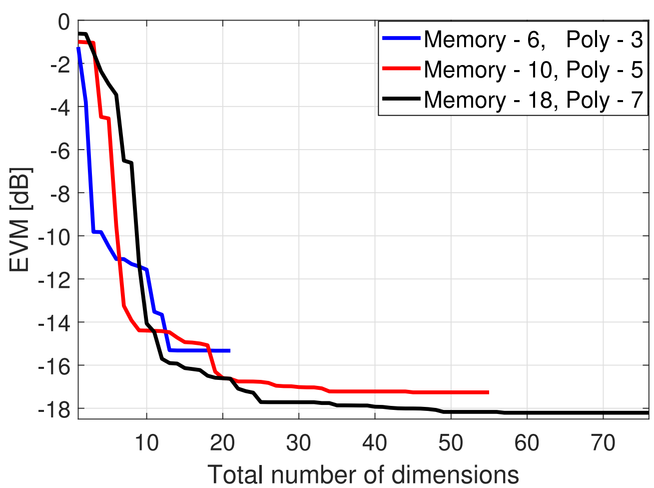
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Dimensions-Reduced Volterra Digital Pre-Distortion Based On Orthogonal Basis for Band-Limited Nonlinear Opto-Electronic Components

Volume 11, Number 1, February 2019

Hananel Faig
Yaron Yoffe
Eyal Wohlgemuth
Dan Sadot



DOI: 10.1109/JPHOT.2019.2892400

IEEE photonics SOCIETY

IEEE

Dimensions-Reduced Volterra Digital Pre-Distortion Based On Orthogonal Basis for Band-Limited Nonlinear Opto-Electronic Components

Hananel Faig,[1,2] **Yaron Yoffe**,[1] **Eyal Wohlgemuth**,[1] **and Dan Sadot**[1]

[1]Department of Electrical and Computer Engineering, Ben-Gurion University, Beershebe 8499000, Israel
[2]Mellanox Technologies Ltd., Sunnyvale, CA 94085 USA

DOI:10.1109/JPHOT.2019.2892400

Manuscript received December 18, 2018; accepted December 30, 2018. Date of publication January 11, 2019; date of current version February 1, 2019. This work was supported by the PetaCloud consortium and Mellanox Technologies Ltd. Corresponding author: Hananel Faig (e-mail: faig@post.bgu.ac.il).

Abstract: Optical communication systems often include nonlinear components, such as Mach–Zehnder modulators (MZMs). The components' nonlinearity may have a deleterious effect on the system performance, especially when the memory effect is included. One of the most common methods to mitigate the resultant distortion effects is digital predistortion (DPD) based on the Volterra polynomial model. Typically, the DPD coefficients are extracted using the least squares approach. However, naive implementation of the Volterra polynomial model usually introduces significant complexity due to the large number of model coefficients. Here, we propose the use of orthogonal polynomial basis functions for efficient DPD implementation. The orthogonal basis enables the estimation of each coefficient separately, which provides a significant computational gain. Furthermore, the amount of dominant terms in the orthogonal basis representation is significantly lower, which leads to dramatic relaxation in the DPD implementation. In addition to the analytical modeling approach, a nonparametric method is developed by applying eigenvalue decomposition on the correlation matrix, which is required for the case of dependent random variables, or for the case of unknown probability distribution functions. MZM-based lab experiments and simulations were performed, which indicated a potential saving of 50%–80% in the amount of useful dimensions.

Index Terms: Fiber optics communication systems, Volterra series, nonlinear digital pre-distortion, orthogonal basis.

1. Introduction

The commonly used models for the components of a communication system are linear. However, in most practical short reach optical communication systems for data center connections (sub-100 km), nonlinear components are included, such as the vertical-cavity surface-emitting laser (VCSEL) at low or high current, as described in its current-power (I-L) curve [1], as well as a Mach-Zehnder modulator (MZM) which has a sine voltage-power (V-L) curve [2]. The difficulties that arise from ignoring this aspect are especially severe when memory effects are included in the model [3]. Therefore, solutions have been sought to overcome the resultant distortions.

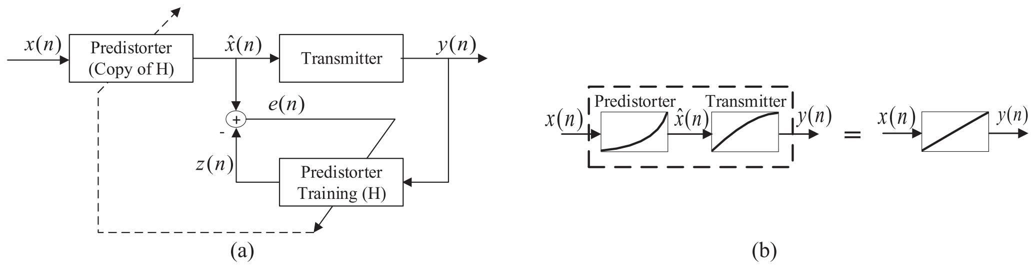


Fig. 1. (a) Block diagram of DPD learning concept. (b) The conversion functions when the DPD is used. The linear relationship can be achieved using the inverse conversion function.

In the field of communication systems there has been widespread research on linearization usign either digital pre-distortion (DPD) at the transmitter or digital post-distortion at the receiver. The DPD approach has some significant advantages over the digital post-distortion approach [4]:

1) The SNR at the transmitter is significantly higher than that at the receiver. Hence, the coefficients estimation is more accurate.
2) The Volterra coefficients estimation is much more challenging at the receiver side, as additional distortions and noises mechanisms are involved.
3) In some cases, the nonlinear effect only occurs at the transmitter. Therefore, it is easier to deal with the resultant distortion before the signal is further distorted by additional system effects.

The DPD concept is based on finding the system's inverse function. Then, the signal is digitally pre-distorted such that the concatenation of DPD and system distortion is linear. Fig. 1 presents a DPD system that is optimized based on the indirect learning approach [5]. In the *learning phase*, the predistorter is bypassed; thus, $\widehat{x}(n) = x(n)$ and the "predistorter training" block learns the transmitter inverse system to achieve $z(n) \rightarrow x(n)$, where $z(n)$ is the reconstructed signal, based on the output measurements. $e(n)$ is the error signal between the reconstructed signal and the original one. In the *operational phase*, the transmitter inverse system is copied to the "predistorter" block to achieve $y(n) \rightarrow x(n)$. In this phase, $\widehat{x}(n)$ is the DPD output that is being transmitted. The DPD calculation process might be continuous based on the back channel (for unstable systems), or "once in a lifetime" (for stable systems).

