Equivalent Baseband Models and Corresponding Digital Predistortion for Compensating Dynamic Passband Nonlinearities in Phase-Amplitude Modulation-Demodulation Schemes

As Published: 10.1109/TSP.2018.2871385

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: https://hdl.handle.net/1721.1/136348

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Equivalent Baseband Models and Corresponding Digital Predistortion for Compensating Dynamic Passband Nonlinearities in Phase-Amplitude Modulation-Demodulation Schemes

Omer Tanovic, *Member, IEEE,* Alexandre Megretski, Yan Li, Vladimir Stojanovic, *Senior Member, IEEE,* and Mitra Osqui

***Abstract*—We consider equivalent baseband representation of transmission circuits, in the form of a nonlinear dynamical system S in discrete time (DT) defined by a series interconnection of a phase-amplitude modulator, a nonlinear dynamical system F in continuous time (CT), and an ideal demodulator. We show that when F is a CT Volterra series model, the resulting S is a series interconnection of a DT Volterra series model of same degree and equivalent memory depth, and an LTI system with special properties. The result suggests a new, non-obvious, analytically motivated structure of digital pre-compensation of analog nonlinear distortions such as those caused by power amplifiers in digital communication systems. The baseband model and the corresponding digital compensation structure readily extend to OFDM modulation. MATLAB simulation is used to verify proposed equivalent baseband model and demonstrate effectiveness of the new compensation scheme, as compared to the standard Volterra series approach.**

***Index Terms*—Power amplifiers, predistortion, baseband, RF signals, phase modulation, amplitude modulation.**

I. Introduction

In modern communications systems, demand for high-throughput data transmission has resulted in the use of complex modulation/demodulation schemes for enhanced spectral efficiency (e.g., Orthogonal Frequency Division Multiplexing - OFDM). The highly varying envelope signals, generated by these schemes, force RF transmitter power amplifiers (PA) to operate over a large portion of their transfer curves, generating in/out of band spectral content which degrades system linearity. A common way to make the PA (and correspondingly the whole signal chain) behave linearly is to back-off PA's input level, which in turn results in reduced power efficiency [1]. This motivates the search for a method which would help increase both linearity and power efficiency. Digital compensation offers an attractive approach to designing electronic devices with superior characteristics, and it is not a surprise that it has been used for PA linearization as well. Nonlinear distortion in an analog system can be compensated with a pre-distorter or a post-compensator system. In particular, a pre-distorter inverts nonlinear behavior of the analog part, and is usually implemented as a digital system. Techniques which employ such systems are called digital predistortion (DPD) techniques, and they can produce highly linear transmitter circuits [1]-[3].

Structure of a digital predistorter usually depends on the choice of a behavioral model used to describe the PA [4], and consequently on the corresponding equivalent baseband model. First attempts to mitigate PA's nonlinear effects by employing DPD used simple memoryless models in order to describe PA's behavior [5]. With the increase of signal bandwidth over time, it has been recognized that short and long memory effects play significant role in PA's behavior [6], and should be incorporated into the model. Since then several memory baseband models and corresponding predistorters have been proposed to compensate memory effects: memory polynomials [7], [8], Hammerstein and Wiener DPD models [9], [10], pruned Volterra series [11], generalized memory polynomials [12], dynamic deviation reduction-based Volterra models [13], [14], as well as the most recent neural networks based behavioral models [15]. In behavioral modeling based on memory polynomials, it is commonly assumed that all branches in the model (i.e., monomials of equal degree) are generated with the same memory requirements. Since memoryless nonlinear effects dominate in terms of degree, the above choice leads to unnecessary complexity of the PA model. Several models were proposed in which strong memoryless nonlinear effects and weak memory dynamic effects are decoupled in order to reduce model complexity while keeping good modeling/linearization performance, such as twin nonlinear two-box (TNTB) [16] and three-box (PLUME) models [17].

In the above papers, an emphasis is put on capturing the whole range of the PA output signal's spectrum, which is proportional to the order of nonlinearity of the RF PA, and is in practice taken to be about five times the input bandwidth [4]. In wideband communication systems this would make the linearization bandwidth very large, and hence would put a significant burden on the system design (e.g., would re-

Manuscript received January 31, 2018; revised June 7, 2018 and August 8, 2018. This work was supported by DARPA Award No. W911NF-10-1-0088. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Youngchul Sung.

O. Tanovic and A. Megretski are with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA email: otanovic@mit.edu.

Y. Li is with NanoSemi Inc., Waltham, MA 02451, USA.

V. M. Stojanovic is with the Department of Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA 94720, USA.

M. Osqui is with Analog Devices|Lyric Labs, Cambridge, MA 02142, USA.

quire very high-speed data converters). Since these restrictions limit applicability of conventional models in the forthcoming wideband systems, it is beneficial to investigate model dynamics when the PA's output is also limited in bandwidth, and in that case DPD would ideally mitigate distortion in that frequency band (also called the observation band). Such band-limited baseband model and its corresponding DPD were investigated in [18], and promising experimental results were shown. Theoretical analysis shown in [18] follows the same modeling approach as the conventional baseband models (dynamic deviation reduction-based Volterra series modeling). Due to the bandpass filtering operation applied on the PA output, long (possibly infinite) memory dynamic behavior is now present, which implies that standard modeling methods, such as memory polynomials or dynamic deviation reduction-based Volterra series modeling, might be too general to efficiently describe this new structure, and therefore not well suited for practical implementations (long memory condition in nonlinear models would require exponentially large number of coefficients to be implemented, which is too expensive for hardware realization). Modeling and linearization of PAs using band-limited observation path was used in [19]-[21] but the corresponding baseband equivalent models and DPDs were derived for the full-band PA output, and are therefore fundamentally different from the band-limited models investigated in this work.

In this paper, we develop an explicit expression of the equivalent baseband model, for the case when the passband nonlinearity can be described by a Volterra series model with fixed degree and memory depth. We show that the equivalent baseband model can be written as a series interconnection of a fixed degree/short memory discrete-time Volterra model, and a long memory discrete-time LTI system which can be viewed as a bank of *reconstruction filters*. Moreover, this equivalent baseband model reveals that order of nonlinearity and, more importantly, memory of the underlying nonlinear system are preserved when passing from passband to baseband. Frequency responses of the above reconstruction filters exhibit discontinuities at frequency values $\pm\pi$, making their unit sample responses infinitely long. This discontinuity is mainly due to the lack of symmetry of the frequency response of the reconstruction filter with respect to the carrier frequency. Nevertheless, the frequency responses are shown to be smooth inside the interval $(-\pi, \pi)$, and thus approximable by low order FIR filters on compact subsets of $(-\pi, \pi)$. Length (i.e, number of taps) of these approximate FIR filters depends on the ratio of the signal to observation bandwidth (i.e., the amount of observation bandwidth which is occupied by the useful signal). Relatively low memory/degree requirements of the nonlinear (Volterra) subsystem, as well as good approximability by FIR filters of the linear subsystem, allow for potentially efficient hardware implementation of the corresponding baseband model. Based on the derived model, we propose a non-obvious, analytically motivated structure of digital precompensation of RF PA nonlinearities. The two-box decomposition of the equivalent baseband model that we propose in this paper is not motivated nor does it rely on decoupling strong from weak nonlinear effects. It rather relies on decoupling short memory nonlinear effects from long memory linear effects, and is therefore fundamentally different from those in [16]-[17] or [20]-[21]. Equivalent model structures for discrete-time Volterra systems similar to the one presented in this paper were derived in [22]-[23]. The work presented here mostly differs from the above in that model dynamics are investigated over a limited bandwidth of the PA output, which, as described earlier, leads to significantly different dynamic behavior of the underlying system, and therefore to a distinct baseband equivalent model.

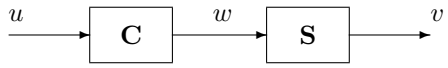
This paper is organized as follows. In Section III we present motivation for considering the problem of equivalent baseband modeling and digital predistortion, and give mathematical description of the system under consideration. Main result is stated and proven in Section IV, in which an explicit expression of the equivalent baseband model is given. In Section V we provide some additional discussion on advantages of the proposed method, and its extension to OFDM modulation. System model validation, as well as DPD design and its performance are demonstrated by MATLAB simulation results presented in Section VI. Finally, we conclude the paper in Section VII.

II. Notation and Terminology

j is a fixed square root of -1. $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{Z}_+$ and $\mathbb{N}$ are the standard sets of complex, real, integer, non-negative integer and positive integer numbers. $[1 : n]$, for $n \in \mathbb{N}$, is the set $\{1, \ldots, n\}$. X^d, for a set X, is the set of all d-tuples $(x_1, \ldots, x_d)$ with $x_i \in X$. For a finite set S, $|S|$ denotes the number of elements in S. Scalar real-valued and complex-valued *continuous-time (CT) signals* are uniformly bounded square integrable functions $\mathbb{R} \to \mathbb{R}$ and $\mathbb{R} \to \mathbb{C}$, respectively. The set of all real-valued CT signals is denoted by $\mathcal{L}$, while the set of all complex-valued CT signals is denoted by $\mathcal{L}_{\mathbb{C}}$. n-element-vector-valued *discrete-time (DT) signals* are the elements of ℓ_n (or simply ℓ for $n = 1$), the set of all square summable functions $\mathbb{Z} \to \mathbb{C}^n$. For $w \in \ell$, $w[n]$ denotes the value of w at $n \in \mathbb{Z}$. In contrast, $x(t)$ refers to the value of $x \in \mathcal{L}$ (or $x \in \mathcal{L}_{\mathbb{C}}$) at $t \in \mathbb{R}$. The Fourier transform $\mathcal{F}$ applies to both CT and DT signals. For $x \in \mathcal{L}$ (or $x \in \mathcal{L}_{\mathbb{C}}$), its Fourier transform $X = \mathcal{F}x$ is a square integrable function $X : \mathbb{R} \to \mathbb{C}$. For $x \in \ell_n$, the Fourier transform $X = \mathcal{F}x$ is a 2π-periodic function $X : \mathbb{R} \to \mathbb{C}^n$, square integrable on its period. *Systems* are viewed as functions $\mathcal{L} \to \mathcal{L}$, $\mathcal{L} \to \mathcal{L}_{\mathbb{C}}$, $\mathcal{L}_{\mathbb{C}} \to \mathcal{L}$, $\mathcal{L}_{\mathbb{C}} \to \ell$, $\ell \to \mathcal{L}_{\mathbb{C}}$, or $\ell_k \to \ell_m$. $\mathbf{G}f$ denotes the response of system $\mathbf{G}$ to signal f (even when $\mathbf{G}$ is not linear), and the *series composition* $\mathbf{K} = \mathbf{Q}\mathbf{G}$ of systems $\mathbf{Q}$ and $\mathbf{G}$ is the system mapping f to $\mathbf{Q}(\mathbf{G}f)$. A system $\mathbf{G} : \mathcal{L} \to \mathcal{L}$ (or $\mathbf{G} : \ell \to \ell$) is said to be *linear and time invariant* (LTI) with *frequency response* $H : \mathbb{R} \to \mathbb{C}$ when $\mathcal{F}\mathbf{G}x = H \cdot \mathcal{F}x$ for all $x \in \mathcal{L}$ (respectively $x \in \ell$). When $\mathbf{G} : \ell \to \ell$, its frequency response H is 2π-periodic and signal $h \in \ell$ such that $H = \mathcal{F}h$ is called the unit sample response of $\mathbf{G}$. We use shorthand notation $X = X(\omega)$ (or $X = X(\Omega)$) to denote the Fourier transform of signal $x \in \mathcal{L}$ or $x \in \mathcal{L}_{\mathbb{C}}$ (respectively $x \in \ell_n$), instead of a standard notation $X = X(j\omega)$ (respectively $X = X(e^{j\Omega})$).

III. Motivation and Problem Setup

In this paper, a digital compensator is viewed as a system $\mathbf{C}: \ell \to \ell$. More specifically, a *pre*-compensator $\mathbf{C}: \ell \to \ell$ designed for a device modeled by a system $\mathbf{S}: \ell \to \mathcal{L}$ (or $\mathbf{S}: \ell \to \ell$) aims to make the composition $\mathbf{SC}$, as shown on the block diagram below,



conform to a set of desired specifications. (In the simplest scenario, the objective is to make $\mathbf{SC}$ as close to the identity map as possible, in order to cancel the distortions introduced by $\mathbf{S}$.)

A common element in digital compensator design algorithms is selection of *compensator structure*, which usually means specifying a finite sequence $\tilde{\mathbf{C}} = (\mathbf{C}_1, \dots \mathbf{C}_N)$ of systems $\mathbf{C}_k: \ell \to \ell$, and restricting the actual compensator $\mathbf{C}$ to have the form

$$\mathbf{C} = \sum_{k=1}^{N} a_k \mathbf{C}_k, \qquad a_k \in \mathbb{C},$$

i.e., to be a linear combination of the elements of $\tilde{\mathbf{C}}$. Once the *basis* sequence $\tilde{\mathbf{C}}$ is fixed, the design usually reduces to a straightforward *least squares optimization* of the coefficients $a_k \in \mathbb{C}$. A popular choice is for the systems $\mathbf{C}_k$ to be some *Volterra monomials*, i.e., to map their input $u = u[n]$ to the outputs $w_k = w_k[n]$ according to the polynomial formulae

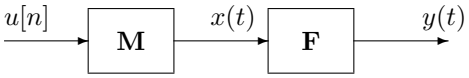
$$w_k[n] = \prod_{j=1}^{d_r(k)} \text{Re } u[n - n^r_{k,j}] \prod_{j=1}^{d_i(k)} \text{Im } u[n - n^i_{k,j}],$$

where the integers $d_r(k)$ and $d_i(k)$ (respectively, $n^r_{k,j}$ and $n^i_{k,j}$) will be referred to as *degrees* (respectively, *delays*). In this case, every linear combination $\mathbf{C}$ of $\mathbf{C}_k$ is a *DT Volterra series* [27], i.e., a DT system mapping signal inputs $u \in \ell$ to outputs $w \in \ell$ according to the polynomial expression

$$w[n] = \sum_{k=1}^{N} a_k \prod_{j=1}^{d_r(k)} \text{Re } u[n - n^r_{k,j}] \prod_{j=1}^{d_i(k)} \text{Im } u[n - n^i_{k,j}].$$

Selecting a proper *compensator structure* is a major challenge in compensator design: a basis which is too simple will not be capable of canceling the distortions well, while a form that is too complex will consume excessive power and space in its hardware implementation. Having an insight into the compensator basis selection can be very valuable.

