge significantly limits the potential of the SISO linearization technique. The challenges responsible for the situation are as follows:

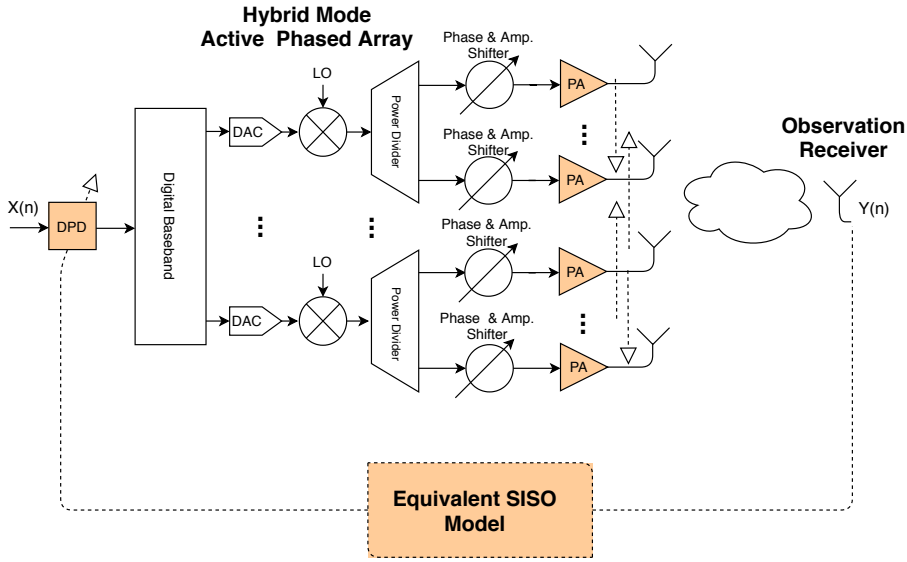


Fig. B.1: Concept illustration of the digital pre-distortion for hybrid beam-forming based on the equivalent SISO model.

1. 5G systems are using the so-called hybrid beam-forming (Fig. B.1) where the number of analog RF chains is higher than the number of digital receivers (for current consumption reductions). In such case a direct digital control of each analog RF chain is not possible and the system may not directly know the output of each amplifier. So an alternative linearization method is needed.

2. The high level of integration and large number of analog chains make the placement of feedback circuits for each branch, which is used for single antenna DPD, very challenging.

3. The signal bandwidth is rapidly increasing and it will place enormous demands on bandwidth of the analog feedback receiver used for linearization as well as sampling rate of the analog to digital converters. The feedback receiver should have a wider bandwidth than a standard receiver in order to record the distortion side-bands.

4. Using multiple transmitters and multiple antennas introduces crosstalk at PA's inputs and outputs. For mitigating the impact of these crosstalks the complexity of the algorithm is expected to increase [5] and avoiding a complexity explosion of the algorithm is another challenge.

In this paper insight into the linearization mechanisms in a 5G millimeter wave (mmWave) active phased array (APA) over the air (OTA) setup is provided through measurements. It is shown that the DPD model based on the signal captured by a single observation receiver in the far-field is able to linearize a set of PAs in an active array. The measurement set-up in this work includes an up-conversion from sub-6 GHz to mmWave which is also the general approach used by 5G manufactures in order to reuse the existing technology for mmWave. The trade-off analysis in this paper treats the following cases:

- Impact of up-conversion from sub-6 GHz into mmWave on linearization of APA.
- Impact of steering angle on trained beam in boresight.
- Comparison between a linearized SISO and a backed-off system.

This paper is organized as follows: Section 1 is the introduction. Section 2 presents active array linearization topologies in state-of-the-art solutions. Section 3 presents crosstalk mitigation methods. Section 4 describes linearization of an active phased array. Measurement results of a 4x4 array as a two-port system are provided in section 5. Finally, the conclusion of this work is presented in section 7.

2 Active Array Linearization topologies, state-of-the-art solutions

A simple and most cost effective solution is to observe only the output of a single PA and assume that all PAs are similar which is presented in [6] and [7]. The drawback of this approach is a reduced performance.

An alternative scheme is to use an observation receiver per PA and linearize according to some averaging principle [8]. This approach is an expensive approach and requires as many observation receivers as PAs.

Another method presented in [9], [10] and [11], suggesting that the output of individual branches be combined and sampled for linearization in order to include the actual

performance of each PA branch. Same approach has been suggested by [12] and [13], where they combine all branches and establish a "virtual" and "rotated" main beam. The authors show that by linearizing with a rotated observation signal, the distortion in the direction of the main beam is minimized. The method of combining the outputs requires a feedback signal from each PA output. Although this method demonstrates significant linearity improvement, it still needs bulky feedback circuits which may be impractical when using highly integrated active arrays in 5G mmWave.

Recently researches have presented the idea of SISO modelling where the entire transmitter has been considered as a two port system as illustrated in Fig. B.1. We have presented a DPD technique for linearization of the antenna array in presence of crosstalk [14], using only one external observation antenna for observing the combined signal in the far field. A similar approach has been introduced by [15] and [16]. In practice, this can be implemented as part of the receiver section of the same device (i.e. diversity receiver) which has been presented by [17]. This kind of adaptive on-line OTA-DPD based on the diversity feedback uses an iterative procedure to eliminate the uncorrelated components from the feedback signal for accurate DPD. The concept has been verified by measurements with good results for small scale arrays but for large scale arrays only simulation results are available. The adaptive on-line DPD is a promising approach and needs to be investigated in more detail by industry and academia in the next years. Nevertheless, investigation of replacing the feedback antenna with a far-field observation receiver in order to analyse the impact of load modulation due to crosstalk is an important topic for the linearization approach and has been investigated in [18].

3 Crosstalk mitigation, state-of-the-art solutions

The topologies described in 2 explain mainly how the response of the amplifiers is measured for DPD implementation and describe the trade-off of the systems in terms of efficiency, cost and size. The discussion about criteria by which the DPD algorithm is optimized in order to mitigate the crosstalk is another important topic. Crosstalk as coupling from one branch to another, in transmitters in an active antenna array can be categorized as two types: before PA and after PA. This is mainly due to RF leakage through the common local oscillator or coupling between different transmit paths because of the electromagnetic coupling, respectively [19].

When no isolators are present at the PAs' outputs then the PAs get a direct impact from antenna mismatch and the mutual coupling between the antennas [20]. For most practical configurations, mutual coupling is difficult to predict analytically but must be taken into account because of its significant contribution [21]. As a consequence of antenna mismatch and mutual coupling at the PA to antenna interface, signals will travel in both directions and the RF behavior of the PA must be described by a dual input dual output behavioral model. Fig. B.2 shows an RF beam former including L antenna elements where a_{1k} is the incoming signal to the amplifier, b_{2k} is the output

from the amplifier and a_{2k} is the reflected signal from the antenna array at the k'th branch.

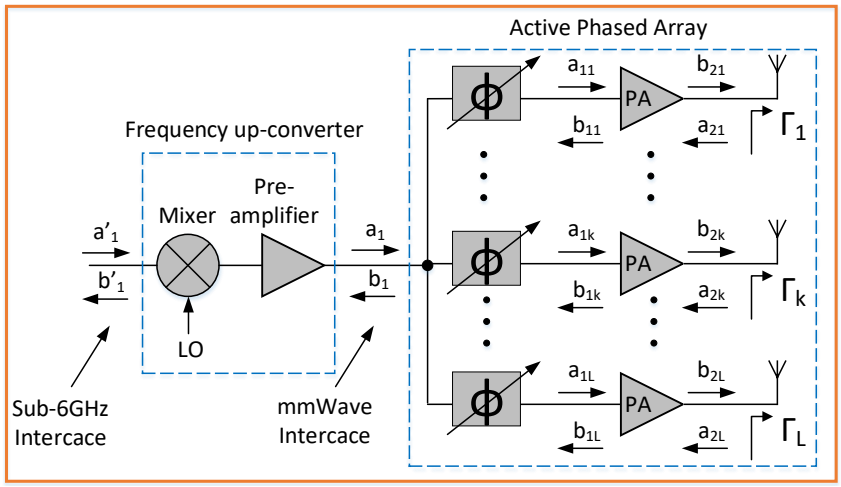


Fig. B.2: Conceptual illustration of mutual coupling and reflection coefficients of system model.

Since the array elements are electromagnetically coupled, the waves fed to the antennas are also coupled back to the output ports of the PAs. This effect creates an apparent variable load at the output of each PA, depending on the operation of the transmitter [22]. Then the effective reflection coefficient (Γ_k) for the k'th element is given by:

$$\begin{bmatrix} a_{21} \\ \cdots \\ a_{2k} \\ \cdots \\ a_{2L} \end{bmatrix} = \begin{bmatrix} s_{11} & \cdots & s_{1k} & \cdots & s_{1L} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{k1} & \cdots & s_{kk} & \cdots & s_{kL} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{L1} & \cdots & s_{Lk} & \cdots & s_{LL} \end{bmatrix} . \begin{bmatrix} b_{21} \\ \cdots \\ b_{2k} \\ \cdots \\ b_{2L} \end{bmatrix} \tag{B.1}$$

$$a_{2k} = \begin{bmatrix} s_{k1} & \cdots & s_{kk} & \cdots & s_{kL} \end{bmatrix} . \begin{bmatrix} b_{21} \\ \cdots \\ b_{2k} \\ \cdots \\ b_{2L} \end{bmatrix} \tag{B.2}$$

$$\Gamma_k = \frac{a_{2k}}{b_{2k}} = \frac{\sum_{i=1}^{L} S_{ki} b_{2i}}{b_{2k}} \tag{B.3}$$

Where elements in the b_2 vector are the complex coefficients describing the input to the antenna and elements in the S_k vector are the complex coefficients describing the relationship between a_{2k} and the output signal b_{2k} and determined by the characteristics of the antenna array. Since (Γ_k) is not only dependent on the reflection from the k'th element but also on the coupling from other elements in the array, then the output impedance of the k'th element is changing and as a result the linear and nonlinear behavior of it, the so-called load modulation.

Furthermore, the behavior of each PA cannot be fully described solely as a function of its input, it will change according to the coupled signal. The input signal a_{1k} is ideally a phase shifted version of the input signal a_1. But due to gain variation of the phase shifters over phase shift setting and impact of the reflected signal b_{1k}, each PA can be driven at a different input levels for different steering angles [22].

To account for both the load modulation and steering angle dependency for each PA in the array, the dual-input PA model has been introduced in [23] and [24] where both signals, a_{1k} and a_{2k} are included in the nonlinear function. These works demonstrated good results by applying DPD to the beam forming array. However, the method requires that the s-parameters for the antenna are known and it needs feedback from each antenna. The high number of PAs and the compact size of modules makes this approach very challenging.

We have in [14] presented a system level SISO DPD technique for linearization of the antenna array in presence of crosstalk and similar work has been presented in [15]. The reported measurement results are limited to implementations at sub-6 GHz and small array sizes, and challenges specific to DPD when applied to mmWave arrays still remain.

As mentioned in section 1, a general approach used by 5G manufactures is to use frequency up-conversion from sub-6 GHz. As illustrated in Fig. B.2 the input signal to the APA, a_1, is ideally a frequency up-converted part of the sub-6 GHz signal a_1'. But in reality there would be mismatch on output and input of the frequency up-conversion block due to reflected signals, b_1 and b_1', as well as distortion in the up-conversion mixer and pre-amplifier. Including the impact of these blocks to the system model makes the DPD algorithm more complex. In this work the impact of these blocks to the applied DPD is evaluated through measurements.

4 Linearization of Active Phased Array

In this section the linearization method using the radiated far-field signal of an active array at a single observation receiver is described. The assumption is that the nonlinearity of the PA is the main source of distortion and not the crosstalk. Equation (B.4) represents the applied memory polynomial model (MPM) which is a deviation of the Hammerstein model and has been proven effective for removing nonlinearity and

memory effect [25]:

$$y(n) = \sum_{k=1}^{K} \sum_{m=0}^{M} a_{km} x(n-m) |x(n-m)|^{k-1} \tag{B.4}$$

Where a_{km} is the 2-D array of filters and power series coefficients of the amplifier, K is nonlinearity order of the memory polynomial and M is the highest memory depth.

Since the a_{km} coefficients are linear weighting of nonlinear signals then these coefficients can be found using the least-square type algorithm. The easiest way to formulate such an algorithm is to first collect the coefficients in a J x 1 vector denoted ω, where J is the total number of coefficients. Then the model output can be expressed using the following equation written in vector form.

$$\widetilde{y} = P\omega \tag{B.5}$$

where:

- $\widetilde{y}$ is a N x 1 vector representing an estimate of the amplifier actual output.
- P is a N x J matrix where N is the number of samples and J is equal to M times K .
- ω is a vector with J x 1 coefficients.

The inverse of this model used for pre-distortion is then:

$$\widetilde{x} = Rw \tag{B.6}$$

where R is defined similarly to P now with $y(n-m)$ replacing $x(n-m)$ in equation (B.4). The input is now estimated from the output samples and the estimation error can be calculated as:

$$e = x - \widetilde{x} \tag{B.7}$$

The best estimate for getting a_{km} coefficients is to use a least square solution which minimizes the squared error:

$$w = (R^H R)^{-1} R^H x \tag{B.8}$$

In this work the parameter extraction was performed using the Moore-Penrose pseudo-inverse because this technique provides a more robust solution to the system and avoids instability in parameter extraction due to the eventually high condition number of the model matrix [3].

5 Measurement Results of 4x4 Array as two-ports system

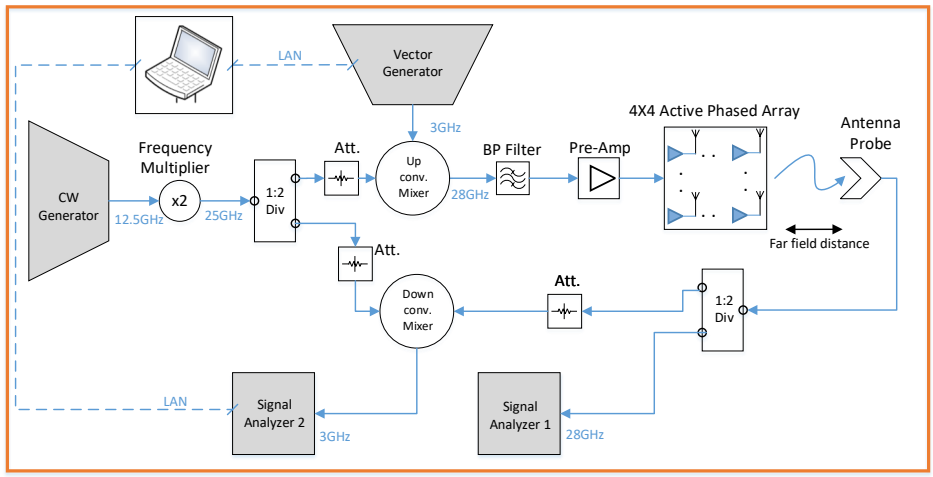


Fig. B.3: Block diagram of OTA SISO memory polynomial model based digital pre-distortion for the 28 GHz active phased array.

5.1 Measurement setup

The block diagram of the measurement setup for the 4x4 array is shown in Fig. B.3 and the actual measurement set-up is illustrated in Fig. B.4. The input source for the measurements is a 3 GHz LTE10 signal, compliant with the 3GPP downlink orthogonal frequency-division multiplexing (OFDM) modulation with a peak to average power ratio of 10.6 dB from the signal generator. For up-conversion, an unmodulated signal of 12.5 GHz has been frequency doubled to 25 GHz and fed into a power divider in order to be used as local oscillator (LO) signal for both up-conversion and down-conversion. A 28 GHz band-pass filter is used to select the up-converted modulated signal and suppress the LO leakage and image frequency signals. In order to avoid any nonlinearity in the multiplier and up-converter, the signal levels in these stages are kept in the linear operating ranges of these devices according to their specifications and then amplified by a pre-amplifier in order to reach the level necessary for driving the active array into compression. The pre-amplifier has a gain of 30 dB and at output levels above 8 dBm (including cable loss), it deteriorates the linearity of the signal which is shown in the

measurement result in next section. The 28 GHz signal is fed to an AMOTECH A0404 which includes four Anokiwave AWMF-0158 [26]. This device integrates 16 branches of attenuators and phase shifters plus PAs and 16 patch antennas in a 4x4 active phased array.

The diameter of the active array antenna is approximately 4 cm which at 28 GHz results in to a far-field distance of:

$$\frac{2D^2}{\lambda} = 30.5 \quad cm \tag{B.9}$$

Where D is the diameter of the antenna and λ is the wavelength. The main beam signal is captured by the observation horn antenna placed 44 cm away which is well above the far-field distance of the device.

Problems with reflections from the surroundings are not observed since the distance between the active array and the horn antenna is short, so the reflected signal from the rack and the set-up are much weaker than the desired signal. Furthermore the active array and horn antenna have very good directivity so reflections are expected to be well attenuated. Therefore measurement without an anechoic chamber is expected to be no issue.

The captured signal is split into two branches in order to both be analysed with the signal analyzer for monitoring the actual adjacent channel power ratio (ACPR) at 28 GHz and be down-converted to a 3 GHz signal and captured by another signal analyser for getting access to I and Q data. The input power from the signal generator is adjusted in order to get an root mean square (RMS) level up to 8 dBm into AWMF-0158 which according to [26] drives the PAs into compression.

The steps of the experiment are:

1. I and Q data of the modulated signal from the vector signal generator were acquired using a sample rate of 100 MHz at the signal analyzer and recorded by a computer.

2. The recorded I and Q data are loaded into the vector signal generator, generating a modulated signal at 3 GHz which is then up-converted to 28 GHz. The active array is excited using this modulated signal and the output, $y(n)$, is captured at the observation point by the receiver antenna probe. This signal is then down-converted and acquired by the signal analyzer and the I and Q data recorded into the computer.

3. The input and the recorded output signals are up-sampled to a finer resolution, time-aligned using cross-correlation, down-sampled and then used for the predistorter identification.

4. The pre-distorted signal is now generated by the memory polynomial with a memory depth of M = 8 and a linearity order of N = 5 using recorded $x(n)$ and $y(n)$.

5. Once the predistorter is identified, the signal is uploaded to the vector signal generator and again applied to the APA. The corresponding output signal is recorded and the power spectral density (PSD) and error vector magnitude (EVM) are calculated based on the recorded I and Q samples.
6. Output power and ACPR of the 28 GHz signal are directly measured by the signal analyzer.

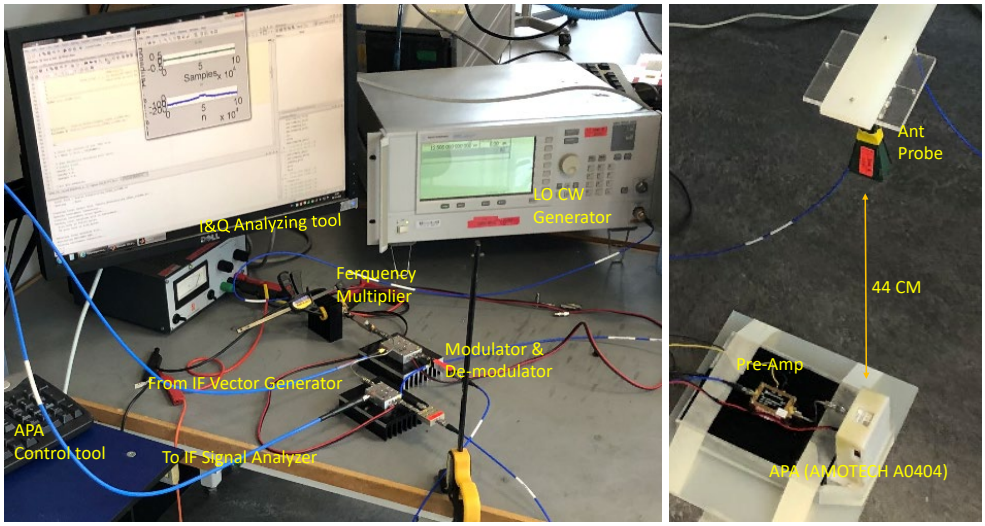


Fig. B.4: OTA SISO measurement setup of the 4x4 active phased array; using up-conversion from 3 GHz to 28 GHz.

5.2 Measurement results

The measurement has been done in 3 cases: APA input at 6 dBm, 8 dBm and 10 dBm respectively. According to the APA data sheet it has a 1 dB compression point at around 8 dBm input. Table B.1 shows the EVM and ACPR measurement results with and without DPD for these 3 cases. The best result is achieved in case 2 where ACPR and EVM are improved by respectively 7 dB and 2.3 % using the applied DPD. As expected the linearity improvement for case 1 with 6 dBm input is not as good as case 2 because the APA is not enough in compression. Fig. B.5 and Fig. B.6 show the measurement results of PSD and amplitude to amplitude (AMAM) distortion gain for these 3 cases. The small linearity improvement in case 3 could be explained due to nonlinearity of pre-APA blocks.

Table B.1: EVM and ACPR measurement results without and with DPD.

Measurements	EVM[%] (wo/w)	ACPR[dB] Lower(wo/w)	ACPR[dB] Upper(wo/w)
case 1	6.21/5.10	-40.03/-43.84	-40.11/-43.53
case 2	6.21/3.96	-38.27/-45.19	-38.08/-44.94
case 3	7.20/5.84	-36.60/-42.48	-36.69/-41.87

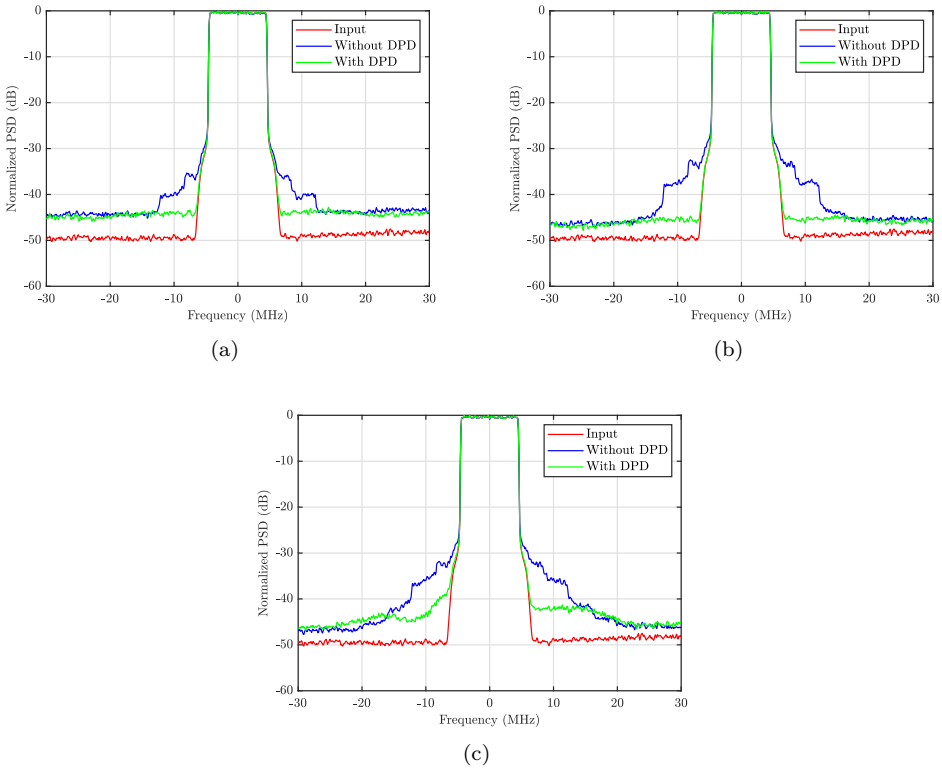


Fig. B.5: Measured power Spectral density with and without pre-distortion: (a) case 1 with input power of 6 dBm; (b) case 2 with input power of 8 dBm; (c) case 3 with input power of 10 dBm.

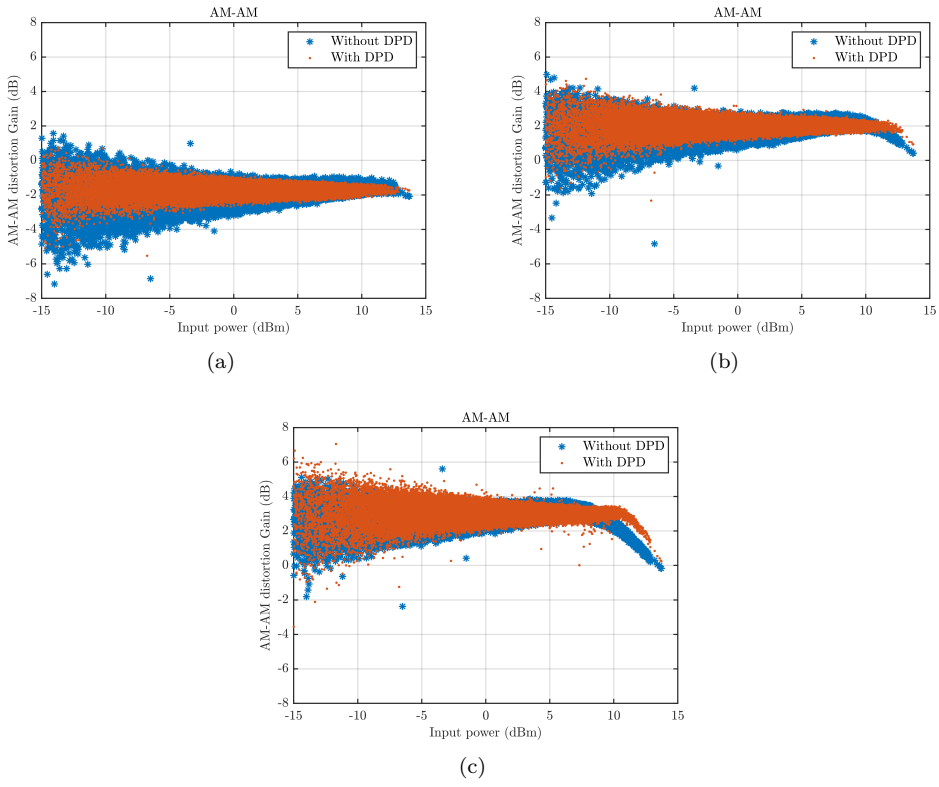


Fig. B.6: Measured AM-AM distortion gain with and without pre-distortion: (a) case 1 with input power of 6 dBm ; (b) case 2 with input power of 8 dBm; (c) case 3 with input power of 10 dBm.

5.3 Impact of up-conversion from sub-6 GHz into mmWave on linearization of APA

The linearity of the up-conversion blocks are expected to have an impact on SISO DPD since the nonlinearity could partly be from a pre-APA block if it is in compression.

To investigate the pre-APA blocks nonlinearity, the ACPR of the system after the up-conversion mixer and after the pre-amplifier are measured and the PSD plots are shown in Fig. B.7 and Fig. B.8 respectively and in tables B.2 and B.3 the EVM and ACPR results are listed.

The results indicate that the signal is linear after the up-conversion mixer but the pre-amplifier is running into high compression in case 3, and as a consequence the ACPR is increased. The applied DPD is not capable to mitigate the nonlinearity caused by

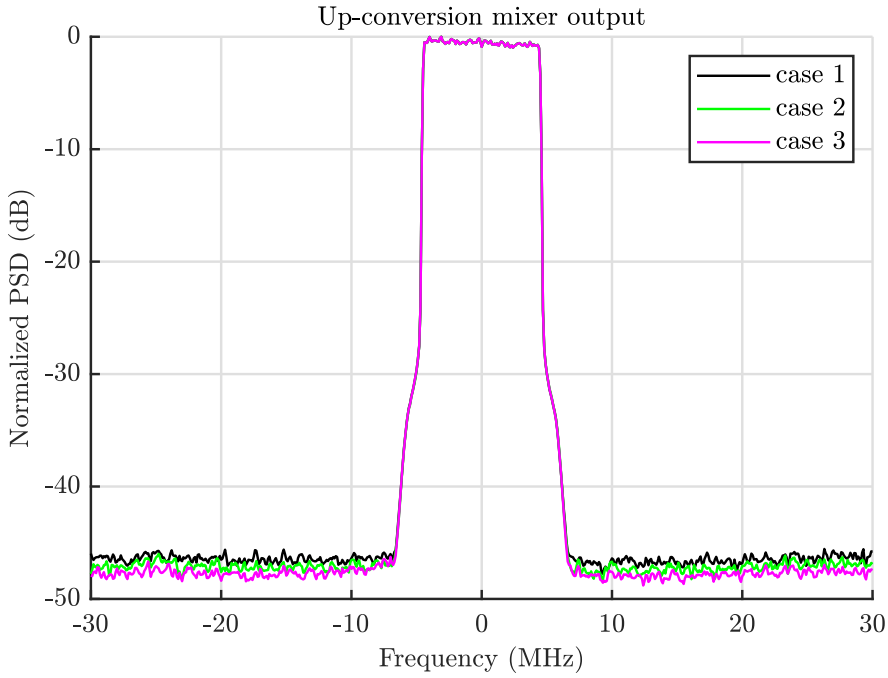


Fig. B.7: Measured ACPR at the up-conversion Mixer out: case 1 (black) corresponds to 6 dBm at the APA input; case 2 (green) corresponds to 8 dBm at the APA input; case 3 (magenta) corresponds to 10 dBm at the APA input.

