FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING
DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

MASTER'S THESIS

AUGMENTED-LSTM AND 1D-CNN-LSTM BASED DPD MODELS FOR LINEARIZATION OF WIDEBAND POWER AMPLIFIERS

Author	Rathnayake Anusha Nadeeshan
Supervisor	Prof. Nandana Rajatheva
Second Examiner	Dr. Pekka Pirinen
Technical Advisor	Hossein Rezaei

April 2023

Anusha Nadeeshan R. (2023) Augmented-LSTM and 1D-CNN-LSTM based DPD Models for Linearization of Wideband Power Amplifiers. University of Oulu, Faculty of Information Technology and Electrical Engineering, Degree Programme in Wireless Communications Engineering, 41 p.

ABSTRACT

Artificial Neural Networks (ANNs) have gained popularity in modeling the nonlinear behavior of wideband power amplifiers. Recently, modern researchers have used two types of neural network architectures, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), to model power amplifier behavior and compensate for power amplifier distortion. Each architecture has its own advantages and limitations. In light of these, this study proposes two digital pre-distortion (DPD) models based on LSTM and CNN.

The first proposed model is an augmented LSTM model, which effectively reduces distortion in wideband power amplifiers. The measurement results demonstrate that the proposed augmented LSTM model provides better linearization performance than existing state-of-the-art DPDs designed using ANNs.

The second proposed model is a 1D-CNN-LSTM model that simplifies the augmented LSTM model by integrating a CNN layer before the LSTM layer. This integration reduces the number of input features to the LSTM layer, resulting in a low-complexity linearization for wideband PAs. The measurement results show that the 1D-CNN-LSTM model provides comparable results to the augmented LSTM model.

In summary, this study proposes two novel DPD models based on LSTM and CNN, which effectively reduce distortion and provide low-complexity linearization for wideband PAs. The measurement results demonstrate that both models offer comparable performance to existing state-of-the-art DPDs designed using ANNs.

Keywords: Augmented-LSTM, digital predistortion, DPD, 1D-CNN, LSTM, power amplifier, PA, artificial neural network.

TABLE OF CONTENTS

FOREWORD

This master's thesis investigates two neural network-based digital predistortion models, namely augmented-LSTM and 1D-CNN-LSTM. The aim of this research was to explore new models for the linearization of wideband power amplifiers, and I am honored to have been able to conduct this study at the Centre for Wireless Communications (CWC) of the University of Oulu, Finland. I would like to express my deepest gratitude to my thesis supevior, Prof. Premanandana Rajatheva and my technical advisor, Hossein Rezaei for their invaluable guidance and support throughout the course of my research. Their expertise and encouragement have been invaluable, and I feel extremely fortunate to have had the opportunity to work under their mentorship. My sincere thanks also go to my peer researcher, Mr. Lakshan, whose assistance has been invaluable in the course of this research. I am also immensely grateful to all the professors at the University of Oulu for their valuable assistance and support, which has helped me to broaden my horizons and achieve academic excellence. And finally, I would like to express my heartfelt appreciation to my family and friends for their unwavering love and support during my master's degree. Without their encouragement and motivation, this achievement would not have been possible. Once again, thank you all for your invaluable contributions and support.

Oulu, 28th April, 2023

Rathnayake Anusha Nadeeshan

LIST OF ABBREVIATIONS AND SYMBOLS

Acronyms

1D	One Dimensional
2D	One Dimensional
5G	Fifth Generation
ACLR	Adjacent Channel Leakage Ratio
ACPR	Adjacent Channel Power Ratio
ADC	Analog to Digital Converter
AM/AM	Amplitude Modulation to Amplitude Modulation
AM/PM	Amplitude Modulation to Phase Modulation
ANN	Artificial Neural Network
ARVTDNN	Augmented Real-Valued Time Delay Neural Network
CF	Crest Factor
CFR	Crest Factor Reduction
CNN	Convolutional Neural Network
DC	Direct Current
DNN	Deep Neural Network
DPD	Digital Predistortion
EVM	Error Vector Magnitude
FC	Fully Connected
FIR	Finite Impulse Response
GHz	Giga Hertz
GMP	Generalized Memory Polynomial
I	In-phase
IBO	Input Back-Off
ILA	Indirect Learning Architecture
IMD	Inter Modulation Distortion
IQ	In-phase and Quadrature
LDMOS	Laterally Diffused Metal Oxide Semiconductor
LMS	Least Mean Square
LS	Least Square
LSTM	Long-Short Term Memory
MHz	Mega Hertz
MP	Memory Polynomial
MSE	Mean Square Error
NMSE	Normalized Mean Square Error
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAPR	Peak to Average Power Ratio
PSD	Power Spectral Density
Q	Quadrature
QAM	Quadrature Amplitude Modulation
ReLU	Rectified Linear Unit

RF	Radio Frequency
RLS	Recursive Least Square
RMS	Root Mean Square
RNN	Recurrent Neural Network
RVFTDNN	Real-Valued Focused Time-Delay Neural Network
RVTDCNN	Real-Valued Time Delay Convolutional Neural Network
RVTDNN	Real-Valued Time Delay Neural Network
SNL	Static Non Linearity
SV	Simplified Volterra

Symbols

Ψ	Phase of the signal
ω_0	Carrier frequency
ϕ_0	Initial phase
h_p	Parameters of the Volterra model
dB	Decibel
dBc	Decibels relative to carrier
K	Nonlinearity order
M	Memory depth
$a_{k,m}$	Kernels of the memory polynomial model
Y_{RF}	Output RF signal
x_{RF}	Input RF signal
X_{in}	Input to the NN model
h_k	Convolution kernel
g_k	Convolution output

Operators

$\lvert \cdot \rvert$	Absolute value
$\circledast$	Convolution operator
$\prod$	Product operator
$\sum$	Sum operator

1 INTRODUCTION

Power amplifiers (PAs) are an indispensables component of wireless communication systems as they provide the necessary power to transmit signals [1, 2]. However, due to their nonlinear characteristics and memory effects [3–5], PAs can cause spectrum regrowth. This phenomenon increases interference with adjacent channels and degrades the quality of the transmitted signal. This issue is particularly concerning for 5G wireless communication systems as they have increased bandwidth and signal complexity [2]. Therefore, there is always a compromise between linearity and efficiency of the PAs, which means that PAs cannot be power efficient and linear simultaneously.

To overcome the issue of spectrum regrowth, a technique known as digital predistortion (DPD) [2, 6] can be employed. DPD is a digital signal processing technique that involves predicting the distortion introduced by the amplifier and generating a pre-distorted version of the input signal that cancels out the distortion during amplification. As a result, the cascaded DPD-PA system becomes ideally linear, ensuring that the quality of the transmitted signal is not degraded.

Traditional DPD models, such as the Volterra series [7] model and its simplified versions, including the memory polynomial (MP) model [8], the generalized memory polynomial (GMP) model [9], and the simplified Volterra (SV) model [10], have been shown to be effective in addressing the nonlinearities of PAs for low-bandwidth signals. However, these models become less effective as the signal bandwidth increases due to the high correlation between polynomial bases [2].

Another popular approach used by modern researchers to model the nonlinearities of power amplifiers (PAs) is the utilization of artificial neural networks (ANNs). In reference [6], the authors developed a shallow Neural Network (NN)-based predistorter to compensate for PA nonlinearity, and subsequently, a deep NN-based DPD was proposed in [11]. Over time, many ANN-based DPD models have been suggested, including long short-term memory (LSTM) [12, 13] networks, recurrent neural networks (RNNs) [14], and convolutional neural networks (CNNs) [1, 2, 15]. However, models such as DNNs, RNNs, and LSTM networks have the drawback of high complexity. While CNN models have been proposed as a solution to this complexity problem, they have less capability to capture memory features [13].

To overcome these limitations, a novel augmented-LSTM model has been proposed to model the predistorter of wideband PAs with higher accuracy and lower complexity than other DNN models. The proposed model utilizes the LSTM layer's capability to leverage time series information of the PA data, as demonstrated in prior works [12, 13, 16, 17]. The key difference between the LSTM-DNN model proposed in [12] and the proposed augmented-LSTM model lies in the input features used to model the inverse model of the PA. Additionally, to simplify the proposed augmented-LSTM model, a 1D-CNN and LSTM-based model has been introduced in the latter part of the study. This technique employs a 1D-CNN layer before the LSTM layer to capture essential features from the same time stamp, reducing the number of input features to the LSTM layer. Then, the LSTM layer leverages the time-series information. This approach aims to combine the strengths of both techniques to develop an efficient model for wideband PA predistortion with low complexity.

1.1 Thesis Contribution

The objective of the study was to design a digital pre-distorter (DPD) using artificial neural network to compensate for the nonlinearities of a wideband power amplifier and achieve a better Adjacent Channel Power Ratio (ACPR)/ Adjacent Channel Leakage Ratio (ACLR). Two DPD models are proposed in this study, and the augmented-LSTM model utilized a LSTM layer to model the non-linearities of the power amplifier. The ability of the LSTM layer to handle sequential data is utilized to capture the time series information of input data in this method.

To simplify the proposed augmented-LSTM model, a 1D-CNN-LSTM model is proposed. In this model, a 1D-CNN layer was utilized before the LSTM layer to capture the useful features in the same memory effect and then the LSTM layer is used to exploit time series information.

1.2 Thesis Outline

The rest of this thesis is organized as follows:

- **Chapter 2:** Chapter 2 includes explanations of concepts, background theories, and a summary of prior research.
- **Chapter 3:** This chapter provides a detailed explanation of the proposed models and their structures, as well as discusses how they can be extended to include DPD design.
- **Chapter 4:** Chapter 4 discusses the training, evaluation and the measurement results of the proposed models.
- **Chapter 5:** In this chapter, the results are presented in summary, and potential areas for future research are discussed.

2 BACKGROUND

This chapter discusses all theoretical concepts related to this study. Non-linear behaviour of the power amplifier still remains a challenge in modern wireless systems. Therefore, to develop a proper technique to mitigate the non-linear behaviour is a major requirement. To achieve high system performance the behavior of the PA should be carefully analyzed and compensated.

2.1 Power Amplifiers

The power amplifier (PA) is a crucial component in wireless transmitters as it amplifies the input signal's power to ensure it reaches the receiver with sufficient strength. The main function of the PA is to convert Direct Current (DC) power into Radio Frequency (RF) power, as illustrated in Fig. 2.1.

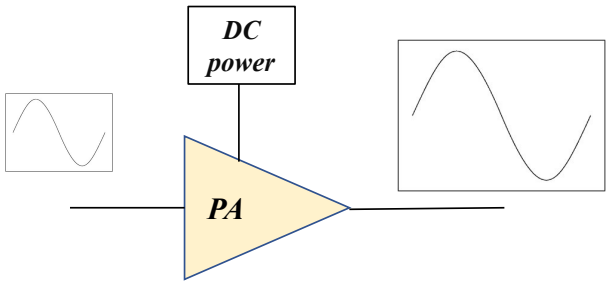

Figure 2.1. Power Amplifier operation.

For a given a constant DC power supply, the PA's output power should increase proportionally to the input power. Hence, this process should be linear and denoted as

$$Y_{RF} = g * x_{RF}(t). \tag{1}$$

However, in the real world, the PA's output power saturates after reaching a certain level, causing the PA to exhibit nonlinear behavior as shown in Fig. 2.2. Consequently, the PA's output becomes a nonlinear function of its input shown as

$$Y_{RF} = g_{nl} * x_{RF}(t). \tag{2}$$

To characterize the nonlinearity of a power amplifier, two primary types of nonlinear distortion can be employed based on the AM/AM and AM/PM characteristics [18]. The input to the PA can be described as;

$$x_{RF}(t) = a(t)e^{j\varphi(t)}. \tag{3}$$

where, $a(t)$ is the envelope and $\varphi(t)$ is the phase of the input signal, $j = \sqrt{-1}$.

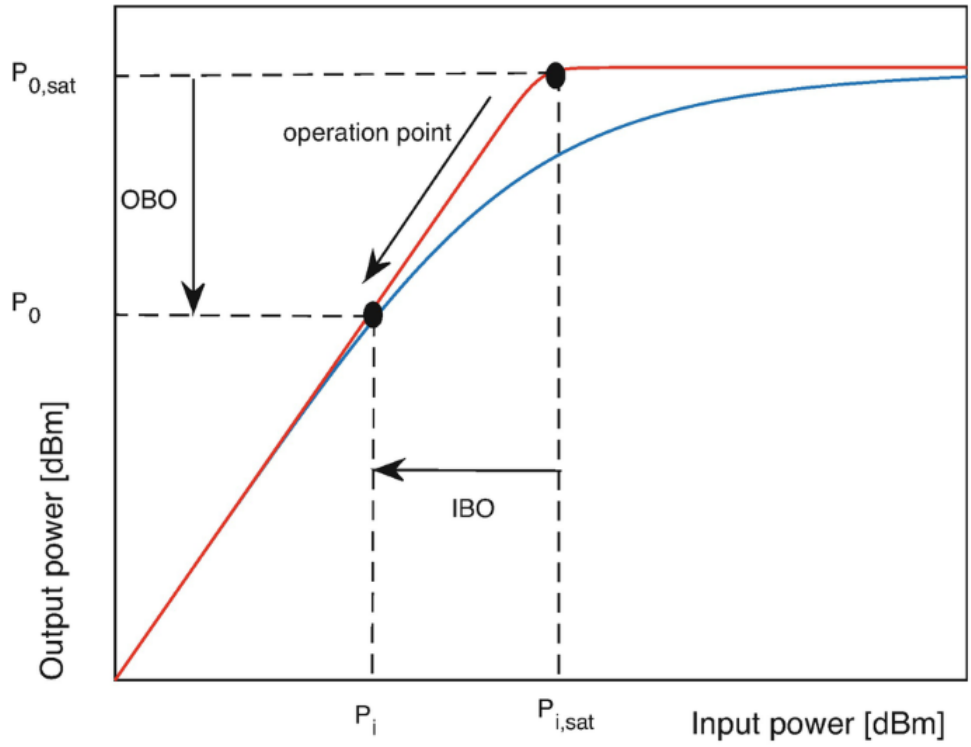

Figure 2.2. Power Amplifier input-output characteristics.