In this paper we establish that a certain special structure is good enough to compensate for imperfect modulation. We consider modulation systems represented by the block diagram



where $\mathbf{M}: \ell \to \mathcal{L}$ is the *ideal* modulator, and $\mathbf{F}: \mathcal{L} \to \mathcal{L}$ is a CT dynamical system used to represent linear and nonlinear distortion in the modulator and power amplifier circuits. We consider the ideal modulator of the form $\mathbf{M} = \mathbf{XZ}$, where $\mathbf{Z}: \ell \to \mathcal{L}_{\mathbb{C}}$ is the *zero order hold* (ZOH) map $u[\cdot] \mapsto x_0(\cdot)$:

$$x_0(t) = \sum_n p(t - nT) u[n], \quad p(t) = \begin{cases} 1, & t \in [0, T), \\ 0, & t \notin [0, T) \end{cases} \tag{1}$$

with fixed sampling interval length $T > 0$ and $\mathbf{X}: \mathcal{L}_{\mathbb{C}} \to \mathcal{L}$ is the *mixer* map

$$x_0(\cdot) \mapsto x(\cdot): \quad x(t) = 2\text{Re}[\exp(j\omega_c t) x_0(t)] \tag{2}$$

with modulation-to-sampling frequency ratio $M \in \mathbb{N}$, i.e., with $\omega_c = 2\pi M/T$. It should be noted that in real applications, there is no constraint on M to be integer valued. We assume $M \in \mathbb{N}$ for convenience, since it simplifies, already complex, derivations as will be seen in section IV. For non-integer valued M, time-invariance of certain subsystems is lost, but, in most applications, the introduced time-dependence is not significant enough to impact the overall system structure.

While it is commonly assumed that the modulation/digital-to-analog conversion (DAC) is realized in a way which includes low-pass filtering of the DAC output, there has been significant body of work recently focusing on the so called "all-digital transmitters" (ADT) that employ ZOH modulation and have no specific low-pass filtering of the DAC output (see e.g., [24]-[26] and references therein). ADT utilizes power amplifiers in switched-mode operation (SMPA) for increased power efficiency. SMPAs, due to their specific mode of operation (transistors work as switches), require piece-wise constant input signals, or piece-wise constant envelope signals in the case of burst-mode transmitters. In such setup, ZOH modulation is a natural method for converting digital RF/baseband into analog signals suitable for driving SMPAs. In Section V.B., we discuss how results similar to the ones stated and proven under assumption of a ZOH modulator, hold for more general system models which include low-pass filtering after ZOH.

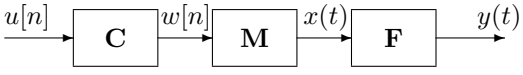
We are particularly interested in the case when $\mathbf{F}$ is described by the *CT Volterra series model*

$$y(t) = b_0 + \sum_{k=1}^{N_b} b_k \prod_{i=1}^{\beta_k} x(t - t_{k,i}), \tag{3}$$

where $N_b \in \mathbb{N}$, $b_k \in \mathbb{R}$, $\beta_k \in \mathbb{N}$, $t_{k,i} \geq 0$ are parameters. (In a similar fashion, it is possible to consider input-output relations in which the finite sum in (3) is replaced by an integral/infinite sum). One expects that the memory of $\mathbf{F}$ is not long, compared to T, i.e., that $\max t_{k,i}/T$ is not much larger than 1.

As a rule, the spectrum of the DT input $u \in \ell$ of the modulator is carefully shaped at a pre-processing stage to guarantee desired characteristics of the modulated signal $x = \mathbf{M}u$. However, when the distortion $\mathbf{F}$ is not linear, the spectrum of the $y = \mathbf{F}x$ could be damaged substantially, leading to violations of error vector magnitude (EVM) and adjacent channel leakage ratio (ACLR) specifications [14].

Consider the possibility of repairing the spectrum of y by pre-distorting the digital input $u \in \ell$ by a compensator $\mathbf{C}: \ell \to \ell$, as shown on the block diagram below:



The desired effect of inserting $\mathbf{C}$ is cancellation of the distortion caused by $\mathbf{F}$. Naturally, since $\mathbf{C}$ acts in the baseband (i.e., in discrete time), there is no chance that $\mathbf{C}$ will achieve a complete correction, i.e., that the series composition $\mathbf{FMC}$ of $\mathbf{F}$, $\mathbf{M}$, and $\mathbf{C}$ will be identical to $\mathbf{M}$. However, in principle, it is sometimes possible to make the frequency contents of $\mathbf{M}u$ and $\mathbf{FMC}u$ to be identical within the CT frequency band $(\omega_c - \omega_b, \omega_c + \omega_b)$, where $\omega_b = \pi/T$ is the Nyquist frequency [28], [29]. To this end, let $\mathbf{H}: \mathcal{L} \to \mathcal{L}$ denote the ideal band-pass filter with frequency response $H(\omega) = 1$ for $|\omega_c - |\omega|| < \omega_b$ and $H(\omega) = 0$ otherwise. Let $\mathbf{D}: \mathcal{L} \to \ell$ be the *ideal de-modulator* operating on the band selected by $\mathbf{H}$, i.e., the linear system for which the series composition $\mathbf{DHM}$ is the identity function. Let $\mathbf{S} = \mathbf{DHFM}$ be the series composition of $\mathbf{D}$, $\mathbf{H}$, $\mathbf{F}$, and $\mathbf{M}$, i.e., the DT system with input $w = w[n]$ and output $v = v[n]$ shown on the block diagram in Fig. 1.

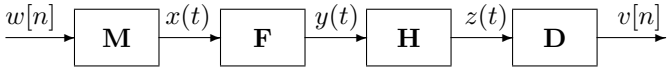


Fig. 1. Block diagram of $\mathbf{S} = \mathbf{DHFM}$.

By construction, the ideal compensator $\mathbf{C}$ should be the inverse $\mathbf{C} = \mathbf{S}^{-1}$ of $\mathbf{S}$, as long as the inverse does exist. A key question answered in this paper is "what to expect from system $\mathbf{S}$?". If one assumes that the continuous-time distortion subsystem $\mathbf{F}$ is simple enough, what does this say about $\mathbf{S}$?

This paper provides an explicit expression for $\mathbf{S}$ in the case when $\mathbf{F}$ is given in the CT Volterra series form (3) with *degree* $d = \max \beta_k$ and *depth* $t_{max} = \max t_{k,i}$. The result reveals that, even though $\mathbf{S}$ tends to have infinitely long memory (due to the ideal band-pass filter $\mathbf{H}$ being involved in the construction of $\mathbf{S}$), it can be represented as a series interconnection $\mathbf{S} = \mathbf{LV}$, where $\mathbf{V}: \ell \to \ell_N$ maps complex scalar-valued input $w \in \ell$ to real vector-valued output $s \in \ell_N$ in such a way that the k-th scalar component $s_k[n]$ of $s[n] \in \mathbb{R}^N$ is given by

$$s_k[n] = \prod_{i=0}^{m} (\text{Re } w[n-i])^{\alpha_i} \prod_{i=0}^{m} (\text{Im } w[n-i])^{\beta_i},$$

$$\alpha_i, \beta_i \in \mathbb{Z}_+, \qquad \sum_{i=0}^{m} \alpha_i + \sum_{i=0}^{m} \beta_i \le d,$$

m is the minimal integer not smaller than t_{max}/T, and $\mathbf{L}: \ell_N \to \ell$ is an LTI system. Moreover, $\mathbf{L}$ can be shown to have a good approximation of the form $\mathbf{L} \approx \mathbf{L}_0 X$, where X is a static gain matrix, and $\mathbf{L}_0$ is an LTI model which does not depend on b_k and $t_{k,i}$ (see Fig. 2). In other words, $\mathbf{S}$ can be well approximated by combining a Volterra series model with short memory, and a *fixed* (long memory) LTI system, as long as the memory depth t_{max} of $\mathbf{F}$ is short, relative to the sampling time T.

In most applications, with an appropriate scaling and time delay, the system $\mathbf{S}$ to be inverted can be viewed as a small perturbation of identity, i.e., $\mathbf{S} = \mathbf{I} + \mathbf{\Delta}$. When $\mathbf{\Delta}$ is "small" in an appropriate sense (e.g., has small incremental L2 gain[1] $\|\mathbf{\Delta}\| \ll 1$), the inverse of $\mathbf{S}$ can be well approximated by $\mathbf{S}^{-1} \approx \mathbf{I} - \mathbf{\Delta} = 2\mathbf{I} - \mathbf{S}$. Hence the result of this paper suggests a specific structure of the compensator (predistorter) $\mathbf{C} \approx \mathbf{I} - \mathbf{\Delta} = 2\mathbf{I} - \mathbf{S}$. In other words, a plain Volterra monomials structure is, in general, not good enough for $\mathbf{C}$, as it lacks the capacity to implement the long-memory LTI post-filter $\mathbf{L}$. Instead, $\mathbf{C}$ should be sought in the form $\mathbf{C} = \mathbf{I} - \mathbf{L_0} X \mathbf{V}$, where $\mathbf{V}$ is the system generating all Volterra series monomials of a limited depth and limited degree, $\mathbf{L}_0$ is a *fixed* LTI system with a very long time constant, and X is a matrix of coefficients to be optimized to fit the data available.

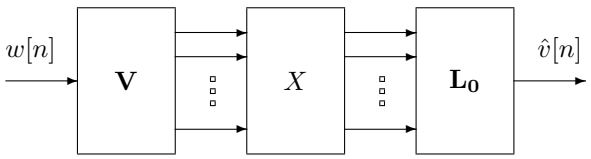


Fig. 2. System $\mathbf{S}$ can be well approximated by the model shown in this block diagram.

IV. Main Result

A. Ideal Demodulator

By definition, demodulator $\mathbf{D}$ should "invert" the operation of $\mathbf{M}$, i.e., $\mathbf{DM} = \mathbf{I}$ should hold. In communications literature, demodulation is usually described as downconversion of the passband signal, followed by low-pass (anti-aliasing) filtering (LPF) and sampling [30]. When modulator system $\mathbf{M}$ produces output signal with significant side-lobes (as is the case with ZOH) the above described demodulation process leads to $\mathbf{DM} \neq \mathbf{I}$. If the ratio $M = f_c/f_s$ is very large, the in-band spectral distortion that results from the folding of ZOH side-lobes would be negligible compared to signal in-band energy, which implies $\mathbf{DM} \approx \mathbf{I}$. As M decreases, this distortion becomes more notable and might match that caused by the PA nonlinearity. This implies that distortions introduced by non-ideal demodulation could mask good performance of digital predistortion. For that reason, in this paper, we apply demodulation which recovers the input signal, without introducing distortion. We call this operation *ideal demodulation*, and in the following derive its mathematical model.

The most commonly known expression for the ideal demodulator inverts not $\mathbf{M} = \mathbf{XZ}$ but $\mathbf{M}_0 = \mathbf{XH}_0\mathbf{Z}$, i.e., the modulator which inserts $\mathbf{H}_0$, the *ideal low-pass filter* for the baseband, between zero-order hold $\mathbf{Z}$ and mixer $\mathbf{X}$, where $\mathbf{H}_0$ is the CT LTI system with frequency response $H_0(\omega) = 1$ for $|\omega| < \omega_b$ and $H_0(\omega) = 0$ otherwise.

[1]Incremental L2 gain (as well as other similarly defined system "gains") can be viewed as a measure of "size" of a system, or more precisely as a measure of sensitivity of a system response to its input. With Euclidean norm (i.e., L2-norm) $\|\xi\|$ of an element $\xi = \begin{bmatrix}\xi_1 & \dots & \xi_k\end{bmatrix} \in \mathbb{C}^k$ defined by $\|\xi\|^2 = \sum_{i=1}^{k} |\xi_i|^2$, the incremental L2 gain $\|\mathbf{G}\|$ of a DT system $\mathbf{G}: \ell_n \to \ell_m$ can be defined as the maximal upper bound on $\gamma \ge 0$ such that

$$\sum_{n \in \mathbb{Z}} \|y_1[n] - y_2[n]\|^2 \le \gamma^2 \sum_{n \in \mathbb{Z}} \|w_1[n] - w_2[n]\|^2,$$

for all $w_1, w_2 \in \ell_n$, where $y_1 = \mathbf{G}w_1$ and $y_2 = \mathbf{G}w_2$.

Specifically, let $\mathbf{X}_c : \mathcal{L} \to \mathcal{L}_{\mathbb{C}}$ be the *dual mixer* mapping $x(\cdot)$ to $e(t) = \exp(-j\omega_c t)x(t)$. Let $\mathbf{E} : \mathcal{L}_{\mathbb{C}} \to \ell$ be the *sampler* map $g(\cdot) \mapsto w[\cdot]$ such that $w[n] = g(nT)$. Finally, let $\mathbf{A}_0$ be the DT LTI system with frequency response A_0 defined by $A_0(\Omega) = P(\Omega/T)^{-1}$ for $|\Omega| < \pi$, where P is the Fourier transform of $p = p(t)$ (1). Then the composition $\mathbf{A}_0\mathbf{E}\mathbf{H}_0\mathbf{X}_c\mathbf{H}\mathbf{M}_0$ is an identity map. Equivalently, $\mathbf{A}_0\mathbf{E}\mathbf{H}_0\mathbf{X}_c$ is the ideal demodulator for $\mathbf{M}_0$.

For the modulation map $\mathbf{M} = \mathbf{X}\mathbf{Z}$ considered in this paper, the ideal demodulator has the form $\mathbf{A}\mathbf{E}\mathbf{H}_0\mathbf{X}_c$, where $\mathbf{A} : \ell \to \ell$ is the linear system mapping $w \in \ell$ to $s \in \ell$ according to

$$\begin{aligned} \mathrm{Re}(s) &= \mathbf{A}_{rr}\mathrm{Re}(w) + \mathbf{A}_{ri}\mathrm{Im}(w), \\ \mathrm{Im}(s) &= \mathbf{A}_{ir}\mathrm{Re}(w) + \mathbf{A}_{ii}\mathrm{Im}(w), \end{aligned} \tag{4}$$

and $\mathbf{A}_{rr}$, $\mathbf{A}_{ri}$, $\mathbf{A}_{ir}$, $\mathbf{A}_{ii}$ are LTI systems with frequency responses $A_{rr} = (P_0 - P_i)Q$, $A_{ir} = A_{ri} = -P_qQ$, $A_{ii} = (P_0 + P_i)Q$, where $Q = (P_0^2 - P_i^2 - P_q^2)^{-1}$, $P_i = (P^+ + P^-)/2$, $P_q = (P^+ - P^-)/2j$, and 2π-periodic functions $P_0, P^+, P^- : \mathbb{R} \to \mathbb{C}$ are defined for $|\Omega| < \pi$ by

$$P_0(\Omega) = P(\Omega/T),\ P^+(\Omega) = P_0(\Omega+\theta),\ P^-(\Omega) = P_0(\Omega-\theta)$$

with $\theta = 4\pi M$.

