


Article

# Experimental Demonstration and Performance Enhancement of 5G NR Multiband Radio over Fiber System Using Optimized Digital Predistortion

Muhammad Usman Hadi <sup>1,\*</sup>, Muhammad Awais <sup>2</sup>, Mohsin Raza <sup>2</sup>, Muhammad Ikram Ashraf <sup>3</sup> and Jian Song <sup>4</sup><sup>1</sup> School of Engineering, Ulster University, Newtownabbey BT37 0QB, UK<sup>2</sup> Faculty of Computer Engineering, Edge Hill University, Ormskirk L39 4QP, UK; m.awais@edgehill.ac.uk (M.A.); Razam@edgehill.ac.uk (M.R.)<sup>3</sup> Nokia Bell Labs, FI-02610 Espoo, Finland; ikram.ashraf@nokia-bell-labs.com<sup>4</sup> Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark; jians@es.aau.dk

\* Correspondence: usmanhadi@ieee.org

**Abstract:** This paper presents an experimental realization of multiband 5G new radio (NR) optical front haul (OFH) based radio over fiber (RoF) system using digital predistortion (DPD). A novel magnitude-selective affine (MSA) based DPD method is proposed for the complexity reduction and performance enhancement of RoF link followed by its comparison with the canonical piece wise linearization (CPWL), decomposed vector rotation method (DVR) and generalized memory polynomial (GMP) methods. Similarly, a detailed study is shown followed by the implementation proposal of novel neural network (NN) for DPD followed by its comparison with MSA, CPWL, DVR and GMP methods. In the experimental testbed, 5G NR standard at 20 GHz with 50 MHz bandwidth and flexible-waveform signal at 3 GHz with 20 MHz bandwidth is used to cover enhanced mobile broad band and small cells scenarios. A dual drive Mach Zehnder Modulator having two distinct radio frequency signals modulates a 1310 nm optical carrier using distributed feedback laser for 22 km of standard single mode fiber. The experimental results are presented in terms of adjacent channel power ratio (ACPR), error vector magnitude (EVM), number of estimated coefficients and multiplications. The study aims to identify those novel methods such as MSA DPD are a good candidate to deploy in real time scenarios for DPD in comparison to NN based DPD which have a slightly better performance as compared to the proposed MSA method but has a higher complexity levels. Both, proposed methods, MSA and NN are meeting the 3GPP Release 17 requirements.



**Citation:** Hadi, M.U.; Awais, M.; Raza, M.; Ashraf, M.I.; Song, J. Experimental Demonstration and Performance Enhancement of 5G NR Multiband Radio over Fiber System Using Optimized Digital Predistortion. *Appl. Sci.* **2021**, *11*, 11624. <http://doi.org/10.3390/app112411624>

Academic Editor: Giuseppe Rizzelli

Received: 4 November 2021

Accepted: 1 December 2021

Published: 7 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

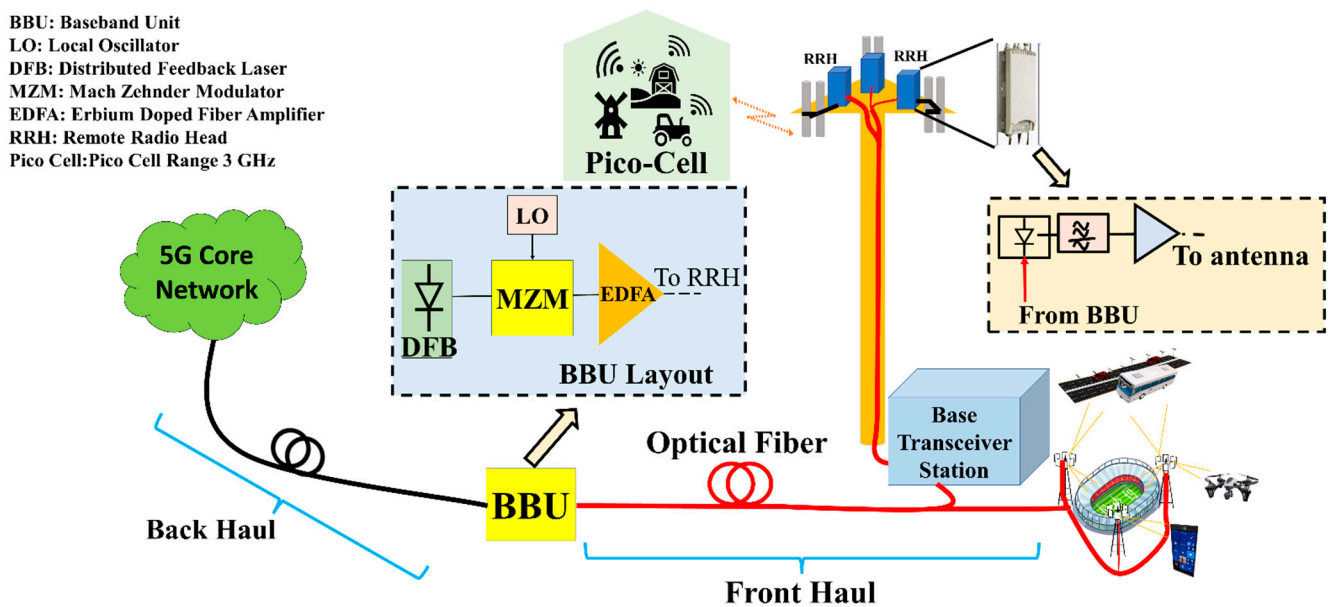


**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** digital predistortion; magnitude selective affine; radio over fiber; neural network; error vector magnitude; adjacent channel power ratio

## 1. Introduction

With recent advances in 5G and beyond, the accelerating growth in base stations (BS) has led to centralization of radio access network (RAN) [1,2], which decreases the capital expenditure as it leads to simplifications in network management [2]. To facilitate C-RAN, a fronthaul (FH) connects base band units (BBU) to remote radio heads (RRH) (see Figure 1). With 5G in roll out stage in most part of the developed world, the microwave photonics-based solutions such as Radio over Fiber (RoF) have a higher significance connecting the BBUs with RRUs [3–5] due to advantages such as cost effectiveness; immunity to electromagnetic disturbance, broader bandwidth and increasing the wireless links reach for all type of distances ranging from short to long.



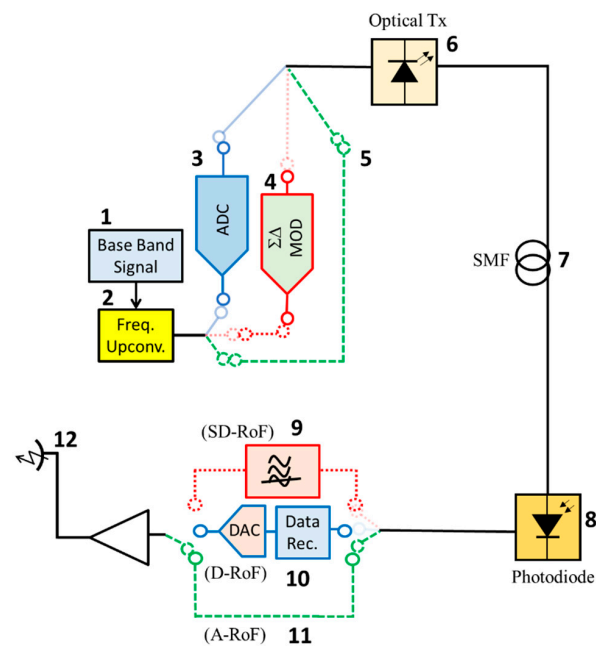
**Figure 1.** Block diagram of C-RAN based 5G application system comprising of Backhaul, Fronthaul and application scenarios such as stadiums, railways, transport, factories, etc.

There have been various versions of RoF such as Analog Radio over Fiber (A-RoF) [6,7], Digital Radio over Fiber (D-RoF) [8,9], Sigma Delta Radio over Fiber (SD-RoF) [10–12] and other variants that have been proposed recently (see Figure 2). Up to an extent, A-RoF links are the simplest, uncomplicated and economical solution, however, they suffer from nonlinearities arising due to signal impairments and devices involved such as laser modules, fibers and photodiodes. The other solutions comprise of utilizing D-RoF or SD-RoF. Considering D-RoF systems, the requirements of conversion from analog to digital (ADC) and digital to analog converters (DAC) makes the process very expensive. In addition, due to high data rate capacity and requirement of high bandwidth, common public radio interface (CPRI) restrictions are faced. CPRI bottleneck can be overcome by exploiting the SD-RoF. Here, ADCs and DACs are not required as it is based on the sigma-delta modulation which utilizes one bit of ADC but the method is complex hence not preferred. Moreover, the quantization noise is high for 1 bit that requires an additional band pass filter (BPF) at the RRH. However, this additional complexity is not alone to be handled for S-DRoF implementation, the addition of BPF results in additional amplitude and phase noise that requires additional solution for removal of these additional nonlinearities [13].

From this it is evident that exploiting other schemes (D-RoF/SD-RoF) is cumbersome. Hence, comparatively, owing to the legacy, infrastructure, and cost effectiveness of the A-RoF systems, they are the better choice for optical fronthaul (OFH). Now then, our best alternative is to counter the nonlinearities of the A-RoF system. Given that we can do this in a simple and practical manner, the RoF technique proves to be extremely advantageous.

Mitigating the nonlinearities of RoF transmission is essential to utilize the system to its best potential and has become an important subject. Within all these different domains, to counter the prevalent issues, a lot of techniques have been exploited. Amongst them, the one that have been utilized extensively are discussed in Section 2 under literature review.

The contribution of nonlinearities from the laser and to some extent, the photodiode part is important as transmission quality decreases and also, the interference with channels nearby is triggered. However, while considering the long-range networks, the nonlinearities due to the combination of fiber chromatic dispersion and laser frequency chirp are usually the main cause of signal impairment [7]. The Orthogonal Frequency Division Modulated (OFDM) signals, such as the emphasized fifth generation (5G) signal, is also liable to these distortions due to high peak-to average power ratio (PAPR).



**Figure 2.** Block diagram of A-RoF, D-RoF and  $\Sigma\Delta$ -RoF downlinks. Blocks active for A-RoF: 1-2-5-6-7-8-11-12. Blocks active for D-RoF: 1-2-3-6-7-8-10-12. Blocks active for  $\Sigma\Delta$ -RoF: 1-2-4-6-7-8-9-12. BB: Base Band, Freq. Upconv.: Frequency Upconverter, Optical Tx: Transmitter,  $\Sigma\Delta$ MOD: Sigma-Delta Modulator, Data Rec.: Data Recovery block.

To the best of the authors knowledge, this paper introduces the nonlinearity and signal impairment compensation for OFH based RoF systems utilizing 5G NR based RoF systems. Following a very detailed literature review on the nonlinearities mitigation in Section 2, the novelties of this article are manifold:

1. Firstly, a multiband 5G NR signals are employed in the experimental testbed to cover enhanced mobile broad band (eMBB) scenarios and small cells for 3 GHz and 20 GHz, respectively.
2. A robust DPD technique utilizing negative feedback iteration is shown. The proposed DPD identification method has a relative lower computational complexity as compared to other learning architectures. In the proposed method of DPD identification, firstly, a DPD signal will be identified followed by the estimation of the DPD parameters.
3. The linearization performed is not limited to our previous proposed methods coming from volterra series but includes the “out of box” approach that includes CPWL and DVR method. In addition, a novel magnitude selective affine (MSA) method is proposed that reduces the overheads of complexity such as multiplications in CPWL architecture, however achieving similar efficiency as of CPWL.
4. In addition, a simple optimized neural network (ONN) based DPD algorithm is proposed that is an upgradation of our previously proposed DPD based method utilizing deep learning to perform linearization of 50 MHz 5G New Radio (NR) based RoF links. The NN DPD method is executed by a different type of training which does not utilize Indirect Learning Architecture (ILA). Initially we emulate the generic RoF link using a RoF NN and then train the proposed DPD ONN using this, by backpropagating the errors.

For the first time, a comparative experimental study has been held where MP, GMP, DVR, CPWL, MSA and ONN are compared for 5G NR multiband signal. The performance is evaluated in terms of Adjacent Channel Power Ratio (ACPR) Error Vector Magnitude (EVM) and complexity in terms of multiplications and coefficients requirements. The summary of this work is shown in the Figure 3 where overall summary of each respective section is shown.

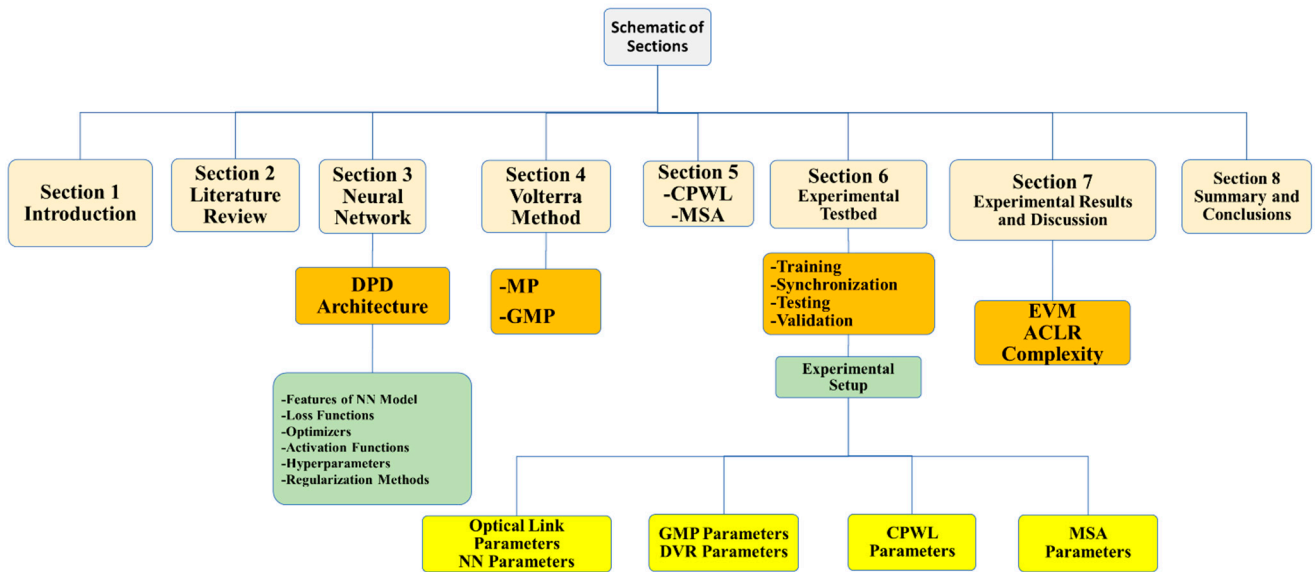


Figure 3. Organization of our work in this paper (Sections 1–8).

