Integrating Volterra Series Model and Deep Neural Networks to Equalize Nonlinear Power Amplifiers

Robert J. Thompson and Xiaohua Li

Department of Electrical and Computer Engineering
State University of New York at Binghamton
Binghamton, NY 13902
Email: {rthomps5,xli}@binghamton.edu

***Abstract*—The nonlinearity of power amplifiers (PAs) has been one of the severe constraints to the performance of modern wireless transceivers. This problem is even more challenging for the fifth generation (5G) cellular system since 5G signals have extremely high peak to average power ratio. This paper develops nonlinear equalizers that exploit both deep neural networks (DNNs) and Volterra series models to mitigate PA nonlinear distortions. The DNN equalizer architecture consists of multiple one-dimension convolutional layers. The input features are designed according to the Volterra series model of nonlinear PAs. This enables the DNN equalizer to mitigate nonlinear PA distortions more effectively while avoiding over-fitting under limited training data. Experiments are conducted with both simulated data based on a Doherty nonlinear PA model and real measurement data obtained from a highly nonlinear cable TV PA. The results demonstrate that the proposed DNN equalizer has superior performance over conventional nonlinear equalization approaches.**

***Index Terms*—Nonlinear power amplifier, Volterra series, equalization, power efficiency, deep neural network**

I. Introduction

Most modern wireless communication systems, including the fifth generation (5G) cellular systems, use multi-carrier or OFDM (orthogonal frequency division multiplexing) modulations whose signals have extremely high peak to average power ratio (PAPR). This makes it challenging to enhance the efficiency of power amplifiers (PAs). Signals with high PAPR need linear power amplifier response in order to reduce distortion. Nevertheless, PAs have the optimal power efficiency only in the nonlinear saturated response region. In practice, the PAs in the wireless transceivers have to work with high output backoff (OBO) in order to suppress nonlinear distortions, which unfortunately results in severe reduction of power efficiency [1]. This problem, originated from the nonlinearity of PAs, has been one of the major constraints to enhance the power efficiency of modern communication systems.

Various strategies have been investigated to mitigate this problem. The first strategy is to reduce the PAPR of the transmitted signals. Many techniques have been developed for this purpose, such as signal clipping, peak cancellation, error waveform subtraction [2]. For OFDM signals, pilot tones and unmodulated subcarriers can be exploited to reduce PAPR with some special pre-coding techniques [3].

The second strategy is to linearize the PAs at the transmitters. One of the dominating practices today is to insert a digital pre-distorter (DPD) before the PA, which distorts the signals appropriately so as to compensate for the nonlinear PA response [4] [5] [6]. DPD has been applied widely in many modern transmitters [2].

The third strategy is to mitigate the nonlinear PA distortions at the receivers via post-distorter equalization [7] [8] [9]. The solution presented in [10] develops a Bayesian signal detection algorithm based on the nonlinear response of the PAs. However, this method works for the simple "AM-AM AM-PM" nonlinear PA model only. Alternatively, as a powerful nonlinear modeling tool, artificial neural networks have also be studied for both nonlinear modeling of PAs [11] [12] and nonlinear equalization [13] [14] [15].

One of the major design goals for the 5G systems is to make the communication systems more power efficient. This needs efficient PAs, which is unfortunately more challenging since 5G signals have much higher PAPR and wider bandwidth [16] [17]. This is a especially severe problem for cost and battery limited devices in massive machine-type communications or internet of things (IoT). Existing nonlinear PA mitigation strategies may not be sufficient enough. We can reduce PAPR to some extent only. DPD is too complex and costly for small and low-cost 5G devices. Existing DPD and equalization techniques have moderate nonlinear distortion compensation capabilities only.

As a matter of fact, the nonlinear equalization strategy is more attractive to massive MIMO and millimeter wave transmissions due to the large number of PAs needed [18] [19] [20]. Millimeter wave transmissions require much higher transmission power to compensate for severe signal attenuation. Considering the extremely high requirement on PA power efficiency and the large number of PAs in a transmitter, the current practice of using DPD may not be appropriate due to implementation complexity and cost.