B. Equivalent Baseband Model

Before stating the main result of this paper, let us introduce some additional notation. For $d \in \mathbb{N}$ and $\tau = (\tau_1, \ldots, \tau_d) \in [0,\infty)^d$ let $\mathbf{F}_\tau : \mathcal{L} \to \mathcal{L}$ be the CT system mapping input $x \in \mathcal{L}$ to the output $y \in \mathcal{L}$ according to

$$y(t) = x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_d). \tag{5}$$

In the rest of this section, many expressions will contain products of the above type, where the complex-valued signal x can be written as $x = I + jQ$, with I and Q representing its real and imaginary part, respectively. Therefore, such terms, as in (5), would correspond to products of delayed real and imaginary parts of x. As will be shown later (e.g., in (12)), the factors in these products can be classified into four groups: real and imaginary parts of two differently delayed versions of x. This explains appearance of the index set $[1:4]$ which will be used to encode these four groups of signals.

For every tuple $\mathbf{m} = (m_1, \ldots, m_d) \in [1:4]^d$ and integer $l \in [1:4]$ let $S^l_{\mathbf{m}}$ be the set of all indices i for which $m_i = l$, i.e., $S^l_{\mathbf{m}} = \{i \in [1:d] :\ m_i = l\}$. Furthermore, define

$$N^1_{\mathbf{m}} = |S^1_{\mathbf{m}} \cup S^2_{\mathbf{m}}|, \quad N^2_{\mathbf{m}} = |S^3_{\mathbf{m}} \cup S^4_{\mathbf{m}}|.$$

Clearly $N^1_{\mathbf{m}} + N^2_{\mathbf{m}} = d$ for every $\mathbf{m} \in [1:4]^d$. Let $R^c_{\mathbf{m}} = \{-1,1\}^{N^1_{\mathbf{m}}}$ and $R^s_{\mathbf{m}} = \{-1,1\}^{N^2_{\mathbf{m}}}$. Let $(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ denote the standard scalar product in $\mathbb{R}^d$. Define the maps $\tilde{\sigma}, \sigma : \mathbb{R}^d \to \mathbb{R}$ by $\tilde{\sigma}(x) = \sum_{i=1}^d x_i$ and $\sigma(x) = \tilde{\sigma}(x) - 1$. For a given $\mathbf{m} \in [1:4]^d$ let $\Pi : \mathbb{R}^{N^2_{\mathbf{m}}} \to \mathbb{R}$ be defined as $\Pi(x) = \prod_{i=1}^{N^2_{\mathbf{m}}} x_i$. For $i \in \{1,2\}$, define projection operators $\mathcal{P}^i_{\mathbf{m}} : \mathbb{R}^d \to \mathbb{R}^{N^i_{\mathbf{m}}}$ by $\mathcal{P}^i_{\mathbf{m}}x = \begin{bmatrix} x_{n_1} & \cdots & x_{n_{N^i_{\mathbf{m}}}} \end{bmatrix}^T$ where $\{n_1, \ldots, n_{N^i_{\mathbf{m}}}\} = S^{2i-1}_{\mathbf{m}} \cup S^{2i}_{\mathbf{m}}$, $n_1 < \cdots < n_{N^i_{\mathbf{m}}}$. The following example should elucidate the above, somewhat involved, notation. Let $d = 7$, $\mathbf{m} = (3,1,4,2,1,3,1) \in [1:4]^7$. Then

$$S^1_{\mathbf{m}} = \{2,5,7\}, \quad S^2_{\mathbf{m}} = \{4\}, \quad S^3_{\mathbf{m}} = \{1,6\}, \quad S^4_{\mathbf{m}} = \{3\},$$

$$N^1_{\mathbf{m}} = |S^1_{\mathbf{m}} \cup S^2_{\mathbf{m}}| = 4, \quad N^2_{\mathbf{m}} = |S^3_{\mathbf{m}} \cup S^4_{\mathbf{m}}| = 3,$$

$$R^c_{\mathbf{m}} = \{-1,1\}^4, \quad R^2_{\mathbf{m}} = \{-1,1\}^3,$$

$$\mathcal{P}^1_{\mathbf{m}}x = \begin{bmatrix} x_2 & x_4 & x_5 & x_7 \end{bmatrix}^T, \quad \mathcal{P}^2_{\mathbf{m}}x = \begin{bmatrix} x_1 & x_3 & x_6 \end{bmatrix}^T.$$

Given a vector $\tau \in [0,\infty)^d$ let $\mathbf{k}$ be the unique vector in $(\mathbb{N} \cup \{0\})^d$ such that $\tau = \mathbf{k}T + \tau'$ and $\tau' \in [0,T)^d$.

Let $\theta : \mathbb{R} \to \{0,1\}$ denote the Heaviside step function $\theta(t) = 0$ for $t < 0$, $\theta(t) = 1$ for $t \geq 0$. For $T \in (0,\infty)$ let $p(t) = \theta(t) - \theta(t-T)$ denote the basic pulse shape of the zero-order hold (ZOH) system with sampling time T. Given $\mathbf{m} \in [1:4]^d$ and $\tau' \in [0,T)^d$ define

$$\tau^{\mathbf{m}}_{min} = \begin{cases} \max_{i \in S^2_{\mathbf{m}} \cup S^4_{\mathbf{m}}} \tau'_i, & |S^2_{\mathbf{m}} \cup S^4_{\mathbf{m}}| > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

and

$$\tau^{\mathbf{m}}_{max} = \begin{cases} \min_{i \in S^1_{\mathbf{m}} \cup S^3_{\mathbf{m}}} \tau'_i, & |S^1_{\mathbf{m}} \cup S^3_{\mathbf{m}}| > 0, \\ T, & \text{otherwise}. \end{cases} \tag{7}$$

Let $p_{\mathbf{m},\tau} : \mathbb{R} \to \mathbb{R}$ be the continuous time signal defined by

$$p_{\mathbf{m},\tau}(t) = \begin{cases} \theta(t - \tau^{\mathbf{m}}_{min}) - \theta(t - \tau^{\mathbf{m}}_{max}), & \tau^{\mathbf{m}}_{min} < \tau^{\mathbf{m}}_{max} \\ 0, & \text{otherwise}, \end{cases}. \tag{8}$$

We denote its Fourier transform by $P_{\mathbf{m},\tau}(\omega)$.

As can be seen from (3), the general CT Volterra model is a linear combination of subsystems $\mathbf{F}_\tau$, with different values of τ. Therefore, in order to establish the desired decomposition $\mathbf{S} = \mathbf{L}\mathbf{V}$ it is sufficient to consider the case $\mathbf{S}_\tau = \mathbf{D}\mathbf{H}\mathbf{F}_\tau\mathbf{M}$ with an arbitrary, but fixed, τ. The following theorem provides description of the underlying subsystems $\mathbf{V}$ and $\mathbf{L}$ of the aforementioned decomposition.

Theorem 4.1. For $\tau \in [0,\infty)^d$, the system $\mathbf{D}\mathbf{H}\mathbf{F}_\tau\mathbf{M}$ maps $w \in \ell$ to

$$v = \mathbf{A}u \in \ell, \quad \text{with} \quad u = \sum_{\mathbf{m} \in [1:4]^d} s_{\mathbf{m},\mathbf{k}} * g_{\mathbf{m}},$$

where

$$\begin{aligned} s_{\mathbf{m},\mathbf{k}}[n] = \prod_{i \in S^1_{\mathbf{m}}} I[n-k_i-1] \cdot \prod_{i \in S^2_{\mathbf{m}}} I[n-k_i] \cdot \\ \cdot \prod_{i \in S^3_{\mathbf{m}}} Q[n-k_i-1] \cdot \prod_{i \in S^4_{\mathbf{m}}} Q[n-k_i], \end{aligned}$$

$$I[n] = \mathrm{Re}(w[n]), \quad Q[n] = \mathrm{Im}(w[n]),$$

and the sequences (unit sample responses) $g_{\mathbf{m}} = g_{\mathbf{m}}[n]$ are defined by their Fourier transforms

$$G_{\mathbf{m}}(\Omega) = \frac{(j)^{N^2_{\mathbf{m}}}}{2^d} \sum_{r_c \in R^c_{\mathbf{m}}} \sum_{r_s \in R^s_{\mathbf{m}}} \Pi(r_c) P_{\mathbf{m},\tau}\left(\tilde{\Omega}\right) e^{-j\omega_c\tilde{\tau}}, \tag{9}$$

$$\tilde{\tau} = (r_c, \mathcal{P}^1_{\mathbf{m}}\tau') + (r_s, \mathcal{P}^2_{\mathbf{m}}\tau'),$$

$$\tilde{\Omega} = \frac{\Omega}{T} - \omega_c \sum_{i=1}^{N^1_{\mathbf{m}}} r_c(i) - \omega_c \sum_{l=1}^{N^2_{\mathbf{m}}} r_s(l) + \omega_c.$$

Proof. We first state and prove the following Lemma, that is a special case of Theorem 4.1, in which τ ranges over $[0,T)^d$ (instead of $\tau \in [0,\infty)^d$), and hence $\mathbf{k} = \mathbf{0}$. The proof of Theorem 4.1 then readily follows from this Lemma.

Lemma 4.2. The DT system $\mathbf{DHF}_\tau\mathbf{M}$ with $\tau \in [0,T)^d$ maps $w \in \ell$ to

$$v = \mathbf{A}u \in \ell, \quad \text{with} \quad u = \sum_{\mathbf{m}\in[1:4]^d} s_{\mathbf{m}} * g_{\mathbf{m}},$$

where

$$s_{\mathbf{m}}[n] = (I[n-1])^{|S^1_{\mathbf{m}}|}(I[n])^{|S^2_{\mathbf{m}}|}(Q[n-1])^{|S^3_{\mathbf{m}}|}(Q[n])^{|S^4_{\mathbf{m}}|},$$
$$I[n] = \mathrm{Re}(w[n]), \quad Q[n] = \mathrm{Im}(w[n]),$$

and the sequences $g_{\mathbf{m}}$ are as defined in Theorem 4.1.

Proof. A block diagram of system $\mathbf{DHF}_\tau\mathbf{M}$ is shown in Fig. 3, where $\mathbf{M}$ and $\mathbf{D}$ are decomposed into elementary subsystems as defined in the previous section. The proof of Lemma 4.2 consists of two steps. First, we express signal y (see Fig. 3) as a function of I, Q, ω_c, T and τ, which would imply description of $s_{\mathbf{m}}$ as given in the statement of Lemma 4.2. Then, by finding relationship between the Fourier transforms of signals y and u, we determine frequency responses $G_{\mathbf{m}}$, which concludes the proof of Lemma 4.2.

Consider first the case $d = 1$ (i.e., $\mathbf{F}_\tau$ is just a delay by $\tau \in [0,T)$). By definition, the outputs w_c, x and y of $\mathbf{Z}, \mathbf{X}$ and $\mathbf{F}$, respectively, are given by

$$w_c(t) = \frac{1}{T}\sum_n w[n]p(t-nT) = \underbrace{\frac{1}{T}\sum_n I[n]p(t-nT)}_{I_c(t)} + \underbrace{\frac{j}{T}\sum_n Q[n]p(t-nT)}_{jQ_c(t)}.$$
$$x(t) = (\mathbf{X}w_c)(t) = \mathrm{Re}\{\exp(j\omega_c t)w_c(t)\},$$
$$y(t) = I_c(t-\tau)\cos(\omega_c t - \omega_c\tau) - Q_c(t-\tau)\sin(\omega_c t - \omega_c\tau). \tag{10}$$

Consider the representation $p(t) = p_{1,\tau}(t) + p_{2,\tau}(t)$, where

$$p_{1,\tau}(t) = \theta(t) - \theta(t-\tau), \qquad p_{2,\tau}(t) = \theta(t-\tau) - \theta(t-T).$$

Let $\mathbf{Z_1} : \ell \to \mathcal{L}_\mathbb{C}$ and $\mathbf{Z_2} : \ell \to \mathcal{L}_\mathbb{C}$ be the pulse amplitude modulators with pulse shapes $p_{1,\tau}$ and $p_{2,\tau}$, respectively. Let $\mathbf{B}$ denote the backshift function mapping $x \in \ell$ to $y = \mathbf{B}x \in \ell$, defined by $y[n] = x[n-1]$. Then

$$I_c(t-\tau) = e_{1,\tau}(t) + e_{2,\tau}(t), \quad Q_c(t-\tau) = e_{3,\tau}(t) + e_{4,\tau}(t), \tag{11}$$

where

$$e_{1,\tau} = \mathbf{Z_1}\mathbf{B}I, \text{ i.e., } e_{1,\tau}(t) = \frac{1}{T}\sum_n I[n-1]p_{1,\tau}(t-nT),$$
$$e_{2,\tau} = \mathbf{Z_2}I, \text{ i.e., } e_{2,\tau}(t) = \frac{1}{T}\sum_n I[n]p_{2,\tau}(t-nT),$$
$$e_{3,\tau} = \mathbf{Z_1}\mathbf{B}Q, \text{ i.e., } e_{3,\tau}(t) = \frac{1}{T}\sum_n Q[n-1]p_{1,\tau}(t-nT),$$
$$e_{4,\tau} = \mathbf{Z_2}Q, \text{ i.e., } e_{4,\tau}(t) = \frac{1}{T}\sum_n Q[n]p_{2,\tau}(t-nT). \tag{12}$$

According to (10)-(12), the output $y(t)$ of $\mathbf{F}_\tau$ can be expressed as:

$$y(t) = f_1(t) + f_2(t) + f_3(t) + f_4(t),$$

where

$$f_i(t) = \begin{cases} e_{i,\tau}(t)\cos(\omega_c t - \omega_c\tau), & i = 1,2 \\ -e_{i,\tau}(t)\sin(\omega_c t - \omega_c\tau), & i = 3,4 \end{cases}. \tag{13}$$

Therefore, subsystem $\mathbf{F}_\tau\mathbf{M}$, mapping $w[n]$ to $y(t)$, can be represented as a parallel interconnection of amplitude modulated delayed and undelayed in-phase and quadrature components of $w[n]$. This is shown in Fig. 4, where $\tilde{\mathbf{D}} = \mathbf{E}\mathbf{H_0}\mathbf{X_c}\mathbf{H}$.