One common DPD approach is based on the Volterra series, or one of its degenerated versions, such as the memory polynomials model, which can be used to describe a wide variety of non-linear systems with memory. The Volterra series is a variant of a multimodal Taylor series, where each dimension represents a memory element. However, beacuse an analytical calculation of the coefficients is complicated, estimation methods are being used. The most common single-step estimation method is linear least squares (LS), which utilizes the fact that the Volterra model is linear with respect to its coefficients. However, this approach suffers from two major drawbacks:

1) **Non-orthogonality** – In most cases, the elements of the Volterra series are the elementary polynomial basis (x, x^2...). These elements are highly correlated. Therefore, a joint estimation of all the elements is required to avoid large estimation errors. Consequently, a single-step estimation leads to a large computational complexity.
2) **Condition Number** – The LS estimation method requires a matrix inversion. The numerical inversion of a matrix is affected by its condition number (CN). If the CN is large, the whole estimation will be inaccurate [6]. The CN of the Volterra matrix reaches typical values of CN $> 10^{20}$ [7]. Consequently, a fixed-point implementation may result in a severe degradation of the estimation accuracy.

In order to overcome the numerical instability of large CN systems orthogonal basis construction of the Volterra series matrix has been proposed [7]. It was shown that by using an orthogonal basis, the CN could be decreased by several order of magnitudes. This orthogonal basis approach was proposed for the case of a uniformly distributed one-dimensional (1-D) random variable (RV). Later [8], Raich proposed an orthogonal basis for a Gaussian distributed 1-D RV. Yang [9] proposed an orthogonal basis for a uniformly distributed 2-D independent RVs. Using Yang's method, a 1-D RV basis can be expanded for any X-D RVs case. However, the X-D RVs must be independent.

Additional methods for reducing the computational complexity have focused on degenerated Volterra series such as memory polynomials [10], a "near-diagonality" pruning model [11], etc. [12]. However, the memory polynomials performance may be limited in cases where nonlinear cross-terms are dominant [13], e.g. in systems that include a combination of reflections and nonlinearity. Other approaches have focused on reducing the calculation complexity by methods such as the use of parallel filters [14]. It should be noted that, grading by the series coefficient magnitudes is impossible, because of the high power elements of the Volterra series, e.g. if the dynamic range is [−5, 5], the high order moment elements will have large values, and, in turn, their coefficients will have small values, with the oppsite true in the case of [−0.5, 0.5].

In this work, the use of orthogonal polynomial basis functions for estimating nonlinear components with memory is proposed and extensively analyzed, following the preliminary publication of this method [15]–[16]. Significant advantages are outlined and discussed. Most importantly, it is shown that a dramatic complexity reduction may be achieved by dimensions selection and reduction. Furthermore, we propose a computational method for calculating the orthogonal basis in general cases such as for an unknown nonlinear distortion or multi-dimensional RVs that are dependent, which is relevant in the case of a system with memory. For such general cases, the use of a diagonalization method for the correlation matrix is proposed.

The rest of the paper is organized as follows. In Section 2, the major benefits of orthogonal basis usage and the process to find it using eigenvalue decomposition (EVD) are presented. Section 3 presents the simulation results for an MZM model. Section 4 presents the experimental results for an MZM. This is followed by a discussion and conclusions, which are drawn in Section 5.

2. Orthogonal Basis

An orthogonal basis for the purpose of CN reduction was proposed in [7]–[9]. However, an orthogonal basis offers additional advantages, as outlined below.

2.1 Parallel Least Squares

In its most general form, the real-valued discrete-time Volterra series can be written as follows:

$$x(n) = h_0 + \sum_{k=1}^{K} \sum_{t_1=0}^{M} \dots \sum_{t_k=0}^{M} h_{k(t_1 \dots t_k)} \prod_{j=1}^{k} y\left(n - t_j\right) \tag{1}$$

where K is the nonlinearity order and M is the memory order. For example, in the case of a system with a memory order of 1 and a nonlinearity order of 2, Eq. (1) degenerates to the following:

$$x(n) = h_0 + h_1 y(n) + h_2 y(n-1) + h_3 y^2(n) + h_4 y^2(n-1) + h_5 y(n) y(n-1) \tag{2}$$

In this example, the two RVs, $y(n)$, $y(n-1)$ yield six Volterra dimensions (series elements).

Following the series representation, the system and its inverse are linear with respect to the Volterra coefficients, and hence it can be described using a matrix form:

$$X = \boldsymbol{Y} H_{inverse} \tag{3}$$

where X is the input signal, $H_{inverse}$ is the vector of the inverse function coefficients and $\boldsymbol{Y}$ is the Volterra matrix of the system output samples y:

$$\boldsymbol{Y} = \begin{bmatrix} \prod_{m=0}^{M} y_{m,1}^{k_m} & \cdots & y_{-1,p} y_{0,1} & y_{-1,1} & y_{0,1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \prod_{m=0}^{M} y_{m,p}^{k_m} & \cdots & y_{-1,p} y_{0,p} & y_{-1,p} & y_{0,p} & 1 \end{bmatrix} \tag{4}$$

where P is the number of samples, $\sum_{m=0}^{M} k_m \leq K, y_{m,j} = y(n_j - m)$ is the output sample in time n and $j \in [1, P]$. The rows of $\boldsymbol{Y}$ represent the time samples, and the columns represent the Volterra elements. $\boldsymbol{Y}$ is a matrix of size $[P$ times $\rho]$, where ρ is number of elements representing all the possible combinations of the memory and polynomial orders including the cross-terms:

$$\rho = \sum_{k=0}^{K} \binom{k+M-1}{k} = \frac{(K+M)!}{K!M!} \tag{5}$$

Since the model is linear, the Volterra coefficients $H_{inverse}$ can be estimated using the LS estimation method:

$$\widetilde{H}_{inverse} = \left(\boldsymbol{Y}^T \boldsymbol{Y}\right)^{-1} \boldsymbol{Y}^T X \tag{6}$$