The nonlinear output of the PA, which is the distorted $x_{RF}(t)$ is given by

$$Y_{RF}(t) = g_{nl}[a(t)]e^{j\varphi(t)+f_{nl1}(a(t))}, \tag{4}$$

where, g_{nl} and f_{nl} functions represent the nonlinear behavior of PA in terms AM/AM and AM/PM distortions.

Fig. 2.3 illustrates the PA's AM/AM characteristics of the PA data used in this study, which demonstrates its nonlinearity. Additionally, Fig. 2.4 and Fig. 2.5 display the PA's gain vs. input power and phase vs. input power, respectively, providing a better understanding of the PA's nonlinearity.

In order to prevent nonlinear behavior, it is necessary to operate the power amplifier (PA) in its linear mode. However, as wireless systems continue to evolve and provide higher data rates and support more users, compact constellations and multiple access techniques based on orthogonal frequency division multiplexing (OFDM) are being used [19]. When operating a PA in the linear region with such signals, which have a high peak-to-average-power ratio (PAPR), the power efficiency is reduced. Therefore, it is not possible for PAs to be both power efficient and linear at the same time.

2.1.1 Static nonlinearity and Dynamic nonlinearity

Current behavioral models and digital predistorters are designed to account for memory effects, in addition to static distortions, in order to accurately model the nonlinearity of power amplifiers (PAs) [20]. Static distortion is primarily caused by the nonlinear behavior of the PA transistor and is the main source of distortion, resulting in spectral

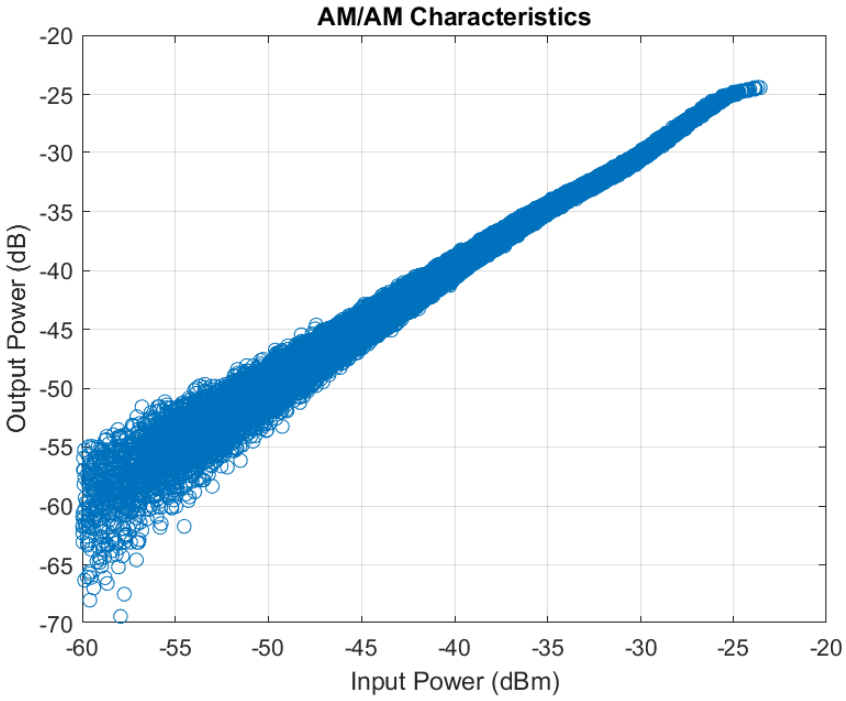

Figure 2.3. Power Amplifier AM/AM characteristics.

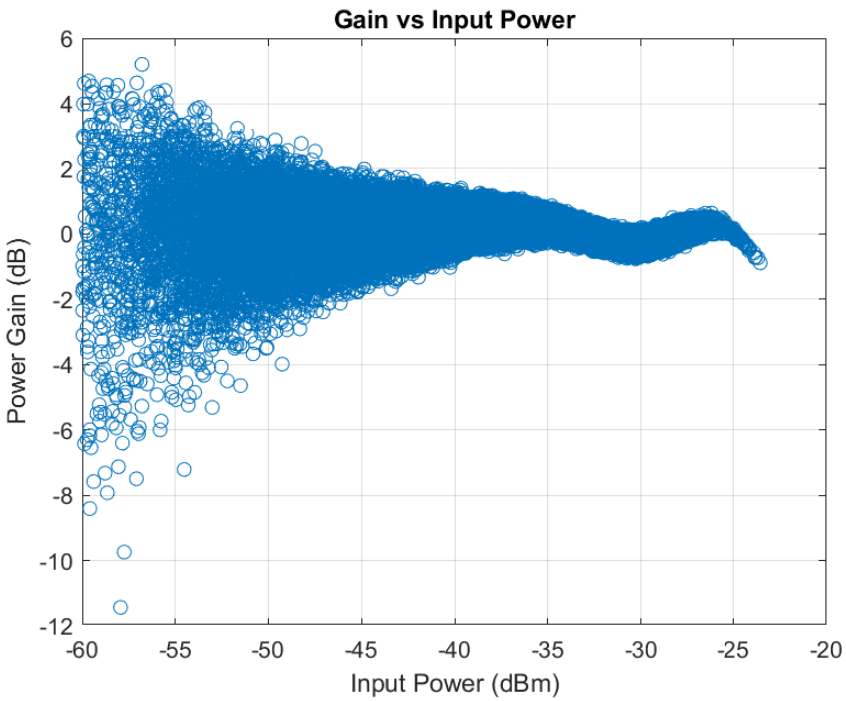

Figure 2.4. Power Amplifier gain vs. input power.

regrowth. Dynamic distortions, which refer to deviations from static characteristics over time, mainly arise from long-term and short-term memory effects. Long-term memory effects are primarily due to temperature variations and the bias network of the PA, while short-term memory effects are due to the frequency response of the input and output

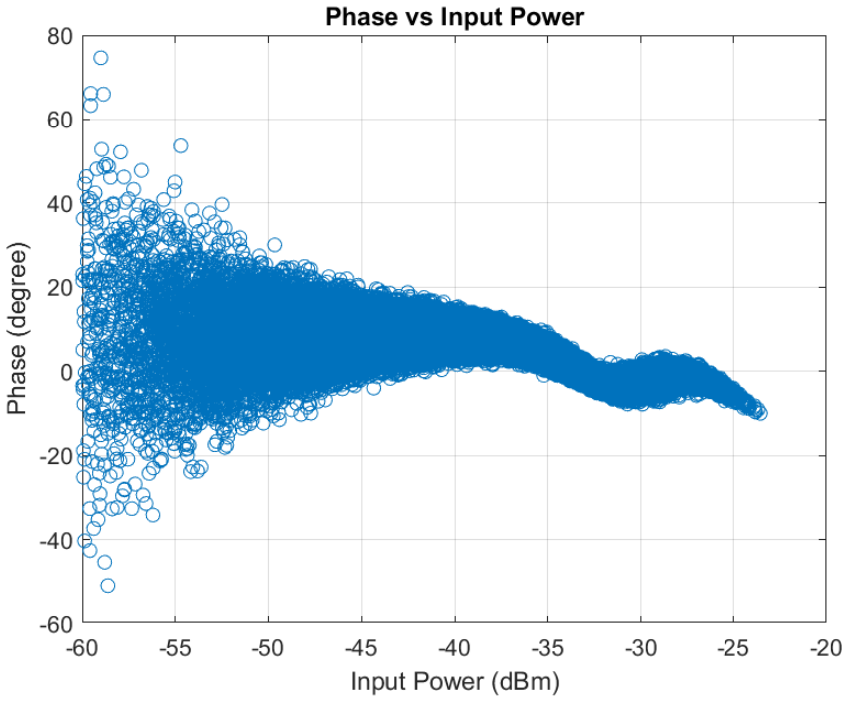

Figure 2.5. Power Amplifier phase vs. input power.

impedance matching networks of the PA. These memory effects can be seen as frequency selective spectrum. At low output power levels, the memory effect caused by dynamic distortions is more pronounced compared to static nonlinearity. However, at high input power levels, the impact of dynamic distortions on the overall behavior of the power amplifier is much less significant than that of static distortions [21].

2.1.2 Power amplifier distortions

Harmonic distortions

Harmonic distortion arises when a power amplifier is driven by a sinusoidal signal, due to nonlinearity in its transfer function. This causes unwanted harmonics in the output signal that are multiples of the fundamental frequency of the input signal. Typically, the fundamental frequency is much higher than the bandwidth of the signal, so harmonic distortion is not usually critical since the harmonics are filtered out by the antenna circuitry.

Inter-modulation distortions

Inter-modulation distortion occurs when a power amplifier is fed with a signal containing two or more frequencies, which creates unwanted frequencies in the output signal. This happens because the input frequencies interact with each other, leading to the production of new frequencies. For instance, if we have two input frequencies, f_1 and f_2, the output can contain frequencies like $f_1 - f_2$ and $f_2 - f_1$ (known as second order IMD products) or $2f_1 - f_2$ and $2f_2 - f_1$ (known as third order IMD products).

Inter-modulation (IM) products can occur within both the transmission (TX) and reception (RX) bands, making them more critical than harmonics.

Envelope distortions

In a more general I/Q modulated signal, the envelope is not constant and can vary over time. Therefore, the input signal can be represented as

$$x_{RF}(t,\tau) = A(\tau)cos(\omega_0 t + \phi_0(\tau)) \tag{5}$$

where $x_{RF}(t,\tau)$ represents the RF signal as a function of time (t) and envelope variation (τ). The signal has an amplitude that varies with time and is denoted by $A(\tau)$, while the carrier frequency is denoted by ω_0 and the initial phase by $\phi_0(\tau)$.

Envelope distortion can result in a phenomenon called spectral regrowth, which is a significant concern in communication systems. Spectral regrowth occurs when the non-linear distortion caused by envelope distortion creates additional spectral components that fall outside and around the allocated channel bandwidth.

Since envelope distortion occurs in the vicinity of the allocated channel, it can lead to spectral components spreading into adjacent channels, causing interference and disrupting communication. Therefore, mitigating envelope distortion is critical in communication systems to ensure that the signal remains within the allocated channel and does not interfere with neighboring channels.

2.1.3 Behavioral modeling of power amplifier

Behavioral modeling is a technique that involves mapping the input-output behavior of a power amplifier (PA) through a mathematical function, based on sampled measured data as shown in Fig. 2.6.

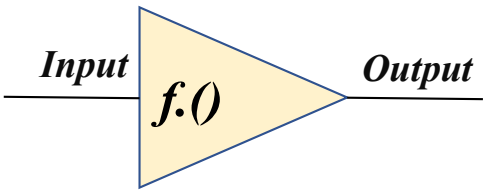

Figure 2.6. Behavioral Modeling of Power Amplifier.

Unlike other modeling techniques that require knowledge of the internal circuit, behavioral modeling can be performed without such knowledge and is sometimes referred to as black-box modeling.

One of the key advantages of behavioral models is that they are easy to characterize and can provide high accuracy for a given set of operating conditions, such as power, bandwidth, and temperature. Additionally, behavioral models have desirable properties that make them useful for a variety of applications. For instance, they can be implemented in digital circuits, and can model both static and dynamic distortion.

Overall, behavioral modeling is a powerful tool that allows designers to accurately predict the behavior of a PA without the need for complex and time-consuming circuit analysis. This makes it a valuable technique for designing and optimizing communication

systems. To include memory in the behavioral model, Finite Impulse Response (FIR) filters can be utilized as below.

- **Weiner Model**

  FIR filter is utilized before the static nonlinearity as shown in Fig. 2.7.

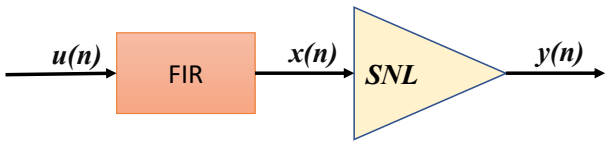

Figure 2.7. Wiener Model.

- **Hammerstein Model**

  Static Non Linearity (SNL) followed by the FIR filter as demonstrated in Fig. 2.8.

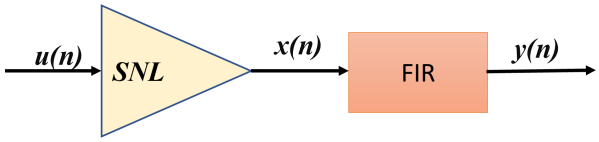

Figure 2.8. Hammerstein Model.

- **Weiner-Hammerstein Model**

  As shown in Fig. 2.9, Wiener-Hammerstein similar to the PA structure:
  input matching circuit –> transistor–> output matching circuit.

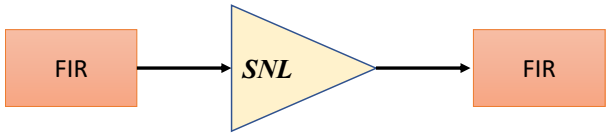

Figure 2.9. Wiener-Hammerstein Model.

Volterra Series Model

The Volterra series is a powerful tool for accurately modeling nonlinear functions with memory effects, as demonstrated in various studies [7]. In this series, the relationship between the input and output signals can be represented mathematically as [20]

$$x_{out}(n) = \sum_{k=1}^{K} \sum_{i_1=0}^{M} \dots \sum_{i_p=0}^{M} h_p(i_1, \dots\dots, i_p) \prod_{j=1}^{k} x_{in}(n - i_j) \tag{6}$$

where $h_p(i_1, \dots\dots, i_p)$ are the parameters of the Volterra model, K is the nonlinearity order of the model, and M is the memory depth.

Memory Polynomial Model

The Memory Polynomial (MP) model [8] is a simplified version of the Volterra series, developed to reduce its computational complexity while maintaining a high degree of accuracy for many practical scenarios. Due to its reliable accuracy, the memory polynomial model is widely used in Power Amplifier (PA) behavioral modeling.

The mathematical relationship between the input and output signals in the memory polynomial model can be represented using

$$x_{out}(n) = \sum_{k=1}^{K} \sum_{m=0}^{M} a_{k,m} x(n-m) |x(n-m)|^{k-1}. \tag{7}$$

In (7), $x(n)$ represents the input signal, while $x_{out}(n)$ represents the output of the Memory Polynomial (MP) model. The parameter K denotes the maximum non-linearity order, while M and $a_{k,m}$ represent the memory depth and the kernels of the model, respectively. The $|.|$ operator is used to denote the absolute value of a quantity.