Suppose now that order d of $\mathbf{F}_\tau$ is an arbitrary positive integer larger than 1, i.e., that $\mathbf{F}_\tau : x \mapsto y$ defined by $y(t) = x(t-\tau_1)\cdot\cdots\cdot x(t-\tau_d)$. Then the output y of $\mathbf{F}_\tau$ can be written as a product

$$y(t) = y_1(t)\cdot y_2(t)\cdot\ldots\cdot y_d(t), \tag{14}$$

where, for all $i \in [1:d]$,

$$y_i(t) = I_c(t-\tau_i)\cos(\omega_c t - \omega_c\tau_i) - Q_c(t-\tau_i)\sin(\omega_c t - \omega_c\tau_i).$$

This implies that, for each i, signal $y_i(t)$ can be represented as the output of subsystem $\mathbf{F}_{\tau_i}\mathbf{M}$, where $\mathbf{F}_{\tau_\mathbf{i}}$ is just a simple delay, as discussed in the case $d = 1$. Therefore, system $\mathbf{F}_\tau\mathbf{M}$, mapping w to y, can be described as shown in Fig. 5. Hence, by using the same notation as in Figs. 4 and 5, signal $y(t)$ can be expressed as

$$y(t) = \prod_{i=1}^d [f_1^i(t) + f_2^i(t) + f_3^i(t) + f_4^i(t)] = \sum_{\mathbf{m}\in[1:4]^d} f_{\mathbf{m}}(t), \tag{15}$$

where

$$f_{\mathbf{m}}(t) = f^1_{m_1}(t)\cdot\ldots\cdot f^d_{m_d}(t), \quad \forall \mathbf{m} \in [1:4]^d.$$

Here components m_i of $\mathbf{m} = (m_1,\ m_2,\ \ldots,\ m_d) \in [1:4]^d$, determine which signal f^i_j, $j \in [1:4]$ from (13) participates as a product factor in $f_{\mathbf{m}}(t)$. With signals $e_{m_i,\tau_i}(t)$ as defined in (12), it follows that summands in (15) can be written as

$$f_{\mathbf{m}}(t) = (-1)^{N^2_{\mathbf{m}}}\prod_{i=1}^d e_{m_i,\tau_i}(t)\cdot \prod_{k\in S^1_{\mathbf{m}}\cup S^2_{\mathbf{m}}}\cos(\omega_c t - \omega_c\tau_k)\cdot\prod_{l\in S^3_{\mathbf{m}}\cup S^4_{\mathbf{m}}}\sin(\omega_c t - \omega_c\tau_l). \tag{16}$$

Products of cosines and sines in (16), denoted Π_c and Π_s, can be expressed as sums of complex exponents:

$$\Pi_c = \frac{1}{2^{N^1_{\mathbf{m}}}}\sum_{r\in R^c_{\mathbf{m}}} e^{j\omega_c\bar{\sigma}(r)t}\cdot e^{-j\omega_c(r,\mathcal{P}^1_{\mathbf{m}}\tau)}, \tag{17}$$

$$\Pi_s = \frac{1}{(2j)^{N^2_{\mathbf{m}}}}\sum_{r\in R^s_{\mathbf{m}}} \Pi(r)\cdot e^{j\omega_c\bar{\sigma}(r)t}\cdot e^{-j\omega_c(r,\mathcal{P}^2_{\mathbf{m}}\tau)}. \tag{18}$$

Recall that signals $e_{m_i,\tau_i}(t)$ are obtained by applying pulse amplitude modulation with pulse signals $p_{1,\tau_i}(t)$ or $p_{2,\tau_i}(t)$ on in-phase or quadrature components I and Q of the input signal (or their delayed counterparts $\mathbf{B}I$ and $\mathbf{B}Q$). Let $e_{\mathbf{m},\tau}(t)$ be the product of signals $e_{m_i,\tau_i}(t)$ (as given in (16)). We now derive

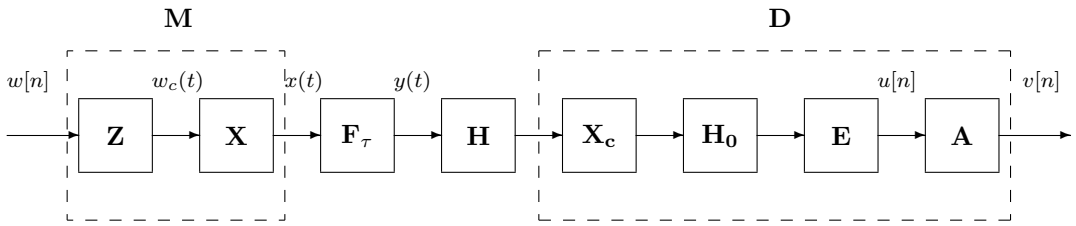


Fig. 3. Block diagram of system $\mathbf{S}_\tau = \mathbf{DHF}_\tau\mathbf{M}$, with all corresponding subsystems.

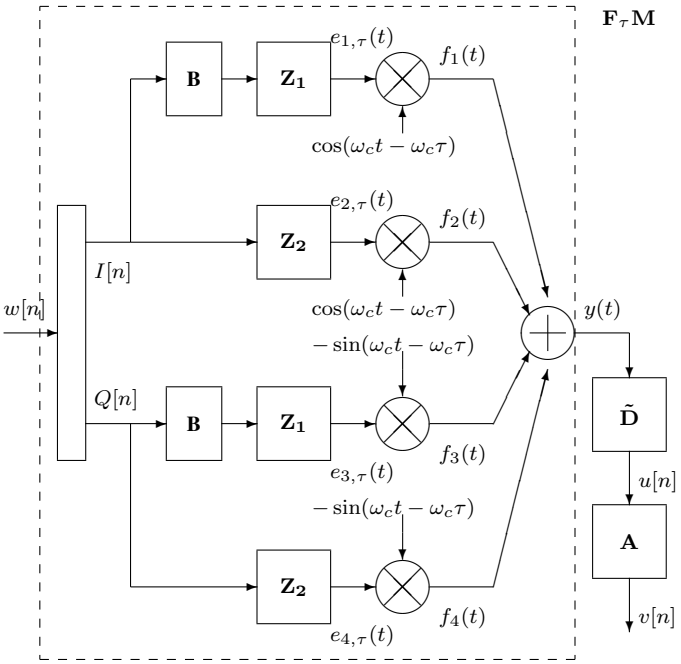


Fig. 4. Equivalent representation of system $\mathbf{DHF}_\tau\mathbf{M}$.

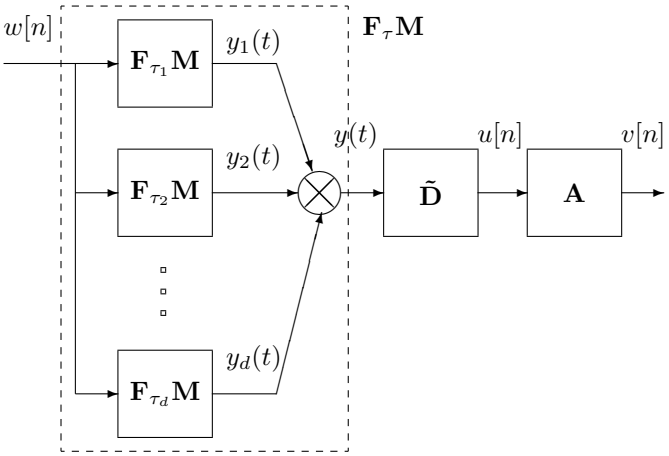


Fig. 5. System $\mathbf{F}_\tau\mathbf{M}$ as an interconnection of subsystems $\mathbf{F}_{\tau_i}\mathbf{M}$.

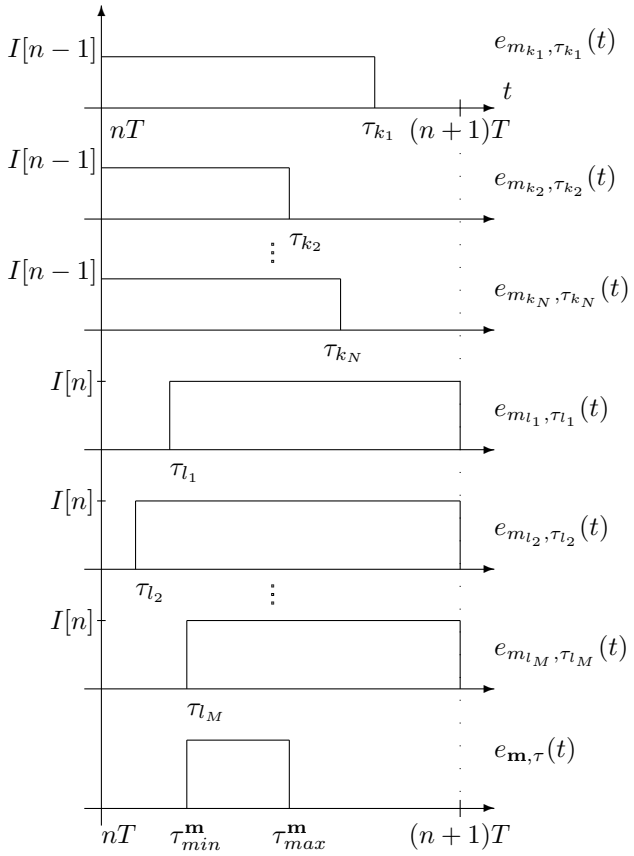


Fig. 6. Signal $e_{\mathbf{m},\tau}$ for $S^1_{\mathbf{m}} \cup S^3_{\mathbf{m}} = \{k_1, k_2, \ldots, k_N\}$ and $S^2_{\mathbf{m}} \cup S^4_{\mathbf{m}} = \{l_1, l_2, \ldots, l_M\}$, where $N + M = d$.

an expression for $e_{\mathbf{m},\tau}(t)$ as a function of signals $I, Q, \mathbf{B}I$ and $\mathbf{B}Q$. We first investigate $e_{\mathbf{m},\tau}(t)$ for $t \in [nT, (n+1)T)$, with $n > 1$ an integer. For a fixed $\mathbf{m} \in [1:4]^d$ and $\tau \in (0,T)^d$ let $S^1_{\mathbf{m}} \cup S^3_{\mathbf{m}}$, $S^2_{\mathbf{m}} \cup S^4_{\mathbf{m}}$, $\tau^{\mathbf{m}}_{min}$ and $\tau^{\mathbf{m}}_{max}$ be as defined at the beginning of this section. It follows that $e_{\mathbf{m},\tau}(t) = 0$ for all $t \in [nT, (n+1)T)$ if $\tau^{\mathbf{m}}_{min} > \tau^{\mathbf{m}}_{max}$. Otherwise it is nonzero for $t \in [nT + \tau^{\mathbf{m}}_{min}, nT + \tau^{\mathbf{m}}_{max})$. This is depicted in Fig. 6 (for the sake of simplicity, only in-phase components I and $\mathbf{B}I$ are considered, but in general signals Q and $\mathbf{B}Q$ would appear as well). It follows from the above discussion that signal $e_{\mathbf{m},\tau}(t)$ can be expressed as

$$e_{\mathbf{m},\tau}(t) = \sum_{n=-\infty}^{\infty} s_{\mathbf{m}}[n] p_{\mathbf{m},\tau}(t - nT), \tag{19}$$

where $p_{\mathbf{m},\tau}(t)$ was defined in (8), and DT signal $s_{\mathbf{m}} = s_{\mathbf{m}}[n]$

is defined as

$$s_{\mathbf{m}}[n] = I[n]^{|S^1_{\mathbf{m}}|} \cdot I[n-1]^{|S^2_{\mathbf{m}}|} \cdot Q[n]^{|S^3_{\mathbf{m}}|} \cdot Q[n-1]^{|S^4_{\mathbf{m}}|}.$$

From (16)-(19), it follows that $f_{\mathbf{m}}(t)$ can be written as

$$f_{\mathbf{m}}(t) = \sum_{r_c \in R^c_{\mathbf{m}}} \sum_{r_s \in R^s_{\mathbf{m}}} C_{r_c,r_s} \cdot e^{j\sigma(r_c,r_s)\omega_c t} \cdot \cdot \sum_{n=-\infty}^{\infty} s_{\mathbf{m}}[n] p_{\mathbf{m},\tau}(t-nT), \quad (20)$$

where $\sigma(r_c, r_s) = \sigma(r) = \sum_{i=1}^{N^1_{\mathbf{m}}} r_c(i) + \sum_{l=1}^{N^2_{\mathbf{m}}} r_s(l)$, and

$$C_{r_c,r_s} = \frac{(j)^{N^2_{\mathbf{m}}}}{2^d} \cdot e^{-j\omega_c[(r_c,\mathcal{P}^1_{\mathbf{m}}\tau)+(r_s,\mathcal{P}^2_{\mathbf{m}}\tau)]} \cdot \Pi(r_s), \quad (21)$$

depends only on $\mathbf{m}$. Therefore, the output signal y of system $\mathbf{F}_\tau\mathbf{M}$, can be expressed in terms of I, Q, ω_c. T and τ by plugging expression (20), for $f_{\mathbf{m}}(t)$, into (15). Thus we have found an explicit input-output description of system $\mathbf{F}_\tau\mathbf{M}$, which concludes the first part of the proof.

In order to find explicit expressions for the frequency responses $G_{\mathbf{m}}$, we first determine frequency domain relationship between signals u and y (see Fig. 3). Recall that $u = \tilde{\mathbf{D}}y = \mathbf{E}\mathbf{H_0}\mathbf{X_c}\mathbf{H}y$. Let $U(\Omega)$ and $Y(\omega)$ denote the Fourier transforms of signals $u[n]$ and $y(t)$ respectively. Also let $H(\omega)$ and $H_0(\omega)$ be the frequency responses of ideal band-pass and low-pass filters $\mathbf{H}$ and $\mathbf{H_0}$, given by

$$H(\omega) = \begin{cases} 1, & |\omega_c - |\omega|| \le \pi/T \\ 0, & \text{otherwise} \end{cases}, H_0(\omega) = \begin{cases} 1, & |\omega| \le \pi/T \\ 0, & \text{otherwise} \end{cases}. \quad (22)$$

The following sequence of equalities hold

$$\begin{aligned} \mathcal{F}\{\mathbf{H}y\} &= Y(\omega)H(\omega), \\ \mathcal{F}\{\mathbf{X_c}\mathbf{H}y\} &= Y(\omega+\omega_c)H(\omega+\omega_c), \\ \mathcal{F}\{\mathbf{H_0}\mathbf{X_c}\mathbf{H}y\} &= Y(\omega+\omega_c)H(\omega+\omega_c)H_o(\omega), \\ U(\Omega) &= Y\left(\frac{\Omega}{T}+\omega_c\right)H\left(\frac{\Omega}{T}+\omega_c\right)H_0\left(\frac{\Omega}{T}\right). \end{aligned}$$

From the definition of $H(\omega)$ and $H_0(\omega)$, $U(\Omega)$ simplifies to

$$U(\Omega) = Y\left(\frac{\Omega}{T}+\omega_c\right), \quad (23)$$

which gives frequency domain relationship between y and u.