Table B.2: EVM and ACPR measurement results at the up-conversion Mixer's output.

Measurements	EVM[%]	ACPR[dB]	
		Lower	Upper
case 1	0.89	-46.06	-46.35
case 2	0.89	-46.68	-47.01
case 3	0.89	-47.01	-47.50

pre-amplifier in case 3 properly. One reason could be a time delay between pre-amplifier and APA which needs further investigations.

5.4 Impact of beam angle

In section 3 the impact of mutual coupling between the antennas at PAs' outputs has been discussed and it was mentioned that the impedance presented at the output of

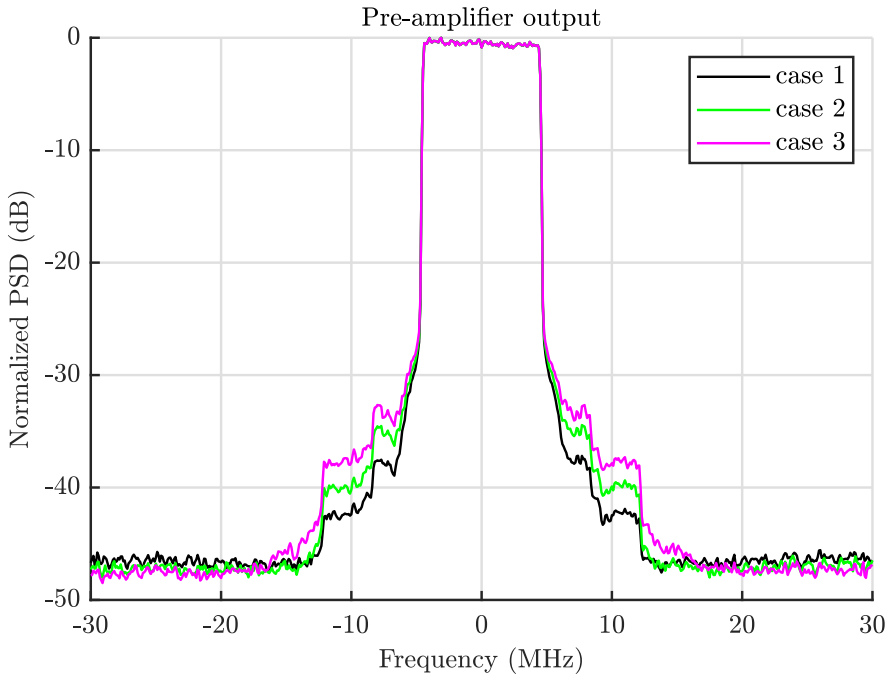
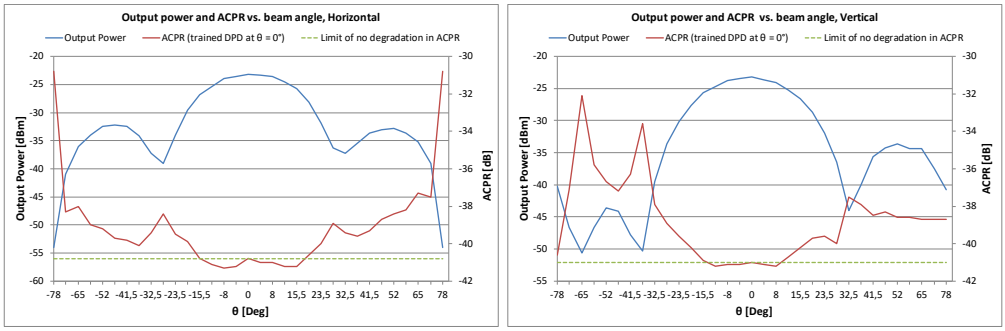


Fig. B.8: Measured ACPR at the pre-amplifier out: case 1 (black) corresponds to 6 dBm at the APA input; case 2 (green) corresponds to 8 dBm at the APA input; case 3 (magenta) corresponds to 10 dBm at the APA input.

Table B.3: EVM and ACPR measurement results at the pre-amplifier output.

Measurements	EVM[%]	ACPR[dB]	
		Lower	Upper
case 1	0.89	-42.23	-42.29
case 2	0.90	-40.51	-40.52
case 3	0.89	-38.50	-38.60

each PA depends on the mutual coupling between antennas. In this section this impact is proven through measurements. The procedure is as follow: the placement of the observation receiver antenna has been kept fixed at maximum received signal (θ = 0 degree). The DPD has been trained and the pre-distorted input has been detected at this position. Then while using this pre-distorted input, the main beam of the APA has been shifted from θ = -78 to +78 degrees in approximately 5 degrees step using



(a) (b)

Fig. B.9: Impact of beam angle on linearization, (a) horizontal beam steering, and (b) vertical beam steering.

the code-book and software tools of AMOTECH A0404. The measurement is done for both horizontal and vertical steering angles of the main beam. The measurement result is shown in Fig. B.9. The magnitude of the beam captured by the fixed antenna probe is varying by changing the beam direction as expected. Fig. B.9b shows the magnitude and ACPR on the left side are worse than on the right side. This is due to the asymmetrical structure of the actual AMOTECH A0404 device where the placement of the connector has an influence on the beam and is not a general issue.

However a single trained DPD is not sufficient for maintaining a low ACPR in a wide range of steering angles. To maintain an ACPR level below 41 dBc across the steering angle, a new training after approximately ±15 degree shift of the main beam is required. This can be explained as the effect of mutual coupling of the highly integrated antennas in the array and due to variation in input levels because of gain variation of phase shifters. Fig. B.9 also shows that there is a symmetry of the array and the trained DPD at $+\theta$ can be used for a steering angle of $-\theta$ which reduces the number of training steps required for linearization across the entire steering range.

5.5 Comparison between linearized SISO and backed-off system

In order to minimize the complexity it is desired to keep the trained pre-distorted signal as input for a range of output powers if possible. At the same time it is desired that the linearized system has not degraded performance compared to a backed-off system around the back-off point. To investigate this, the trained pre-distorted signals in case 1 and case 2 of section 5 have been reused for a set of output powers of ±6 dB and the resulting ACPRs are compared with the case without pre-distortion. As it is shown in Fig. B.10, the linearity compared to a backed-off system, is always better for output

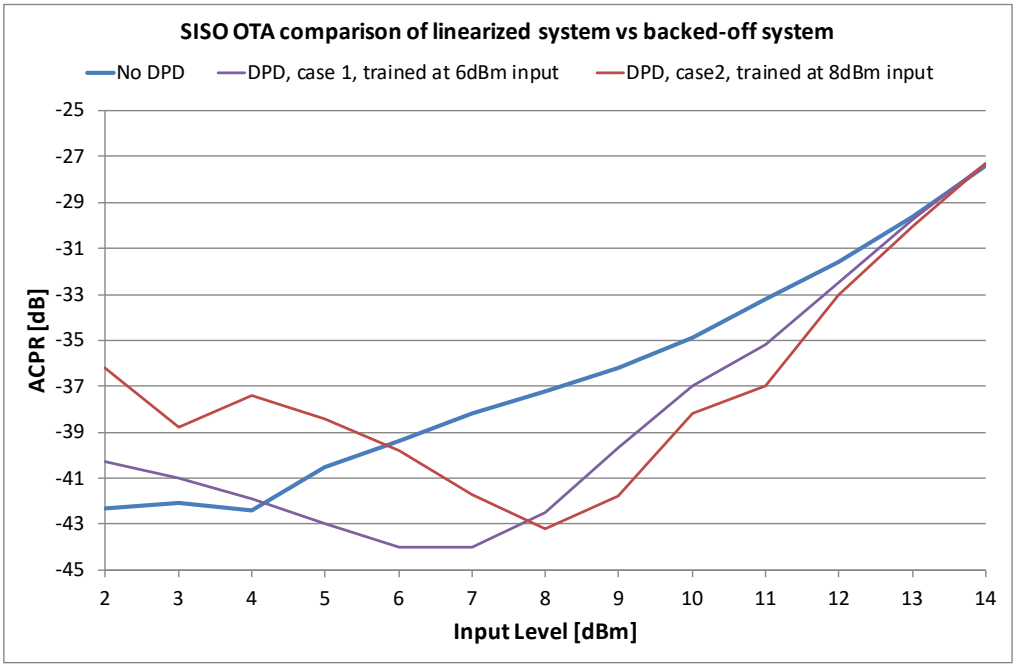


Fig. B.10: Comparison of a linearized system to a backed-of system, in terms of ACPR, at varying input back-off levels.

powers above the trained point and down to 2 dB below this point.

6 Conclusion

In this paper the SISO OTA linearization trade-offs of a 4x4 active array running at 28 GHz, modulated with an LTE10 DL OFDM signal have been investigated. The results indicate:

i. there is a trade-off between the complexity of the applied DPD algorithm and the linearity of the pre-APA blocks under test. Increasing the gain of these pre-APA blocks, which can affect their linearity, will limit the capability of a less complex DPD algorithm. With the applied OTA DPD, up to 7dB improvement of ACPR and 2.3 % improvement of EVM are achieved with minimum complexity of the algorithm.

ii. the linearized beam is sensitive to beam angle. The trained beam cannot be reused for beam angles above a certain limit which could be explained due to mutual coupling and crosstalk between antennas and due to variation in input levels. A new trained beam is required in order to avoid ACPR degradation when changing beam

steering angle. In this investigation, for maintaining the ACPR level across the steering angle, a new training after approximately ±15 degree shift of the main beam is required.

iii. near the back-off region, reusing a set of trained coefficients, the linearized system is out-performing the backed-off system from above the trained output level and down to a certain level below that. The achieved result in this work is 2dB below the trained level.

Future work based on the presented SISO OTA technique may include complexity analysis of the DPD algorithm in the case of distorted pre-APA blocks and the impact of channel properties on the SISO OTA linearization.

References

[1] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results. *IEEE communications magazine*, 52(2):106–113, 2014.

[2] S Crips. RF power amplifiers for wireless communications, artech house, 1999.

[3] Felice Francesco Tafuri, Cataldo Guaragnella, Marco Fiore, and Torben Larsen. Linearization of RF power amplifiers using an enhanced memory polynomial predistorter. In *NORCHIP 2012*, pages 1–4. IEEE, 2012.

[4] Fadhel M Ghannouchi, Mayada Younes, and Meenakshi Rawat. Distortion and impairments mitigation and compensation of single-and multi-band wireless transmitters. *IET Microwaves, Antennas & Propagation*, 7(7):518–534, 2013.

[5] Mahmoud Alizadeh, Peter Händel, and Daniel Rönnow. Behavioral modeling and digital pre-distortion techniques for RF pas in a 3× 3 mimo system. *International journal of microwave and wireless technologies*, 11(10):989–999, 2019.

[6] Ling Liu, Wenhua Chen, Liyong Ma, and He Sun. Single-pa-feedback digital predistortion for beamforming mimo transmitter. In *2016 IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, volume 2, pages 573–575. IEEE, 2016.

[7] Han Yan and Danijela Cabric. Digital predistortion for hybrid precoding architecture in millimeter-wave massive mimo systems. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3479–3483. IEEE, 2017.

[8] Sangil Lee, Minhyun Kim, Youngwook Sirl, Eui-Rim Jeong, Soonil Hong, Seongjin Kim, and Yong H Lee. Digital predistortion for power amplifiers in hybrid mimo systems with antenna subarrays. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2015.

[9] Xin Liu, Qian Zhang, Wenhua Chen, Haigang Feng, Long Chen, Fadhel M Ghannouchi, and Zhenghe Feng. Beam-oriented digital predistortion for 5g massive mimo hybrid beamforming transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(7):3419–3432, 2018.

[10] Nuutti Tervo, Janne Aikio, Tommi Tuovinen, Timo Rahkonen, and Aarno Parssinen. Digital predistortion of amplitude varying phased array utilising over-the-air combining. In *2017 IEEE MTT-S International Microwave Symposium (IMS)*, pages 1165–1168. IEEE, 2017.

[11] Sungho Choi and Eui-Rim Jeong. Digital predistortion based on combined feedback in mimo transmitters. *IEEE communications letters*, 16(10):1572–1575, 2012.

[12] Mahmoud Abdelaziz, Lauri Anttila, Alberto Brihuega, Fredrik Tufvesson, and Mikko Valkama. Digital predistortion for hybrid mimo transmitters. *IEEE Journal of Selected Topics in Signal Processing*, 12(3):445–454, 2018.

[13] Chan-Won Park, Eui-Rim Jeong, and Ji-Hoon Kim. A new digital predistortion technique for analog beamforming systems. *IEICE Electronics Express*, 13(2):20150998–20150998, 2016.

[14] Feridoon Jalili, Martin H Nielsen, Ming Shen, Ole K Jensen, Jan H Mikkelsen, and Gert F Pedersen. Linearization of active transmitter arrays in presence of antenna crosstalk for 5g systems. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, pages 1–5. IEEE, 2019.

[15] Katharina Hausmair, Ulf Gustavsson, Christian Fager, and Thomas Eriksson. Modeling and linearization of multi-antenna transmitters using over-the-air measurements. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE, 2018.

[16] Alberto Brihuega, Mahmoud Abdelaziz, Lauri Anttila, Matias Turunen, Markus Allén, Thomas Eriksson, and Mikko Valkama. Piecewise digital predistortion fo mmwave active antenna arrays: Algorithms and measurements. *arXiv preprint arXiv:2003.06348*, 2020.

[17] Xin Liu, Wenhua Chen, Long Chen, Fadhel M Ghannouchi, and Zhenghe Feng. Linearization for hybrid beamforming array utilizing embedded over-the-air diversity

feedbacks. *IEEE Transactions on Microwave Theory and Techniques*, 67(12):5235–5248, 2019.

[18] Eric Ng, Yehia Beltagy, Giovanni Scarlato, Ahmed Ben Ayed, Patrick Mitran, and Slim Boumaiza. Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets. *IEEE Transactions on Microwave Theory and Techniques*, 67(11):4479–4492, 2019.

[19] Abubaker Abdelhafiz, Laleh Behjat, Fadhel M Ghannouchi, Mohamed Helaoui, and Oualid Hammi. A high-performance complexity reduced behavioral model and digital predistorter for mimo systems with crosstalk. *IEEE Transactions on Communications*, 64(5):1996–2004, 2016.

[20] Zahidul Islam Shahin, et. al. Efficient dpd coefficient extraction for compensating antenna crosstalk and mismatch effects in advanced antenna system. Department of Electrical and Information Technology LTH, Lund University SE-221 00 Lund, Sweden, 2017.

[21] Constantine A Balanis. *Antenna theory: analysis and design*. John wiley & sons, 2016.

[22] Filipe M Barradas, Pedro M Tomé, José M Gomes, Telmo R Cunha, Pedro M Cabral, and José C Pedro. Power, linearity, and efficiency prediction for mimo arrays with antenna coupling. *IEEE Transactions on Microwave Theory and Techniques*, 65(12):5284–5297, 2017.

[23] Christian Fager, Thomas Eriksson, Filipe Barradas, Katharina Hausmair, Telmo Cunha, and Jose Carlos Pedro. Linearity and efficiency in 5g transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5g active antenna transmitter context. *IEEE Microwave Magazine*, 20(5):35–49, 2019.

[24] Katharina Hausmair, Per N Landin, Ulf Gustavsson, Christian Fager, and Thomas Eriksson. Digital predistortion for multi-antenna transmitters affected by antenna crosstalk. *IEEE Transactions on Microwave Theory and Techniques*, 66(3):1524–1535, 2017.

[25] Dennis R Morgan, Zhengxiang Ma, Jaehyeong Kim, Michael G Zierdt, and John Pastalan. A generalized memory polynomial model for digital predistortion of RF power amplifiers. *IEEE Transactions on signal processing*, 54(10):3852–3860, 2006.

[26] Anokiwave. Awmf-0158 28 ghz silicon 5g tx/rx quad core ic. `https://www.anokiwave.com/products/awmf-0158/index.html`, 2019.

Paper C

Antenna Array Inter-Element Coupling impact on Linearization of Active Phased Array

Feridoon Jalili, Daniel E. Serup, Ondrej Franek, Ming Shen and Gert F. Pedersen

The paper has been published in the
International Symposium on Networks, Computers and Communications (ISNCC)
Year: 2021 | Conference Paper | Publisher: IEEE.

Abstract

When moving to high integration scenarios in the mmWave active phased array (APA), the isolation between the array and the power amplifiers (PAs) is often removed. Since the array elements are electromagnetically coupled, the waves fed to the antennas are also driving the output ports of the PAs. This effect creates a possible variable load at the output of each PA, depending on steering angle and radiated power at each element in the array. In this system, the behavior of each PA cannot be fully described solely as a function of its input, as it will change according to the coupled signal. In this paper the inter-element coupling of a 4x4 planar patch antenna array and the resulting variation in output impedance of the PA connected to each patch is simulated by using computer simulation tool (CST). The impact of the inter-element coupling has been verified on linearization procedure of the active array.

1 Introduction

For radio links, high power and linear operation are required to achieve high data-rates while maintaining spectral efficiency and low current consumption. To fulfill the latest requirements for high data rate in the 5th generation of mobile and satellite communication, integrated steerable active phased array is a preferred solution [1].

In order to improve the efficiency of the active array, PA needs to operate as close to saturation region as possible. For achieving optimum output power and efficiency, output impedance of the the PA is not necessarily 50 ohm but needs be matched to optimum impedance [2]. In active array each PA is directly connected to an antenna without any isolator in between. Therefore the PAs' outputs impedance get a direct impact from antenna mismatch and the mutual coupling between the antennas [3]. The mismatch and mutual coupling between antenna elements of an array are not easy to foresee by mathematical model but still need to be considered due to their significant impact [4].

As a consequence the waves presents at PA output will travel in both directions and between the antenna elements. Fig. C.1 shows block diagram of a generic antenna array including L antenna elements where a_{1k} is the incoming signal to the amplifier, b_{1k} is the reflected signal from PA input, b_{2k} is the output from the amplifier, and a_{2k} is the reflected signal from the antenna array at the k'th branch.

There is electromagnetic coupling between the elements of active arrays. Therefore the power injected to one element of the array will be coupled back to other elements and consequently to other PAs outputs. The magnitude of the coupling is dependent on the steering angle and the radiated power level. As illustrated in figure C.1, for the k'th element of the array the effective reflection coefficient, (Γ_k), is defined as:

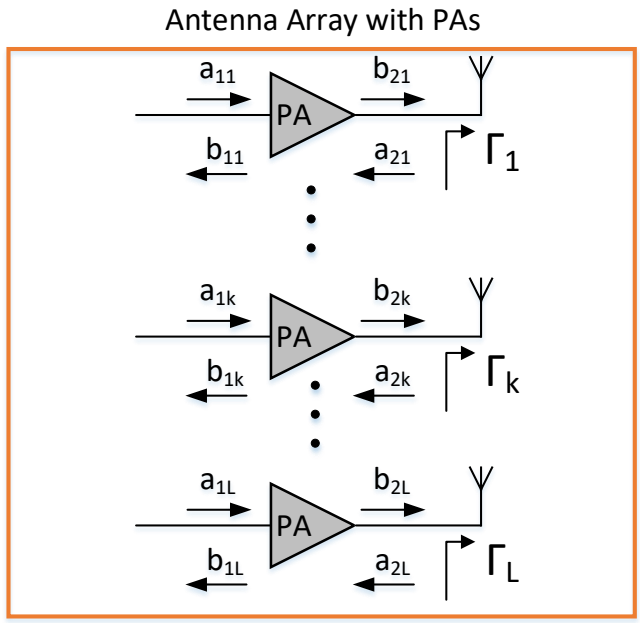


Fig. C.1: Conceptual illustration of mutual coupling and reflection coefficients of antenna array including PAs.

$$\Gamma_k = \frac{a_{2k}}{b_{2k}} = \frac{\sum_{i=1}^{L} Z_{ki} b_{2i}}{b_{2k}} \tag{C.1}$$

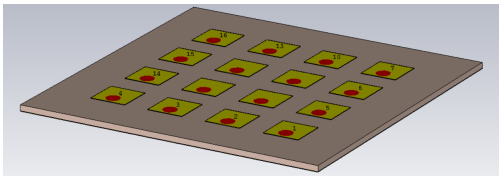
Where the coupling between the k'th and i'th antennas are defined by the coefficients in scattering matrix, Z_{ki}, which are set by the characteristics of the antenna array. The elements in b_{2i} vector are complex coefficients for the the input to the antenna. Linearity of the active device connected to the k'th element of the antenna is related to the reflection coefficient Γ_k. This parameter is not only dependent on the reflection from the k'th elemnts but also dependent on the coupling from all other elements which is defined as load modulation. The impact of this load modulation on the digital predistortion of the active phased array is described in this paper through simulation and verification.

This paper is organized as follows: Section 1 is the introduction. Section 2 is about the numerical modeling of the antenna array with CST Studio. Section 3 explains the simulation results. Section 4 describes the mutual coupling impact on digital predistortion and finally, the conclusion of this work is presented in section 5.

2 Numerical modeling of the antenna array with CST Studio

A 4x4 planar patch antenna array is designed for this simulation. By placing 16 excitation pins and by appropriately phasing of the 16 feeds, a linear polarization with different steering angle can be generated. The substrate is a low-loss Rogers RO4403 with a thickness of 0.406 mm. The 3D geometry and design parameters of the substrate are shown in Fig. C.2

An optimization of the patch dimensions and feed position was performed to improve the return loss of each patch to be approximately -25 dB at 28 GHz with sequential excitation. The time domain solver in CST MICROWAVE STUDIO® (MWS) was used with a Hexahedral mesh and a perfect matched layer (PML) boundary condition for the simulation.



(a)

Name	Expression	Value
pat_sep	= 10.7 * 0.4	4.28
Patch_Width	= Parch_length	2.56
Parch_length	= 2.56	2.56
Ground_Size	= 22	22
Substrate_Height	= 0.406	0.406
Substrate_Size	= Ground_Size	22
Feeding_Point_Lenght	= 0.60	0.60
Feed_Impedance	= 50	50
Feeding_Point_Width	= 0.00	0.00
TT	= 0.035	0.035
Patch_Rotation	= 0	0
pat_num	= 4	4

(b)

Fig. C.2: 4x4 planar patch antenna array: (a) Geometry; (b) PCB design parameters.

3 Simulation results

3.1 Sequential excitation

In the first step the 4x4 patch antenna is driven sequential i.e. one port is exited and all other ports are terminated with 50-ohm impedance. The S-parameters for port 1 with sequential excitation is shown in Fig. C.3 as an example. All other ports behaves almost similar to port 1.

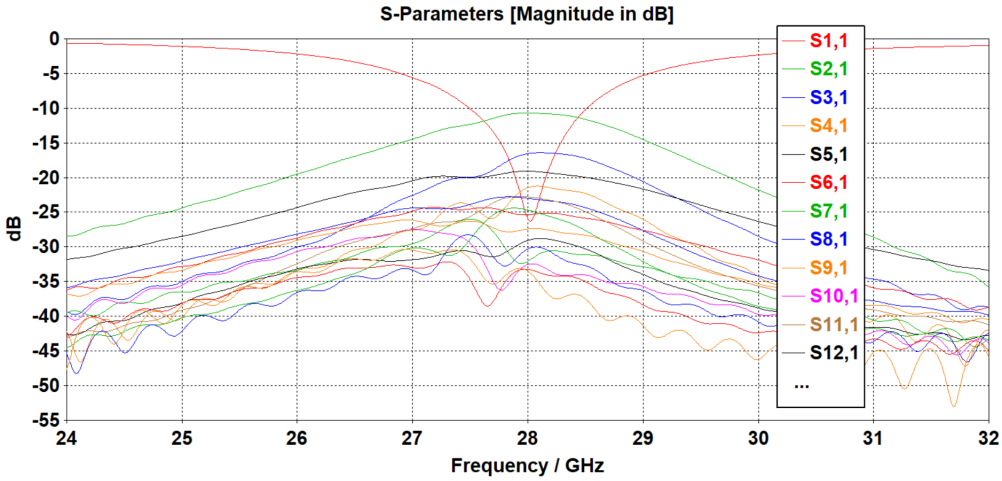
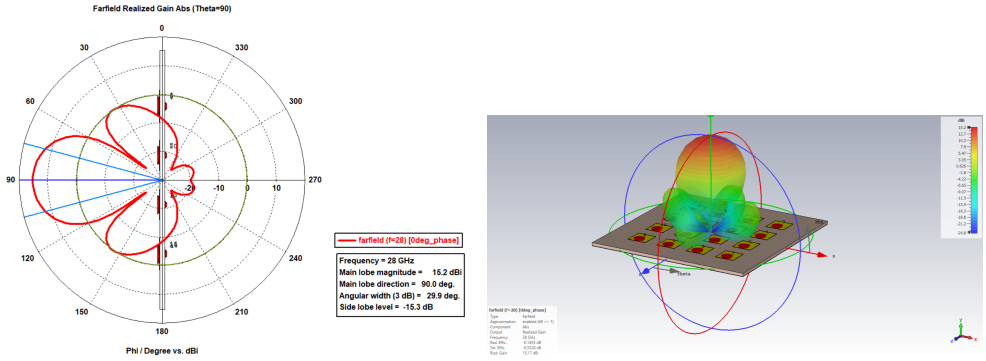


Fig. C.3: Sequential S-parameters of each patch while the other 15 patches are terminated with matched impedance.

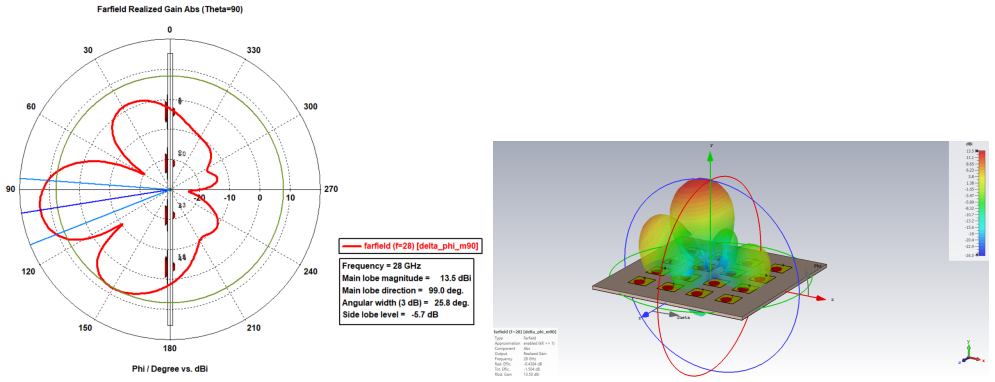
3.2 Simultaneous excitation

While exiting all ports, the steering angle is shifted horizontally in step of 3 degrees step, ie. 0, ±3, ±6, ±9, ±12 and ±15 degrees by adding appropriate phase shift to the excitation signal to each patch. The combined polar and 3D far-field results with steering angle of 0 degree and 9 degree are illustrated in Fig. C.4 and Fig. C.5 as examples.



(a) (b)

Fig. C.4: Far-field simulation results with steering angle = 0 degree: (a) Polar; (b) 3D.



(a) (b)

Fig. C.5: Far-field simulation results with steering angle = 9 degree: (a) Polar; (b) 3D.