## 2. Literature Review

In this section, the linearization methodologies that have been implied for the RoF systems are discussed. The linearization methodologies consist of Electrical, Optical and Machine Learning Methods. Electrical methods are subdivided in to analog and digital methods. Digital methods are subdivided into predistortion and postdistortion methods. Optical methods consist of largely dual wavelength, singular polarization, mixed polarization, etc., while a newer avenue of linearization has been coined that belongs to machine learning. The higher schematic of these methods is summarized in the Figure 4.

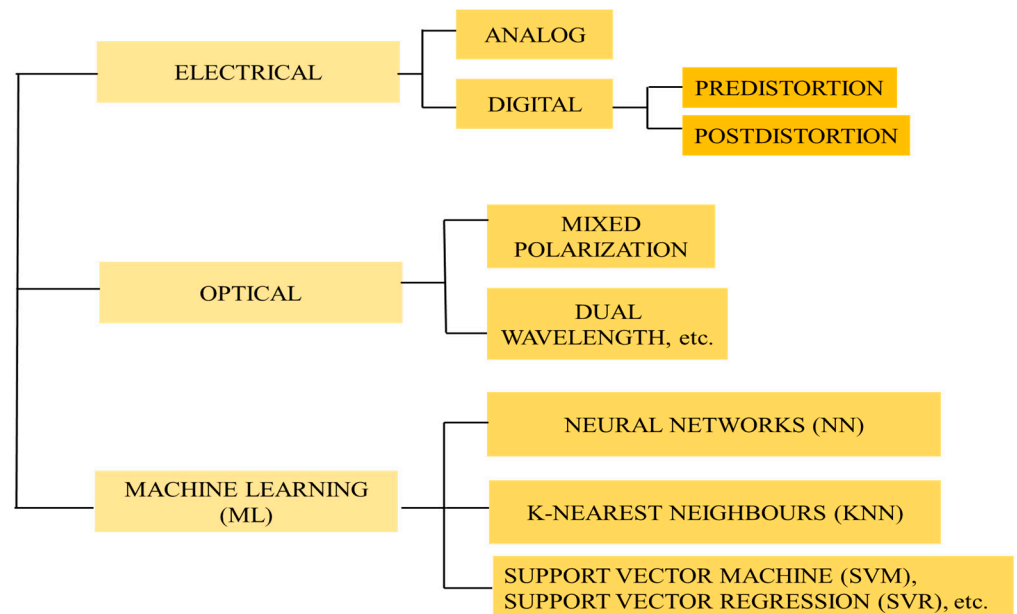


Figure 4. Different types of linearization.

A detailed literature review is enlisted for the mitigation of the impairments of RoF system that are categorized in Table 1. The table summarizes the method employed, type of linearization, category, parameters evaluated, advantages and disadvantages of respective methods.

**Table 1.** Methods exploited to mitigate nonlinearities.

No.	Author	Type	Subcategory	Parameter	Linearization	Pros	Disadvantages
1	Draa et al.; Chen et al. [14,15]	Electrical	Analog Predistortion	IMD3	Complete RoF system (Laser, photodiode, LNA)	Phase maintenance for IMD3 generated components	Perplexing suppression of second order nonlinear distortion for large bandwidth
2	Hass et al. [16]	Optical	Mixed Polarization	Second/third order nonlinear distortion	Complete RoF system	Suppression of second and third order nonlinearities	RoF transmission must be polarization dependent. Compression of linear components to some extent.
3	Zhu et al. [17]	Optical	Dual wavelength linearization (DWL)	Second/third order nonlinear distortion	Complete RoF system	Suppression of second and third order nonlinearities	Wavelength dependent transmission, i.e., suppression of nonlinear components for anti-phased wavelengths only.
4	Ghannouchi et al. [18]	Digital	Digital Predistortion (DPD)	Third order nonlinearities	Power Amplifier	Wideband linearization achievable	DSP required is complicated. Consumption of power is huge.
5	Duan et al. [19]	Digital	DPD	ACPR, EVM	Laser	Added accuracy with less DSP requirements	High complexity and calculation time for digital linearization for higher nonlinearity order and memory depth.
6	Pei et al. [20]	Digital	DPD	ACPR	Laser/RoF	Higher suppression in ACPR by 15 dB	Feedback complexity.
7	Lam et al. [21]	Digital	Digital Post Processing	ACPR, BER	RoF	All order nonlinear distortion components significantly compressed.	High speed digitizer required. Only applicable to uplinks. Deployment of DPP at RRH side, so prototype price passes to customer side, adding complexity to the RRH.
8	Hekkala et al. [22]	Digital	DPD	ACPR, EVM	Laser only	Less complexity and over head	EVM improved by 0.3%. Link length not considered. Intermediate frequency used for DPD. RoF link was not composed of laserfiber-photodiode only. Attenuators and amplifiers were used so perhaps the signal impairments were corrected due to these components. Limited to sinusoid I/P signal.
9	Hadi et al. [23]	Digital	DPD	C/HD2, IIP2, IIP3	Combination of fiber dispersion and laser chirp	Linearizes links up to tens of km	Limited to sinusoidal (single/dual) I/P tones. No ACPR, EVM shown.
10	Vieira et al. [24]	Digital	DPD	EVM	Laser	Utilization of OFDM signal with 5 MHz bandwidth	Only magnitude (AM/AM) linearization shown. RoF link is not generic, contains 10 dB attenuator.
11	Hekkala et al. [25]	Digital	DPD	ACPR, BER	Laser	Utilization of OFDM signal with 12.5 MHz bandwidth	Joint compensation proposed for PA and RoF hence reducing improvement.
12	Mateo et al. [26]	Digital	DPD	EVM, ACPR	RoF system	Utilization of LTE 20 MHz signal	RoF link is not generic; contains signal amplification by LNA and a PA. Unrealistic feedback for 10 km length; requires an uplink for the linearization of downlink.
13	Mateo et al. [27]	Digital	DPD	NMSE, ACPR	RoF system	Utilization of LTE 20 MHz with 16 QAM modulation	PSD of O/P with and without DPD is different after normalization, DPD reduced the bandwidth of the signal. DFB laser has not been pushed to higher RF I/P powers to see the efficacy of the predistorter.
14	Mateo et al. [28]	Digital	DPD	ACPR, EVM	RoF system	Linearization considering ideal and no feedback	Results are attenuation dependent, i.e., results similar to ideal case can be obtained with proper attenuation using different optimization algorithms.
15	Roselli et al. [29]	Electrical	Analog Predistortion	IMD3	Laser	Maintenance of phase for generated IMD3 components	Large scale production difficult as variant predistorter required per RoF transmitter.

Table 1. Cont.

No.	Author	Type	Subcategory	Parameter	Linearization	Pros	Disadvantages
16	R. B. Childs et al. [30]	Electrical	Analog Predistortion	IMD3	Laser	Accurate correction	Tough synthesis of arbitrary transfer functions in the time domain; due to coupling between different paths, simultaneous correction of both, second and third order is not possible.
17	Veiga et al. [31]	Electrical	Analog Predistortion	IMD3	Laser	Phase maintenance is easy	To compensate arbitrary bandwidth limitation in the frequency domain.
18	Hadi et al. [32]	Digital	Direct DPD	EVM, ACPR, IMD	Laser	Only requires transient chirp coefficient, no exhaustive training.	Limited to laser; works only for kilometers of length; complex behavior.
19	Hadi et al. [33]	Digital	DPD for short link	NMSE, ACPR	Laser	Feasible closed loop DPD	Training is time consuming and complex
20	Hadi et al. [34]	Digital	DFB based	NMSE, ACPR, EVM	Laser	Utilization of DVR, GMP, MP	High complexity. I/Q imbalance.
21	Hadi et al. [35]	Digital	ML-NN based	ACPR, EVM	Laser	Learns nonlinearities of model in contrast with Volterra based methods, avoids requirement of explicit designing of distortion causes	Limited to LTE framework. Upgradation to 5G ambiguous.

Along with the numerous techniques listed above, mitigation of the impairments of RoF system have also been studied previously [24–35]. A feedforward scheme has also been analyzed [26,27] but is complex and to counter these problems, optical methods such as dual parallel modulation [28], mixed polarization [29,30], etc. have been presented. In addition to the nonlinearities of optical fiber/modulators, RoF is liable to other issues such as amplified emission noise and phase noise [32,33] that further limit performance of RoF increasing proportionally with RF carrier frequency increase.

Moreover, Digital Signal Processing (DSP) methodologies have been exploited as well owing to their adaptive flexibility. A lot of such techniques have been studied to counter the phase noise issue [36–38], of optical modulator [39,40]. Digital pre-distortion is presented for laser resonance, improving results in [41]. Apart from DSP approaches, Digitized RoF (DRoF) has also been utilized for nonlinearity mitigation [42–45].

Another important concept proposed, Sigma Delta RoF (SD-RoF) is a method involving the plus points of the DRoF in addition to its own implementation [46,47]. This is especially advantageous when utilizing optical transmitters having low linearity. On the other hand, considering Machine Learning as a domain of nonlinearity compensation, a lot of models and methodologies have been proposed such as the KNN/SVM [48]; SVM [49,50]; k-means [51]; Convolution Neural Network (CNN) and Binary-CNN (BCNN) [52,53] and Fully-CNN (FCNN) [54–61].

### 3. Neural Network Based DPD Architecture

Neural Networks (NN) are sophisticated networks, similar to the suggested NN based DPD model and they require extensive training data. This model is then cascaded with the RoF link, but its output is not known. However, in the case of an RoF link, the output is known hence, we form an RoF NN model and train it to mirror the original RoF link. Once we have formulated this RoF NN, we can now backpropagate through it and update the parameters in the suggested NN DPD model.

Supposing that the considered RoF link has a transfer function  $H(n)$  and an output signal  $y(n)$ ; and a baseband signal  $x(n)$  must be sent through it, DPD aims to calculate the inverse transfer function of this RoF link denoted by  $\hat{I}^{-1}$ , whose output will then be denoted by  $\hat{x}(n)$ .

This can be expressed by:

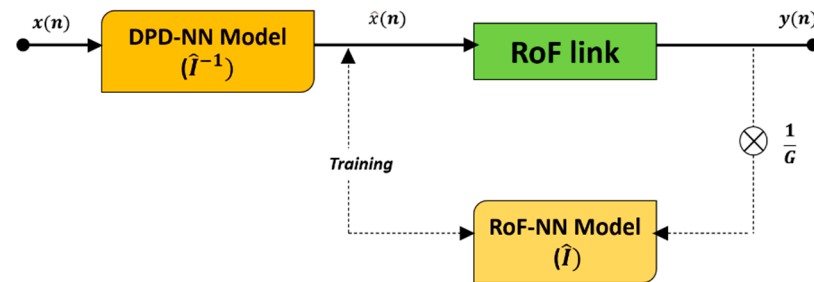
$$\hat{x}(n) = \hat{I}^{-1}(x(n)) \quad (1)$$

While,

$$y(n) = Gx(n) = I(\hat{x}(n)) \quad (2)$$

where  $G$  is the gain. The NN here calculates  $\hat{I}^{-1}$  utilized for predistortion. A direct training cannot be implemented for construction of the NN for DPD as the ideal  $\hat{x}(n)$  is not known.

The option analyzed for performance assessment is illustrated in Figure 5 [1].



**Figure 5.** Block diagram of RoF system utilizing RoF NN model and NN based DPD system. The RoF NN model  $\hat{I}$  is obtained upon transmission of I/P and O/P through the generic RoF link. Once achieved, we train  $\hat{I}^{-1}$  by backpropagating error through  $\hat{I}$ . Then, linearization is performed by connecting this DPD-RoF model to generic RoF link. Predistortion is held in a digital baseband eliminating DACs and ADCs.

Initially, the RoF link is emulated by the second NN. The generic RoF link has  $\hat{x}(n)$  as input and  $\frac{y(n)}{G}$  as output data, considering which, training is executed with the regression based NN model for them. This results in learning of the NN and now an approximate transfer function  $\hat{I}$  can be identified by it. Upon the formation of the RoF NN model, the model weights are fixed after which we connect it with the NN DPD model. Now, we use the original input, i.e.,  $x(n)$  and output as training data to calculate error using a loss function. We then backpropagate it via  $\hat{I}$  to train  $\hat{I}^{-1}$ .

### 3.1. Features of the NN Model

This section will discuss all the features of the Neural networks that are essential to be used in DPD based RoF system. The summary of the topics discussed in this section is given as follows. Firstly, Section 3.1.1 discusses the loss functions while Section 3.1.2 talks about optimizers. Section 3.1.3 talks about activation functions and Section 3.1.4 talks about hyperparameters. Section 3.1.5 discusses the regularization methods and Section 3.1.6 talks about are characteristics of NN.

The NN are layered structures that have been inspired from the human nervous system and are an adequate choice due to their powerful accuracy at approximating any nonlinear function and learning the relationship between their inputs and outputs. The basic working involves an input signal provided at the input layer, processing which takes place in the hidden layer/layers and the output is produced by the last (output layer). Each layer represents a cluster of neurons which are not connected to one another within the layer but are connected to the neurons of the next layer.

#### 3.1.1. Loss Function

Prior to optimizing, it is essential that, the error is estimated for the current state of the model. For this, selecting an error function, normally known as loss function is necessary as it helps calculate the loss of the model. After this, the weights of the model are updated which minimizes this loss function for further evaluations. Loss functions broadly classify into:

- i. Regression Loss.
- ii. Classification Loss.

### Regression Loss Functions

Regression loss functions are divided into following:

#### Mean Squared Error (MSE)

MSE calculates the average of the squared differences between actual and predicted value. Outcomes are always positive values irrespective of the sign of the actual and predicted values. The minimum value it can result in is 0.0. Note that, larger differences cause more error due to squaring of the value, penalizing the model for them.

#### Mean Squared Logarithmic Error (MSLE)

As the name suggests, here, the natural logarithm of each of the predicted values is calculated and then the MSE is calculated. MSLE relaxes the punishing effect of large differences in large, predicted values.

#### Mean Absolute Error (MAE)

The MAE is more powerful. It calculates the average of the absolute difference between the predicted and actual values.

### Classification Loss Functions:

Classification loss functions are divided into following:

#### Binary Cross-Entropy

This function is used when our target values belong to the set  $\{0, 1\}$ . A score which summarizes the average difference between the predicted and actual probability distributions is calculated and is then minimized. The least score that can be achieved by this function is 0.

#### Hinge Loss

This function is used when our target values belong to the set  $\{-1, 1\}$ . This function basically promotes correctness of the sign of values and will allot more error for differences in signs between predicted values and actual values.