In this paper, we develop a novel deep neural network (DNN)-based nonlinear equalizer to equalize the PA distorted signals at the receiver. We exploit the Volterra series nonlinearity model to construct the input features of the DNN, which can help the DNN converge rapidly to the desired nonlinear response under limited training data and training time. Conventionally, Volterra series and neural networks are studied as two separate techniques for nonlinear PAs [2]. By integrating these two techniques together and exploiting the recent advancements in DNN, we can develop nonlinear equalizers with superior equalization performance, which is demonstrated in this paper by simulations and experiments.

The remainder of this paper is organized as follows. Nonlinear PA models are introduced in Section II. In Section III, the DNN-based nonlinear equalization scheme is presented. Experiments and conclusions are given in Sections IV and V, respectively.

II. Nonlinear Power Amplifier Models

Consider the baseband discrete model of the PA $y(n) = f(x(n), x(n-1), \cdots)$, where $x(n)$ is the input signal, $y(n)$ is the output signal, and $f(\cdot)$ is some nonlinear function. The simplest nonlinear PA model is the "AM-AM AM-PM" model. Let the amplitude of the input signal be $V_x = E[|x(n)|]$, where $E[\cdot]$ denotes short-term expectation or average. The output sample $y(n)$'s amplitude $V_y = E[y(n)]$ and additional phase change $\psi_y = E[\angle y(n)]$ depend on V_x in nonlinear ways as

$$V_y = \frac{gV_x}{\left(1 + \frac{gV_x}{c}\right)^{\frac{1}{2\sigma}}}, \quad \psi_y = \frac{\alpha V_x^p}{1 + \left(\frac{V_x}{\beta}\right)^q}, \tag{1}$$

where g is the linear gain, σ is the smoothness factor, and c denotes the saturation magnitude of the PA. Typical examples of these parameters are $g = 4.65$, $\sigma = 0.81$, $c = 0.58$, $\alpha = 2560$, $\beta = 0.114$, $p = 2.4$, and $q = 2.3$, which are used in the PA models regulated by IEEE 802.11ad task group (TG) [10].

More accurate models should take into consideration the fact that nonlinearity leads to memory effects. In this case, Volterra series are typically used to model PAs [4] [21]. A general model is [5]

$$y(n) = \sum_{d=0}^{D} \sum_{k=1}^{P} b_{kd} x(n-d) \left| x(n-d) \right|^{k-1} \tag{2}$$

with up to Pth order nonlinearity and up to D step memory.

It can be shown that only odd-order nonlinearity (i.e., odd k) is necessary because even-order nonlinearity falls outside of the passband and will be filtered out by the receiver bandpass filters [2]. To illustrate this phenomenon, we can consider some simple examples where the input signal $x(n)$ consists of a few single frequency components only. Omitting the memory effects, if $x(n)$ is a single frequency signal, i.e., $x(n) = V_0 \cos(a_0 + \phi)$, where $a_0 = 2\pi f_0 n$. Then, the output signal can be written as

$$\begin{aligned} y(n) = {} & c_1 V_0 \cos(a_0 + \phi + \psi_1) + \\ & \left(\frac{3}{4} c_3 V_0^3 + \frac{5}{8} c_5 V_0^5\right) \cos(a_0 + \phi + \psi_3 + \psi_5) && (3) \\ & + \frac{1}{2} c_2 V_0^2 + \frac{3}{8} c_4 V_0^4 && (4) \\ & + \left(\frac{1}{2} c_2 V_0^2 + \frac{1}{2} c_4 V_0^4\right) \cos(2a_0 + 2\phi + 2\psi_2 + 2\psi_4) \cdots && (5) \end{aligned}$$

where the first line (3) is the inband response with AM-AM & AM-PM nonlinear effects, the second line (4) is the DC bias, and the third line (5) includes all the higher frequency harmonics. At the receiving side, we may just have (3) left because all the other items will be canceled by bandpass filtering.

If $x(n)$ consists of two frequencies, i.e., $x(n) = V_1 \cos(a_1 + \phi_1) + V_2 \cos(a_2 + \phi_2)$, where $a_i = 2\pi f_i n$, then the inband response includes many more items, such as the first order items $c_i V_i \cos(a_i + \phi_i + \psi_i)$, the third order items $c_3(V_i^3 + V_i V_j^2) \cos(a_i + \phi_i + \psi_i)$, the fifth order items $c_5(V_i^5 + V_i V_j^4 + V_i^3 V_j^2) \cos(a_i + \phi_i + \psi_i)$, for $i, j \in \{1, 2\}$. There are also intermodulation items that consist of $na_i \pm ma_j$ as long as they are within the passband of the bandpass filter, such as $(V_i^2 V_j + V_i^2 V_j^3 + V_1^4 V_j) \cos(2a_i - a_j + 2\phi_i - \phi_j + 2\psi_i - \psi_j)$. There are many other higher order items with frequencies na_i, $n(a_i \pm a_j)$, or $na_i + ma_j$, that can not pass the passband filter. One of the important observations is that the contents that can pass the passband filter consist of odd-order nonlinearity only.