Generalized Memory Polynomial Model

The Generalized Memory Polynomial (GMP) [9, 22] is an extended version of the memory polynomial that considers memory cross-terms to enhance its memory modeling capabilities. The mathematical formulation of the GMP is shown as

$$\begin{aligned} x_{out}(n) = & \sum_{k=1}^{K_a} \sum_{m=0}^{M_a} a_{k,m} x(n-m) |x(n-m)|^{k-1} \\ & + \sum_{k=1}^{K_b} \sum_{m=0}^{M_b} \sum_{l=1}^{L_b} b_{k,m,l} x(n-m) |x(n-m-l)|^{k-1} \\ & + \sum_{k=1}^{K_c} \sum_{m=0}^{M_c} \sum_{l=1}^{L_c} c_{k,m,l} x(n-m) |x(n-m+l)|^{k-1}. \end{aligned} \tag{8}$$

Equation (8) involves two signals: $x(n)$ as the input and $x_{out}(n)$ as the output produced by the Generalized Memory Polynomial (GMP) model. The GMP model has several parameters, including K_a, M_a, and $a_{k,m}$, which respectively represent the maximum non-linearity order, the memory depth, and the kernels of the model for the aligned term between the input and its envelope. Additionally, the model has parameters K_b, M_b, L_b, and $b_{k,m}$, which respectively represent the maximum non-linearity order, the memory depth, the lagging cross-term index, and the kernels of the model for the input and its lagging envelope terms. Finally, the model also has parameters K_c, M_c, L_c, and $c_{k,m}$, which respectively represent the maximum non-linearity order, the memory depth, the lagging cross-term index, and the kernels of the model for the input and its leading envelope terms.

As expected, the Generalized Memory Polynomial (GMP) model produces more accurate modeling results than the Memory Polynomial (MP) model. However, it should be noted that the GMP model is more complex than the MP model.

Neural Network based models

Neural networks have become a popular choice for modeling power amplifiers due to their adaptive nature and universal approximation capability [23–27]. Various neural network architectures have been proposed over time to account for memory effects. The single-input-single-output feedforward model is a basic neural network that can be used to model power amplifier nonlinearities [25,28]. To separately extract the amplitude and phase responses, a polar feedforward NN was introduced [26], as illustrated in Fig. 2.10. Another commonly used model for power amplifier distortion is the Cartesian NN, also known as the real-valued feedforward NN, whose structure is shown in Fig. 2.11 [28].

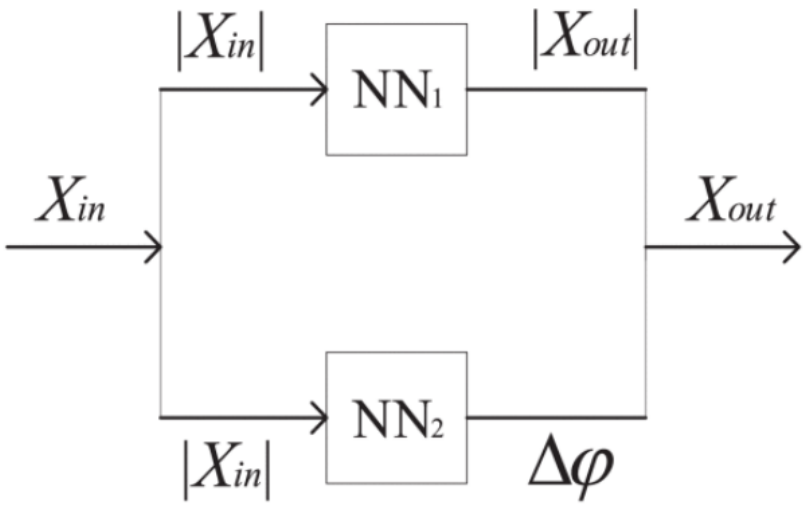

Figure 2.10. Polar topology [6].

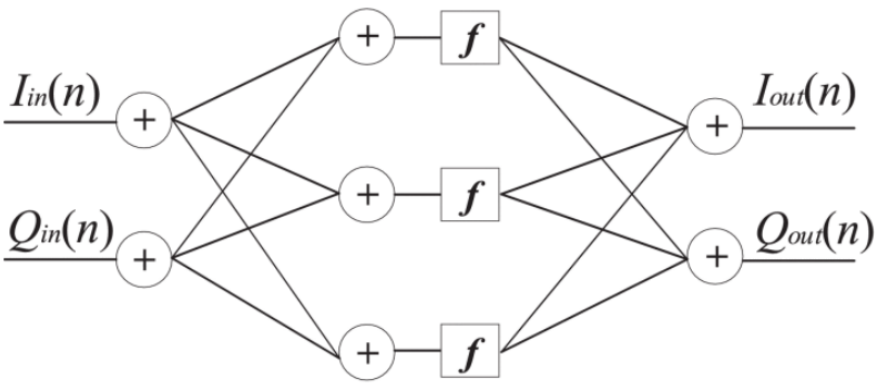

Figure 2.11. Cartesian topology [6].

In recent years, both shallow neural networks with fewer hidden layers [6] and deep neural networks (DNNs) with multiple hidden layers [11] have been used for power amplifier modeling and digital predistortion. While the simpler network structure and training process of shallow neural network models can be advantageous, their fewer

hidden layers limit their ability to capture the nonlinear effects of power amplifiers compared to DNN-based models. However, high complexity can be an issue for DNN-based models in high-bandwidth situations. To address this, LSTM networks [12,13] and RNN networks [14] have been introduced to capture time-series information of the input signal. The LSTM-DNN model [12] consists of an LSTM layer and two fully connected layers. To further reduce the complexity of this model, the LSTM-CNN model was proposed [13].

Modelling accuracy of the behavioral models can be compared using Normalized Mean Squared Error (NMSE) given by

$$NMSE = 20log_{10}\frac{rms(error)}{rms(measured\_signal)} \tag{9}$$

$$error = measured\_signal - model\_output. \tag{10}$$

2.2 Power amplifier linearization

As mentioned earlier, a RF power amplifier is a nonlinear active device. Apart from this, the transfer function of the PA is time variant. PA can be made linear by backing off the input power until the signal is within the linear region of the PA. This is known as Input Back-off (IBO). But this will reduce the efficiency of PA. There are different PA linearization techniques to linearize PAs. Some of the examples are, feedback linearization, feedforward linearization and digital predistortion. By linearizing the PA, the output of the PA can be increased, which results in higher power efficiency. Therefore, PA can be designed with smaller maximum output power, which will reduce the expense of PA. It also leads to less power consumption and hence cheaper operation. Therefore, linearizing PA is very important.

2.2.1 Challenges for PA linearization

In modern communication networks, information is encoded using both amplitude and phase modulation techniques. However, the amplitude of the modulated signal is not constant and can exhibit large peaks, leading to a high Peak-to-Average Power Ratio (PAPR). This high PAPR makes the signal more susceptible to nonlinearities in the PA. To linearize the PA, the average output power must be backed off by the Crest Factor (CR), which is equal to the square root of PAPR. This requires a high input back-off when dealing with high PAPR. By reducing PAPR, the average power for a fixed peak power can be increased, which helps to improve efficiency. Various Crest Factor Reduction (CFR) techniques, such as clipping, can be used for modulated signals. However, the Error Vector Magnitude (EVM) requirements limit the degree of PAPR reduction that can be achieved. Furthermore, the transfer functions of actual PAs are more complicated and time-variant. This means that the PA can start to become nonlinear much earlier than the expected saturation point. Additionally, the memory effects of actual PAs are significant.

Another effect of nonlinearity in PAs is that they produce harmonics and intermodulation among subcarriers. This result in a distorted spectrum that is much

wider than the transmitted signal. While it is possible to remove the harmonics using a transmit filter, the intermodulation products may still extend into adjacent channels.

2.2.2 Digital Pre-distortion (DPD)

There are various techniques used for linearizing PAs, and one such method is DPD. A block diagram of a DPD-based linear PA is shown in Fig 2.12.

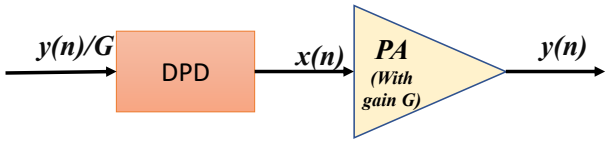

Figure 2.12. Block diagram of DPD based linear PA.

DPD is a technique that is similar to feed-back linearization. In DPD, the parameters are updated using information obtained from the feedback signal of the PA output, which is sent back to DPD via an Analog to Digital Converter (ADC).

The purpose of DPD is to train the parameters to obtain the inverse response of the PA. This enables DPD to modify the input signal of the PA by applying an inverse transfer function, resulting in better performance. DPD can be trained using various approaches, such as Least Mean Square (LMS), Least Square (LS), and Recursive Least Square (RLS). However, the main challenge of DPD lies in its computational complexity. There is always a trade-off between the computational complexity of DPD and the efficiency gains of the PA.

There are primarily three architectures used for DPD learning. These architectures are the Open Loop Architecture, the Indirect Learning Architecture (ILA), and the Direct Learning Architecture.

Open Loop Architecture

In an open loop architecture shown in Fig. 2.13, the forward model is typically trained first and then inverted to obtain the inverse model. However, this approach may not be accurate when dealing with tasks that involve memory, and therefore, it is rarely used in practical applications.

Indirect Learning Architecture (ILA)

In the ILA shown in Fig. 2.14, the first step is to train a post-inverse model using input-output data pairs obtained from the PA. This post-inverse model is then utilized as the digital pre-distortion (DPD) model, assuming that the post-inverse is equivalent to the pre-inverse. This method is widely used because it offers excellent performance and is relatively easy to implement.

Direct Learning Architecture

The direct learning architecture shows in Fig. 2.15 trains the DPD by minimizing the difference between the ideal signal and the distorted output signal produced by the power

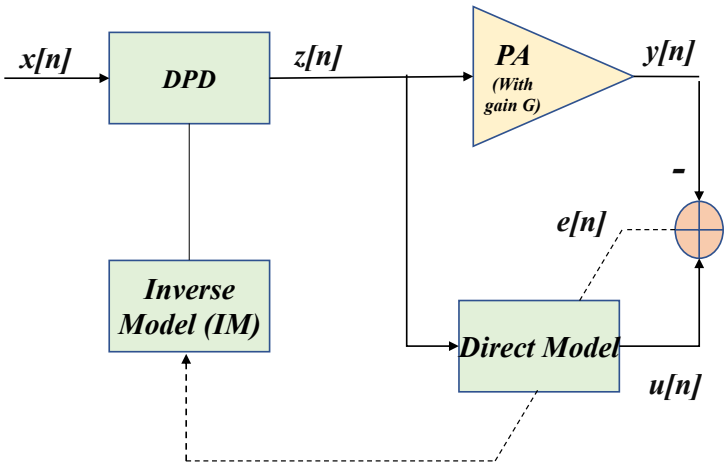

Figure 2.13. Open Loop Architecture.

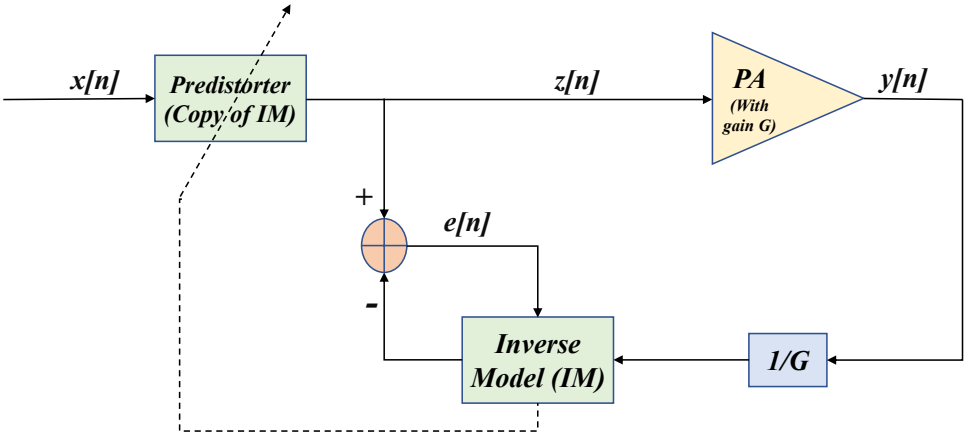

Figure 2.14. Indirect Learning Architecture.

amplifier (PA). While this approach typically yields better results, it may also exhibit stability issues in the long term and require a longer time to converge.

Assessing Linearization performance

The quality of linearization should be evaluated based on both in-band and out-of-band measurements. In-band quality is typically assessed using the Error Vector Magnitude (EVM) metric, which is calculated as follows:

$$EVM(\%) = \sqrt{\frac{P_{error}}{P_{ref}}} * 100. \tag{11}$$

Here, the power of the error signal P_{error} is determined by taking the difference between the ideal transmit signal and the corresponding complex samples at the output of the PA after amplitude and phase equalization. P_{ref} is the power of the ideal constellation symbols.

Out-of-band quality can be evaluated by measuring the adjacent channel interference caused by a signal. The Adjacent Channel Leakage Ratio (ACLR) is a common measure of the amount of unwanted signal power that leaks from a transmitter's allocated channel

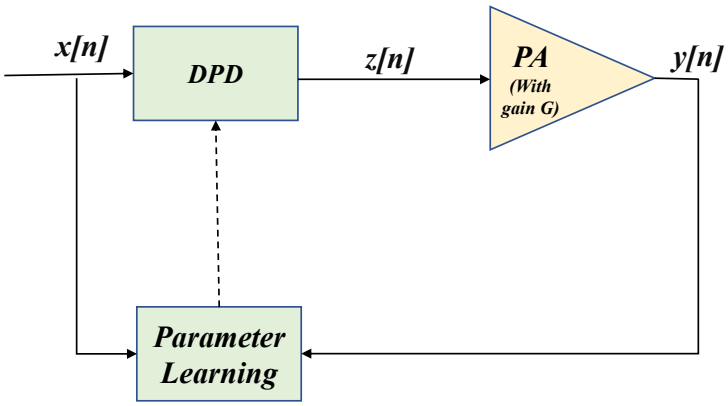

Figure 2.15. Direct Learning Architecture.

into the adjacent channels. ACLR is expressed in decibels (dB) and is calculated as the ratio of the power in a given adjacent channel to the power in the allocated channel, expressed in dBc as

$$ACLR(dB) = 10log_{10}\frac{P_{adjc\_chn}}{P_{alloc\_chn}}. \tag{12}$$

Overall, both in-band and out-of-band measurements are essential for evaluating the quality of linearization in communication systems. In-band measurements assess the accuracy of the transmitted signal within its assigned frequency band, while out-of-band measurements assess the interference caused by the signal in adjacent frequency channels.