Next we express $Y(\omega)$ in terms of $S_{\mathbf{m}}(\Omega) = \mathcal{F}\{s_{\mathbf{m}}[n]\}$. For the sake of simplicity, we assume that $y(t)$ is equal to $f_{\mathbf{m}}(t)$ for some fixed $\mathbf{m}$, i.e., we omit the sum in (15). Since $\sigma(r) \in \mathbb{Z}$ and $\omega_c T = 2\pi n$, $n \in \mathbb{Z}$, it follows from (20) that

$$Y(\omega) = S_{\mathbf{m}}(\omega T) \cdot \sum_{r_c \in R^c_{\mathbf{m}}} \sum_{r_s \in R^s_{\mathbf{m}}} C_{r_c,r_s} P_{\mathbf{m},\tau}\left(\omega - \sigma(r)\omega_c\right). \quad (24)$$

It now follows from (23) and (24) that

$$U(\Omega) = S_{\mathbf{m}}(\Omega) \sum_{r_c \in R^c_{\mathbf{m}}} \sum_{r_s \in R^s_{\mathbf{m}}} C_{r_c,r_s} P_{\mathbf{m},\tau}\left(\tilde{\Omega}\right),$$

where $\tilde{\Omega} = \frac{\Omega}{T} - \sigma(r)\omega_c + \omega_c$ and C_{r_c,r_s} as defined in (21). Therefore, the frequency response $G_{\mathbf{m}}(\Omega)$ of a LTI system mapping $s_{\mathbf{m}}$ to u is given by

$$G_{\mathbf{m}}(\Omega) = \sum_{r_c \in R^c_{\mathbf{m}}} \sum_{r_s \in R^s_{\mathbf{m}}} C_{r_c,r_s} P_{\mathbf{m},\tau}\left(\frac{\Omega}{T} - \sigma(r)\omega_c + \omega_c\right). \quad (25)$$

This concludes the proof of Lemma 4.2. □

In Lemma 4.2, it was assumed that $\tau_i \in [0, T),\ \forall i \in [1:d]$, but in general τ_i can take any positive real value depending on the depth of (2), i.e., vector $\mathbf{k}$ associated with τ is not necessarily the zero vector. Suppose now that $\tau = \mathbf{k}T + \tau'$, where $\tau' \in [0, T)^d$, and $\mathbf{k} \neq \mathbf{0}$. In the rest of this proof we adopt the same notation for corresponding signals and systems as in the proof of Lemma 4.2. It is clear, in this case, that mapping from y to u is identical to the one derived for $\tau \in [0, T)$. Therefore, in order to prove the statement of Theorem 4.1 it suffices to determine a new relationship between signals w and y.

Let first $d = 1$, i.e., $\tau = kT + \tau'$, with $k \in \mathbb{N}$ and $\tau' \in [0, T)$. Analogously to the case in the proof of Lemma 4.2, it follows that signal y can be expressed as

$$y(t) = [e_{1,\tau}(t) + e_{2,\tau}(t)]\cos(\omega_c t - \omega_c\tau) - - [e_{3,\tau}(t) + e_{4,\tau}(t)]\sin(\omega_c t - \omega_c\tau),$$

where

$$\begin{aligned} e_{1,\tau} &= \mathbf{Z_1}\mathbf{B}^{k+1}I, \qquad e_{2,\tau} = \mathbf{Z_2}\mathbf{B}^{k}I, \\ e_{3,\tau} &= \mathbf{Z_1}\mathbf{B}^{k+1}Q, \qquad e_{4,\tau} = \mathbf{Z_2}\mathbf{B}^{k}Q, \quad \text{i.e.,} \\ e_{1,\tau}(t) &= \frac{1}{T}\sum_{n=-\infty}^{\infty} I[n-k-1]p_{1,\tau'}(t-nT), \\ e_{2,\tau}(t) &= \frac{1}{T}\sum_{n=-\infty}^{\infty} I[n-k]p_{2,\tau'}(t-nT), \\ e_{3,\tau}(t) &= \frac{1}{T}\sum_{n=-\infty}^{\infty} Q[n-k-1]p_{1,\tau'}(t-nT), \\ e_{4,\tau}(t) &= \frac{1}{T}\sum_{n=-\infty}^{\infty} Q[n-k]p_{2,\tau'}(t-nT). \end{aligned} \quad (26)$$

Here $\mathbf{B}^k$ denotes the composition of $\mathbf{B}$ with itself k times, i.e., $\mathbf{B}^k : x \mapsto y$ such that $y[n] = x[n-k]$.

For $d > 1$, reasoning similar to that in the proof of Lemma 4.2 (see (14)-(19)), leads to the following expression for $e_{\mathbf{m},\tau}$:

$$e_{\mathbf{m},\tau}(t) = \sum_{n=-\infty}^{\infty} s_{\mathbf{m},\mathbf{k}}[n] p_{\mathbf{m},\tau'}(t-nT), \quad (27)$$

where

$$s_{\mathbf{m},\mathbf{k}}[n] = \prod_{i \in S^1_{\mathbf{m}}} I[n-k_i-1] \cdot \prod_{i \in S^2_{\mathbf{m}}} I[n-k_i] \cdot \cdot \prod_{i \in S^3_{\mathbf{m}}} Q[n-k_i-1] \cdot \prod_{i \in S^4_{\mathbf{m}}} Q[n-k_i], \quad (28)$$

and $p_{\mathbf{m},\tau'}(t)$ as defined in (8). Let $S_{\mathbf{m},\mathbf{k}} = S_{\mathbf{m},\mathbf{k}}(\Omega)$ be the Fourier transform of $s_{\mathbf{m},\mathbf{k}}$. With (27) at hand, it is straightforward to find the analytic expression for $U = \mathcal{F}u$, in

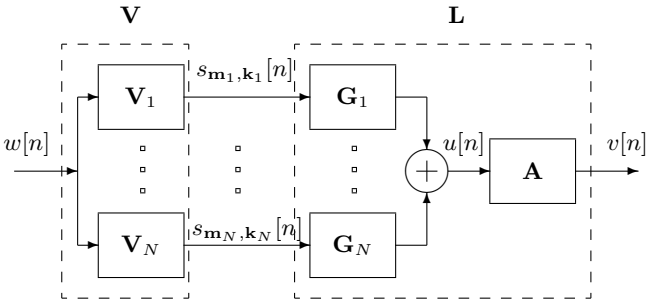


Fig. 7. Block diagram depicting the novel equivalent baseband model structure as defined in Theorem 4.1.

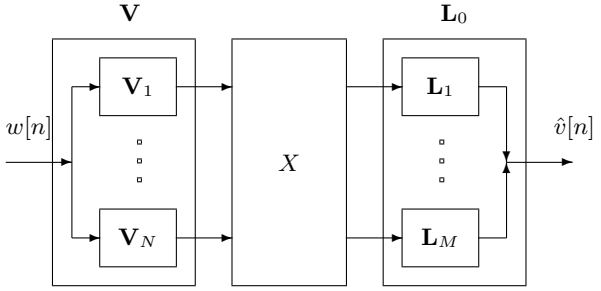


Fig. 8. Detailed block diagram of an approximate model $\hat{\mathbf{S}} = \mathbf{L}_0 X \mathbf{V}$ of $\mathbf{S}$.

terms of $S_{\mathbf{m},\mathbf{k}}$. Similarly to (20)-(24), the Fourier transform $Y = \mathcal{F}y$, can be written as

$$Y(\omega) = S_{\mathbf{m},\mathbf{k}}(\omega T) \cdot \sum_{r_c \in R^c_{\mathbf{m}}} \sum_{r_s \in R^s_{\mathbf{m}}} C_{r_c,r_s} P_{\mathbf{m},\tau}\left(\omega - \sigma(r)\omega_c\right), \tag{29}$$

where C_{r_c,r_s} as defined in (21), with τ replaced by τ'. It follows from (23) and (29) that

$$U(\Omega) = S_{\mathbf{m},\mathbf{k}}(\Omega) G_{\mathbf{m}}(\Omega), \tag{30}$$

with $G_{\mathbf{m}}(\Omega)$ as defined in (25). Statement of the theorem now immediately follows from the above equality. This concludes the proof. □

Block diagram of the proposed equivalent baseband model of $\mathbf{S}_\tau = \mathbf{DHF}_\tau\mathbf{M}$, as suggested by the statement of Theorem 4.1, is shown in Fig. 7. Therefore, system $\mathbf{S}_\tau$ can be represented as a series interconnection $\mathbf{S}_\tau = \mathbf{LV}$, of systems $\mathbf{V}$ and $\mathbf{L}$, where components $\mathbf{V}_i$ of $\mathbf{V}$ map complex scalar-valued input $w \in \ell$ to real scalar-valued outputs $s_{\mathbf{m}_i,\mathbf{k}_i} \in \ell$ as defined in (28), and $\mathbf{L}$ maps real vector-valued input $s = (s_{\mathbf{m}_i,\mathbf{k}_i})_{i=1}^N \in \ell^N$ to complex scalar-valued output $v \in \ell$ as defined in (25) and (4).

It is not hard to see, from (25) and (4), that frequency responses of the reconstruction filters $\mathbf{L}_i = \mathbf{AG}_i$ are discontinuous at frequencies $\Omega = \pm\pi$. Indeed, this is to be expected, since $\mathbf{L}$ represents a baseband equivalent of the linear part of the response of system $\mathbf{F}$ over the frequency interval $(\omega_c - \pi/T, \omega_c + \pi/T)$, which is, in general, not symmetric with respect to ω_c. Therefore, the frequency responses $L_i(\Omega)$ of $\mathbf{L}_i$ should, in general, be discontinuous at $\Omega = \pm\pi$. This further implies infinite memory of $\mathbf{L}_i$ in time-domain. That is, the unit sample responses of $\mathbf{L}_i$ are of infinite length, which is impracticable for hardware implementation. Nevertheless, from (25) and (4), it follows that $L_i(\omega)$ are smooth functions of Ω on the interval $(-\pi, \pi)$, and hence can be well approximated by polynomials (or some other appropriate basis) on compact subintervals of $(-\pi, \pi)$. This suggests an approximate model $\hat{\mathbf{S}} = \mathbf{L}_0 X \mathbf{V}$ of $\mathbf{S}$, as shown in Fig. 8. In this case, frequency responses of components of $\mathbf{L}_0$ are elements of a polynomial basis (1, $j\Omega$, Ω^2, etc). and X is a matrix of coefficients. By increasing degree of the polynomial approximating $\mathbf{L}$ (i.e., increasing the number of components in $\mathbf{L}_0$), the modeling error $||\mathbf{S} - \hat{\mathbf{S}}||$ can be made arbitrarily small.

V. Discussion

A. Effects of oversampling

The analytical result of this paper suggests a special structure of a digital pre-distortion compensator which appears to be, in first approximation, both necessary and sufficient to match the discrete time dynamics resulting from combining modulation and demodulation with a dynamic non-linearity in continuous time. The "necessity" somewhat relies on the input signal u having "full" spectrum. Let us describe this in more detail. As reported in the previous section, frequency responses $L_i(\Omega)$ of $\mathbf{L_i}$, $i = \{1, \ldots, M\}$, are discontinuous at frequencies $\Omega = \pm\pi$, but are smooth on the interval $(-\pi, \pi)$ and therefore approximable by some appropriate basis (e.g., polynomials in Ω) on compact subintervals of $(-\pi, \pi)$. Moreover, these basis elements (e.g. frequency responses are monomials in Ω, in the case of polynomial basis) can be well approximated by FIR filters. When desired approximation accuracy is fixed, order of such an FIR filter mainly depends on the ratio of the baseband signal bandwidth B_w and baseband sampling rate f_{dac} (i.e., rate of the DAC). Let us denote this ratio as $\xi = B_w/f_{dac} = B_w T$ (e.g., $\xi = 1$ corresponds to the case when spectrum of the baseband input signal occupies the whole Nyquist band). For ξ close to 1, if one wants to achieve good approximation of $\mathbf{L_i}$, the frequency responses of the approximate FIR filters must have sharp transition region (see Fig. 9a), which leads to high order (large number of taps) of the corresponding filters. Therefore, memory of the approximate reconstruction filter $\mathbf{L_0}$, and correspondingly approximate system $\hat{\mathbf{S}}$, will be long. Contrary to that, if ξ is relatively small compared to 1 (e.g., $1/2$ as in Fig. 9b), in order to transmit symbol information without distortion, the reconstruction filter $\mathbf{L_0}$ has to match the frequency response of the ideal baseband model LTI filter $\mathbf{L}$ only on the effective band defined by ξ, as shown in Fig. 9b. Therefore, reconstruction filters $\mathbf{L_0}$ can have smooth transitions and are realizable with low order FIR filters, which suggest that $\hat{\mathbf{S}}$ will have short memory as well.

Standard practice in transceiver design is to oversample baseband signal (symbols), and shape its spectrum (samples), before it is modulated onto a carrier [30]. In the case of large oversampling ratios (OSR), from symbol to sample space, the effective band of the signal containing symbol information is small compared to the rate of a transceiver DAC, i.e., ratio ξ is much smaller than 1. This implies that a plain Volterra

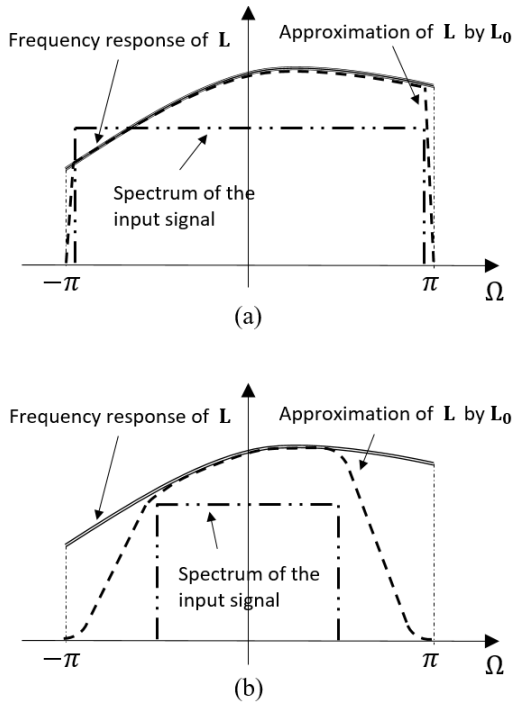


Fig. 9. Simplified spectral diagrams that show how memory of the approximate reconstruction filter $\mathbf{L_0}$ depends on ratio ξ.

structure with relatively short memory can capture dynamics of such a system well enough. A common implementation of amplitude-phase modulation will frequently employ a signal component separation approach (also known as out-phasing), such as LINC [31], where the low-pass signal u is decomposed into two components of constant amplitude, $u = u_1 + u_2$, $|u_1[n]| \equiv |u_2[n]| = \text{const}$, after which the components u_i are fed into two separate modulators, to produce continuous time outputs y_1, y_2, to be combined into a single output $y = y_1 + y_2$. Even when u is band-limited, the resulting components u_1, u_2 are not, and the full range of modulator's nonlinearity is likely to be engaged when producing y_1 and y_2. Furthermore, in forthcoming wideband communication systems, OSR is limited by the speed that digital baseband and DAC are able to sustain, and low OSR is more likely to be encountered, therefore emphasizing significance of a baseband model derived in this paper.