#### Squared Hinge Loss

This is a function obtained by modifications in the hinge loss function. It merely calculates the square of the hinge loss score. This causes the error function to smoothen making it easy to work with numerically. It smoothenes the surface of the error function and makes it easy to work with numerically. As stated in the case of hinge loss, values must belong to the set  $\{-1, 1\}$ .

#### Multi-Class Cross-Entropy

Hinge loss is considered for binary classification similarly, multi-class cross entropy loss is considered for multi-class classification. Every class is given a distinct integer value. A score for all classes is summarized by calculating the average difference between the predicted and actual probability distributions. The minimum score that can be obtained is 0.

### 3.1.2. Optimizers

The process of optimizing loss functions improves the function of the model. The procedure to minimize/maximize a mathematical function is called optimization and utilizes optimizers. Optimizers work by changing the parameters of NNs such as the learning rates or weights. The types of optimizers are:

#### Gradient Descent (GD)

The GD algorithm works on the principle that the direction opposite to that of the calculated gradient will locate the lower point/surface of the considered function, i.e., the potential minima. It then repeatedly takes steps in that direction with every iteration. In the case of GD, we require large memory as we calculate the gradient over the entire dataset after which the updates are performed meaning that, for huge data, it may take years to converge.



### Stochastic Gradient Descent (SGD)

To make up for the problem faced by the GD algorithm, SGD was developed. The SGD algorithm works by computing derivatives by considering only one point at once. This solves the large memory requirement problem; however, time is still an issue as one epoch of SGD takes more time when compared to the GD algorithm.

### Mini Batch Stochastic Gradient Descent (MB-SGD)

This algorithm was developed to overcome time complexity problems. MB-SGD as the name suggests, works by considering a small batch of points or a small subset of the whole data at once to compute the derivatives. It takes longer to converge than GD does and additionally, its weight updates are noisy.

### SGD with Momentum

SGD with momentum (momentum for short) calculates another parameter, momentum in addition to the gradient in every step. This momentum is the cumulative movement from all previous steps, overcoming the issue of noisy weight updates in MB-SGD by denoising gradients. It moves faster owing to the accumulated momentum which helps it prevent local minima and plateau regions. In addition, the algorithm Nesterov Accelerated Gradient (NAG) is similar to momentum except, it is a future considering method. The cost function is evaluated using future parameters rather than the current ones to prevent the chance of skipping the minima in case of high momentum value.

### Adaptive Gradient Descent (AdaGrad)

AdaGrad, works by accumulating the sum of gradient 'squared'. As the name suggests the learning rate is adaptive and doesn't need to be tuned manually. The adaptive learning rate property helps escape typical complexities of non-convex functions such as saddle points as they take a straight path compared to methods such as GD, SGD, MB-SGD and even momentum. There are chances that these methods become stuck at saddle points and never converge to minima. Unfortunately, AdaGrad is very slow as the sum it accumulates continues to grow resulting in the learning rate to become very small, this consequently leads to vanishing of the gradient.

### Adaptive Delta (AdaDelta)

As we know, AdaGrad is a slow converging method due to its large value of the accumulated sum of gradient squared. AdaDelta was developed to resolve this by modifying certain aspects of AdaGrad. Here, rather than storing all the past gradients inefficiently, we recursively define the sum of gradients as a decaying average of all the past squared gradients. This allows AdaDelta to continue learning even after multiple updates.

### RMSprop (Root Mean Square Propagation)

RMSprop, similar to AdaDelta, was developed to counter AdaGrad's problem of vanishing learning rates. RMSProp fixes the issue by adding a decay factor which emphasizes on recent gradients neglecting the older ones. An exponentially decaying average of squared gradients is used to divide the learning rate. AdaGrad might keep up with RMSProp initially, but the sum of gradients squared for AdaGrad accumulate and become huge, eventually AdaGrad practically stops moving, in contrast with RMSProp, smartly stored the squares due to the decay rate. This makes it faster than AdaGrad and it impedes search in direction of oscillations.

### Adaptive Moment Estimation (ADAM)

The Adam algorithm combines the features of both Momentum and RMSProp meaning that, along with accumulating the exponentially decaying average of the previous squared gradients, it also holds exponentially decaying average of previous gradients as well. Adam is an efficient optimizer due to the component of speed and the capability to

adapt gradients acquired from momentum and RMSProp, respectively. This is also the reason it is the most used optimizer.

### Comparison of Optimizers

There are two metrics to determine the efficacy of an optimizer, speed of convergence to the global minimum and generalization (the model's performance on new data). Performance of the optimizers and hence their choice is also dependent on the type of data provided, based on that, two types of functions exist- convex and non-convex (Figure 4). In case of the latter, optimizers must be carefully chosen as it has multiple hurdles where our algorithm might become stuck, some of which we will discuss ahead. The performance of mentioned optimizers on different obstacles faced on non-convex functions are discussed below (see Table 2).

**Table 2.** Comparison of Optimizers.

Obstacle Type	Brief Definition	Convergence Speed (Descending Order of Converging Speed)
Saddle point	Point on the surface of a function which is not a local extremum of the function but the slopes in orthogonal directions are all zero.	ADAM, RMSprop, AdaDelta, AdaGrad, Momentum.
Plateau region	Plateau of a function is a part of its domain where the function has constant value.	ADAM, Momentum
Local minima	Minimum of a function other than the global minimum (extremum).	ADAM, RMSprop, Momentum, AdaGrad

From the table, we see that ADAM has performed comparatively better than others, especially in terms of convergence speed which is why it is the most used and most efficient optimizer.

### 3.1.3. Activation Functions

The activation or capability of response of a node is not pre-defined. This is deduced with the help of an Activation. It does this by building a relationship between various weights and biases affecting the node and applying the relationship to the node as a function, hence determining responses. It also helps it learn the complex patterns of data. They transform the incoming signals in the node into an output signal which will be used in the next layer of the network or will be the output. The different types of activation functions are summarized in the table below (Table 3).

**Table 3.** Activation Functions.

Activation Function	Mathematical Representation	Brief Description
Binary Step	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	only has values 1 or 0 as output.
Linear	$y(x) = x$	Linear over entire range $(-\infty, \infty)$ .
Sigmoid/Logistic	$y(x) = \frac{1}{1+e^{-x}}$	Range (0, 1)
ReLu (Rectified Linear Unit)	$f(x) = \max(0, x)$	Range [0 to $\infty$ )
Leaky ReLu	$f(x) = \max(0.01 * x, x)$	Range $(-\infty, \infty)$ .
Softmax	$\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$	Range (0, 1) and sum of O/Ps = 1
tanh/Hyperbolic	$y(x) = \frac{\sinh x}{\cosh x}$	Range $(-1, 1)$

When considering ReLu, the possibility of a vanishing gradient is lower, in contrast with the gradient of sigmoids, which grow smaller with an increase in the absolute value of x.

ReLU has a constant gradient, helping it learn faster making it the better choice, provided we can work in the positive range of input values.

### 3.1.4. Hyperparameters

The implementation of algorithms on data has certain pre-requisites. There are several hyperparameters that must be defined for the functionality of our network. These are overall functionality defining components and are not to be confused with the internal model parameters. This section discusses the hyperparameters definitions in Table 4.

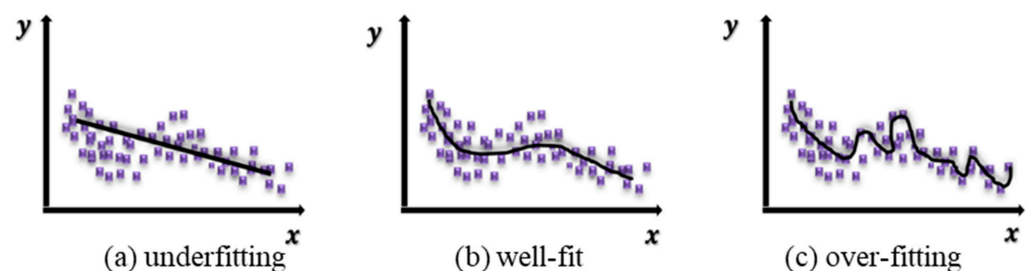
**Table 4.** Definition of Hyperparameters.

Hyperparameter	Definition	Explanation
Batch size	Number of samples processed while updating the internal parameters of a model.	GD = Size of Training Set; SGD = 1; B-SGD = 1 < Batch Size < Size of Training Set
Epoch	Number of complete passes through the samples being trained.	For MB-SGD if batch size = 25, then number of complete passes through these 25 samples is number of epochs.
Regularization	Penalizes weight matrices to resolve overfitting and underfitting	Methods-L1, L2, Dropout.

Note that there are no pre-defined fixed values for these hyperparameters. There are some conventionally used values, but they are not mandatory and attaining a value suitable to a function is a trial-and-error method.

### 3.1.5. Different Methods of Regularization

All datasets are divided into two sections, namely, the training data and the testing data. Sometimes, we come upon conditions where models do not result in good performance on testing data despite performing well on training data. These situations arise due to underfitting and overfitting in Figure 6. Regularization is the procedure in which weight matrices of nodes are penalized, hence overcoming these problems.



**Figure 6.** Graphical representation of the training curve (black) learning the training data (purple, scattered) in three labelled cases.

### L2 and L1 Regularization

L1 and L2 are simple and widely known types. A regularization term is considered while updating the cost function in this method. This regularization term causes a reduction in the value of the weight matrices as they result in a simpler model, significantly reducing the problem of overfitting.

In L2, we have:

$$Loss = error + \lambda \sum_{i=1}^N |w_i^2| \tag{3}$$

where  $\lambda$  is the regularization parameter (which is updated for more precise outcomes) and  $w$  is weight. The L2 method is referred to as a weight decay procedure because it compels the weights to decay towards zero although never exactly becoming zero.

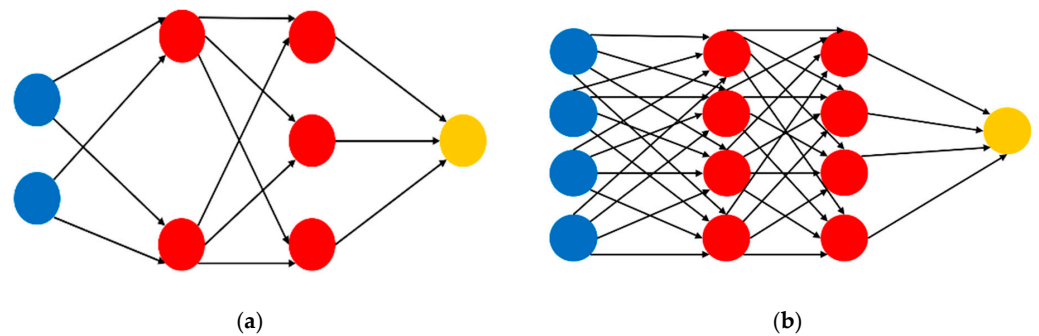
In L1, we have:

$$Loss = error + \lambda \sum_{i=1}^N |w_i| \tag{4}$$

As observed in Equation (4) here, the absolute weight value is considered. Contrary to the L2 method, the weights here can become zero which is beneficial for compression of models.

### Dropout

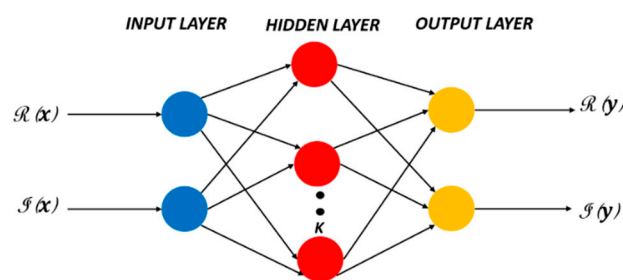
Dropout is the most used type of regularization due to its efficient results. It performs several iterations on the NN and at every iteration, random nodes are selected and removed along with all of their input and output signals/connections as seen in Figure 7. This results in selection of distinct nodes and hence distinct outputs per iteration. More randomness is expressed by splitting one network into its subsets leading to better performance when compared to a single dense model. The method can be implemented on input as well as hidden layers making it an adequate choice for large NNs.



**Figure 7.** Schematic diagram of complete densely connected in (a) NN and (b) representation of network upon selection and removal of random nodes, after any one iteration.

### 3.1.6. Characteristics of the NN Models

Based on the discussion made so far, we have built a foundation of all components of a Neural Network and established that the two neural network models we require here are, first, the DPD NN, used to predistort our original RoF link and second, the emulated RoF NN, which is essential for training this DPD NN. The NN employed here contains  $N$  hidden-layers,  $K$  neurons per hidden-layer and is a feedforward fully connected network. The symbolic structure of the employed NN is shown below in Figure 8. The NN has two inputs and two outputs, in view of the complex nature of baseband signals representing both the real and imaginary parts of the signal. The ReLu activation function has been utilized for at least one of the multiple hidden layers (owing to its lower complexity).



**Figure 8.** Schematic diagram of the feedforward fully connected NN structure used, with one of the  $N$  number of hidden layers having  $K$  neurons per hidden layer.  $\mathcal{R}(x)$  and  $\mathcal{R}(y)$  are real I/P and O/P;  $\mathcal{I}(x)$  and  $\mathcal{I}(y)$  are imaginary I/P and O/P, respectively.

We represent the O/P for the first hidden-layer as follows:

$$l_1(n) = f\left(W_1 \begin{bmatrix} \mathbb{R}(x) \\ \mathbb{I}(x) \end{bmatrix} + b_1\right) \tag{5}$$

where,  $l_1$  expresses the first hidden O/P layer,  $f$  is the activation function (nonlinear) and  $W_1$  is the weight and  $b_1$  is the bias for the first output layer in the network.

We represent the general output for the  $i$ th layer as:

$$l_i(n) = f(W_i h_{i-1}(n) + b_i) \tag{6}$$

where,  $i \in \mathbb{N} : 2 \leq i \leq N$ .

Then, the final output after  $N$  hidden layers will be:

$$\hat{x}(n) = W_{N+1} h_N + b_{N+1} \tag{7}$$

### Training Algorithm

The algorithm used to train the NN DPD model is given below. The algorithm utilizes MSE as the loss function, the optimizer used is ADAM and backpropagation is used for updating the weights. The process is repeated for a certain number of iterations ( $Z$ ) to refine performance.

As mentioned earlier, we train our emulated RoF NN model using the I/P and O/P of the original RoF link and once this model is acquired, we connect the NN DPD model to it. Then, upon convergence of this training, we connect the actual RoF link with this DPD NN and proceed with its predistortion (see pseudocode in Algorithm 1).

