Doctoral Thesis

Advanced Signal Processing Techniques for the Modeling and Linearization of Wireless Communication Systems

Author: Juan Antonio Becerra González

Advisor: Luis Javier Reina Tosina

Co-advisor: Carlos Crespo Cadenas

Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, Spain 2018

Doctoral Thesis

Advanced Signal Processing Techniques for the Modeling and Linearization of Wireless Communication Systems

Author:

Juan Antonio Becerra González

Advisor:

Luis Javier Reina Tosina

Profesor Titular

Co-advisor:

Carlos Crespo Cadenas

Profesor Titular

Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, Spain 2018

Doctoral Thesis: Advanced Signal Processing Techniques for the Modeling and Linearization of Wireless Communication Systems

Author: Juan Antonio Becerra González
Advisor: Luis Javier Reina Tosina
Co-advisor: Carlos Crespo Cadenas

El tribunal nombrado para juzgar la Tesis arriba indicada, compuesto por los siguientes doctores:

Presidente:

Vocales:

Secretario:

acuerdan otorgarle la calificación de:

El Secretario del Tribunal

Fecha:

A mis padres, Juan Antonio y Lucía.
To my parents, Juan Antonio and Lucía.

Acknowledgements

I would like to express my sincerest gratitude to my research group, María José, Javier, and Carlos. Thank you very much for everything you taught me and the patience you had with me. I appreciate all the guidance during these years and the unconditional support you have shown to every kind of idea. To the young grasshopper of the family, Abraham, for bringing strength, energy, and motivation in this final stage.

To my colleagues in the Department of Signal Theory and Communications, for making me feel at home and for everything I learned from them. Special thanks to Irene Fondón, Uxi, Iván, Eva, Irene Santos and Javier Olías.

To Michel, for his patience while teaching the basic concepts on which this thesis is based. To Deepa, for proving that connection with other people is above language barriers.

To my colleagues at the *Université de Limoges* and the University of Delaware, for the new dimensions they gave to this project.

To Javier Payán and Juanjo, for the design of this template that has been so simple to use. To Esperanza, for the tremendous logistical support at the time of doing paperwork with the University. To María and José, for all the cheering while preparing the gasoline that accelerates the research.

To Aurora, a friend more than a sister.

Finally, I would like to close the acknowledgments with the same sentence that I started my final undergraduate project nine years ago: "*To my parents, for all that I am, and for all that I do not*".

Agradecimientos (Acknowledgements)

Me gustaría comenzar mis agradecimientos expresando mi más sincera gratitud a mis compañeros de investigación, María José, Javier y Carlos. Mil gracias por todo lo que me habéis enseñado y la paciencia que habéis tenido conmigo. Os agradezco enormemente la guía durante todos estos años y el apoyo incondicional que habéis mostrado a todo tipo de ideas. Al pequeño de la familia, Abraham, por traer fuerza, energía y motivación en esta etapa final.

A mis compañeros del Departamento de Teoría la Señal y Comunicaciones, por hacerme sentir como en casa y por todo lo que he aprendido de ellos. Agradecimiento especial a Irene Fondón, Uxi, Iván, Eva, Irene Santos y Javier Olías.

A Michel, por su paciencia al enseñarme los conceptos básicos en los que se fundamenta esta tesis. A Deepa, por demostrar que la conexión con otras personas está por encima del lenguaje.

A mis compañeros de estancia en la *Université de Limoges* y la *University of Delaware*, por las nuevas dimensiones que han aportado a este proyecto.

A Javier Payán y Juanjo, por el diseño de esta plantilla que tan sencilla ha sido de usar. A Esperanza, por el infinito soporte logístico a la hora de hacer trámites con la Universidad. A María y Jose, por esos ánimos mientras preparan la gasolina que acelera la investigación.

A Aurora, más amiga que hermana.

Por último, me gustaría cerrar los agradecimientos con la misma frase que empecé el Proyecto Fin de Carrera hace ya nueve años: "*A mis padres, por todo lo que soy, y por todo lo que no*".

Abstract

New standards of wireless digital communications are pushing the design of power amplifiers towards challenging constraints in terms of linearity and efficiency. While these new systems demand active devices to work near saturation seeking power efficiency, their inherent nonlinearity may drive the overall system to inadequate performance in out-of-band emissions and in-band distortion. The need of digital compensation techniques and the evolution in the design of new digital signal processing architectures make digital predistortion (DPD) a convenient linearization approach.

Digital predistorters usually rely on behavioral models such as memory polynomials (MPs), generalized memory polynomials (GMPs), dynamic deviation reduction-based (DDR), etc. Volterra system identification suffers the "curse of dimensionality", since the complexity tends to grow exponentially in the number of coefficients as the memory and nonlinear order become larger.

This Thesis is focused primarily on contributing to the modeling and linearization of wireless communication systems. The main covered topics are the Volterra-Parafac and the General Volterra Model for Complex Systems which address the structure of the DPD and the structured compressed-sensing Volterra series and a method for linearization in a range of average working operation levels, which focuses in how the coefficients of the models have to be retrieved.

Resumen (Abstract)

Los nuevos estándares de comunicaciones digitales inalámbricas están impulsando el diseño de amplificadores de potencia con unas condiciones límites en términos de linealidad y eficiencia. Si bien estos nuevos sistemas exigen que los dispositivos activos trabajen cerca de la zona de saturación en busca de la eficiencia energética, la no linealidad inherente puede producir que el sistema muestre prestaciones inadecuadas en emisiones fuera de banda y distorsión en banda. La necesidad de técnicas digitales de compensación y la evolución en el diseño de nuevas arquitecturas de procesamiento de señales digitales posicionan a la predistorsión digital (DPD) como un enfoque práctico.

Los predistorsionadores digitales se suelen basar en modelos de comportamiento como el memory polynomial (MP), el generalized memory polynomial (GMP) y el dynamic deviation reduction-based (DDR), etc. Los modelos de Volterra sufren la llamada "maldición de la dimensionalidad", ya que su complejidad tiende a crecer de forma exponencial a medida que el orden y la profundidad de memoria crecen.

Esta tesis se centra principalmente en contribuir a la rama de conocimiento que enmarca el modelado y linealización de sistemas de comunicación inalámbrica. Los principales temas tratados son el modelo Volterra-Parafac y el modelo general de Volterra para sistemas complejos, los cuales tratan la estructura del DPD y las series de Volterra estructuradas con compressed-sensing y un método para la linealización en un rango de potencias de operación, que se centran en cómo los coeficientes de los modelos deben ser obtenidos.

Short Contents

Contents

List of Acronyms

List of Acronyms

ACPR Adjacent Channel Power Ratio

ACEPR Adjacent Channel Error Power Ratio

BIC Bayesian Information Criterion

DDR Dynamic Deviation Reduction

DPD Digital Predistortion

DUT Device Under Test

EVM Error Vector Magnitude

FPGA Field-Programmable Gate Arrays

FV Full Volterra

GMP Generalized Memory Polynomial

LASSO Least Absolute Shrinkage and Selection Operator

LMS Least Mean Squares

LS Least Squares

LTE Long Term Evolution

MP Memory Polynomial

NMSE Normalized Mean Square Error

OFDM Orthogonal Frequency Division Multiplexing

PA Power Amplifier

PAPR Peak-to-Average Power Ratio

PRSS Penalized Residual Sum of Squares

QAM Quadrature Amplitude Modulation

RSS Residual Sum of Squares

VSA Vector Signal Analyzer

VSG Vector Signal Generator

Notation

t	Continuous-time variable
k	Discrete-time variable
n	Iteration index
$\mathbf{x}$	Input signal in vector notation
$\mathbf{X}$	Measurement matrix
$\mathbf{y}$	Output signal in vector notation
$\mathbf{h}$	Volterra kernel vector
$\hat{\mathbf{y}}$	Estimated output signal in vector notation
$x(t)$	Input signal in continuous time t
$x[k]$	Input signal in discrete time k
h_i	Volterra kernel of i-th order
f	Continuous-frequency variable
$X(f)$	Signal in frequency domain
ε	Error signal
$\mathbf{A}^H$	Hermitian of matrix $\mathbf{A}$
$\mathbf{A}^+$	Moore-Penrose pseudoinverse of matrix $\mathbf{A}$

1 Overview

1.1 Introduction

Reducing the consumption of electrical energy is nowadays an issue of great worldwide concern, not just because of the economic savings that this implies, but also because of the reduction in pollution associated with the production of electricity. It is estimated that telecommunications represent about an 8% of the total global energy consumption [1]. In addition, the trends of annual increase in electricity use by communications networks indicate that this demand will continue to grow at a rapid rate [1]. This effect is due to the fact that operators are continuously installing new stations to offer novel wireless broadband services. In addition to the increase in energy prices, the growth in telecommunication emissions is expected to reach 350 metric tons of carbon dioxide equivalent in 2020 [2]. This scenario of environmental and economic impact has gathered a rising interest that translates into a movement towards the so-called green communications, in which special attention is paid to the improvement of the energy efficiency in communication devices.

Mobile base stations use a significant amount of energy to transmit information. In particular, the device that consumes most of the energy is the Power Amplifier (PA). PAs have a particular feature in their operation: their efficiency is greater at high powers, where the distortion appears, which is an undesirable effect. Therefore, if the device needs to be efficient, it is necessary to operate at a high power. On the other hand, new communication standards within the fourth and fifth generation (4G and 5G), define signals with a high Peak-to-Average Power Ratio (PAPR). These excursion points of power together with the distortion of the PAs at a high power make it necessary to decrease the operating point, with the corresponding decrease in efficiency of the communications device. This effect captured the interest of the scientific community, a scenario in which Digital Predistortion (DPD) was presented as a very promising solution. This way, all the desirable requirements for a communications system are met: operation at high powers for ensuring energy efficiency and a high fidelity device, which allows the use of the new transmission systems.

The evolvement of this scenario is characterized by the requirement of an efficient spectrum to deliver those data rates in such a scarce resource. Linearity in the blocks of the

transmitter of communication systems becomes a critical point in their design. The PA is one of the nonlinear components that causes the final transmission data to be affected and also not to comply with the linearity requirements of the standards. One possible solution for this is the use of the PA in a low power region, where the device becomes more linear. This is inefficient in terms of power, since the PA becomes more power-efficient when it is operating in its saturation region. Nonlinear effects cause distortion that generates spectral regrowth, out-of-band emissions, and in-band emissions. Hence, linearization techniques should be developed to enable the simultaneous achievement of a high power-efficiency and an efficient usage of the spectrum.

One of the most popular linearizing technique is DPD. These systems basically consist of a digital transformation of the PA input signal, in which the cascade effect of the predistorter and the nonlinearity of the amplifier results in a linear behavior. DPDs usually rely on Volterra series, which can be considered like the extension of a Taylor polynomial to a complex time-varying input, to represent the nonlinearity. Volterra series are usually characterized by their order and memory depth. The memory depth models the dependence of the output with the input in not just the same instant but also in the past. The order enables the system to represent a nonlinear behavior. Volterra series suffer the *curse of dimensionality*, since their number of parameters rapidly grows with the order and memory depth.

1.2 Aims and Scope

Considering the previous work made in this research field, the main aim of this Thesis is contributing to the study and development of new signal processing techniques applied to the modeling and linearization of wireless communication systems. This general objective can be split in the following primary aims:

- Development of novel models for the PA behavior.
- Development of new techniques for obtaining the model coefficients.
- Experimental characterization of the nonlinear mitigation performance of these algorithms.

The scope of the nonlinear mitigation is wide, therefore some boundaries have been adopted in order to make its amplitude reasonable.

Nonlinear phenomena can be studied in an enormous diversity of systems. The types of systems herein studied are the power amplifier and the modulator, both from the perspective of its nonlinear behavior. The input signals of interest are digitally-modulated communication signals, mainly focusing on those that support 4G and 5G. The kind of models that have been worked are Volterra series and, when it comes to its prunning, the signal processing techniques that have been applied are *greedy algorithms* and those based on information theory.

1.3 Thesis Outline

This Thesis is developed through seven chapters, starting from the current introduction and motivation of this thesis. First, chapter 2 establishes the theoretical and experimental frameworks. The following chapters are divided into two conceptual groups. Chapters 3 and 4 belong to the discussion of which structure performs a better representation of the DPD. In these, the models of the Volterra-Parafac family are first discussed and next the complex Volterra series (CVS) is presented. Following the modeling part of this Thesis, the next chapters are focused on retrieving the important coefficients of the model. In chapter 5, a heuristic unsupervised method that takes into account the structural information of the model for retrieving the coefficients is shown, followed by a statistical method that also performs the sparse regression in chapter 6. Chapter 7 includes the experimental design and results for the theoretical part of the Thesis. Finally, in chapter 8 conclusions are drawn and future work in this field is discussed.

2 Materials and Methods

Empires die, but Euclid's theorems keep their youth forever.

VITO VOLTERRA

This chapter describes the theory on which the rest of this Thesis is based. First, in section 2.1, the Volterra series theory is introduced followed by a description of the performance metrics in section 2.2. The extraction of model parameters and DPD identification is reviewed in 2.3 and the experimental testbenches used for the experimental part are described in section 2.4.

2.1 Behavioral Modeling Background

Nonlinear distortion can be defined as the signal components, other than the original signal, which are produced by nonlinear transformation of an input signal [3]. The Volterra series theory, which models nonlinear distortion, was developed by Vito Volterra in the late 19th century [4]. This theory was first applied by [5], where the response of a nonlinear device to noise was analyzed. The continuous-time Volterra filter can be interpreted as the natural expansion of a Taylor series in which the ability of capturing memory effects has been enabled. The continuous time output response $y(t)$ to input $x(t)$ depends linearly on the Volterra coefficients as

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\tau_1)x(t-\tau_1)d\tau_1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_n(\tau_1,\tau_2,\ldots,\tau_n)x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)d\tau_1 d\tau_2\ldots d\tau_n + \ldots, \tag{2.1}$$

where h_0 is a constant and $h_k(\tau_1,\tau_2,\ldots,\tau_k)$ is the k-th order Volterra kernel. The causal discrete-time Volterra filter is equivalently described by

$$\begin{aligned} y[k] = h_0 + \sum_{k_1=0}^{\infty} h_1(k_1)x[k-k_1] + \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} h_2(k_1,k_2)x[k-k_1]x[k-k_2] + \\ \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_p=0}^{\infty} h_p(k_1,k_2,\ldots,k_p)x[k-k_1]x[k-k_2]\ldots x[k-k_p] + \ldots \end{aligned} \tag{2.2}$$

Volterra series expansions have the following properties [6]:

Linearity with respect to the kernel coefficients: Although Volterra series represent a nonlinear behaviour between the input and the output of the system, the kernels are linear with respect to the output of the system.