The columns of $\boldsymbol{Y}$ are the series elements. Therefore, under the assumption of ergodicity in the second moment, $\boldsymbol{Y}^T \boldsymbol{Y}$ is the correlation matrix. In a case where the orthogonal basis is being used, the correlation matrix should be diagonal. Thus, the LS formula degenerates to the following:

$$\widetilde{H}_{new} = \boldsymbol{\Lambda}^{-1} \boldsymbol{\theta}^T X \tag{7}$$

where $\widetilde{H}_{new}$ is the estimated coefficients vector, $\boldsymbol{\Lambda}^{-1}$ is a diagonal matrix of the pseudo inverse matrix, $\boldsymbol{\theta}^T$ is the orthogonal projection matrix and X is the input vector. Hence, we obtain the following:

$$\widetilde{H}_{new_q} = \boldsymbol{\Lambda}_{q,q}^{-1} \sum_{i=0}^{P} \boldsymbol{\theta}_{q,i}^T X_i \tag{8}$$

where q is the element index. This result indicates that it is possible to estimate any coefficient separately, which makes sense considering the assumption that with an orthogonal basis, the estimation error in any dimension is orthogonal to that in other dimensions. We now propose the use of the parallel LS (PLS) method, where each coefficient is estimated separately. Using an orthogonal basis, we can perform PLS with multiple scalar inversions instead of LS with one matrix inversion.

Using the PLS approach, and assuming that the orthogonal basis is known [17], the number of required resources is decreased, as shown in Appendix A. This significantly reduces the amount of computation required for the estimation of the series elements during the *learning phase*, e.g. bring up and tracking processes. Fig. 2(a) presents the required multiplications for the estimation of the Volterra coefficients. It is clearly observed that there is a significant amount of saving for a case with a large number of samples or series with a large number of elements.

2.2 Dimensions Reduction

The main implementation challenge with a Volterra-based DPD is its complexity [18]. The number of Volterra series elements grows quickly with the memory and polynomial orders [19]. The use of Volterra-based compensation in high-speed communication systems can be relevant only with efficient, real-time and low-power methods [20]–[21]. Hence, dimensions reduction methods are critical to allow effective pre-distorters.

One of the main challenges for dimensions reduction in the case of a non-orthogonal basis is identifying the dominant elements (dimensions) [22]. This challenge can be resolved using the orthogonal basis. Each dimension contribution is orthogonal to the other. Hence, the dimensions are graded, and the dominant dimensions are allocated and selected. Consequently, the total number of dominant dimensions is significantly reduced. This method enables the use of DPD in high-speed communication, as it dramatically reduces the real-time operations in the high-speed data path.

The basic dimension grading is based on the dimension variance [23]. Since the correlation between different dimensions is zero, the dimension's variance reflects the amount of information carried by each dimension. Thus, the dimensions with low variance can be neglected, resulting in a negligible effect on the DPD performance.

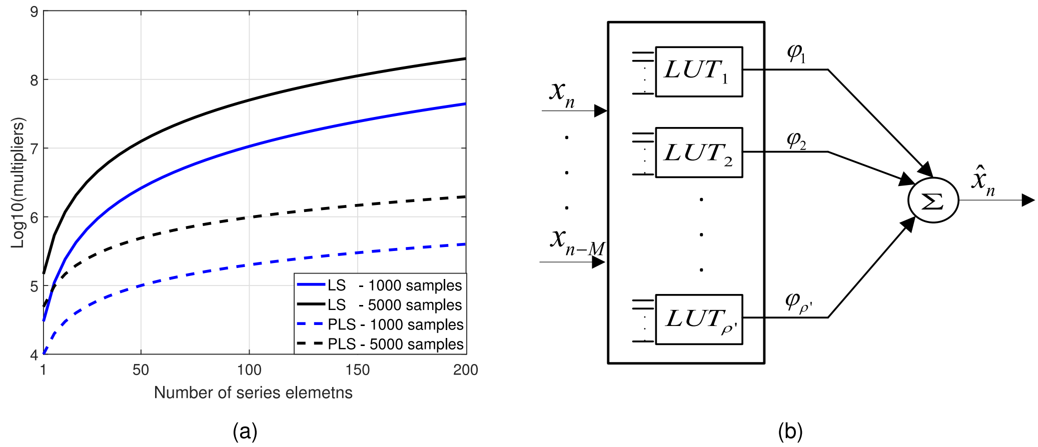


Fig. 2. (a) Number of multipliers required for estimation. (b) The DPD implementation. The LUT calculation can be performed "offline" by the EVD process. On the other hand, the ρ' LUT input-output translation and following summations should be implemented "online", when ρ' is the number of the most significant dimensions.

It should be noted that this method can be applied to any variant of the Volterra series including the memory polynomials case. In addition, the algorithm is signal-independent. Thus, it can be applied to higher-order modulation systems that use nonlinear equalizers, such as in [24]–[25].

2.3 Orthogonal Basis Estimation

As previously mentioned, earlier works proposed the use of an orthogonal basis for the Volterra series coefficients estimation. However, these methods were limited to specific distributions, i.e., uniform or Gaussian. These distributions are easier to analyze because they have well-known appropriate orthogonal polynomials. In addition, they are of interest because the input of a common high-speed digital communication system is uniformly distributed, and the system output is usually Gaussian, as it usually contains inter symbol interference (ISI).

Nevertheless, some interesting cases include different distributions. For example, in the case of a short memory channel, the output will be neither Gaussian nor uniform as the central limit theorem does not hold. Another example is the case of a nonlinear system where the output distribution obeys the transformation formula of the input distribution.