---

#### Algorithm 1. DPD training

---

```

 $\hat{x}(n) \leftarrow x(n)$ 
for  $i \leq Z$  do
 $y(n) \leftarrow I(\hat{x}(n))$ : //RoF-Transmission
 $\hat{I} \leftarrow \text{Train on } \hat{x}(n), \frac{y(n)}{G}$  //Updating RoF NN
//Freeze NN weights of  $\hat{I}$ 
 $\hat{I} \leftarrow \text{Train on } x(n)$ . //Use  $\hat{I}^{-1}(\hat{I}(x(n)))$ 
 $\hat{x}(n) \leftarrow \hat{I}^{-1}(x(n))$ : //Predistort
end for

```

---

### 4. Comparison with Volterra Method

It will be interesting to see the comparison of NN DPD methodology with conventional GMP method that has been validated in [1,6,34]. MP/GMP are the most viable solutions that have been used for DPD.

Volterra series-based models are very commonly used. Upon conversion of signals from electrical to optical domain and vice-versa, certain memory effects are introduced. To take these effects into account, Volterra Series is considered.

We represent Volterra series as:

$$\check{y}(n) = \sum_{k=1}^{+\infty} \int_0^{+\infty} \dots \int_0^{+\infty} h_k(\tau_1, \dots, \tau_k) \prod_{r=1}^k \check{x}(t - \tau_r) d\tau_r \tag{8}$$

where  $\check{y}(n)$  represents RF out put,  $\check{x}(n)$  represents RF input signal and  $h_m$  is the  $m$ th order Volterra kernel. RF signal is down converted to baseband, and we obtain the envelope through a low pass filter. Then, with input  $\check{x}(n)$  baseband image in discrete time is:

$$y(n) = \sum_{k=0}^K \sum_{q_1=0}^{Q-1} \sum_{q_2=l_1}^{Q-1} \dots \sum_{q_{2k+1}}^{Q-1} h_{2k+1}(l_1, \dots, l_{2k+1}) \prod_{r=1}^{k+1} \check{x}(t - l_j) \prod_{r=k+2}^{2k+1} x^*(n - l_j) \tag{9}$$

where,  $K$  is the nonlinearity order and  $Q$  is the memory depth.

#### 4.1. Memory Polynomial Model (MP)

The Memory polynomial (MP) model is capable of resolving both memory effects and nonlinearities simultaneously. It is also referred to as diagonal Volterra model as all its diagonal terms are non-zero. It is the middle ground between a memoryless model and a full Volterra model due to existence of diagonal memory (non-diagonal terms are zero). The MP model is generally used as it is less complex compared to Volterra series. It can also emulate the nonlinear behavior of Power Amplifiers to an extent which is why it has been used to model Power Amplifiers previously. The equation is as follows [24–35]:

$$y(n) = \sum_{k=0}^{K-1} \sum_{q=0}^{Q-1} C_{kq} x(n-l) |x(n-l)|^k \tag{10}$$

where  $K$  is non-linearity order,  $Q$  is the memory depth,  $y(n)$  is input sequence of the predistorter,  $x(n)$  represents baseband input and  $c_{kq}$  is the model coefficient.

#### 4.2. Generalized Memory Polynomial (GMP)

GMP has been productively utilized for DPD linearization of Power Amplifiers. Here, we will now use it to aid linearization of RoF links using the Digital predistortion method. If we observe the equation below, we see that, the GMP model, in contrast with MP, holds both, memory of diagonal terms along with crossing terms as well, which is why it outperforms MP and in the coming sections, we will evaluate comparisons with GMP only having established it is superior to MP [1,24–35]:

$$\begin{aligned} y(n) &= \sum_{k=0}^{K_a-1} \sum_{q=0}^{Q_a-1} c_{kq} x(n-q) |x(n-q)|^k \\ &+ \sum_{k=1}^{K_b} \sum_{q=0}^{Q_b-1} \sum_{r=1}^{R_b} d_{kqr} x(n-q) |x(n-q-r)|^k \\ &+ \sum_{k=1}^{K_c} \sum_{q=0}^{Q_c-1} \sum_{r=1}^{R_c} e_{kqr} x(n-q) |x(n-q+r)|^k \end{aligned} \tag{11}$$

where  $x(n)$  is the DPD input and  $y(n)$  is the DPD output.

The complex coefficients  $c_{kq}$ ,  $d_{kqr}$  and  $e_{kqr}$  denote signal and the envelope; signal and lagging envelope and signal and leading envelope, respectively.  $K_a$ ,  $K_b$ ,  $K_c$  are the maximum nonlinearity orders,  $Q_a$ ,  $Q_b$ ,  $Q_c$  are the memory depths.  $q$ ,  $r$  represents the indices of the memory  $k$  is nonlinearity index and  $R_c$  and  $R_b$  exhibit the leading and lagging delay tap lengths, respectively. Since, in [62–71], it was established that GMP is better as compared to MP, therefore for simplicity, evaluation with GMP is included in this paper.

### 5. Models Based on Canonical Piece Wise Linearization

The DPD linearization has been achieved with Volterra based methods as discussed in Section 3. However, “out of the box” approach which can achieve better performance is always interesting addition to the topic. Recently, it was shown in [26–28] that out of all the possible architectures, CPWL method outperforms the other models such as memory polynomial (MP) and generalized memory polynomial (GMP). CPWL is an obvious choice due to the performance enhancement that it brings, however it has a lot of complexity and overheads. The CPWL model can be expressed as [28]:

$$\begin{aligned} y(n) &= \sum_{m=0}^M \sum_{k=0}^K \sum_{l=1}^L c_{m,k,l}^{(1)} \left| |x(n-k)|^2 - \beta_l \right| x(n-m-k) \\ &+ \sum_{m=1}^M \sum_{k=0}^K \sum_{l=1}^L c_{m,k,l}^{(2)} \left| |x(n-k)|^2 - \beta_l \right| x^2(n-k) x^*(n-m-k) \\ &+ \sum_{m=1}^M \sum_{k=0}^K \sum_{l=1}^L c_{m,k,l}^{(3)} \left| |x(n-k)|^2 - \beta_l \right| x(n-k) |x(n-m-k)|^2 \\ &+ \sum_{m=1}^M \sum_{k=0}^K \sum_{l=1}^L c_{m,k,l}^{(4)} \left| |x(n-k)|^2 - \beta_l \right| x^*(n-k) x^2(n-m-k) \end{aligned} \tag{12}$$

Here, the baseband input is represented by  $x(n)$  and the output baseband signal is represented by  $y(n)$ ,  $K$  is the FIR length filter,  $M$  is represented by memory depth,  $L$  is the number of partitions in the CPWL,  $\beta_l$  shows the threshold while  $c_{m,k,l}^{(1)}, c_{m,k,l}^{(2)}, c_{m,k,l}^{(3)}, c_{m,k,l}^{(4)}$  presents the model coefficients.

In the Equation (1), there are many orders of multiplications and additions which will add a lot of overhead in terms of complexity and utilization if hardware resources during DPD implementation, the most important of them is dedicated hardware adders and multipliers.

### 5.1. Decomposed Vector Rotation (DVR) Model

It is derived from the canonical piecewise-linear (CPWL) function and is a modified form of it. The model's nonlinear function is constructed from piecewise vector decomposition and is entirely different to the one used in previously discussed Volterra series. The model is more flexible and can perform better even with a comparatively small coefficient number than the conventional models. DVR is expressed as:

$$\begin{aligned}
 y(n) = & \sum_{i=0}^{Q_{DVR}} a_i |x(n-i)| + \sum_{k=1}^{K_{DVR}} \sum_{i=0}^{Q_{DVR}} a_{ki,1} ||x(n-i) - \beta_k| e^{j\theta(n-1)} \\
 + & \sum_{k=1}^{K_{DVR}} \sum_{i=0}^{Q_{DVR}} a_{ki,21} ||x(n-i) - \beta_k| e^{j\theta(n-1)} \cdot |x(n)| + \sum_{k=1}^{K_{DVR}} \sum_{i=0}^{Q_{DVR}} a_{ki,22} ||x(n-i) - \beta_k| \cdot x(n) \\
 + & \sum_{k=1}^{K_{DVR}} \sum_{i=0}^{Q_{DVR}} a_{ki,23} ||x(n-i) - \beta_k| \cdot x(n-i) + \sum_{k=1}^{K_{DVR}} \sum_{i=0}^{Q_{DVR}} a_{ki,24} ||x(n-i) - \beta_k| \cdot x(n-i)
 \end{aligned} \tag{13}$$

where  $x(n)$  is I/P and  $y(n)$  is O/P.

$K_{DVR}$  is elements in the partition and  $Q_{DVR}$  is the memory depth.  $\beta_k$  represents the breakpoint.

### 5.2. Magnitude Selective Affine (MSA) Based Linearization

Extending the previous work reported for DVR and CPWL, the objective of this work is to further reduce the overheads and complexity of the CPWL method by proposing a magnitude selective affine (MSA) function-based model. The advantage of this method is that it requires only a single linear operation for the selected zone leading to a lower complexity and simpler structure.

In order to optimize the operations, the coefficients in the zone that have similar magnitude can be coupled together [72,73]. The comparison between threshold function and magnitude of input samples can select which zone the samples will be falling and which affine functions can be utilized. This simplification will result in the reduction of the CPWL complex operation. Therefore, we can rewrite the first term of CPWL function in Equation (13) in a simplified way in terms of MSA function that can be expressed as:

$$\begin{aligned}
 & \sum_{m=0}^M \sum_{k=0}^K \sum_{l=1}^L c_{m,k,l}^{(1)} \left| |x(n-k)|^2 - \beta_l \right| x(n-m-k) \\
 = & \sum_{m=0}^M \sum_{k=0}^K x(n-m-k) \left( \sum_{l=1}^L c_{m,k,l}^{(1)} \left| |x(n-k)|^2 - \beta_l \right| \right) \\
 = & \sum_{m=0}^M \sum_{k=0}^K u_{m,k}^{(1)}(n-k) x(n-m-k)
 \end{aligned} \tag{14}$$

$$\begin{aligned}
 u_{m,k}^{(1)}(n-k) = & \sum_{l=1}^L c_{m,k,l}^{(1)} \left| |x(n-k)|^2 - \beta_l \right| \\
 = & \begin{cases} A_{m,k,1}^{(1)} |x(n-k)|^2 + B_{m,k,1}^{(1)}, & 0 \leq |x(n-k)|^2 < \beta_1 \\ \vdots \\ A_{m,k,L}^{(1)} |x(n-k)|^2 + B_{m,k,L}^{(1)}, & \beta_{L-1} \leq |x(n-k)|^2 < \beta_L \end{cases}
 \end{aligned} \tag{15}$$

Here in Equation (15),  $A_{m,k,l}^{(1)}$  and  $B_{m,k,l}^{(1)}$  are the linear model coefficients defined for each zone of the MSA function  $u_{m,k}^{(1)}(\cdot)$ . The example of hardware implementation is shown in Figure 9. The simplification shown in Equation (15) leads to this realization that input power terms without any magnitude are compared with the thresholds for the offset and linear gain selection for the MSA function. This leads to removal of square root calculation operation. The overall model of Equation (12) in terms of MSA function can be written as:

$$\begin{aligned}
 y(n) &= \sum_{m=0}^M \sum_{k=0}^K u_{m,k}^{(1)}(n-k)x(n-m-k) \\
 &+ \sum_{m=1}^M \sum_{k=0}^K u_{m,k}^{(2)}(n-k)x^2(n-k)x^*(n-m-k) \\
 &+ \sum_{m=1}^M \sum_{k=0}^K u_{m,k}^{(3)}(n-k)x(n-k)|x(n-m-k)|^2 \\
 &+ \sum_{m=1}^M \sum_{k=0}^K u_{m,k}^{(4)}(n-k)x^*(n-k)x^2(n-m-k)
 \end{aligned} \tag{16}$$

$$\begin{aligned}
 u_{m,k}^{(i)}(n-k) &= \sum_{l=1}^L c_{m,k,l}^{(i)} \left| |x(n-k)|^2 - \beta_l \right| \\
 &= \begin{cases} A_{m,k,1}^{(i)} |x(n-k)|^2 + B_{m,k,1}^{(i)}, & 0 \leq |x(n-k)|^2 < \beta_1 \\ \vdots \\ A_{m,k,L}^{(i)} |x(n-k)|^2 + B_{m,k,L}^{(i)}, & \beta_{L-1} \leq |x(n-k)|^2 < \beta_L \end{cases}
 \end{aligned} \tag{17}$$

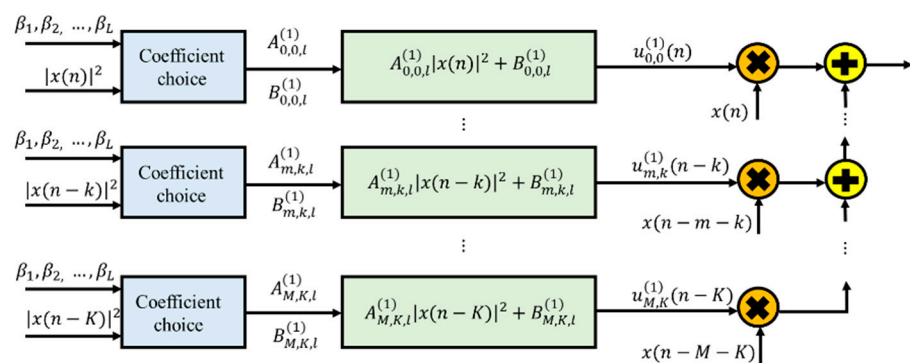


Figure 9. Hardware implementation example using MSA.

### 5.3. Negative Feedback Iteration based Modelling Approach

For the computation of DPD model coefficients, as shown in the Figure 10 with negative feedback iteration technique, the predistortion coefficients are calculated in the training phase.  $z(n)$  is the input of the predistorter, coming through the output of the RoF  $y(n)$  here,  $z(n) = \frac{y(n)}{G}$  and  $G$  is the link's gain. There are two main steps in this technique. The first step is to establish a negative feedback iteration to obtain an input signal that can be regarded as a DPD signal followed by the second step which consists of calculating the DPD model parameters. The shared feedback path is adopted to observe the output information in both frequency bands. In this case, the negative feedback iteration is performed on the lower band and upper band by turns.