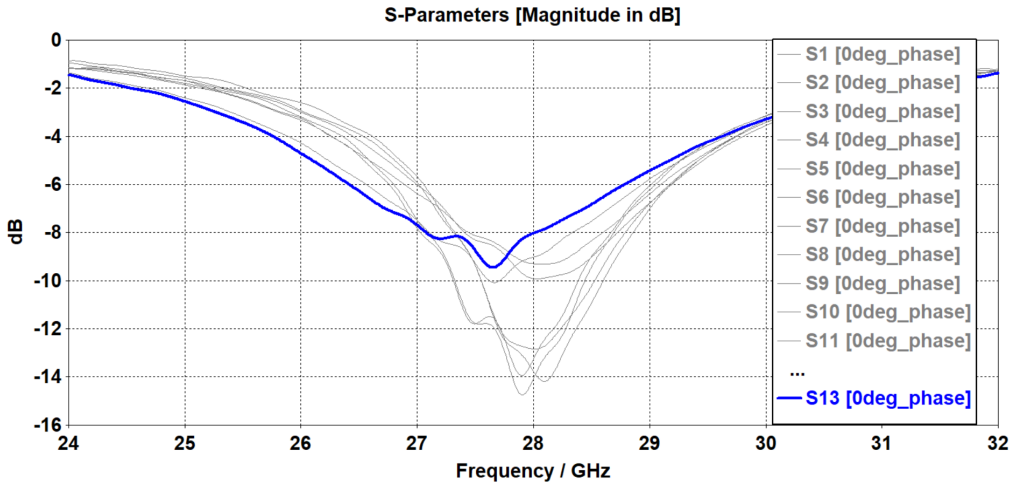
The resulting F-parameters for each patch while all patches are simultaneously exited is illustrated in Fig. C.6. The S-parameters for e.g. patch number one of the array is calculated by simulation tool as following:

$$S_1 = \frac{b_1}{a_1} = \frac{\sum_{i=1}^{16} S_{1i} a_i}{a_1} \tag{C.2}$$

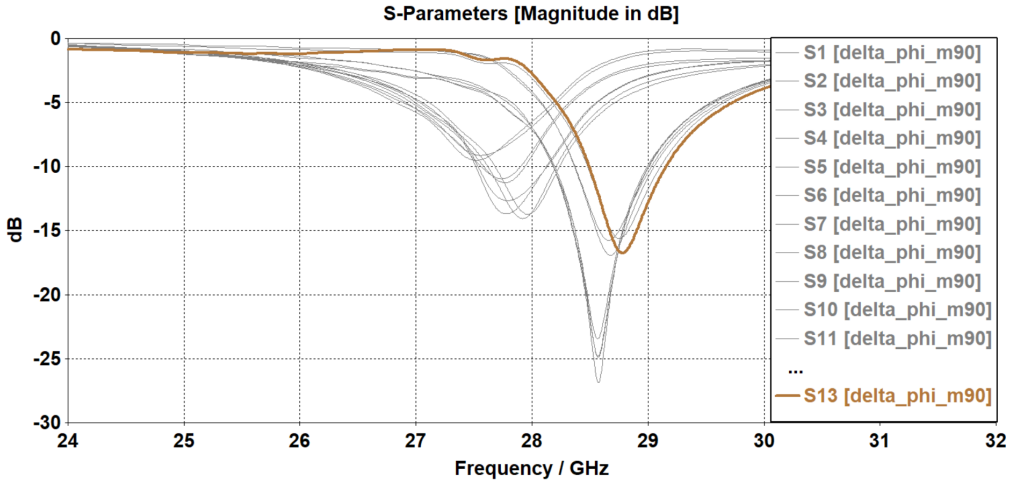
where a_1 is the exited signal into antenna and b_1 is the reflected signal from the antenna array. This is equivalent to the reflection coefficient in (3). As it is shown in Fig. C.6 for the case of 0 degree beam steering at 28 GHz frequency, taking patch number 13 as worst performed, the reflection coefficient is -8 dB, where it changes to -2.5 dB in case of 9 degree beam steering. This change of 5.5 dB varies the PA's output impedance and obviously impacts into the calibration and linearization procedure for whole array.

4 Mutual coupling impact on digital pre-distortion

In section 1 the impact of mutual coupling between the antennas at PAs' outputs has been discussed and there is shown that the impedance presented at the output of each PA depends on the mutual coupling between antennas. In order to improve the efficiency of transmitter, normally the PA is driven to compression which on the other hand results into increase of 3rd order inter-modulation which is magnified by adjacent channel leakage ratio (ACLR). Digital pre-distortion is often used as a method to keep the magnitude of ACLR as low as possible while driving PA into compression. In [5], [6]



(a)



(b)

Fig. C.6: S-parameter results for each patch, all ports simultaneously exited. (a) Steering angle = 0 degree; (b) Steering angle = 9 degree.

and [7] an investigation for improving ACLR by using digital pre-distortion of the main beam has been demonstrated.

4.1 Measurement setup

Fig. C.7 illustrates the block diagram of the measurement setup for the 4×4 active phased array (APA) and the actual measurement setup is in Fig. C.8. The intermediate frequency (IF) signal is a 3 GHz LTE10, compliant with the 3GPP downlink orthogonal frequency-division multiplexing (OFDM) modulation. It has a peak to average power ratio (PAPR) of 10.6 dB. An unmodulated signal of 12.5 GHz has been frequency-doubled to 25 GHz and fed into a power divider to be used as a local oscillator (LO) signal for both up-conversion and down-conversion. The 28 GHz bandpass filtered signal is linearly amplified by a pre-amplifier and drives the APA (AAiPK428GC-A0404) into compression. The APA includes four Anokiwave AWMF-0158 which includes four Anokiwave AWMF-0158 [8]. This device integrates 16 branches of attenuators and phase shifters plus PAs and 16 patch antennas in a 4×4 array. The main beam signal is captured by an observation horn antenna and then down-converted to IF, collected by a spectrum analyzer and converted to a baseband signal. The baseband signal is processed by MATLAB. The training and testing of the proposed method are implemented using TensorFlow 1.14 via the Keras API in Python 3.7.6.

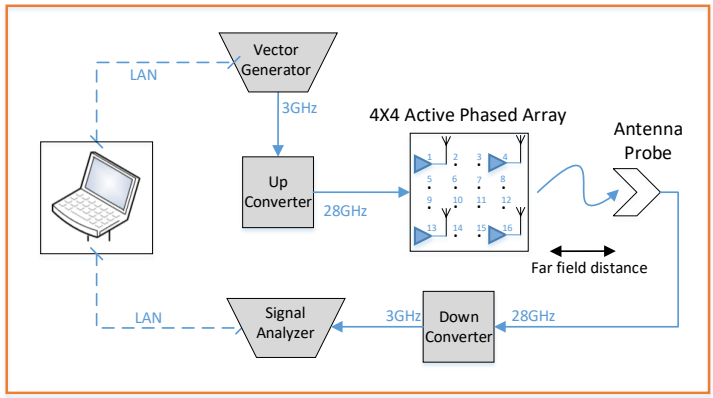


Fig. C.7: The block diagram of the measurement setup for the 4x4 array. The far field distance is 48 cm.

4.2 Measurements procedure

The steps of the measurements are:

- A vector signal generator is used to sample the 10 MHz bandwidth signal with a sampling rate of 100 MHz and 1E5 samples of I and Q data are recorded.

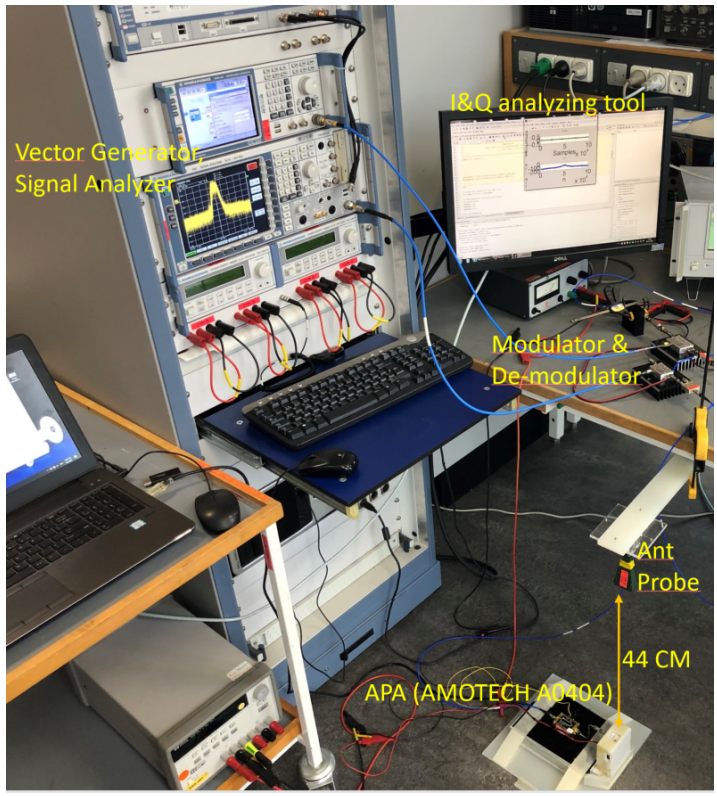


Fig. C.8: Measurement set-up.

- The vector signal generator uses these recorded I and Q data to generate a modulated signal at 3 GHz. After up-conversion by external mixer to 28GHz, the signal is fed to the active phased array and received by observation antenna. A down-conversion mixer is used for converting the 28 GHz signal to a 3 GHz signal, which is sampled by a vector signal analyzer and converted to the baseband I and Q data.
- After time-alignment by cross-correlation, the recorded input and output I and Q data are used in memory polynomial model.
- The memory polynomial model uses a nonlinearity order of 8 and a memory depth of 8 for generating the pre-distorted signal.
- A modulated 28 GHz RF signal based on the pre-distorted I and Q data is generated and inserted to the active array.

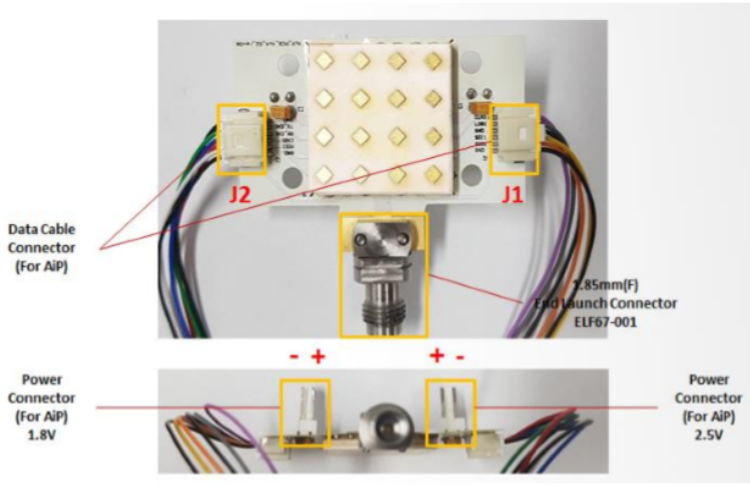


Fig. C.9: AAiPK428GC-A0404 [9] evaluation board used for measurements.

- Output power and ACLR of the 28 GHz signal are measured by the signal analyzer to quantify the linearization performance.

4.3 Measurements results

Each element of the antenna array is exited by a phase shifter and a PA. The main beam of the array has been shifted from θ = -78 to +78 degrees in approximately 5 degrees step using the code-book and software tools of AMOTECH AAiPK428GC-A0404 (using AWMF-0158 [8]). The used evaluation board for test is shown in Fig. C.9.

An observation receiver is placed in far-field to capture the steered beam. The placement of the observation receiver antenna has been kept fixed at maximum received signal at $\theta = 0$ degree at X-Y coordinate system. The measurement is done for both horizontal and vertical steering angles of the main beam. The measurement result is shown in Fig. C.10. The magnitude of the beam captured by the fixed antenna probe is varying by changing the beam direction as expected. However due to load mismatch while changing steering angle, the ACLR varies significantly.

During linearization of the active array, it is desired to keep the coefficients of the trained digital pre-distorted signal unchanged as long as possible since a new training is costly due to power consumption. However this experiment shows that a single trained DPD is not sufficient for maintaining a low ACLR in a wide range of steering angles. To avoid an ACLR improvement be degraded by more than 1dB across the steering angle, a new training after approximately $\pm$28 degree shift of the main beam is required. For allowing only 0.5 dB degradation, a new training after approximately 5 degrees shift

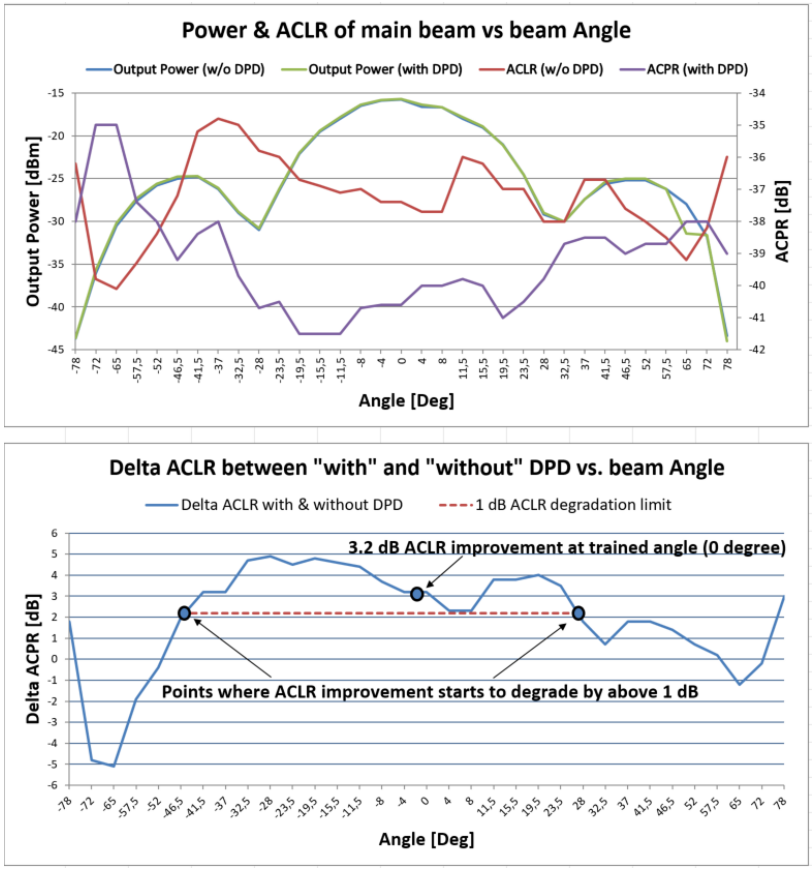


Fig. C.10: Impact of beam angle on linearization.

of the main beam is required which is a hard limit and requires many several training during beam steering.This can be explained as the effect of mutual coupling of the highly integrated antennas in the array as the simulation results in section 3 also has indicated.

5 Conclusion

This paper shows a methodology for modeling the active phased array and the behavior of the steering beam when there is no isolation between the PAs and the patches of a active antenna array. Simulation results showing a clear change of the reflection

coefficient of each patch antenna of the active array due to the mutual coupling among the antenna elements. This variation of input impedance of each element of antenna array results into change in output power and efficiency of the PAs since these parameters are mainly determined by the output impedance conditions. Simulation results in this paper shows that it is possible to magnify the input impedance of each patch of the array while all elements are exited simultaneously and it is possible to predict the behavior with strongly correlated signals.

The simulated effect has been validated by measurement on 4x4 active phased array. The mutual coupling affects, among others, the linearization of active array using digital pre-distortion. Since the load impedance of each PA is changing by the steering angle and because the pre-distortion algorithm cannot predict the loading condition then it affects the performance of the algorithm. This is illustrated in section 3 where a new training after approximately ±28 degree shift of the main beam is required.

References

[1] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results. *IEEE communications magazine*, 52(2):106–113, 2014.

[2] S Crips. RF power amplifiers for wireless communications, artech house, 1999.

[3] Zahidul Islam Shahin, et. al. Efficient dpd coefficient extraction for compensating antenna crosstalk and mismatch effects in advanced antenna system. Department of Electrical and Information Technology LTH, Lund University SE-221 00 Lund, Sweden, 2017.

[4] Constantine A Balanis. *Antenna theory: analysis and design*. John wiley & sons, 2016.

[5] Feridoon Jalili, Martin H Nielsen, Ming Shen, Ole K Jensen, Jan H Mikkelsen, and Gert F Pedersen. Linearization of active transmitter arrays in presence of antenna crosstalk for 5g systems. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, pages 1–5. IEEE, 2019.

[6] Feridoon Jalili, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Ming Shen, and Gert F Pedersen. Linearization trade-offs in a 5g mmwave active phased array ota setup. *Ieee Access*, 8:110669–110677, 2020.

[7] Mahmoud Alizadeh, Peter Händel, and Daniel Rönnow. Behavioral modeling and digital pre-distortion techniques for RF pas in a 3× 3 mimo system. *International journal of microwave and wireless technologies*, 11(10):989–999, 2019.

[8] Anokiwave. Awmf-0158 28 ghz silicon 5g tx/rx quad core ic. https://www.anokiwave.com/products/awmf-0158/index.html, 2019.

[9] Amotech. A 5G antenna that implements beam-steering with a 4 x 4 antenna element and 8 x 8 antenna, 2020.

Paper D

Tuning of Deep Neural Networks for Over-The-Air Linearization of Highly Nonlinear Wide-Band Active Phased Arrays

Feridoon Jalili, Felice F. Tafuri, Ole K. Jensen, ingyue Chen, Ming Shen and Gert F. Pedersen

The paper has been published in the
International Symposium on Networks, Computers and Communications (ISNCC)
Year: 2021 | Conference Paper | Publisher: IEEE.

In this paper, we demonstrate how a deep neural network (DNN) can be used to compensate for nonlinearities and distortion effects introduced by the latest technology of 5G transmitters. A linearization approach based on neural networks can successfully cope with the challenges introduced not only by the high nonlinearity, wide bandwidth, and high frequency but also with challenges due to inter-PA crosstalk and load modulation. The device-under-test used in this experiment, is a state-of-the-art 5G 4×4 active phased array (APA) operating in highly nonlinear regions at 28 GHz with a 100 MHz wide 3GPP base station signal and with OTA measured signals used for training. Using the proposed DNN based linearization technique, an improvement of 11 % in error vector magnitude (EVM) and 10 dB suppresion in adjacent channel leakage ratio (ACLR) are achieved which demonstrates the promising potential of this technique for emerging broadband communication systems such as 5G/6G and low earth orbit (LEO) satellite networks.

1 Introduction

The next generation of 5G new radio (NR) and satellite communications use multi-input multi-output (MIMO) systems together with active phased array (APA) transmitters in order to fulfill the capacity and data rate requirements. Due to limited space, lack of sufficient isolation between power amplifiers and antennas, large number of the active devices and high peak-to-average-ratio of 5G transmitter signals, the nonlinearity of the APAs is going to be a major challenge. On the other hand there is a huge demand for high efficiency and low power consumption which do not allow a power back-off of the active devices, so operation in highly compressed mode is desired despite the impact of non-linearities. To overcome the nonlinearity challenge which is not easily handled by the conventional linearization techniques, deep neural networks (DNN) appear to be a good solution due to their ability to learn any arbitrary nonlinear function according to the universal approximation theorem [1].

Artificial neural networks (ANN) were successfully used in the state-of-the-art to model single-input single-output (SISO) power amplifiers (PAs) [2]. The usage of an ANN architecture with multiple hidden layers, which is a system typically addressed as DNN, is here proposed to tackle the additional complexity of modern APAs. For example in [3] a DNN was used for mitigating imperfection such as imbalance of the in-phase and quadrature signal as well as DC offset correction for a wide band active phased array transmitter. Another approach where only the magnitudes of input signals are processed by nonlinear operations, and the phase information is recovered with linear weighting operations is introduced in [4]. The issue of mutual coupling effects at the output of PAs in a active phased array during beam steering, the so-called load modulation, has been analyzed and a DNN based solution for mitigating the imperfections has been introduced in [5] and the performance has been compared with the conventional memory

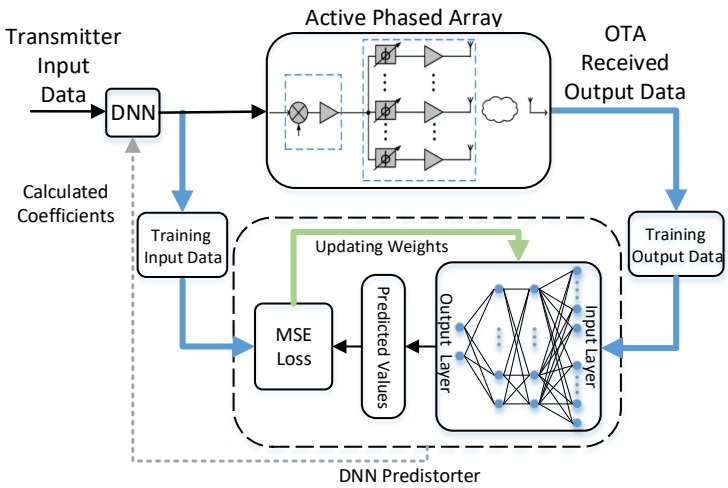


Fig. D.1: Proposed active array transmitter with DNN based linearization.

polynomial model method.

However, the state-of-the art still did not propose an analysis of the complexity vs. performance trade-off intrinsic of the usage of a high-complexity model such a DNN for modeling and linearization of modern APAs. This paper presents such an analysis based on measurements from a cutting edge device-under-test. The paper shows clearly the impact of the number of the DNN hidden layers and number of neurons in the linearization performance and we presents guidelines to optimize the complexity of the DNN to minimize the computational cost of its implementation. The proposed active array with DNN based linearization is shown in Fig. D.1.

This paper is organized as follows: Section 2 presents the linearization technique. Section 3 explains the measurement and tuning procedure and finally, the conclusion of this work is presented in section 7.

2 Linearization Technique

2.1 Deep Neural Network Model

The DNN model is not a linear-in-parameter (LIP) model. Consequently, the parameters have to be identified by a nonlinear training process. As illustrated in Fig. D.2, if i(t) and o(t), are presenting the complex equivalent base-band input and output signals of the APA, then the objective of training the DNN is to find the inverse of the nonlinearity

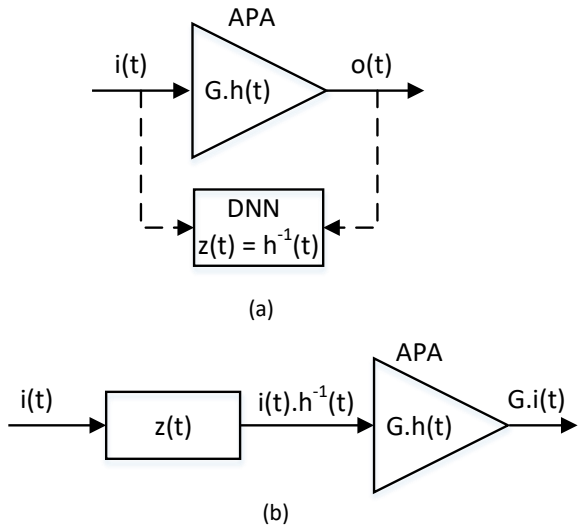


Fig. D.2: Linearization technique based on a deep neural network: (a) Configuration for model training, where G is constant gain and h(t) is the nonlinear function; (b) Configuration of applying trained DNN for pre-distortion.

function [6], h(t), in (D.1):

$$\mathbf{o}(t) = \mathbf{G} \cdot \mathbf{h}(t) \cdot \mathbf{i}(t). \tag{D.1}$$

Step by step the DNN learns the nonlinearity function by calculating the weights during training. This procedure continues until the difference between the output and the estimated output is below a defined threshold value. In the end of training the inverse nonlinearity function, z(t), is an optimal estimation of $\mathbf{h}^{-1}(t)$. When the pre-distorted signal, $\mathbf{i}(t) \cdot \mathbf{h}^{-1}(t)$, is used as the input to the APA then ideally the output will be $\mathbf{G} \cdot \mathbf{i}(t)$ which represents the linearized output.

2.2 Training Procedure

The architecture of the proposed DNN is illustrated in Fig. D.3. A feed-forward fully connected (FC) structure with a variable number of dense layers, hidden layers and neurons is constructed. The optimum number of each block is determined by iterative steps with target of achieving the best performance. The input data, $i_I[n]$ and $i_Q[n]$ and output data, $o_I[n-M]$ and $o_Q[n-M]$, used for training the DNN, are separated into in-phase and quadrature data where n is the total number of samples and M is the memory depth. The relation between input and output data of each fully connected

layer is [3]:

$$\mathbf{o}_{(j)} = \mathbf{W}_{(j)} \cdot \mathbf{i}_{(j)} + \mathbf{B}_{(j)}, \tag{D.2}$$

where j is the j-th FC layer. For an input layer of K neurons and output of L neurons then $\mathbf{W}_{(j)}$ is an $L \times K$ matrix, $\mathbf{i}_{(j)}$ is a $K \times 1$ vector, and B_j is an $L \times 1$ vector. Equation (D.2) describes the function of each dense layer defined as (a) in Fig. D.3. Using back-propagation, the characteristic weights and biases of each FC layer are then optimized. The adaptive moment estimator (Adam) algorithm is used for optimization. Using gradient descent algorithm in this optimizer makes it computationally efficient through using momentum and randomized batches to avoid local minima [7].

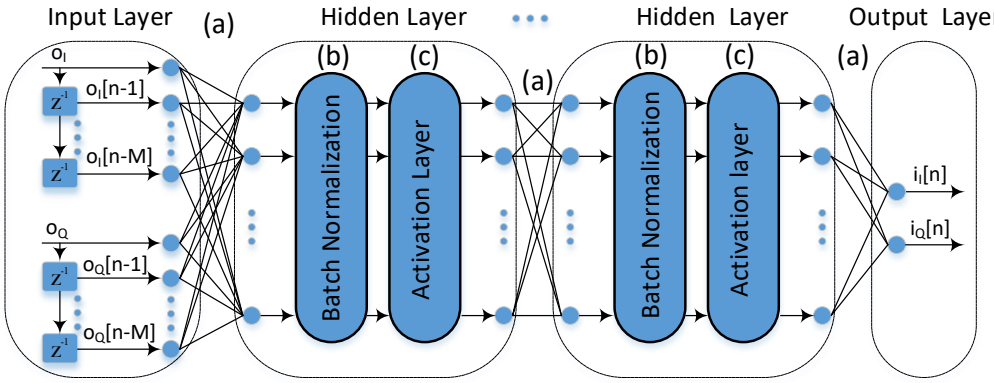


Fig. D.3: The proposed neural network architecture including dense layer (a), batch normalization layer (b) and activation layer (c). Variable parameters include memory depth M, number of hidden layers and neurons in each hidden layer.

3 Measurement and Tuning Procedure

3.1 Over-The-Air Setup

The block diagram of the setup used for capturing the IQ input and output data is shown in Fig. D.4 and the actual laboratory set-up is illustrated in Fig. D.5.

The Amotech AAiPK428GC-A0404 [8], which includes four Anokiwave AWMF-0158 [9], integrates 16 branches of attenuators, phase shifters and PAs and 16 patch antennas in a 4x4 active phased array. A host PC is used for capturing and up-loading the IQ samples. The setup includes up- and down-conversion from sub 6 GHz to 28

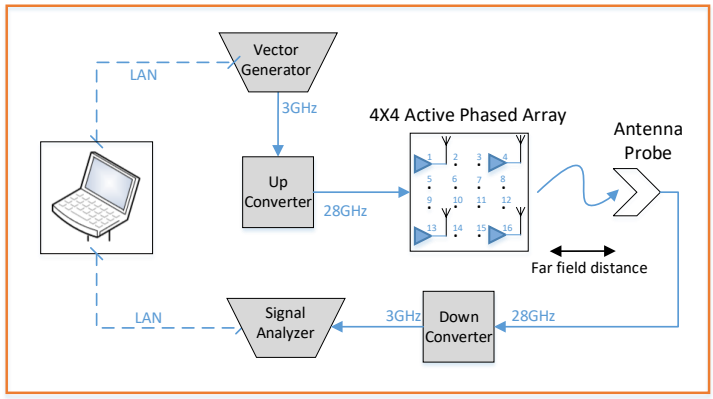


Fig. D.4: The block diagram of the measurement setup for the 4×4 array.

GHz and pre-amplification stages. All components are calibrated to run in their linear operating regions except the APA which is running in the highly nonlinear region and its nonlinearity is the main impairment factor. For controlling the main beam of the array the code-book and software tools of Amotech have been used.