B. Impact of low-pass filtering after zero order hold DAC

In Section III, we assume that digital-to-analog conversion (DAC) at the modulator $\mathbf{M}$ is performed using a zero order hold system. However, a common assumption in many communication systems is that DAC is modeled as a series interconnection of a ZOH and a low-pass filter, in order to suppress the high frequency harmonics caused by discontinuous nature of the ZOH pulse shape. Now, let this new modulator $\mathbf{M_0}$ be defined as a series interconnection $\mathbf{M_0} = \mathbf{X}\mathbf{H_{dac}}\mathbf{Z}$, where $\mathbf{X}$ and $\mathbf{Z}$ are the mixing and zero order hold systems, as defined in (1)-(2), and $\mathbf{H_{dac}}$ is an LTI system with memory equal to $m_0 > 0$. Let the memory and degree of the passband nonlinearity $\mathbf{F}$ be equal to m and d, respectively. Block

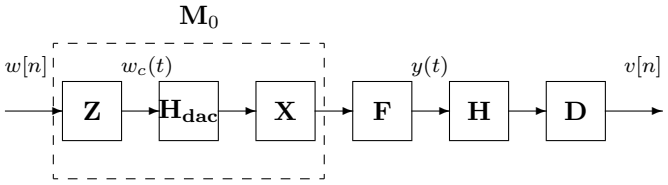
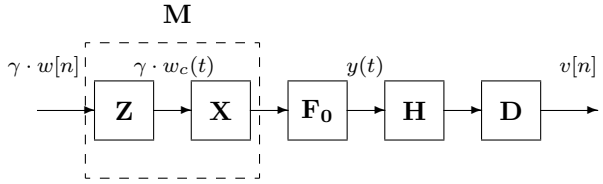


Fig. 10. Block diagram of a modified system $\mathbf{S_0}$.

diagram of a modified system $\mathbf{S_0} = \mathbf{DHFM_0}$ is shown in Fig. 10. We argue that system $\mathbf{S_0}$ has an equivalent baseband model decomposition similar to that from Theorem 4.1, where the nonlinear part $\mathbf{V}$ has the same maximal degree but its memory is equal to $\lceil (m+m_0)/T \rceil$ (as opposed to $\lceil m/T \rceil$ in Theorem 4.1). Due to space constraints, we give only a simplified argument on why this should hold. Let us assume that the impulse response of $\mathbf{H_{dac}}$ is given by $h_{dac}(t) = c \cdot \delta(t - m_0)$ where $m_0 > 0, c \in \mathbb{R}$ and $\delta = \delta(t)$ is the Dirac delta function. Since

$$w_c(t - m_0)e^{j\omega_c t} = e^{j\omega_c m_0} w_c(t - m_0)e^{j\omega_c (t-m_0)},$$

it follows that $\mathbf{H_{dac}X} = \gamma \mathbf{X H_{dac}}$ where $\gamma = c \cdot e^{j\omega_c m_0}$, that is, systems $\mathbf{H_{dac}}$ and $\mathbf{X}$ commute up to a scaling factor. Hence, the system in Fig. 10 is equivalent to the following



where $\mathbf{F_0}$ is a series interconnection of a delay-by-m_0 and system $\mathbf{F}$. As a result, the memory and degree of $\mathbf{F_0}$ are equal to $m + m_0$ and d, respectively. This implies that $\mathbf{S_0}$ is equivalent to $\mathbf{S} = \mathbf{DHFM}$ from Section III, where $\mathbf{F}$ is replaced by $\mathbf{F_0}$ (with an additional scaling in baseband), and, according to Theorem 4.1, can be decomposed as $\mathbf{LV}$ where the maximal memory and degree of Volterra monomials in the nonlinear DT system $\mathbf{V}$ are equal to $\lceil (m + m_0)/T \rceil$ and d, respectively. It should be noted here that in the case of an arbitrary low-pass filter $\mathbf{H_{dac}}$ the particular formulas from Theorem 4.1 would not hold anymore (i.e. nonlinear terms in $\mathbf{V}$ and expressions for reconstruction filters in $\mathbf{L}$ would change), but the general system structure is preserved.

C. Extension to OFDM

Orthogonal frequency-division multiplexing (OFDM) is a multicarrier digital modulation scheme that has been the dominant technology for broadband multicarrier communications in the last decade. Compared with single-carrier digital modulation, by increasing the effective symbol length and employing many carriers for transmission, OFDM theoretically eliminates the problem of multi-path channel fading, which is the main type of disturbance on a terrestrial transmission path. It also mitigates low spectrum efficiency, impulse noise, and frequency selective fading [32]. One of the major drawbacks of OFDM is the relatively large Peak-to-Average

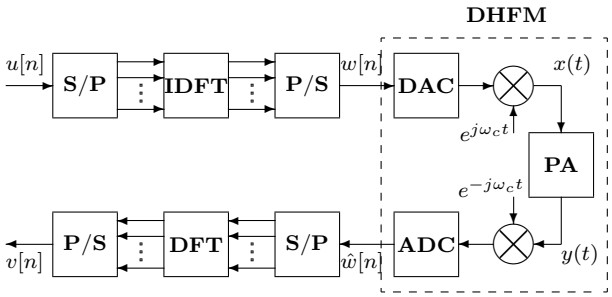


Fig. 11. Block diagram of a typical implementation of OFDM.

Power Ratio (PAPR) [33]. This makes OFDM very sensitive to the nonlinear distortion introduced by high PA, which causes in-band as well as out-of-band (i.e., adjacent channel) radiation, decreasing spectral efficiency [34]. For that reason, linearization techniques play very important role in OFDM and have been studied extensively (see e.g., [35]).

Fig. 11 shows a block diagram of the typical implementation of an N-carrier OFDM system. Input stream of symbols $u[n]$, with bandwidth B, is first converted into blocks of length N by serial-to-parallel conversion, which are then fed to an N-point inverse FFT block. Output of this block is then transformed with a parallel-to-serial converter into a stream of N samples $w[n]$, and further converted to analog domain and used to modulate a single carrier. As can be seen from Figure 11, sequence $w[n]$ is fed into a system which can be modeled as $\mathbf{S} = \mathbf{DHFM}$, i.e., the model investigated in the previous section. Since the choice of input symbols' values (e.g., QPSK, QAM, etc.), was not relevant to the derivation of the baseband model from Section IV, and hence input symbols can be arbitrary bounded complex numbers. Therefore, $w[n]$ can be considered as a legitimate input sequence to a system modeled as $\mathbf{DHFM}$. This implies that our baseband model, and its corresponding DPD structure, can be possibly used for distortion reduction in OFDM modulation applications.

VI. Simulation Results

In this section, aided by MATLAB simulations, we validate the proposed (approximate) baseband model and illustrate performance of a simple DPD structure based on this model. Simulations were performed for various models of passband nonlinearity, as explained below.

The general simulation parameters, used to obtain all results presented in this section, were chosen as follows. The input symbol sequence was generated as a 64QAM signal at 20MS/sec rate. The choice of 64QAM encoded baseband data is solely due to the ease of generating and processing such sequences in Matlab (as opposed to e.g., LTE signals). For the purpose of producing training and validation data, system $\mathbf{S}$ is simulated for two cases of baseband (or sampling) frequency of $f_{dac_1} = 20$MS/sec and $f_{dac_2} = 25$MS/sec. Therefore, the input signal bandwidth occupies 100% and 80% of observed bandwidth, respectively. After digital-to-analog conversion, the baseband input signal is up-converted to a passband carrier frequency of $f_c = 1$GHz, by a standard quadrature modulation (i.e., multiplication of I and Q components with $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$, respectively, where t is the time variable). Digital simulation of the passband part of system $\mathbf{S}$ was done by representing continuous-time signals by their discrete-time counterparts, obtained by sampling with very high sampling rate of $f_{ct} = 100 f_c$, so that spectral aliasing effects do not significantly affect modeling or linearization performance. Two cases of passband nonlinearity $\mathbf{F}$ were considered, and are described in detail in the next section. In order to capture the PA's output, bandpass filter $\mathbf{H}$, as well as the low-pass filter used in demodulation system $\mathbf{D}$, were realized as ideal rectangular filters. This was done in simulation by zeroing out the frequency response of the input signal outside the frequency band of interest. The length of input symbol sequences used for training, validation and DPD performance evaluation was $N_{symb} = 16,384$. Both model validation and DPD performance evaluation were done by Monte Carlo simulation with 100 simulation runs.

A. Passband Nonlinearity Model

Two models of passband nonlinearity were used to confirm the proposed baseband model, and evaluate performance of a compensator based on this model. In the first case, a cubic term, similar to those from Theorem 4.1, was added to the identity function in order to model passband nonlinearity $\mathbf{F}$, which is in this case defined as:

$$(\mathbf{F}x)(t) = x(t) - \delta \cdot x(t-\tau_1)x(t-\tau_2)x(t-\tau_3), \qquad (31)$$

where, without loss of generality, $0 \le \tau_1 \le \tau_2 \le \tau_3$ and $\delta > 0$ is a parameter controlling magnitude of distortion $\mathbf{\Delta}$ in $\mathbf{S} = \mathbf{I} + \mathbf{\Delta}$.

In the second case, subsystem $\mathbf{F}$ is modeled by a series interconnection of a linear combination of the identity and simple delay, and Cann's model [36], which is frequently used for behavioral modeling of power amplifiers:

$$(\mathbf{F}x)(t) = f\left((1-\rho)x(t) + \rho x(t-\tau)\right), \qquad (32)$$

where

$$f(a) = \frac{ga}{\left[1 + \left(\frac{ga}{L}\right)^s\right]^{\frac{1}{s}}}, \qquad (33)$$

is Cann's model with fixed parameters $L, s, g > 0$. In this case, parameter $\rho \in (0,1)$ controls the proportion of memoryless dependence in this passband nonlinearity.

B. Model Selection

Results from Section IV suggest that the approximate baseband model should be searched for within a family of models $\hat{\mathbf{S}} = \mathbf{L}_0 X \mathbf{V}$, as shown on the block diagram in Fig 8. Abbreviation L_0XV will be used in all figures and tables to denote our model. Reconstruction filters $\mathbf{L}$ are approximated by the basis $(\mathbf{L_{01}}, \mathbf{L_{02}}, \mathbf{L_{03}})$, where LTI subsystems $\mathbf{L}_{0i}, i \in \{1,2,3\}$, have frequency responses L_{0i} defined by

$$L_{01}(\Omega) = 1, \quad L_{02}(\Omega) = j\Omega, \quad L_{03}(\Omega) = \Omega^2, \quad \forall \Omega \in [-\pi, \pi),$$

and 2π-periodically extended for other values of Ω. In actual simulation, each basis element $\mathbf{L}_{0i}$ is realized (or, more precisely, approximated) as an FIR filter with 30 taps.

TABLE I
PARAMETER SELECTION FOR DIFFERENT APPROXIMATION MODELS

Cubic nonlinearity model					
Model	L_0XV	Volt. 1	Volt. 2	Volt. 3	Volt. 4
d	3	3	5	5	3
m_b	1	1	4	2	4
m_f	0	0	0	2	4
Modified Cann's model					
Model	L_0XV	Volt. 1	Volt. 2	Volt. 3	Volt. 4
d	7	7	7	5	5
m_b	1	1	2	4	2
m_f	0	0	0	0	2

Nonlinear system $\mathbf{V}$ is comprised of all Volterra monomials up to order d, with backward and forward memory m_b and m_f, respectively, i.e.,

$$(\mathbf{V}_i w)[n] = \prod_{k=m_b}^{m_f} I[n-k]^{\alpha_i(k)} \prod_{k=m_b}^{m_f} Q[n-k]^{\beta_i(k)},$$

for all

$$\alpha_i(k), \beta_i(k) \in \mathbb{Z}_+, \qquad \sum_{k=m_b}^{m_f} \alpha_i(k) + \sum_{k=m_b}^{m_f} \beta_i(k) \leq d,$$

where $I[n] = \mathrm{R}e\ w[n]$ and $Q[n] = \mathrm{I}m\ w[n]$.

The ability of our model to approximate system $\mathbf{S}$ is compared to that of a widely used model obtained by employing general Volterra series structure [3]:

$$(\hat{\mathbf{S}}_{\mathbf{v}} w)[n] = \sum_{(\alpha_i,\beta_i)} c_i \prod_{k=-m_b}^{m_f} I[n-k]^{\alpha_i(k)} \prod_{k=-m_b}^{m_f} Q[n-k]^{\beta_i(k)},$$

$$\alpha_i(k), \beta_i(k) \in \mathbb{Z}_+, \qquad \sum_{k=-m_b}^{m_f} \alpha_i(k) + \sum_{k=-m_b}^{m_f} \beta_i(k) \leq d,$$

and $(\alpha_i, \beta_i) = (\alpha_i(m_b), \ldots, \alpha_i(m_f), \beta_i(m_b), \ldots, \beta_i(m_f))$. We should remark here that the standard practice in literature on equivalent baseband modeling, and corresponding digital predistortion (see e.g., [8] or [12]), is to assume much simpler approximate Volterra model, which is comprised only of monomials with odd degree $\sum_{k=-m_b}^{m_f} \alpha_i(k) + \sum_{k=-m_b}^{m_f} \beta_i(k)$. This is justified by the assumed low-pass filtering (LPF) operation both after digital-to-analog conversion (DAC) at the transmitter side, and before demodulation and analog-to-digital conversion (ADC) at the receiver side. In this paper, an LPF after DAC is not assumed and therefore a full Volterra series model has to be taken into consideration if we hope to successfully approximate system $\mathbf{S}$.

For each passband nonlinearity, we consider four different Volterra based models, by varying degree d, and forward and backward memory depths m_f and m_b, respectively, as given in Table I. For each model, the corresponding coefficients (matrix X for our model, and vector (c_i) for Volterra models) are found by simple least squares optimization to fit the input-output data available. It should be noted that fitting has to be done for both real and imaginary part of $w[n]$.