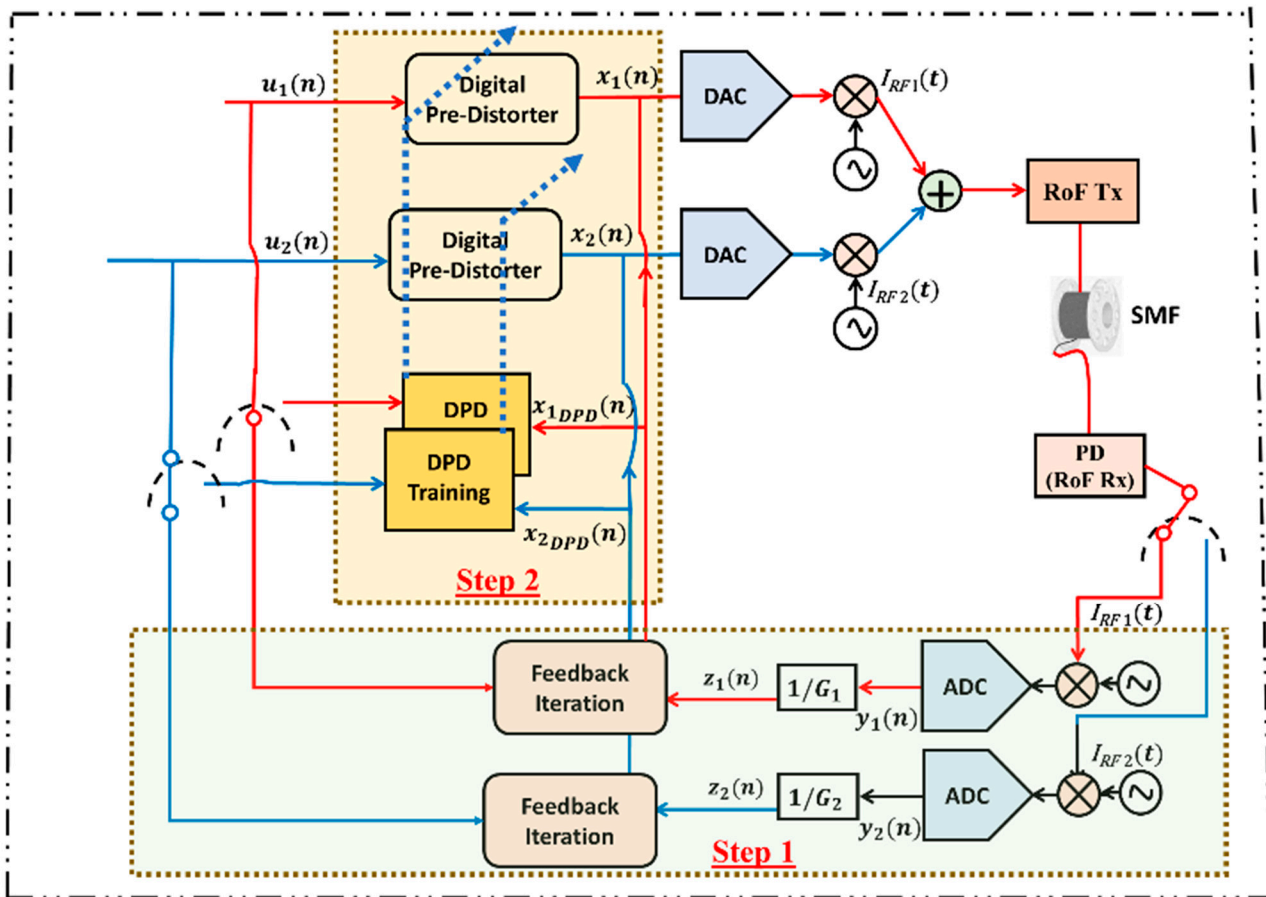


Figure 10. Block diagram representing ILA architecture for Multi Band DPD utilizing Generalized Memory Polynomial (GMP), Canonical Piecewise Linearization (CPWL) and Magnitude Selective Affine (MSA) methodology.

#### 5.4. Estimation Algorithm

Estimation is initiated by the collection of coefficients  $c_{kq}$ ,  $d_{kqr}$  and  $e_{kqr}$ , into a  $R \times 1$  vector  $v$ . here, the total number of coefficients is represented by  $R$  and  $v$  is related with a signal whose time is sampled over the same period. Coefficients  $c_{21}$  are associated to the signal  $x(n-1)|x(n-1)|^2$ .  $Z$  a  $N \times R$  matrix represents the collection of all such vectors. The predistorter training block output upon convergence will be  $z_p(n) = x(n)$  hence,  $z(n) = u(n)$ .

For  $N$  number of samples, output is:

$$z_p = Zv \tag{18}$$

where  $Z = [Z(1), \dots, Z(N)]^T$  and  $z_p = [z_p(1), \dots, z_p(N)]^T$  and  $v$  is as mentioned earlier, a  $R \times 1$  vector containing set of coefficients  $c_{kq}$ ,  $d_{kqr}$  and  $e_{kqr}$ .

The LS solution is the solution for the equation represented by:

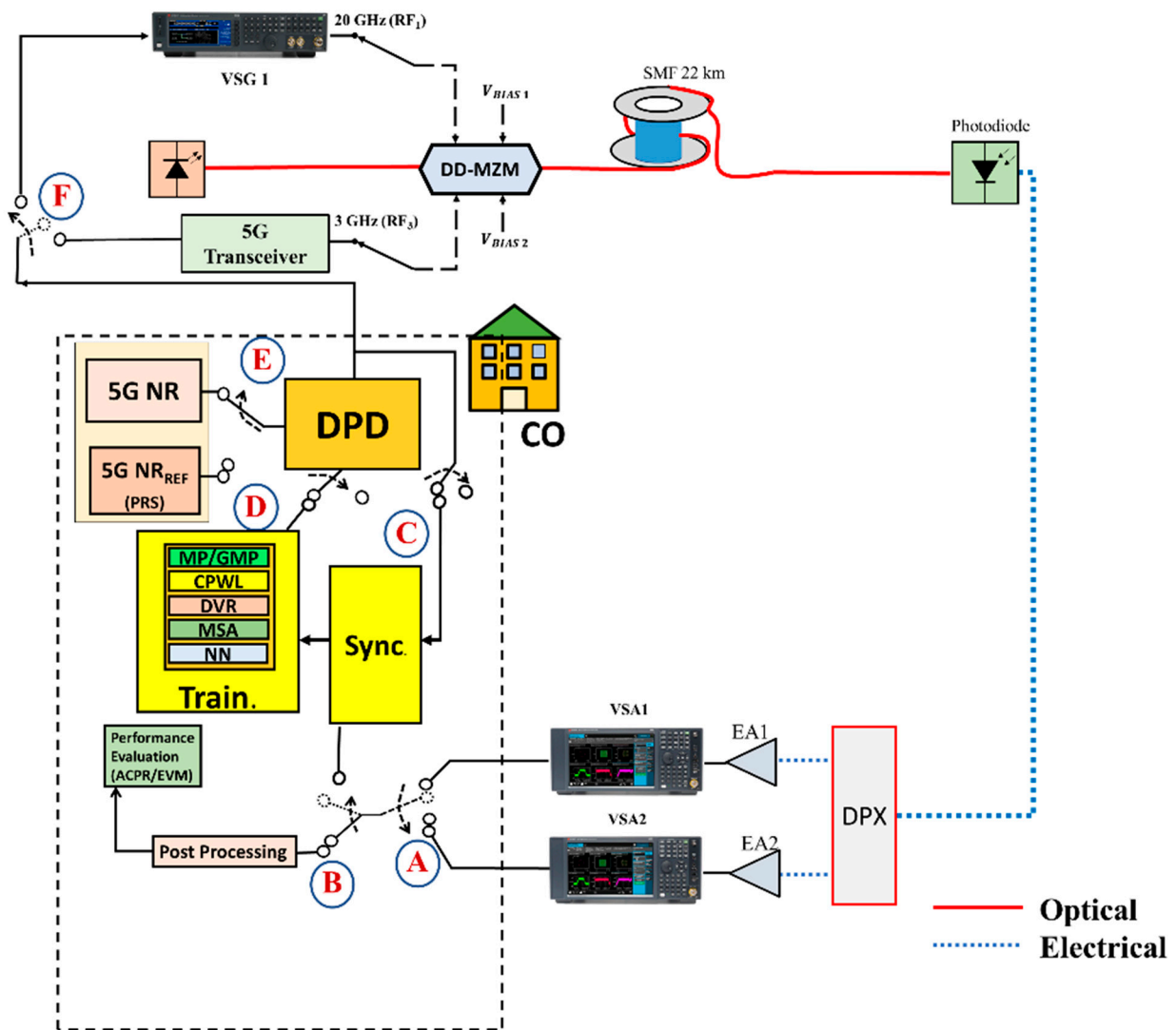
$$[Z^H Z] \hat{v} = Z^H x \tag{19}$$

Solution minimizing the cost function is represented by:

$$C = \sum_{n=1}^N |z_p(n) - x(n)|^2 \tag{20}$$

### 6. Experimental Setup

For the validation of this technique, a multiband 5G NR scenario for out/in-door environments working at 3 GHz (20 MHz bandwidth) and 20 GHz (50 MHz bandwidth), which was discussed in our previous work [35], but no DPD was implemented. As an upgradation of this architecture, the setup is integrated with a multiband DPD block to this setup for enhancing the performance of this link. The setup shown in Figure 11 comprises of a 1310 nm optical carrier is modulated by a MZM working with two distinct RF-driven signals and a 1310 nm DFB laser. VSG1 provides RF1 which is a 5G NR waveform at 20 GHz while 5G transceiver provides RF2 which is a 3 GHz flexible (O/G/F-FDM) signal. The process of DPD can be divided in to three main phases.



**Figure 11.** Block Diagram of Experimental Analog Multi Band 5G-NR-OFH RoF. The following Switches have following functionalities: A: Selection between VSA 1 and 2. B: Selection between post processing block for performance or Synchronization block. C: Connected to Synch. Block for performing synchronization followed by DPD training. D: DPD blocks connection with training DPD block E: Required for time synchronization (TS) procedure. F: Required for performing validation to launch DPD inputs to the VSGs. For DPD training phase, architectures can be selected that are requires such as MSA, CPWL, ONN, etc.

In the first phase, we upconvert these signals at the respective carrier frequencies of 3 and 20 GHz, respectively, one after the other followed by the passing of these signals through 22 km of Standard Single-Mode Fiber (SSMF) and photodetector (0.71 A/W and

40 GHz bandwidth) receives the signal and converts the received signal back to electrical domain. Since, the multiband needs to be isolated separately, an amplification stage is added. Followed by a diplexer (DPX) that separates the 20 GHz and 3 GHz signals. The signals then go to distinct vector signal analyzers (VSA). Here each VSA outputs is fed to the post processing block for performance evaluation. This step is carried out without DPD, this means that the output is evaluated without DPD process in this step.

In the second phase called as DPD training phase, the DPD operation depicted in the Figure 10 is utilized in this section and training is employed unless the error converges.

In simple words, DPD ensures that the phase and amplitude responses are inverse to that obtained at electrical amplifiers EA1 and EA2, respectively. The architectures such as GMP, CPWL, MSA and ONN can be utilized as per the user requirement and comparative requirements.

In order to achieve synchronization for the received input and the output waveform, we exploit PRS (Position reference signal) presented in the 5G NR framework. The bandwidth for PRS is taken to be 20 MHz/106 resource blocks. The received and output reference transmitted signal are correlated in time domain and the PDP (power delay profile) is passed through the maximum block to find the strongest path of arrival.

Moving on to the third phase, the predistorted basebands signal are passed into the DPD block upconverting the signals at their carrier frequency by their respective VSGs after which they are fed into the optical link. Then, the signal that is received at photodiode is passed through diplexer DPX to isolate the respective multi-bands at the VSA passes to the DPD training stage. In the phase where DPD validation is carried out, we move the switches to the opposite direction. The evaluation for 5G NR frames is achieved which involves predistortion followed by passing them to the VSG. The RoF link's nonlinearities slowly fluctuate owing to thermal effects and component ageing, from which we deduce that real-time processing in the adaptation is inessential. The parameters utilized are summarized in Table 5 that have been used previously in [35] and other state of the art [20–28].

**Table 5.** Optical Link Parameters.

Parameters	Values
5G NR Waveforms	$f_c = 3$ and 20 GHz Flexible G/F-OFDM Constellation type = 256 QAM
Laser	Wavelength = 1310 nm Transmitter Type = Mach Zehnder Modulator
Optical Fiber	Type = SSMF Fiber Dispersion = $16 \frac{\text{ps}}{\text{nmkm}}$ Fiber Distance = 22 km Attenuation = $0.42 \frac{\text{dB}}{\text{km}}$
Photodiode	Responsivity = 0.71 A/W Bandwidth = 40 GHz

The parameters of different architectures that are utilized in this study for the experimental setup are given in Tables 6–8, respectively. Table 6 enlists the parameters of the NN that results in the optimized performance of NN. The parameters discussed previously in Section 3.1 are utilized to define the structure of the proposed Optimized Neural Network and their values are mentioned in the Table 6. The last section of the table evaluates the complexity aspect of the NN by evaluating its expressions in terms of its coefficients. For a comparative analysis, the parameters of GMP and DVR are also mentioned below in Table 7. The proposed MSA-DPD technique and CPWL without modification is used with  $M = 3$  and  $K = L = 4$ . Similarly, for comparison, we have used GMP method previously used in [34,65] with the parameters  $K = Q = R = 3$ . For NN,  $N = 10$ ,  $K = 30$  is utilized.

**Table 6.** ONN parameters.

<b>Architecture</b>		<b>Parameters</b>
<b>Component</b>		<b>Value/Specification</b>
Optimizer		ADAM
Activation function		ReLU
O/P layer activation function		Softmax
Loss- function		Mean Square Error (MSE)
No. of Hidden-Layers $N$		10
No. of Neurons per layer $K$		30
<b>Component</b>		<b>Value/Specification</b>
Regularization method		Dropout
Regularization factor		0.001
Learning rate		0.01
Batch size		16, 32, 64, 128, 256, 512, 1024
Number of epochs		100
Validation split		0.4
<b>Training</b>		
<b>Component</b>		<b>Value/Specification</b>
Number of training samples		1,000,000
Number of testing samples		1,000,000
<b>Complexity</b>		<b>Value</b>
<b>Equation</b>		
$(N - 1) K^2 + (4 + N) K + 6$		8526

**Table 7.** GMP and DVR parameters.

<b>GMP</b>		<b>Parameters</b>
<b>Component</b>		<b>Value</b>
Nonlinearity Order ( $K$ )		3
Memory depth ( $Q$ )		3
Delay taps length ( $R$ )		3
<b>DVR</b>		
<b>Component</b>		<b>Value</b>
Elements in partition ( $K$ )		3
Memory depth ( $Q$ )		3
Breakpoint ( $\beta$ )		1

**Table 8.** CPWL and MSA parameters.

<b>CPWL</b>		<b>Value</b>
Memory Depth ( $M$ )		3
Number of partitions ( $L$ )		4
Length of the FIR filter ( $K$ )		4
<b>MSA</b>		<b>Value</b>
Memory Depth ( $M$ )		3
Number of partitions ( $L$ )		4
Length of the FIR filter ( $K$ )		4

## 7. Experimental Results and Discussion

In this section, the results are discussed for the experimental setup discussed in the earlier section. The Mean Square Error (MSE) is one of the ways to estimate the accuracy of the estimation of coefficients for different architectures utilized. The MSE when No DPD is applied is measured to be  $-27$  dB, while for GMP it is  $-30$  dB. The value reaches as low as  $-35$  dB for CPWL and MSA while NN has a MSE of  $-39$  dB.