3.2 Parameter Tuning

A 100 MHz bandwidth 5G 3GPP downlink OFDM modulated signal at 28 GHz frequency is used as input to the APA. Both input and output signals at base band level are captured with sampling frequency of 600 MHz. For parameter tuning, 100k IQ samples of input and output of the active array are used, where 70% of the data was randomly chosen for training and the remaining 30% for testing. The training and validation of the proposed technique are implemented using TensorFlow 1.14 through the Keras API in Python 3.8.4 environment. Table D.1 shows how the DNN is configured and which parameters are tuned in which ranges. The memory depth, the number of neurons in each layer and the number of hidden layers should be set appropriately. If these numbers are too small, then the DNN cannot get the right features for the nonlinearity model and there is a risk of underfitting. The same is valid if these numbers are too large which results in overfitting. To avoid this, the memory depth has been chosen to a small number, and then the other parameters have been initialized to get the best linearization parameters in terms of ACLR and EVM. Using the input and output data for each power level, we train the DNN to reach the minimum MSE of less than 1E-6 as shown in Fig. D.6. The optimization is continued by tuning the memory depth and keeping other parameters unchanged. Fig. D.7a shows the ACLR and EVM results

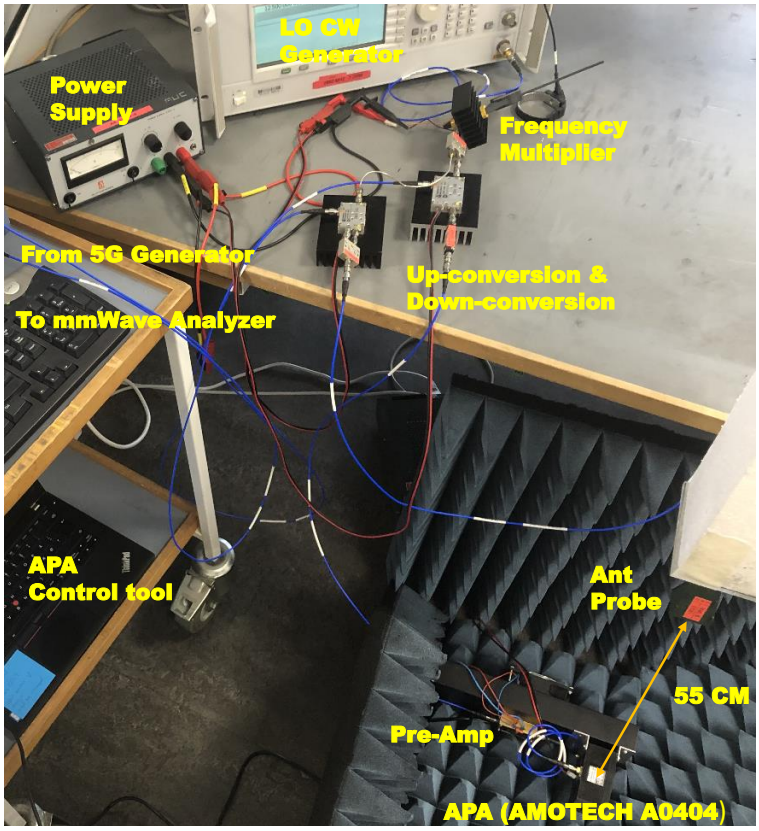


Fig. D.5: Measurement setup in Lab.

for different numbers of memory depths. The best performance is achieved by setting memory depth to 14. The experiment clearly indicates that, for the highly nonlinear use case, using an optimal DNN memory depth value had significant impact on the in-band distortion (EVM) linearization performance. Increasing the memory depth above the optimal value did not provide any benefit.

Having the memory depth optimized to 14, the number of neurons is tuned for best linearization which shows that there is a minor impact from the number of neurons if it is at least 150. In the end the number of hidden layers is tuned and, as shown in Fig. D.7b, a hidden layer number above 3 cannot improve the linearization performance. Since the additional improvement due to the number of hidden layers going from 1 to 3 is very small, then for achieving faster tuning and for reducing the number of multiplications in a real application, 1 hidden layer is chosen.

Table D.1: Configurations and Training for optimizing Memory depth.

Optimizer	Adam
Activation function	ReLU
Cost function	MSE
Accelerator	BN
Hidden layers	1 to 10
Neurons	50 to 600
Epoch	50
Batch size	100
Initial learning rate	0.1
Minimum loss	10^{-6}
Training data	70%
Validation data	30%
Memory Depth	1 to 23
Sampling frequency	600 MHz

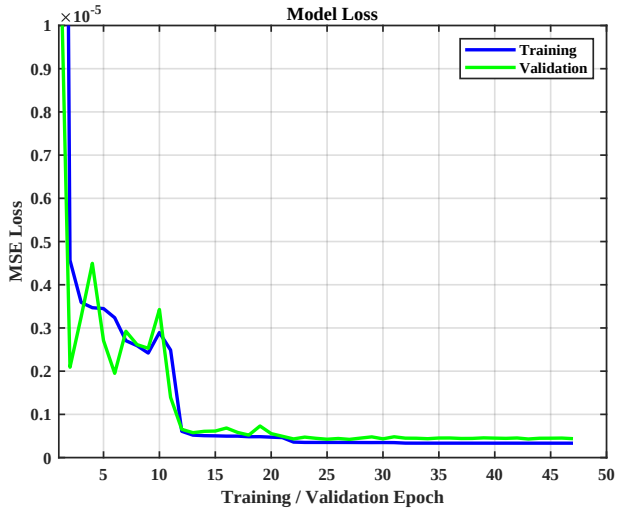
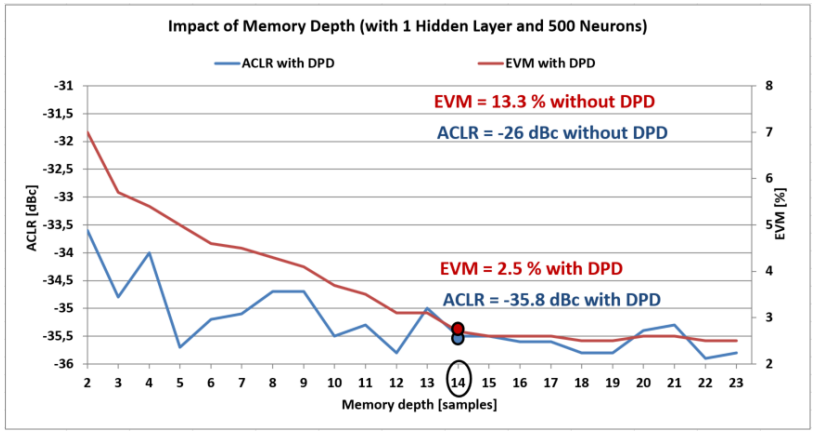


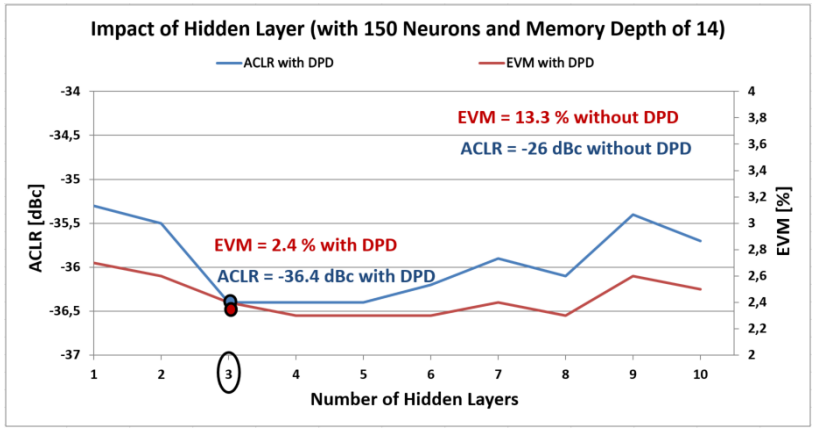
Fig. D.6: Model loss vs. number of training / validation epochs.

3.3 DNN linearization results based on tuned parameters

Fig. D.8 shows the linearization results in terms of power spectral density (PSD), amplitude to amplitude (AM-AM) and amplitude to phase (AM-PM) distortions for the active phased array. Training data and validation data are based on the best tuning



(a)



(b)

Fig. D.7: Results of tuning of the DNN parameters: (a) tuning of the memory depth; (b) tuning of the number of hidden layers.

parameters. Beside the good improvement of ACLR, Fig. D.8a, the proposed DNN effectively improves the memory effect of the system as illustrated in Fig. D.8b and Fig. D.8c. According to our experiments, the DNN optimization procedure was best performed by fixing the optimal memory depth first and the optimal number of neurons and layers afterwards.

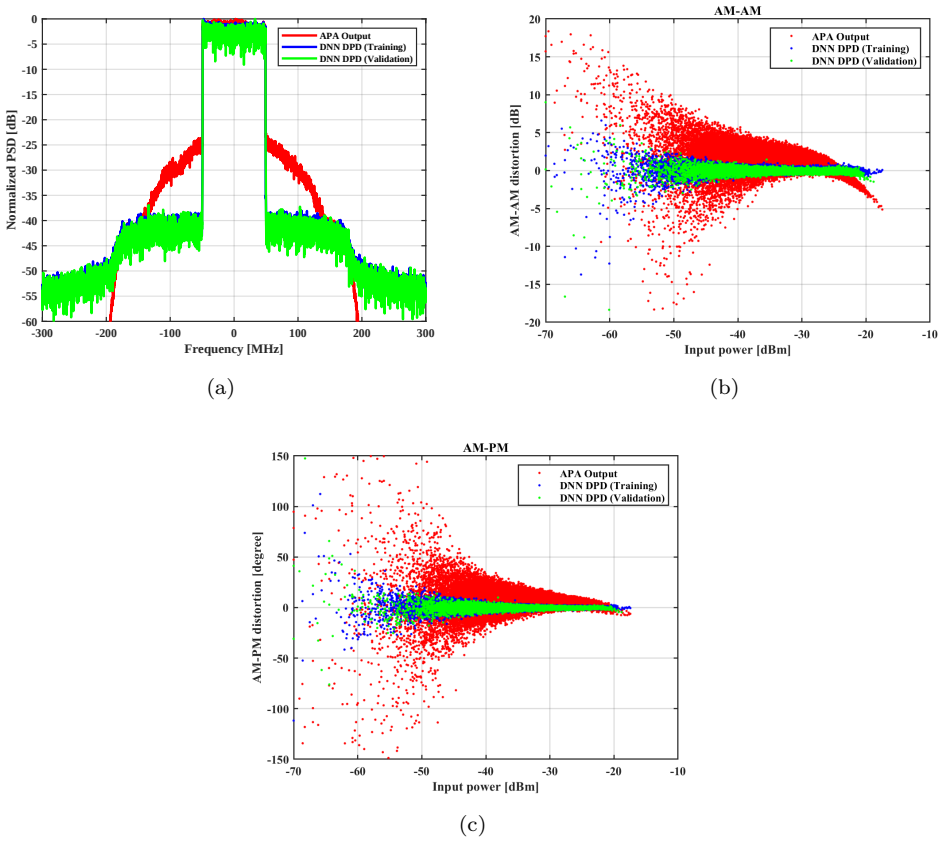


Fig. D.8: Linearization result using the tuned parameters values, memory depth = 14, neurons = 150 and hidden layer = 1: (a) Power spectral density (PSD); (b) AM-AM distortion; (c) AM-PM distortion.

4 Conclusion

This paper demonstrated the feasibility of applying a deep neural network (DNN) based technique for the digital pre-distortion of a highly nonlinear active phased array (APA). The experiments were carried out using a state-of-the-art 4×4 APA transmitter together with a 100 MHz wide band 3GPP OFDM base-station signal as input.

The authors proposed an optimization procedure to minimize the DNN number of coefficients and computational cost while obtaining the best linearization performance. The proposed procedure allowed to find the optimal values to be used for the DNN memory depth, number of layers and number of neurons.

The proposed DNN linearization technique applied on a highly nonlinear 4×4 active phased array showed an improvement of EVM and ACLR of 11 % and 10 dB, respectively, which exhibit the promising potential of this technique for future broad-band communication systems.

References

[1] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[2] Farouk Mkadem and Slim Boumaiza. Physically inspired neural network model for rf power amplifier behavioral modeling and digital predistortion. *IEEE Transactions on Microwave Theory and Techniques*, 59(4):913–923, 2011.

[3] Praveen Jaraut, Meenakshi Rawat, and Fadhel M Ghannouchi. Composite neural network digital predistortion model for joint mitigation of crosstalk, i/q imbalance, nonlinearity in mimo transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(11):5011–5020, 2018.

[4] Yikang Zhang, Yue Li, Falin Liu, and Anding Zhu. Vector decomposition based time-delay neural network behavioral model for digital predistortion of rf power amplifiers. *IEEE Access*, 7:91559–91568, 2019.

[5] Alberto Brihuega, Lauri Anttila, and Mikko Valkama. Neural-network-based digital predistortion for active antenna arrays under load modulation. *IEEE Microwave and Wireless Components Letters*, 30(8):843–846, 2020.

[6] Andrzej S Ciminski. Recurrent neural networks usefulness in digital pre-distortion of power amplifiers. In *15th International Conference on Microwaves, Radar and Wireless Communications (IEEE Cat. No. 04EX824)*, volume 1, pages 249–252. IEEE, 2004.

[7] Diederik P Kingma and Jimmy Ba. Adam: A methodfor stochastic optimization. In *International Conference onLearning Representations (ICLR)*, 2015.

[8] Amotech. AAiP28G-A0808-EVB-R01_190321c, 2020.

[9] Anokiwave. Awmf-0158 28 ghz silicon 5g tx/rx quad core ic. `https://www.anokiwave.com/products/awmf-0158/index.html`, 2019.

Paper E

Highly Nonlinear and Wide-Band mmWave Active Array OTA Linearization Using Neural Network

Feridoon Jalili, Yufeng Zhang, Markko Hintsala, Qingyue Chen, Ming Shen and Gert F. Pedersen

The paper has been published in the
IET Microwaves, Antennas & Propagation | 14 December 2021 | Journal Article.

Abstract

This paper proposes a neural network (NN) based over-the-air (OTA) linearization technique for a highly nonlinear and wide-band mmWave active phased array (APA) transmitter and compares it with the conventional memory polynomial model (MPM) based technique. The proposed NN effectively learns the distinctive nonlinear distortions, which may not easily fit to existing MPM solutions, and can therefore successfully cope with the challenges introduced by the high nonlinearity and wide bandwidth. The proposed technique has been evaluated using a state-of-the-art 4x4 APA operating in highly nonlinear regions at 28 GHz with a 100 MHz wide 3GPP base-station signal as input. Experimental results show the predistortion signal generated by the NN exhibits peak to average power ratio (PAPR) much lower than the one generated by MPM and consequently superior linearization performance in terms of adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) for high nonlinearity cases. Using the proposed NN-based linearization technique, an improvement of 5 dB ACLR and 7 % points in EVM are achieved, which demonstrates the promising potential of this technique for emerging broadband communication systems such as 5G/6G and low earth orbit (LEO) satellite networks. [1]

1 Introduction

Recent wireless communication systems operating at mmWave are using active phased array (APA) transmitters together with multi-input multi-output (MIMO) systems to improve the system capacity and data rates of the wireless networks. The traditional linearization techniques like digital predistortion (DPD) which are mainly based on the memory polynomial model (MPM) are facing new challenges [2] such as:

- Highly integrated front-ends and a large number of PA's do not allow placement of feedback circuits for each branch so a single input single output (SISO) over-the-air (OTA) model is needed.
- mmWave frequencies and wide-band linearization must be handled.
- The increasing desire for high power efficiency requires linearization of power-efficient but highly nonlinear APAs.

Several solutions using a SISO model and modified DPD algorithms for linearization have been proposed to combat the above challenges [3–7]. A SISO modeling where the entire transmitter has been considered as a two-port system has been presented in [8–11] using an observation receiver in far-field together with using a MPM-based DPD technique for linearization of the antenna array in presence of crosstalk. In [12], the potential mismatches between PAs have been compensated, so that they all exhibit

the very same behavior. By doing so, it is possible to provide linearization in all directions with a single DPD, in contrast to linearizing the main beam only. However, compensating the mismatch requires analog circuits which introduces complexity and delay for large arrays and the potential changes in the PAs' behaviors due to crosstalk. In the present work, the reference signal for DPD learning was obtained through measurements from a far-field test receiver placed on the main beam direction and the focus is on the challenges related to high bandwidth and high nonlinearity.

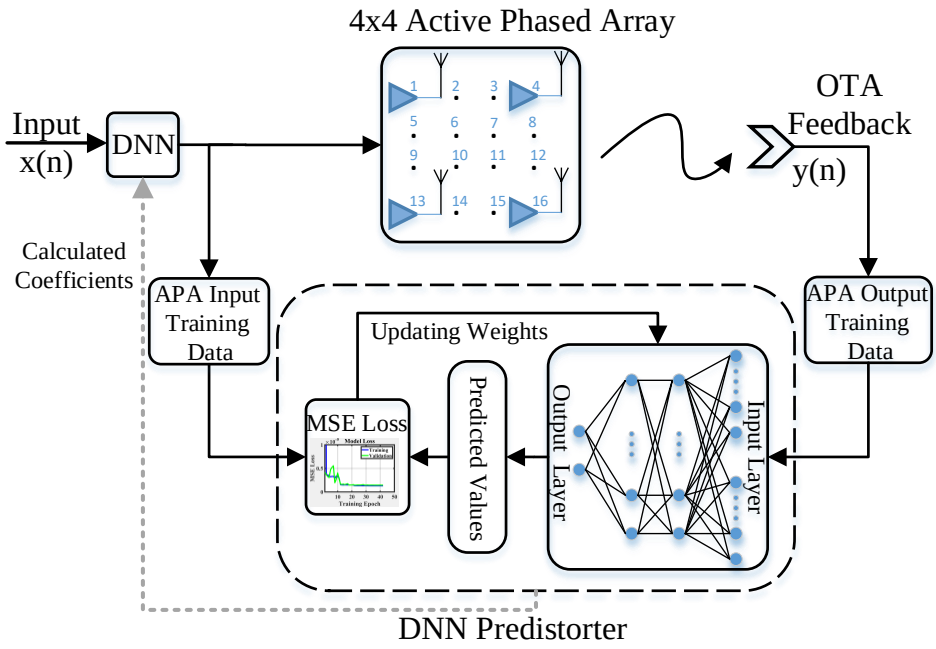


Fig. E.1: Concept illustration of the digital predistortion for APA based on the equivalent SISO model using NN.

For the cases where enhanced power efficiency is required such as Doherty PAs and envelope-tracking based transmitters, the amplifier exhibits different behavior for different power levels. A piecewise model based on a region partition algorithm that takes the actual nonlinear characteristics of the device into account was proposed in [11], which gives significantly better linearization than the general memory polynomial models. However, memory modeling capabilities may be compromised in piecewise models as the different sub-models operate independently, whereas memory effects may involve samples belonging to different sub-regions. A new piecewise model for PAs based on the mixture of experts (ME) approach, which builds on a probabilistic model that al-

lows the different sub-models to cooperate has been presented in [13]. It demonstrates a model which outperforms previous piecewise modeling methods. The ME approach is a promising technique and is a highly valid approach to be compared with the NN approach in future work. The challenges as high bandwidth and high nonlinearity lead to huge complexity and explosion of MPM-based algorithms. The Volterra series model approach, which is commonly used in MPM approaches, is preferable if the order of the nonlinearity is not too high (e.g. third or possibly fifth order) [14]. With very high nonlinearity order, the MPM approach is not practical because of the increased complexity and consequent increase of the number of unknown kernel coefficients in the model [15]. Neural networks (NNs) are well known to be able to learn any arbitrary nonlinear function according to the universal approximation theorem [16]. Several state-of-the-art linearization techniques based on NN have recently been introduced. A solution for performance imperfections such as crosstalk, power amplifier (PA) non-linearities along with modulator imperfections like in-phase and quadrature (I/Q) imbalance and DC-offset for a wide-band direct-conversion transmitter has been recently introduced in [17, 18]. A similar approach, where only the magnitude of input signal undergoes a nonlinear operation, and the phase information is recovered with a linear weighting operation has been introduced in [19].

For wideband signals, in particular, the memory effects have significant impact [20, 21]. To take care of memory effects, two dynamic neural structures have been proposed in the NN literature [1]. In the first structure, recurrent neural networks (RNNs), utilizes feed-forward and feedback signal processing. In another structure, a real-valued time-delay neural network (RVTDNN), combines I/Q processing with input time-delay lines to handle memory effects, whereas RNN uses output-to-input time-delay lines. Reference [22] indicates that RVTDNNs offer superior performance and easy baseband implementation when used for inverse modeling of PAs with strong non-linearities and memory effects. For the high nonlinearity cases, the model needs a low learning rate during training at the cost of the training time. In the present paper, we are using the so-called batch normalization (BN) together with the hidden layer in order to use a higher learning rate and reduce the training time. Furthermore, the proposed RVTDNN uses the rectified linear units (ReLU) activation function, which is less computationally expensive than hyperbolic tangent (Tanh) and Sigmoid because it involves simpler mathematical operations [23]. We are proposing a NN using only one hidden layer and a minimum number of neurons to make it comparable with conventional MPM. The proposed RVTDNN is applied to linearize highly nonlinear multi-PA devices-under-test (DUTs) such as the active phased array. We are using the proposed NN model for a 5G DUT that includes complex interactions between the PAs in the array, such as load-modulation. Measurements quantifying this impact are included in section 6. Finally, for the first time, to the best knowledge of the authors, a pre-distortion scheme based on the RVTDNN was validated using a real 5G test-bed environment with a minimum number of neurons and layers together with the ReLU activation function to keep the

cost and size of the device during implementation as low as possible.

Fig. E.1 illustrates the digital predistortion concept for the APA based on the equivalent SISO model using the proposed neural network. The mapping relationship between the order of memory depth, the number of hidden layers and the number of neurons to the corresponding required linearity have been analysed. The optimum levels for the parameters in each block have been identified, verified through measurements, and then the performance and the complexity are compared with the applied MPM based DPD using the same laboratory setup.

This paper is organized as follows: Section 1 is the introduction. Section 2 describes the MPM based approach. Section 2 is about the NN linearization technique. Section 3 explains the NN training and parameter tuning, section 5 is an investigation of complexity and section 6 is about the measurement results. A discussion on comparison between measurements results of MPM and NN approaches is included in section 7 and finally, the conclusion of this work is presented in section 7.

2 MPM based approach

The classical approach to modeling the full behaviour of a nonlinear device is by the Volterra series, Eq. (E.1), which describes the relation between the output and input signals in discrete time:

$$\begin{aligned} y[n] &= \sum_{k=1}^{K} \sum_{m_1=0}^{M-1} \cdots \sum_{m_k=0}^{M-1} h_k[m_1, \ldots, m_k] \prod_{j=1}^{k} x[n-m_j] \\ &= \sum_{m_1=0}^{M-1} h_1[m_1] x[n-m_1] \\ &+ \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} h_2[m_1, m_2] x[n-m_1] x[n-m_2] \\ &+ \ldots \\ &+ \sum_{m_1=0}^{M-1} \cdots \sum_{m_K=0}^{M-1} h_K[m_1, \ldots, m_K] \prod_{j=1}^{K} x[n-m_j], \end{aligned} \tag{E.1}$$

where K is the order of the nonlinearity, M is the memory depth and $h_k(m_1, \ldots, m_k)$ are the parameters of the model, which are often referred to as the "Volterra kernels" in literature. The n^{th} sample of the input signal $x[n]$ is mixed with the $M-1$ preceding samples at each of the k^{th} Volterra kernel. In other words, the k^{th} kernel includes all possible combinations of k time shifts of the input signal, which includes all types of memory effects. For this reason the Volterra series is considered as the most complete

model, but the computational complexity of the model is very high [24]. A much less complex model is the MPM which is widely used for linearization. Eq. (B.4) represents the applied MPM which is a deviation of the Hammerstein model and has been proven effective for removing nonlinearity and memory effect [25]:

$$y(n) = \sum_{k=1}^{K} \sum_{m=0}^{M} a_{km} x(n-m) |x(n-m)|^{k-1}, \tag{E.2}$$

where a_{km} is the 2-D array of filters and power series coefficients of the active device, K is the nonlinearity order of the memory polynomial model and M is the highest memory depth. a_{km} coefficients are the linear weighting of nonlinear signals and these coefficients are calculated by using the least-squares type algorithm. The generalized memory polynomial, which combines the memory polynomial with cross terms between the signal and lagging and/or leading exponentiated envelope terms is presented in [25]. This model shows a slightly improved linearization effect but on the cost of complexity which needs to be compared with a more complex neural network model i.e. long short-term memory (LSTM) neural network techniques [26]. In this work, we introduce the comparison between a MPM model based on Eq. (E.2) and a simple neural network model to relax the overall complexity.

In [9], we have provided a detailed insight into the linearization mechanisms for an APA based on the MPM model. A similar approach has been used for constructing the predistorted signals for the different nonlinearity cases of the APA in actual work. The same captured input and output I and Q samples are used for both MPM and NN techniques.

3 NN Linearization Technique

3.1 NN Model

The SISO model where the entire transmitter has been considered as a two-port system has been described in [9]. This model uses only one external antenna for observing the combined signal in the far field. Similarly in the present work, the entire OTA beam-forming set-up is considered as a SISO model with the APA as the main source of nonlinearity. The NN is used as an inverse system for such a model and it is trained using the measured input and output data. Once the training is completed the inverse model is used as predistorter for the SISO model, as seen in Fig. E.2. If the output and the input of the SISO model are set to the I and Q training data, $y(t)$, and true values, $x(t)$, respectively, then the NN needs to be trained to capture the nonlinearity of the model by generating the inverse of the nonlinearity function, $h(t)$, given in Eq. (E.3):

$$y(t) = G \cdot h(t) \cdot x(t), \tag{E.3}$$

where G is the constant gain.

The aim of the training is to calculate the weights such that the NN gradually learns the nonlinearity of the SISO model during the training procedure. When the cost is under the specified threshold or no longer converges, then the training step is finished.

After training, the functionality of the NN, denoted as $u(t)$ in Fig. E.2, is an optimal estimation of $h^{-1}(t)$. Ideally using the predistorted signal, $x(t) \cdot h^{-1}(t)$, as input to the SISO model, the output will be a linearized function defined as $G \cdot x(t)$.

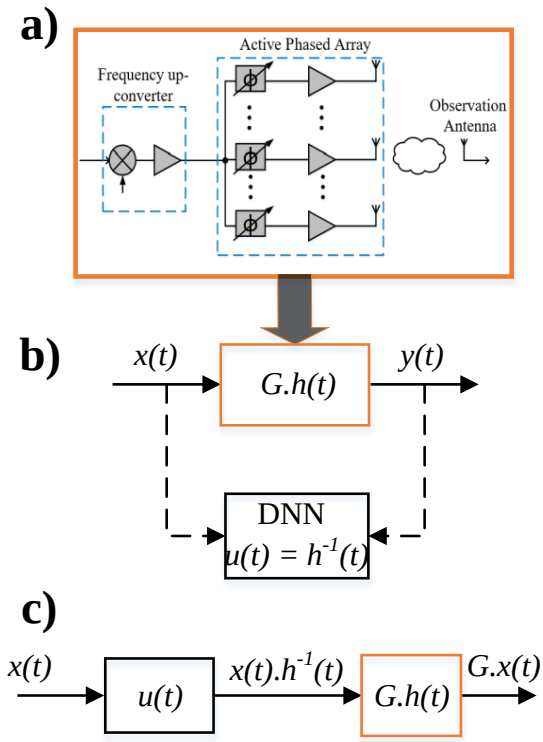


Fig. E.2: Linearization technique based on a neural network. a) Beam-forming behavioral SISO model; b) NN based on equivalent SISO model where G is constant gain and $h(t)$ is the nonlinear function; c) Applying trained NN in predistortion and linearization.

3.2 NN Training

The proposed NN shown in Fig. E.3 uses a feedforward fully connected (FC) structure. Based on the interconnection pattern or architecture, we can distinguish between

feedforward networks (FNNs) and recurrent (or feedback) networks (RNNs) [27]. The feedforward network is considered since it is the most used NN and according to the universal approximation theorem, it can approximate any nonlinear function with any desired error [28]. A FC structure in a densely populated NN may increase requirements for hardware resources, but in many applications, the weight of some interconnections can be set to zero without loss of accuracy, which results in sparsely connected layers [27]. The sparse structure is out of the scope of this work. The input and output data are separated as $y_I[n-M]$, $y_Q[n-M]$, $\hat{x}_I[n]$ and $\hat{x}_Q[n]$ where n is the number of the I and Q data used in the training. The wide-band memory effects are modeled by the delayed replica up to memory depth of M. The weights, $\mathbf{W}_{(i)}$, and biases, $\mathbf{B}_{(i)}$ for the vector expressing the relation between input and output of each FC layer is defined as:

$$\mathbf{y}_{(i)} = \mathbf{W}_{(i)} \cdot \mathbf{x}_{(i)} + \mathbf{B}_{(i)}, \tag{E.4}$$

where i is the i-th FC layer. For an input layer of L neurons and output of P neurons then $\mathbf{x}_{(i)}$ is a $L \times 1$ vector, $\mathbf{W}_{(i)}$ is a $P \times L$ matrix and $\mathbf{B_i}$ is a $P \times 1$ vector. Each dense layer which is defined as (a) in Fig. E.3, can be described using Eq. (E.4). The weights and biases of each FC layer are distinctive and are optimized by back-propagation. The optimization algorithm used in this work is the adaptive moment estimator (Adam) [29]. It is based on a gradient descent algorithm that gets more computationally efficient by using momentum and randomized batches to avoid local minima. The batch size is the number of training samples used for estimating the error gradient. A batch size, e.g. 50, means that 50 samples of the training samples are used for estimating the error gradient before the weights are updated. Another parameter, called training epoch, shows how many passes have been done through the training samples with a randomly selected group of batches. The training procedure is summarized in Table E.1.