C. Performance Evaluation

As a measure of quality of the approximate model, normalized mean square error (NMSE) metric is used, which, for a given input w, is defined as

$$\mathrm{NMSE}(\hat{\mathbf{S}}, w) = 20 \log_{10} \left(\frac{||\mathbf{S}w - \hat{\mathbf{S}}w||_2}{||\mathbf{S}w||_2} \right),$$

where $\mathbf{S}$ is the true system, and $\hat{\mathbf{S}}$ is the approximate model. We also evaluate performance of a simple compensator based on the proposed model. We assume that parameters δ (case 1) and ρ (case 2) are relatively small, so that the inverse $\mathbf{S^{-1}}$ of $\mathbf{S}$ can be well approximated by $2\mathbf{I} - \mathbf{S}$, as discussed in previous sections. Then our goal is to build a compensator $\mathbf{C} = \widehat{\mathbf{S^{-1}}} = \mathbf{2I} - \hat{\mathbf{S}}$, where $\hat{\mathbf{S}}$ is again sought in the family of models as described above (i.e., $\hat{\mathbf{S}}$ is fit according to the above procedure and substituted into $\mathbf{2I} - \hat{\mathbf{S}}$). Compensator performance is measured in terms of output Error Vector Magnitude (EVM) and Adjacent-Channel-Leakage-Ratio (ACLR) [30], which are defined, for a given input w, as

$$\mathrm{EVM}(\mathbf{C}, w) = 20 \log_{10} \left(\frac{||w - \mathbf{S}\mathbf{C}w||_2}{||w||_2} \right),$$

$$\mathrm{ACLR}(\mathbf{C}, w) = 10 \log_{10} \left(\frac{\frac{1}{\xi} \int_{I_1} |W(\Omega)|^2 \Omega}{\frac{1}{1-\xi} \int_{I_2} |W(\Omega)|^2 \Omega} \right),$$

where W is the Fourier transform of w, intervals $I_1 = (-\frac{\pi\xi}{2}, \frac{\pi\xi}{2})$ and $I_2 = (-\pi, -\frac{\pi\xi}{2}) \cup (\frac{\pi\xi}{2}, \pi)$ with $\xi = B_w T_s$ where B_w is the input signal bandwidth.

In order to evaluate performance of the proposed model we have performed various simulations, varying different model and simulation parameters, as described in corresponding subsections. We compare approximation capability of the proposed model, as well as performance of a DPD based on it, with those obtained by employing pure Volterra series models, where the corresponding model parameter values are given in Table I.

1) Effects of changing parameters δ and ρ: For the cubic passband nonlinearity model (31) we fix delays to $\tau = \begin{bmatrix} 0.2T_s & 0.3T_s & 0.4T_s \end{bmatrix}$, where $f_{dac} = 1/T_s$ is the DAC rate, and vary parameter δ to simulate various degrees of distortion. In the case of Cann's model defined in (33), we take arbitrary, but fixed, parameter values $\tau = 0.5T_s$, $L = 1$, $s = 8$ and $g = 1$. The level of passband distortion introduced by $\mathbf{F}$ is controlled by varying parameter ρ. Simulations are run for two values of sampling/DAC frequency $f_{dac1} = 20$MHz and $f_{dac1} = 25$MHz. This is equivalent to setting parameter ξ to $\xi = 1$ and $\xi = 0.8$, respectively.

In case of $f_{dac1} = 20$MHz (i.e. when information bearing signal occupies the whole observation band), approximation results for different models are shown in Figs. 12 and 13. It can be seen that our model significantly outperforms Volterra models, and results in approximation error of 0.1% or better in all cases. This result was to be expected, since model shown in Fig. 8 can approximate arbitrarily close the original system $\mathbf{S}$. This is not the case with Volterra series models, due to inherently long (or more precisely infinite) memory introduced

by the LTI part of $\mathbf{S}$. Even if we use a non-causal Volterra series model (i.e., $m_f \neq 0$), which is expected to better capture the true dynamics of $\mathbf{S}$, we are still unable to get good fitting of $\mathbf{S}$ (though NMSE decreases with the increase of m_f). We get similar performance results for the proposed model in the case of $f_{dac2} = 25$MHz (though we do not report those to avoid repetition), with slightly better performance of Volterra based models than in the case $\xi = 1$. This is in accordance with presented theoretical results and discussions given in sections IV and V: as ratio ξ decreases, memory of $\mathbf{L_0}$ decreases, and Volterra models approach performance of the ideal model proposed in this paper.

Comparison of different DPD structures in terms of output EVM, for the case of $f_{dac1} = 20$MHz, is shown in Figs. 14 and 15. As a baseline, we also plot output EVM for the case when no compensation is used. As can be seen from Figs. 14 and 15, the proposed DPD model outperforms other compensators for a wide range of parameter values. In the case of a cubic passband nonlinearity, as parameter δ increases, the DPD performance decreases and approaches that of Volterra series based models. This is caused by the relatively simple structure of the compensator: as δ increases, the inverse $(\mathbf{I}+\mathbf{\Delta})^{-1}$ is not approximated close enough by the linear term $\mathbf{I}-\mathbf{\Delta}$. Hence, simple compensator $\mathbf{C} = \mathbf{2I} - \hat{\mathbf{S}}$ is not capable of approximating well the inverse $\mathbf{S^{-1}}$ (even though $\hat{\mathbf{S}}$ still approximates $\mathbf{S}$ very well). When $\xi = 1$, the observed (i.e., linearized) bandwidth of the PA output signal is equal to the baseband signal bandwidth, and therefore ACLR is not a meaningful metric of performance since there is no adjacent channel. Results in terms of NMSE and EVM for the case of $\xi = 0.8$, are similar to those for $\xi = 1$, and are omitted. Again, a slight increase in modeling/linearization capability of Volterra models is noticeable when going from $\xi = 1$ to $\xi = 0.8$. ACLR results for both cases of passband nonlinearity are shown in Figs. 16 and 17. Comparison of output signal power spectral density for various DPD models is shown in Fig. 18. As can be seen, about 10dB in ACLR improvement is achieved with the proposed DPD, while Volterra based DPDs were unable to clear much distortion. Again, this is expected since simulated Volterra models do not have enough memory capability to approximate well frequency response of $\mathbf{L}$ close to the points $\Omega = \pm\pi$ of discontinuity.

In the next two subsections, due to space constraints and without loss of generality, we report only results for cubic nonlinearity. Results for the other type of nonlinearity follow the same trends and do not bring any new information and can therefore be omitted.

2) Effects of changing parameter ξ: Now we fix model parameters to $\delta = 0.1$ and $\tau = \begin{bmatrix}0.2T_s & 0.3T_s & 0.4T_s\end{bmatrix}$, and vary ratio ξ from 0.5 to 1. The results in terms of EVM and ACLR are shown in Figs. 19 and 20. As is expected, for small values of ξ, all models perform similarly, since the memory of $\mathbf{L_0}$ is not large. As ξ increases, our model achieves the best performance since it is the only one capable of approximating long memory effects.

3) Effects of increasing memory of $\mathbf{F}$: In this case, we fix parameters $\delta = 0.1$ and $\xi = 1$, and vary memory of the passband nonlinearity $\mathbf{F}$ (and accordingly memory of our

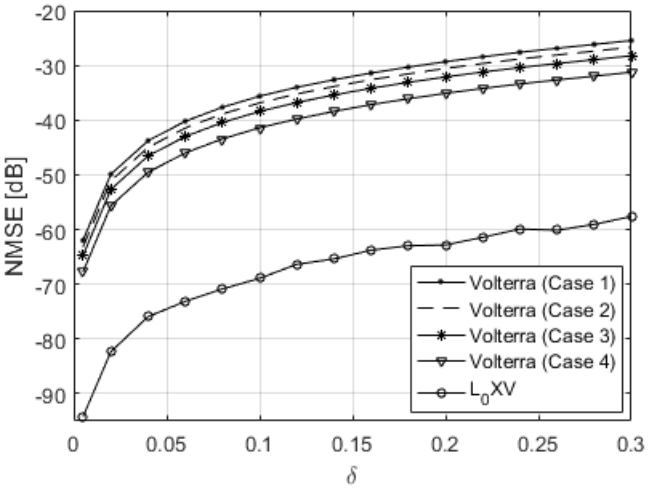


Fig. 12. NMSE of approximation for different models in the case of cubic passband nonlinearity.

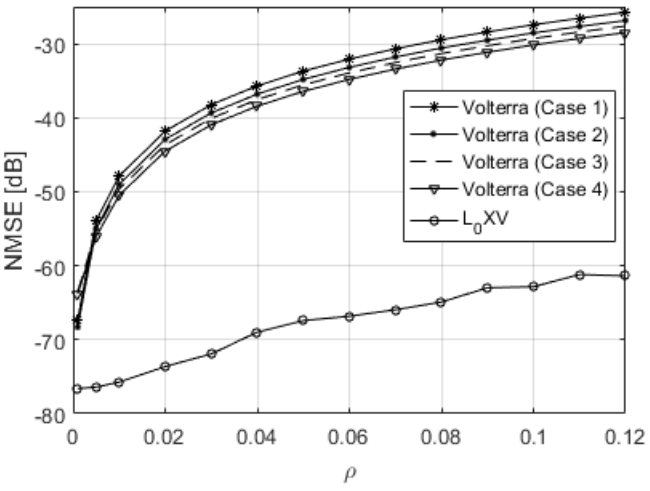


Fig. 13. NMSE of approximation for different models in the case of modified Cann's nonlinearity model.

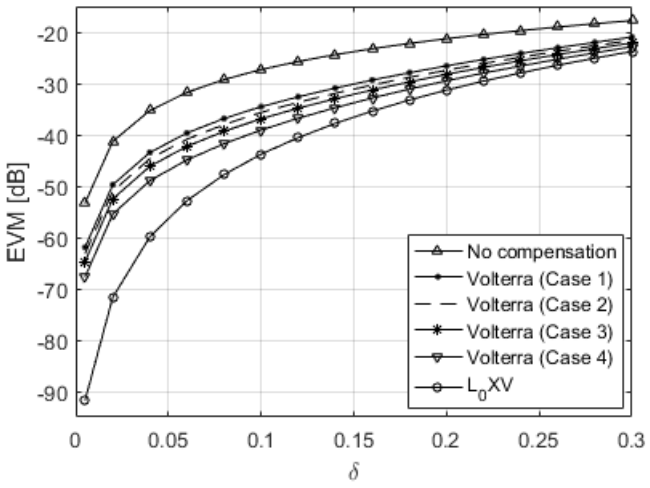


Fig. 14. Output EVM, for different DPD structures, as a function of parameter δ (in the case of cubic passband nonlinearity).

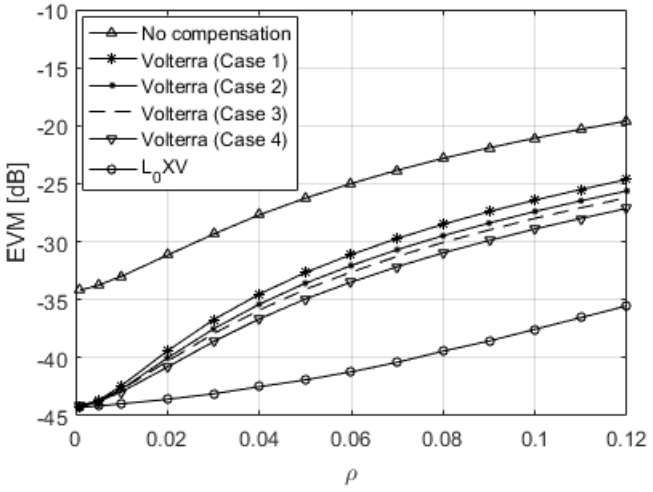


Fig. 15. Output EVM, for different DPD structures, as a function of parameter ρ (in the case of modified Cann's nonlinearity model).

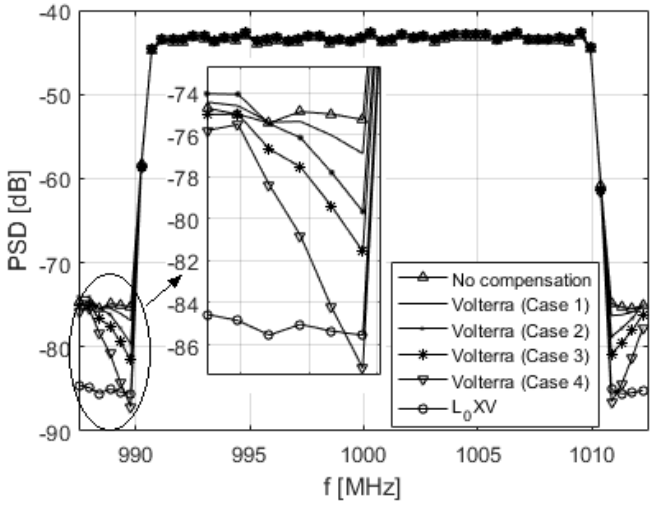


Fig. 18. PSD of the PA output, for different DPD structures (in the case of cubic passband nonlinearity).

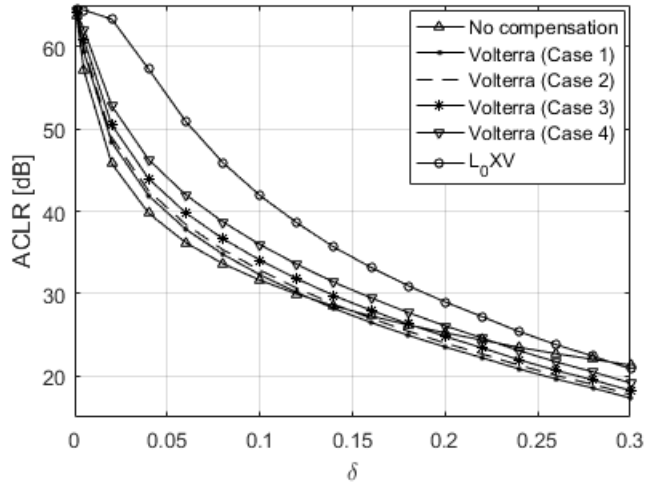


Fig. 16. Output ACLR, for different DPD structures, as a function of parameter δ (in the case of cubic passband nonlinearity).

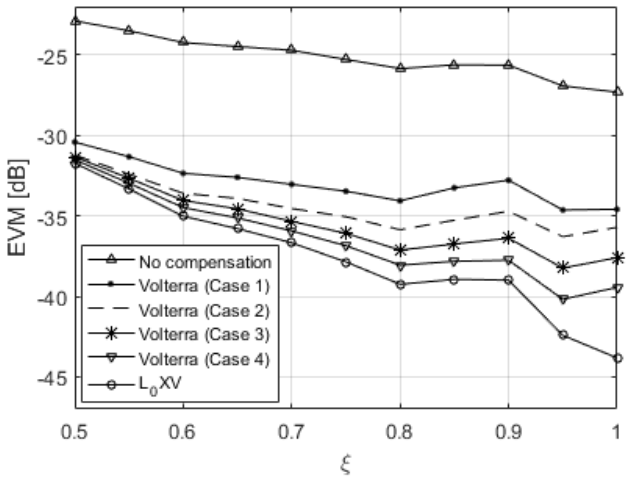


Fig. 19. Output EVM, for different DPD structures, as a function of the ratio $\xi = B_w/f_{dac}$ (in the case of cubic passband nonlinearity).