Real-Valued Focused Time-Delay Neural Network (RVFTDNN) based Digital Predistorter

The RVFTDNN model [28], as depicted in Fig. 2.16, is composed of four layers, including an input layer, an output layer, and two hidden layers. The input to the model is comprised of both present and delayed I and Q values of the input signal and denoted as

$$\begin{aligned} x(n) =& [I_{\text{in}}(n), I_{\text{in}}(n-1), \ldots, I_{\text{in}}(n-M), \\ & Q_{\text{in}}(n), Q_{\text{in}}(n-1), \ldots, Q_{\text{in}}(n-M)]. \end{aligned} \tag{13}$$

Augmented Real-Valued Time-Delay Neural Network (ARVTDNN) based Digital Predistorter

The ARVTDNN model [6] shown in Fig. 2.17 has 3 layers namely input layer, output layer and one hidden layer. Input of the model is consists of current and delayed samples of IQ data and envelope dependent terms of the input represented as

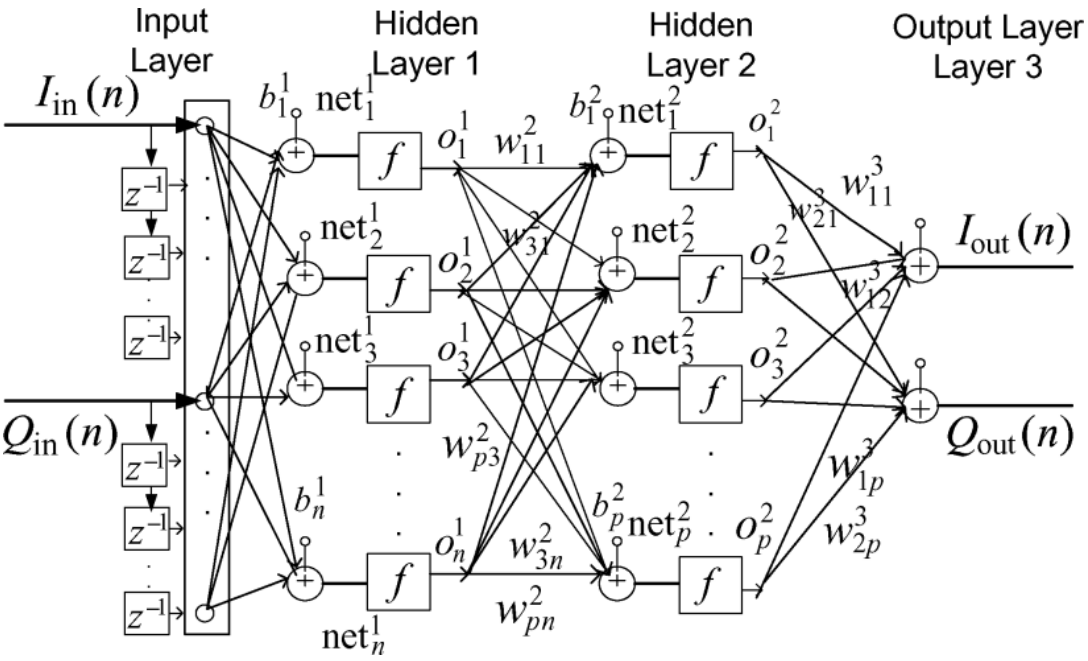

Figure 2.16. Structure of the RVFTDNN DPD model [28].

$$
\begin{aligned}
x(n) = [&I_{\text{in}}(n), I_{\text{in}}(n-1), \ldots, I_{\text{in}}(n-M),\\
&Q_{\text{in}}(n), Q_{\text{in}}(n-1), \ldots, Q_{\text{in}}(n-M)\\
&|X_{\text{in}}(n)|, |X_{\text{in}}(n-1)|, \ldots, |X_{\text{in}}(n-M)|,\\
&|X_{\text{in}}(n)|^2, |X_{\text{in}}(n-1)|^2, \ldots, |X_{\text{in}}(n-M)|^2,\\
&|X_{\text{in}}(n)|^3, |X_{\text{in}}(n-1)|^3, \ldots, |X_{\text{in}}(n-M)|^3]
\end{aligned} \tag{14}
$$

where $X_{in}(n)$ represents the current input signal and $I_{in}(n)$ and $Q_{in}(n)$ are the real and imaginary components of complex signal $X_{in}(n)$, respectively. $|X_{in}(n)|$ represents the amplitude of signal $X_{in}(n)$. $I_{in}(n-k)$ and $Q_{in}(n-k)$, $k = 1,2,\ldots,m$ represent the real and imaginary components of delayed samples and $|X_{in}(n-k)|$, $k = 1,2,\ldots,m$ denotes the amplitudes of delayed samples. Symbol M denotes the memory depth.

Deep Neural Network (DNN) based Digital Predistorter

A Deep Neural Network (DNN) consists of multiple hidden layers between its input layer and output layer. The generalization capability of the neural network increases exponentially with the depth of its layers [11]. In [11], the authors implemented a DNN-DPD with $L \geq 3$, which included one input layer, one output layer, and $L-2$ hidden layers. The number of neurons in the hidden layers was set to be equal to the number of neurons in the input layer, as shown in Fig. 2.18. In the DNN-DPD model, the input layer contained current I,Q samples and their delayed samples to handle memory distortion.

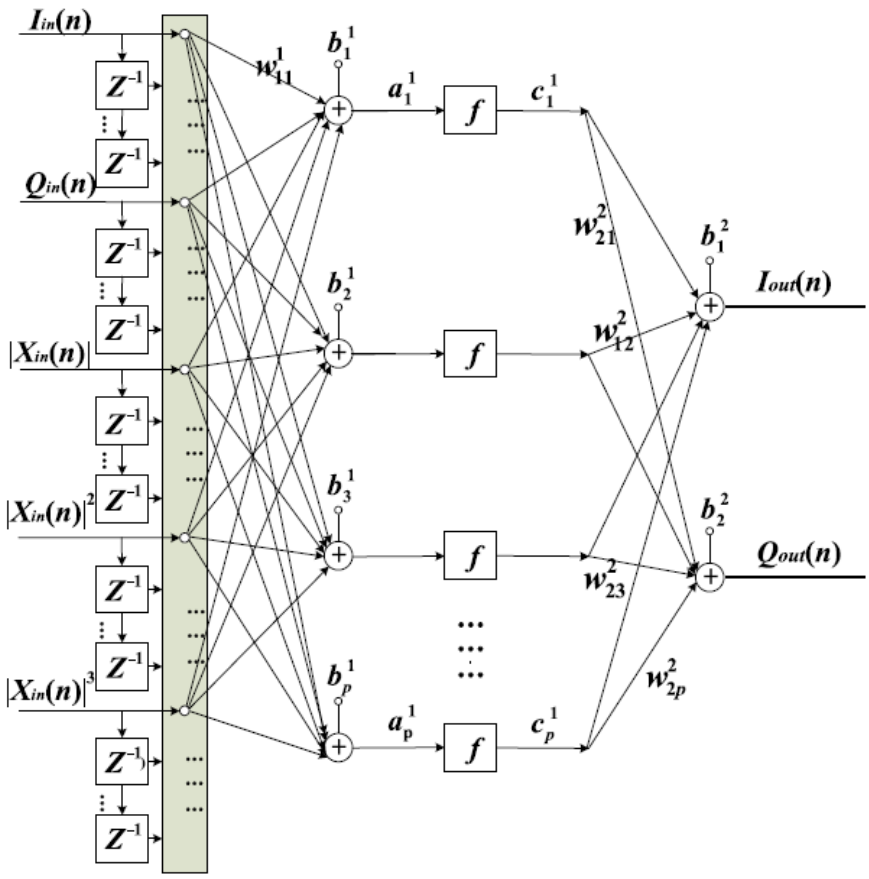

Figure 2.17. Structure of the ARVTDNN DPD model [6].

Convolutional Neural Network (CNN) based Digital Predistorter

The Real-Valued Time-Delay Convolutional Neural Network (RVTDCNN) model [1] comprises four layers: an input layer, an output layer, a pre-designed filter layer, and an FC layer, as shown in Fig. 2.19. The pre-designed filter layer uses a convolutional layer to capture the important features and characteristics of the input data. The input to the model is a combination of I/Q components and the envelope-dependent terms of the current and delayed signals. The 1D input data is mapped to a 2D matrix to make it suitable for the convolution process. The input matrix can be expressed as follows:

$$\begin{aligned} x(n) =& [I_{\text{in}}(n), I_{\text{in}}(n-1), \ldots, I_{\text{in}}(n-M), \\ & Q_{\text{in}}(n), Q_{\text{in}}(n-1), \ldots, Q_{\text{in}}(n-M) \\ & |X_{\text{in}}(n)|, |X_{\text{in}}(n-1)|, \ldots, |X_{\text{in}}(n-M)|, \\ & |X_{\text{in}}(n)|^2, |X_{\text{in}}(n-1)|^2, \ldots, |X_{\text{in}}(n-M)|^2, \\ & |X_{\text{in}}(n)|^3, |X_{\text{in}}(n-1)|^3, \ldots, |X_{\text{in}}(n-M)|^3] \end{aligned} \tag{15}$$

where $X_{in}(n)$ represents the current input signal and $I_{in}(n)$ and $Q_{in}(n)$ are the real and imaginary components of complex signal $X_{in}(n)$, respectively. $|X_{in}(n)|$ represents the amplitude of signal $X_{in}(n)$. $I_{in}(n-k)$ and $Q_{in}(n-k)$, $k = 1,2,\ldots,m$ represent the real

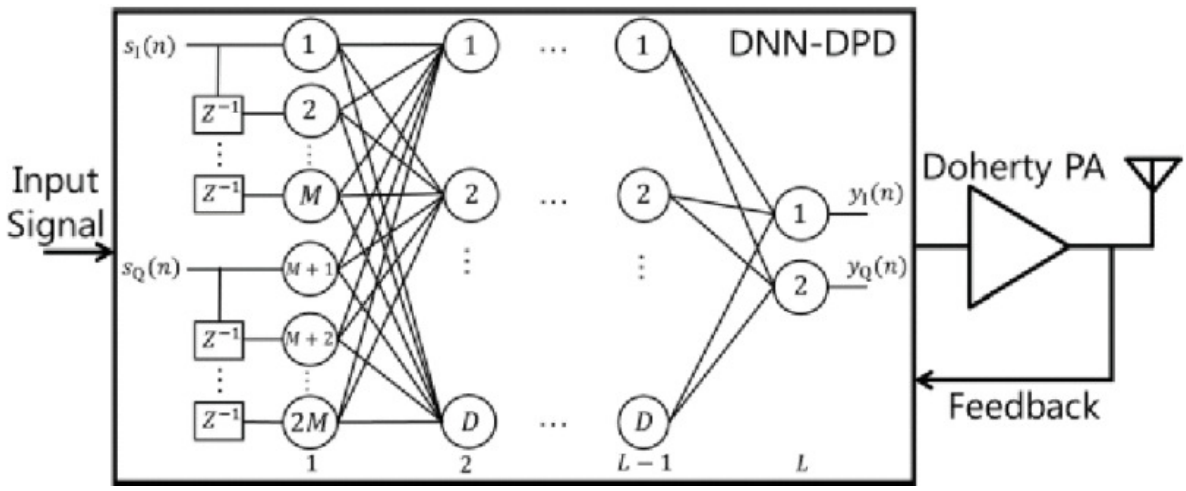

Figure 2.18. Structure of the DNN DPD model [11].

and imaginary components of delayed samples and $|X_{in}(n-k)|$, $k = 1,2,\ldots,m$ denotes the amplitudes of delayed samples. Symbol M denotes the memory depth. where $x(n)$ represents the current input signal and $I_{in}(n)$ and $Q_{in}(n)$ are the real and imaginary components of complex signal $X_{in}(n)$, respectively. $|x(n)|$ represents the amplitude of signal $X_{in}(n)$. $I_{in}(n-k)$ and $Q_{in}(n-k)$, $k = 1,2,\ldots,m$ represent the real and imaginary components of delayed samples and $|x(n-k)|$, $k = 1,2,\ldots,m$ denotes the amplitudes of delayed samples. Symbol M denotes the memory depth.

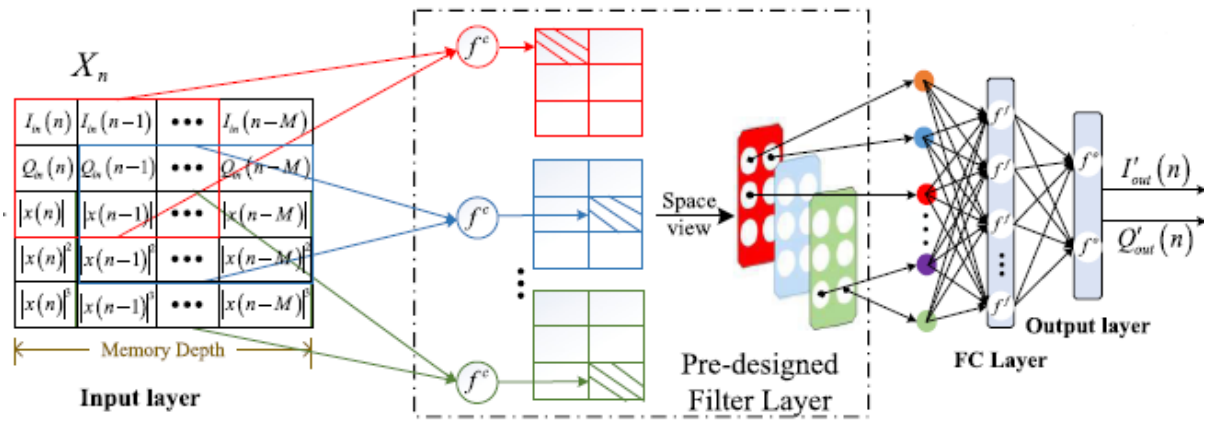

Figure 2.19. Structure of the RVTDCNN DPD model [1].