Symmetry of the kernels: The permutation of the indices in a Volterra series kernels results in the same combination of input samples.

Multidimensional convolution property: A p-th order Volterra kernel may be written as a p-dimensional convolution.

Stability property: A Volterra kernel of order p is bounded-input bounded-output (BIBO) stable if

$$\sum_{k_1=0}^{\infty}\cdots\sum_{k_p=0}^{\infty} |h_p(k_1,\ldots,k_p)| < \infty. \tag{2.3}$$

High kernel complexity: The number of parameters in Volterra series increases rapidly with the polynomial order and memory. This is generally known as the *curse of dimensionality*.

2.1.1 Volterra Series Models

PAs are passband devices. By assuming that the bandwidth of the signal is small with respect to the center frequency, instead of using passband samples (at high sample rates), baseband samples (at lower sample rates) can be used. The relation between the passband signal $\tilde{x}(t)$ and its baseband formulation $x(t)$ follows

$$\tilde{x}(t) = \frac{e^{j2\pi f_0 t}x(t) + e^{-j2\pi f_0 t}x^*(t)}{2}, \tag{2.4}$$

where $x^*(t)$ is the complex conjugate of $x(t)$. This complex baseband representation is used in the rest of this Thesis. The discrete baseband representation of the general Volterra

series has the structure of

$$y[k] = \sum_{p_1=0}^{P} \sum_{q_1=0}^{Q} \sum_{q_2=q_1}^{Q} \cdots \sum_{q_p=q_{p-1}}^{Q} \sum_{q_{p+1}=0}^{Q} \sum_{q_{p+2}=q_{p+1}}^{Q} \cdots \sum_{q_{2p-1}=q_{2p-2}}^{Q} h_{2p-1}(q_1,q_2,\ldots,q_{2p-1}) \prod_{j_1=1}^{p} x[k-q_{j_1}] \prod_{j_2=p+1}^{2p-1} x^*[k-q_{j_2}]. \tag{2.5}$$

In the literature, this model is referred to as the Full Volterra (FV) model. In a practical application, the number of coefficients of this most-general model increases rapidly with the order $2P-1$ and the memory depth Q [7]. Because of this, pruned versions have been extensively researched and developed.

2.2 Performance of Behavioral Models

The aim we pursue when a system is characterized by a behavioral model is to obtain an transfer function that is able to represent the output precisely. For that purpose, several performance metrics are defined and widely used [8]. In this section, these key indicators are highlighted.

2.2.1 Metrics of Performance

The most widely used performance indicator is the Normalized Mean Square Error (NMSE). The NMSE represents the ratio between the error power and the reference signal power, and it measures the modeling accuracy capability of the model. It is defined as

$$\text{NMSE} = \frac{\|\mathbf{y}_{\text{model}} - \mathbf{y}_{\text{ref}}\|_2^2}{\|\mathbf{y}_{\text{ref}}\|_2^2}, \tag{2.6}$$

where $\mathbf{y}_{\text{model}}$ is the output of the model (stacking the samples in vector form) for an input signal, $\mathbf{y}_{\text{ref}}$ is the reference signal vector and

$$\|\mathbf{V}\|_\ell = \left(\sum_{i=0}^{N_v} (v_i)^\ell \right)^{\frac{1}{\ell}} \tag{2.7}$$

is the ℓ norm of the vector $\mathbf{V} = [v_1, v_2, ..., v_{N_v}]$.

The Adjacent Channel Power Ratio (ACPR) is also extensively applied to get the model performance. The ACPR takes into account the ratio between the power in the main operation band to the power in the adjacent bands in the spectrum. Depending on which band we are taking as a reference, the ACPR is usually tagged with a plus or a minus symbol and a number. The $+$ symbol indicates the upper channel and the $-$ symbol stands

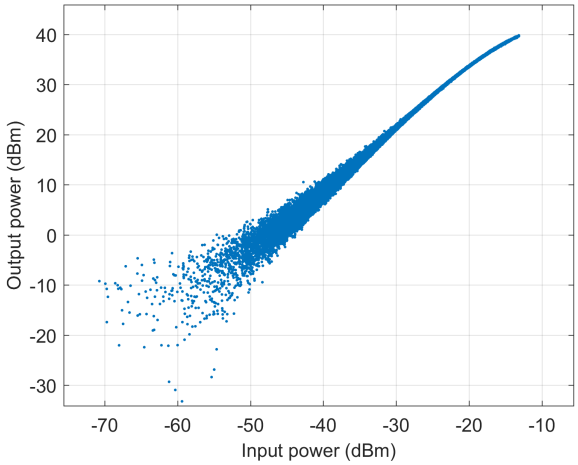

Figure 2.1 Example of an AM/AM characteristic.

for the lower channel. The number indicates the channel (1 for the first, 2 for the second and so on). The ACPR is defined as

$$\mathrm{ACPR}_{(+/-)K} = \frac{\int_{\mathrm{channel}_0} |Y(f)|^2 df}{\int_{\mathrm{channel}_{(+/-)K}} |Y(f)|^2 df}, \tag{2.8}$$

where $Y(f)$ represents the spectrum of the signal y. The ACPR is calculated over the error signal when it is taken as a performance indicator of a modeling and it is calculated over the linearized signal when the evaluation of the linearization capabilities is performed.

The AM/AM and the AM/PM characteristics are the graphical representations of the instantaneous power of the output and the instantaneous phase of the output versus the instantaneous power of the input. An example of these representations are shown in figures 2.1 and 2.3. The AM/AM characteristic is sometimes represented using the instantaneous gain instead of the instantaneous input power (see figure 2.2). The AM/AM plot gives relevant information about the behaviour of the system, as

Linear gain The linear gain of the system is the slope of the AM/AM.

Compressed gain The compressed gain is the gain at the highest input power.

Memory effects Memory effects are evidenced by the width of the cloud at medium-low powers.

Nonlinear effects Nonlinear effects are shown in this representation as the average evolution of the output power with the input power.

The Error Vector Magnitude (EVM) is a performance indicator of in-band distortion for digitally modulated signals. It is calculated over the received constellation with respect

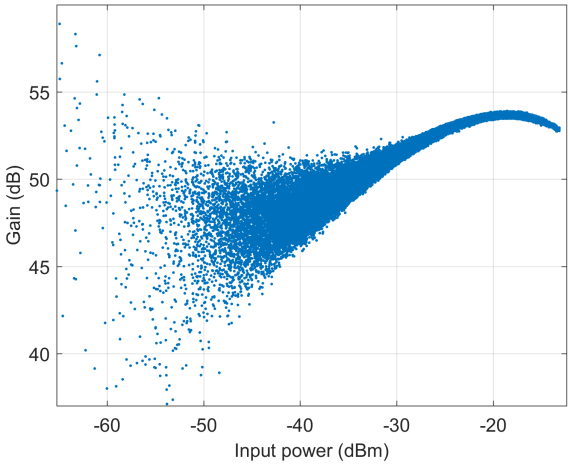

Figure 2.2 Example of an AM/Gain characteristic.

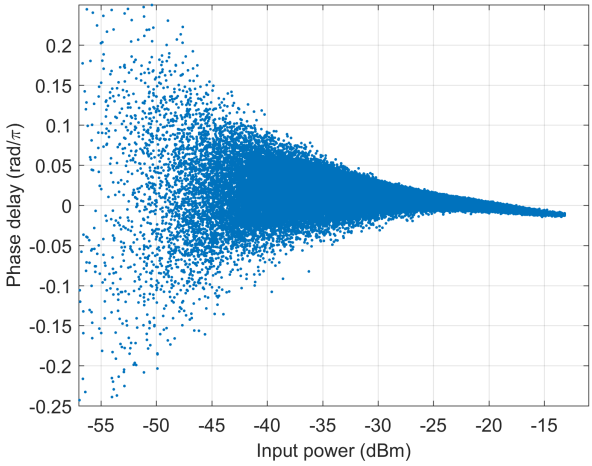

Figure 2.3 Example of an AM/PM characteristic.

to the ideal constellation at the transmitter. It measures how far are the points from the ideal locations, giving a measure of the error in the constellation space.

$$\text{EVM} = \frac{P_{\text{error}}}{P_{\text{reference}}}, \tag{2.9}$$

where P_{error} is the amplitude of the error vector in the constellation space and $P_{\text{reference}}$ is the coordinate associated to the transmitted symbol.

2.3 Model Parameters and DPD Identification

2.3.1 Digital Predistortion

The basic principle of digital predistortion is to obtain a block that will be placed in front of the nonlinear PA so that the combined action of these two systems produces a scaled version of the input [9].

The input-output characteristic of a typical PA is shown in figure 2.4. In a PA, when the signal level is low, the response is close to linear. At high signal levels, the PA produces compression and generates distortion. When the PA is not able to deliver more power, it is in saturation. The input-output response of a DPD is also presented. While the PA shows a compressive characteristic, the DPD has an expansive characteristic that is the symmetrical complementary of the PA with respect to a linear gain. The combination of these two characteristics results in a linear relationship. The inverse characteristic of the DPD will only work up to the saturation power of the PA, therefore we need to limit the input signal level to prevent the PA becoming saturated.

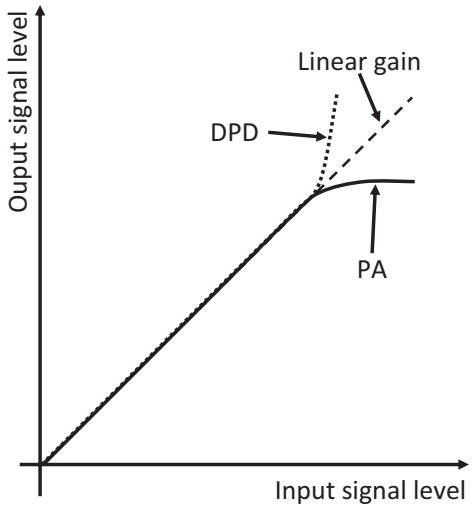

Figure 2.4 Input-output characteristic of a saturated PA and a DPD.

There is a need of upsampling in the input of the DPD. Since it generates harmonics in anti-phase to compensate the nonlinear behavior of the PA, the Nyquist sample rate is no longer valid. The sampling frequency needs to accommodate the new generated spectral bands. The oversample rates will depend on the bandwidth of the input signal and the DPD capability we are expecting. The typical oversampling rate is over five times the signal bandwidth.

Digital predistortion is commonly considered as a traditional closed-loop digital adaptive problem, in which the controller (the DPD) is tuned by some algorithm.

In an indirect learning scheme [10], we aim at finding a postdistorter, that is, a block that creates the linear input to the system at its output when it has in its input a scaled version of the PA output. The scale is usually the average power gain of the PA, i.e., the DPD has an expected gain of about the unit. This is required for stability of the DPD loop. The estimation of the DPD coefficients may be performed at once through Least Squares (LS) or with an adaptive technique. These algorithms are introduced in the next sections.

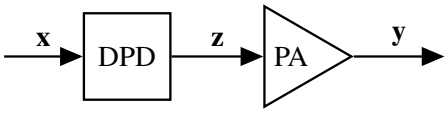

Figure 2.5 Schematic of a predistorter and a power amplifier.

2.3.2 Least Squares (LS)

The LS regression, which appeared formally formulated at the beginning of the 19th century [11], is one of the most common techniques for regressing a linear set of equations, since the error is penalized quadratically and it has only one solution. In the case of Volterra series, although the relation between the input $\mathbf{x} = [x(n), x(n-1), ..., x(n-M-1)]$, where M is the maximum number of samples taken into account and the output $\mathbf{y} = [y(n), y(n-1), ..., y(n-M-1)]$ is nonlinear by its definition, there exists a linear regression between the vector of coefficients $\mathbf{h}$ and the output $\mathbf{y}$, where the columns of the measurement matrix $\mathbf{X}$ make the transformation from one space to another. The measurement matrix structure is model-dependent. Its columns, $\mathbf{X}_{\{i\}}$, are the regressors of the model and usually take the form of some kind of function of the input complex envelope $x(k)$. For example, the Memory Polynomial (MP) model [12] generates pth-order regressors lagged q samples with the form $x(k-q)|x(k-q)|^{P-1}$. Its Pth-order and maximum memory Q measurement matrix has the following structure

$$\mathbf{X}_{\mathrm{MP}} = \begin{bmatrix} x(k) & \cdots & x(k)|x(k)|^{P-1} & \cdots & x(k-Q)|x(k-Q)|^{P-1} \\ x(k-1) & \cdots & x(k-1)|x(k-1)|^{P-1} & \cdots & x(k-Q-1)|x(k-Q-1)|^{P-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(k-(M-1)) & \cdots & x(k-(M-1))|x(k-(M-1))|^{P-1} & \cdots & x(k-Q-(M-1))|x(k-Q-(M-1))|^{P-1} \end{bmatrix}. \tag{2.10}$$

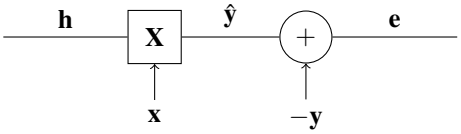

Figure 2.6 Block diagram of Volterra series represented as a measurement process where $\mathbf{h}$ is the vector of Volterra coefficients to be estimated.