If $x(n)$ consists of three or more frequencies, we can have similar observations, albeit the expressions are more complex. Let the input signal $x(n)$ be

$$x(n) = \sum_{i=1}^{3} V_i \cos(a_i), \quad a_i = 2\pi f_i n. \tag{6}$$

Based on [22], the nonlinear distorted output response $y(n) = f(x(n))$ can be written as

$$y(n) = \sum_{i=1}^{\infty} k_i x^i(n), \tag{7}$$

where k_i represents the gain coefficients for the ith order components. The 1st order component is simply $k_1 x(n)$. The 2nd order component includes the DC component, the sum/difference of beat components, and the second-order harmonic components. Specifically,

$$k_2 x^2(n) = g_{2,0} + g_{2,1}(n) + g_{2,2}(n), \tag{8}$$

where

$$\begin{aligned} g_{2,0} &= \sum_{i=1}^{3} V_i^2 / 2, \\ g_{2,1}(n) &= \sum_{i=1}^{3} \sum_{j \neq i} V_i V_j \cos(a_i \pm a_j), \\ g_{2,2}(n) &= \sum_{i=1}^{3} V_i^2 \cos(2a_i)/2. \end{aligned}$$

The 3rd order component includes the third-order harmonic components $g_{3,1}(n)$, the third intermodulation beat components $g_{3,2}(n)$, the triple beat components $g_{3,3}(n)$, the self-compression/expansion components $g_{3,4}(n)$, and the cross-compression/exansion components $g_{3,5}(n)$. Details are skipped due to page limits.

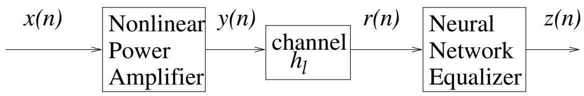

Fig. 1: System block diagram with nonlinear power amplifier and deep neural network equalizer.

III. DNN-BASED NONLINEAR EQUALIZATION

A. Nonlinear equalizer models

To mitigate the PA nonlinear distortions, we can apply nonlinear equalizers at the receivers. Obviously, we can still use the Volterra series model to analyze the response of nonlinear equalizers. One of the differences from (2) is that the even order nonlinearity may still be included and may increase the nonlinear mitigation effects [5].

Consider the system block diagram of nonlinear equalization shown in Fig. 1. Let the received signal be

$$r(n) = \sum_{\ell=0}^{L} h_\ell y(n-\ell) + v(n), \tag{9}$$

where h_ℓ is the finite-impulse response (FIR) channel coefficients and $v(n)$ is additive white Gaussian noise (AWGN). With the received sample sequence $r(n)$, a nonlinear equalizer will generate $z(n)$ as the estimated symbols.

If the PA has only slight nonlinearity as modeled by the simple "AM-AM AM-PM" model (1), we can stack received samples $r(n)$ together into $M+1$ dimensional vectors $\mathbf{r}(n) = [r(n), \cdots, r(n-M)]^T$, where $(\cdot)^T$ denotes transpose, and write the received samples in vector form as

$$\mathbf{r}(n) = \mathbf{H}\mathbf{G}(n)\mathbf{x}(n) + \mathbf{v}(n), \tag{10}$$

where $\mathbf{H}$ is an $(M+1)\times(M+L+1)$ dimensional channel matrix

$$\mathbf{H} = \begin{bmatrix} h_0 & \cdots & h_L & & \\ & \ddots & & \ddots & \\ & & h_0 & \cdots & h_L \end{bmatrix} \tag{11}$$