In addition to MSE, the wellness of the proposed methods is compared and are presented in form of Adjacent Channel Leakage or Power Ratio (ACLR/ACPR) and Error Vector Magnitude (EVM).

### 7.1. Error Vector Magnitude

Error Vector Magnitude is the most common performance metrics that is utilized for performance evaluation for this study item in 3GPP. EVM determines the difference between the 'expected' value of the symbol in demodulated form to 'actual' value of the demodulated received symbol. EVM can be mathematically written as [5]:

$$EVM (\%) = \sqrt{\frac{\frac{1}{M} \sum_{m=1}^M |S_m - S_{0,m}|^2}{\frac{1}{M} \sum_{m=1}^M |S_m|^2}} \quad (21)$$

where  $M$  is the quantity of constellation symbols,  $S_m$  is the real symbol of the constellation associated with the symbol "m" and  $S_{0,m}$  is the real symbol associated with  $S_m$ . The 3GPP has set an EVM limit for 256 QAM to be 3.5% [74].

In Figure 12a, the Error Vector Magnitude EVM is represented by sweeping the RF input power. It is evident that MSA-DPD results in EVM reduction to <3% as compared to 5% obtained with GMP. The MSA-DPD has a slight improvement as compared to CPWL, but this is not the significant contribution, we expect to have similar improvement but with smaller complexity. In addition to this, DPD NN has a better performance as compared to MSA-DPD by 1%. It can be seen that DPD NN has a better performance as compared to MSA technique between 0 and 5 dB, however, MSA has overall better performance from -15 to 0 dB as compared to all other techniques. Similarly, In Figure 12b, EVM is reported for all the comparative methodologies employed for different flexible waveforms that has been used in 5G NR framework for 0 dBm RF input power. Clearly, MSA-DPD has almost similar performance as compared to CPWL knowing MSA reduces the complexity as compared to CPWL. It can be seen that MSA has a better reduction by 1% as compared to CPWL. When compared to NN, the NN has better performance as compared to MSA. It is clearly observable that NN has the best performance.

### 7.2. Adjacent Channel Leakage Ratio

ACLR also called as Adjacent Channel Power Ratio (ACPR) is a quantity that determines the distortion components outside the useful signal bandwidth. It is expressed as [5]:

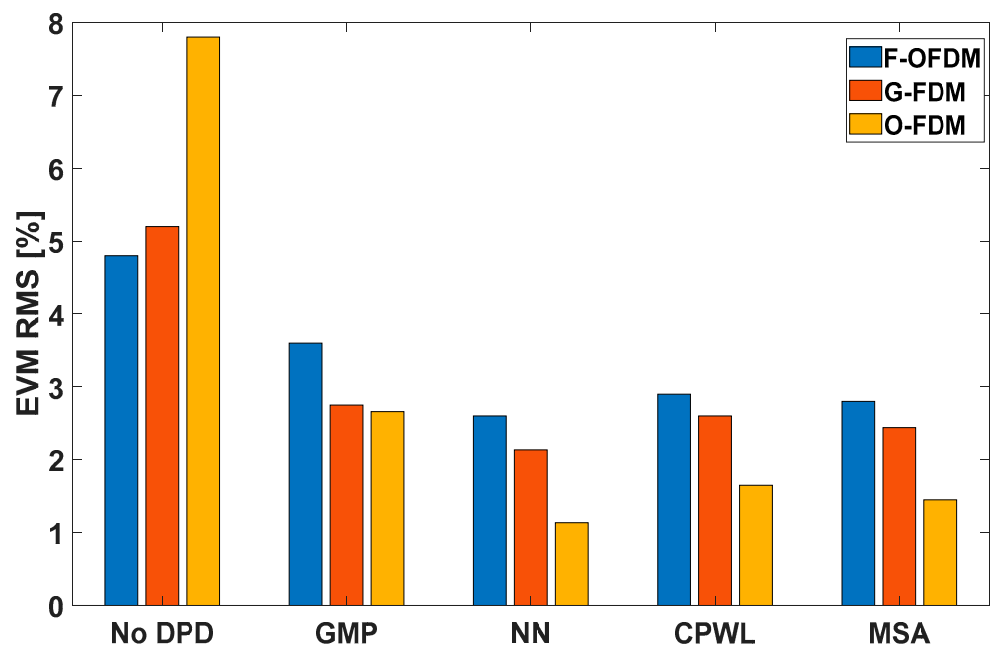
$$ACLR_{dBc} = 10 \log_{10} \left[ \frac{\int_{ab_l}^{ab_u} T(f) df}{\int_{ub_l}^{ub_u} T(f) df} \right] \quad (22)$$

where  $T(f)$  denotes Power Spectral Density (PSD) of the output signal while  $ab_u$  and  $ab_l$  are the upper and lower frequency limits of the adjacent channel;  $ub_l$  and  $ub_u$  are the frequency bounds of useful bands.

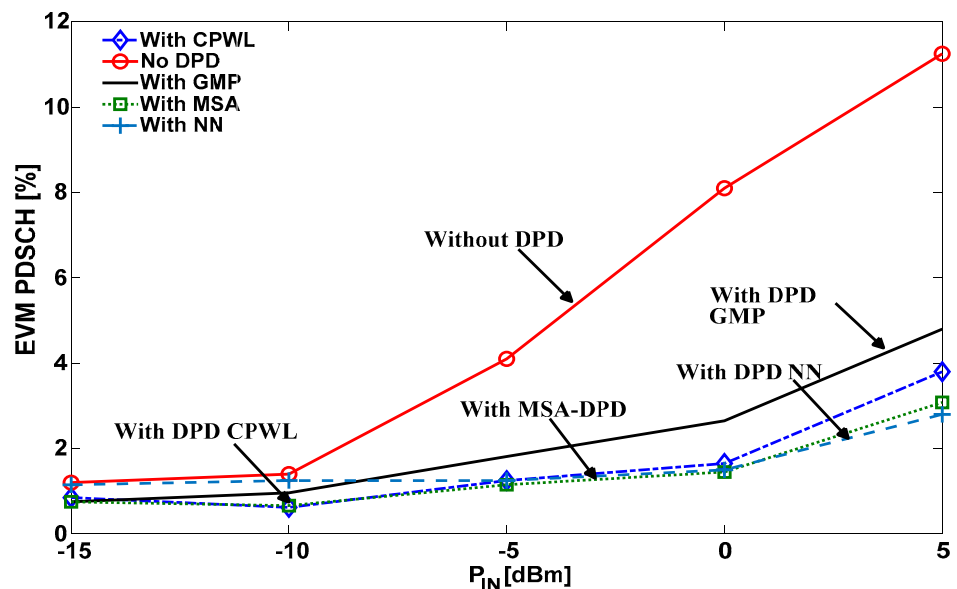
In addition to EVM, we look at the ACLR also called as MIA. In Figure 13a, ACLR for given input power variations and for different methodologies is represented. We can see that, at an RF I/P power of 0 dBm, the ACPR value for no DPD utilization is around -28 dBc, for DPD-GMP the value is around -41 dBc, for DPD-CPWL, it is around -44 dBc, for MSA-DPD, it is around -45 dBc and finally for DPD-NN, the value is around -48 dBc keeping the ACPR below -45 dBc set by 3GPP [74]. With this it is obvious that DPD-NN has achieved significantly 3 dBc lower ACPR value as compared to MSA-DPD which has the best performance as compared to other Volterra series methods.

Similarly, in Figure 13b, the spectral density is shown. The power spectral density (PSD) with and without DPD at 3 GHz is shown. It is observed that MSA technique is achieving good performance as compared to CPWL while the NN DPD has a little advantage over the MSA with spectral regrowth suppression in the 10 to 15 MHz zones.

Similarly, in Figure 13c, the electrical spectra are shown, there are two main components that show the carrier signals at 3 and 20 GHz, respectively. It represents that the aim of the proposed MSA-DPD technique further reduces the complexity and implementation of the DPD methodology and enhance the performance of the link as compared to CPWL, GMP and when no DPD is applied. Therefore, it is observable that ACPR is reduced with the proposed MSA-DPD methods as compared to CPWL and GMP method by a good proportion keeping the ACPR below  $-45$  dBc set by 3GPP [74]. It is important to observe that MSA-DPD performs a little better than CPWL, however, the performance gain is not the only benefit, but the complexity reduction is the most important benefit of the proposed MSA-DPD technique (discussed in Section 7.3).

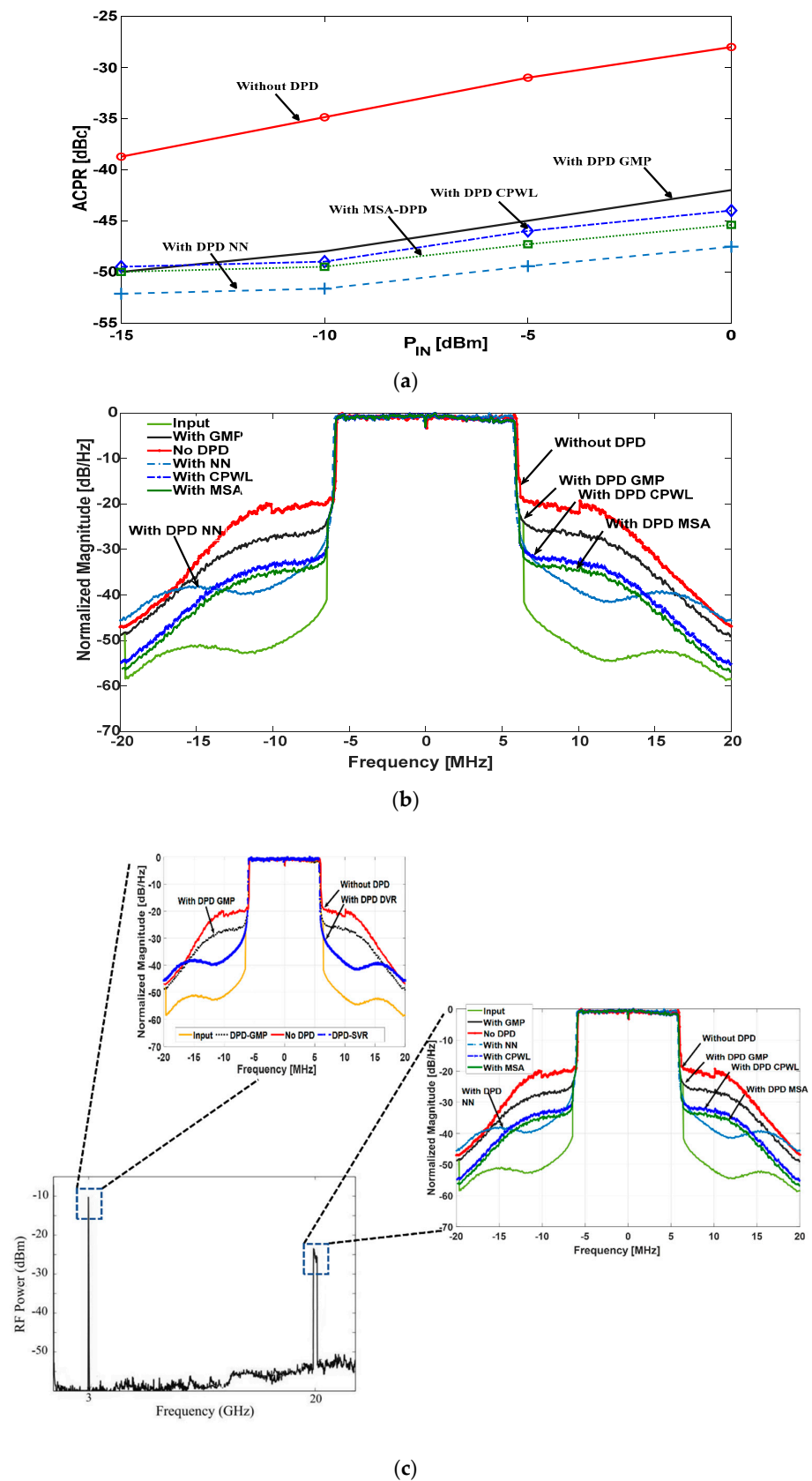


(a)



(b)

**Figure 12.** EVM results (a) for DPD methods utilized in this study vs. No DPD case. DPD with CPWL (blue), GMP (black), MSA (green), NN (light blue) and No DPD (red). (b) 5G NR OFH performance DPD efficacy with MSA, CPWL, GMP and without DPD for flexible 5G transceiver waveforms.



**Figure 13.** 5G NR OFH performance (a) ACPR vs. varying RF input power. (b) DPD efficacy with MSA, CPWL, GMP and without DPD in terms of Power Spectral Density/Spectral Regrowth. (c) Electrical Spectra showing the two carrier frequencies.

### 7.3. Complexity Considerations

The complexity reduction that MSA-DPD brings with gaining similar performance as compared to CPWL method is significant contribution. The complexity of the DPD model construction of the models is shown in Table 9, which is mainly measured by the required number of real multipliers, as the multipliers take up most of the hardware resources. Table 9 signifies that MSA DPD (220 multiplications) has much lesser complexity as compared to CPWL (880 multiplications). The advanced variations in Volterra series can be obtained by changing memory depth and nonlinearity order to higher numbers. However, the computational complexity has to be considered as shown in Equation (11). This means that while selecting the DPD model and its complexity, a smart trade-off between complexity and performance can be made accordingly. For a comparative evaluation of NN and GMP methods in terms of complexity and performance, we evaluate expressions for each method in term of its complexity. The number of coefficients and multiplications in CPWL architecture are higher as compared to GMP, however, the complexity in CPWL results in better performance as compared to GMP (as seen in results section in Figures 11 and 12). The reduction in complexity is brought into the play as proposed in MSA method. The performance of MSA achieves similar performance as compared to CPWL, however, the complexity in terms of coefficients in MSA is reduced from 520 to 260 and multiplications are reduced from 880 to 220. Similarly, looking at NN architecture, the performance (see Figures 11 and 12) is comparable to MSA which is in similar proportions to NN, however, the number of coefficients rise to 8526 and number of multiplications reached to 8224. NN complexity is a challenging issue, and this can be reduced as shown in this work that a very limited number of hidden layers and neuron per layer are employed. The multiplications and complexity of the NN increases exponentially with the increase in number of hidden layers and neurons per layer, depending on the application, MSA with much less complexity and similar based DPD can be employed as they achieve a similar performance as compared to NN.