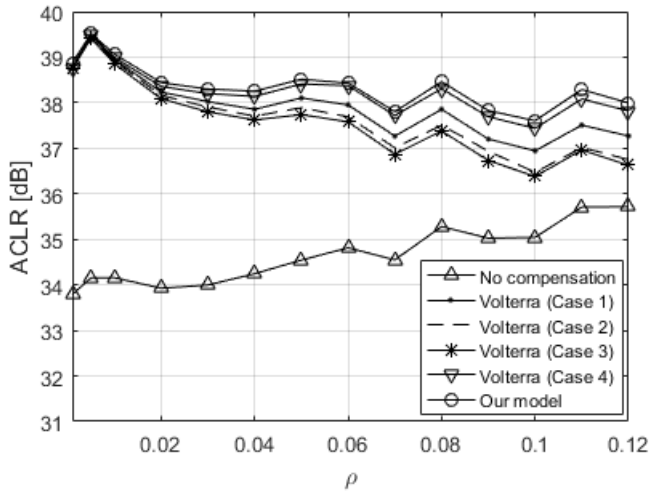


Fig. 17. Output ACLR, for different DPD structures, as a function of parameter ρ (in the case of modified Cann's nonlinearity model).

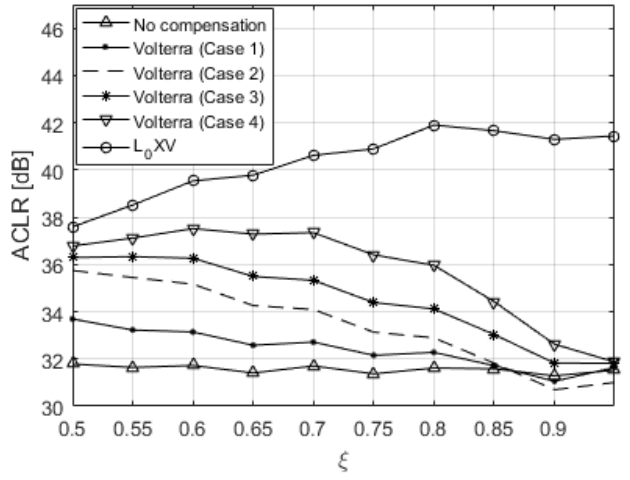


Fig. 20. Output ACLR, for different DPD structures, as a function of the ratio $\xi = B_w/f_{dac}$ (in the case of cubic passband nonlinearity).

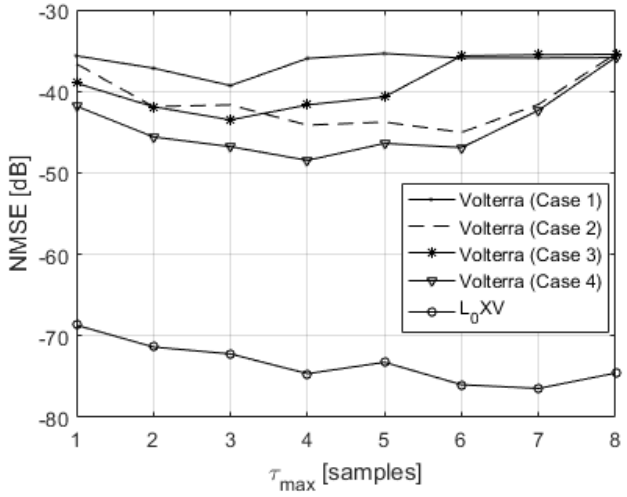


Fig. 21. Output NMSE, for different DPD structures, as a function of the maximal delay τ_{max} (in the case of cubic passband nonlinearity).

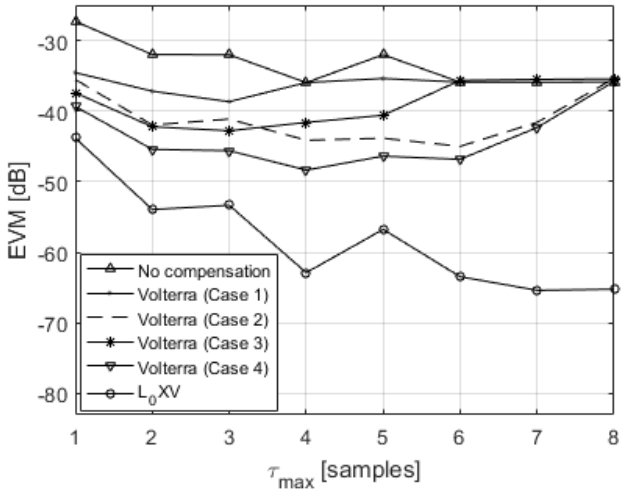


Fig. 22. Output EVM, for different DPD structures, as a function of the maximal delay τ_{max} (in the case of cubic passband nonlinearity).

model). Maximal delay $\tau_{max} = \max_{i\in\{1,2,3\}} \tau_i$ is varied from 1 sample to 7 samples, i.e. from $\tau_{max} \in (0, T_s]$ to $\tau_{max} \in (6T_s, 7T_s]$. The results in terms of NMSE and EVM are shown in Figs. 21 and 22 (there is no reported ACLR since $\xi = 1$). Results are again as expected, i.e., for large values of τ_{max} Volterra models struggle with approximating well the nonlinearity $\mathbf{F}$ due to their limited memory capability. Our model is again capable of successfully modeling and linearizing the PA output for all values of parameter τ_{max}.

4) Model Complexity: Advantage of the proposed compensator structure is not only in better compensation performance, but also in that it achieves this performance in a more efficient way. That is, significantly lower number of Volterra monomials (basis elements) is needed in order to represent the nonlinear part of the compensator. Table II shows a comparison in the number of basis elements for different compensator structures (parameter values for the corresponding passband nonlinearities were fixed to $\delta = 0.04$ and $\rho = 0.06$). For each passband nonlinearity case, numbers in the first row

TABLE II
COMPLEXITY OF DIFFERENT COMPENSATOR MODELS IN TERMS OF THE NUMBER OF COEFFICIENTS THAT ARE NEEDED FOR HARDWARE IMPLEMENTATION OF THE NONLINEAR PART.

Cubic nonlinearity model					
Model	L_0XV	Volt. 1	Volt. 2	Volt. 3	Volt. 4
# of basis elements	70	70	6006	6006	2660
# of significant basis elements	42	14	2058	1190	48
Modified Cann's model					
Model	L_0XV	Volt. 1	Volt. 2	Volt. 3	Volt. 4
# of basis elements	660	660	3432	6006	6006
# of significant basis elements	253	528	2914	1881	449

in Table II represent the total number of basis elements of the corresponding model. The second row shows the actual number of basis elements used to build the compensator, that is, least squares optimization yields many nonzero coefficients, but only a subset of those are considered significant and thus used in an actual compensator implementation. In this case coefficient is considered significant if its value falls above a certain threshold t_0, where t_0 is chosen such that increase in EVM after zeroing non-significant coefficients is not larger than 1% of the best achievable EVM (i.e., when all basis elements are used). From Table II we can see that, even for the best (in terms of EVM) Volterra structure, more basis elements are needed in order to implement the compensator (and its performance is still below the one achieved by our model). It should be noted though that the low complexity advantage of the proposed model gets lost as memory of passband nonlinearity increases. This is due to an increase in the number of nonlinear basis elements in $\mathbf{V}$, which then approaches that of a plain Volterra model.

We should mention here that in [37] decomposition to a short memory nonlinear system and a long memory linear system, for a system architecture similar to the one proposed in our paper, but with a different modulator model, has been verified by simulations which combine Cadence Spectre Circuit Simulator and Matlab (PA is simulated in Spectre, a phase modulator is realized with verilog-A model in Spectre, and other processing blocks/subsystems are realized in Matlab). Similarly, in [38], a DPD based on the model proposed in our paper has been implemented on an FPGA and successfully tested for an outphasing transmitter at Q-band (45GHz) and for an RF PA with 1.97GHz carrier.

VII. CONCLUSION

A new explicit expression of the equivalent baseband model has been presented, under assumption that the passband nonlinearity can be described by a Volterra series model with fixed degree and memory depth. It has been shown that order of nonlinearity and, more importantly, memory of the underlying nonlinear system are preserved when passing from passband to baseband. This results in a novel equivalent baseband model which is a series interconnection of a fixed degree/short mem-

ory Volterra model, and a long memory LTI system. The result suggests a new, non-obvious, analytically motivated structure of digital precompensation of passband nonlinear distortions caused by power amplifiers. It also follows that the memory, and therefore complexity, of an approximate baseband model, and its corresponding digital predistorter, increases as the ratio of baseband signal bandwidth to the observed bandwidth increases. This suggests that proposed model should be best utilized under assumption of a full input signal bandwidth (that is, no oversampling of the baseband signal). Unlike conventional DPD implementations which have long memory specifications in the corresponding nonlinear subsystem, a DPD model proposed in this paper exploits the underlying system structure to model long memory requirements in terms of high order FIR filters, which are relatively simple for digital implementation. Presented theoretical analysis is supported by simulation results. Future work would include verifying proposed method for different modulator models (e.g., when zero-order hold DAC is cascaded with a low-pass filter).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. The authors are grateful to Dr. Yehuda Avniel for bringing researchers from vastly different backgrounds to work together on the tasks that led to the writing of this paper.

REFERENCES

[1] P. B. Kennington, *High linearity RF amplifier design*. Norwood, MA: Artech House, 2000.

[2] S. C. Cripps, *Advanced techniques in RF power amplifier design*. Norwood, MA: Artech House, 2002.

[3] J. Vuolevi, and T. Rahkonen, *Distortion in RF Power Amplifiers*. Norwood, MA: Artech House, 2003.

[4] J. C. Pedro and S. A. Mass, "A Comparative Overview of Microwave and Wireless Power-Amplifier Behavioral Modeling Approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, No. 4, pp. 1150-1163, April 2005.

[5] A. A. M. Saleh, and J. Salz, "Adaptive linearization of power amplifiers in digital radio systems," *Bell Syst. Tech. J.*, vol. 62, no. 4, pp. 1019-1033, April 1983.

[6] W. Bösch, and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *IEEE Trans. Microw. Theory Techn.*, vol. 37, pp. 1885-1890, December 1989.

[7] J. Kim, and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol 37, no. 23, pp. 1417-1418, November 2001.

[8] 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 Trans. Commun.*, vol. 52, no. 1, pp.159-165, January 2004.

[9] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York: Wiley, 2000.

[10] T. Liu, S. Boumaiza and F. M. Ghannouchi, "Augmented hammerstein predistorter for linearization of broad-band wireless transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 4, pp. 1340-1349, June 2006.

[11] A. Zhu, and T. Brazil, "Behavioral modeling of RF power amplifiers based on pruned Volterra series," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 12, pp. 563-565, December 2004.

[12] D. R. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852-3860, October 2006.

[13] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, No. 12, pp. 4323-4332., December 2006.

[14] 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 Trans. Microw. Theory Techn.*, vol. 56, No. 7, pp. 1524-1534., July 2008.

[15] 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 Trans. Microw. Theory Techn.*, vol. 58, No. 1, pp. 95-104, January 2010.

[16] O. Hammi and F. M. Ghannouchi, "Twin Nonlinear Two-Box Models for Power Amplifiers and Transmitters Exhibiting Memory Effects With Application to Digital Predistortion," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 8, pp. 530-532, Aug. 2009.

[17] M. Younes, O. Hammi, A. Kwan and F. M. Ghannouchi, "An Accurate Complexity-Reduced PLUME Model for Behavioral Modeling and Digital Predistortion of RF Power Amplifiers," *IEEE Trans. Indust. Electron.*, vol. 58, no. 4, pp. 1397-1405, April 2011.

[18] C. Yu, L. Guan, and A. Zhu, "Band-Limited Volterra Series-Based Digital Predistortion for Wideband RF Power Amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, No. 12, pp. 4198-4208, December 2012.

[19] L. Ding, F. Mujica and Z. Yang, "Digital predistortion using direct learning with reduced bandwidth feedback," *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, Seattle, WA, 2013, pp. 1-3.

[20] O. Hammi, A. Kwan, S. Bensmida, K. A. Morris and F. M. Ghannouchi, "A Digital Predistortion System With Extended Correction Bandwidth With Application to LTE-A Nonlinear Power Amplifiers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 12, pp. 3487-3495, Dec. 2014.

[21] Y. Liu, W. Pan, S. Shao and Y. Tang, "A General Digital Predistortion Architecture Using Constrained Feedback Bandwidth for Wideband Power Amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 5, pp. 1544-1555, May 2015.

[22] G. M. Raz, and B. D. Van Veen, "Baseband Volterra filters for implementing carrier based nonlinearities," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 103-114, January 1998.

[23] M. Morhac, "A fast algorithm of nonlinear Volterra filtering," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2353-2356, Oct 1991.

[24] P. A. J. Nuyts, B. Francois, W. Dehaene and P. Reynaert, "A CMOS Burst-Mode Transmitter With Watt-Level RF PA and Flexible Fully Digital Front-End," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 10, pp. 613-617, Oct. 2012.

[25] K. Hausmair, S. Chi, P. Singerl and C. Vogel, "Aliasing-Free Digital Pulse-Width Modulation for Burst-Mode RF Transmitters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 2, pp. 415-427, Feb. 2013.

[26] M. T. Pasha, M. F. U. Haque, J. Ahmad and T. Johansson, "A Modified All-Digital Polar PWM Transmitter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 2, pp. 758-768, Feb. 2018.

[27] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems*. reprint ed. Malabar, FL: Krieger, 2006.

[28] W. Frank, "Sampling requirements for Volterra system identification," *IEEE Signal Process. Lett.*, vol. 3, no. 9, pp. 266-268, September 1996.

[29] J.Tsimbinos, and K.V.Lever, "Input Nyquist sampling suffices to identify and compensate nonlinear systems," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2833-2837, Oct. 1998.

[30] J. G. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill, 2007.

[31] D. C. Cox, "Linear amplification with nonlinear components," *IEEE Trans. Commun.*, vol. 22, no. 12, pp. 1942-1945, Dec. 1974.

[32] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[33] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun.*, vol. 52, pp. 5-65, March 2005.

[34] Q. Shi, "OFDM in bandpass nonlinearity," *IEEE Trans. Consumer Electron.*, vol. 42, pp. 253-258, August 1996.

[35] A. N. D'Andrea, V. Lottici, and R. Reggiannini, "Nonlinear predistortion of OFDM signals over frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 49, no. 5, pp. 837-843, May 2001.

[36] A. J. Cann, "Nonlinearity Model With Variable Knee Sharpness," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 16, no. 6, pp. 874-877, November 1980.

[37] Y. Li, *Digital Assistance Design for Analog Systems: Digital Baseband for Outphasing Power Amplifiers*, PhD Thesis, Massachusetts Institute of Technology, June 2013.

[38] Z. Li, *Efficient Baseband Design and Implementation for High-Throughput Transmitters*, PhD Thesis, Massachusetts Institute of Technology, September 2015.