Table E.1: Algorithm used for NN training.

Algorithm 1 NN Training process
i: Generate n samples of IQ data of $x[n]$ and $y[n]$
ii: Update weights and biases given by (E.4) using 70 % of n samples
iii: Continue updating until finding the optimum $u(t) = h^{-1}(t)$
iv: Validate the model using 30 % of n samples
v: If the cost function of validation is ok, then freeze the model
vi: Construct the predistorted signal, $x(t) * h^{-1}(t)$

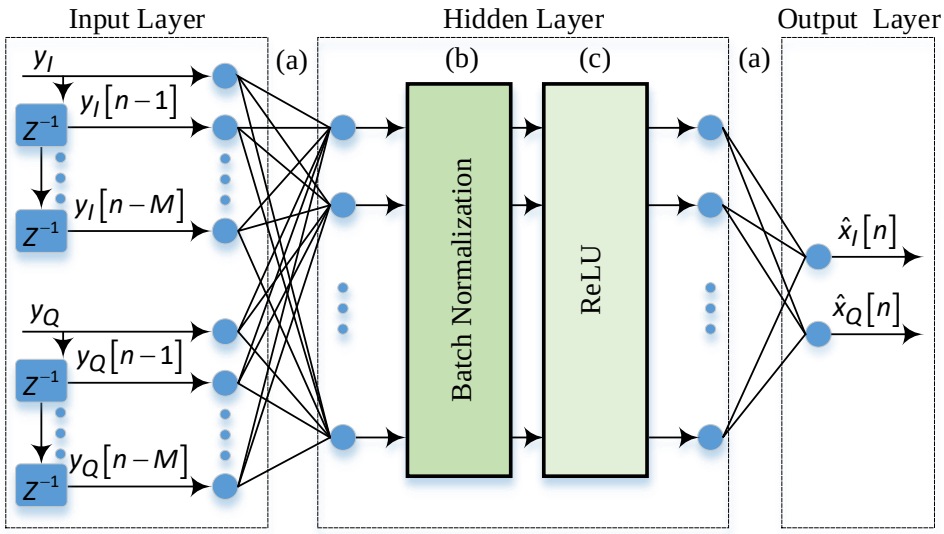


Fig. E.3: Structure of the neural network architecture proposed in this paper. The hidden layer consists of two dense layers which are batch normalization layer and ReLU layer.

3.3 Accelerating NN Training

For the models used in high nonlinearity cases, the training needs a low learning rate which on the other hand increases the training time. Therefore in each hidden layer, there is also an accelerator, the so-called BN layer, shown as block (b) in Fig. E.3. The BN layer allows using much higher learning rates which will accelerate the training and reduce the time cost significantly [30].

The BN layer normalizes the mean and variance of the outputs of the dense layer to 0 and 1 and introduces a new mean and variance. The output of the BN layer, $\hat{\mathbf{y}}_{(i)}$ is expressed by:

$$\hat{\mathbf{y}}_{(i)} = \gamma \frac{\mathbf{y}_{(i)} - \mathbb{E}[\mathbf{y}_{(i)}]}{\sqrt{\mathrm{Var}[\mathbf{y}_{(i)}] + \epsilon}} + \beta, \tag{E.5}$$

where γ and β are the new learnable mean and variance parameters and ϵ is a constant parameter to prevent the equation be infinite and is set to 0.001.

3.4 Activation Function

For the NN to be able to fit an arbitrary nonlinear function, the nonlinearity is introduced in the form of an activation function which is shown as block (c) in Fig. D.3.

Both Tanh and ReLU are evaluated as activation functions in this work, where ReLU has been chosen as the activation function due to its performance which is described in section 4. ReLU introduces a nonlinearity by deactivating negative inputs, adding sparsity to the model, and accelerating convergence [31]. ReLU is defined as:

$$f_{\mathrm{ReLU}}(u) = \max(0, u), \tag{E.6}$$

where u is the input to the activation function. In this way the output of a hidden layer can be expressed as Eq. (E.7):

$$\mathbf{x}_{(i+1)} = \max\left(0, \hat{\mathbf{y}}_{(i)}\right), \tag{E.7}$$

where $\mathbf{x}_{(i+1)}$ is the input of the next FC layer. With the sequential structure, the inputs of the subsequent hidden layer can be described in terms of the current hidden layer. This procedure which goes from the first layer to the last layer is called forward propagation.

3.5 Cost Function

There are different ways, based on the type of the problem, to evaluate the difference between the real output and the estimated output, the so-called cost function. In this work the effects of two kinds of cost functions, Huber cost and mean square error (MSE) cost have been investigated. In the Huber cost function, instead of minimizing the cost function, $|x_i - \hat{x}_i|$, the smooth cost function, L_1, is used for regression because it is robust against gross errors [32]. The smooth L_1 cost function is defined as,

$$L_1(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{B}\sum_{i=1}^{B} \varepsilon_i, \tag{E.8}$$

where B is the batch size and ε_i is defined as a combination of the squared error and absolute error,

$$\varepsilon_i = \begin{cases} 0.5(x_i - \hat{x}_i)^2, & \text{if} \quad |x_i - \hat{x}_i| < 1, \\ |x_i - \hat{x}_i| - 0.5, & \text{otherwise.} \end{cases} \tag{E.9}$$

The corresponding MSE cost function is defined as:

$$\mathrm{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{B}\sum_{i=1}^{B} (x_i - \hat{x}_i)^2. \tag{E.10}$$

4 NN Training and Parameter Tuning

4.1 Training Process

The concept of the proposed application is illustrated in Fig. D.1 and the configuration of the designed NN is shown in Fig. D.3. As shown in Fig. D.1, the parameters of the NN are updated step by step by reducing the losses between outputs of the NN (i.e., predicted values) and the reference inputs. The NN can gradually learn features hidden in training data for classification or regression missions. Generally, if the NN is trained as a classifier, the cross-entropy function is a commonly used cost function. For linearization of active circuits where the NN needs to learn the distinctive nonlinear distortions, it is trained as the regression model.

Considering the memory effect of the active array, the memory depth, M, has a direct impact on the number of neurons in the input layer. So there is a trade-off between the size of M and the linearization performance. Since the output training data and the reference input data are complex values (I and Q), the number of neurons of the input layer and output layer is set to $2 \cdot M$ and 2, respectively.

Adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) are used as metrics for choosing the desired parameter values in each training step. The ACLR describes the power of the leakage in the adjacent channel compared to the in-band channel power and is defined as:

$$\mathrm{ACLR}_{\mathrm{dB}} = 10 \log_{10} \left(\frac{P_{\mathrm{adj}}}{P_{\text{in-band}}} \right), \tag{E.11}$$

where P_{adj} and $P_{\text{in-band}}$ are the powers of the adjacent channel and the main channel, respectively. In this way, the signal integrity can be directly assessed in the frequency domain. The left side ACLR is used for evaluation through the experiments in this paper. Since ACLR only measures distributed power in different channels, another metric for in-band signal quality, EVM, in terms of percentage, is calculated as:

$$\mathrm{EVM}_{\%} = \sqrt{\frac{P(\text{in-band})_{\mathrm{error}}}{P(\text{in-band})_{\mathrm{ref}}}} \cdot 100_{\%}, \tag{E.12}$$

where $P(\text{in-band})_{\mathrm{error}}$ and $P(\text{in-band})_{\mathrm{ref}}$ are the powers of the error vector and ideal signal vector in I and Q plane, respectively.

All operations are realized using Python 3.8.4 on Visual Studio Code. The NN is built and trained using Keras 2.3.0-tf, and the version of Tensorflow is 2.2.0.

4.2 Parameter Tuning

For parameter tuning, 100k I and Q samples of input and output of the active array are captured, where 70% of the data was randomly chosen for training and the remaining 30% for testing.

The memory depth, the number of neurons in each layer and the number of hidden layers should be set appropriately. If these numbers are too small, then NN cannot get the right features for the nonlinearity model and there is a risk of underfitting, the same is valid if these numbers are too large which results in overfitting. To avoid this, the memory depth has been chosen to a low number, and then the other parameters have been initialized to get the best linearization parameter in terms of ACLR and EVM. For achieving faster tuning and for reducing the number of multiplications in a real application, 1 hidden layer is chosen. The optimization was continued by tuning the memory depth and keeping other parameters unchanged. Table E.2 shows how the NN is configured together with the ACLR and EVM results for the different number of memory depths. The best performance is achieved by setting memory depth to 5.

Table E.2: Configurations and Training for optimizing Memory depth.

Optimizer	Adam				
Activation function	ReLU				
Cost function	Huber				
Accelerator	BN				
Hidden layers	1				
Neurons	100				
Epoch	50				
Batch size	100				
Initial learning rate	0.1				
Minimum cost	$1E{-}6$				
Training Data	70%				
Validation data	30%				
Memory Depth	3	4	**5**	6	7
ACLR improvement	6.2 dB	6.3 dB	**6.5 dB**	6.2 dB	5.5 dB
EVM improvement	7.4%	8%	**8%**	8.1%	8%

Having memory depth fixed to 5, then other optimization parameters as activation function, cost function, batch size and the number of epochs have been tuned. The results are listed in Table E.3 and based on those parameters, the ReLU and the MSE are chosen for the activation function and the cost function, respectively. Further evaluation on the number of epochs shows that approximately 25 epochs are enough for the algorithm to reach its minimum cost of 1E -6 and the cost function will not improve further as shown in Fig. E.4.

4.3 NN simulation results

Fig. E.5 shows the simulation results of power spectral density (PSD), amplitude to amplitude (AM-AM) and amplitude to phase (AM-PM) distortions for the active array

Table E.3: NN parameter optimization. Parameters include activation function, cost function, batch size and epochs size.

Memory depth = 5				
Parameter optimization*				
Activation	Cost	Batch size	Epochs	ACLR improvement
ReLU	MSE	100	50	7.03 dB
ReLU	Huber	100	50	6.50 dB
ReLU	MSE	20	50	6.04 dB
Tanh	Huber	100	50	5.52 dB
ReLU	MSE	1000	500	3.03 dB
ReLU	MSE	1000	50	1.84 dB

*: Sorted based on decreasing ACLR improvement.

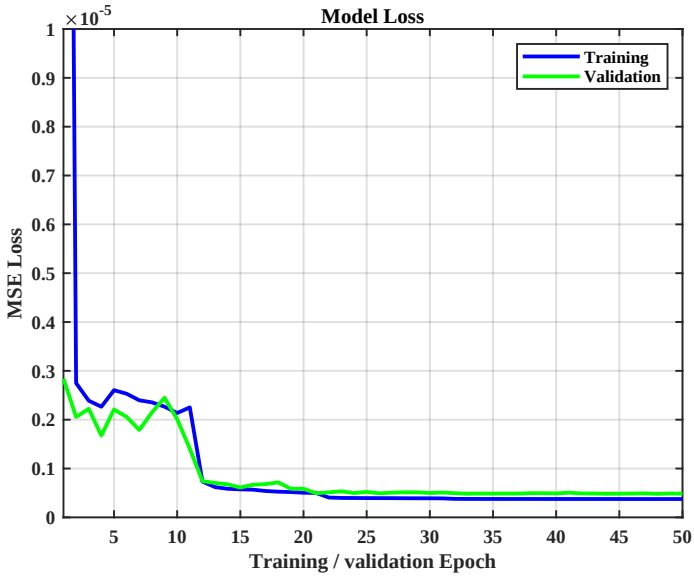


Fig. E.4: Model loss vs. number of training / validation epochs.

output. The parameters used for simulation are based on the best tuning parameters from Table B.2.

Several sets of predistorted signals have been trained based on the final model and have been used for characterizing the efficiency of the model versus the level of nonlinearity in the active array. These results are discussed in the next section.

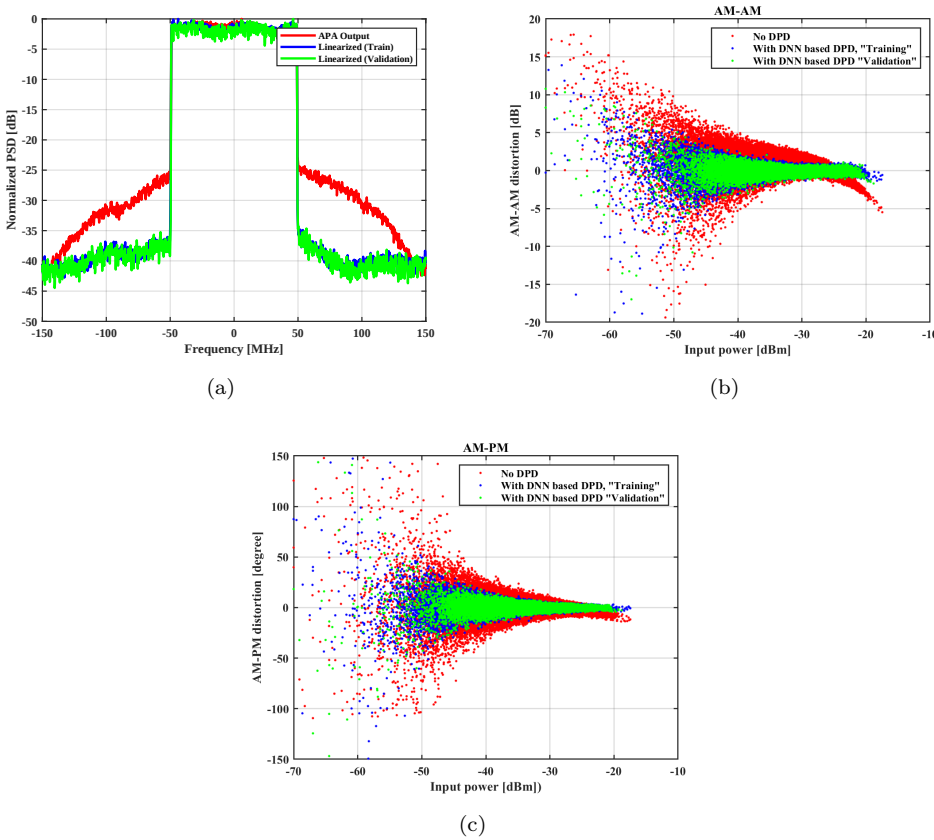


Fig. E.5: Simulated linearization results with and without NN predistortion: (a) Power spectral density; (b) AM-AM distortion; (c) AM-PM distortion.

5 Complexity Analysis

5.1 MP Predistortion

Eq. (B.4) models the behavior of the PA, which means the APA output can be estimated from the inputs. For predistortion, the inverse model is needed, which means the input should be estimated based on the output which is implemented by switching the input

and outputs:

$$x[n] = \sum_{m=0}^{M} \sum_{k=1}^{K} a_{mk} y[n-m] \cdot |y[n-m]|^{k-1}, \tag{E.13}$$

where $x[n]$ is the estimated input to the APA. This can be written as a vector-vector product on the form:

$$\mathbf{x} = \mathbf{r}\mathbf{w}, \tag{E.14}$$

where

$$\mathbf{r} = \begin{bmatrix} y[n-0] \cdot |y[n-0]|^{1-1} & \cdots & y[n-M] \cdot |y[n-M]|^{K-1} \end{bmatrix} \tag{E.15}$$

Since the absolute value, $|\cdot|$, requires 3 multiplications the complexity of finding $\mathbf{r}$ is thus:

$$\begin{aligned} C_{\text{r,complex}} &= (M+1) \cdot (0+3+4+5+\cdots+k+1) \\ &= \sum_{i=3}^{K+1} i \cdot (M+1) \end{aligned} \tag{E.16}$$

The complexity of the vector-vector product $x[n] = rw$ is simply:

$$C_{\text{Mul,MP,complex,vector}} = (M+1)K, \tag{E.17}$$

$$C_{\text{Add,MP,complex,vector}} = (M+1)K - 1. \tag{E.18}$$

where $C_{\text{Mul,MP,complex,vector}}$ is the number of complex multiplications and $C_{\text{Add,MP,complex,vector}}$ is the number of complex additions. The total number of complex multiplications becomes:

$$C_{\text{Mul,MP,complex,total}} = (M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1). \tag{E.19}$$

A complex multiplication takes four real multiplications and two real additions and a complex addition involves two real additions. This means the total complexity of the MPM predistortion in real operations is:

$$C_{\text{Mul,MP}} = 4 \cdot [(M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1)]. \tag{E.20}$$

$$C_{\text{Add,MP}} = 2 \cdot ([(M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1)] + [(M+1)K - 1]). \tag{E.21}$$

5.2 NN Predistortion

The complexity analysis is made with a starting point in Eq. (E.4) with L as the number of outputs of the previous layer, and P as the number of inputs to the next layer. If only fully connected layers with equal amounts of neurons are considered, the problem can be further reduced as $P = L$. Between each fully connected layer, there are P^2 multiplications and P^2 additions. The number of operations between the input layer and the first hidden layer is $2MP$ multiplications and additions, where M is the memory depth. Between the last hidden layer and the output layer, there are $2P$ multiplications and additions. Thus, the total amount of multiplications and additions is:

$$C_{\text{Mul,NN}} = C_{\text{Add,NN}} = 2MP + (J-1)P^2 + 2P, \tag{E.22}$$

where J is the number of hidden layers. Eq. E.22 shows that complexity scales quadratically with the number of neurons if there is more than one hidden layer. The complexity grows linearly with the number of neurons if only a single hidden layer is used. According to the universal approximation theorem, a single hidden layer can be used for arbitrary function approximation, so for applications where low complexity is required, a single hidden layer may be desirable.

5.3 Complexity Comparison

For the MPM the predistorted signals based on Eq. (B.4) with various nonlinearity orders, K, and memory depths, M, have been constructed and evaluated in the lab and the optimal values of K = 5 and M = 8 have been chosen. The NN is trained to reach the minimum MSE of approximately 1E-6 as an example shown in Fig. E.4. By sweeping the NN parameters, 1 hidden layer and 100 neurons and a memory depth of 5 have been chosen. Table E.4 shows the computational effort in term of multiplications and additions based on Eq. (E.20), Eq. (E.21) and Eq. (E.22). Although the number of multiplications is higher in the case of NN compare to MPM, the absolute number is still very low and besides NN has superior linearization performance which is shown in section 6.

Table E.4: Computational effort in term of multiplications and additions for MPM and NN.

	K	M	Layer	Neurons	Multiplications	Additions
MPM	5	8	-	-	828	502
NN	-	5	1	100	1200	1200

6 Measurement Results

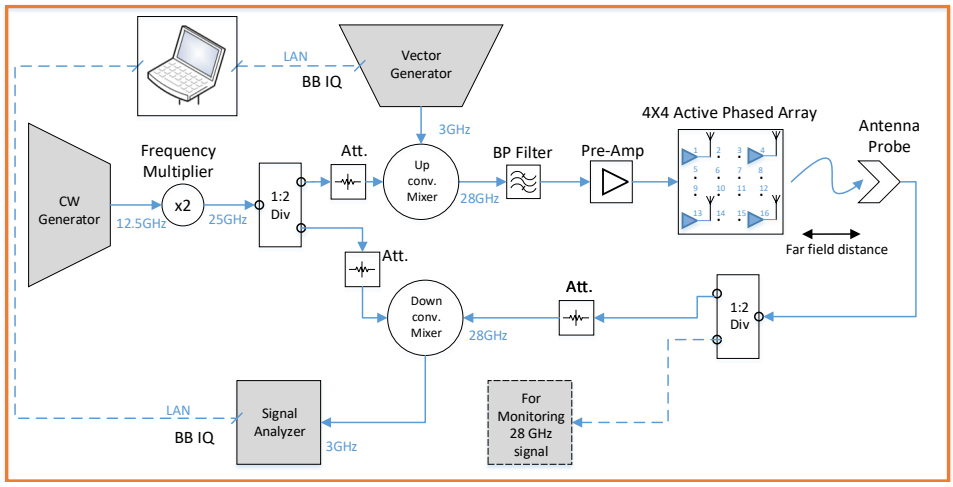
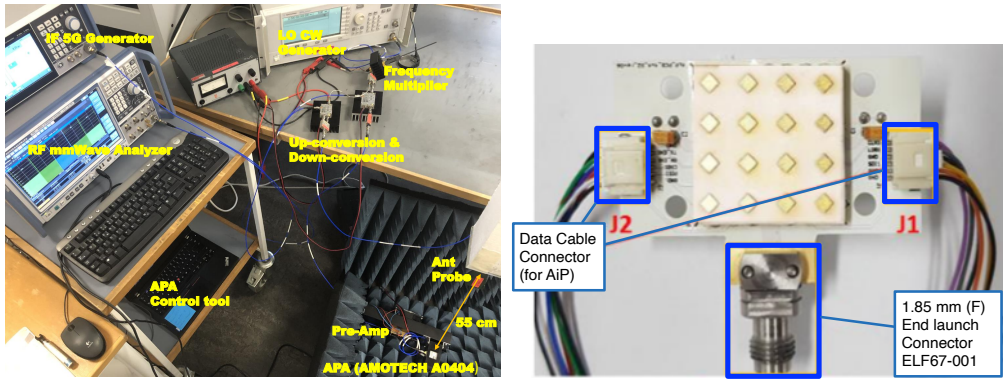


Fig. E.6: The block diagram of the measurement setup for the 4x4 array.



(a) (b)

Fig. E.7: (a): Measurement setup in Lab; (b): AAiPK428GC-A0404 evaluation board used for measurements [33].

6.1 OTA Measurement Setup

The block diagram of the measurement setup is shown in Fig. E.6 and the actual laboratory measurement set-up is illustrated in Fig. 11.

The R&S SMBV100B Vector Signal Generator and its arbitrary waveform generator function generate the TX input IF signal, centered at 3 GHz, which is a 100 MHz

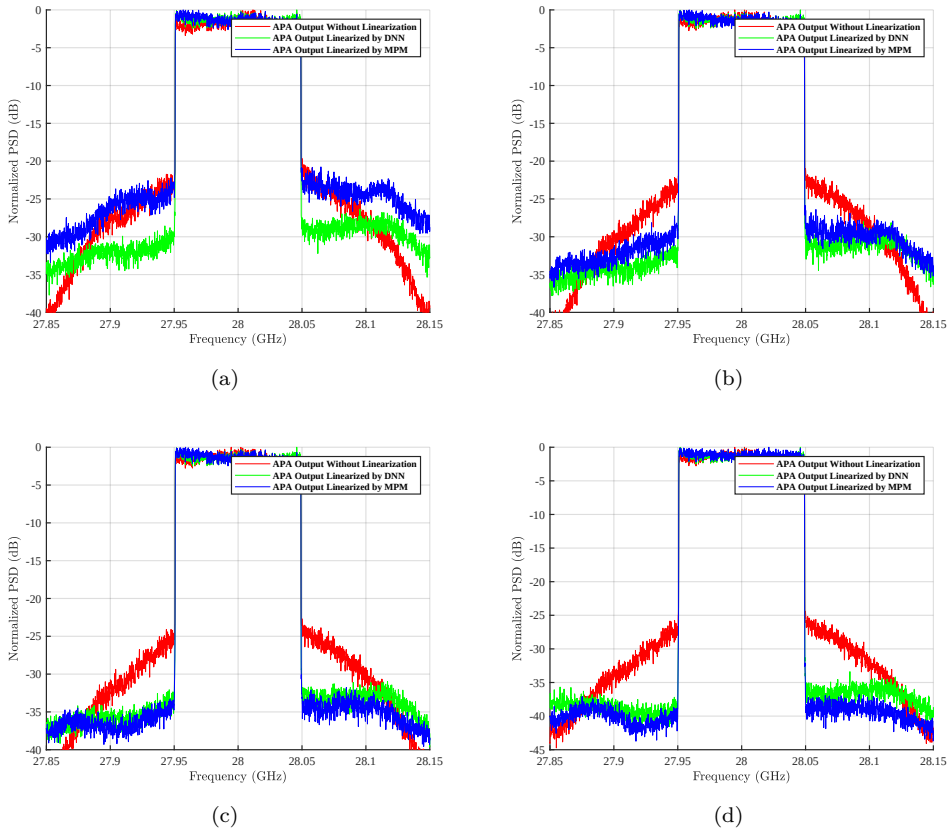


Fig. E.8: Measured OTA spectra with NN and MPM based linearization at different main beam power: (a) Very high nonlinearity case (main beam power = +34 dBm); (b) High nonlinearity case (main beam power = +33 dBm); (c) Medium nonlinearity case (main beam power = +32 dBm); (d) Low nonlinearity case (main beam power = +31 dBm).

bandwidth 5G NR signal. It is a 3GPP downlink OFDM modulated waveform with 64-QAM sub-carrier modulation, sub-carrier spacing of 60 kHz and 1584 active sub-carriers. With an oversampling factor of 6, the sample rate of the transmitter and receiver signals is 600 MHz. The peak to average power ratio (PAPR) of the input signal, after capturing and loading to the generator, is 11.6 dB. For up-conversion and down-conversion, an un-modulated signal of 12.5 GHz has been generated by an Agilent E3247C and frequency-doubled to 25 GHz using a MITEQ-MAX2M200400 and fed into a power divider to be used as a local oscillator (LO) signal. Two active mixers, KTX321840 and KRX321840,

operating in their highly linear region, are utilized for up converting the IF signal to the 28-GHz carrier frequency and for down-converting the signal back to IF. A 28 GHz band-pass filter is used to select the up-converted modulated signal and suppress the LO leakage and image frequency signals. The Ducommum APH-26063325 is used as a pre-amplifier. The pre-amplifier is a high power device and while operating more than 10 dB below its compression point, the output is linear and the power is sufficient to drive the 4× 4 APA, AAiPK428GC-A0404 [33], close to its saturated region. The APA includes four Anokiwave AWMF-0158 [34], integrates 16 branches of attenuators, phase shifters, PAs and 16 patch antennas in a 4×4 active phased array. The APA is designed for a typical main beam power of +33 dBm.

The diagonal length of the active array antenna is approximately 4 cm which at 28 GHz results in a far-field distance of:

$$\frac{2D^2}{\lambda} = 30.5 \quad \text{cm}, \tag{E.23}$$

where D is the diagonal length of the antenna and λ is the wavelength. The main beam signal is captured by the observation horn antenna placed 55 cm away which is well above the far-field distance of the device.

After down-conversion to IF, the signal is captured by the R&S FSW Signal and Spectrum Analyzer and converted to base-band. A host PC running in Matlab and using the R&S ARB Toolbox is used for capturing and uploading the I and Q samples. The measurement setup is power calibrated in order to keep all other components in their linear operating regions and the only source of nonlinearity is related to the active phased array. For controlling the main beam of the array the code-book and software tools from Amotech [33] have been used. For MPM, 4 predistorted signals, one for each power level, have been constructed based on a memory depth of 8 and nonlinearity order of 5. For NN, four corresponding predistorted signals based on the parameters in Table B.2 have been constructed and used as the input to the APA for the measurements.

6.2 OTA Measurement Results

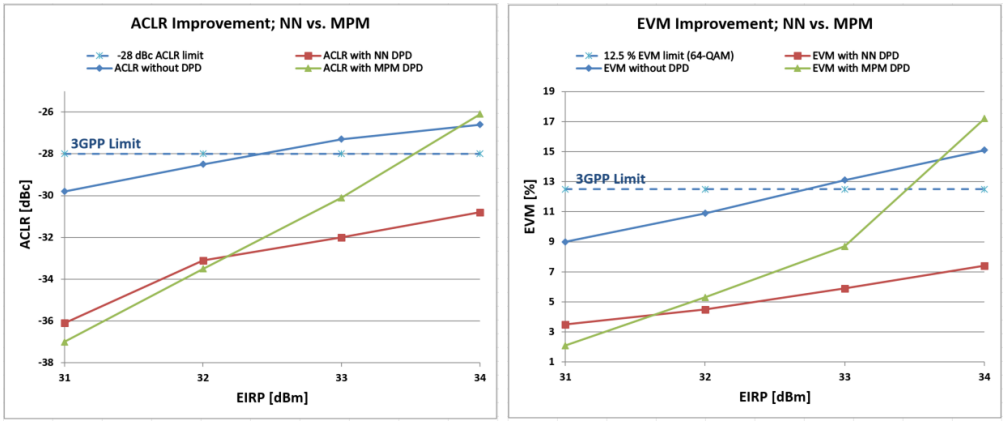
In this section, we present the experimental results of using the NN-based DPD and compare that with MPM-based DPD. In both cases, we do the measurements on four different power levels where the APA is driven into compression. The radiated power has been adjusted to get the ACLR slightly worse than the limit for systems operating at FR2 which is -28 dBc [35]. The main beam power of the APA is the sum of the transmitter power and gain of the antenna array:

$$P_{\text{main-beam}_{\text{APA}}} = P_{\text{TX}_{\text{APA}}} + G_{\text{Ant}_{\text{APA}}}, \tag{E.24}$$

and is measured as the power at the observation horn antenna placed 55 cm away, adding the propagation loss and subtracting the gain of the observation horn antenna.