For any other distribution, an orthogonal polynomial basis for a 1-D RV can be analytically calculated by applying the Gram-Schmidt process [26], and an orthogonal basis for an X-D independent RVs can be analytically calculated using the Yang expansion [9]. Still, two main issues remain:

1) **Output distribution** - The analytical calculation requires a knowledge of the RV's probability distribution function (PDF). By the PDF transformation formula [27], the distribution after the nonlinear effect depends on the input distribution and system nonlinearity function whereas, the nonlinear distortion is usually unknown. Hence, the orthogonal basis cannot be calculated by the Gram-Schmidt process.
2) **Samples dependence** - Even if an orthogonal basis for a 1-D RV can be built, if there is dependency between the RVs, the Yang expansion to an X-D RV does not hold, as shown in Appendix B. In the DPD estimation process, the X-D RVs represent the system output samples, which are dependent on the system memory effect.

Here, we propose a method to evaluate the orthogonal basis for X-D dependent RVs by applying EVD on the correlation matrix [28]. This method is similar to the principal components analysis (PCA) process, which leads to an uncorrelated basis. However, it should be noted that the columns of the Volterra matrix (Eq. (4)) are non-zero mean. Therefore, applying PCA to this matrix will not lead to an orthogonal basis, because uncorrelated non-zero mean RVs are non-orthogonal.

The eigenvalues resulting from the EVD process are the second moments of the Volterra dimensions and the eigenvectors matrix is the linear mapping from the original Volterra basis to the new orthogonal basis.

It should be noted that for cases with an unknown orthogonal basis, the advantage of operation saving by the PLS is decreased as a result of the orthogonalization process [29]. However, the significant operation saving by the dimensions reduction remains.

2.4 DPD Implementation

Using all the three sub-sections above, a dimensions-reduced DPD can be applied using the following four steps:

1) *Step 1:* the Volterra matrix is built using the non-orthogonal basis.
2) *Step 2:* the eigenvalues and eigenvectors are found by the EVD on the correlation matrix.
3) *Step 3:* only the ρ' most significant dimensions are chosen, by their variances weights, and their rotation matrix is built using the appropriate eigenvectors.
4) *Step 4:* ρ' look-up tables (LUTs) of the projections on the orthogonal basis and their appropriate coefficients are built.

The use of LUTs is possible since the transmitted symbols are selected out of a finite dictionary set, e.g. a set of four in the case of PAM-4. The proposed DPD scheme and implementation block diagram are presented in Fig. 2(b) and summarized as follows:

$$\widehat{x} = \sum_{i=1}^{\rho'} \varphi_i = \sum_{i=1}^{\rho'} \widetilde{H}_{orth_i} \cdot (V(X)\boldsymbol{E}_i) \tag{9}$$

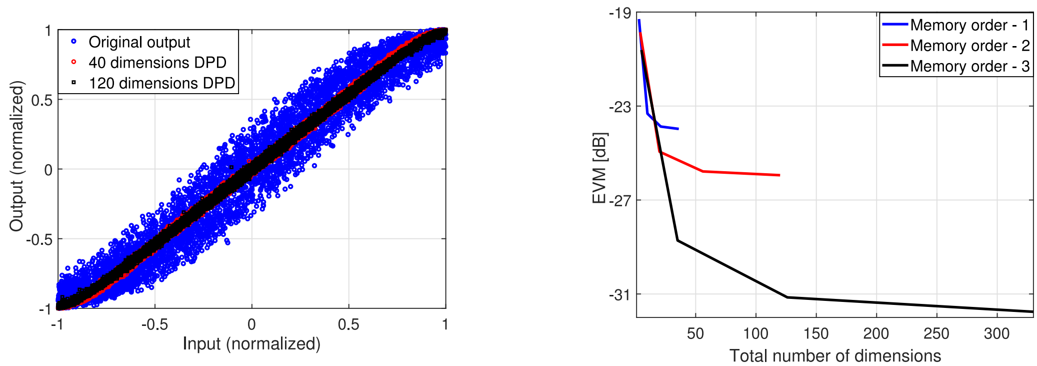
where the DPD output $\widehat{x}$ of input signal X is generated by the sum of all the weighted projections φ. Each projection consists of the inverse system non-orthogonal Volterra representation $V(X)$ projected on the orthogonal basis by the rotation matrix $\boldsymbol{E}$. The projections are weighted by $\widetilde{H}_{orth}$, which is the weights vector, estimated by the PLS process.

In Fig. 2(b), the LUT implements the operation $\widetilde{H}_{orth_i} \cdot (V(X)\boldsymbol{E}_i)$ of Eq. (9) "offline". The input to the LUT is the current symbol including all its M history elements $[x(n), \ldots, x(n-M)]$. The LUT output is defined by the number of the most significant dimensions ρ'. In the case of an input signal with a dictionary size of N, the total LUTs space is $N^M \rho'$, where N^M is the number of each LUT input combinations, and ρ' is the number of LUTs outputs. It should be emphasized that the LUT calculation is done "offline" during the tracking flow, while the LUT input-output translation and the following summations are performed "online" on the data path.

3. Simulation Analysis

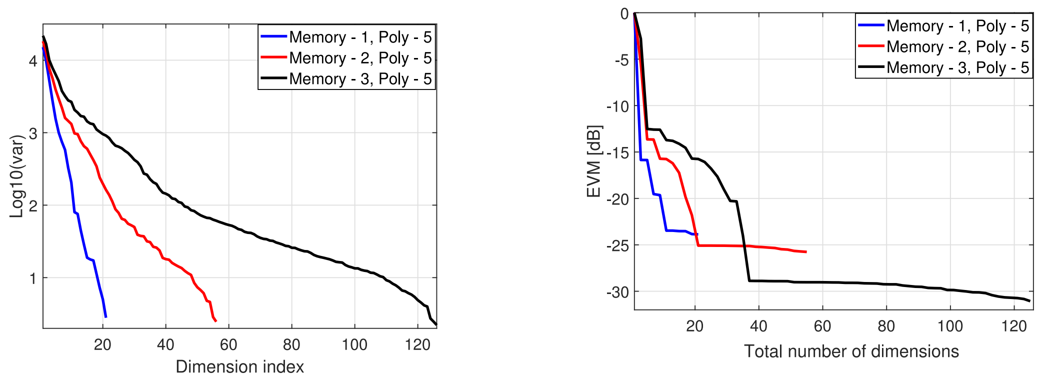
A common transmitter technology being used in coherent systems is based on MZM, which is inherently nonlinear due to its sinusoid conversion function, resulting from the interferometric process. In addition, in high-speed transmission systems, the MZM also introduces a memory effect as a result of its bandwidth-limitation [30]. In the simulation, the MZM model is based on the Wiener-Hammerstein methodology [31]. A set of numerical simulations was carried out by transmitting analog signals through an MZM input-output transfer function and analyzing the resulting distorted signals. For the case of an input signal limited by the range [−1, 1], the overall MZM conversion behavior is presented in Fig. 3(a) by the blue dots. The spreading around the sinusoid results from the memory effect generated by the finite impulse response (FIR) vector [1, 0.3, 0.15, 0.1, 0.05], which represents a typical band-limited transmitter.