and $\mathbf{G}(n) = \mathrm{diag}\{V_{y(n)}e^{j\psi_{y(n)}}, \cdots, V_{y(n-M-L)}e^{j\psi_{y(n-M-L)}}\}$ is an $(M+L+1)\times(M+L+1)$ dimensional diagonal matrix which consists of the nonlinear PA responses, $\mathbf{x}(n) = [x(n), \cdots, x(n-M-L)]^T$, and $\mathbf{v}(n) = [v(n), \cdots, v(n-M)]^T$. To equalize the received signal, we apply a nonlinear equalizer with the form

$$\mathbf{f}^T = G'(n)[f_0, \cdots, f_M] \tag{12}$$

where $[f_0, \cdots, f_M]\mathbf{H} \approx [0, \cdots, 1, \cdots, 0]$ is to equalize the propagation channel, and $G'(n) \approx \frac{1}{V_{y(n-d)}}e^{-j\psi_{y(n-d)}}$ is to equalize the nonlinear PA response. Let $\hat{r}(n)$ be the output of the first linear equalization step. The second nonlinear equalization step can be implemented as a maximum likelihood estimation problem, i.e., $z(n) = \arg\min_{\forall x(n)} |\hat{r}(n) - V_y e^{j\psi_y}x(n)|^2$. This gives the output

$$z(n) = \mathbf{f}^T\mathbf{r}(n) \approx x(n-d) \tag{13}$$

with certain equalization delay d.

Both the channel coefficients h_ℓ and the nonlinear PA responses V_y, ψ_y can be estimated via training. So does the channel equalizer $\mathbf{f}^T$. Because the PA nonlinearity is significant for large signal amplitude only, we can apply small-amplitude training signals $x(n)$ first to estimate the channel h_ℓ and the channel equalizer $[f_0, \cdots, f_M]$. We can then remove the channel $\mathbf{H}$ from (10) with the first step linear channel equalization. Because the matrix $\mathbf{G}(n)$ is diagonal, we can easily estimate $\mathbf{G}(n)$ with regular training and then estimate the transmitted symbols as outlined in (13).

For more complex nonlinear PA responses, such as (2), we can conduct channel equalization similarly as (12). First, we can still apply small-amplitude training signals to estimate $[f_0, \cdots, f_M]$ so as to equalize the channel h_ℓ. This linear channel equalization step gives $\hat{r}(n) \approx y(n)$. We can then focus on studying the equalization of nonlinear distortion of PA, which can in general be conducted with the maximum likelihood method,

$$\{\hat{x}(n): n = 0, \cdots, N\} = \arg\min_{\{x(n)\}} \sum_{n=0}^{N} |\hat{r}(n) - \hat{y}(n)|^2, \tag{14}$$

where $\hat{r}(n)$ is the sequence after the linear channel equalization, $\hat{y}(n)$ is the sequence reconstructed by using the sequence $x(n)$ and the nonlinear PA response parameters b_{kd} based on (2), and N is the total number of symbols. The optimization problem (14) can be solved with the Viterbi sequence estimation algorithm if the memory length of the PA is small enough and the PA nonlinear response is known to the receiver.

In case the PA nonlinear response can not be estimated, the equalization of nonlinear PA response is challenging. In this case, one of the ways is to use the conventional Volterra series equalizer, which approximates $G'(n)$ with a Volterra series model. Similar to (2), this gives

$$z(n) = \sum_{d=0}^{D}\sum_{k=1}^{P} g_{kd}\hat{r}(n-d)\left|\hat{r}(n-d)\right|^{k-1}. \tag{15}$$

The objective of the Volterra series equalizer design is to design g_{kd} such that $z(n) \approx x(n-\ell)$ for some equalization delay ℓ.

Similarly as the DPD design of [5], based on the Volterra series model (15), we can estimate the coefficients g_{kd} by casting the estimation into a least squares problem

$$\min_{\{g_{kd}\}} \sum_{n=L}^{N} \left| x(n-L) - \sum_{d=0}^{D}\sum_{k=1}^{P} g_{kd}\hat{r}(n-d)|\hat{r}(n-d)|^{k-1} \right|^2, \tag{16}$$

with training symbols $x(n)$ and received samples $\hat{r}(n)$. Note that only the coefficients g_{kd} are needed to be estimated, and these coefficients are linear with respect to $\hat{r}(n)$ and $x(n)$. Define the vector $\mathbf{a} = [g_{00}, g_{01}, \cdots, g_{PD}]^T$, and the vector