1D Convolutional Neural Network based Digital Predistorter

This low computational 1D CNN-based model [2] is an extension of RVTDCNN designed to reduce computational complexity. The proposed model achieves this by decomposing the two-dimensional convolution kernel into two types of one-dimensional convolutional kernels. The architecture of the model is depicted in Figure 2.20.

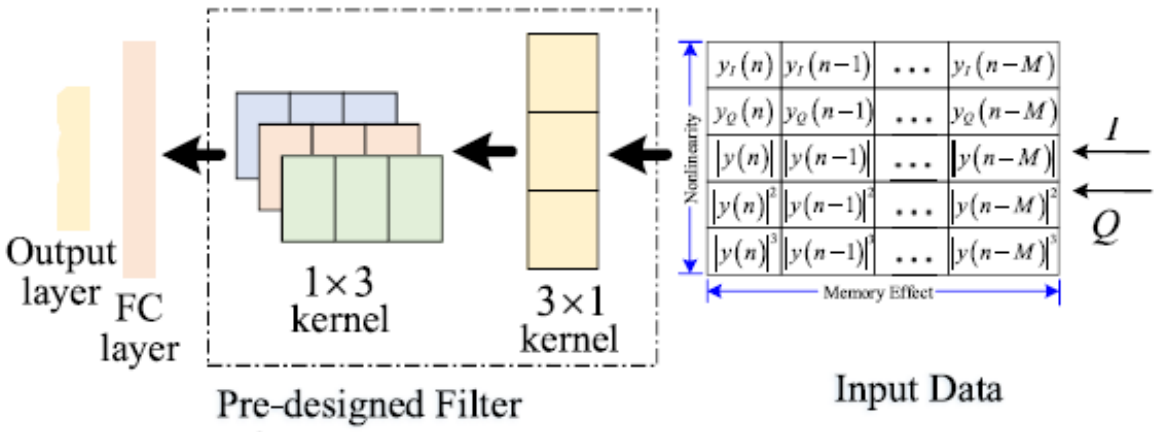

Figure 2.20. Structure of the 1D-CNN DPD model [2].

LSTM based Digital Predistorter

The LSTM-DNN model [12] shown in Fig. 2.21 is a combination of RNNs and shallow NN models. The model comprises of five layers: an input layer, an output layer, a LSTM layer, and two fully connected layers. The input to the model includes the I/Q components of present and past inputs, as well as past outputs, which are expressed as

$$\begin{aligned} X_{in}(n) =& [I_{in}(n), I_{in}(n-1) \dots I_{in}(n-m), \mathcal{Q}_{in}(n), \\ & \mathcal{Q}_{in}(n-1), \dots \mathcal{Q}_{in}(n-m), I_{out}(n), I_{out}(n-1), \dots \\ & I_{out}(n-t), \mathcal{Q}_{out}(n), \mathcal{Q}_{out}(n-1) \dots \mathcal{Q}_{out}(n-t)] \end{aligned} \tag{16}$$

where, m is memory depth in input, and t is memory delay length in output.

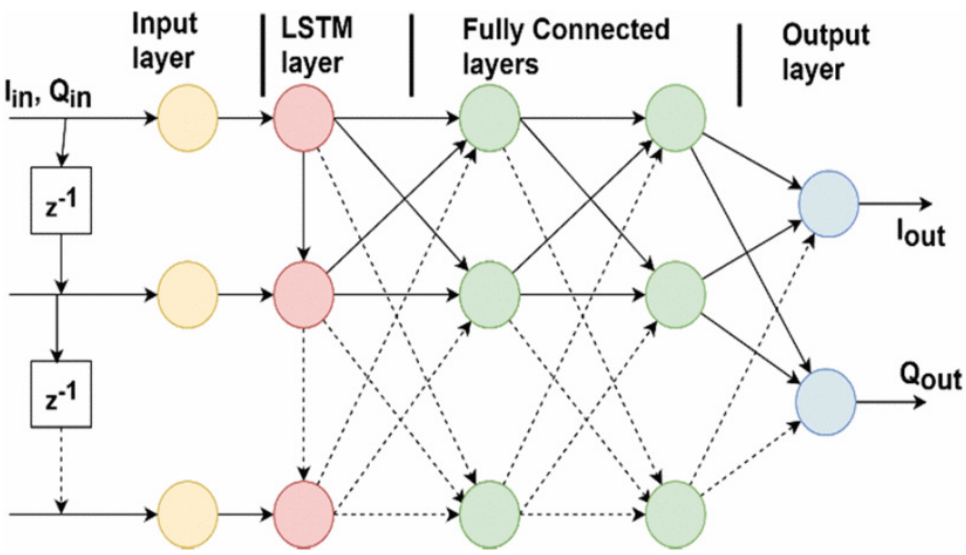

Figure 2.21. Structure of the LSTM-DNN DPD model [12].

Outputs I_{out} and Q_{out} are formulated as

$$I_{out}(n) = f_1(X_{in}(n)) \tag{17}$$

$$\mathcal{Q}_{out}(n) = f_2(X_{in}(n)) \tag{18}$$

where X_{in} is the input and f_1 and f_2 are unknown functions approximated by the proposed LSTM-DNN model.

3 PROPOSED MODELS

Two Digital Pre-Distortion (DPD) models are proposed in this thesis to linearize wideband power amplifiers. This chapter provides a detailed description of the structure and behavior of these two proposed models: the augmented-LSTM model and the 1D-CNN-LSTM model.

3.1 The Proposed augmented-LSTM Model

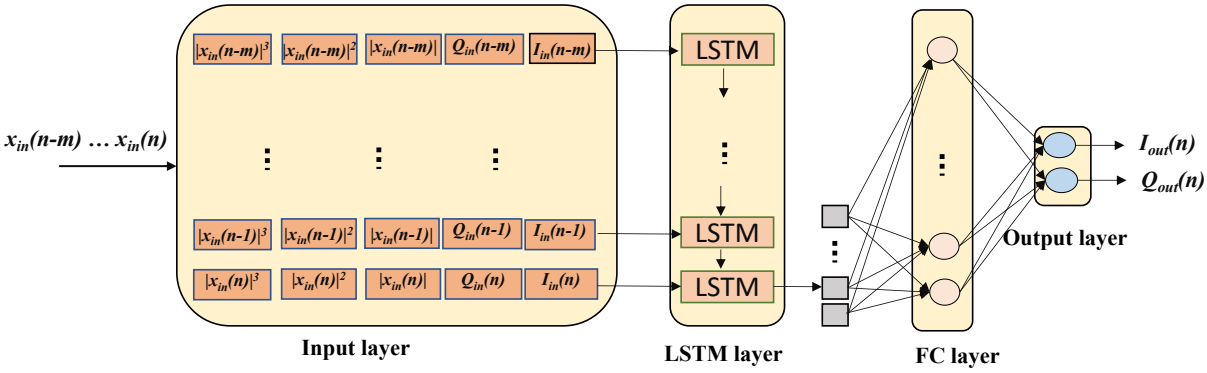

Figure 3.1. Structure of the Proposed augmented-LSTM model.

The proposed augmented-LSTM model utilizes more input features to model the DPD model in comparison to the LSTM-DNN model proposed in [12]. In addition to the real (I) and imaginary (Q) components used in LSTM-DNN model to model the predistortion, the proposed augmented-LSTM model also utilizes the amplitude of the input signal, similar to previous work in [1, 6].

With additional input features, the model can potentially learn more complex relationships between input and output which lead to improved performance. The additional features used in the proposed model are amplitude terms which can provide significant information to the model to enhance the accuracy.

The structure of the proposed augmented-LSTM model consists of four main layers; an input layer, a LSTM layer, a fully connected layer and an output layer as shown in Fig. 3.1 . The input layer serves as the initial point of data processing, where the input signal is normalized to a zero mean and a unity standard deviation. This is a common preprocessing step in neural networks and it has many benefits.

- **To mitigate the potential problem of gradient explosion**

  When the gradients in a neural network become too large, the network can struggle to converge to an optimal solution. This phenomenon is known as gradient explosion. Normalizing the input to have a zero mean and unit variance can help prevent this problem from occurring.

- **To improve performance**

  Normalizing the input signal to have a zero mean and unit variance ensures that the neural network is not biased towards particular features or dimensions of the

input signal. This can enable the network to identify patterns more effectively and make more accurate predictions.

- **To make training more efficient**

  Normalizing the input signal can accelerate the training process, as it enables the neural network to learn faster and converge to an optimal solution in fewer iterations.

The input to the model is a collection of real and imaginary (I/Q) components and envelop-dependent terms shown as

$$X_{in}(n) = \begin{bmatrix} I_{in}(n) & I_{in}(n-1) & \dots & I_{in}(n-m) \\ Q_{in}(n) & Q_{in}(n-1) & \dots & Q_{in}(n-m) \\ |x_{in}(n)| & |x_{in}n-1)| & \dots & |x_{in}(n-m)| \\ |x_{in}(n)|^2 & |x_{in}(n-1)|^2 & \dots & |x_{in}(n-m)|^2 \\ |x_{in}(n)|^3 & |x_{in}(n-1)|^3 & \dots & |x_{in}(n-m)|^3 \end{bmatrix} \tag{19}$$

where $x_{in}(n)$ represents the current input signal and $I_{in}(n)$ and $Q_{in}(n)$ are the real and imaginary components of complex signal $x_{in}(n)$, respectively. $|x_{in}(n)|$ represents the amplitude of signal $x_{in}(n)$. $I_{in}(n-k)$ and $Q_{in}(n-k)$, $k = 1,2,\ldots,m$ represent the real and imaginary components of delayed samples and $|x_{in}(n-k)|$, $k = 1,2,\ldots,m$ denotes the amplitudes of delayed samples. Symbol m denotes the memory depth.

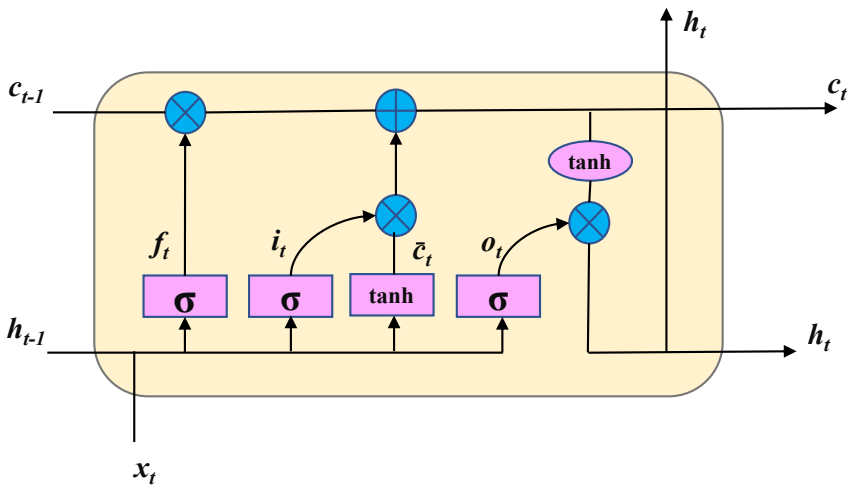

Figure 3.2. Internal structure of LSTM unit.

The input layer passes the processed data to the LSTM layer, which consists of a collection of LSTM units. The LSTM layer is a type of recurrent neural network layer that is specifically designed to handle sequential data, allowing it to model long-term dependencies in the input data. This is achieved through a special internal structure consisting of gates and memory cells, as shown in Fig. 3.2.

The LSTM layer is composed of several components, including the input gate (i_t), forget gate (f_t), memory cells, and output gate. The input gate determines how much weight to assign to new input and how much of the previous memory cell state to retain. The forget gate decides how much of the previous memory cell state to keep or discard at each time step. Meanwhile, the memory cells store relevant information over time, and the

output gate determines how much of the memory cell state to output as the final result of the LSTM unit. By combining these components, the LSTM layer can selectively retain or discard information from previous time steps, enabling it to handle sequential data and model long-term dependencies more effectively than a standard RNN unit. Additionally, the LSTM layer addresses the vanishing gradient problem that is often encountered in RNNs. To summarize, the LSTM layer generates an output sequence based on the input data sequence, incorporating information from both current and delayed inputs. This allows it to extract time-series information [12, 13]. The activation function used in this layer is the hyperbolic tangent activation function, *tanh*, which helps improve its performance.

The output generated by the LSTM layer is then passed to the fully connected layer. Here, the last timestamp of the sequence output from the LSTM layer is selected and pass to the fully connected layer. The fully connected layer, also known as the dense layer, is a standard neural network layer where each neuron receives input from every neuron in the previous layer and is connected to every neuron in the following layers. Then the activation function *tanh* is used to the output of the fully connected layer to introduce non- linearity into the output. Finally, the desired in-phase and quadrature values are produced by the output layer, which consists of two neurons corresponding to real and imaginary components of the output signal.

3.2 The Proposed 1D-CNN-LSTM Model

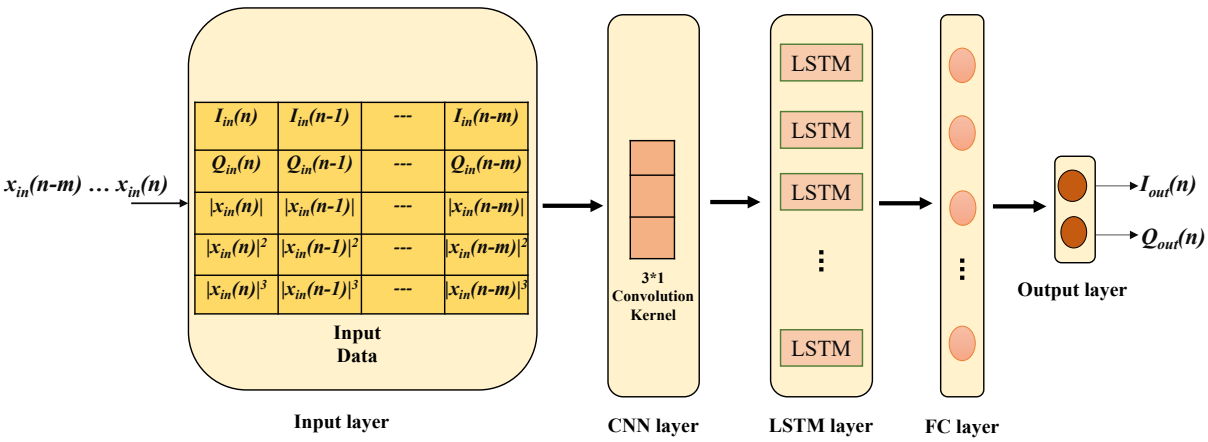

Figure 3.3. Structure of the Proposed 1D-CNN-LSTM model.