**Table 9.** Complexity Comparisons.

DPD Method	Coefficients	# Coefficients	Multiplications
GMP	$K_a(Q_a + 1) + K_b(Q_b + 1)R_b + K_c(Q_c + 1)R_c$	84	244
CPWL	$(4M + 1)(K + 1)L$	260	$(14M + 2)(K + 1)L = 880$
MSA	$2(4M + 1)(K + 1)L$	520	$(14M + 2)(K + 1) = 220$
NN	$(N - 1)K^2 + (4 + N)K + 6$	8526	$(N - 1)K^2 + 4K + 4 = 8224$

### 7.4. Real Time Implementation

In a realistic NN scenario, linearization methodology is carried out at the Central Office (CO) where the BBUs are placed and a periodical re-training of the DPD system is in this case necessary, requiring however a negligible time with respect to the time of normal operation of the RoF system. Recently, a Xilinx DPD kit has been developed that can be used for this purpose [75].

Bringing feedback signal from the BS to RAU is one of the main challenging task in the adaptive recompense of the RoF link. This is due to possible nonlinearity of the feedback link; actually, it can be as nonlinear as the RoF link, which is being compensated for. The present work is based on the fact that the predistorter sees only the non-idealities that it needs to compensate. Since it is assumed that the nonlinear feedback connection is uncompensated, therefore, it would destroy the performance of the compensation. In simple words, it means that an approach is utilized where the RoF link is first compensated for using a post distorter and known training signal from the RAU is used here. After that, the already compensated for downlink RoF link can be used as a feedback connection for the compensation [76]. However, as a possible feedback scenario, Figure 14 represents a realistic implementation of an adaptive DPD that shifts complex processing of the signals at the Central Office/Base Transmit Station (CO-BTS).



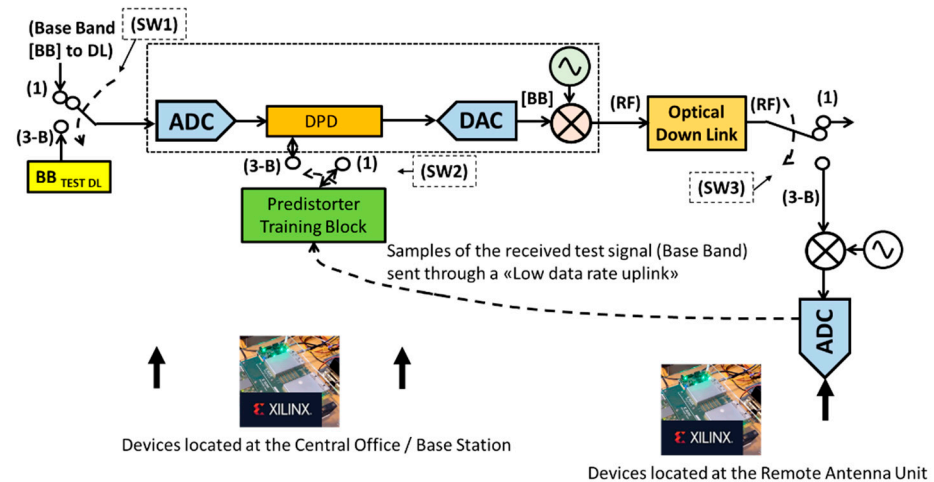


Figure 14. Conceptual Possible Realization of an Adaptive Predistortion Scheme.

It should be noted that DPD works as a black box, it counter acts the overall nonlinearities of the system including that of MZM (laser), fiber and photodiode. For tens of km, the combined effect of laser chirp and fiber dispersion becomes a major nonlinearity issue after tens of km [5]. Therefore, laser and possibly photodiode are the primary source of nonlinearity which is mitigated in this proposed bench. In future, it will be interesting to increase the length of fiber and linearize the link by mitigating the fiber nonlinearities such as Kerr effect.

Indeed, with the higher modulation format and higher bandwidth similar to multiple LTE carriers or 5G new radio (NR) waveforms as discussed, they would lead to higher complexity of DPD operation due to stronger PAPR. Concomitantly, the elevation in bandwidth will lead to overall increase in the base-band memory of the system model. Nevertheless, the evaluated models are still valid. However, higher values of the  $Q$  and  $K$  will be indispensable as compared to the considered case.

To summarize the discussion above, Table 10 lists the values of ACPR, MSE and EVM for all the utilized methodologies.

Table 10. Results Summary for ACPR (0 dBm), EVM (5 dBm) and MSE.

Methodology	ACPR (dBc)	EVM (%)	MSE (dB)
No-DPD	−28	11	−27
DPD-GMP	−41	5	−30
DPD-CPWL	−44	4	−35
DPD-MSA	−45	3.5	−35
DPD-ONN	−48	2.7	−39

## 8. Conclusions

This work summarizes a successful realization of 5G NR multiband OFH with a novel unprecedented DPD solutions for reducing RoF nonlinearities using the conventional volterra based methods and further exploiting deep learning methods to further improve results using Neural Networks. Firstly, a novel MSA DPD method has been proposed which reduces the complexity of CPWL method and reaches similar performance with reduction of complexity overheads and multiplications by 75%. The article also explains the theoretical foundations and elements required for building a Neural Network. The 5G NR multiband signals at 3 GHz and 20 GHz are employed to 22 km fiber length. The proposed MSA-DPD method results in reduction of ACPR from −28 dBc to −45 dBc while NN results in ACPR reduction to −48 dBc. Similarly, 11% of EVM is reduced to 3.5% with MSA method and NN leads to 2.7% at RF input power of 5 dBm. The results signify that proposed MSA-DPD method reduces the signal impairments in better proportions

as compared to GMP method and CPWL. The estimated multiplication operations from CPWL to MSA are reduced from 880 to 220 leading to much less complexity and overheads meeting the standardization requirements set by 3GPP Release 17. MSA with much less complexity and similar performance to NN can be employed for DPD depending on the application scenario.

**Author Contributions:** Conceptualization, M.U.H.; methodology, M.U.H. and J.S.; software, M.U.H. and M.R.; validation, M.I.A., M.A. and M.U.H.; formal analysis, M.U.H.; investigation, M.U.H. and M.A.; resources, M.I.A. and M.U.H.; data curation, J.S.; writing—original draft preparation, M.U.H. and M.R.; writing—review and editing, M.U.H., M.A., M.R., M.I.A., J.S.; visualization, J.S.; supervision, M.U.H. and M.I.A.; project administration, M.A.; funding acquisition, M.R. and M.U.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** Authors would like to thank Isha Mittal and Samridhi Mishra for their efforts in experimentation and data analysis.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hadi, M.; Awais, M.; Raza, M.; Khurshid, K.; Jung, H. Neural Network DPD for Aggrandizing SM-VCSEL-SSMF-Based Radio over Fiber Link Performance. *Photonics* **2021**, *8*, 19. [CrossRef]
- 3GPP, User Equipment (UE) Radio Transmission and Reception; Part 3: Range 1 and Range 2 Interworking Operation with Other Radios: 2019, TS 38.101-3 Version 16.0.0 Release 16. Available online: [https://www.etsi.org/deliver/etsi\\_ts/138100\\_138199/13810103/16.06.00\\_60/ts\\_13810103v160600p.pdf](https://www.etsi.org/deliver/etsi_ts/138100_138199/13810103/16.06.00_60/ts_13810103v160600p.pdf) (accessed on 5 November 2021).
- Nanni, J.; Polleux, J.-L.; Algani, C.; Rusticelli, S.; Perini, F.; Tartarini, G. VCSEL-based radio-over-G652 fiber system for short-/medium-range MFH solutions. *IEEE/OSA J. Lightwave Technol.* **2018**, *36*, 4430–4437. [CrossRef]
- Khurshid, K.; Khan, A.A.; Siddiqui, H.; Rashid, I.; Hadi, M.U. Big Data Assisted CRAN Enabled 5G SON Architecture. *J. ICT Res. Appl.* **2019**, *13*, 93–106. [CrossRef]
- Hadi, M.U.; Nanni, J.; Polleux, J.-L.; Traverso, P.A.; Tartarini, G. Direct digital predistortion technique for the compensation of laser chirp and fiber dispersion in long haul radio over fiber links. *Opt. Quantum Electron.* **2019**, *51*, 205. [CrossRef]
- Hadi, M.U.; Murtaza, G. Enhancing distributed feedback-standard single mode fiber-radio over fiber links performance by neural network digital predistortion. *Microw. Opt. Technol. Lett.* **2021**, *63*, 1558–1565. [CrossRef]
- Zhang, X.; Zhu, R.; Shen, D.; Liu, T. Linearization Technologies for Broadband Radio-Over-Fiber Transmission Systems. *Photonics* **2014**, *1*, 455–472. [CrossRef]
- Li, H.; Bauwelinck, J.; Demeester, P.; Torfs, G.; Verplaetse, M.; Verbist, J.; Van Kerrebrouck, J.; Breyne, L.; Wu, C.-Y.; Bogaert, L.; et al. Real-Time 100-GS/s Sigma-Delta Modulator for All-Digital Radio-Over-Fiber Transmission. *J. Lightwave Technol.* **2020**, *38*, 386–393. [CrossRef]
- Hadi, M.U.; Jung, H.; Ghaffar, S.; Traverso, P.A.; Tartarini, G. Optimized digital radio over fiber system for medium range communication. *Opt. Commun.* **2019**, *443*, 177–185. [CrossRef]
- Wang, J.; Jia, Z.; Campos, L.A.; Knittle, C.; Jia, S. Delta-Sigma Modulation for Next Generation Fronthaul Interface. *J. Lightwave Technol.* **2018**, *37*, 2838–2850. [CrossRef]
- Hadi, M.U.; Traverso, P.A.; Tartarini, G.; Jung, H. Experimental characterization of Sigma Delta Radio over fiber system for 5G C-RAN downlink. *ICT Express* **2020**, *6*, 23–27. [CrossRef]
- Hadi, M.U.; Jung, H.; Traverso, P.A.; Tartarini, G. Experimental evaluation of real-time sigma-delta radio over fiber system for fronthaul applications. *Int. J. Microw. Wirel. Technol.* **2020**, *13*, 756–765. [CrossRef]
- Hadi, M.U.; Kausar, S.; Mittal, I. Efficient Pre-BPF based Sigma Delta Radio over Fiber System for 5G NR Fronthauls. In Proceedings of the IEEE EUROCON 2021—19th International Conference on Smart Technologies, Lviv, Ukraine, 6–8 July 2021; pp. 308–311.
- Draa, M.N.; Hastings, A.S.; Williams, K.J. Comparison of photodiode nonlinearity measurement systems. *Opt. Express* **2011**, *19*, 12635–12645. [CrossRef] [PubMed]
- Chen, Z.; Yan, L.; Pan, W.; Luo, B.; Zou, X.; Guo, Y.; Jiang, H.; Zhou, T. SFDR enhancement in analog photonic links by simultaneous compensation for dispersion and nonlinearity. *OSA Opt. Expr.* **2013**, *21*, 20999–21009. [CrossRef]
- Haas, B.M.; Murphy, T. A Simple, Linearized, Phase-Modulated Analog Optical Transmission System. *IEEE Photonics Technol. Lett.* **2007**, *19*, 729–731. [CrossRef]