Taking into account this linear relation and following the block system shown in figure 2.6, the measurement equation $\hat{\mathbf{y}} = \mathbf{X}\mathbf{h}$, the error vector can be defined as

$$\varepsilon = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{h}. \tag{2.11}$$

Minimizing the norm of the error is performed by finding the minimum of

$$\begin{aligned} \|\varepsilon(\mathbf{h})\|^2 = \varepsilon^H \varepsilon &= (\mathbf{y}^H - \mathbf{h}^H \mathbf{X}^H)(\mathbf{y} - \mathbf{X}\mathbf{h}) \\ &= \mathbf{y}\mathbf{y}^H - \mathbf{y}^H \mathbf{X}\mathbf{h} - \mathbf{h}^H \mathbf{X}^H \mathbf{y} + \mathbf{h}^H \mathbf{X}^H \mathbf{X}\mathbf{h}, \end{aligned} \tag{2.12}$$

where $\mathbf{A}^H$ is the hermitian transpose of matrix $\mathbf{A}$. Taking the derivative of this equation with respect to the kernel vector and setting it to zero,

$$\frac{\partial \|\varepsilon(\mathbf{h})\|^2}{\partial \mathbf{h}} = -2\mathbf{X}^H \mathbf{y} + 2\mathbf{X}^H \mathbf{X}\mathbf{h} = 0, \tag{2.13}$$

we get the so-called *deterministic normal equation*,

$$\mathbf{X}^H \mathbf{y} = \mathbf{X}^H \mathbf{X}\hat{\mathbf{h}}, \tag{2.14}$$

where $\mathbf{h}$ has become $\hat{\mathbf{h}}$ since its value is the estimated optimum value after setting the derivative equal to zero. Please note that $\mathbf{X}$ is not generally square, therefore it is not invertible. Nevertheless, the product $\mathbf{X}^H \mathbf{X}$ is an invertible square matrix, allowing us to rearrange

$$\hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y}. \tag{2.15}$$

The term $\mathbf{X}^+ = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H$ is commonly known as the Moore–Penrose pseudoinverse, which it is a generalization of the inverse matrix for non-square matrices. In a LS regression, the estimate error ε is orthogonal to the columns of the measurement matrix.

2.3.3 Ridge Regression

Ridge regression is similar to LS, with the exception that the Penalized Residual Sum of Squares (PRSS) is minimized instead of the Residual Sum of Squares (RSS).

$$PRSS = \|\mathbf{y} - \mathbf{X}\mathbf{h}\|_2^2 + \lambda \|\mathbf{h}\|_2^2 = \mathrm{RSS} + \lambda \|\mathbf{h}\|_2^2, \tag{2.16}$$

where λ is the tuning parameter. Ridge regression trades off two different criteria. The first term seeks coefficients that fit the data well, producing a small RSS. The shrinkage penalty $\lambda \|\mathbf{h}\|_2^2$ has the effect of shrinking the coefficients towards zero. The tuning parameter λ controls the impact of both terms on the regression. When $\lambda = 0$, the shrinkage penalty has no effect and the equation becomes the LS regression. When $\lambda \to \infty$, the penalty grows and the estimates approach zero. Ridge regression produces a set of estimates instead of a single solution as the LS regression does, therefore a selection procedure for λ is critical. As the solution depends on the scaling of the predictors, a good practice is to standardize

the predictors by dividing them by its ℓ_2 norm:

$$\bar{\mathbf{X}}_i = \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|_2}. \tag{2.17}$$

The advantage of Ridge regression over LS is the existence of the tuning parameter and that it allows to control the bias-variance trade-off. As λ increases, the flexibility of the Ridge regression decreases, leading to decreased variance but increased bias.

In cases where the number of samples is higher than the number of regressors, as it is with Volterra series, Ridge regression can perform well by trading off a small increase in bias for a large decrease in variance.

An alternative formulation for Ridge regression is solving the problem

$$\underset{h}{\text{minimize}} \|\mathbf{y} - \mathbf{X}\mathbf{h}\|_2^2 \quad \text{subject to} \quad \|\mathbf{h}\|_2^2 \leq s. \tag{2.18}$$

The estimator of a Ridge regression has a closed form that follows

$$\hat{\mathbf{h}}_{\text{Ridge}} = (\mathbf{X}^H\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^H\mathbf{y}, \tag{2.19}$$

where $\mathbf{I}$ is the identity matrix.

2.3.4 The LASSO

The main disadvantage of Ridge regression is that all the p predictors are included in the final model. The penalty $\lambda\|\mathbf{h}\|_1$ will shrink all the coefficients towards zero, but it will not set any of them exactly to zero, except for $\lambda = \infty$.

The LASSO is an alternative to the Ridge regression that overcomes this disadvantage.

$$\|\mathbf{y} - \mathbf{X}\mathbf{h}\|_2^2 + \lambda\|\mathbf{h}\|_1 = \text{RSS} + \lambda\|\mathbf{h}\|_1. \tag{2.20}$$

The formulation is very similar to the Ridge regression but the penalty includes the ℓ_1 norm instead of the ℓ_2.

The LASSO shrinks the coefficients towards zero. However, in comparison with the Ridge regression, the ℓ_1 penalty forces the coefficients to be exactly equal to zero when the tuning parameter λ is large. The LASSO yields sparse models —which involve a subset of variables—. In the case of the LASSO regression the selection of λ is critical as well.

The equivalent to the formulation in 2.19 to the LASSO is

$$\underset{h}{\text{minimize}} \left\{ \|\mathbf{y} - \mathbf{X}\mathbf{h}\|_2^2 + \lambda\|\mathbf{h}\|_1 \right\} \quad \text{subject to} \quad \|\mathbf{h}\|_1 \leq s. \tag{2.21}$$

In a two-dimensional plane with coordinates β_1 and β_2, Ridge regression seeks the solution with the constraint $\beta_1^2 + \beta_2^2 \leq s$, while the LASSO constraint has the form $|\beta_1| + |\beta_2| \leq s$. Figure 2.7 shows the constraints for both regressions. The LS solutions are marked as $\hat{h}$ and all the points in the same ellipse have a common value in RSS. The RSS increases as the ellipses are further from the LS solution. The Ridge and LASSO solutions are the first point at which an ellipse contacts the constraint region. The constraint of the LASSO has a diamond shape and the Ridge regression's is sphere-shaped. Since Ridge

has no sharp points, this intersection will not generally occur in the axis -equivalent to sparse solutions-. However, the LASSO constraint has corners in the axis which will often intercept the ellipse at an axis. When this occurs, those coefficients will be equal to zero. In higher dimensions, many of the coefficients may be equal to zero simultaneously.

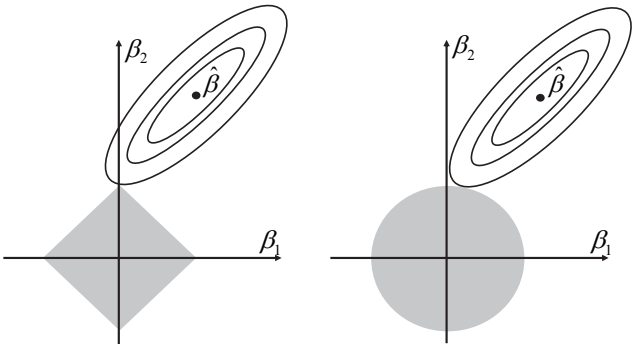

Figure 2.7 Contours of the error and constraint functions for the LASSO (left) and Ridge regression (right) in a space with coordinates β_1 and β_2. The solid areas are the constraint regions, $|\beta_1| + |\beta_2| \leq s$ and $\beta_1^2 + \beta_2^2 \leq s$, while the ellipses are the contours of the RSS. Figure adapted from [13].

2.3.5 Selection of the Tuning Parameter

The general technique for selecting a value for the tuning parameter is known as cross-validation. Cross-validation is a simple procedure that consists on choosing a grid of λ values and computing the cross-validation error for each value of λ. The λ that gives the lowest cross-validation error is chosen and the model is re-fit with all the available observations.

2.3.6 Adaptive Optimization

Adaptive optimization is a technique where the estimation of the kernel vector is updated dynamically in the time. It has the advantage of being more robust to the change of conditions of the system. This family of algorithms is able to follow the effects that may appear with a change of any variable -signal type, power level, temperature, etc.- with the counterpart of constantly update the model of the system. The scheme is very similar to that of the LS, but taking into account only the last N samples of each of the signals. In this topic, the index n is introduced to express the dependency with the time instant.

2.3.7 Steepest Descent

The block diagram followed in this topic is shown in figure 2.8. Please note that it has the same structure than figure 2.6 with the addition of dependence on the time or iteration variable n. For the derivation of the algorithm, we depart from the measurement process equation that takes the form

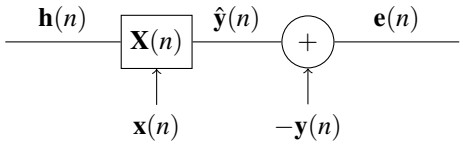

Figure 2.8 Block diagram of Volterra series represented as a measurement process where **h** is the vector of Volterra coefficients to be estimated.

$$\hat{\mathbf{y}}(n) = \mathbf{h}^H(n)\mathbf{X}(n), \tag{2.22}$$

where $\hat{\mathbf{y}}(n) \in \mathbb{C}^M$ is the estimated output in the iteration n, $\mathbf{h}(n) \in \mathbb{C}^N$ is the Volterra vector with the N coefficients of the model, $\mathbf{X}(n) \in \mathbb{C}^{M\times N}$ is the measurement matrix, whose N columns corresponds to the regressors of the model formed with a section of the last N samples of the input signal $\mathbf{x}(n) = [x(n)x(n-1)\ldots x(n-M+1)]$. The optimization error is defined as

$$\boldsymbol{\varepsilon}(n) = \mathbf{y}(n) - \hat{\mathbf{y}}(n) = \mathbf{y}(n) - \mathbf{h}^H(n)\mathbf{X}(n). \tag{2.23}$$

Considering the ℓ_2-norm as the cost function to minimize,

$$J(n) = E\left[\|\boldsymbol{\varepsilon}(n)\|^2\right] = E\left[\left(\mathbf{y}(n) - \mathbf{h}^H(n)\mathbf{X}(n)\right)\left(\mathbf{y}^H(n) - \mathbf{X}^H(n)\mathbf{h}(n)\right)\right], \tag{2.24}$$

and handling the terms considering the random character of $\mathbf{X}(n)$ and $\mathbf{y}(n)$,

$$\begin{aligned} J(n) = {} & E\left[\mathbf{y}(n)\mathbf{y}^H(n)\right] \\ & - E\left[\mathbf{y}(n)\mathbf{X}^H(n)\right]\mathbf{h}(n) \\ & - \mathbf{h}^H(n)E\left[\mathbf{X}(n)\mathbf{y}^H(n)\right] \\ & + \mathbf{h}^H(n)E\left[\mathbf{X}(n)\mathbf{X}^H(n)\right]\mathbf{h}(n). \end{aligned} \tag{2.25}$$

The final cost function remains as

$$J(n) = \sigma_y^2 + \mathbf{p}^H\mathbf{h}(n) - \mathbf{h}^H(n)\mathbf{p} + \mathbf{h}^H(n)\mathbf{R}\mathbf{h}(n), \tag{2.26}$$

where $\mathbf{p}$ is the cross-correlation between $\mathbf{X}(n)$ and $\mathbf{y}^H(n)$ and $\mathbf{R}$ is the correlation matrix of $\mathbf{X}(n)$. If we set the direction of the kernel vector update to $-\nabla J(n) = 2\mathbf{p} - 2\mathbf{R}\mathbf{h}(n)$,

$$\begin{aligned} \mathbf{h}(n+1) &= \mathbf{h}(n) + \frac{1}{2}\mu\left[-\nabla J(n)\right] \\ &= \mathbf{h}(n) + \mu\left[\mathbf{p} - \mathbf{R}\mathbf{h}(n)\right], \end{aligned} \tag{2.27}$$

where μ is the step size, which can be demonstrated that for the algorithm to converge, it should have a value in the range of $0 < \mu < \frac{2}{\lambda_{\max}}$, being $\lambda_{\max}$ is the maximum eigenvalue of the correlation matrix $\mathbf{R}$.

2.3.8 LMS Algorithm

The *a priori* information required by the steepest descent algorithm is not always known. One possible approach for this issue is to use estimated values for the cross-correlation $\mathbf{p}$ and the correlation matrix $\mathbf{R}$. The resulting algorithm is known as LMS, developed in the 60s by Professor Bernard Widrow and his first PhD student Ted Hoff [14]. Using the following instantaneous estimates:

$$\begin{aligned} \hat{\mathbf{R}}(n) &= \mathbf{X}(n)\mathbf{X}^H(n), \\ \hat{\mathbf{p}}(n) &= \mathbf{X}(n)\mathbf{y}^H(n). \end{aligned} \tag{2.28}$$

Utilizing the direction of the update

$$-\nabla J(n) = -2\mathbf{X}(n)\mathbf{y}^H(n) + 2\mathbf{X}(n)\mathbf{X}^H(n)\mathbf{h}(n) = -2\mathbf{X}(n)\varepsilon^H(n) \tag{2.29}$$

in the update of the estimated Volterra vector

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu\mathbf{X}(n)\varepsilon^H(n), \tag{2.30}$$

we get the LMS update equation. This update is extremely simple and it has the property of averaging the large variance that the instantaneous estimates may have. Due to this simplicity and good performance of the algorithm, it is considered a standard benchmark against which other algorithms are compared.

2.4 Experimental Testbenches

The algorithms developed in this thesis have been probed through two different experimental setups. Both facilities, which are maintained by the University of Seville and the Chalmers University of Technology are described in this section.

2.4.1 University of Seville Testbench

The measurement platform of the University of Seville is shown in figure 2.9. It is based on a evaluation board of the CGH40010 GaN HEMT transistor, from Cree Inc., driven by a SMU200A Vector Signal Generator (VSG) [15] from Rohde & Schwarz, and a PXA N9030A Vector Signal Analyzer (VSA) [16] from Agilent Technologies that records the output with a maximum sampling frequency of 100 MHz. The VSG allows to modulate the carrier with arbitrary baseband waveforms hence it allows to use a custom signal while driving the Device Under Test (DUT). In the receiver side, the RF signal is down-converted to baseband and the appropriate range, span and resolution bandwidth are set to optimize the measurement dynamic range.

The DUT is the cascade of a ZHL4W driver amplifier from Minicircuits and a CGH40010 GaN HEMT transistor from Cree Inc. in its evaluation board, shown in figure 2.12. The ZHL42W is operating in its linear region and acts as a driver for the CGH40010.

The set of equipments are controlled through either General-Purpose Instrumentation Bus (GP-IB) or Registered Jack 45 (RJ45) physical channels by using the Standard Commands for Programmable Instruments (SCPI). A custom Matlab script controls the sending of the signal to the VSG and the settings to reproduce it in the transmitter. In the receiver, the same configuration is applied to recover the signal properly. Next, the signal is downloaded to the Matlab workspace and it is postprocessed to adequate it to the algorithms. The postprocessing consists on the following operations:

- Normalization: scaling of the signal taking into account the attenuations and gains of the measurement chain.
- Time alignment: synchronization of the input and output signals in time.
- Partition into several datasets: sets for identification and validation of the models.