The predistorted signals are fed as input to the APA and the corresponding ACLRs and EVMs are measured for each case and each technique.

The APA is driven in four different linearity cases with main beam power to be 34 dBm, 33 dBm, 32 dBm and 31 dBm for very high, high, medium and low power cases, respectively. Due to the limited output power of the 4x4 APA, we are not able to increase the main beam power further due to risk of damage to the device. However, with the main beam power of 34 dBm, the device is in the saturated region and the ACLR is approximately 2 dB worse than the 3GPP limit and suitable for our analysis. Measured OTA spectra with NN and MPM-based linearization for four different cases are shown in Fig. E.8 and the ACLR and EVM improvements are illustrated in Fig. E.9. Here we can see that for higher nonlinearities, the NN is performing better than MPM. In the case of low nonlinearity then MPM performance is equal or slightly better, but it is worth pointing out that linearization is less meaningful for relatively linear and less power-efficient operations. The results demonstrate that the proposed NN is capable of effectively learning the distinctive highly nonlinear distortions, which may not easily fit to existing MPM solutions.



(a) (b)

Fig. E.9: Comparison of NN vs. MPM: (a) ACLR comparison; (b) EVM comparison.

6.3 ACLR and Total Radiated Power (TRP)

Even though an existing reference claims the distortion is beam-formed in the same direction as the intended signal with a multi-antenna transmitter in the single-user case [5], quantitative results are desired to evaluate if using main beam ACLR is a valid method for characterizing the linearization performance of a beam steerable array. For

evaluating this, we performed a TRP ACLR measurement and compared the results with main beam direction measurements. The TRP is defined from the integration of signal power over the angular domains. The estimated TRP for a discrete set of measured directions is defined as [35]:

$$\mathrm{TRP}_{\mathrm{Estimate}} = \frac{\pi}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathrm{EIRP}(\phi_n, \theta_m) \cdot \sin\,\theta_m, \tag{E.25}$$

where N and M are the number of azimuth angles, ϕ_n, and elevation angles, θ_n, respectively and $\mathrm{EIRP}(\phi_n, \theta_m)$ is the radiated power in each angular case as a sum of both linear polarizations. The TRP-ACLR in linear scale is calculated as total radiated power of the adjacent channel divided by the total radiated power of the in-band channel:

$$\begin{aligned} \text{TRP-ACLR} &= \frac{\frac{\pi}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathrm{P}_{adj.ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m}{\frac{\pi}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathrm{P}_{ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m} \\ &= \frac{\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathrm{P}_{adj.ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m}{\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathrm{P}_{ch.}(\phi_n, \theta_m) \cdot \sin\,\theta_m} \end{aligned} \tag{E.26}$$

The block diagram and lab-setup for the measurements are shown in Fig. E.10 and E.11. The following procedure is applied for all specific angles θ and ϕ:

1. Place the APA at the positioner and align the coordinate system.
2. Align the beam of the APA to the desired beam steering angle.
3. Measure the main channel power and adjacent channel power using a spectrum analyzer.
4. Repeat steps 1-3 for all directions in the TRP measurement grid.
5. Calculate TRP-ACLR according to (E.26).

The steering angle of the APA is set to 0 degrees. The position of the APA is changed by θ from 0 to 180 degrees in steps of 10 degrees and for each step, the ϕ angle is changed from -90 degrees to 90 degrees in 20-degree steps. The in-band channel power and the ACLR power for each angular position have been measured and the TRP-ACLR level has been calculated according to (*E.26*) to be -35.0 dBc. For the same set-up, the main-beam-only level of ACLR is measured to be -33.3 dBc. In our work, we assume the main beam pointing is maintained in communication. Moreover, since the measured difference is less than 2 dB, the main beam ACLR is chosen as the metric for experimental validation in this work.

6.4 ACLR and Beam Directions

Due to interactions between the PAs in the array, the linearized beam is sensitive to steering angle, so the impact of beam steering needs to be quantified. A single trained DPD is not sufficient for maintaining a low distortion in a wide range of steering angles. To maintain a low level of distortion across the steering angle, a new training after some degrees shift (based on the actual setup) of the main beam is required [9]. The same investigation has been done in [4] with the same conclusion. Furthermore, remarkable results in [36] show that the NN is capable of modeling the correlation between the nonlinear distortion characteristics among different beams. This allows providing consistently good linearization regardless of the beamforming direction, thus avoiding the necessity of executing continuous digital predistortion parameter learning. In this paper, we have quantified the load modulation impact by measuring on an over-the-air test setup in a compact antenna test range (CATR). Fig. E.10 shows the setup used for measuring the APA output over the air in the CATR.

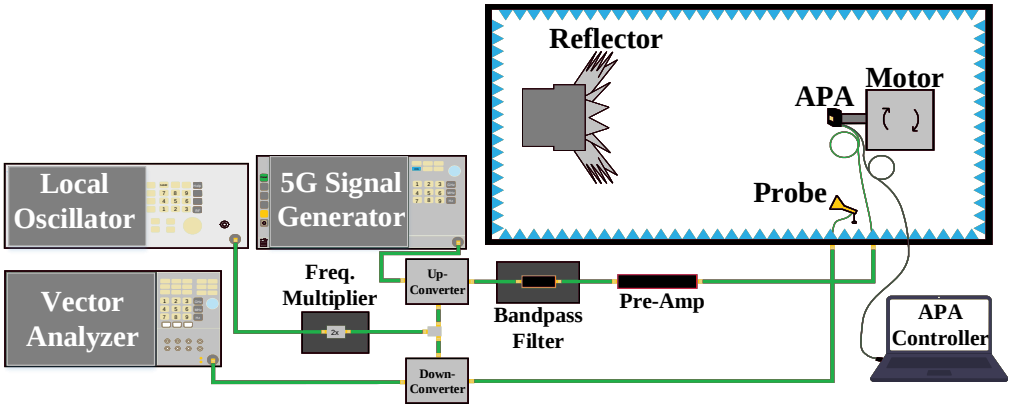


Fig. E.10: Block diagram of the measurement setup for measuring APA in compact antenna test range.

In this experiment, seventeen different values of the steering angle θ were used in the range of -78 to +78 degrees with a step of approximately 8 degrees. The following procedure is applied firstly to capture the nonlinear data for all angles and secondly the linearized data for all angles. For all specific angles θ_1 to θ_{17} following steps are used:

1. Adjust the steering angle to θ_i according to code-book and software tools.
2. Adjust the mechanical angle accordingly.
3. Measure input/output data for each steering angle.

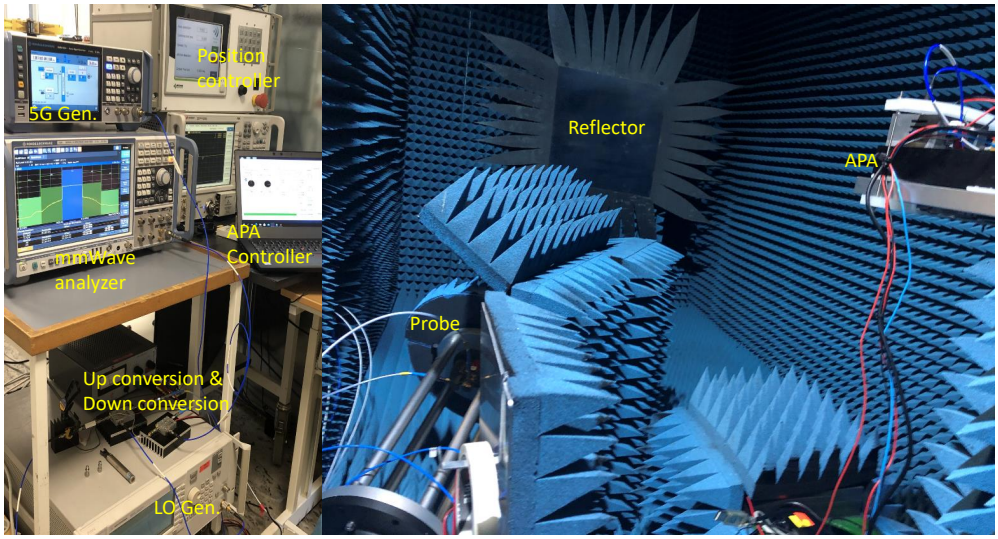


Fig. E.11: TRP measurement setup using compact antenna test range.

4. Make the MPM and the NN pre-distorters based on pre-distortion coefficients obtained from measurements of 0 degree steering angle.

5. Use the pre-distorter as input and repeat steps 1-3.

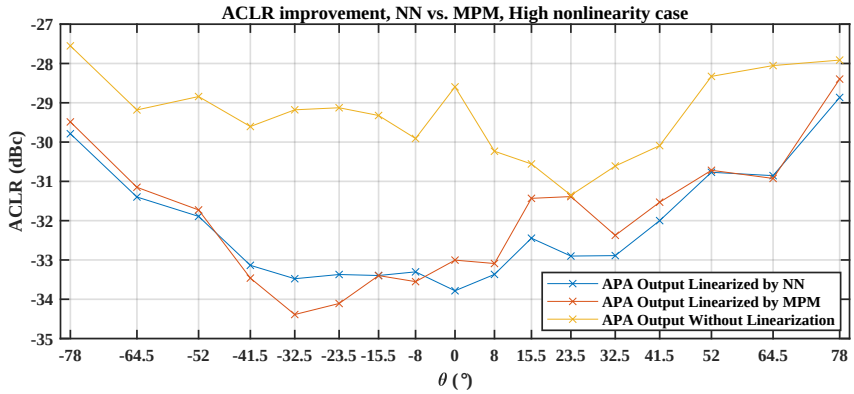


Fig. E.12: Measured ACLR performance of MPM and NN based DPD vs. steering angle using the OTA setup in compact antenna test range.

The results are shown in Fig. E.12, as measured for the same power level corresponding to the highly nonlinear use case depicted in Fig. E.8b. The ACLR of the APA without linearization is varying with the steering angle due to changes in radiation patterns and because of load modulation. Furthermore, the ACLR improvement rate for the linearized signals is varying with steering angle also as a result of load modulation.

6.5 Time Domain Comparison of NN and MPM Predistortion Signals

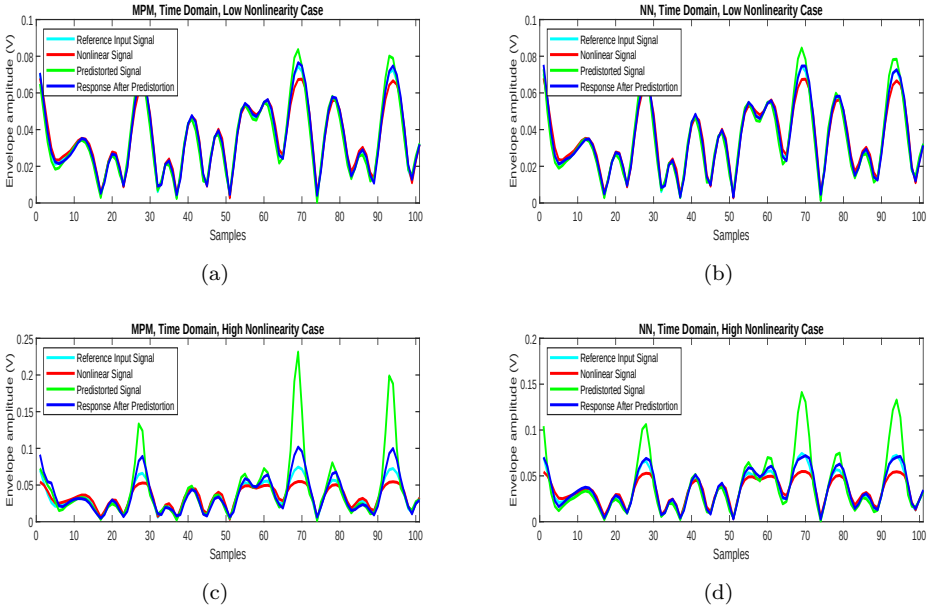


Fig. E.13: Time domain representation of predistorted signal: (a) MPM low nonlinearity case; (b) NN low nonlinearity case; (c) MPM high nonlinearity case; (b) NN high nonlinearity case.

In this section, we compare the predistorted signals of the NN and MPM DPD to understand better what the NN does differently. A possible explanation for this can be found by inspecting the signals in the time domain. Fig. E.13 shows the complex envelope of the reference input signal, the nonlinear signal, the predistorted signal and response after predistortion for the two power levels, 31 dBm and 34 dBm indicated as low and high linearity cases respectively. The gain is normalized to 0 dB for comparison.

The predistorted signal, as expected, has extra gain to counteract the decreasing nonlinear gain at the points where the nonlinear signal is in compression, which is illus-

trated as the high peaks in the time domain. As a consequence when the predistorted signal is applied, the response of the predistorted signal should ideally end up on top of the reference signal. This is exactly what happens in Fig. E.13a and b, where there is almost no difference between the reference input signal and the measured response after predistortion. The case of high nonlinearities can be seen in Fig. E.13c and d, where the difference between the reference input signal and the response after predistortion can now be easily observed. It is clearly seen from E.13c, that the MPM technique overcompensates the compression. When comparing the NN approach with the polynomial one the polynomials have inherent local approximating properties in contrast to the global approximation capability of NNs when modeling strongly nonlinear systems. Therefore NN may adapt better to extrapolating beyond the zone exploited for parameter extraction [37].

Although this effect is still under investigation, we see clearly the impact of the predistorted signal's PAPR on overall linearization and as a consequence, the shortcoming of the conventional and less complex MPM for linearizing highly nonlinear 5G modulated signals whereas a less complex NN based technique can do the job satisfactorily.

7 Discussion

The phenomenon of difference in the PAPR in MPM vs NN DPD was observed for all measurements where the output power of the APA in our setup is above +32 dBm. During experiments, we kept the PAPR of the input signal as defined in 3GPP i.e. 11.3 dB without providing any clipping and filtering method to reduce the PAPR. The root cause of the difference between MPM and NN approaches can be explained as:

1. The MPM nonlinearity kernel is a polynomial. For high nonlinearities high order polynomials are necessary. In our MPM we used a nonlinearity order equal to 5 and a memory depth of 8. When trying to linearize a highly compressed, deeply saturated PA characteristic, such a high order polynomial nonlinear function, quickly explodes at the upper side of the input amplitude range, thus causing the peaks of the predistorted signal to reach extremely high values, and hence leading to a huge increase in the predistorted signal PAPR compared to the PAPR of the original input modulated signal.

2. The nonlinear kernel of the NN does not contain inherently such an "explosion" effect for high amplitudes. It is important to keep in mind that the proposed RVTDNN structure is based on supervised learning. While we train the NN we are actually using low envelope fluctuations, i.e. the desired I and Q at the output layer in Fig. E.3, which allows the NN to learn the characteristics for a signal with low envelope fluctuations. This is also related to the fact that long-term memory is built into the RVTDNN through supervised learning. This kind of

long-term memory can be used to simulate the slow dynamic changes of nonlinear characteristics of the PA over time, mentioned in [1].

8 Conclusion

This paper presents how a neural network (NN) based linearization technique behaves on the digital predistortion (DPD) of a highly nonlinear active phased array (APA) using a wideband 3GPP 5G mmWave base-station transmitter signal and compares it to the used memory polynomial model (MPM) technique. The proposed design is implemented in a state-of-the-art 4×4 APA and a setup using up- and down-conversion from sub 6 GHz to 28 GHz and having high nonlinearity of the active phased array as the main impairment factor. The NN is built and trained using a Python simulation environment. The performance of the optimal NN predistorter was assessed with measurement results and compared to the MPM-based DPD technique. Measurement results on the proposed NN technique show that in the case of very high nonlinearity with an adjacent channel leakage ratio (ACLR) of -26 dBc, the predistortion signal generated by the NN exhibits peak to average power ratio (PAPR) much lower than the one generated by MPM and consequently is still capable to linearize the APA where it is not possible for the actual MPM technique. The proposed NN-based DPD technique applied on a highly nonlinear APA with an ACLR of -28 dBc shows an improvement of error vector magnitude (EVM) of 7.2 % points and ACLR of 4.7 dB. For the same set-up, an MPM-based DPD can only achieve an improvement of EVM and ACLR of 4.4 % points and 2.8 dB respectively. In the future, we may include an investigation of the robustness of NN-based linearization due to the steering angle and the impact of channel properties for the high nonlinearity cases.

Acknowledgements

The authors would like to acknowledge Jakob G. Brask, Kasper B. Olesen and Lauge F. Dyring, all of the University of Aalborg, Denmark, for providing technical and software support during OTA measurements.

References

[1] Taijun Liu, Slim Boumaiza, and Fadhel M Ghannouchi. Dynamic behavioral modeling of 3g power amplifiers using real-valued time-delay neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 52(3):1025–1033, 2004.

[2] Ulf Gustavsson, Pål Frenger, Christian Fager, Thomas Eriksson, Herbert Zirath, Franz Dielacher, Christoph Studer, Aarno Pärssinen, Ricardo Correia, João Nuno

Matos, et al. Implementation challenges and opportunities in beyond-5G and 6G communication. *IEEE Journal of Microwaves*, 1(1):86–100, 2021.

[3] Mahmoud Alizadeh, Peter Händel, and Daniel Rönnow. Behavioral modeling and digital pre-distortion techniques for RF PAs in a 3× 3 MIMO system. *International journal of microwave and wireless technologies*, 11(10):989–999, 2019.

[4] Eric Ng, Yehia Beltagy, Giovanni Scarlato, Ahmed Ben Ayed, Patrick Mitran, and Slim Boumaiza. Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets. *IEEE Transactions on Microwave Theory and Techniques*, 67(11):4479–4492, 2019.

[5] Christian Fager, Thomas Eriksson, Filipe Barradas, Katharina Hausmair, Telmo Cunha, and Jose Carlos Pedro. Linearity and efficiency in 5G transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5G active antenna transmitter context. *IEEE Microwave Magazine*, 20(5):35–49, 2019.

[6] Nuutti Tervo, Janne Aikio, Tommi Tuovinen, Timo Rahkonen, and Aarno Parssinen. Digital predistortion of amplitude varying phased array utilising over-the-air combining. In *2017 IEEE MTT-S International Microwave Symposium (IMS)*, pages 1165–1168. IEEE, 2017.

[7] Xin Liu, Qian Zhang, Wenhua Chen, Haigang Feng, Long Chen, Fadhel M Ghannouchi, and Zhenghe Feng. Beam-oriented digital predistortion for 5G massive mimo hybrid beamforming transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(7):3419–3432, 2018.

[8] Feridoon Jalili, Martin H Nielsen, Ming Shen, Ole K Jensen, Jan H Mikkelsen, and Gert F Pedersen. Linearization of active transmitter arrays in presence of antenna crosstalk for 5G systems. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, pages 1–5. IEEE, 2019.

[9] Feridoon Jalili, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Ming Shen, and Gert F Pedersen. Linearization trade-offs in a 5G mmwave active phased array ota setup. *IEEE Access*, 8:110669–110677, 2020.

[10] Katharina Hausmair, Ulf Gustavsson, Christian Fager, and Thomas Eriksson. Modeling and linearization of multi-antenna transmitters using over-the-air measurements. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE, 2018.

[11] Alberto Brihuega, Mahmoud Abdelaziz, Lauri Anttila, Matias Turunen, Markus Allén, Thomas Eriksson, and Mikko Valkama. Piecewise digital predistortion for

mmwave active antenna arrays: Algorithms and measurements. *IEEE Transactions on Microwave Theory and Techniques*, 68(9):4000–4017, 2020.

[12] Chao Yu, Jianxin Jing, Han Shao, Zhi Hao Jiang, Pinpin Yan, Xiao-Wei Zhu, Wei Hong, and Anding Zhu. Full-angle digital predistortion of 5G millimeter-wave massive mimo transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 67(7):2847–2860, 2019.

[13] Alberto Brihuega, Mahmoud Abdelaziz, Lauri Anttila, Yue Li, Anding Zhu, and Mikko Valkama. Mixture of experts approach for piecewise modeling and linearization of rf power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 2021.

[14] YH Ku. The volterra & wiener theories of nonlinear systems: Martin schetzen. 531 pages, diagrams, illustr., John Wiley, New York, 1980., 1982.

[15] Changsoo Eun and Edward J Powers. A new volterra predistorter based on the indirect learning architecture. *IEEE Transactions on Signal Processing*, 45(1):223–227, 1997.

[16] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[17] Dongming Wang, Mohsin Aziz, Mohamed Helaoui, and Fadhel M Ghannouchi. Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters. *IEEE transactions on neural networks and learning systems*, 30(1):242–254, 2018.

[18] Praveen Jaraut, Meenakshi Rawat, and Fadhel M Ghannouchi. Composite neural network digital predistortion model for joint mitigation of crosstalk, I/Q imbalance, nonlinearity in MIMO transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(11):5011–5020, 2018.

[19] Yikang Zhang, Yue Li, Falin Liu, and Anding Zhu. Vector decomposition based time-delay neural network behavioral model for digital predistortion of rf power amplifiers. *IEEE Access*, 7:91559–91568, 2019.

[20] Joel HK Vuolevi, Timo Rahkonen, and Jani PA Manninen. Measurement technique for characterizing memory effects in rf power amplifiers. *IEEE Transactions on microwave theory and techniques*, 49(8):1383–1389, 2001.

[21] Qingyue Chen, Yufeng Zhang, Feridoon Jalili, Zhugang Wang, Yonghui Huang, Yubo Wang, Ying Liu, Gert Frølund Pedersen, and Ming Shen. Robust digital signal recovery for leo satellite communications subject to high snr variation and transmitter memory effects. *IEEE Access*, 2021.

[22] Meenakshi Rawat, Karun Rawat, and Fadhel M Ghannouchi. Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 58(1):95–104, 2009.

[23] Changliu Liu, Tomer Arnon, Christopher Lazarus, Christopher Strong, Clark Barrett, and Mykel J Kochenderfer. Algorithms for verifying deep neural networks. *arXiv preprint arXiv:1903.06758*, 2019.

[24] Fadhel M Ghannouchi, Oualid Hammi, and Mohamed Helaoui. *Behavioral modeling and predistortion of wideband wireless transmitters*. John Wiley & Sons, 2015.

[25] Dennis R Morgan, Zhengxiang Ma, Jaehyeong Kim, Michael G Zierdt, and John Pastalan. A generalized memory polynomial model for digital predistortion of RF power amplifiers. *IEEE Transactions on signal processing*, 54(10):3852–3860, 2006.

[26] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[27] Pere L Gilabert, David López-Bueno, Thi Quynh Anh Pham, and Gabriel Montoro. Machine learning for digital front-end: a comprehensive overview. *Machine Learning for Future Wireless Communications*, pages 327–381, 2020.

[28] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[29] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[30] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[31] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[32] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[33] Amotech. *AAiP28G-A0808-EVB-R01_190321c*, 2020.

[34] Anokiwave. *AWMF-0158 28 GHz Silicon 5G Tx/Rx Quad Core IC*, 2019.

[35] Evolved Universal Terrestrial Radio Access. Base station radio transmission and reception. volume 36, pages 96–97, 2018.

[36] Alberto Brihuega, Lauri Anttila, and Mikko Valkama. Neural-network-based digital predistortion for active antenna arrays under load modulation. *IEEE Microwave and Wireless Components Letters*, 30(8):843–846, 2020.

[37] Pere Lluís Gilabert Pinal. *Multi look-up table digital predistortion for RF power amplifier linearization.* Universitat Politecnica de Catalunya, 2008.

Paper F

Complexity Analysis of Artificial Neural Networks Used for Active Phased Array Linearization

Feridoon Jalili, Felice F. Tafuri, Kasper B. Olesen, Lauge F. Dyring, Jakob G. Brask, Ming Shen and Gert F. Pedersen

The paper has been submitted to the
2022 IEEE MTT-S International Wireless Symposium (IWS) Year: 2022 | Conference Paper | Publisher: IEEE.

Abstract

This paper describes a complexity analysis of an artificial neural network (ANN) which allows illustrating the trade-offs in choosing model parameters of digital predistorters for multiple power levels of a transmitter system and can be used for applications such as look-up table based digital predistortion. The ANN architecture is based on real-value time-delayed neural networks (RVTDNNs). In this work, the authors are providing an analysis that includes the number of neurons and the number of time-delay lines versus linearization quality, in the form of minimum in-band and out-of-band distortions. For each case, the implementation complexity in form of the number of multiplications and additions has been quantified and the optimum model has been compared with a traditional memory polynomial model (MPM) by measurements. We demonstrate the performance of the ANN method by using a 28 GHz, 100 MHz bandwidth, 3GPP base-station over-the-air (OTA) setup. For the measurement, a sub 6 GHz signal is up-converted to an RF frequency of 28 GHz and used in a state-of-the-art 4×4 steerable active phased array (APA) transmitter.

1 Introduction

Active phased array (APA) transmitters including multiple antennas operating at mmWave frequencies, which are used in the recent wireless communication systems, are facing new challenges in forms of high bandwidth, high nonlinearity and mutual coupling between antennas. Digital predistortion (DPD) techniques based on conventional techniques such as the memory polynomial model (MPM) can not easily handle these new challenges without increasing the complexity. Rich experimental results are showing that the artificial neural network (ANN) based DPD approach requires fewer parameters when compared to MPM based approach, whose number of coefficients increases exponentially with the number of antennas [1]. However, an indication of how the ANN architecture compares with the classical MPM based DPD approach in terms of complexity, specifically for the recent proposed 5G multiple-input multiple-output (MIMO) transmitters architecture with wide bandwidth and high nonlinearity, is still an open question.

In this paper, we focus on the complexity of the ANN network for DPD in terms of the number of multiplications, and consequently implementation cost. For doing this, the number of multiplications, which is related to the number of neurons at each hidden layer and the depth of the delay line, have been bench-marked. Fig. F.1 illustrates the proposed ANN-based linearization technique. We define the mapping relationship between the order of time-delays, the number of hidden layers and the number of neurons to the corresponding required linearity. We developed a predistorted signal modeling (PSM) tool for finding the required number of hidden layers, time-delays and neurons

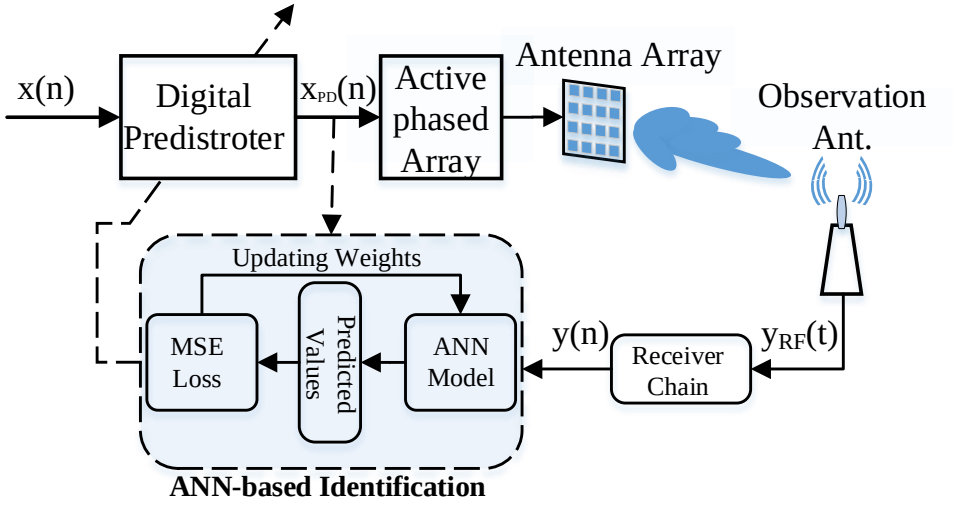


Fig. F.1: Proposed active array transmitter with ANN based linearization.

for achieving a target linearity performance. We afterward evaluate the optimal configuration found by PSM during measurements. Finally, to the best of the authors' knowledge, the first time a complexity analysis which considers the number of multiplications between ANN and conventional MPM has been introduced using a highly nonlinear and wide-band 5G test-bed measured over-the-air (OTA).

2 The proposed ANN Model

For wideband signals, in particular, the memory effects have a significant impact. There are generally two dynamic neural structures for taking care of memory effects [2]. The first structure, recurrent neural networks (RNNs), utilizes feed-forward and feedback signal processing and uses output-to-input time-delays lines. In another structure, a time-delay neural network (TDNN), combines I/Q processing with input time-delay lines to handle memory effects. To extract amplitude and phase information from modulated complex waveforms, ANNs need to consider operating with either complex-valued (CV) input signals, weights and activation outputs, or real-valued (RV) double-inputs double-outputs (and real weights and activation outputs), i.e. in the form of multiple I and Q components. CV operation leads to heavy calculations and a longer training phase [3] and therefore we have chosen the RV concept in our model. The real-valued time-delay neural networks (RVTDNNs) offer superior performance and easy baseband implementation when used for inverse modeling of PAs with strong nonlinearities and

memory effects [1].