$\mathbf{x} = [x(0), \cdots, x(N-L)]^T$. Define the $(N-L+1) \times DP$ data matrix

$$\mathbf{B} = \begin{bmatrix} \hat{r}(L) & \hat{r}(L)|\hat{r}(L)| & \cdots & \hat{r}(L-D)|\hat{r}(L-D)|^{P-1} \\ \vdots & & & \vdots \\ \hat{r}(N) & \hat{r}(N)|\hat{r}(N)| & \cdots & \hat{r}(N-D)|\hat{r}(N-D)|^{P-1} \end{bmatrix}. \tag{17}$$

Then (16) becomes

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{B}\mathbf{a}\|^2. \tag{18}$$

The solution to (18) is

$$\mathbf{a} = \mathbf{B}^{+}\mathbf{x}, \tag{19}$$

where $\mathbf{B}^{+} = (\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}$ is the pseudo-inverse of the matrix $\mathbf{B}$. From (19), we can obtain the Volterra series equalizer coefficients g_{kd}.

One of the major problems for the Volterra series equalizer is that it is hard to determine the order sizes, i.e., the values of D and P. Even for a nonlinear PA with slight nonlinear effects (i.e., small D and P in (2)), the length of D and P for Volterra series equalizer may be extremely long in order for (15) to have sufficient nonlinearity mitigation capability.

A potential way to resolve this problem is to apply artificial neural networks to fit the nonlinear equalizer response (15). Neural networks can fit arbitrary nonlinearity and can realize this with potentially small sizes. Nevertheless, in conventional neural network equalizers such as [14] [15], the input (features) to the neural networks was simply a time-delayed vector $[r(n), \cdots, r(n-M)]$. Although neural networks may have the capability to learn the nonlinear effects specified in (15), in practice the training may not necessarily converge to the desirable solutions due to local minimum and limited training data. In addition, conventional neural network equalizers were all feed-forward networks with fully connected layers only, which often suffer from problems like shallow network architecture and over-fitting.

B. Volterra-based DNN equalizer

We propose to use deep neural network to implement the nonlinear equalizer in the receiver, which can mitigate the nonlinear effects of the received signals due to not only PAs but also nonlinear channels and propagations. The architecture of the DNN equalizer is shown in Fig. 2. Different from [10], we use multi-layer convolutional neural networks (CNNs). Different from conventional neural network predistorters proposed in [6], we use neural networks as equalizers at the receivers. Different from conventional neural network equalizers such as those proposed in [14] [15], in our DNN equalizer, we use CNN and the input features in X are not only the linear delayed samples $r(n)$. But rather, we apply the Volterra series models to create input features.

To simplify presentation, according to Section III-A, we assume that the linear channel $\mathbf{H}$ has already been equalized by a linear equalizer, whose the output signal is $r(n)$. According to

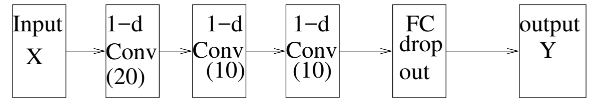

Fig. 2: Block diagram of DNN equalizer.

Volterra series representation of nonlinear functions, the input-output response of the nonlinear equalizer can be written as

$$z(n) = \sum_{k=1}^{P} \sum_{d_1=0}^{D} \cdots \sum_{d_k=0}^{D} f_{d_1,\cdots,d_k} \prod_{i=1}^{k} r(n-d_i). \tag{20}$$

One of major problems is that the number of coefficients $f_{d_1,\cdots,d_k}$ increases exponentially with the increase of memory length D and nonlinearity order P. There are many different ways to develop more efficient Volterra series representations with reduced number of coefficients. For example, in [23], the authors exploit the fact that higher-order terms do not contribute significantly to the memory effects of PAs to reduce the memory depth d when the nonlinearity order k increases. This technique can drastically reduce the total number of coefficients. In [24] [25] [26], the authors developed the dynamic deviation model to reduce the full Volterra series model (20) to the following simplified one

$$\begin{aligned} z(n) =& z_s(n) + z_d(n) = \sum_{k=1}^{P} f_{k,0} r^k(n) \\ &+ \sum_{k=1}^{P}\sum_{j=1}^{k} r^{k-j}(n) \sum_{d_1=0}^{D} \cdots \sum_{d_j=d_{j-1}}^{D} f_{k,j} \prod_{i=1}^{j} r(n-d_i) \end{aligned} \tag{21}$$

where $z_s(n)$ is the static term, and $z_d(n)$ is the dynamic term that includes all the memory effects. We can see that the total number of coefficients can be much reduced by controlling the dynamic order j which is a selectable parameter.