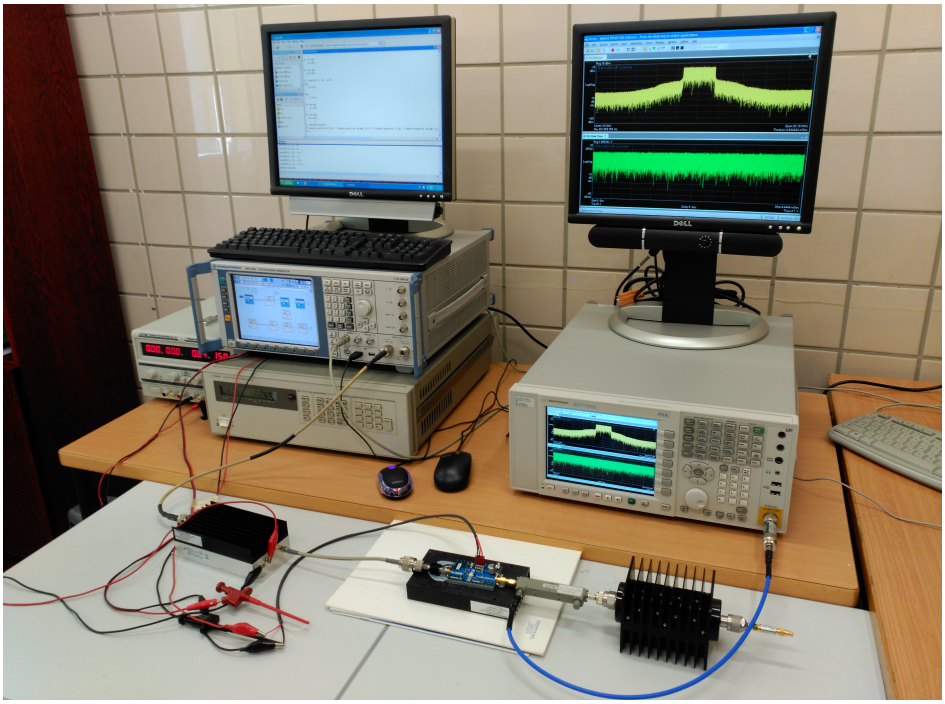

Figure 2.9 Experimental test bench. From left to right. Front: ZHL42W Minicircuits' driver amplifier, CGH40010 Cree's evaluation board, coupler and power load. Back: SMU200A vector signal generator, PXA N9030A vector signal analyzer.

2.4.2 Chalmers University of Technology Weblab

The weblab for PA DPD and characterization maintained by the Chalmers University of Technology [17] is shown in figure 2.11. This online laboratory has a web interface that allows to take measurements directly from the Matlab workspace and compare distinct

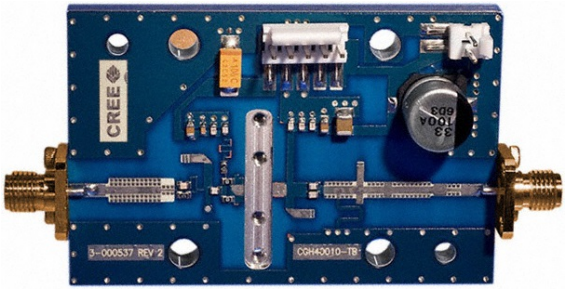

Figure 2.10 Evaluation board of the CGH40010 GaN HEMT transistor.

approaches to PA modeling and predistortion in a common test bench following the diagram in figure 2.12. The weblab setup consists of the PXI chassis PXIe-1082 from National Instruments Inc., which has the vector signal transceiver PXIe-5646R installed with a maximum instantaneous bandwidth of 200 MHz and is capable of reproducing the signal uploaded to the web interface. The DUT is the GaN PA CGH40006P in its test board. The output of the PA is attenuated 30 dB and then connected to the VST receiver.

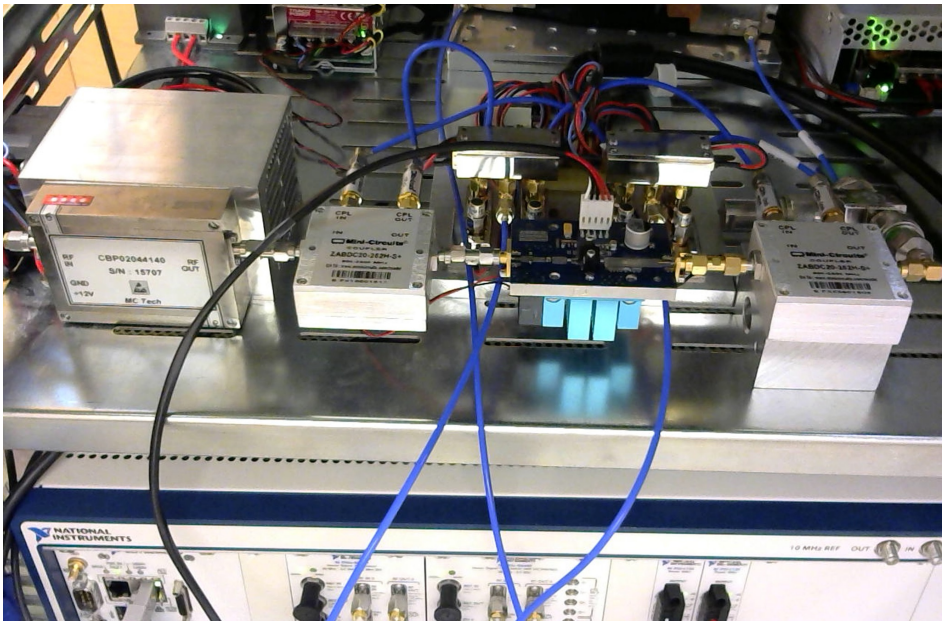

Figure 2.11 Picture of the Chalmers University of Technology weblab.

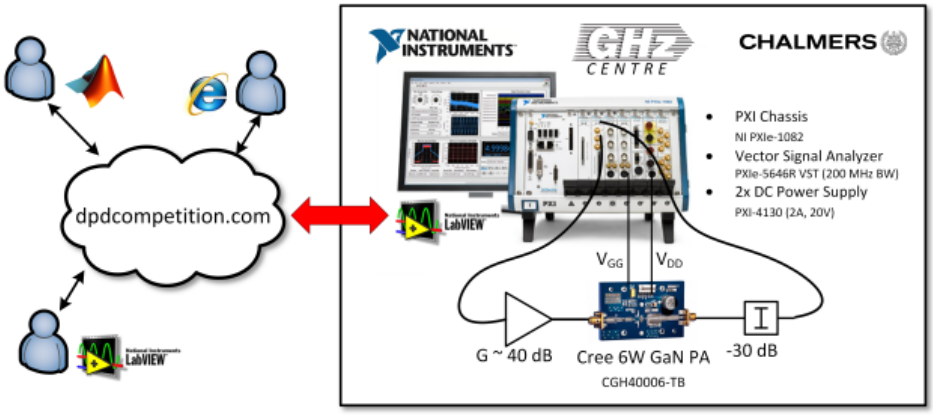

Figure 2.12 Diagram of the Chalmers University of Technology weblab.

3 Application of Efficient Frequency Domain Techniques to Volterra-Parafac Digital Predistorters

3.1 Introduction

Deployment of modern wireless communication systems, based on spectrally efficient modulation schemes, is perhaps the principal agent that has pushed the renaissance of new ultra-linear and highly efficient PAs. A great number of these wireless standards (WiFi, WiMAX, LTE, DVB-T, etc.) use Orthogonal Frequency Division Multiplexing (OFDM) and Quadrature Amplitude Modulation (QAM) signals with high PAPR that makes necessary the application of amplitude-reduction approaches and linearization techniques. In particular, the use of a DPD in the communication transmitter has demonstrated an outstanding capability to procure linear compensation of the PA nonlinearities with no degradation of the efficiency.

The main purpose of the DPD is the reduction of the spectral regrowth to provide a low ACPR, but in-band and out-of-band linearization are often contending parts in the optimization process, and a complete design must also include a high-quality in-channel equalization to comply with the EVM specifications of the wireless communications systems. A second important goal when the DPD is applied to an actual communications system is to provide a good tradeoff between the necessary accuracy of the transmitted signal and its computational complexity.

In the majority of DPDs, the proposed architecture is based on different time-domain Volterra behavioral models: the Generalized Memory Polynomial (GMP) model [18], the Dynamic Deviation Reduction (DDR) model [19], nonlinear two-box models [20], and others. An alternative structure that exploits the symmetric Parafac decomposition of the

Volterra kernels viewed as tensors has been demonstrated in [21]. This Volterra-Parafac (VP) structure can be represented as an arrangement of parallel branches formed with cascaded linear filters and static homogeneous nonlinearities. Recently, a new baseband VP model with a reduced parametric complexity has been derived by using a doubly symmetric Parafac decomposition [22]–[23].

Based on the encouraging results of PA linearization in [24], this chapter is dedicated to demonstrate the design of two Volterra-Parafac DPDs under a frequency domain perspective and show their improvement capacity, mainly in terms of computational complexity. In the first approach, the DPD operates on the discrete-time samples of the input waveform to produce the predistorter signal. Block processing and frequency-domain techniques are applied to improve the computational efficiency. We refer to this technique as block-processed Volterra-Parafac (B-VP) DPD and it has been favourably compared to a DPD designed with a high-standard model, the GMP. In the second case, the predistorter acts directly on the data symbols, exploiting an attractive feature of the VP model in the particular case of OFDM systems, i.e., its ability to integrate the M-QAM symbols within the signal processing technique. In this approach, designated as frequency-domain Volterra-Parafac (FD-VP) DPD, the DPD input data is the block of M-QAM symbols corresponding to the active subcarriers in the spectral domain, and the output is the discrete-time predistorter signal.

The discussion of the Volterra-Parafac approach, block adaptive filtering and frequency-domain techniques applied to VP models is developed in section 3.2. In section 3.3, the computational complexity of the model is analyzed.

3.2 Volterra-Parafac Models and Frequency-Domain Techniques

Making use of the fact that the kernels of the full Volterra behavioral model can be seen as symmetric multidimensional arrays, the tensorial Parafac decomposition has been applied to introduce the VP representation [21]. Since in this work we use the corresponding discrete-time baseband structure published in [23] to model the DPD, we dedicate the following subsection to expose a brief overview of this VP representation and the parameter identification procedure.

3.2.1 The Volterra-Parafac Baseband Model

The VP baseband model is composed of several branches, each one containing a low-pass parallel Wiener structure with an nth-order baseband nonlinearity, as it is shown in figure.3.1. The input to the DPD is given by the complex-envelope samples of the wireless signal $x[k]$, and the output is the complex-envelope of the predistorted signal $u[k]$. The number of branches for each nonlinear order is equal to the symmetrical rank of the nth-order kernel, R_n, and the filter for the particular branch (n,r) is defined by its weights $w_{n0}^{(r)}[k],\cdots,w_{nQ_n}^{(r)}[k]$. Therefore, the output of the filter at discrete time instant k is given by the convolution

$$z_n^{(r)}[k] = \mathbf{w}_n^{(r)T}[k]\mathbf{x}_n[k], \tag{3.1}$$

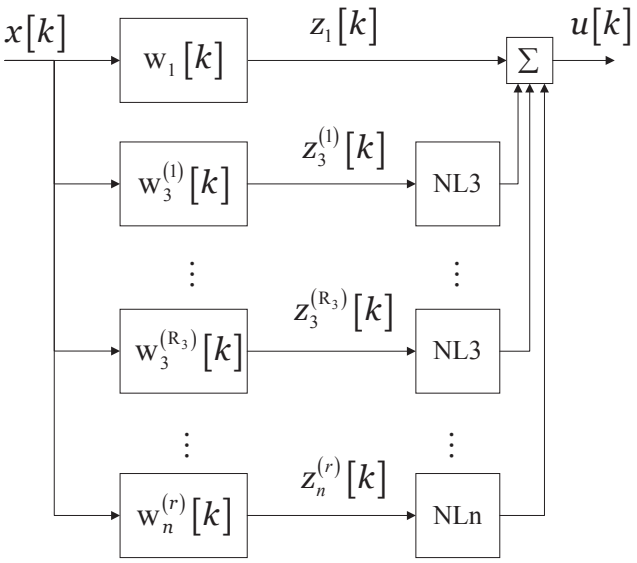

Figure 3.1 Structure of the Volterra-Parafac baseband model.

where $\mathbf{x}_n[k] = [x[k], x[k-1], \cdots, x[k-Q_n]]^T$ is the input vector associated to the nth-order Volterra regression term, and the weight vector $\mathbf{w}_n^{(r)}[k]$ is defined as

$$\mathbf{w}_n^{(r)}[k] = \left[w_{n0}^{(r)}[k] \cdots w_{nQ_n}^{(r)}[k] \right]^T. \tag{3.2}$$

If we write the output of the nonlinearity in the branch (n,r) as

$$u_n^{(r)}[k] = \left| z_n^{(r)}[k] \right|^{n-1} z_n^{(r)}[k], \tag{3.3}$$

and collect the contribution of all the branches, the output of the VP model can be enunciated as

$$u[k] = \sum_{n=1}^{N}{}' \sum_{r=1}^{R_n} u_n^{(r)}[k] = \sum_{n=1}^{N}{}' \sum_{r=1}^{R_n} \left| z_n^{(r)}[k] \right|^{n-1} z_n^{(r)}[k]. \tag{3.4}$$

The primma in the first sum indicates that only the contribution of the odd-order terms is included. Since the nonlinearity in each branch is already provided with the VP model, it is only necessary to identify the parameters of the filters to evaluate the predistorter output. The goal is that the output $y[k]$ produced by the PA when it is driven by $u[k]$, is equal to the linearized signal $y_{des}[k]$. The complex least mean square (CLMS) algorithm can be implemented to estimate $\mathbf{w}_n^{(r)}[k]$ by defining the error with respect to the desired output, $e[k] = y_{des}[k] - y[k]$, and minimizing the real-valued cost function $J(\mathbf{w}[k]) = \frac{1}{2}|e[k]|^2$.