As the number of input features in a model increases, so does its complexity. This can make it challenging to understand which features are contributing to the model's predictions, thereby hindering improvements to the model. To simplify the the proposed augmented-LSTM model, a convolution layer is utilized before the LSTM layer as shown in Fig. 3.3 to reduce the input size of the LSTM layer. The main advantage of using a CNN layer in the network is parameter sharing. In a CNN layer, same kernel is applied to the every part of the input, enabling parameter sharing. This helps to reduce the number of parameters in the model, making training more efficient and reducing the risk of overfitting. By using a less number of input features to the LSTM layer, the overall

complexity of the model can be reduced. Hence the idea of this 1D-CNN-LSTM model is to reduce the complexity and to achieve comparable linearization results.

The proposed 1D-CNN-LSTM model is composed of several key layers; input layer, 1D-CNN layer, LSTM layer, fully connected layer and output layer. The Input layer process and normalize the input data and passes the normalized data to 1D convolution layer. The 1D convolution layer performs the convolution operation where it slides a small filter over the input data and computes the dot products between the overlapping sections of the Input. As shown in Fig. 3.4, the CNN layer uses a $3*1$ convolution kernel to extract the useful features in the same memory effect.

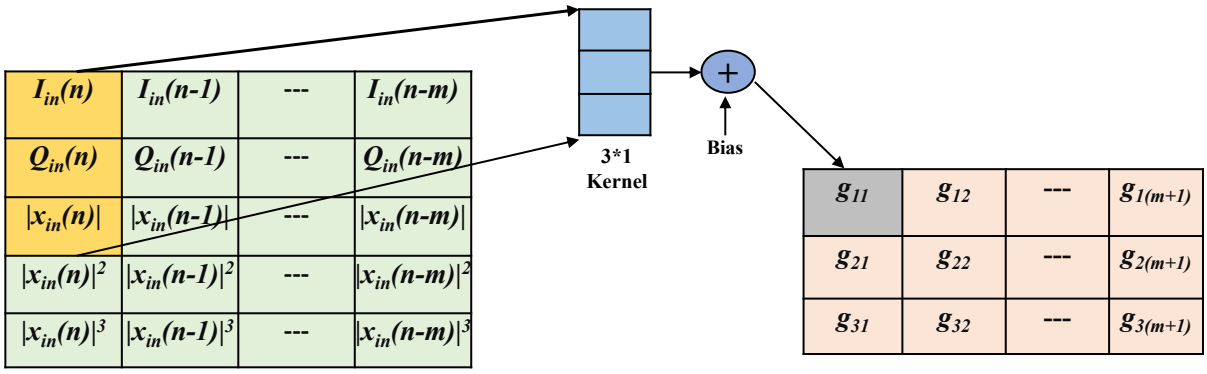

Figure 3.4. Convolution diagram.

Convolution operation on the input is shown as

$$
\begin{aligned}
g_k &= X_{in}(n) \circledast h_k \\
&= \begin{bmatrix} I_{in}(n) & I_{in}(n-1) & \dots & I_{in}(n-m) \\ Q_{in}(n) & Q_{in}(n-1) & \dots & Q_{in}(n-m) \\ |x_{in}(n)| & |x_{in}n-1)| & \dots & |x_{in}(n-m)| \\ |x_{in}(n)|^2 & |x_{in}(n-1)|^2 & \dots & |x_{in}(n-m)|^2 \\ |x_{in}(n)|^3 & |x_{in}(n-1)|^3 & \dots & |x_{in}(n-m)|^3 \end{bmatrix} \circledast h_k
\end{aligned} \tag{20}
$$

where g_k represents the 1D convolution output and h_k represent the coefficients of the 1D convolution kernel. Symbol $\circledast$ denotes the convolution operation.

Result of the convolution operation is as

$$
g_k = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1(m+1)} \\ g_{21} & g_{22} & \cdots & g_{2(m+1)} \\ g_{31} & g_{32} & \cdots & g_{3(m+1)} \end{bmatrix} \tag{21}
$$

where, g_{11} to $g_{3(m+1)}$ represent the convolution results as shown in Fig. 3.4. The output of the CNN layer is then passed to the LSTM layer to exploit time series information. The LSTM layer generates an output sequence by processing the input sequence. The last timestamp of this output sequence is then selected and passed to a fully connected layer. Finally, the output layer consisting of two neurons, provides the desired output related to I and Q components.

3.3 Extension to DPD

In this work, the inverse modeling of PA's nonlinear function was performed to model the DPD. The indirect learning architecture (ILA) [1] shown in Fig. 3.5, a popular method for extracting DPD models was employed for this purpose. The architecture involves training a neural network model to learn the inverse mapping of the non-linear function of the power amplifier. The inverse model was trained by using the output data of PA as

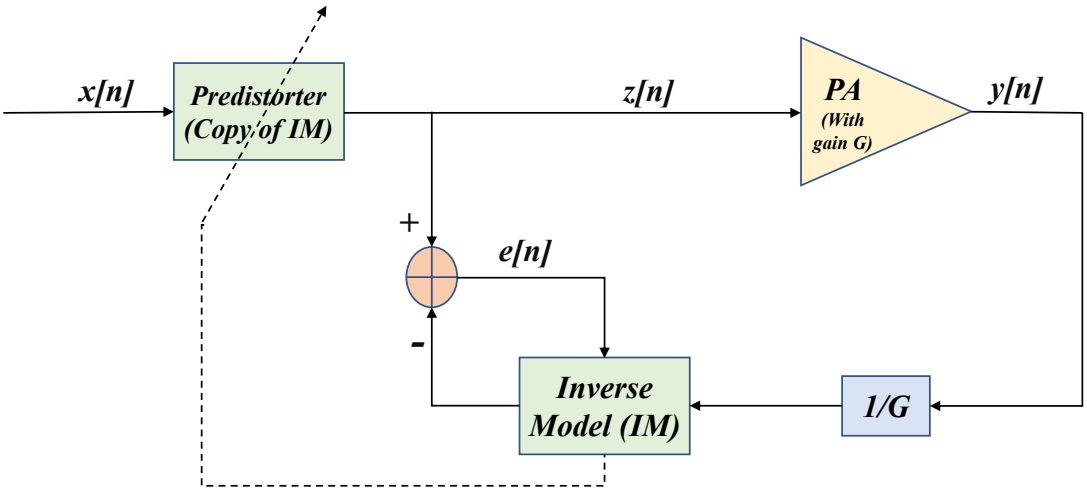

Figure 3.5. Indirect Learning Architecture.

the input data to the inverse model and input data of PA as the output data of the inverse model. Then using the trained DPD model, the main path DPD was updated to linearize the PA. Therefore, the input data matrix of PA's inverse model can be interpreted as

$$Y_n = \begin{bmatrix} I_{out}(n) & I_{out}(n-1) & \dots & I_{out}(n-m) \\ Q_{out}(n) & Q_{out}(n-1) & \dots & Q_{out}(n-m) \\ |y_{out}(n)| & |y_{out}n-1)| & \dots & |y_{out}(n-m)| \\ |y_{out}(n)|^2 & |y_{out}(n-1)|^2 & \dots & |y_{out}(n-m)|^2 \\ |y_{out}(n)|^3 & |y_{out}(n-1)|^3 & \dots & |y_{out}(n-m)|^3 \end{bmatrix} \tag{22}$$

where $y_{out}(n)$ represent the output signal of the PA and $I_{out}(n)$ and $Q_{out}(n)$ are the real and imaginary components of complex signal $y_{out}(n)$.

4 TRAINING, EVALUATION AND MEASUREMENT RESULTS

This chapter provides a comprehensive overview of the training procedure used to develop the proposed algorithm. It provides information on the training platforms used, the training parameters, and the evaluation criteria. Additionally, the latter part of the chapter compares the measurement results obtained from the proposed models with other Artificial Neural Network (ANN) based DPD models.

4.1 Training and evaluation of the proposed models

Training and evaluation of the proposed algorithm were performed using the power amplifier measurements data set provided by the Mathworks, Inc. (www.mathworks.com) [29]. According to [29], the data set was recorded using a NXP Airfast LDMOS Doherty power amplifier and the test signal employed was a 5G-like OFDM waveform, with each subcarrier carrying 16-QAM symbols. The specifications of the data set are listed in the Table 4.1.

Table 4.1. Characteristics of the training and testing data.

Parameter	**Value**
Operating frequency	3.6 - 3.8 GHz
Operating Bandwidth	100 MHz
Gain of the PA	29 dB

The training workflow is shown in Fig. 4.1 and the evaluation workflow is shown in Fig. 4.2. First the DPD was trained using the PA's input and output signals. Then the trained DPD was used to linearize the PA.

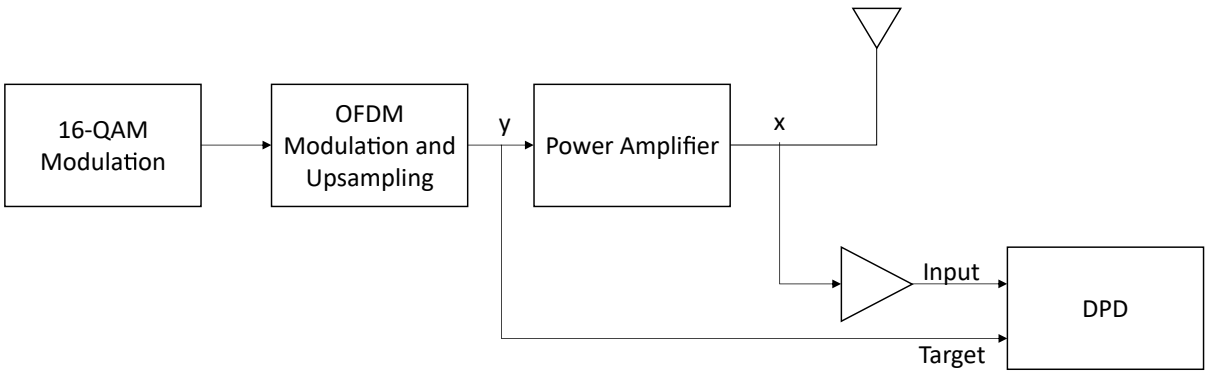

Figure 4.1. Training Workflow.

In the training algorithm, Adam optimization algorithm [30] was employed as the optimizing function for parameter optimization of the model. The Adam optimization algorithm is a computationally efficient technique that uses adaptive learning rates and

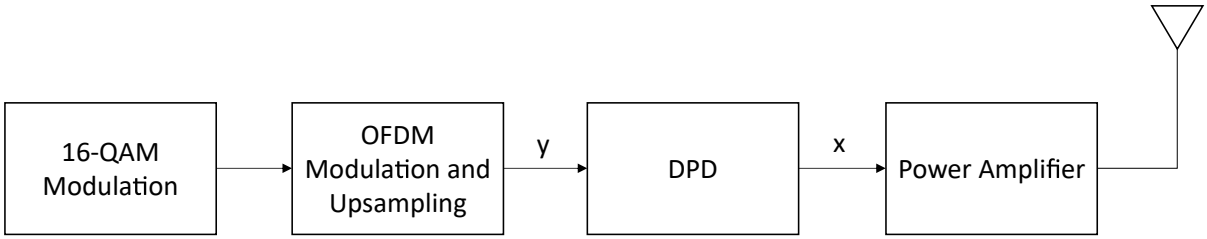

Figure 4.2. Evaluation Workflow.

momentum to update the model parameters during training. Mean Squared Error (MSE) was utilized as the cost function to measure the error between actual and predicted values. The MSE function calculates the average squared error between the actual and predicted values. To assess the performance of the training process and to prevent model from overfitting, a validation set was utilized. To further prevent overfitting, we implemented an early stopping criterion. This criterion stops the training process if the validation accuracy does not improve for a pre-specified number of validations (validation patience). This ensures that the model does not become overly specialized to the training data and can generalize well to new data.

4.2 Measurement results and Discussion

To evaluate the effectiveness of the proposed DPD model, it was implemented using the PyTorch framework. To evaluate the linearization performance of the DPD model, a comparison was made between the Normalized Power Spectral Density (PSD) of the actual PA output signal and the PA output signal obtained using the proposed augmented-LSTM DPD model. Fig. 4.3 illustrates the normalized PSD graphs, which demonstrate the linearization performance of the augmented-LSTM model.

The compensating effects of the proposed augmented-LSTM model on gain distortion and phase distortion are shown in Fig. 4.4 and Fig. 4.5, respectively. To further evaluate performance, Normalized Mean Squared Error (NMSE) and Adjacent Channel Power Ratio (ACPR) were calculated. The proposed DPD model significantly improves the NMSE from -22.17 dB to -33.49 dB and improves the ACPR performance from -26.69 dBc to -40.04 dBc (decibels relative to carrier). These results indicate that the proposed augmented-LSTM model has significant impact on reducing PA distortion.

In order to compare the linearization performance of the proposed augmented-LSTM model with other artificial neural network (ANN) models, several DPD models were also developed using the PyTorch framework. These models include LSTM-DNN [12], DNN [11], and ARVTDNN [6]. The number of neurons in the hidden layers of the LSTM-DNN network was set to [LSTM FC1 FC2] = [10 7 5], as described in the [12]. The number of neurons in the hidden layer of the ARVTDNN model was set to 17, as suggested in the [6], and the DNN model had a hidden layer structure of [17 17 17]. All of the models were trained using the same algorithm that was used to train the proposed augmented-LSTM model. By comparing the linearization performance of these different models, a more comprehensive understanding of their relative strengths and weaknesses can be gained.