For the MZM presented above, several sets of truncated Volterra series representations of the inverse system were estimated using both the elementary polynomial basis and the newly proposed orthogonal basis. The first set of truncated elementary Volterra series includes several combinations. Three memory order cases [1, 2, 3] were investigated, where for each memory order the polynomial order increases from 1 to 5, and therefore the total number of series elements



(a) (b)

Fig. 3. (a) MZM conversion behavior with and without DPD. (b) EVM versus the total number of dimensions. Longer series yields better EVMs.



(a) (b)

Fig. 4. (a) Dimensions variances versus the dimension index for the orthogonal basis case. (b) EVM versus total number of dimensions after dimension ordering.

is increased according to Eq. (5). Fig. 3(b) presents the error vector magnitude (EVM), which is associated with $e(n)$ of Fig. 1(a) during the *learning phase* (where $e(n)$ is defined as the difference between the non-distorted signal and the reconstructed signal after distortion), as a function of the total number of series elements. Longer series yield better EVMs. However, as the dimensions are non-orthogonal, the most significant dimensions cannot be identified, and therefore shorter DPD cannot be performed.

To select the appropriate polynomial order, the Taylor expansion of the MZM sinusoid was analyzed:

$$\sin(x) = \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \tag{10}$$

From Eq. (10), it can be seen that the series elements weights decrease by two orders of magnitude per polynomial order. Consequently, a polynomial order of five was selected.

In turn, the variances of the dimensions of the orthogonal basis were sorted and are presented in Fig. 4(a). Due to the fact that the dimensions are sorted in the orthogonal representation case, the most dominant ones can be selected, and the impact of the dimensions reduction can be analyzed. This is summarized in Fig. 4(b), where the EVM is presented versus the total number of dimensions that are used in the orthogonal expansion. It should be noted that following the orthogonality principle, the contribution of each dimension to the series is orthogonal to the others.

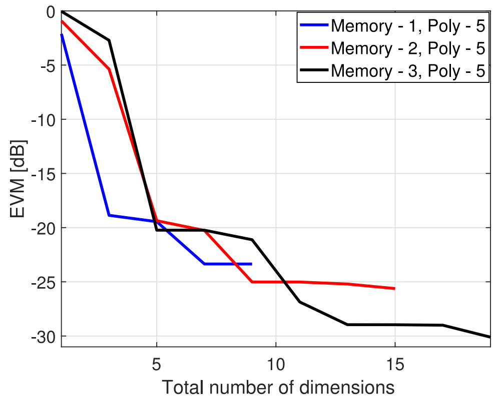


Fig. 5. EVM versus total number of dimensions for memory polynomials series.

Hence, the EVM can be further improved by adding more dimensions without the need to recalculate the previous ones. Fig. 4(b) reveals that for the case of polynomial order 5 and memory orders of 3, 2 and 1, the total numbers of required dimensions are significantly reduced to 38, 21 and 10, respectively. This indicates that a saving of more than 50% of the dimensions and associated hardware is achieved.

Using the orthogonal basis, the DPD can be implemented (*operational phase* of Fig. 1(a)) based on the selected most dominant dimensions. Fig. 3(a) presents the original input-output relationship of the simulated transmitter. By applying the DPD, the linear input-output relationship can be achieved as presented by the red and black dots in the figure. Here, the original Volterra was based on a memory order of 3 and a polynomial order of 5, projected on the orthogonal basis, and the most significant dimensions were selected. It is clear from Fig. 3(a) that both cases with DPDs of 40 and 120 dimensions DPD lead to almost identical results, demonstrating a 66% saving in the number of dimensions. This result agrees with the EVM estimation of Fig. 4(b), which almost does not improve beyond the use of 40 dimensions.

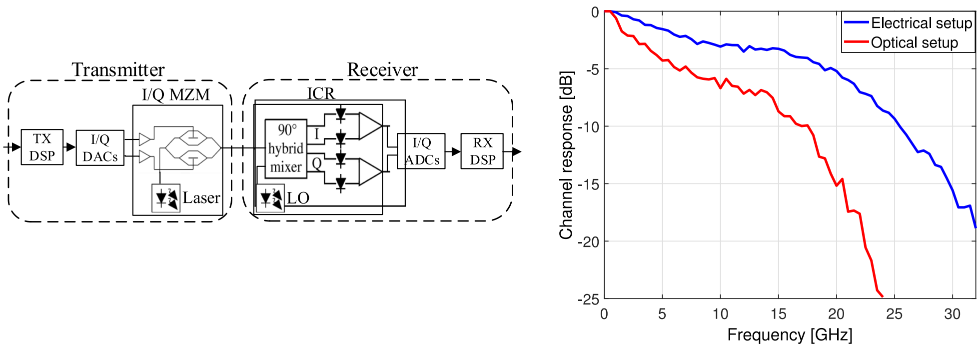
A similar comparison analysis was performed while using the degenerated Volterra form of a memory polynomials series. This is summarized in Fig. 5. As mentioned in Section 2.2, using the proposed orthogonal basis leads to a significant improvement for any variant of the Volterra series, including the memory polynomials variant.