Observe that the cost function depends on the adjustable weight vector $\mathbf{w}[k]$, defined as

$$\mathbf{w}[k] = \left[\mathbf{w}_1^{(1)T}[k]\ \mathbf{w}_3^{(1)T}[k]\cdots\mathbf{w}_3^{(R_3)T}[k]\cdots\mathbf{w}_N^{(1)T}[k]\cdots\mathbf{w}_N^{(R_N)T}[k]\right]^T. \tag{3.5}$$

A necessary and sufficient condition for $J(\mathbf{w})$ to have a stationary point is

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}^*} = -\frac{1}{2}e^*[k]\frac{\partial y(\mathbf{z})}{\partial \mathbf{w}^*} - \frac{1}{2}e[k]\frac{\partial y^*(\mathbf{z})}{\partial \mathbf{w}^*} = 0, \tag{3.6}$$

where the dependence of the output signal with the aggregated vector $\mathbf{z}[k] = \left[z_1^{(1)}[k]\cdots z_N^{(N)}[k]\right]^T$ is explicited.

Applying the steepest descent algorithm, we obtain the update equation for the estimated parameters:

$$\hat{\mathbf{w}}^T[k+1] = \hat{\mathbf{w}}^T[k] - \boldsymbol{\mu}\left.\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}^*}\right|_{\hat{\mathbf{w}}[k]}. \tag{3.7}$$

The diagonal matrix $\boldsymbol{\mu}$ contains the step sizes $\mu_n^{(r)}$ that control the convergence speed and the steady-state properties of the CLMS algorithm for each branch (n,r). The derivative can be computed using the chain rule by regarding $y[k]$ as a bivariate function $y(\mathbf{z},\mathbf{z}^*)$ and treating $\mathbf{z}$ and $\mathbf{z}^*$ as independent variables. The update equation for the branch (n,r) adopts the form

$$\mathbf{w}_n^{(r)}[k+1] = \mathbf{w}_n^{(r)}[k] + \mu_n^{(r)}\boldsymbol{\varepsilon}_n^{(r)}[k]\mathbf{x}_n^*[k]. \tag{3.8}$$

The defined variable

$$\boldsymbol{\varepsilon}_n^{(r)}[k] = \frac{n+1}{4}e[k]|z_n^{(r)}[k]|^{n-1} + \frac{n-1}{4}e^*[k]|z_n^{(r)}[k]|^{n-3}(z_n^{(r)}[k])^2 \tag{3.9}$$

can be intuitively interpreted as the mapping of the output error $e[k]$ onto the (n,r)-branch filter output.

Once the update procedure (3.8) has converged, the system output is easily evaluated by means of the convolutions (3.1) and the nonlinear memoryless equation (3.4). The direct application of this procedure in a sample-by-sample way is expensive from a computational cost perspective, hence the usefulness of the more efficient block processing techniques in the frequency domain becomes evident.

3.2.2 Block Processing and FFT Algorithm Applied to the VP Structure (B-VP DPD)

An immediate strategy to reduce the computational complexity of the filter convolutions is to incorporate, together with the block processing technique, the fast Fourier transform (FFT) algorithms that efficiently perform the convolutions. Assuming this perspective in the estimation procedure, we can segment the input signal in blocks of L samples each, and keep the filter parameters fixed until a new block is received. For the sake of clarity, from now on the input sequence will be written as $x_k[l] = x[kL+l]$, where $l = 0,1,\ldots L-1$ is the time index inside the kth block, and the filter parameters of the current block will be denoted as $\mathbf{w}_{n,k}^{(r)}$. This approach allows to execute the update equation at a lower sampling

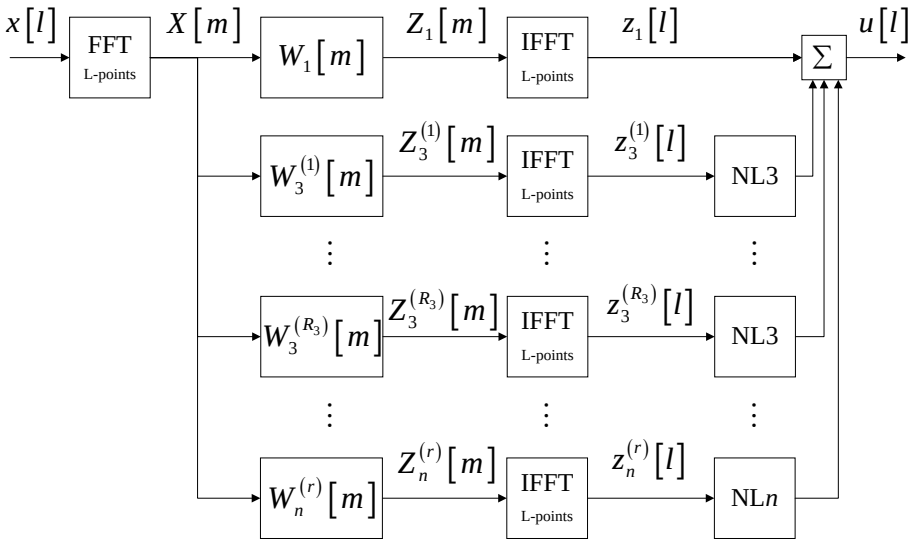

Figure 3.2 Volterra-Parafac structure and implementation of block processing with efficient frequency-domain techniques (B-VP DPD). The block index k is omitted.

rate, following the recursion for the block $k+1$. If the definition $\varepsilon_{n,k}^{(r)}[l] = \varepsilon_n^{(r)}[kL+l]$ is used, this update equation can be written as

$$\mathbf{w}_{n,k+1}^{(r)} = \mathbf{w}_{n,k}^{(r)} + \mu_n^{(r)} \sum_{l=0}^{L-1} \varepsilon_{n,k}^{(r)}[l]\mathbf{x}_k^*[l]. \tag{3.10}$$

It is well-known that block processing can be implemented in a computationally more efficient manner by performing the parameter estimation in the frequency domain using the FFT algorithm. We have selected the filter dimension equal to the block length L and followed the procedure given by the unconstrained method [25], [26].

The resulting structure of this approach is depicted in figure 3.2. The input signal $x[k]$ is accumulated in a memory buffer to form L-point data blocks and transformed by the Fourier transform matrix, denoted as $\mathbf{F}$. Let us define the input signal diagonal matrix for the current block k as

$$\mathbf{X}_k = \text{diag}\{\mathbf{F}[x_k[0], \cdots, x_k[L-1]]^T\} \tag{3.11}$$

and the frequency-domain weight vector for each filter as

$$\mathbf{W}_{n,k}^{(r)} = [W_{n,k}^{(r)}[0], \cdots, W_{n,k}^{(r)}[L-1]]^T = \mathbf{F}\mathbf{w}_{n,k}^{(r)}, \tag{3.12}$$

then, the frequency-domain output vector of the filter (n,r) can be written as

$$\mathbf{Z}_{n,k}^{(r)} = \mathbf{X}_k \mathbf{W}_{n,k}^{(r)}. \tag{3.13}$$

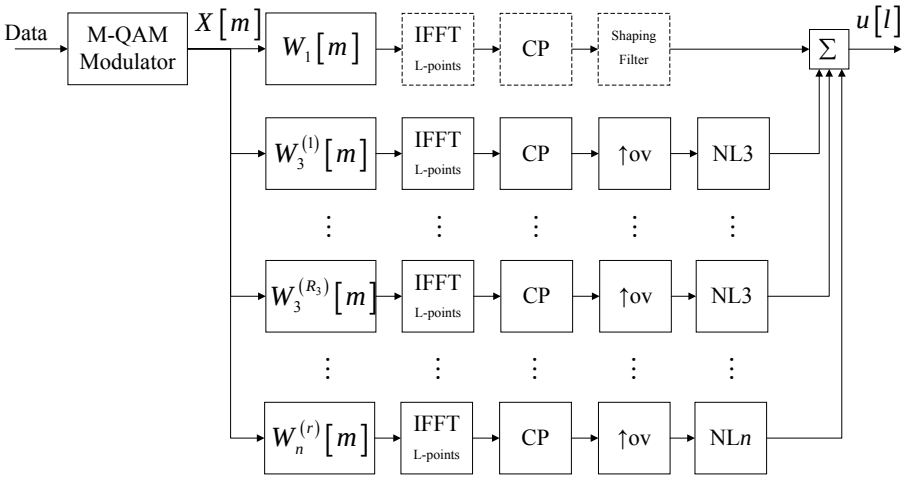

Figure 3.3 A DPD with Volterra-Parafac structure for OFDM systems (FD-VP DPD). The block index k is omitted.

Once the parameters $\mathbf{W}_{n,k}^{(r)}$ have been identified, the computation is implemented by using efficient frequency-domain techniques and inverse-transforming the filter output to the time domain. After applying the corresponding homogeneous nth-order baseband nonlinearity, the predistorted output $u_k[l]$ is straightforwardly computed. If this B-VP approach is used in the implementation of the DPD, the model operates over the time-domain input signal to produce the time-domain predistorted signal $u_k[l]$.

3.2.3 Particularization to OFDM Systems (FD-VP DPD)

In OFDM systems, the transmitter generates the signal by assembling blocks of the input M-QAM symbols, and mapping them into subcarriers in the frequency domain. Therefore, the use of a DPD with the output in the time domain, but operating directly on the input symbols of the kth block, denoted as $X_k(m)$, seems very attractive. With reference to figure 3.3, the M-QAM modulator generates the symbol sequence, and the linear branch processes these symbols by filtering, transforming them to the time domain with an inverse FFT (IFFT), adding the cyclic prefix (CP) and filtering to conform a band-limited spectrum. The modules that are represented with a dashed contour in the figure have been integrated in the linear branch of the DPD, but these elements are inherent of a typical OFDM transmitter.

In this second approach, the estimation procedure can be performed following the general form of the frequency-domain update algorithm, given by

$$\mathbf{W}_{n,k+1}^{(r)} = \mathbf{W}_{n,k}^{(r)} + \boldsymbol{\mu}_n^{(r)} \mathbf{X}_k^* \boldsymbol{\varepsilon}_{n,k}^{(r)}. \tag{3.14}$$

The vector $\boldsymbol{\varepsilon}_{n,k}^{(r)}$ represents the frequency-domain transformation of the error vector in the (n,r)-branch output, $\boldsymbol{\mu}_n^{(r)}$ is a diagonal matrix with the step sizes, and $\mathbf{X}_k$ refers to (3.11).

The computation of the error term used in (3.14) depends on an output error like in (3.9), but here this error shall compensate for the frequency-domain error between the OFDM symbols passed through the PA model ($\mathbf{Y}_k$) and the linearized OFDM symbols ($\mathbf{X}_k$), expressed as:

$$E_k(m) = Y_k(m)/G_c - X_k(m), \tag{3.15}$$

where m stands for the frequency bins ($m = 0, \ldots, L$), k is the block index, and G_c is the target linearized gain.

Since the input symbols $X_k(m)$ and the parameters $W_{n,k}^{(r)}(m)$ are specified in the frequency domain, it is not necessary to compute the FFTs (3.11) and (3.12), and the filter operation is performed by a simple product. In each nonlinear branch, an inverse transform to the time domain and insertion of the CP modules are included. In the same form, an oversampling operation is applied before the homogeneous nth-order memoryless nonlinear block. This oversampling is necessary to match the sampling rate produced by the spectrum shaping filter of the linear branch. Although in this FD-VP DPD structure, the input band is approximately equal to the bandwidth of the OFDM signal, the approach is sufficient to guarantee in-band equalization as well as minimization of the spectral regrowth in the adjacent channels.

3.3 Complexity Assessment

A common measure to evaluate the complexity of a behavioral model for DPD is the number of floating point operations (FLOPs) [7]. While both the identification complexity, running complexity, and adaptation complexity may be distinguished, the running complexity is the most demanding term. Its consideration entails the complexity associated with the construction of the basis functions and filtering the basis with the model kernels.

To set the notation, the complexity associated with the Fourier transform or inverse Fourier transform of a block of L points will be denoted as C_{FFT-L}. Other elementary operations such as the complex-to-complex product, the complex-to-real product, the real-to-real product, the complex summation, and the square-modulus will be expressed in terms of their complexity as $C_{prodc-c}$, $C_{prodc-r}$, $C_{prodr-r}$, C_{sumc}, and $C_{|\cdot|^2}$, respectively. The number of FLOPs corresponding to these elementary operations is detailed in Table 1 of [7]. As general remarks, considering that both of the proposed approaches yield a block of L samples each time, in order to compare their complexity with a conventional DPD computed in the time-domain, the per-sample complexity will be calculated by dividing the complexity associated with a block of L samples over the block size. On the other hand, let us recall that filtering a block of L points in the frequency domain with a filter of L taps involves L complex products only, instead of L complex products and $L-1$ complex summations that would be required, per sample, to compute the convolution in the time domain.

Regarding the per-sample complexity of the static nonlinearities, for order $n = 2p+1$ ($p = 1, \ldots, P$), $p-1$ real-to-real products are required, which refer to products among the square-modulus terms to yield the $(n-1)$-th power of the absolute value, in addition to one complex-to-real product. Let us also recall that the branch-summation operator block for both approaches involves a number of complex summations equal to the number of

branches minus one. In order to derive compact expressions for the complexity, the same rank R will be assumed for all the nonlinear orders.

3.3.1 B-VP DPD Approach

For the construction of the basis functions, the following operations are required:

- A Fourier transform, corresponding to the time-domain to frequency-domain conversion of the input signal.
- L complex-to-complex products per order and rank, resulting from the computation of $\mathbf{Z}_{n,k}^{(r)}$ in (3.13).
- An L-point IFFT per order and rank.
- L square-modulus per order and rank (with the exception of the linear branch).