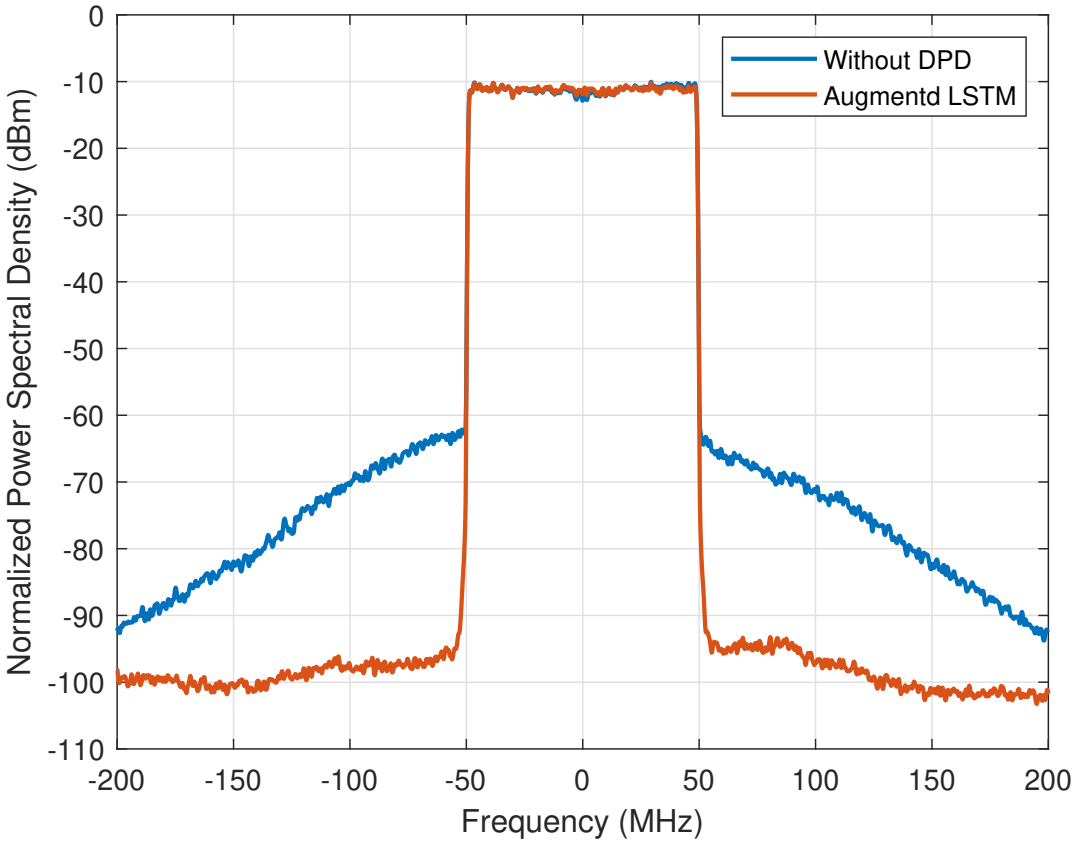

Figure 4.3. Linearization performance of the augmented-LSTM model.

Fig. 4.6 compares the normalized PSD of the proposed augmented-LSTM model with other ANN DPD models and memory polynomial DPD. Table 4.2 further compares the linearization performances and complexity through the use of NMSE, ACPR and number of model coefficients. The results suggest that the proposed LSTM model has superior linearization performances compared to other ANN DPD models.

The 1D-CNN-LSTM DPD model was proposed as a means of reducing complexity, and it achieved an impressive NMSE value of -33.31 dB. Furthermore, the ACPR performance was significantly improved from -26.69 dBc to -39.79 dBc. Fig. 4.7 provides a graphical representation of the linearization performances of both the 1D-CNN-LSTM DPD model and the augmented-LSTM model. The compensating effects of the proposed 1D-CNN-LSTM model on gain distortion and phase distortion are shown in Fig. 4.8 and Fig. 4.9, respectively. Table 4.3 provides a comparison of the achieved NMSE, ACPR value, and number of model coefficients for both the proposed 1D-CNN-LSTM model and augmented-LSTM model. These results demonstrate that integrating the CNN layer before the LSTM layer can maintain comparable linearization results while reducing the overall complexity of the model. This suggests that the 1D-CNN-LSTM DPD model may be a promising approach for addressing the challenge of complexity in DPD modeling. Furthermore, the improved performance in terms of NMSE and ACPR indicates that the 1D-CNN-LSTM DPD model is highly effective at reducing PA distortion.

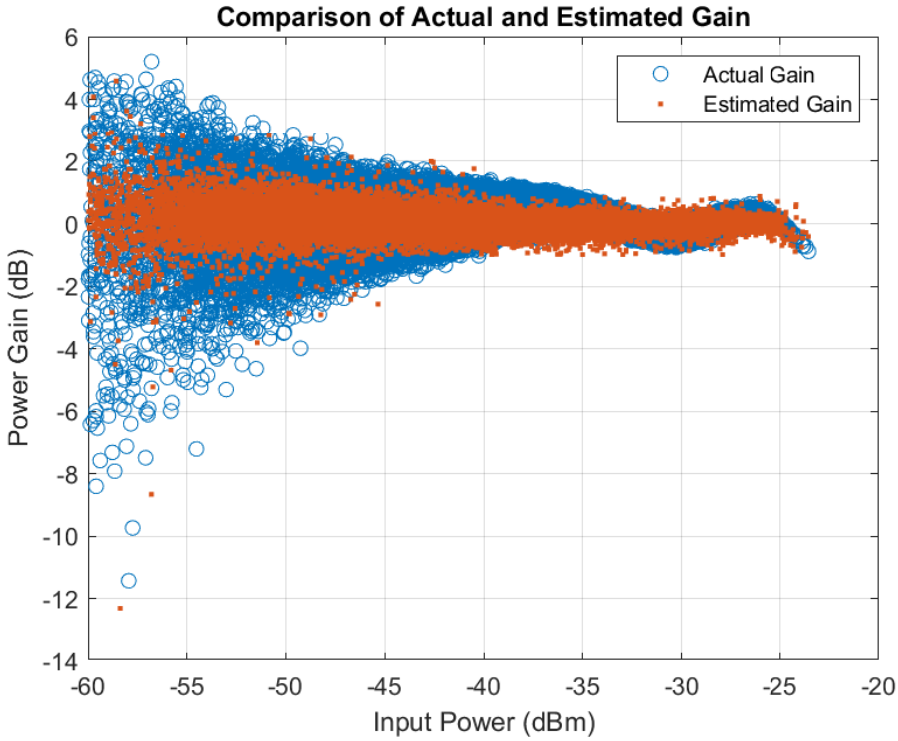

Figure 4.4. Gain characteristics.

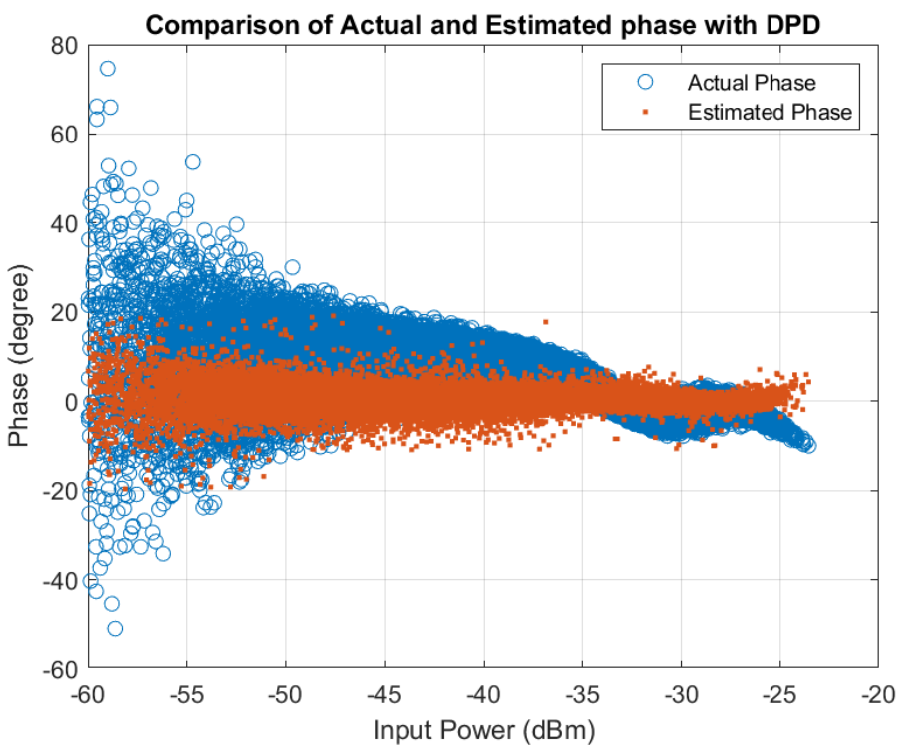

Figure 4.5. Phase characteristics.

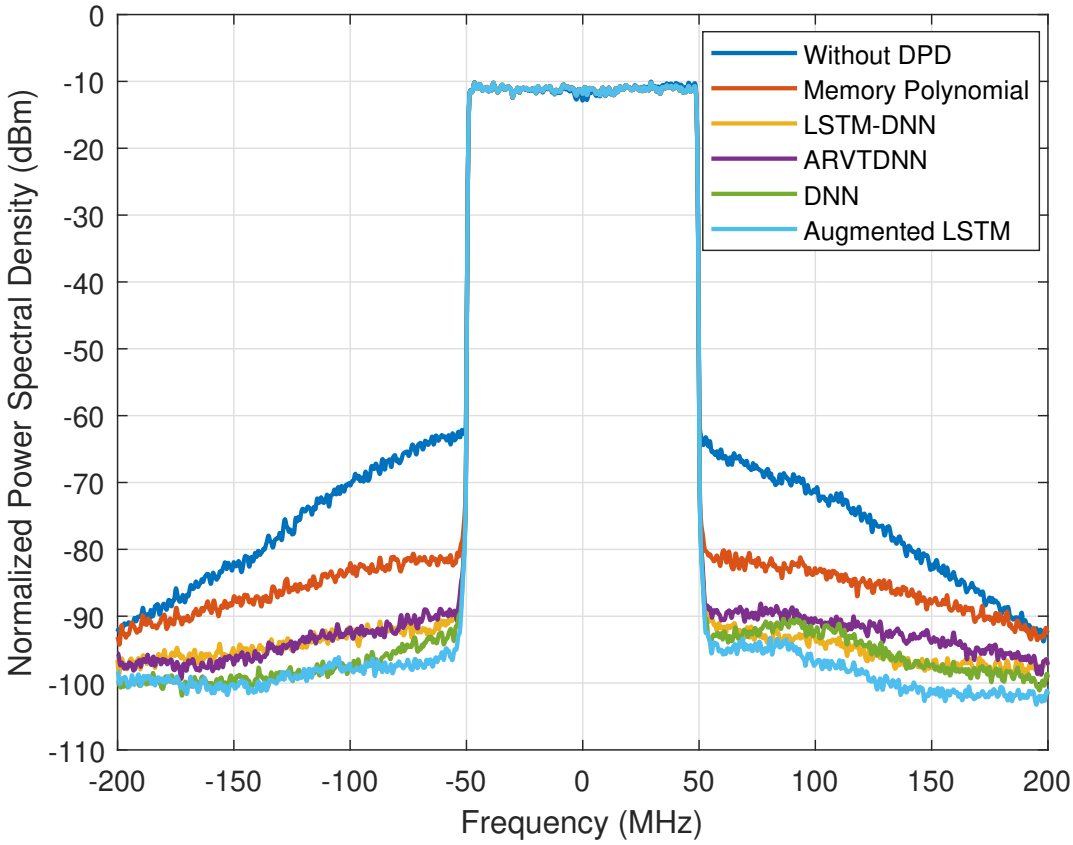

Figure 4.6. Linearization performance of PA using various DPD models.

Table 4.2. Comparison of performances and complexity of the augmented-LSTM model and other DPD models.

Model	**NMSE(dB)**	**ACPR (dBc) (-/+ 25 MHz)**	**Number of model coefficients**
Without DPD	-22.17	-26.19/-27.19	N/A
Memory Polynomial	-29.59	-34.76/-34.69	N/A
ARVTDNN [6]	-32.29	-38.54/-37.86	563
DNN [11]	-33.36	-40.15/-39.09	869
LSTM-DNN [12]	-32.79	-39.24/-38.45	689
Augmented-LSTM	-33.49	-41.01/-39.60	559

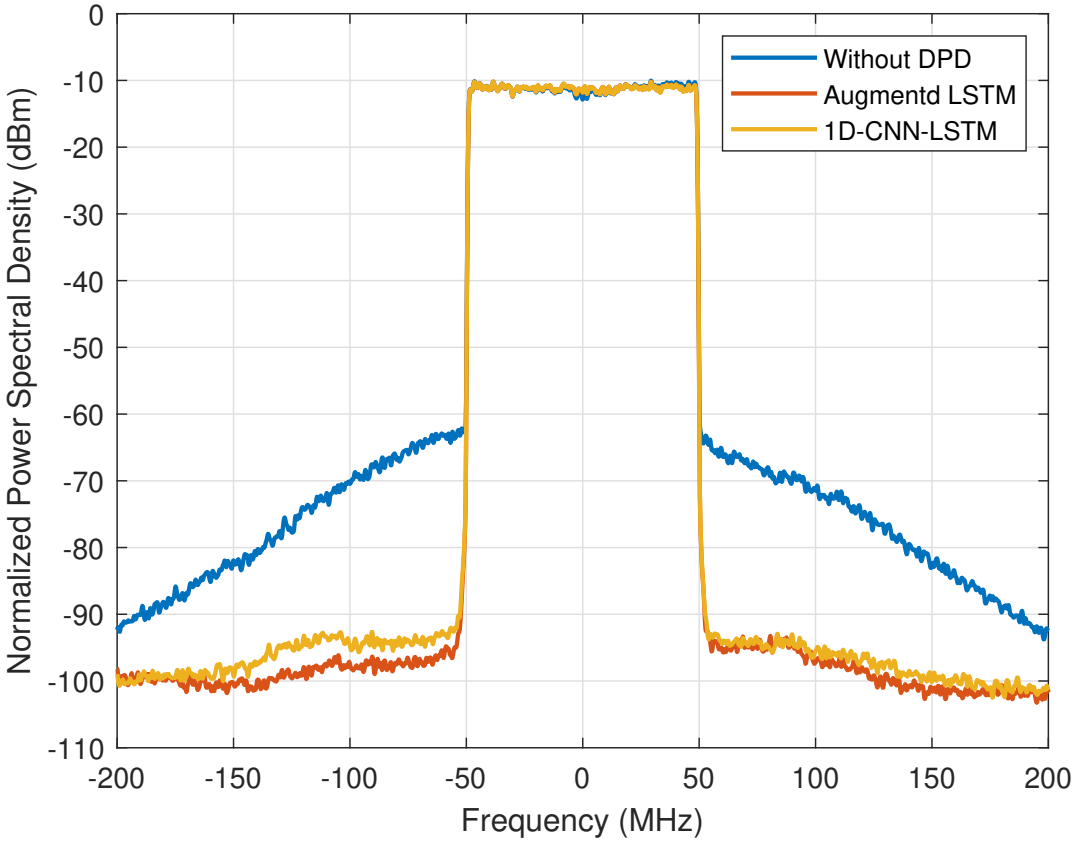

Figure 4.7. Linearization performance of PA using 1D-CNN-LSTM DPD.