2.1 Implementation of the proposed ANN

Fig. F.2 shows the proposed ANN model based on RVFTDNN used in this work where a different number of memory taps can be assessed and the same taps configuration is employed between input and feedback signals regardless of the physics to be modeled [4, 5]. The model has been implemented using the Tensorflow Keras package in Python. The proposed ANN has a fully-connected structure which is characterized by the input-output relationship between the hidden layers [6]:

$$\mathbf{y}^{(k)} = f\left(\mathbf{W}\mathbf{x}^{(k-1)} + \mathbf{B}\right), \tag{F.1}$$

where f is the activation function, $\mathbf{y}^{(k)} \in \mathbb{R}^{P\times 1}$ are the output values of the k'th layer, $\mathbf{W} \in \mathbb{R}^{P\times Q}$ are trainable coefficients, $\mathbf{x}^{(k-1)} \in \mathbb{R}^{Q\times 1}$ are the outputs of the previous layers and $\mathbf{B} \in \mathbb{R}^{P\times 1}$ are trainable biases. Thus, Q is the number of outputs of the previous layer, and P is the number of inputs to the next layer. The activation function, denoted as f in Fig. F.2 is what makes the ANN able to fit arbitrary nonlinear functions. It uses the ReLU activation function, which is a computationally efficient nonlinear function and which is piecewise linear [7]. The ReLU activation function is defined by:

$$\sigma_{\mathrm{ReLU}}(x) = \max(0, x) \tag{F.2}$$

ReLU activation function introduces nonlinearity by setting negative inputs to 0, which also adds sparsity to the ANN and can simplify the computations.

2.2 Complexity analysis of the proposed ANN

The complexity analysis is made with a starting point in Eq. F.1. Assuming only fully connected layers with equal amounts of neurons are considered, the problem can be further reduced as $P = Q$. Between each fully connected layer, there are P^2 multiplications and P^2 additions. The number of operations between the input layer and the first hidden layer is $2MP$ multiplications and additions, where M is the number of time-delays. Between the last hidden layer and the output layer, there are $2P$ multiplications and additions. Thus, the total amount of multiplications and additions is:

$$C_{\mathrm{m,DNN}} = C_{\mathrm{a,DNN}} = 2MP + (J-1)P^2 + 2P, \tag{F.3}$$

where J is the number of hidden layers.

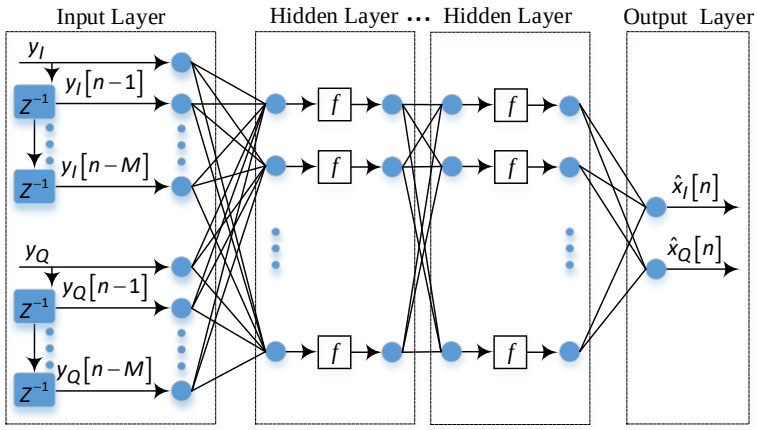


Fig. F.2: The proposed artificial neural networks model.

3 Complexity analysis of MPM

The applied MPM, which is a deviation of the Hammerstein model and has been proven effective for removing nonlinearity and memory effects, is represented in Eq. F.4 [8]:

$$x[n] = \sum_{m=0}^{M} \sum_{k=1}^{K} a_{mk} y[n-m] \cdot |y[n-m]|^{k-1}, \tag{F.4}$$

where $x[n]$ is the estimated input to the PA, and a_{mk} are the coefficients that describe the relationship between the output and input of the PA, given for all combinations of memory depth and nonlinearity orders. This technique for inverting the PAs nonlinear response by directly modeling the inverse model is called postdistortion, as the model is directly made with the input/output relationship of a known data set from the APA. The validity of this technique is given in [8]. A detailed insight into the linearization mechanisms for an APA based on the MPM model is provided in [9]. Based on a direct implementation of Eq. F.4 the total amount of real multiplications, $C_{\mathrm{m,MPM}}$, can be defined as:

$$C_{\mathrm{m,MPM}} = 4[(M+1)K + \sum_{i=3}^{K+1} i \cdot (M+1)], \tag{F.5}$$

4 OTA Measurement Setup

The setup of the OTA measurement in compact antenna test range (CATR) chamber is shown Fig. F.3.

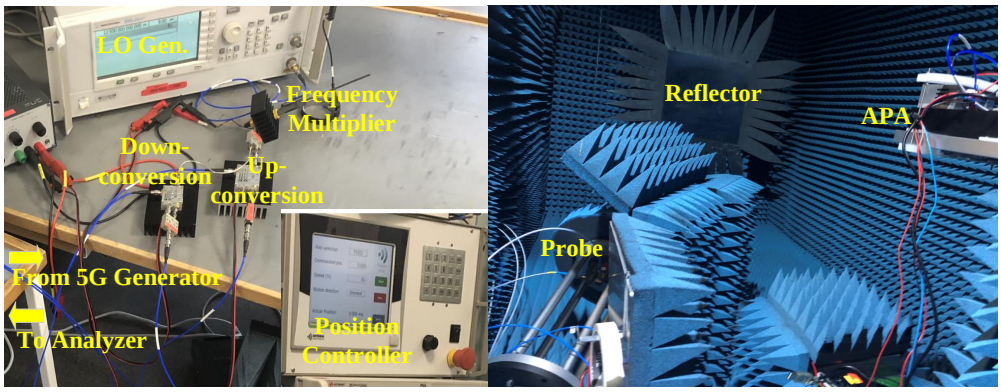


Fig. F.3: Measurement setup using compact antenna test range chamber.

The APA device is based on Amotech AAiPK428GC-A0404 which includes four Anokiwave AWMF-0158 transceivers [10]. It integrates 16 branches of attenuators, phase shifters and PAs and 16 patch antennas in a 4×4 active phased array. A host PC is used for capturing and uploading the IQ samples. The measurement setup is power calibrated in order to keep all other components in their linear operating regions and the only source of nonlinearity is related to the active phased array. For controlling the main beam of the array the code-book and software tools of Amotech have been used.

5 Optimization of the ANN Parameters

In order to reduce the complexity of the ANN, the minimum number of time-delays and neurons need to be identified while the desired linearization levels in terms of adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) are maintained. It is chosen to restrict the DNN to a single hidden layer to reduce the complexity, as the complexity is squared on the number of hidden layers (Eq. F.3). 100 k I/Q samples of the input and output signals are captured, time-aligned, and used to train a number of ANN predistorters. The dynamic range of the time-delays is from 3 to 17, and the range of neurons is from 40 to 480. The optimization results are assessed by training the network to use 70 % of the I/Q data for the training and 30 % for the validation. Fig. F.4 shows the optimization results, where the optimal choice is a trade-off between the ACLR and the number of multiplications. By keeping the number of time-delays

to 5 and the number of neurons to 80, it is possible to achieve an average left / right sides ACLR of −36.8 dBc corresponding to 960 multiplications. Increasing the number of time-delays to 15 will increase the number of the multiplications to 2560, without further improvement of ACLR.

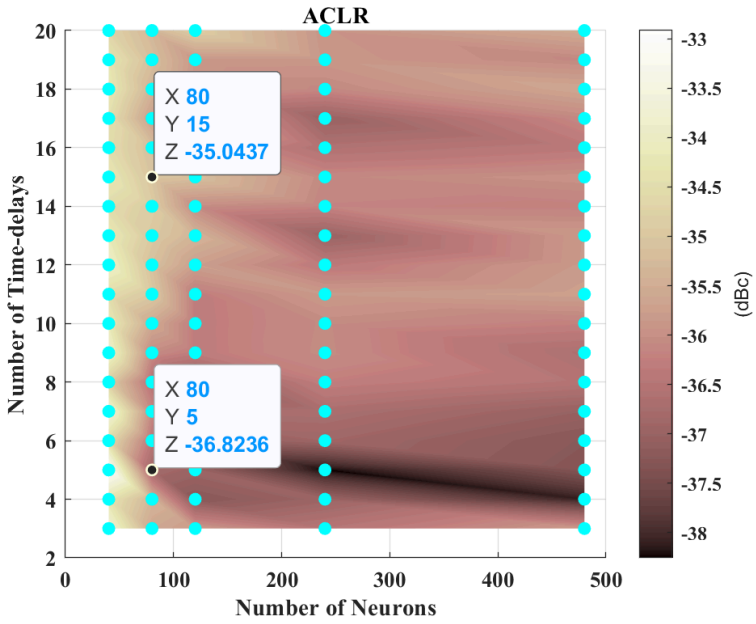


Fig. F.4: The ACLR values obtained by varying ANN parameters.

6 Measurement Results

In this section, we present the experimental results of using the ANN-based DPD and compare that with MPM-based DPD. In both cases, the predistorted signals from the ANN and the MPM models are fed as input to the APA, and the corresponding ACLRs and EVMs are measured for each case and each technique. For the MPM, a nonlinearity order of 5 and memory depth of 8 has been chosen as the optimum model parameters based on results in [9]. For ANN, based on the optimization results in Section 5, time-delays of 5 and neurons-number of 80 have been chosen. The measured OTA spectra are shown in Fig. F.5, which indicate comparable results between these two techniques.

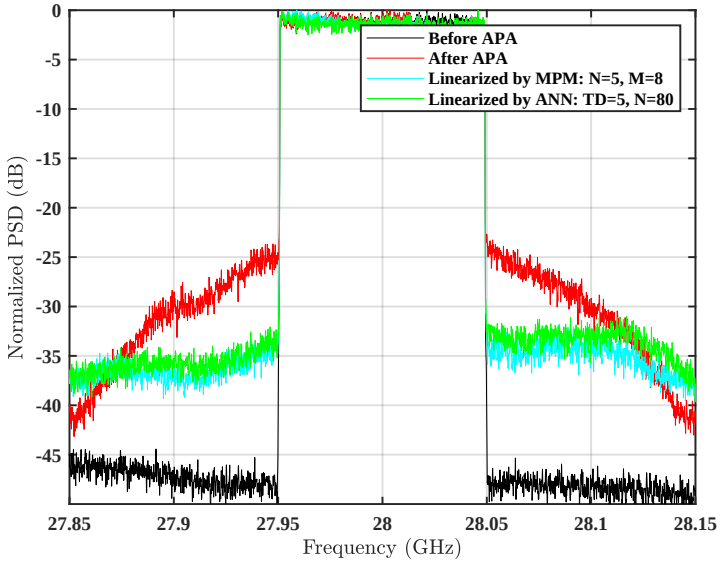


Fig. F.5: Measured OTA spectra, ANN-based vs. MPM-based linearization.

7 Conclusion

This paper presents a complexity analysis of a proposed real-value time-delay neural networks (RVTDNNs) used for digital predistortion (DPD) of a 5G active phased array transmitter. The complexity analysis can be used for applications such as look-up table based DPD. We introduced an optimization procedure to achieve the desired linearity with a minimum number of multiplications. The achieved linearity is compared with the conventional memory polynomial model (MPM) in terms of complexity. In our opinion, the DPD complexity-performance trade-off is correlated to the hardware we are using and to the multiple-input multiple-output (MIMO) nature of our APA. However, the methodology can be used by System Engineers to assess the optimal complexity-performance trade-off to be used when designing their DPD. The strength of the ANN model is its ability to provide the target linearity performance without necessarily increasing the complexity in the forms of the number of neurons, hidden layers and time-delays. A future work based on the proposed RVTDNN model may use transfer-learning approach for further reduction of complexity and training time.

References

[1] Praveen Jaraut, Meenakshi Rawat, and Fadhel M Ghannouchi. Composite neural network digital predistortion model for joint mitigation of crosstalk, i/q imbalance, nonlinearity in mimo transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(11):5011–5020, 2018.

[2] Taijun Liu, Slim Boumaiza, and Fadhel M Ghannouchi. Dynamic behavioral modeling of 3g power amplifiers using real-valued time-delay neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 52(3):1025–1033, 2004.

[3] Pere L Gilabert, David López-Bueno, Thi Quynh Anh Pham, and Gabriel Montoro. Machine learning for digital front-end: a comprehensive overview. *Machine Learning for Future Wireless Communications*, pages 327–381, 2020.

[4] Feridoon Jalili, Yufeng Zhang, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Qingyue Chen, Ming Shen, and Gert F. Pedersen. Tuning of deep neural networks for over-the-air linearization of highly nonlinear wide-band active phased arrays. In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, pages 1–4, 2021.

[5] Jakob Gjedsted Brask, Kasper Bruun Olesen, Lauge Føns Dyring, Arun Yadav, Feridoon Jalili, Yufeng Zhang, and Ming Shen. Deep digital signal recovery for leo satellite communication in presence of system perturbations. In *2021 IEEE MTT-S International Wireless Symposium (IWS)*, pages 1–3. IEEE, 2021.

[6] Timothy O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

[7] Xavier Glorot, Antoine Bordes, and Y. Bengio. Deep sparse rectifier neural networks. *Journal of Machine Learning Research*, 15, 2010.

[8] Dennis R Morgan, Zhengxiang Ma, Jaehyeong Kim, Michael G Zierdt, and John Pastalan. A generalized memory polynomial model for digital predistortion of RF power amplifiers. *IEEE Transactions on signal processing*, 54(10):3852–3860, 2006.

[9] Feridoon Jalili, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Ming Shen, and Gert F Pedersen. Linearization trade-offs in a 5g mmwave active phased array ota setup. *IEEE Access*, 8:110669–110677, 2020.

[10] Anokiwave. AWMF-0158 28 GHz Silicon 5G Tx/Rx Quad Core IC. `https://www.anokiwave.com/products/awmf-0158/index.html`, 2019.

Paper G

Bandwidth Scalable Behavioral Modeling using Neural Network based on Transfer Learning

Feridoon Jalili, Felice F. Tafuri, Qingyue Chen, Ming Shen and Gert F. Pedersen

The paper to be submitted to the
To be submitted to the IEEE Access Year: 2022 | Volume: 8 | Journal Article.

Abstract

This paper proposes a transfer learning neural network (TLNN) approach for digital pre-distortion (DPD) of mm-Wave active phased arrays (APA) operated under variable signal bandwidth regimes. Compared with the conventional ANN method, the proposed approach can achieve similar linearization performance with much lower computational complexity by transferring part of a trained model from one bandwidth to another bandwidth. In the recently introduced 5G, the increased signal bandwidth triggers considerable memory effects in the APA. Moreover, dealing with different signal bandwidths typically requires a time-consuming recalculation of the predistorter parameters. In this paper, the authors propose to have those challenges solved by using a DPD model based on the transfer learning method. The proposed approach was validated with over-the-air (OTA) measurements on an APA excited with signals of varying bandwidth, namely from 20 MHz to 100 MHz. Experimental results show a significant reduction in the training time while ensuring good linearization performance. With the applied TLNN DPD, 8.5 dB improvement of adjacent channel leakage ratio (ACLR) and 8.6 % points improvement of error vector magnitude (EVM) is achieved. Under the variable bandwidth regime, the complexity of the DPD model in terms of the number of multiplications is reduced from 199168 to 160. The proposed TLNN DPD proved to be robust concerning variation in the bandwidth of the APA excitation signal.

1 Introduction

Active phased array (APA) transmitters including multiple antennas operating at mmWave frequencies, which are used in the recent wireless communication systems, are facing new challenges in forms of high bandwidth, high nonlinearity and mutual coupling between antennas together with dynamic change of the bandwidth. Digital predistortion (DPD) techniques based on conventional methods can not easily handle these new challenges without increasing the computational complexity. Together with the wide bandwidth, 5G has introduced a dynamic bandwidth selection which requires the mobile transmitter to quickly adapt to different operating conditions. Dynamic bandwidth selection together with the impact from the transmission channel makes the need for reusing the adjusted parameters defined for calibration, linearization, etc. highly important [1]. The transmission quality of the communication system is to a high degree dependent on how well it can dynamically change the bandwidth and power level with minimum cost in terms of speed and cost. The state-of-the-art (SoA) DPD systems deployed by the industry have excellent performance for relatively steady conditions where the bandwidth and power are not rapidly changed. For the cases with a rapid change of transmission parameters and environment, the existing DPD methods need to update a huge amount of coefficients which potentially can make the system complex and slow.

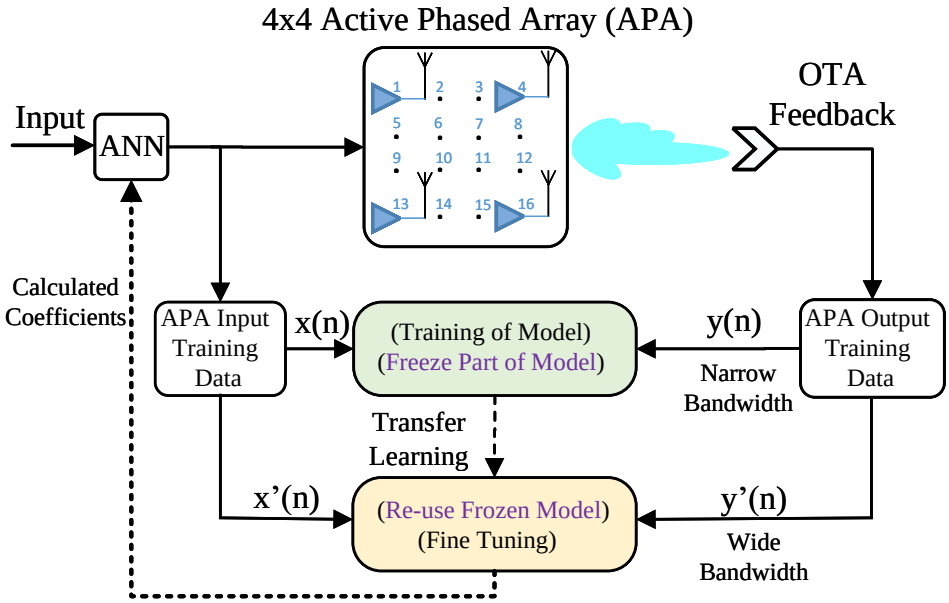


Fig. G.1: TLNN-based linearization model.

Artificial neural networks (ANN) have been widely used in the modeling of nonlinear devices because of its good approximation ability to nonlinear functions [2]. For wide bandwidth signals, in particular, the memory effects have a significant impact. There are generally two dynamic neural network structures for taking care of memory effects [3]. The first structure, recurrent neural networks (RNNs), utilizes feed-forward and feedback signal processing and uses output-to-input time-delays lines. In another structure, a time-delay neural network (TDNN), combines I/Q processing with input time-delay lines to handle memory effects. In order to extract amplitude and phase information from modulated complex wave-forms, ANNs need to consider operating with either complex-valued (CV) input signals, weights and activation outputs, or real-valued (RV) double-inputs double-outputs (and real weights and activation outputs), i.e. in the form of multiple I and Q components. CV operation leads to heavy calculations and a longer training phase [4] and therefore the proposed model in this work uses the RV concept. The real-valued time-delay neural networks (RVTDNNs) offer superior performance and easy baseband implementation when used for inverse modeling of PAs with strong nonlinearities and memory effects [5].
However, by increasing the bandwidth and nonlinearity, the RVTDNN requires a higher input dimension, i.e. larger number of IQ data, and more hidden layers which make the model slow. Several works based on transfer learning have been introduced to cope

with these challenges [5, 6]. The study of transfer learning is motivated by the fact that one can intelligently apply knowledge learned previously to solve new problems faster or with better solutions [7]. A similar problem also lies in the way of other dense ANN networks with several layers and neurons used in image recognition [8] and channel estimation [9, 10]. In these works, the transfer learning techniques grant the models the ability of rapid image recognition and channel estimation by leveraging prior knowledge. Inspired by these works, this paper investigates applying transfer learning DPD for bandwidth-scalable active phased arrays. Fig. G.1 shows the block diagram of the actual transfer learning neural network (TLNN) linearization technique. Part of the narrow bandwidth model from the previous training has been transferred and combined with the fine-tuning layers to make the new model for the wide bandwidth.

This paper is organized as follows: Section 1 is the introduction. Section 2 presents the proposed linearization method. The measurement setup is in section 3.1. The optimization of the pre-designed model and the reference model is described in section 5. Section 5 is about transfer learning implementation. Bandwidth-scalable predistortion results are shown in section 6 and finally, the conclusion of this work is presented in section 7.

2 Proposed TLNN Linearization Method

This section describes the selected model for linearization, the data structure and architecture of the model together with a complexity analysis of the proposed neural network.

2.1 SISO model for TLNN-based linearization

Several modified DPD algorithms have been introduced to combat the challenges raised by the recently introduced hardware configuration for 5G mmWave transmitter based on the active phased array [11–13]. A single input single output (SISO) model where the entire transmitter has been considered as a two-port system has been presented using an observation receiver in far-field In [14–16]. An MPM-based DPD technique based on this SISO model has been used for the linearization of the antenna array in presence of crosstalk. It has been shown that the trained DPD is able to mitigate the impact of cross-talk at PAs outputs, which is also called load modulation, in a limited range of steering angle. The step size for reusing the trained model is dependent on the target specification of linearity and the amount of coupling among the branches of the active phased array which again is dependent on the size of the array and the distance between the patches [17]. The potential mismatches between PAs can be compensated so that they all exhibit the very same behavior which is presented in [18]. In this way, linearization in all directions can be achieved with a single DPD, in contrast to linearizing the main beam only. However, this approach requiring analog circuits for

compensating the mismatch in each branch which may introduce high complexity and delay for large arrays and the potential changes in the PAs' behaviors due to crosstalk. In the present work, based on the SISO model, the reference signal for DPD identification is obtained through far-field measurements of an observation antenna placed at the main beam direction, Fig. D.1, and the focus here is on the challenges related to high bandwidth and dynamic bandwidth behaviour.

2.2 Data structure of the model

The data structure of the exploited TLNN is shown in Fig. G.2, where $y_I(n)$ and $y_Q(n)$ are the I/Q components of input to the ANN and $\hat{x}_I(n)$ and $\hat{x}_Q(n)$ are the I/Q components of the output of the network. The data format of the source and target datasets is the same, and the inputs and outputs are represented as:

$$\begin{aligned} Y_n = &[y_I(n), y_I(n-1), ..., y_I(n-M), \\ &y_Q(n), y_Q(n-1), ..., y_Q(n-M)] \end{aligned} \tag{G.1}$$

and

$$X_n = [\hat{x}_I(n), \hat{x}_Q(n)]\,, \tag{G.2}$$

where M denotes the number of delay lines at the input of the network. The procedure for training is as follows: a set of source datasets, e.g. measured IQ samples of a 5G signal with 20 MHz channel bandwidth, are used for offline training. Part of the network is then used as a transfer learning model for the target dataset, which is a 5G signal that can have the same or different channel bandwidth. As illustrated in Fig. G.2, the first k layers of the model, FC_k, are used for extracting the nonlinear characteristics of the APA in low bandwidth cases and are frozen after executing offline training. The output of the frozen layers, T_n, is written as:

$$T_n = f^{frozen}\left(X_n\right). \tag{G.3}$$

Here, $f^{frozen}(.)$ indicates the function representing the frozen layer. The block diagram in Fig. G.2 represents a generic implementation of the TL concept.

2.3 Transfer learning DPD architecture

The proposed DPD architecture used in this work is based on RVTDNN, where an arbitrary number of memory taps can be assessed [5]. The same taps configuration is employed between input and feedback signals regardless of the physics to be modeled. The proposed architecture has a fully-connected structure and the input-output relationship between the hidden layers is defined as [19–21]:

$$\mathbf{y}^{(j)} = f\left(\mathbf{W}\mathbf{x}^{(j-1)} + \mathbf{B}\right), \tag{G.4}$$

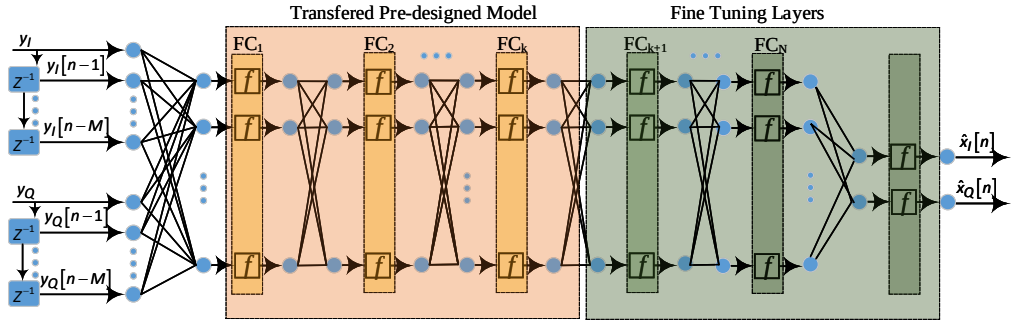


Fig. G.2: The proposed TLNN model based on RVTDNN. The transferred pre-design model is the frozen model from the previous training and is combined with the fine-tuning layers to make the new model.

where j is the j-th fully connected layer and $f(.)$ is the activation function and $\mathbf{y}^{(j)}$ is a $P \times 1$ vector representing the output values of the j-th layer, $\mathbf{W}$ is a $P \times Q$ matrix representing the trainable coefficients, $\mathbf{x}^{(j-1)}$ is a $Q \times 1$ vector representing the outputs of the previous layers and $\mathbf{B}$ is a $P \times 1$ vector representing the trainable biases. Thus, the number of outputs of the previous layer is defined by Q, and the number of inputs to the next layer is defined as P. By using the activation function, denoted as f in Fig. G.2, any arbitrary nonlinear functions can be fitted. The proposed RVTDNN architecture uses the rectified linear units (ReLU) activation function, which is less computationally expensive than hyperbolic tangent (Tanh) and Sigmoid because it involves simpler mathematical operations [22, 23]. The ReLU activation function is defined as:

$$\sigma_{\mathrm{ReLU}}(x) = \max(0, x) \tag{G.5}$$

The ReLU activation function introduces nonlinearity by setting negative inputs to 0, which also adds sparsity to the ANN and can simplify the computations. The fine-tuning layers denoted by z, where $z = N - k$, are defined as transferred layers (TL). The output of the i-th fine-tuning layers, $(TL)_i$, is written as:

$$(TL)_i = f_1\left(w_i^T \cdot (TL)_{i-1} + b_i\right), i = 1, 2, ..., z \tag{G.6}$$

where w_i^T and b_i are the weights and biases of the i-th transfer layer and the final output, Y_n' is defined as:

$$Y_n' = f_2(w_{out}^T \cdot (TL)_z + b_{out}), \tag{G.7}$$

where w_{out}^T and b_{out} denote the weights and biases of the output layer and $(TL)_z$ is the output of z-th transfer layer. $f_1(.)$ and $f_2(.)$ are the activation functions which can be chosen differently. In the presented work, both activation functions are of the ReLu type. The experimental dataset is divided into a training set and a validation set at 70%

and 30%, respectively. The weights and biases of the network are learned by choice of an appropriate loss function. The two most used loss functions for regression tasks are mean square error (MSE) loss and Huber loss. The Huber loss is a robust loss function used for a wide range of regression tasks [24] and it is used for the presented work. The Huber loss function behaves quadratic for small residuals and linearly for large residuals and is defined as [25]:

$$L_\delta(Y'_n, Y_n) = \begin{cases} \frac{1}{2}(Y'_n - Y_n)^2 & for\,|Y'_n - Y_n| \leq \delta \\ \delta\,|Y'_n - Y_n| - \frac{1}{2}\delta^2 & otherwise \end{cases}, \tag{G.8}$$

where δ, set to 1, is the parameter of Huber loss. Y'_n and Y_n denote the observation and prediction values, respectively. Through backward propagation and using the Adam optimization algorithm, the local minimum is approached. The measured data are collected and uploaded by using MATLAB. The ANN is built and trained using the Keras 2.3.0-tf package in Python.

2.4 Complexity of the proposed ANN

The complexity analysis is made with a starting point in Eq. G.4, assuming only fully connected layers with equal amounts of neurons and $P = Q$. Between each fully connected layer, there are P^2 multiplications. The number of operations between the input layer and the first hidden layer is $2MP$ multiplications, where M is the number of time delays and P is the number of neurons. There are $2P$ multiplications between the last hidden layer and the output layer. The total amount of multiplications are:

$$C_{\mathrm{m,ANN}} = C_{\mathrm{a,ANN}} = 2MP + (J-1)P^2 + 2P, \tag{G.9}$$

where the number of hidden layers is defined by J.