In summary, the complexity associated with the construction of the basis functions for a block of L samples can be expressed as:

$$C_{basis,B-VP} = C_{FFT-L} + R(P+1)\left[LC_{prodc-c} + C_{FFT-L}\right] + LRPC_{|\cdot|^2}. \quad (3.16)$$

Filtering itself corresponds to passing the block samples through the static nonlinearities plus branch summation. This filtering corresponds to a per-sample complexity given by:

$$C_{filt,B-VP} = R\frac{P}{2}(P-1)C_{prodr-r} + PRC_{prodc-r} + [R(P+1)-1]C_{sumc-c}. \quad (3.17)$$

The per-sample FLOPs of a B-VP DPD structure of order $N = 2P+1$ and rank R can be obtained by dividing (3.16) over the block size L and combining with (3.17):

$$C_{B-VP} = \frac{1}{L}C_{FFT-L} + R(P+1)\left[C_{prodc-c} + \frac{1}{L}C_{FFT-L}\right] + RPC_{|\cdot|^2} + \\ + R\frac{P}{2}(P-1)C_{prodr-r} + PRC_{prodc-r} + [R(P+1)-1]C_{sumc-c} \quad (3.18)$$

3.3.2 FD-VP DPD Approach

Here, it is assumed that the linear processes associated with the linear branch, except the filtering by the equalizer $\mathbf{W}_{1,k}^{(1)}$, do not contribute to the system complexity, as they belong to the OFDM modulator, i.e. M-QAM symbol mapping, serializer-deserializer, L-point IFFT, CP insertion and spectrum-shape filtering. On the other hand, while CP insertion does not contribute to complexity itself, because it only involves reading a shifted segment of memory, it indirectly affects complexity as it changes the block size, becoming $L' = L + L_{CP}$, where block-size L is augmented with the CP length (L_{CP}). Similarly, oversampling does not increase complexity either, as it is performed through zero-padding, but the interpolation filter represents an additional cost per branch depending on the filter length (M) and oversampling rate (ovs), equivalent to $L' \cdot ovs \cdot C_{filter}$ FLOPs, where $C_{filter} = M \cdot C_{prodc-c} + (M-1)C_{sumc}$. Finally, including the complexity associated

with the complex multiplications required by the equalizer filters $\mathbf{W}_{n,k}^{(r)}$, IFFT blocks of the nonlinear branches, the static nonlinearities and branch summation, the per-sample complexity associated with a FV-VP structure of order $N = 2P+1$ and rank R can be expressed as:

$$\begin{aligned} C_{FD-VP} = \frac{L \cdot R(P+1)}{L' \cdot ovs} C_{prodc-c} + \frac{R \cdot P}{L' \cdot ovs} C_{FFT-L} + \\ + R \cdot P \left[M \cdot C_{prodc-c} + (M-1) C_{sumc-c} \right] + \\ + \frac{1}{2} R \cdot P(P-1) C_{prodr-r} + P \cdot R \cdot C_{prodc-r} + \\ + \frac{R \cdot P}{ovs} C_{|\cdot|^2} + [R(P+1)-1] C_{sumc} \,. \end{aligned} \quad (3.19)$$

4 Formal Deduction of a Volterra Series Model for Complex-valued Systems

4.1 Introduction

Volterra series (VS) is one of the most popular representations to model modern wireless communication systems [27]. The baseband signal is used in the modulator to generate the RF signal, possibly with linear and nonlinear impairments, and the power amplifier (PA) delivers the signal at the desired level, adding further nonlinear distortion. Without nonlinear effects, the complex envelope of the RF output signal can be viewed as a linear transformation of the complex-valued baseband signal $x[k] = x_I[k] + jx_Q[k]$. However, to exploit the complete statistical characterization of data in complex-valued signal processing, access to the information contained in the complementary correlation is required through the introduction of widely linear (WL) transformations [28]–[29]. When nonlinear effects need to be modeled for real-valued systems, the VS approach is a natural choice, however, its extension to the complex-valued case is not a direct mapping, and particular models are usually deduced to match specific problems. Examples of these are the proposals in array processing and beamforming [30]–[31], or in the case of impairments compensation in I/Q modulators [32], but the use of these particular models is not justified to describe the output of other systems. The availability of a general complex-valued Volterra nonlinear model is a relevant topic for the design of nonlinear communication systems. In this chapter, the VS approach to a general nonlinear system with complex-valued signals by using Wirtinger calculus [33] is extended.

4.2 Volterra Models for Real-valued Systems

For a nonlinear system described by a Volterra model with a real-valued signal $x[k]$ applied at the input, the output $y[k]$ can be written as [6]

$$y[k] = h_0 + \sum_{n=1}^{\infty} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_n} h_n[\mathbf{q}_n] \prod_{r=1}^{n} x[k-q_r], \tag{4.1}$$

where $h_n[\mathbf{q}_n]$ is the nth-order Volterra kernel, $\prod_{r=1}^{n} x[k-q_r] = x[k-q_1]x[k-q_2]\cdots x[k-q_n]$, $\mathbf{q}_n = [q_1, q_2, \cdots, q_n]^T$ is a vector of delays of the nth-order term, with $q_r = 0, 1, \cdots, Q_n$ for all r, and $\mathbf{Q}_n = [Q_n, Q_n, \cdots, Q_n]^T$ is the vector of maximum delays.

For bivariate nonlinear systems, the output can be expressed as a double Volterra series, equation (1) of [34]. If the two input signals are $x[k]$ and $z[k]$, the discrete-time output is expressed as

$$\begin{aligned} y[k] = h_{0,0} &+ \sum_{n=1}^{\infty} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{n,0}} h_{n,0}[\mathbf{q}_n] \prod_{r=1}^{n} x[k-q_r] + \sum_{m=1}^{\infty} \sum_{\mathbf{q}_m=\mathbf{0}}^{\mathbf{Q}_{0,m}} h_{0,m}[\mathbf{q}_m] \prod_{s=1}^{m} z[k-q_s] + \\ &+ \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{n,m}} \sum_{\mathbf{p}_m=\mathbf{0}}^{\mathbf{P}_{n,m}} h_{n,m}[\mathbf{q}_n, \mathbf{p}_m] \prod_{r=1}^{n} x[k-q_r] \prod_{s=1}^{m} z[k-p_s]. \end{aligned} \tag{4.2}$$

The vector of delays $\mathbf{p}_m$ and the vector of maximum delays $\mathbf{P}_m$ have been defined as in (4.1). The multidimensional functions $h_{n,0}[\mathbf{q}_n]$ and $h_{0,m}[\mathbf{q}_m]$ are standard Volterra kernels of order n and m, respectively, and the constant $h_{0,0}$ is the zeroth-order kernel. The third group of sums contains the bivariate Volterra kernels $h_{n,m}[\mathbf{q}_n, \mathbf{p}_m]$ and cross products of $x[k-q_r]$ by $z[k-p_s]$.

4.3 Specific Volterra Models for Complex-valued Systems

In wireless communication systems it is necessary a nonlinear baseband model to express the relationship between the input and the output complex envelopes. Let us review three different situations.

- **Baseband PA model**. The baseband model of a power amplifier can be derived from the Volterra series RF model (4.1) [27, 35]. In discrete-time, the baseband Volterra input-output relationship is

$$y[k] = \sum_{m=0}^{\infty} \sum_{\mathbf{q}_{2m+1}=\mathbf{0}}^{\mathbf{Q}_{2m+1}} h_{2m+1}[\mathbf{q}_{2m+1}] \prod_{r=1}^{m+1} x[k-q_r] \prod_{r=m+2}^{2m+1} x^*[k-q_r], \tag{4.3}$$

  where $n = 2m+1$ is the (odd) nonlinear order. Unlike the kernels of the real-valued Volterra case, in this model $h_n[\mathbf{q}_n]$ is symmetric under any permutation of its first $m+1$ indices, and it is also separately symmetric under any permutation of its last m indices. In the literature, the model (4.3) is referred to as the full-Volterra (FV)

model to distinguish it from other pruned baseband Volterra representations, e.g., the memoryless model. The FV model has been deduced exclusively for PAs, and direct application to other systems is not proven.

- **Beamforming**. For the narrowband array processing problem, a proposal was introduced in [30] and [31]. It is advanced by establishing the vector of complex amplitudes of the signals at the output of the sensors, $\mathbf{x}_k$. The input-output relation is defined as

$$y[k] = \sum_{m=1}^{M} \sum_{q=0}^{m} \mathbf{w}_{m,q}^{H} (\mathbf{x}_k^{\otimes q} \otimes \mathbf{x}_k^{*\otimes (m-q)}), \tag{4.4}$$

  where $\mathbf{w}_{m,q}$ is a complex filter and $\otimes$ denotes the Kronecker product. This approach was applied to a third-order Volterra minimum variance distortionless response beamformer, considering only polynomial terms of odd order.

- **I/Q modulators**. A proposal to model linear and nonlinear impairments in I/Q modulators was presented in [32]. The output $u[k]$ is given by

$$u[k] = u_0 + \sum_{n=1}^{\infty} \left\{ \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_n} h_{n,0}[\mathbf{q}_n] \prod_{r=1}^{n} x[k-q_r] + \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_n} h_{0,n}[\mathbf{q}_n] \prod_{r'=1}^{n} x^*[k-q_{r'}] + \right.$$
$$\left. + \sum_{\mu=1}^{n-1} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_n} h_{n-\mu,\mu}[\mathbf{q}_n] \prod_{r=1}^{n-\mu} x[k-q_r] \prod_{r'=n-\mu+1}^{n} x^*[k-q_{r'}] \right\}. \tag{4.5}$$

  This widely nonlinear (WNL) representation can be also viewed as a nonlinear extension of the WL transformation and presents analogies with the WNL Volterra beamformer (4.4).

Although the results reported above have shown a good performance in their respective fields, a particular method cannot be assumed to be valid in another unknown situation. The availability of a general nonlinear model that admits complex-valued input signals is mandatory in the joint modeling of the modulator and the PA in a communications transmitter, for instance.

4.4 A General Volterra Model for Complex Systems

The results concerning the output of nonlinear functions with complex arguments are not new, although they have gone largely unnoticed by the engineering community [29]. In particular, Wirtinger calculus (also denoted as $\mathbb{CR}$-calculus) presents an elegant approach, which allows keeping all computations and derivations in the complex domain with expressions very similar to the real-valued case, and in a more efficient time-consuming manner [36]. In a context where the input of the system x is complex-valued, the corresponding output y depends on x and also on the complex conjugate x^*. In that case, Wirtinger calculus allows treating x and x^* as independent variables (Theorem 1 of [36]), and therefore, the double Volterra series representation can be applied. If we replace z by x^* in (4.2), the complex-valued output of the system is obtained as

$$y[k] = h_{0,0} + \sum_{n=1}^{\infty} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{n,0}} h_{n,0}[\mathbf{q}_n] \prod_{r=1}^{n} x[k-q_r] + \sum_{m=1}^{\infty} \sum_{\mathbf{q}_m=\mathbf{0}}^{\mathbf{Q}_{0,m}} h_{0,m}[\mathbf{q}_m] \prod_{s=1}^{m} x^*[k-q_s] + \\ + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{n,m}} \sum_{\mathbf{p}_m=\mathbf{0}}^{\mathbf{P}_{n,m}} h_{n,m}[\mathbf{q}_n,\mathbf{p}_m] \prod_{r=1}^{n} x[k-q_r] \prod_{s=1}^{m} x^*[k-p_s]. \quad (4.6)$$

Remarking that the order of the cross-terms is given by $\nu = n+m$, the corresponding summation can be calculated by adding all the terms of the same order ν, with $m = 1,2,\cdots,\nu-1$, and then add all orders. If we change the notation in the three sums to designate the order of the terms with the same index n, the relation (4.6) can be written as

$$y[k] = h_{0,0} + \sum_{q_1=0}^{Q_{1,0}} h_{1,0}[q_1]x[k-q_1] + \sum_{q_1=0}^{Q_{0,1}} h_{0,1}[q_1]x^*[k-q_1] + \\ + \sum_{n=2}^{\infty} \left\{ \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{n,0}} h_{n,0}[\mathbf{q}_n] \prod_{r=1}^{n} x[k-q_r] + \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{0,n}} h_{0,n}[\mathbf{q}_n] \prod_{s=1}^{n} x^*[k-q_s] + \right. \\ \left. + \sum_{m=1}^{n-1} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}_{n-m,m}} \sum_{\mathbf{p}_m=\mathbf{0}}^{\mathbf{P}_{n-m,m}} h_{n-m,m}[\mathbf{q}_{n-m},\mathbf{p}_m] \prod_{r=1}^{n-m} x[k-q_r] \prod_{s=1}^{m} x^*[k-p_s] \right\}. \quad (4.7)$$

The complex Volterra series (CVS) model (6.1) is the counterpart of the real-valued Volterra series (4.1) when the input is complex-valued. Note that if the nonlinearity of the system can be neglected, the CVS model is reduced to the WL transformation.

Equation (6.1) can be written in a more compact form. To that aim, let us rearrange the elements of the nth-order tensor $h_{n-m,m}[\mathbf{q}_{n-m},\mathbf{p}_m]$ to form the vector $\mathbf{h}_{n-m,m}$ as

$$\mathbf{h}_{n-m,m} = [h_{n-m,m}[0,0,\cdots,0], h_{n-m,m}[0,0,\cdots,1], \cdots, \\ h_{n-m,m}[0,0,\cdots,P_{n-m,m}], \cdots, h_{n-m,m}[Q_{n-m,m},Q_{n-m,m},\cdots,P_{n-m,m}]]^T. \quad (4.8)$$

Recalling that the first $n-m$ indices have maximum delays $Q_{n-m,m}$ and the other m indices have maximum delays $P_{n-m,m}$ and using the definition $\mathbf{x}_{Q_{n-m,m},k} = [x[k],x[k-1],\cdots,x[k-Q_{n-m,m}]]^T$, (6.1) can be rewritten as

$$y[k] = h_{0,0} + \sum_{n=1}^{\infty} \sum_{m=0}^{n} \mathbf{h}_{n-m,m}^T \mathbf{x}_{Q_{n-m,m},k}^{\otimes(n-m)} \otimes \mathbf{x}_{P_{n-m,m},k}^{*\otimes m}. \quad (4.9)$$

The CVS model is valid to describe the three examples exposed above. In the PA case, only input products with x and x^* appearing $m+1$ and m times, respectively, must be considered and (6.1) is reduced to the baseband FV model. It is important to note that the CVS model has been demonstrated without any assumption on the internal structure of these systems and nonetheless it presents all the terms required for the nonlinear detection and estimation problem in narrowband array processing [30]–[31]. If we consider an array

of N_s narrowband sensors in (6.1), then $Q_{n-m,m} = P_{n-m,m} = N_s - 1$, for all n and $0 \leq m \leq n$. In that case, the CVS model (4.9) reduces to the WNL Volterra beamformer (4.4) for odd-order terms. Finally, in the case of I/Q modulators the WNL representation (4.5) is a particular case of the CVS model. In this chapter we also refer to (6.1) as WNL model.