17. Hraimel, B.; Zhang, X. Characterization and compensation of AM-AM and AM-PM distortion in mixed polarization radio over fiber systems. In Proceedings of the 2012 IEEE MTT-S International Microwave Symposium (IMS), Montreal, QC, Canada, 1–3 June 2012.
18. Zhu, R.; Zhang, X. Linearization of Radio-Over-Fiber Systems by Using Two Lasers with Different Wavelengths. In Proceedings of the 2014 IEEE MTT-S International Microwave Symposium (IMS), Tampa, FL, USA, 1–6 June 2014.
19. Ghannouchi, F.; Younes, M.; Rawat, M. Distortion and impairments mitigation and compensation of single- and multi-band wireless transmitters. *IET Microw. Antennas Propag.* **2013**, *7*, 518–534.
20. Duan, R.; Xu, K.; Dai, J.; Cui, Y.; Wu, J.; Li, Y.; Dai, Y.; Li, J. Linearity improvement based on digital signal processing in in-tensity-modulated analog optical links incorporating photonic frequency down conversion. In Proceedings of the Optical Fiber Communication Conference and Exposition, Los Angeles, LA, USA, 6–8 March 2012.
21. Vieira, L.; Gomes, N.J.; Nkansah, A.; van Dijk, F. Behavioral modeling of radio-over-fiber links using memory polynomials. In Proceedings of the 2010 IEEE Topical Meeting on Microwave Photonics (MWP), Montreal, QC, Canada, 5–9 October 2010; pp. 85–88.
22. Pei, Y.; Xu, K.; Li, J.; Zhang, A.; Dai, Y.; Ji, Y.; Lin, J. Complexity-reduced digital predistortion for subcarrier multiplexed radio over fiber systems transmitting sparse multi-band RF signals. *Opt. Express* **2013**, *21*, 3708–3714. [[CrossRef](#)]
23. Hekkala, A.; Hiiivala, M.; Lasanen, M.; Perttu, J.; Vieira, L.C.; Gomes, N.J.; Nkansah, A. Predistortion of Radio Over Fiber Links: Algorithms, Implementation, and Measurements. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2011**, *59*, 664–672. [[CrossRef](#)]
24. Fuochi, F.; Hadi, M.U.; Nanni, J.; Traverso, P.A.; Tartarini, G. Digital predistortion technique for the compensation of nonlinear effects in radio over fiber links. In Proceedings of the 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI), Bologna, Italy, 7–9 September 2016; pp. 1–6.
25. Vieira, L.C.; Gomes, N.J.; Nkansah, A. An experimental study on digital predistortion for radio-over-fiber links. *Asia Commun. Photonics Conf. Exhib.* **2010**, 126–127. [[CrossRef](#)]
26. Hekkala, A.; Lasanen, M.; Vieira, L.C.; Gomes, N.J.; Nkansah, A. Architectures for Joint Compensation of RoF and PA with Nonideal Feedback. In Proceedings of the 2010 IEEE 71st Vehicular Technology Conference, Taipei, Taiwan, 16–19 May 2010; pp. 1–5.
27. Mateo, C.; Carro, P.L.; Garcia-Ducar, P.; De Mingo, J.; Salinas, I. Experimental evaluation of the feedback loop effects in digital predistortion of a radio-over-fiber system. In Proceedings of the 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, Italy, 4–7 July 2017; pp. 1039–1041.
28. Mateo, C.; Clemente, J.; Garcia-Ducar, P.; Carro, P.L.; de Mingo, J.; Salinas, I. Digital predistortion of a full-duplex Radio-over-Fiber mobile fronthaul link with feedback loop. In Proceedings of the 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26–30 June 2017; pp. 1425–1430.
29. Mateo, C.; Carro, P.L.; García-Dúcar, P.; De Mingo, J.; Salinas, I. Minimization of Feedback Loop Distortions in Digital Predistortion of a Radio-Over-Fiber System with Optimization Algorithms. *IEEE Photonics J.* **2017**, *9*, 7904414. [[CrossRef](#)]
30. Roselli, L.; Borgioni, V.; Zepparelli, F.; Ambrosi, F.; Comez, M.; Faccin, P.; Casini, A. Analog laser predistortion for multiservice radio-over-fiber systems. *J. Lightwave Technol.* **2003**, *21*, 1211–1223. [[CrossRef](#)]
31. Aslam, N.; Xia, K.; Hadi, M.U. Optimal Wireless Charging Inclusive of Intellectual Routing Based on SARSA Learning in Renewable Wireless Sensor Networks. *IEEE Sens. J.* **2019**, *19*, 8340–8351. [[CrossRef](#)]
32. Hraimel, B.; Zhang, X. Low-cost broadband predistortion-linearized singledrive x-cut Mach-Zehnder modulator for ra-dio-over-fiber systems. *Photonics Technol. Lett.* **2012**, *24*, 1571–1573. [[CrossRef](#)]
33. Hadi, M.U.; Khurshid, K.; Basit, A. Pre-BPF based Sigma Delta Radio over Fiber for 5G Sub-6 GHz Fronthauls. In Proceedings of the 2021 International Conference on Communication Technologies (ComTech), Washington, DC, USA, 20–25 June 2021; pp. 13–16. [[CrossRef](#)]
34. Hadi, M.U.; Nanni, J.; Traverso, P.A.; Tartarini, G.; Venard, O.; Baudoin, G.; Polleux, J.-L. Experimental evaluation of digital predistortion for VCSEL-SSMF-based Radio-over-Fiber link. In Proceedings of the 2018 International Topical Meeting on Microwave Photonics (MWP), Toulouse, France, 22–25 October 2018; pp. 1–4.
35. Hadi, M.U.; Kantana, C.; Traverso, P.A.; Tartarini, G.; Venard, O.; Baudoin, G.; Polleux, J. Assessment of digital predistortion methods for DFB-SSMF radio-over-fiber links linearization. *Microw. Opt. Technol. Lett.* **2020**, *62*, 540–546. [[CrossRef](#)]
36. Hadi, M.U.; Awais, M.; Raza, M. Multiband 5G NR-over-Fiber System Using Analog Front Haul. In Proceedings of the 2020 International Topical Meeting on Microwave Photonics (MWP), Matsue, Japan, 24–26 November 2020; pp. 1–4.
37. James, J.; Shen, P.; Nkansah, A.; Liang, X.; Gomes, N.J. Nonlinearity and Noise Effects in Multi-Level Signal Millimeter-Wave Over Fiber Transmission Using Single and Dual Wavelength Modulation. *IEEE Trans. Microw. Theory Tech.* **2010**, *58*, 3189–3198. [[CrossRef](#)]
38. Jeon, D.-H.; Jung, H.-D.; Han, S.-K. Mitigation of dispersion-induced effects using SOA in analog optical transmission. *IEEE Photonics Technol. Lett.* **2002**, *14*, 1166–1168. [[CrossRef](#)]
39. Usman Hadi, M.; Mittal, I.; Siraj, M. 5G NR Fiber Wireless System using Analog Optical Front Haul for Long Reach Applications. In Proceedings of the 2021 International Conference on Communication, Control and Information Sciences (ICCIsc), Online, 24–25 September 2021; pp. 1–4. [[CrossRef](#)]
40. Ismail, T.; Liu, C.-P.; Mitchell, J.; Seeds, A.J. High-Dynamic-Range Wireless-Over-Fiber Link Using Feedforward Linearization. *J. Lightwave Technol.* **2007**, *25*, 3274–3282. [[CrossRef](#)]

41. Park, S.-H.; Choi, Y.-W. Significant suppression of the third intermodulation distortion in transmission system with optical feedforward linearized transmitter. *IEEE Photonics Technol. Lett.* **2005**, *17*, 1280–1282. [[CrossRef](#)]
42. Korotky, S.K.; De Ridder, R. Dual parallel modulation schemes for low-distortion analog optical transmission. *IEEE J. Sel. Areas Commun.* **1990**, *8*, 1377–1381. [[CrossRef](#)]
43. Gamage, P.A.; Nirmalathas, A.; Lim, C.; Novak, D.; Waterhouse, R. Design and Analysis of Digitized RF-Over-Fiber Links. *J. Lightwave Technol.* **2009**, *27*, 2052–2061. [[CrossRef](#)]
44. Zhang, L.; Pang, X.; Ozolins, O.; Udalcovs, A.; Popov, S.; Xiao, S.; Hu, W.; Chen, J. Spectrally efficient digitized radio-over-fiber system with k-means clustering-based multidimensional quantization. *Opt. Lett.* **2018**, *43*, 1546–1549. [[CrossRef](#)]
45. Shiina, R.; Fujiwara, T.; Taniguchi, T.; Ikeda, S. DRoF-based optical video re-transmission of commercial RF broadcast signals. *J. Opt. Commun. Netw.* **2019**, *11*, 559–567. [[CrossRef](#)]
46. Mathe, D.M.; Neto, B.; Oliveira, R.S.; Teixeira, A.L.J.; Costa, J.C.W.A. Adaptive Signal Compression in Intermediate Frequency Band for DRoF Mobile Fronthaul. *IEEE Photonics Technol. Lett.* **2019**, *31*, 579–582. [[CrossRef](#)]
47. Hadi, M.U. Sigma Delta Radio Frequency Transport System for 5G Sub-6 GHz Fronthauls. In Proceedings of the 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), Rupnagar, India, 26–28 November 2020; pp. 596–600.
48. Sezgin, I.C.; Dahlgren, M.; Eriksson, T.; Coldrey, M.; Larsson, C.; Gustavsson, J.; Fager, C. A Low-Complexity Distributed-MIMO Testbed Based on High-Speed Sigma-Delta-Over-Fiber. *IEEE Trans. Microw. Theory Tech.* **2019**, *67*, 2861–2872. [[CrossRef](#)]
49. Hadi, M.U.; Basit, A.; Khurshid, K. MIMO Enabled MultiBand 5G NR FiberWireless using Sigma Delta over FiberTechnology. In Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 12–16 January 2021; pp. 1007–1010.
50. Cui, Y.; Zhang, M.; Wang, D.; Liu, S.; Li, Z.; Chang, G.K. Bit-based support vector machine nonlinear detector for millimeter-wave radio-over-fiber mobile fronthaul systems. *Opt. Express* **2017**, *25*, 26186–26197. [[CrossRef](#)]
51. Hadi, M.U.; Mittal, I. On the Use of SVR based Machine Learning Method for Nonlinearities Mitigation in Short Range Fronthaul Links. In Proceedings of the 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 18–19 June 2021; pp. 628–631.
52. Gonzalez, N.G.; Zibar, D.; Caballero, A.; Monroy, I.T. Experimental 2.5-Gb/s QPSK WDM Phase-Modulated Radio-Over-Fiber Link with Digital Demodulation by a K-Means Algorithm. *IEEE Photonics Technol. Lett.* **2010**, *22*, 335–337. [[CrossRef](#)]
53. Lee, J.; He, J.; Wang, Y.; Fang, C.; Wang, K. Experimental demonstration of millimeter-wave radio-over-fiber system with convolutional neural network and binary convolutional neural network (BCNN). *arXiv* **2020**, arXiv:2001.02018.
54. Lee, J.; He, J.; Wang, K. Neural networks and FPGA hardware accelerators for millimeter-wave radio-over-fiber systems. In Proceedings of the International Conference on Transparent Optical Networks, Bari, Italy, 19–23 July 2020; p. Mo.D1.5.
55. Liu, S.; Wang, X.; Zhang, W.; Shen, G.; Tian, H. An Adaptive Activated ANN Equalizer Applied in Millimeter-Wave RoF Transmission System. *IEEE Photonics Technol. Lett.* **2017**, *29*, 1935–1938. [[CrossRef](#)]
56. Liu, S.; Xu, M.; Wang, J.; Lu, F.; Zhang, W.; Tian, H.; Chang, G.-K. A Multilevel Artificial Neural Network Nonlinear Equalizer for Millimeter-Wave Mobile Fronthaul Systems. *J. Lightwave Technol.* **2017**, *35*, 4406–4417. [[CrossRef](#)]
57. Liu, S.; Alfadhli, Y.M.; Shen, S.; Tian, H.; Chang, G.K. Mitigation of multi-user access impairments in 5G A-RoF-based mobile-fronthaul utilizing machine learning for an artificial neural network nonlinear equalizer. In Proceedings of the Optical Fiber Communication Conference, San Diego, CA, USA, 11–15 March 2018.
58. Hadi, M.U.; Basit, A. Machine Learning for Performance Enhancement in Fronthaul Links for IOT Applications. In Proceedings of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), Islamabad, Pakistan, 20–21 May 2021; pp. 1–5.
59. Liu, E.; Yu, Z.; Yin, C.; Xu, K. Nonlinear distortions compensation based on artificial neural networks in wideband and multi-carrier systems. *IEEE J. Quantum Electron.* **2019**, *55*, 800305. [[CrossRef](#)]
60. Liu, J.; Zou, X.; Bai, W. Performance enhancement of UPMC based radio over fiber system using ANN equalizer. In Proceedings of the Asia Communications and Photonics Conference, Hangzhou, China, 26–29 October 2018.
61. Safari, L.; Baghersalimi, G.; Karami, A.; Kiani, A. On the Equalization of an OFDM-Based Radio-over-Fiber System Using Neural Networks. *Radioengineering* **2017**, *26*, 162–169. [[CrossRef](#)]
62. Zhou, Q.; Lu, F.; Xu, M.; Peng, P.-C.; Liu, S.; Shen, S.; Zhang, R.; Yao, S.; Finkelstein, J.; Chang, G.-K. Enhanced Multi-Level Signal Recovery in Mobile Fronthaul Network Using DNN Decoder. *IEEE Photonics Technol. Lett.* **2018**, *30*, 1511–1514. [[CrossRef](#)]
63. Xu, T.; Shevchenko, A.N.; Zhang, Y.; Jin, C.; Zhao, J.; Liu, T. Information rates in Kerr non-linearity limited optical fiber communication systems. *Opt. Express* **2021**, *29*, 17428–17439. [[CrossRef](#)] [[PubMed](#)]
64. Jin, C.; Shevchenko, N.A.; Li, Z.; Popov, S.; Chen, Y.; Xu, T. Nonlinear Coherent Optical Systems in the Presence of Equalization Enhanced Phase Noise. *J. Lightwave Technol.* **2021**, *39*, 4646–4653. [[CrossRef](#)]
65. Khayatzadeh, R.; Poëtte, J.; Cabon, B. Impact of Phase Noise in 60-GHz Radio-Over-Fiber Communication System Based on Passively Mode-Locked Laser. *J. Lightwave Technol.* **2014**, *32*, 3529–3535. [[CrossRef](#)]
66. Maeda, J.; Katoh, T.; Ebisawa, S. Effect of Fiber Dispersion on Subcarrier QAM Signal in Radio-Over-Fiber Transmission. *J. Lightwave Technol.* **2012**, *30*, 2625–2632. [[CrossRef](#)]
67. Hadi, M.U. Mitigation of nonlinearities in analog radio over fiber links using machine learning approach. *ICT Express* **2021**, *7*, 253–258. [[CrossRef](#)]

68. Vieira, L.C.; Gomes, N.J. Experimental demonstration of digital predistortion for orthogonal frequency-division multi-plexing-radio over fibre links near laser resonance. *IET Optoelectron.* **2015**, *9*, 310–316. [[CrossRef](#)]
69. Hadi, M.U.; Nanni, J.; Venard, O.; Baudoin, G.; Polleux, J.-L.; Traverso, P.A.; Tartarini, G. Linearity Improvement of VCSELs based Radio over Fiber Systems utilizing Digital Predistortion. *Adv. Sci. Technol. Eng. Syst. J.* **2019**, *4*, 156–163. [[CrossRef](#)]
70. Hadi, M.U.; Nanni, J.; Venard, O.; Baudoin, G.; Polleux, J.-L.; Tartarini, G. Practically Feasible Closed-Loop Digital Predistortion for VCSEL-MMF-Based Radio-over-Fiber links. *Radioengineering* **2020**, *29*, 37–43. [[CrossRef](#)]
71. Hadi, M.U.; Basit, A.; Khurshid, K. Nonlinearities Mitigation in Radio over Fiber Links for Beyond 5G C-RAN Applications using Support Vector Machine Approach. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–5.
72. Cao, W.; Li, Y.; Zhu, A. Magnitude-selective affine function based digital predistorter for RF power amplifiers in 5G small-cell transmitters. In Proceedings of the 2017 IEEE MTT-S International Microwave Symposium (IMS), Honolulu, HI, USA, 4–9 June 2017; pp. 1539–1541.
73. Li, Y.; Cao, W.; Zhu, A. Instantaneous sample indexed magnitudes elective affine function-based behavioral model for digital predistortion of RF power amplifiers. *IEEE Trans. Microw. Theory Tech.* **2018**, *66*, 5000–5010.
74. LTE. *Evolved Universal Terrestrial Radio Access (E-UTRA)*; Base Station (BS) Radio Transmission and Reception; ETSI, TS 136 104 V8.2.0; European Telecommunications Standards Institute: Valbonne, France, 2018.
75. Xilinx DPD Design. Available online: [https://www.xilinx.com/publications/prod\\_mktg/pn2061.pdf](https://www.xilinx.com/publications/prod_mktg/pn2061.pdf) (accessed on 1 September 2021).
76. Hadi, M.U.; Traverso, P.A.; Tartarini, G.; Venard, O.; Baudoin, G.; Polleux, J.-L. Digital Predistortion for Linearity Improvement of VCSEL-SSMF-Based Radio-Over-Fiber Links. *IEEE Microw. Wirel. Compon. Lett.* **2019**, *29*, 155–157. [[CrossRef](#)]