4. Experiments

4.1 Experimental Setup

The experiment is based on an end-to-end practical phase-modulation coherent-detection (PM-CD) setup, which is presented in Fig. 6(a). The TX path includes a digital signal processing (DSP) block and a high-speed digital-to-analog converter (DAC) component with a bandwidth (BW) of 16 [GHz] operating at 64 [Gsamples/sec]. The TX DSP consists of a root raised cosine (RRC) pulse shaping filter with roll-off factor $\beta = 0.2$. It is followed by an analog driver and a dual-port MZM coherent transmitter. The laser linewidth is 100 [KHz]. The receiver consists of an integrated coherent receiver (ICR), a high-resolution analog-to-digital converter (ADC) with a BW of 16 [GHz] operating at 64 [Gsamples/sec] and an RX DSP block. The entire system BW (3 [dB] point) is less than 5 [GHz] as presented in Fig. 6(b). The receiver DSP implements all the necessary coherent algorithms (e.g. synchronization, carrier phase estimation (CPE) and frequency estimation). The input signal consists of QPSK symbols, which are transmitted at 32 [GBaud] with 2 samples per symbol (SPS). The received data are recorded at the ADC output at the received side.

Fig. 7 presents the input-output relationship of the full system of Fig. 6(a). The In-phase and Quadrature components are analyzed separately, assuming ideal phase reconstruction by the



(a) (b)

Fig. 6. (a) Experimental setup block diagram of PM-CD system with I/Q MZM. (b) System channel response.

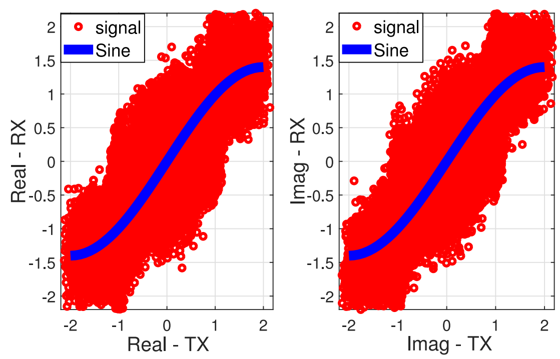


Fig. 7. The entire system conversion behavior.

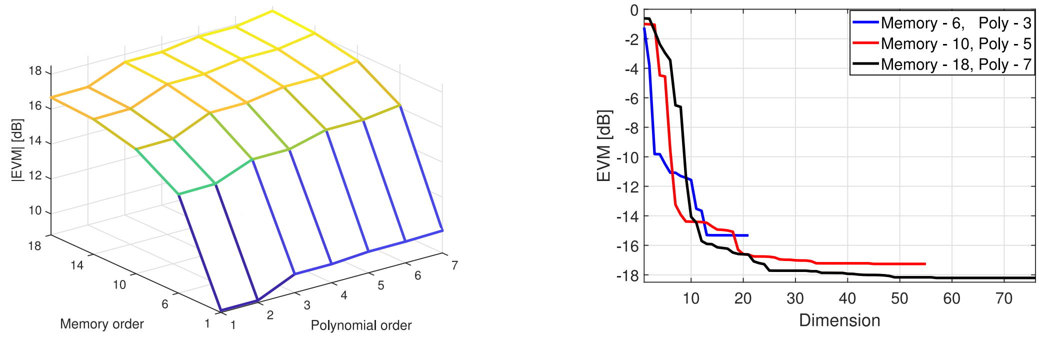
coherent algorithms. As expected, the conversion behavior of the system is similar to a spread sinusoid associated with the MZM [32], while the spreading is attributed to the bandwidth-limitation.

4.2 Experimental Results

As our experimental setup is limited by an extremely low BW, a high series memory order is required. Hence, a full Volterra series representation has a complex implementation, as a result of the large number of series elements. Consequently, the inverse system is estimated using a memory polynomials model, which is commonly used in high-speed cases [33]–[34]. First, the conventional degenerated Volterra series expansion is applied to the received samples using the memory polynomials method. Second, the linear mapping to the orthogonal basis is performed by the EVD process on the correlation matrix of the degenerated Volterra elements. Third, the variances are graded, and the most dominant dimensions are selected. Fourth, the original degenerated Volterra elements are projected on the reduced dimensions orthogonal basis.

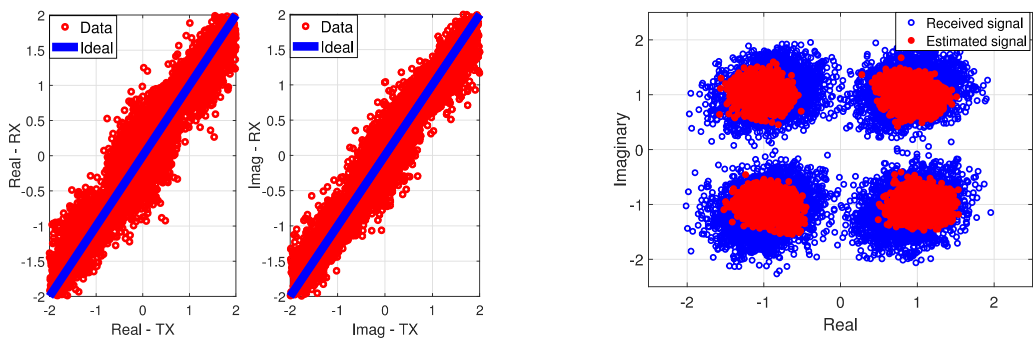
A quantitative analysis is performed on the orthogonal expansion. The post-estimation EVM results are examined for several options of memory and polynomial orders. Fig. 8(a) presents the EVM as a function of the memory and polynomial orders and the following observations are noted. Firstly, the performance is improved only by the odd polynomial orders, which is related to the odd behavior of the nonlinear system (sine behavior of MZM). Secondly, the nonlinear parts yield a performance improvement (EVM reduction) of ~1.5 [dB] over the linear parts, which is in agreement with the results of previous studies [5], [20], which indicated that the nonlinear part introduces a penalty of 1 up to 3 [dB].

The analysis is performed for three cases with the following memory depths and polynomial orders, respectively: [6, 3], [10, 5] and [18, 7].



(a) (b)

Fig. 8. (a) EVM versus memory and polynomial orders. (b) EVM as a function of the total number of dimensions after dimension selection.



(a) (b)

Fig. 9. (a) Input-output relationship after the linearization process. (b) QPSK constellation with and without linearization.

Fig. 8(b) presents the EVM as a function of the total number of dimensions that is being used. It is clearly shown that only 7–18 dimensions are dominant, which leads to a dramatic saving of 50–80% of the total number of the original dimensions. A longer series has a larger sparsity and resulting amount of saving.