Considering the increased model complexity because of its structure with a larger number of parameters, the technique presented in [37] was used to cope with this great number of coefficients, relying on the sparsity assumption for the kernels of the model. Recalling that the WNL model is linear with respect to the coefficients, (4.9) can be rewritten as

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{h} + \mathbf{e}, \tag{4.10}$$

where $\mathbf{y}$ is a column vector with the samples of the output complex-envelope, $\mathbf{e}$ is a noise vector, $\mathbf{h}$ is a vector which arranges sequentially the normalized Volterra coefficients of the model, and $\mathbf{X}$ is a measurement matrix whose columns stack the samples of the model regressors for the same order and delay. The solution to (4.10) is the LS estimate for the kernel vector. If only a few of the regressor coefficients are active, it is possible to apply the Orthogonal Matching Pursuit technique and the Bayesian Information Criterion aimed at determining the active support set of model coefficients [37]. This approach applied to the model provides a reduced-complexity structure which will be referred to as WNL compressed-sensing (WNL-CS) model. The particular model coefficients are still solved by using (6.4), but here it only applies to the subset of active regressors defining the WNL-CS structure.

5 Digital Predistortion of Power Amplifiers Using Structured Compressed-Sensing Volterra Series

5.1 Introduction

Recent advances in digital communication standards require the management of the trade-off between efficiency and nonlinearity. Digital predistortion comes up as a solution that allows the PA to operate near saturation, mitigating the distortion created in new standard signals such as OFDM, which are characterized by a high PAPR. DPDs rely on behavioral models [12], which usually take the form of Volterra series. The most general FV [38] series has a rich set of terms to represent the modeled system, but this number of components is usually very large due to its inherent structure. Given the limited real-time computational capability of nowadays Field-Programmable Gate Arrays (FPGAs), as it is desirable to keep an easy-to-manage number of components, researchers have developed a set of pruning strategies —also called sparse recovery techniques— . These strategies can be either ad-hoc, which include a subset of the FV such as the GMP or DDR among others, or based on information theory, which do not include the information of the intrinsic structure of the model [39]. Structural information based on the algorithm in [37] was incorporated in [40]. In this chapter, we show the structural pruning of Volterra series and validate the method in the DPD application, obtaining a reduced complexity model while keeping the level of performance.

This chapter is organized as follows. The algorithm is first formulated. Then, the experimental design of a DPD application for a commercial PA is presented and results are discussed.

Algorithm 1 Summary of the Structured Compressed-Sensing Algorithm for Volterra Series Models

Input: $n_{\max} > 0$, $\alpha \in [0,1]$, $f(\cdot)$, $\mathbf{X} \in \mathbb{C}^{m \times n}$, $\mathbf{y} \in \mathbb{C}^m$
Output: $S(t)$, n_{opt}
Initialization :
1: $\mathbf{r}^{(\mathbf{0})} \leftarrow \mathbf{y}$, $S(0) \leftarrow \varnothing$
2: **for** $t = 1$ to $n_{\max}$ **do**
3: $\theta_t \leftarrow (1-\alpha) \cdot \max_{i \notin S(t)} \frac{\left|\mathbf{X}_{\{i\}}^H \cdot \mathbf{r}^{(t-1)}\right|}{\|\mathbf{X}_{\{i\}}\|_2}$
4: $\mathbf{i}_{\text{pre}}^{(t)} \leftarrow \left\langle i \left| \frac{\left|\mathbf{X}_{\{i\}}^H \cdot \mathbf{r}^{(t-1)}\right|}{\|\mathbf{X}_{\{i\}}\|_2} > \theta_t \right. \right\rangle$
5: $i^{(t)} \leftarrow \arg\min_{i \in \mathbf{i}_{\text{pre}}^{(t)}} f\left(\mathbf{X}_{\{i\}}\right)$
6: $S(t) \leftarrow S(t) \cup i^{(t)}$
7: $\hat{\mathbf{h}} \leftarrow \left(\mathbf{X}_{S(t)}^H \mathbf{X}_{S(t)}\right)^{-1} \mathbf{X}_{S(t)}^H \mathbf{y}$
8: $\hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S(t)} \hat{\mathbf{h}}$
9: $\mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
10: **end for**
11: $\hat{\sigma}_e^2 = \|\mathbf{y} - \hat{\mathbf{y}}^{(n_{\max})}\|_2^2$
12: $n_{opt} \leftarrow \arg\min_{n_c} \left\{2m \ln \hat{\sigma}_e^2 + 2n_c \ln(2m)\right\}$

5.2 Structured Compressed-Sensing for Volterra Series Models

The structural compressed-sensing algorithm presented in this chapter can be considered a particularization of the stagewise orthogonal matching pursuit (StOMP) algorithm, which selects a fixed number of regressors in each iteration based on a threshold. The improvement consists on the inclusion of a priority function that assigns the significance of the coefficient within the model in the subset defined after the thresholding. The new greedy algorithm for pruning Volterra model matrices taking into account the structural information is summarized in Algorithm 1.

The initialization, which corresponds to line 1 of Algorithm 1, consists on the definition of the residual $\mathbf{r}^{(\mathbf{0})} = \mathbf{y}$, that will be used for keeping the remaining part of the output still to be modeled. The support set $S(0)$ is empty in the first iteration, as no regressor is still selected. In each iteration t, the algorithm calculates the correlation between the residual $\mathbf{r}^{(t)}$ and each of the columns of the measurement matrix $\mathbf{X}$ normalized by its ℓ_2-norm (line 3). A first preselection is performed where all the regressors with absolute value of the correlation greater than a fraction $(1-\alpha) \in [0,1]$ of the maximum are included in the subset $\mathbf{i}_{\text{pre}}^{(t)}$, shown in line 4. When the span α is equal to 0, the selection becomes that of the classic OMP, which chooses only the maximum correlation within all the regressors, and if it is equal to 1, no correlation-based sorting is made and only the structural information of the model is evaluated for this arrangement. Then, the regressor with the lowest score given by the priority function $f(\cdot)$ is included onto the support set. Then, the estimation

of the Volterra vector $\hat{\mathbf{h}}$ is obtained by a LS regression and the estimated output of the model $\hat{\mathbf{y}}^{(t)}$ and the residual $\hat{\mathbf{r}}^{(t)}$ are updated (lines 7-9). An example of this function is given in the experimental design section. Finally, when all the regressors are sorted or a fixed maximum of regressors to sort $n_{\max}$ has been reached, a Bayesian Information Criterion (BIC) is applied to obtain the optimum number of Volterra kernels n_{opt}. The model with the lowest BIC is selected according to line 12, where $\hat{\sigma}_e^2$ is the estimation of the error variance and n_c is the number of components.

6 Transmitter Linearization Adaptable to Power-Varying Operation

6.1 Introduction

The evident interest of the current published work about the linearization of wireless communications transmitters reveals the convenience of adequate behavioral models for the design of DPDs. The conventional baseband Volterra models, such as the FV [27], the MP [12], or the GMP [18], are specific for PA, and have insufficient accuracy to represent more general nonlinear systems with complex-valued input signals.

To comply with the objective of linearizing a transmitter, designers address several challenging issues. In particular, the joint compensation of the I/Q-modulator and the PA impairments [41]–[42], and the problem of concurrent dual-band PAs [43] have been subject of concern in the last years. Another important goal is the design of a DPD that is capable of following the changes in the PA operating conditions. This has been performed by updating the DPD parameters in order to follow the power level changes [44]–[45].

Joint compensation of the I/Q modulator impairments and the PA nonlinearities are addressed in [46] by using a structure composed of a MP model dependent on the signal $x(k)$ and a filter, augmented with a similar second branch dependent on the image signal $x^*(k)$. A similar approach is presented in [47] for the case of a multiple-input multiple-output (MIMO) transmitter. The generalized twin-box model [42] incorporates parallel-Hammerstein (PH)-based branches that ignore the out-of-diagonal kernels and may contain non-significant diagonal terms. In these approaches, some terms belonging to the GMP or the FV models, which can be important, are not included.

In [44], a power-scalable DPD based on a two-box architecture with a set of precalculated memoryless look-up table and a MP function to compensate the memory effects is proposed. The need of a power-adaptive DPD to compensate distortion in level-variable conditions was solved in [45] by adjusting the model parameters in accordance with the changes in the

input power level. The procedure introduces extended scaling factors to track the varying signal power with an interpolation approach. The model extraction in both proposals requires PA measurements at multiple power levels and the number of coefficients is fixed independently of the varying PA nonlinear operation.

In this chapter, we focus on a DPD design to linearize the transmitter in a range of output power levels incorporating a model with a complete set of parameters to jointly compensate the I/Q modulator impairments and the PA nonlinearities. The mathematical deduction of the general complex-valued Volterra series (CVS) representation [48] is based on Wirtinger calculus by considering the nonlinear system dependent on the input $x(k)$ and its complex-conjugate $x^*(k)$, operating these variables as real-valued. Then, the transmitter can be viewed as a two-input system and analyzed using a double Volterra series approach. Under the assumption of a Volterra representation, the CVS model is able to describe the nonlinear behavior of a wireless communications transmitter. The present proposal requires signal acquisition at only a single power operating point and the identified model is directly extended to a wide range of power levels. The rise of the number of coefficients in the CVS model makes necessary a suitable procedure to identify the model parameters in an efficient and robust manner, as in [49]–[37], where a technique for model reduction using the sparse structure of Volterra kernels was introduced. A thresholding procedure contributes to the model effectiveness by reducing the number of coefficients as the system enters into weakly nonlinear modes. In [50], the robustness of the procedure is confirmed by verifying that the set of parameters identified at a given power level are applicable to the accurate estimation of the system output over a wide dynamic range. Here, the proposed approach is applied to the design of a DPD to accomplish the joint linearization of transmitter I/Q modulator and PA impairments, under power-varying conditions.

The next sections of this chapter are organized as follows: After this introduction, section 6.2 presents the rationale for the proposed identification method in regression models. First, subsection 6.2.1 describes the framework of the CVS model in the context of general discrete-time complex-valued nonlinear systems. Next, subsection 6.2.2 gives a detailed theoretical justification of the proposed identification procedure and establishes the power-scalable law for the model parameters. Subsection 6.2.3 reviews related works for the joint compensation of I/Q modulator and PA impairments.

6.2 DPD Structure and Identification Procedure

Compensation of transmitter RF impairments, originated predominantly by I/Q imbalance and nonlinearities, is one of the most important challenges for DPD designers. This problem can be approached by following two different points of view. The first approach is based on the knowledge of how the different blocks are assembled inside the transmitter (see figure 6.1a) and an approximate input-output representation is deduced for the whole system. An alternative perspective, where the predistorter is designed considering the transmitter as a black box, is presented below. For comparison, the first approach is discussed at the end of this section.

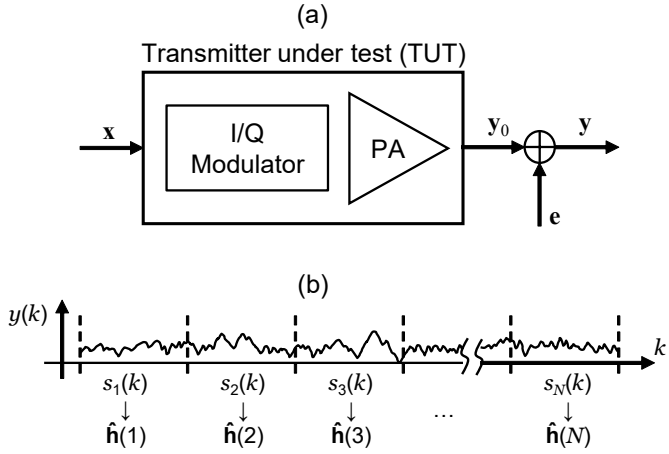

Figure 6.1 Acquisition model with additive noise (a) and signal segmentation (b).

6.2.1 Proposed DPD Model for Joint Mitigation in Transmitters

In the context of Volterra series representation, any given system with complex-valued input can be modeled by the CVS model [48]. Viewed as a black box, the relationship for the input and output complex envelopes, $x(k)$ and $y_0(k)$, of the DPD in a wireless communications system can be expressed as

$$y_0(k) = h_{0,0} + \sum_{n=1}^{\infty} \left\{ \sum_{m=0}^{n} \sum_{\mathbf{q}_n=\mathbf{0}}^{\mathbf{Q}} \sum_{\mathbf{p}_m=\mathbf{0}}^{\mathbf{Q}} h_{n-m,m}(\mathbf{q}_{n-m},\mathbf{p}_m) \times \right. \\ \left. \times \prod_{r=1}^{n-m} x(k-q_r) \prod_{s=1}^{m} x^*(k-p_s) \right\}. \quad (6.1)$$

The vector of delays are $\mathbf{q}_n = [q_1, q_2, \cdots, q_n]^T$ and $\mathbf{p}_m$, defined in a similar way, with a maximum delay Q for all indices. The product of the input signal samples is denoted as $\prod_{r=1}^{n} x(k-q_r) = x(k-q_1)x(k-q_2)\cdots x(k-q_n)$ and the same notation is used for the product of the image samples $x^*(k)$. For $m=0$ and $m=n$, $h_{n,0}(\mathbf{q}_n)$ and $h_{0,n}(\mathbf{p}_n)$ are standard Volterra kernels and the products do not contain $x^*(k-p_s)$ and $x(k-q_r)$, respectively. The factor $h_{n-m,m}(\mathbf{q}_{n-m},\mathbf{p}_m)$ is a Volterra kernel of order n dependent on the two kinds of indices.

The CVS model (6.1) with $h_{0,0}=0$ can be viewed also as a linear combination of the Volterra regressors $\prod_{r=1}^{n-m} x(k-q_r) \prod_{s=1}^{m} x^*(k-p_s)$. The M samples of the input signal can be disposed to form the column vector $\mathbf{x} = [x(0), x(1), \cdots, x(M-1)]^T$ and, if the Volterra regressors are likewise organized to construct the regressor vectors $\boldsymbol{\xi}_i$, the truncated version of (6.1) can be rewritten as

$$\mathbf{y}_0 = \sum_{i=1}^{N_R} h_i \boldsymbol{\xi}_i, \quad (6.2)$$

where the ith regression coefficient h_i comes from the Volterra kernels arranged in an ordered-fashion, $\mathbf{y}_0$ is a vector with the output samples disposed in the same way as $\mathbf{x}$, and N_R is the number of regressor vectors of the model. Examples of Volterra regressor vectors are $\boldsymbol{\xi}_1 = \mathbf{x}$, corresponding to the memoryless linear regressor, and $\boldsymbol{\xi}_{i_0} = [x(-q)|x(-q)|^2, x(1-q)|x(1-q)|^2, \cdots, x(M-1-q)|x(M-1-q)|^2]^T$, corresponding to the third-order memory polynomial regressor with a delay q. The regressor vectors constitute a complete set in the case of the CVS model. Although $\boldsymbol{\xi}_i$ are non-orthogonal, a feasible recursive algorithm to estimate the N_R coefficients of (6.2) is the Orthogonal Matching Pursuit (OMP) [51].