We construct the input features of the DNN based on the model (21). Corresponding to the static term $z_s(n)$, we change it to

$$\hat{z}_s(n) = \sum_{1\le k\le P} f_{k,0} r(n) \left|r(n)\right|^{k-1}. \tag{22}$$

The reason that (22) changes $r^k(n)$ to $r(n)|r(n)|^{k-1}$ is that only the signal frequency within the valid passband is interested. This means the input feature vector X should include terms $r(n)|r(n)|^{k-1}$. Similarly, corresponding to the dynamic term $z_d(n)$, we need to supply $r^{k-j}(n)\prod_{i=1}^{j} r(n-d_i)$ in the features where half of the terms $r(n)$ and $r(n-d_i)$ should be conjugated. For simplicity, in our DNN equalizer, the vector X includes $r(n-q)|r(n-q)|^{k-1}$ for some q and k.

By applying Volterra series components directly as features of the input X, the DNN can develop more complex nonlinear functions with less number of hidden layers and less number of neurons. This will also make the training procedure converge much faster with much less training data.

In Fig. 2, the input X is a tensor formed by the real and imaginary parts of $r(n-q)|r(n-q)|^{k-1}$ with appropriate number of delays q and nonlinearities k. There are three one-dimension convolutional layers, each with 20 or 10 feature maps. After a drop-out layer for regularization, this is followed by a fully connected layer with 20 neurons. Finally there is a fully-connected layer to form the output tensor Y which has two dimensions. The output Y is used to construct the complex $z(n)$, where $z(n) = \hat{x}(n-d)$ for some appropriate delay d. All the convolutional layers and the first fully connected layer use the sigmoid activation function, while the output layer uses the linear activation function. We use the mean square error loss function $L_{loss} = E[|x(n-d) - z(n)|^2]$, where $z(n)$ is replaced by Y and $x(n-d)$ is replaced by training data labels.

IV. Experiment Evaluations

In this section, we present our experiments on applying the Volterra series based DNN equalizer (**Volterra+NN**) for nonlinear PA equalization. We compared the proposed scheme with the following equalization methods: a Volterra series based equalizer (**Volterra**) and a conventional time-delay neural network equalizer (**NN**). The performance metrics are mean square error (MSE) $\sqrt{E[|z(n) - x(n-d)|^2]/E[|x(n-d)|^2]}$ and symbol error rate (SER).

We used both simulated signals and real measurement signals. To generate simulated signals, we used a Doherty nonlinear PA model consists of 3rd and 5th order nonlinearity. Referring to (2), the coefficients $b_{k,q}$ were

$$\begin{aligned}&1.0513 + .0904j, -.068 - .0023j, .0289 - .0054j,\\&-.0542 - .29j, .2234 + .2317j, -.0621 - .0932j,\\&-.9657 - .7028j, -.2451 - .3735j, .1229 + .1508j,\end{aligned}$$

which was used in [5] to simulate a 5th order dominant nonlinear distortion derived from PA devices used in the satellite industry. For real measurement, our measurement signals were obtained from PA devices used in the cable TV (CATV) industry, which are typically dominated by 3rd order nonlinear distortion (NLD). Various levels of nonlinear distortion, in terms of dBc, were generated by adjusting the PAs.

For the Volterra equalizer, we approximate the response of the nonlinear equalizer with delays including 8 pre- and post-main taps and with nonlinearities including even and odd order nonlinearity up to the 5th order. To determine the values of the Volterra coefficients, we transmitted $N = 4,096$ training symbols through the PA and then collected the noisy received samples $r(n)$.

For the conventional time-delay NN equalizer, we applied a feedforward neural network with 80-dimensional input vector X and 5 fully-connected hidden layers with 20, 20, 10, 10, 10 neurons, respectively.

Fig. 3 shows the constellation and MSE of the equalizers outputs. It can be seen that the proposed scheme provides the best performance.

Fig. 4 shows the constellation of 16 QAM equalization over the real PA. The corresponding SER were 0.0067, 0.0027, 0.00025, respectively. It can be seen that the proposed Volterra+NN scheme has the best performance.