Using the results of Fig. 8(b), the dimensions-reduced "linearizer" can be applied. Fig. 9(a) presents the input-output relationship of the system after the linearization process. In this figure, the original basis was built using a memory polynomials with a memory order of 18 and polynomial order of 7, and the 24 most dominant dimensions were selected. Fig. 9(b) presents the resulting constellation diagram of the QPSK symbols. The blue dots represent the system output before linearization, and the red dots represent the pre-compensated received signal. The pre-compensation was implemented by estimating of inverse system using the reduced dimensions orthogonal representation.

The main purpose of the DPD is to compensate for the BW limitation and nonlinearities of the optoelectronic components. It is shown that by applying our proposed DPD, an MZM-based high speed transmission at 32 [Gbaud] can be achieved using optoelectronic components with a dramatically lower cost and an overall bandwidth as low as 5 [GHz].

5. Conclusions

In this paper, we have shown that in the case of a band-limited nonlinear system, Volterra DPD may become implementable if designed with orthogonal basis functions. In addition, the orthogonal basis makes it possible to apply PLS estimation of the Volterra coefficients, discriminating the most important coefficients while neglecting the rest. This, in turn, dramatically reduces the computational complexity and minimizes the hardware resources requirements.

All of these benefits are critical in high-speed communication systems. For cases of independent variables, the orthogonal basis can be calculated analytically. For cases with dependent variables or unknown probability distribution functions, a computational method is proposed using the EVD process on the correlation matrix, and a dimension reduction could be achieved. Detailed sets of simulations and experiments for the QPSK signal over band-limited MZM transmitters are performed. The results indicated that by using the orthogonalization approach, 50–80% of the dimensions could be saved. Our approach is signal-independent. Thus it could also be applied to higher-order modulation systems too. In addition, this method could be used for any variant of the Volterra series such as memory polynomials, which may be commonly used in high-speed optical communication.

Appendix A
PLS Operations Saving

Let us assume that ρ elements are estimated using P samples. Then, the LS equation is as follows:

$$\widetilde{H} = ([\rho X P] \cdot [P X \rho])^{-1} \cdot [\rho X P] \cdot [P X 1] = [\rho X \rho] \cdot [\rho X 1] \quad (11)$$

The right-hand side product requires ρP multiplications and $\rho(P-1)$ additions. The pseudo-inverse matrix has $[\rho X \rho]$ dimensions. Thus, the middle product requires ρ^2 multiplications and $\rho(\rho-1)$ additions. The correlation matrix calculation requires $\rho^2 P$ multiplications and $\rho^2(P-1)$ additions. Thus, the total resource requirements are $\rho(P+\rho+\rho P)$ multiplications and $\rho(P+\rho P-2)$ additions, without the inversion action. The naive matrix inversion requires $O(\rho^3)$ operations.

On the other hand, the PLS approach requires solving the PLS equation to be solved several times:

$$\widetilde{H}_{new_j} = ([1 X P] \cdot [P X 1])^{-1} \cdot [1 X P] \cdot [P X 1] = [1 X 1] \cdot [1 X 1] \quad (12)$$

The right-hand side product requires P multiplications and $(P-1)$ additions. The middle product requires a single multiplication. The correlation matrix requires only P multiplications and $(P-1)$ additions. Therefore, the PLS method requires $\rho(2P+1)$ multiplications and $\rho(2P-2)$ additions, without the diagonalization procedure. When the orthogonal basis is unknown, the EVD requires $O(\rho^3)$ operations for the naive diagonalization process, or $O(\rho^3/3)$ operations for the Cholesky algorithm.

Appendix B
Orthogonal Basis for X-D Independent RVs

The distribution of some RV's distributions have an appropriate orthogonal basis, such as “shifted Legendre polynomials” for a uniform distribution, or “Hermitian polynomials” for a Gaussian distribution.

For 2-D independent RVs, Yang [8] gave the following definition:

$$\Psi_{i,j}(x_1, x_2) = \varphi_{i-j}(x_1)\,\chi_j(x_2) \quad (13)$$

where φ, χ are the appropriate orthogonal basis for a 1-D RV, $i = 1, 2, \ldots$ and $j = 0, 1, \ldots$

Hence:

$$E\left[\Psi_{i,j}(x_1, x_2)\,\Psi_{r,t}(x_1, x_2)\right] = E\left[\varphi_{i-j}(x_1)\,\chi_j(x_2)\,\varphi_{r-t}(x_1)\,\chi_t(x_2)\right] \quad (14)$$

If the RVs are independent, then:

$$E\left[\Psi_{i,j}(x_1, x_2)\,\Psi_{r,t}(x_1, x_2)\right] = E\left[\varphi_{i-j}(x_1)\,\varphi_{r-t}(x_1)\right] E\left[\chi_j(x_2)\,\chi_t(x_2)\right] = \delta\left((i-j)-(r-t)\right)\delta(j-t) \quad (15)$$

Using this approach the expansion of the Yang process to X-D independent RVs is achieved by defining:

$$\Psi_{j_1 \ldots j_n}(x_1 \ldots x_n) = \varphi_{j_1 - \sum_{k=2}^{n} j_k}(x_1) \ldots \chi_j(x_n) \tag{16}$$

The inter-projection between elements gives a product of delta functions. However, if there is dependency between the RVs, the separation of the moments in Eq. (15) does not hold. The memory effect causes dependency between the output samples. Hence, the Yang process cannot be used for DPD estimation for a system with the memory effect.