3 OTA Measurements Setup

The block diagram of the OTA measurements setup using a compact antenna test range (CATR) is shown in Fig. G.3 and the actual laboratory setup is in Fig. G.4. The 5G signal generator and it's arbitrary waveform generator function generates the intermediate frequency (IF) signal for transmitter input. It is centered at 3 GHz and generates an up to 100 MHz bandwidth 5G NR signal. The modulation format for the 100 MHz bandwidth is 3GPP downlink OFDM 64-QAM, sub-carrier spacing of 30 kHz and 3168 active sub-carriers. With 3 sub-carrier in each resource block (RB), it ends up to 1056 RB. The sample rate of the transmitter and receiver signals is 400 MHz which gives an oversampling rate of 4. The peak to average power ratio (PAPR) of the input signal, after capturing and loading to the generator, is 11.6 dB. A 12.5 GHz continued-wave signal of 12.5 GHz has been generated by the local oscillator (LO) generator and multiplied to

25 GHz. This LO signal is used for up-converting of the 3 GHz modulated IF signal to 28 GHz and down-converted it back to 3 GHz. For up-converting the IF signal to the 28 GHz carrier frequency and for down-converting the signal back to IF, two active mixers operating in their highly linear region are utilized. For selecting the up-converted modulated signal and suppressing the LO leakage and image frequency signals, a 28 GHz band-pass filter is used. The pre-amplifier is a high-power device operating more than 10 dB below its compression point. The output signal from the pre-amplifier is highly linear and the signal power is sufficient to drive the 4×4 APA, Amotech AAiPK428GC-A0404 [26], close to its saturated region. The APA device includes four Anokiwave AWMF-0158 transceivers [27] and integrates 16 branches of attenuators, phase shifters and PAs and 16 patch antennas in a 4×4 active phased array. A host PC is used for capturing and uploading the IQ samples. The measurement setup is power calibrated to keep all other components in their linear operating regions and the only source of nonlinearity is related to the active phased array. For controlling the main beam of the array the code-book and software tools of Amotech have been used.

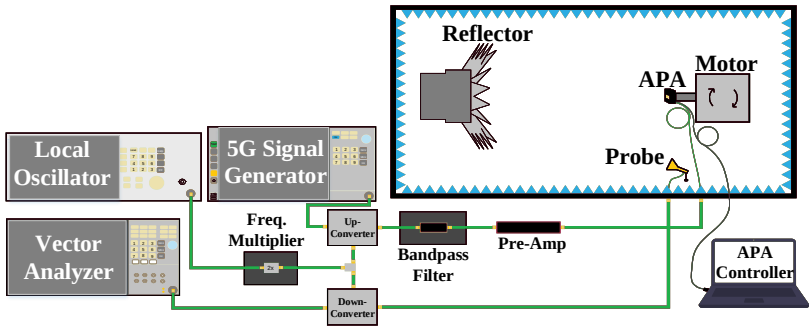


Fig. G.3: block diagram of measurement setup in compact antenna test range chamber.

Fig. G.4: Measurement setup using compact antenna test range chamber.

4 ANN Optimization Results

The ANN optimization methodology presented in [28] was used in this paper. The methodology is applied to an ANN model trained using two signal bandwidth values, namely 20 MHz and 100 MHz. When moving from the classical ANN model to the proposed TLNN approach, part of the optimized 20 MHz model will be frozen and used as the pre-design model for TLNN. The results from ANN optimization of 100 MHz bandwidth are used as the benchmark to compare with the results obtained using TLNN. This chapter includes the ANN optimization procedure and verification results carried out for an RF signal bandwidth of 20 MHz. The target of the ANN optimization is to minimize the number of time-delays and the number of neurons while the desired levels of linearization in terms of the adjacent channel leakage ratio (ACLR) and the error vector magnitude (EVM) are maintained. 100 k I/Q samples of the input and output signals are captured, time-aligned and used to train several ANN predistorters. There are four fully connected hidden layers in the model based on the results obtained in [28] where a number above four couldn't improve the linearization performance anymore. The time-delays parameter is swept from 3 to 20 and the neurons are swept from 40 to 480. The optimization results are assessed by constructing the network to use 70 % of the I/Q data for training and 30 % for validation. Fig. G.5 shows ANN parameter optimization results of linearization of the narrow-bandwidth signal, where the optimal choice is a trade-off between the ACLR, the EVM and the number of multiplications. By keeping the number of time delays to 4 and the number of neurons to 256, it is possible to achieve an ACLR improvement of 13.1 dB, as shown in Fig. G.5a, and EVM improvement of 8.8 % points, Fig. G.5b, while keeping the number of multiplications as low as possible, i.e. app. 199 k, Fig. G.5c. Increasing the number of neurons to higher than 256 will lead to ACLR incremental improvements below 0.4 dB and EVM incremental improvements below 0.2 % points, which we consider negligible for the sake of our optimization procedure as shown in Fig. G.5a-b. There is a clear indication from Fig. G.5c that in a dense network, with several hidden layers, the number of multiplications will increase drastically by the number of neurons. This is in agreement with Equation (G.9) where the number of multiplications increases approximately as the square of the number of neurons when the number of the hidden layers exceed one. So it is important to keep as low as possible the number of neurons for a dense network with several layers, for achieving a lower training time and computational complexity. The spectral power density (PSD) result is in Fig. G.6a which shows the achieved out-of-band improvement obtained by deploying the proposed optimized ANN-based DPD. Fig. G.6b and Fig. G.6c show the in-band AM/AM and AM/PM gain distortions which are related to EVM. These results are perfectly aligned with the expected performance based on based on the proposed optimization procedure, whose results are summarized in Fig. G.5.

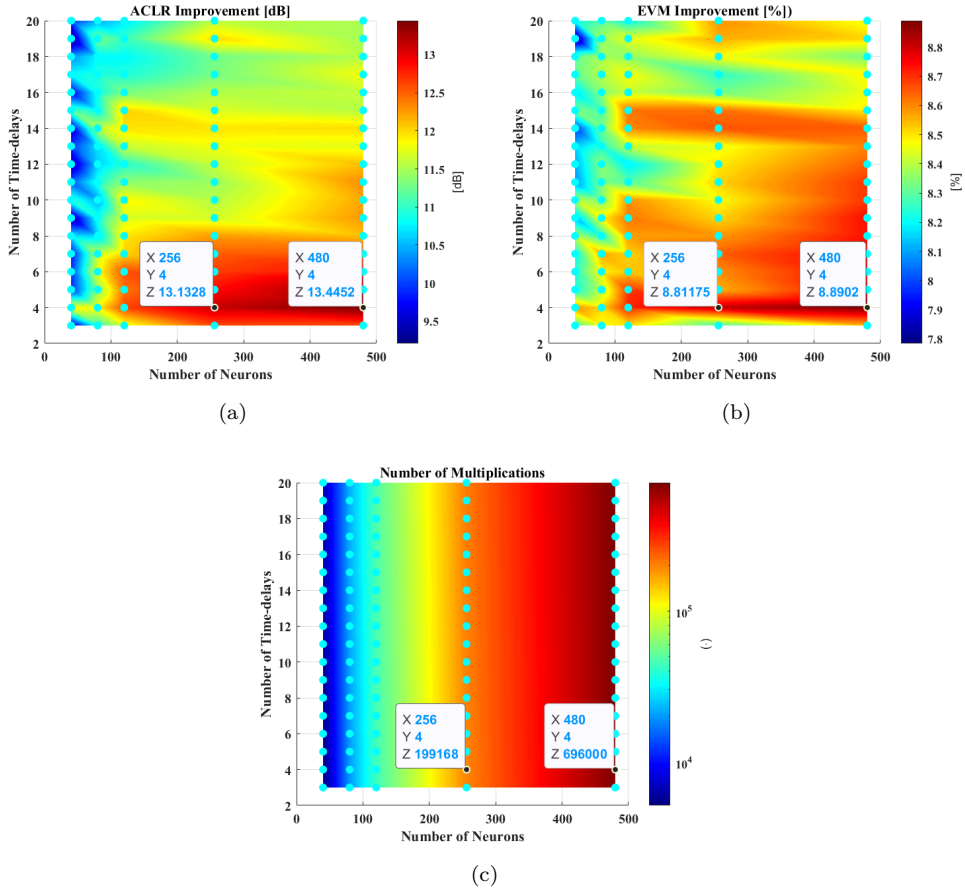


Fig. G.5: ANN parameter optimization results of linearization of the narrow bandwidth signal. (a): The ACLR (average left/right levels) improvements [dB], (b): The EVM improvements [%], (c): The number of required multiplications for each case.

5 Transfer learning implementation

For implementing the transfer learning algorithm, the part of the model of 20 MHz bandwidth is copied and used as a transferred pre-design model for linearization of the 100 MHz bandwidth signal. This is done by freezing three hidden layers from the trained model of 20 MHz bandwidth. The frozen layers are then combined with the fine-tuning layers to build the model for 100 MHz bandwidth. The implemented architecture of

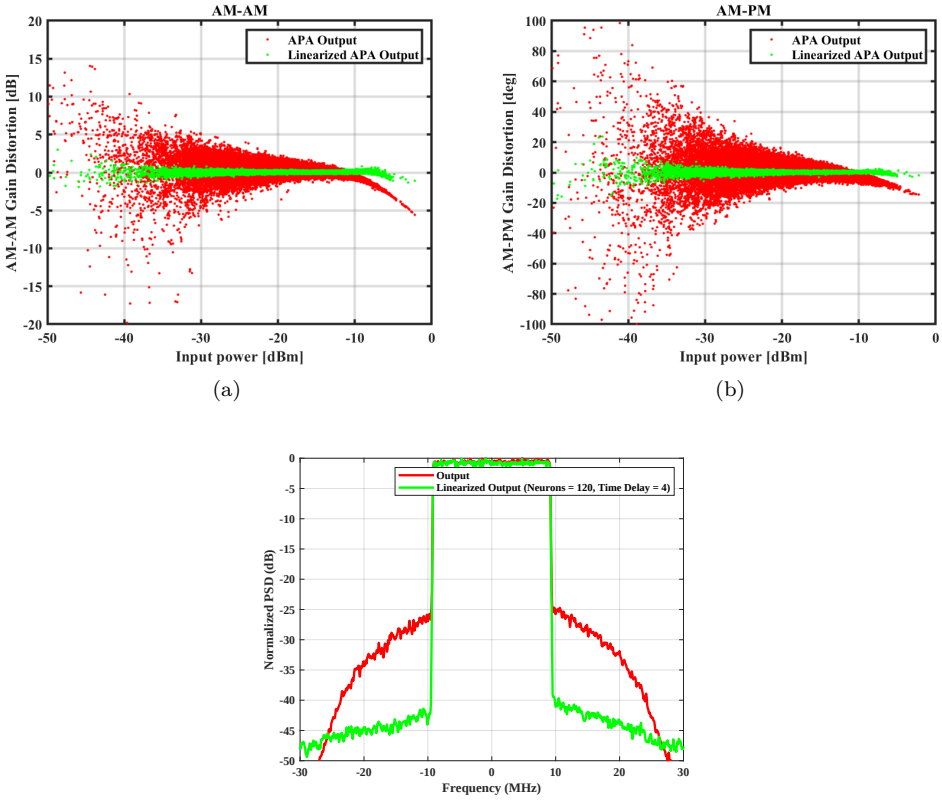


Fig. G.6: ANN-based results for 20 MHz BW using 256 neurons and 4 time delays. (a): The AM/AM gain distortion, (b): The AM/PM gain distortion, (c): The power spectral density.

the TLNN is in shown Fig. G.7. Table G.1 summarizes the implementation procedure used for the proposed method. Table G.2 shows network configuration parameters for regular ANN and TLNN. By using the transfer learning approach, the number of hidden layers is reduced from four to one and the number of neurons is reduced from 256 to 16. Furthermore, the model from one bandwidth is transferred to another bandwidth which means the transferred pre-designed model already includes most of the knowledge of the nonlinear behavioral model of APA.

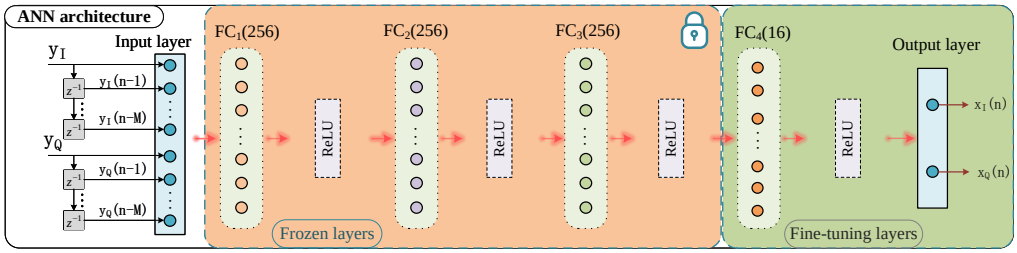


Fig. G.7: The implemented architecture of the TLNN. The transferred pre-design model is the frozen model from the previous training and is combined with the fine-tuning layers to make the new model.

Table G.1: Algorithm used for TLNN training.

Algorithm 1 Training of Regular ANN
i: Generate n samples of IQ data of $x[n]$ and $y[n]$ with 20 MHz BW
ii: Update weights and biases given by Eq. (G.4) using 70 % of n samples
iii: Continue updating until the minimum cost level is reached
iv: Validate the model using 30 % of n samples
v: If the cost function of validation is ok, then freeze the model
vi: Save the first k layers of the ANN as pre-designed model
vii: Repeat steps i-vi by using 100 MHz BW and save it as reference model
Algorithm 2 Training of TLNN
i: Generate l samples of IQ data of $x[l]$ and $y[l]$ of 100 MHz BW
ii: combine the pre-design model with the fine-tuning model
iii: Update weights and biases by using 70 % of l samples
iv: Continue updating until the minimum cost level is reached
v: Validate the model using 30 % of l samples
vi: If the cost function of validation is ok, exit the loop
vii: Update network coefficients

Table G.2: Network configuration parameters for Regular ANN and TLNN.

Model	Regular ANN	**TLNN**
Maximum Epochs	500	200
Minimum Batch Size	500	200
Optimizer	Adam	Adam
Loss Function	Huber loss	Huber loss
Activation Function	Relu	Relu
Initial Learning Rate	0.01	0.01
Early Stop Patience	20 epochs	20 epochs
Minimum Learning Rate	10^{-6}	10^{-6}
Number of delay lines	4	4
Number of hidden layers	4	**1**
Number of Neurons	256 (each hidden layer)	**16**

6 Bandwidth-Scalable Predistortion Results

First, the model for the reference 100 MHz bandwidth based on regular ANN, has been optimized using the same procedure as described in section 4. This model is

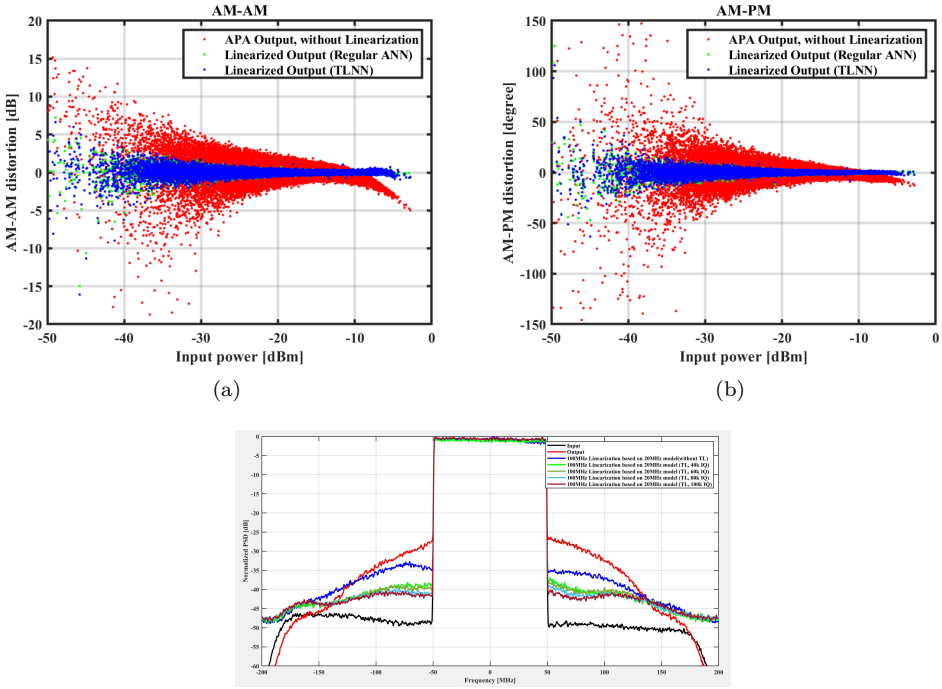


Fig. G.8: Regular ANN vs. TL-ANN for 100 MHz BW. (a): The AM/AM gain distortion, (b): The AM/PM gain distortion, (c) Power spectral density (e.g. TLNN-H1N16, means TLNN with 1 hidden layer and 16 neurons).

Table G.3: Performance comparison between regular ANN and the proposed TLNN for 100 MHz bandwidth signal.
*) ACLR is based on the average of the left and the right sides.

	Number of Multiplications	**EVM (without/with DPD), (Improve.)**	**ACLR *) (without/with DPD), (Improve.)**
Regular ANN (**256 Neurons, 4 hidden layers**)	199168	(10.3/1.6 %), (8.7 %)	(34.7/43.7 dBc), (9 dB)
TLNN (**128 Neurons, 1 hidden layer**)	1280	(10.3/1.8 %), (8.5 %)	(34.7/43.2 dBc), (8.5 dB)
TLNN (**64 Neurons, 1 hidden layer**)	640	(10.3/1.8 %), (8.5 %)	(34.7/43 dBc), (8.3 dB)
TLNN (**32 Neurons, 1 hidden layer**)	320	(10.3/1.7 %), (8.6 %)	(34.7/43.1 dBc), (8.4 dB)
TLNN (**16 Neurons, 1 hidden layer**)	160	(10.3/1.7 %), (8.6 %)	(34.7/43.7 dBc), (8.5 dB)

constructed by using four hidden layers with 256 neurons in each. Linearization results

of this approach are used for bench-marking of the TLNN-based linearization of the 100 MHz bandwidth. For TLNN, the frozen pre-defined model from 20 MHz training and one fully connected fine-tuning hidden layer are used. This model has been verified with four different sets of neurons, 128, 64, 32 and 16 in the fine-tuning layer. The results for each set of neurons are bench-marked with the regular ANN which has four fully connected layers and 256 neurons in each. The structures of the input and output layers of the networks are the same for both regular and TLNN. The number of multiplications based on Equation (G.9) for regular ANN and TLNN are given as:

$$\begin{aligned} C_{\mathrm{m,ANN}} &= 2 * 4 * 256 + (4-1) * 256^2 + 2 * 256 \\ &= 199168, \end{aligned} \tag{G.10}$$

and for TLNN with 1 hidden layer and 16 neurons it will results to:

$$\begin{aligned} C_{\mathrm{m,TNN}} &= 2 * 4 * 16 + 2 * 16 \\ &= 160 \end{aligned} \tag{G.11}$$

6.1 Discussion

A comparison of the verification results in terms of AM/AM and AM/PM distortion gains and power spectral density (PSD) are illustrated in Fig. G.8. The TLNN linearization can provide the same level of linearity as the regular ANN. Detailed performance comparisons between regular ANN and the proposed TL ANN are in Table G.3. These results show that it is possible to achieve the same linearization performance compared to regular ANN, i.e. an EVM improvement of 8.6 % points and ACLR improvement of 9 dB, by using TLNN. Hence the proposed approach proves to be robust versus signal bandwidth and can be used as a bandwidth-scalable linearization technique. On the other hand, TLNN allows reducing the number of the hidden layers (through re-using the frozen model) and the number of neurons which results in relaxation of the computational complexity in terms of the number of multiplications. The outcomes of the performed linearization experiments can be summarized as follows:

1. By using the SoA conventional RVTDNN approach for linearization of the actual APA, we need an ANN DPD of 256 neurons, 4 hidden layers for 20 MHz signal linearization, and another ANN DPD of 256 neurons, 4 hidden layers multiplications for 100 MHz signal linearization.

2. TLNN approach, instead can reuse the model calculated for 20 MHz and need only additional 16 neurons and one layer for 100 MHz signal linearization.

3. For an adaptive DPD, the time to calculate the incremental layers in TLNN is reduced and grants the models the ability to be adaptively re-identified.

4. A clear advantage delivered by the proposed is in terms of the LUT (look-Up Table) size necessary to implement the DPD. Instead of storing two completely different sets of ANN DPD parameters (SoA approach), one for the narrow bandwidth use case and the other for the wide bandwidth use case, system engineers will need to store much fewer parameters for linearizing the wide bandwidth use case, because they can reuse most of the ones calculated for the narrow bandwidth.

7 Conclusion

This paper presented a bandwidth-scalable over-the-air digital predistortion (DPD) of an active phased array transmitter based on a transfer learning neural network (TLNN) method. The proposed methodology allows reducing the hardware implementation complexity in terms of the number of multiplications while ensuring the same linearization performance as a regular ANN. In the proposed method, part of the model is fixed as a pre-designed model, and then an incremental model component was trained and deployed for fine-tuning the remaining adaptation layers to build the final model. This paper demonstrated how such TL technique could be used to implement a bandwidth scalable digital predistorter. The ANN layers identified for one signal bandwidth were reused and enhanced with an incremental neuron layer to allow the ANN predistorter to successfully linearize input signals with wider bandwidths. The proposed linearization technique was validated with measurements on a state-of-the-art 4×4 APA and a setup using up- and down-conversion from sub-6 GHz to 28 GHz for verification. Experimental results showed that our optimized ANN-based DPD could linearize a 20 MHz 5G signal with an EVM improvement of 8.8 % points and an ACLR improvement of 13.3 dB. It was also demonstrated, that by using TL, the same ANN DPD can be reused to linearize a 5G signal with a much wider bandwidth, namely 100 MHz. To do so, only an additional layer of 16 neurons was added on top of the reused ANN DPD. Such an approach allowed us to obtain an EVM improvement of 8.6 % points and an ACLR improvement of 8.5 dB. The proposed approach required only 160 additional multiplications to linearize the new signal bandwidth, compared to the 199168 multiplications necessary if one would use another fully trained ANN predistorter. The reduced complexity allows to bring down the cost of the implementation using digital hardware. Further research is being conducted to make the proposed bandwidth-scalable DPD fully robust concerning the signal bandwidth and other transmitter operating conditions. Our future goal is to enhance the TL methodology to obtain a universal set of parameter that can be fully reused to linearize multiple signal bandwidths. Such a result would allow lowering further the complexity and cost of the DPD implementation on digital hardware.

References

[1] Estefanía Guillena, Wantao Li, Gabriel Montoro, Roberto Quaglia, and Pere L Gilabert. Reconfigurable dpd based on anns for wideband load modulated balanced amplifiers under dynamic operation from 1.8 to 2.4 ghz. *IEEE Transactions on Microwave Theory and Techniques*, 70(1):453–465, 2021.

[2] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[3] Taijun Liu, Slim Boumaiza, and Fadhel M Ghannouchi. Dynamic behavioral modeling of 3g power amplifiers using real-valued time-delay neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 52(3):1025–1033, 2004.

[4] Pere L Gilabert, David López-Bueno, Thi Quynh Anh Pham, and Gabriel Montoro. Machine learning for digital front-end: a comprehensive overview. *Machine Learning for Future Wireless Communications*, pages 327–381, 2020.

[5] Meenakshi Rawat, Karun Rawat, and Fadhel M Ghannouchi. Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 58(1):95–104, 2009.

[6] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[7] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

[8] Philip Smith and Cuixian Chen. Transfer learning with deep cnns for gender recognition and age estimation. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2564–2571. IEEE, 2018.

[9] Yuwen Yang, Feifei Gao, Zhimeng Zhong, Bo Ai, and Ahmed Alkhateeb. Deep transfer learning-based downlink channel prediction for fdd massive mimo systems. *IEEE Transactions on Communications*, 68(12):7485–7497, 2020.

[10] Wesin Alves, Ilan Correa, Nuria González-Prelcic, and Aldebaro Klautau. Deep transfer learning for site-specific channel estimation in low-resolution mmwave mimo. *IEEE Wireless Communications Letters*, 10(7):1424–1428, 2021.

[11] Eric Ng, Yehia Beltagy, Giovanni Scarlato, Ahmed Ben Ayed, Patrick Mitran, and Slim Boumaiza. Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets. *IEEE Transactions on Microwave Theory and Techniques*, 67(11):4479–4492, 2019.

[12] Christian Fager, Thomas Eriksson, Filipe Barradas, Katharina Hausmair, Telmo Cunha, and Jose Carlos Pedro. Linearity and efficiency in 5g transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5g active antenna transmitter context. *IEEE Microwave Magazine*, 20(5):35–49, 2019.

[13] Xin Liu, Qian Zhang, Wenhua Chen, Haigang Feng, Long Chen, Fadhel M Ghannouchi, and Zhenghe Feng. Beam-oriented digital predistortion for 5g massive mimo hybrid beamforming transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 66(7):3419–3432, 2018.

[14] Feridoon Jalili, Martin H Nielsen, Ming Shen, Ole K Jensen, Jan H Mikkelsen, and Gert F Pedersen. Linearization of active transmitter arrays in presence of antenna crosstalk for 5g systems. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, pages 1–5. IEEE, 2019.

[15] Katharina Hausmair, Ulf Gustavsson, Christian Fager, and Thomas Eriksson. Modeling and linearization of multi-antenna transmitters using over-the-air measurements. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE, 2018.

[16] Alberto Brihuega, Mahmoud Abdelaziz, Lauri Anttila, Matias Turunen, Markus Allén, Thomas Eriksson, and Mikko Valkama. Piecewise digital predistortion for mmwave active antenna arrays: Algorithms and measurements. *IEEE Transactions on Microwave Theory and Techniques*, 68(9):4000–4017, 2020.

[17] Feridoon Jalili, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Ming Shen, and Gert F Pedersen. Linearization trade-offs in a 5g mmwave active phased array ota setup. *IEEE Access*, 8:110669–110677, 2020.

[18] Chao Yu, Jianxin Jing, Han Shao, Zhi Hao Jiang, Pinpin Yan, Xiao-Wei Zhu, Wei Hong, and Anding Zhu. Full-angle digital predistortion of 5G millimeter-wave massive mimo transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 67(7):2847–2860, 2019.

[19] Timothy O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

[20] Yufeng Zhang, Zhugang Wang, Yonghui Huang, Wei Wei, Gert Frølund Pedersen, and Ming Shen. A digital signal recovery technique using dnns for leo satellite communication systems. *IEEE Transactions on Industrial Electronics*, 68(7):6141–6151, 2021.

[21] Jakob Gjedsted Brask, Kasper Bruun Olesen, Lauge Føns Dyring, Arun Yadav, Feridoon Jalili, Yufeng Zhang, and Ming Shen. Deep digital signal recovery for leo satellite communication in presence of system perturbations. In *2021 IEEE MTT-S International Wireless Symposium (IWS)*, pages 1–3. IEEE, 2021.

[22] Xavier Glorot, Antoine Bordes, and Y. Bengio. Deep sparse rectifier neural networks. *Journal of Machine Learning Research*, 15, 2010.

[23] Changliu Liu, Tomer Arnon, Christopher Lazarus, Christopher Strong, Clark Barrett, and Mykel J Kochenderfer. Algorithms for verifying deep neural networks. *arXiv preprint arXiv:1903.06758*, 2019.

[24] Gregory P Meyer. An alternative probabilistic interpretation of the huber loss. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 5261–5269, 2021.

[25] Peter J Huber. Robust statistics. In *International encyclopedia of statistical science*, pages 1248–1251. Springer, 2011.

[26] Amotech. AAiP28G A0808 EVB version R01 190321c. `https://www.mrc-gigacomp.com/Amotech.php`, 2019.

[27] Anokiwave. AWMF-0158 28 GHz Silicon 5G Tx/Rx Quad Core IC. `https://www.anokiwave.com/products/awmf-0158/index.html`, 2019.

[28] Feridoon Jalili, Yufeng Zhang, Felice Francesco Tafuri, Ole Kiel Jensen, Yunfeng Li, Qingyue Chen, Ming Shen, and Gert F Pedersen. Tuning of deep neural networks for over-the-air linearization of highly nonlinear wide-band active phased arrays. In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, pages 1–4. IEEE, 2021.

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-883-0

AALBORG UNIVERSITY PRESS