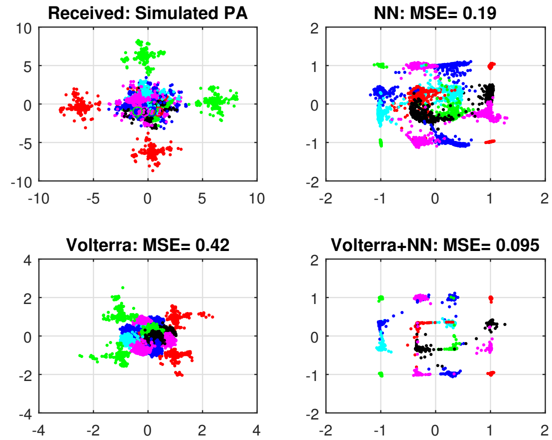

Fig. 3: Constellation of 16QAM over a simulated PA. Top-left: received signal. Bottom-left: Volterra equalizer output. Top-right: time-delayed NN output. Bottom-right: proposed Volterra+NN output.

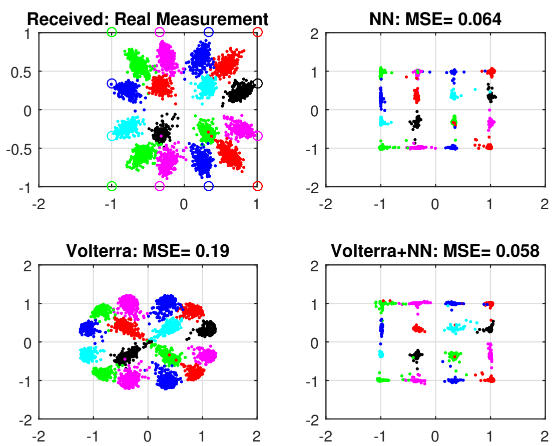

Fig. 4: Constellation of 16QAM over a real PA. Top-left: received signal. Bottom-left: Volterra equalizer output. Top-right: time-delayed NN output. Bottom-right: proposed Volterra+NN output.

Fig. 5 provides MSE measurements for 16-QAM under various nonlinear distortion level dBc. For each 1 dB increase in NLD, the resultant MSE is shown for the "Measured", "Volterra", "NN", and the proposed "Volterra+NN" cases. MSE reduction diminishes appreciably as modulation order increases from QPSK to 64-QAM, but small improvements in MSE have been observed lead to appreciable SER improvement, especially for more complex modulation orders. Unfortunately, $4,096$ symbol sample sizes have limited our measurements to a minimum measurable 0.000244 SER, which represents 1 symbol error out of $4,096$ symbols.

Fig. 6 summarizes equalization performance, which shows the averaged percent reduction/improvement in MSE and SER

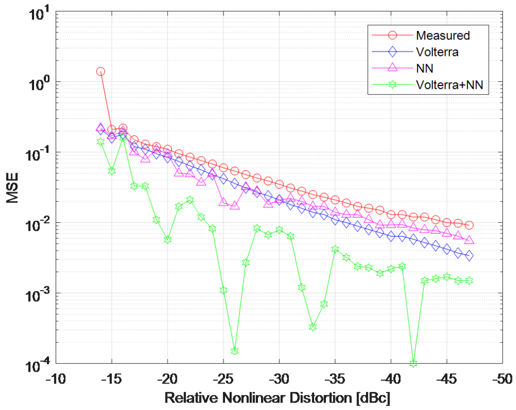

Fig. 5: Comparing three equalization methods for 16-QAM under various NLD levels.

from the NLD impaired data for multiple modulation orders.

	Volterra		NN		Volterra+NN	
	MSE	SER	MSE	SER	MSE	SER
64-QAM	16%	26%	10%	25%	42%	44%
16-QAM	41%	2%	35%	6%	85%	28%
QPSK	57%	0%	100%	0%	100%	0%
AVERAGE	38%	9%	48%	10%	76%	24%

Fig. 6: Comparing MSE/SER improvement in percentage for the three equalization methods. Note that 0% SER improvement for QPSK was because the received signal's SER was already very low.

V. Conclusions

This paper develops a new nonlinear equalization scheme by integrating the Volterra series nonlinear model with deep neural networks. The Volterra series based deep neural network equalizer yields promising results over conventional nonlinear equalization approaches in mitigating nonlinear power amplifier distortions. It can be potentially useful for many 5G communication scenarios.