Table 4.3. Comparison of performances and complexity of the 1D-CNN-LSTM model and the augmented-LSTM model.

Model	**NMSE(dB)**	**ACPR (dBc) (-/+ 25 MHz)**	**Number of model coefficients**
Without DPD	-22.17	-26.19/-27.19	N/A
Augmented-LSTM	-33.49	-41.01/-39.06	559
1D-CNN-LSTM	-33.31	-40.20/-39.37	466

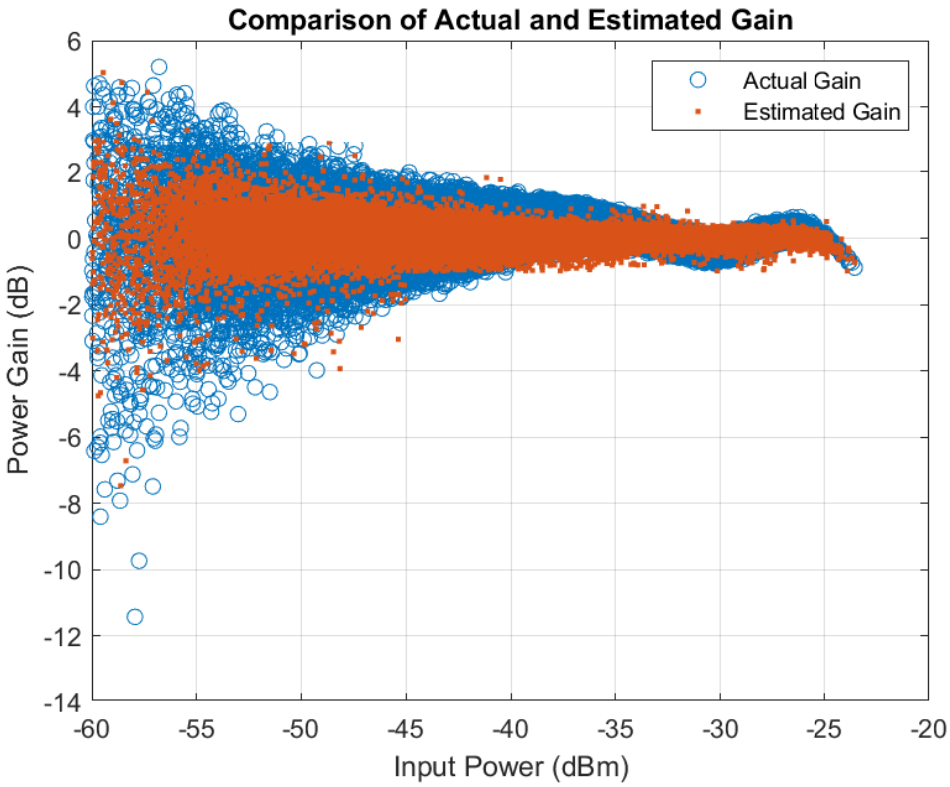

Figure 4.8. Gain characteristics.

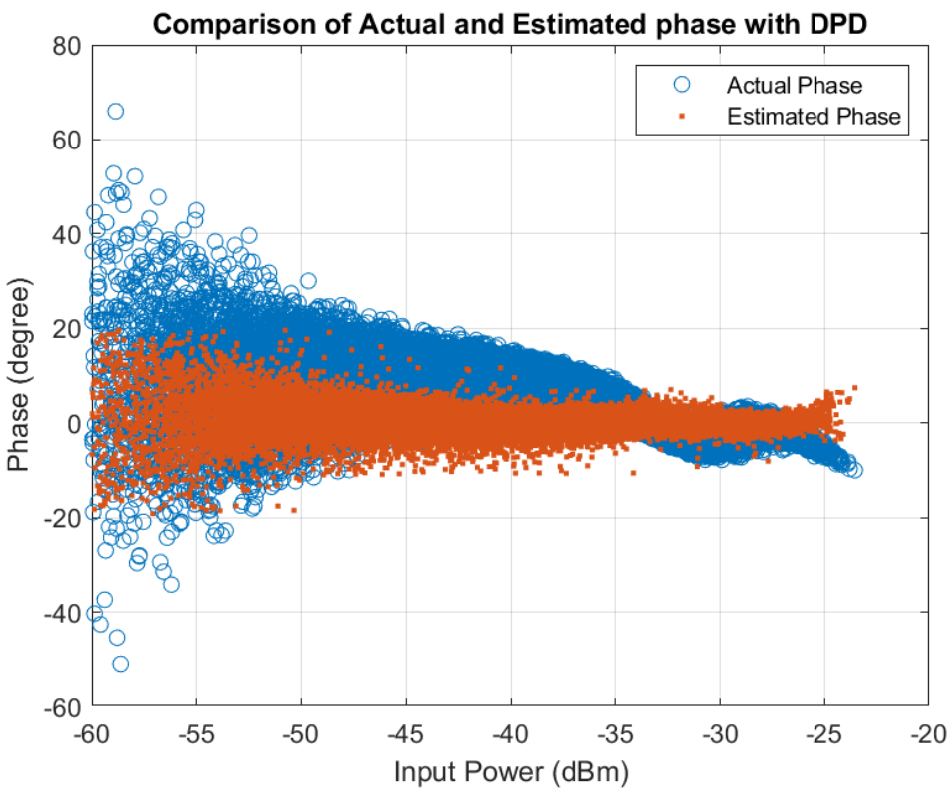

Figure 4.9. Phase characteristics.

5 CONCLUSION AND FUTURE WORK

There are two main outcomes in this study. First, an augmented-LSTM model is proposed to model DPD specifically for wideband PAs. The proposed model uses a comprehensive basis set to enhance the modeling performance, resulting in a significant improvement in NMSE from -22.17 to -33.49 dB and ACPR from -26.69 to -40.04 dBc. In comparison to other NN-based DPD models, the augmented-LSTM model demonstrates better linearization performance. Moreover, the complexity of the augmented-LSTM model is reduced using a 1D-CNN-LSTM model, where a CNN layer is utilized before the LSTM layer to capture useful features from the input data. By using this approach, the number of input features to the LSTM layer is reduced, leading to a decrease in complexity. The simulation results indicate that the 1D-CNN-LSTM model with reduced complexity achieves a comparable linearization performance to that of the augmented-LSTM model. These findings suggest that the 1D-CNN-LSTM DPD model has significant potential to provide more efficient and accurate linearization of PA output signals, which could have practical implications for improving the performance of wireless communication systems.

Further research could be conducted to implement the proposed DPD models in hardware to test their performances. This could involve testing the models on real-world data, evaluating their performance in comparison to other DPD models, and assessing their ability to meet the requirements of different wireless communication systems. Such research could lead to the development of more effective DPD models that can be widely adopted in the wireless communication industry.

6 REFERENCES

[1] Hu X., Liu Z., Yu X., Zhao Y., Chen W., Hu B., Du X., Li X., Helaoui M., Wang W. & Ghannouchi F.M. (2022) Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers. IEEE Transactions on Neural Networks and Learning Systems 33, pp. 3923–3937.

[2] Liu Z., Hu X., Xu L., Wang W. & Ghannouchi F.M. (2022) Low Computational Complexity Digital Predistortion Based on Convolutional Neural Network for Wideband Power Amplifiers. IEEE Transactions on Circuits and Systems II: Express Briefs 69, pp. 1702–1706.

[3] Reina-Tosina J., Allegue-Martínez M., Crespo-Cadenas C., Yu C. & Cruces S. (2015) Behavioral Modeling and Predistortion of Power Amplifiers Under Sparsity Hypothesis. IEEE Transactions on Microwave Theory and Techniques 63, pp. 745–753.

[4] Chani-Cahuana J., Özen M., Fager C. & Eriksson T. (2017) Digital Predistortion Parameter Identification for RF Power Amplifiers Using Real-Valued Output Data. IEEE Transactions on Circuits and Systems II: Express Briefs 64, pp. 1227–1231.

[5] Li Y. & Zhu A. (2020) On-Demand Real-Time Optimizable Dynamic Model Sizing for Digital Predistortion of Broadband RF Power Amplifiers. IEEE Transactions on Microwave Theory and Techniques 68, pp. 2891–2901.

[6] Wang D., Aziz M., Helaoui M. & Ghannouchi F.M. (2019) 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, pp. 242–254.

[7] de Figueiredo R. (1982) The Volterra and Wiener theories of nonlinear systems. Proceedings of the IEEE 70, pp. 316–317.

[8] Ku H. & Kenney J. (2003) Behavioral modeling of nonlinear RF power amplifiers considering memory effects. IEEE Transactions on Microwave Theory and Techniques 51, pp. 2495–2504.

[9] Morgan D., Ma Z., Kim J., Zierdt M. & Pastalan J. (2006) A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers. IEEE Transactions on Signal Processing 54, pp. 3852–3860.

[10] Du T., Yu C., Gao J., Liu Y. & Li S. (2012) A new accurate volterra-based model for behavioral modeling and digital predistortion of RF power amplifiers. Progress In Electromagnetics Research C 29, pp. 205–218.

[11] Hongyo R., Egashira Y., Hone T.M. & Yamaguchi K. (2019) Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers. IEEE Microwave and Wireless Components Letters 29, pp. 146–148.

[12] Phartiyal D. & Rawat M. (2019) LSTM-Deep Neural Networks based Predistortion Linearizer for High Power Amplifiers. In: 2019 National Conference on Communications (NCC), pp. 1–5.

[13] Wang W., Sun L., Liu H. & Feng Y. (2022) LSTM-CNN for Behavioral Modeling and Predistortion of 5G Power Amplifiers. In: 2022 IEEE 9th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE), pp. 28–32.

[14] Gong B., Feng Z., Liu C., Wang J., Zhang C., Pan C. & Xue Y. (2022) A Memorized Recurrent Neural Network Design for Wide Bandwidth PA Linearization. In: 2022 International Conference on Electrical Engineering and Photonics (EExPolytech), pp. 114–117.

[15] De Silva U., Koike-Akino T., Ma R., Yamashita A. & Nakamizo H. (2022) A Modular 1D-CNN Architecture for Real-time Digital Pre-distortion. In: 2022 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR), pp. 79–81.

[16] Wang S., Geng M., Yu C. & Cai J. (2022) Improved behavioral modeling using augmented lstm networks for ultra-broadband mmwave pa. Microwave and Optical Technology Letters 65.

[17] Wang S., Geng M., Yu C. & Cai J. (2021) Behavioral modeling of ultra-broadband mmwave power amplifier based on augmented long-short term memory networks. In: 2021 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM), pp. 1–3.

[18] Moulthrop A., Clark C., Silva C. & Muha M. (1997) A dynamic AM/AM and AM/PM measurement technique. In: 1997 IEEE MTT-S International Microwave Symposium Digest, pp. 1455–1458.

[19] Ghannouchi F.M. & Hammi O. (2009) Behavioral modeling and predistortion. IEEE Microwave Magazine 10, pp. 52–64.

[20] Ghannouchi F.M., Hammi O. & Helaoui M. (2015) Behavioral modelling and predistortion of wideband wireless transmitters. Wiley.

[21] Abdelrahman A.E., Hammi O., Kwan A.K., Zerguine A. & Ghannouchi F.M. (2016) A novel weighted memory polynomial for behavioral modeling and digital predistortion of nonlinear wireless transmitters. IEEE Transactions on Industrial Electronics 63, pp. 1745–1753.

[22] Liu Y.J., Zhou J., Chen W. & Zhou B.H. (2014) A robust augmented complexity-reduced generalized memory polynomial for wideband RF power amplifiers. IEEE Transactions on Industrial Electronics 61, pp. 2389–2401.

[23] Haykin S. (1998) Neural networks: A comprehensive foundation. Macmillan.

[24] Luongvinh D. & Kwon Y. (2005) Behavioral modeling of power amplifiers using fully recurrent neural networks. In: IEEE MTT-S International Microwave Symposium Digest, 2005., pp. 1979–1982.

[25] Ibukahla M., Sombria J., Castanie F. & Bershad N. (1997) Neural networks for modeling nonlinear memoryless communication channels. IEEE Transactions on Communications 45, pp. 768–771.

[26] Benvenuto N., Piazza F. & Uncini A. (1993) A neural network approach to data predistortion with memory in digital radio systems. In: Proceedings of ICC '93 - IEEE International Conference on Communications, pp. 232–236.

[27] Yang J., Gao J. & Jiang F. (2008) Neural network predistortion technique for nonlinear rf amplifiers. In: 2008 International Conference on Computer Science and Software Engineering, pp. 769–772.

[28] Rawat M. & Ghannouchi F.M. (2012) A mutual distortion and impairment compensator for wideband direct-conversion transmitters using neural networks. IEEE Transactions on Broadcasting 58, pp. 168–177.

[29] Data set provided by the MathWorks, Inc., MATLAB version 2022a. http://www.mathworks.com [Accessed on 01-Sep-2022].

[30] Kingma D.P. & Ba J. (2015), Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980v9.