References

[1] J. D. Ingham *et al.*, "32 Gb/s multilevel modulation of an 850 nm VCSEL for next-generation data communication standards," in *Proc. CLEO, Laser Appl. Photon. Appl.*, 2011, Paper. CWJ2.
[2] H. Yang *et al.*, "Evaluation of effects of MZM nonlinearity on QAM and OFDM signals in RoF transmitter," in *Proc. Int. Topical Meeting Microw. Photon./Asia-Pac. Microw. Photon. Conference.*, Sep. 2008, pp. 90–93.
[3] Y. Satuby and M. Orenstein, "Limits of the modulation response of a single-mode proton implanted vcsel," *IEEE Photon. Technol. Lett.*, vol. 10, no. 6, pp. 760–762, Jun. 1998.
[4] I. Teikari, "Digital predistortion linearization methods for RF power amplifiers," Ph.D. dissertation, Dept. Microtechnol Nanotechnol., Teknillinen korkeakoulu, Espoo, Finland, 2008.
[5] P. W. Berenguer *et al.*, "Nonlinear digital pre-distortion of transmitter components," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1739–1745, Apr. 2016.
[6] P. Knupp, "Matrix norms and the condition number: A general framework to improve mesh quality via node-movement," Sandia National Labs., Albuquerque, NM, USA/Sandia National Labs., Livermore, CA, USA, Tech. Rep. SAND99-2542C, 1999.
[7] R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1468–1479, Sep. 2004.
[8] R. Raich and G. T. Zhou, "Orthogonal polynomials for complex Gaussian processes," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2788–2797, Oct. 2004.
[9] G. Yang, F. Liu, L. Li, H. Wang, C. Zhao, and Z. Wang, "2D orthogonal polynomials for concurrent dual-band digital predistortion," in *Proc. IEEE MTT-S Int. Microw. Symp. Dig.*, 2013, pp. 1–3.
[10] L. Ding *et al.*, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.
[11] A. Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned Volterra series," *IEEE Microw. Compon. Lett.*, vol. 14, no. 12, pp. 563–565, Dec. 2004.
[12] F. Mkadem, M. Fares, S. Boumaiza, and J. Wood, "Complexity-reduced Volterra series model for power amplifier digital predistortion," *Analog Integr. Circuits Signal Process.*, vol. 79, no. 2, pp. 331–343, May 2014.
[13] N. Safari, T. Roste, P. Fedorenko, and J. S. Kenney, "An approximation of Volterra series using delay envelopes, applied to digital predistortion of RF power amplifiers with memory effects," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 2, pp. 115–117, Feb. 2008.
[14] A. Zhu, M. Wren, and T. J. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers," in *Proc. IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2003, vol. 2, pp. 787–790.
[15] D. Sadot, Y. Yoffe, H. Feig, and G. Paryanti, "Digital pre-compensation techniques enabling cost-efficient high-order modulation formats transmission," in *Proc. Opt. Fiber Commun. Conf. Expo.*, 2018, Paper Th3E-1.
[16] H. Faig, Y. Yoffe, and D. Sadot, "Dimensions-Reduced Volterra-Based Digital Pre-Distortion for Band-Limitted Nonlinear Components," in *Proc. OSA Signal Process. Photon. Commun.*, Jul. 2018, Paper SpM3G.2.
[17] X. Yu and H. Jiang, "Digital predistortion using adaptive basis functions," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 12, pp. 3317–3327, Dec. 2013.
[18] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 7, pp. 1510–1520, Jun. 2010.
[19] J. Tsimbinos and K. Lever, "Computational complexity of Volterra based nonlinear compensators," *Electron. Lett.*, vol. 32, no. 9, pp. 852–854, 1996.
[20] N. Stojanovic, F. Karinou, Z. Qiang, and C. Prodaniuc, "Volterra and Wiener equalizers for short-reach 100 G PAM-4 applications," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4583–4594, Nov. 2017.
[21] Y. Gao, J. C. Cartledge, S. S.-H. Yam, A. Rezania, and Y. Matsui, "112 Gb/s PAM-4 using a directly modulated laser with linear precompensation and nonlinear post-compensation," in *Proc. Eur. Conf. Opt. Commun.*, 2016, pp. 1–3.
[22] G. Kerschen G, J. C. Golinval, A. F. Vakakis, and L. A. Bergman, "The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview," *Nonlinear Dyn.*, vol. 41, pp. 147–169, 2005.
[23] N. Kambhatla and T. Leen, "Dimension reduction by local principal component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1493–1516, 1997.
[24] G. Shulkind and M. Nazarathy, "Nonlinear digital back propagation compensator for coherent optical OFDM based on factorizing the Volterra series transfer function," *Opt. Express*, vol. 21, no. 11, pp. 13145–13161, 2013.
[25] G. Khanna, B. Spinnler, S. Calabro, E. De Man, and N. Hanik, "A robust adaptive pre-distortion method for optical communication transmitters," *IEEE Photon. Technol. Lett.*, vol. 28, no. 7, pp. 752–755, Apr. 2016.

[26] H. Ying, M. Zhu, J. Zhang, X. Yi, Y. Song, and K. Qiu, "Sparse Volterra model based on optical single side-band NPAM-4 direct-detection system," *Opt. Fiber Technol.*, vol. 40, pp. 180-184, 2018.
[27] G. Malhotra and J. Kamali, "BER analysis of high speed links with nonlinearity%," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, 2015, pp. 635–640.
[28] I. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer-Verlag, 2002.
[29] Y. Chen and I. W. Hunter, "Nonlinear stochastic system identification of skin using Volterra kernels" *Ann. Biomed. Eng.*, vol. 41, no. 4, pp. 847–862, 2013.
[30] A. Napoli, M. M. Mezghanni, S. Calabr, R. Palmer, G. Saathoff, and B. Spinnler, "Digital predistortion techniques for finite extinction ratio IQ Mach–Zehnder modulators," *J. Lightw. Technol.*, vol. 35, no. 19, pp. 4289–4296, Oct. 2017.
[31] A. Y. Kibangou and G. Favier, "Wiener–Hammerstein systems modeling using diagonal Volterra kernels coefficients," *IEEE Trans. Signal Process. Lett.*, vol. 13, no. 6, pp. 381–384, Jun. 2006.
[32] W. Jones, Introduction to optical fiber communication systems. New York, NY, USA: Holt, Rinehart and Winston, 1998.
[33] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-band complexityreduced generalized-memory-polynomial power-amplifier digital predistortion," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 6, pp. 1763–1774, Jun. 2016.
[34] J. Zhang *et al.*, "Memory-polynomial digital pre-distortion for linearity improvement of directly-modulated multi-IF-over-fiber LTE mobile fronthaul," in *Proc. Opt. Fiber Commun. Conf. Expo.*, 2016, pp. 1–3.