References

[1] J.-A. Lucciardi, P. Potier, G. Buscarlet, F. Barrami, and G. Mesnager, "Non-linearized amplifier and advanced mitigation techniques: Dvb-s2x spectral efficiency improvement," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–7.

[2] J. Wood, *Behavioral modeling and linearization of RF power amplifiers*. Artech House, 2014.

[3] C.-L. Wang and Y. Ouyang, "Low-complexity selected mapping schemes for peak-to-average power ratio reduction in ofdm systems," *IEEE Transactions on signal processing*, vol. 53, no. 12, pp. 4652–4660, 2005.

[4] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, no. 23, pp. 1417–1418, 2001.

[5] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Transactions on communications*, vol. 52, no. 1, pp. 159–165, 2004.

[6] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 1, pp. 95–104, 2010.

[7] S. Dimitrov, "Non-linear distortion cancellation and symbol-based equalization in satellite forward links," *IEEE Trans Wireless Commun*, vol. 16, no. 7, pp. 4489–4502, 2017.

[8] D. J. Sebald and J. A. Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3217–3226, 2000.

[9] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Transactions on neural networks*, vol. 4, no. 4, pp. 570–590, 1993.

[10] B. Li, C. Zhao, M. Sun, H. Zhang, Z. Zhou, and A. Nallanathan, "A bayesian approach for nonlinear equalization and signal detection in millimeter-wave communications," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3794–3809, 2015.

[11] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for rf power amplifier behavioral modeling and digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 4, pp. 913–923, 2011.

[12] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic behavioral modeling of 3g power amplifiers using real-valued time-delay neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 1025–1033, 2004.

[13] M. Ibnkahla, "Applications of neural networks to digital communications–a survey," *Signal processing*, vol. 80, no. 7, pp. 1185–1215, 2000.

[14] D.-C. Park and T.-K. J. Jeong, "Complex-bilinear recurrent neural network for equalization of a digital satellite channel," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 711–725, 2002.

[15] A. Uncini, L. Vecci, P. Campolucci, and F. Piazza, "Complex-valued neural networks with adaptive spline activation function for digital-radio-links nonlinear equalization," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 505–514, 1999.

[16] M. S. Sim, M. Chung, D. Kim, J. Chung, D. K. Kim, and C.-B. Chae, "Nonlinear self-interference cancellation for full-duplex radios: From link-level and system-level performance perspectives," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 158–167, 2017.

[17] I. Yoffe and D. Wulich, "Predistorter for mimo system with nonlinear power amplifiers," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3288–3301, 2017.

[18] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive mimo," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 6478–6482.

[19] H. Yan and D. Cabric, "Digital predistortion for hybrid precoding architecture in millimeter-wave massive mimo systems," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3479–3483.

[20] C. Mollén, E. G. Larsson, and T. Eriksson, "Waveforms for the massive mimo downlink: Amplifier efficiency, distortion, and performance," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5050–5063, 2016.

[21] A. Cheaito, M. Crussière, J.-F. Hélard, and Y. Louët, "Quantifying the memory effects of power amplifiers: Evm closed-form derivations of multicarrier signals." *IEEE Wireless Commun. Letters*, vol. 6, no. 1, pp. 34–37, 2017.

[22] K. Simons, *Technical Handbook for CATV Systems, 3rd Edition*. Jerrod Publication No. 436-001-01, 1968.

[23] J. Staudinger, J.-C. Nanan, and J. Wood, "Memory fading volterra series model for high power infrastructure amplifiers," in *Radio and Wireless Symposium (RWS), 2010 IEEE*. IEEE, 2010, pp. 184–187.

[24] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of rf power amplifiers," *IEEE Transactions on microwave theory and techniques*, vol. 54, no. 12, pp. 4323–4332, 2006.

[25] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Open-loop digital predistorter for rf power amplifiers using dynamic deviation reduction-based volterra series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 7, pp. 1524–1534, 2008.

[26] L. Guan and A. Zhu, "Simplified dynamic deviation reduction-based volterra model for doherty power amplifiers," in *Integrated Nonlinear Microwave and Millimetre-Wave Circuits (INMMIC), 2011 Workshop on*. IEEE, 2011, pp. 1–4.