The so-called observation matrix $\mathbf{X}$ and the coefficients vector $\mathbf{h}$ are constructed by stacking all column vectors $\boldsymbol{\xi}_i$ and the model coefficients h_i, respectively. Therefore, (6.2) can be expressed in a compact matrix form [18],[37].

6.2.2 Identification Procedure

In an experimental setup, the vector of the acquired samples, $\mathbf{y}$, is contaminated by equipment distortion and noise. Assuming an approximately distortionless setup, for instance in laboratory conditions with high-performance equipment, the schematic of the acquisition experiment can be modeled as in figure 6.1. Therefore, the input-output relationship is

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}, \tag{6.3}$$

with a zero-mean additive white Gaussian (AWGN) noise $\mathbf{e}$, stationary and complex-valued. The model parameters can be identified by using a direct LS solution, given by

$$\hat{\mathbf{h}} = \left(\mathbf{X}^H \mathbf{X}\right)^{-1} \mathbf{X}^H \mathbf{y} + \mathbf{w}, \tag{6.4}$$

where H represents the Hermitian transpose operation. The difference between the estimated vector $\hat{\mathbf{h}}$ and the true vector $\mathbf{h}$ is originated by the noise process $\mathbf{w}$. Observe that the selection of a model with lacking regressors is another important source of identification error. For example, the FV and the GMP representations, as particular cases of (6.1), are insufficient to model a transmitter with impairments.

The general character of the CVS model provides a complete set of regressors with a high number of coefficients and, at the same time, an associated suitability for a pruning procedure without a presumption on the significant regressors. The application of the Bayesian information criterion (BIC) was proposed in [37, 50, 52] to select the significant parameters. Whereas [52] is based on a simulated annealing algorithm with all the possible variants of the model, the approach in [37] is based on the application of the OMP algorithm to represent the PA output as its projection onto the span of Volterra regressors, and on the BIC rule to discard the irrelevant coefficients, maintaining only the active regressors. It is worth noticing that while LS identification is affected by regressor correlation, the OMP algorithm guarantees the recovery of the exact value of the coefficients in a noiseless environment in a given number of iterations [53].

The alternative procedure applied in [50] is mainly a parameters detection based on statistical hypotheses testing. Referred to figure 6.1b), the entire set of acquired samples is divided in N segments. Focusing on the ith coefficient of the estimated vector $\hat{\mathbf{h}}$, each

segment returns an independent measurement of this particular coefficient. We can define the vector $\tilde{\mathbf{h}}_i$ constructed with the N different realizations of this random variable to decide whether or not the ith coefficient has to be incorporated as an active parameter of the model. We use the Neyman-Pearson (NP) approach to make the decision, based on two hypotheses [54]:

- The measurement is produced by noise,

$$\mathscr{H}_0\text{: if } \tilde{\mathbf{h}}_i = \mathbf{w}_i,$$

- The measurement is produced by the presence of a model coefficient plus noise,

$$\mathscr{H}_1\text{: if } \tilde{\mathbf{h}}_i = \mathbf{h}_i + \mathbf{w}_i.$$

The probability density function under $\mathscr{H}_1$ is

$$p(\tilde{\mathbf{h}}_i;\mathscr{H}_1) = \frac{1}{\pi^N \sigma^{2N}} e^{-\frac{1}{\sigma^2}(\tilde{\mathbf{h}}_i-\mathbf{h}_i)^H(\tilde{\mathbf{h}}_i-\mathbf{h}_i)}, \tag{6.5}$$

where σ^2 is the variance of the complex-valued AWGN $\mathbf{w}_i$. Likewise, the probability density function under $\mathscr{H}_0$ is expressed as

$$p(\tilde{\mathbf{h}}_i;\mathscr{H}_0) = \frac{1}{\pi^N \sigma^{2N}} e^{-\frac{1}{\sigma^2}\tilde{\mathbf{h}}_i^H\tilde{\mathbf{h}}_i}. \tag{6.6}$$

For a particular probability of erroneous measurement induced by noise, the NP theorem states that the probability of true detection is maximized if the detector decides the hypothesis $\mathscr{H}_1$ when the likelihood ratio $L(\tilde{\mathbf{h}}_i)$ exceeds a given threshold γ, i.e.,

$$L(\tilde{\mathbf{h}}_i) = \frac{p(\tilde{\mathbf{h}}_i;\mathscr{H}_1)}{p(\tilde{\mathbf{h}}_i;\mathscr{H}_0)} > \gamma. \tag{6.7}$$

Substituting (6.5) and (6.6) in the likelihood ratio and taking $\ln L(\tilde{\mathbf{h}}_i)$, we obtain

$$\ln L(\tilde{\mathbf{h}}_i) = \frac{2}{\sigma^2}\mathrm{Re}\left(\mathbf{h}_i^H\tilde{\mathbf{h}}_i\right) - \frac{1}{\sigma^2}|\mathbf{h}_i|^2 \tag{6.8}$$

and the decision

$$\mathrm{Re}(\mathbf{h}_i^H\tilde{\mathbf{h}}_i) = \mathrm{Re}\left(\sum_{r=1}^{N} h_i^*(r)\tilde{h}_i(r)\right) > \gamma' \tag{6.9}$$

is equivalent to the likelihood ratio test (6.7). Under steady conditions, any coefficient is an unknown constant $h_i(r) = h_i$, with its estimator being the average value

$$h_i \approx \bar{\tilde{h}}_i = \frac{1}{N}\sum_{r=1}^{N}\tilde{h}_i(r). \tag{6.10}$$

Table 6.1 Summary of the Proposed Identification Procedure

1: Divide the input $x(k)$ and measured output $y(k)$ signal into N segments, each with M samples.
2: Estimate the parameter vectors $\hat{\mathbf{h}}(r)$, $r = 1,\ldots,N$ using the OMP algorithm with the same pre-defined N_R number of components for all models.
3: For all the CVS components, compute the average coefficients $\bar{\tilde{h}}_i$.
4: Sweep the thresholding level to apply the hypothesis test to all coefficients.
5: Take the thresholding level that optimizes the BIC rule.

Operating with (6.9), we can decide $\mathscr{H}_1$ if

$$\mathrm{Re}\left(\bar{\tilde{h}}_i^* \frac{1}{N}\sum_{r=1}^{N}\tilde{h}_i(r)\right) = \mathrm{Re}\left(\bar{\tilde{h}}_i^*\bar{\tilde{h}}_i\right) = |\bar{\tilde{h}}_i|^2 > \frac{\gamma'}{N}. \tag{6.11}$$

There is a trade off between the deficit of model coefficients (if the threshold is high), and the inclusion of undesired noise (if the threshold is chosen too low).

The proposed procedure is as follows. Once the set of coefficients has been estimated by the OMP algorithm for each segment of the acquired signal (see figure 6.1b), the average $\bar{\tilde{h}}_i$ and the hypothesis test (6.11) are computed for all coefficients. The lower the threshold level, the more coefficients (regressors) will be incorporated to the model and the BIC criterion is used to decide the optimum number of regressors n_c [37]. Due to the sparse character of the Volterra kernels, many coefficients can be discarded without a significant loss in accuracy.

Here the BIC rule with its explicit dependence on the NMSE. If the NMSE is expressed in dB, the variance is given by

$$\hat{\sigma}_e^2 = \left(\frac{1}{M}\sum_{m=0}^{M-1}|y(m)|^2\right) \times 10^{\mathrm{N}MSE/10} \tag{6.12}$$

and substituting in (23) of [37], the BIC rule becomes

$$n_{c0} = \arg\min_{n_c}\left\{\mathrm{N}MSE + \frac{n_c}{M}10\log(2M)\right\}. \tag{6.13}$$

This procedure, summarized in Table 6.1, and that published in [37], start with the OMP, a method that estimates the set of coefficients by iteratively adding new components to the coefficients vector. The difference is that in this work the coefficients computed with several segments of the signal are averaged and a further thresholding step is implemented to get the sparse model structure.

According to the results in [50], the normalized coefficients originated in the nonlinearities of the I/Q branches are not dependent on level variations at the modulator output and the coefficients associated to the FV regressors are normalized according to the PA input level [37]. For example, if $h^{(n)}\Big|_{P_{i0}}$ is a FV nth-order normalized coefficient (n odd) for an

input signal with a power level P_{i0} and this level changes to $P = A^2 P_{i0}$ ($A > 0$, real-valued), the corresponding normalized coefficients follow the relation

$$\frac{h^{(n)}\Big|_P}{h^{(n)}\Big|_{P_{i0}}} = \left(\frac{1}{A}\right)^n = \left(\frac{P_{i0}}{P}\right)^{n/2}. \tag{6.14}$$

Once the normalized parameters of the DPD have been computed at a given input level, P_{i0}, they can be straightforwardly scaled to adapt the coefficients to other level P. Odd-order normalized parameters shared with the FV model follow an exponential scaling with the average input power, so that the magnitudes at P_{i0} and P are related as:

$$h^{(n)}\Big|_P = h^{(n)}\Big|_{P_{i0}} \cdot 10^{-\frac{n \cdot \Delta P(\mathrm{dB})}{20}}, \tag{6.15}$$

where $\Delta P(\mathrm{dB}) = P(\mathrm{dBm}) - P_{i0}(\mathrm{dBm})$ and n is the coefficient order. Those normalized coefficients not included in the FV model (e.g., the image or the even-order parameters associated with impairments of the I/Q modulator), are not dependent on the power level. In that case, the way the DPD is adapted to a decrement in power level is by excluding parameters with values below a given threshold, and not recalculating the coefficients [44, 45].

6.2.3 Preceding DPD Models for Joint Mitigation

In the previous procedure, the transmitter is viewed as a black box and the CVS model is adopted for the DPD. Another perspective is to assume the internal architecture of the transmitter (figure 6.1a) and deduce the DPD structure to join the mitigation of PA and I/Q-modulator impairments under reasonable approximations [41]-[47]. In [41], the rationale is to compensate the impairments in the reverse order that they appear. Based on the knowledge of the ensemble arranged by the modulator and the PA, a first MP structure is proposed for the PA DPD and then the modulator compensation is implemented with a widely linear (WL) model. This point of view is further extended to a general FV model [46] and to the augmented complex conjugate (ACC) model for joint mitigation of distortion in MIMO transmitters [47]. The output for the ACC model in a single-input single-output (SISO) transmitter is written as

$$y(k) = \sum_{n=1}^{N}{}' \bar{H}_n\{x(k)\} + \sum_{n=1}^{N}{}' \bar{H}_{cn}\{x^*(k)\} + h_0, \tag{6.16}$$

where the prima symbols in the sums indicate that only odd orders are considered,

$$\bar{H}_n\{x\} = \sum_{q=0}^{Q} h_n(q)|x(k-q)|^{n-1}x(k-q) \tag{6.17}$$

and $\bar{H}_{cn}\{x^*(k)\}$ is defined in the same form.

The two viewpoints discussed above have strengths and weaknesses. When the transmitter composition is exploited, the principal advantage is the achievement of a simpler *a priori* structure for the DPD [47]. This favourable feature is based on the assumption of the particular modulator model, in this case a WL transformation. The WL supposition does not consider the contribution of the nonlinearities in the modulator baseband [55]-[56], but it is a good approximation in cases where the modulator nonlinearity can be neglected. On the other hand, making no assumptions about the internal architecture of the transmitter, the CVS model gives a complete (and huge) set of regressors. However, many of them are negligible because of the inherent sparse characteristics of the DPD. If the CVS representation is complemented with the compressed-sensing technique described in section 6.2.2, the result is a reduced model with only the indispensable number of active coefficients.

7 Results

This chapter covers the experimental design and results of the theory shown in the previous chapters.

7.1 Volterra-Parafac Digital Predistorters

As a proof of concept for the proposed DPD structures, they have been validated experimentally with a commercial PA. The measurement setup is described in 2.4.1.

In order to tradeoff the available bandwidth at the VSG with the occupied bandwidth due to the PA nonlinearity, and ensure an accurate representation of the spectral regrowth at the PA output, the test signal was a 15-MHz OFDM signal generated according to the LTE-downlink standard. A sequence of 16-QAM symbols were assigned to all the active subcarriers and an oversampling rate $ovs = 6$ was used to generate the OFDM signal characterized by a PAPR about 12.0 dB. Acquisition with the signal analyzer was managed by Agilent VSA-89600 software, operating in the raw acquisition mode.

The device under test was the evaluation board of a commercial PA based on the CGH40010 GaN HMET device, from Cree Inc., at a frequency of 3.6 GHz and biased with a quiescent drain-to-source current about 200 mA. Driving the test PA in the range of output levels from $+25$ to $+30$ dBm ensured a nonlinear operating condition characterized by ACPR levels in the range of -40 to -36 dBc, exceeding the requirements of unwanted emissions for the LTE standard, and in-band distortions characterized by EVM values in the range of 3–4%.

7.1.1 Measured Performance for the B-VP DPD

A conventional indirect learning approach was followed for the identification of the inverse PA functions by running the frequency domain version of the adaptive CLMS algorithm over a pair of uncompensated output and input 15-MHz LTE training datasets at an average output level of $+32.4$ dBm. The dc current level was 240.8 mA. The target gain of the linearized PA was carefully selected about 15.4 dB to account for the PAPR of the test signal. An attenuation of about 2.7 dB in the DPD ensured that the peak-envelope-power

(PEP) at the PA input did not exceed the maximum range of the identification dataset, yielding an output level of $+30$ dBm and a dc current about 203 mA. This shift in the actual working operation point with respect to the training level introduced imperfections in the identification process that were removed by repeating the identification procedure with predistorted and compensated datasets in an iterative fashion. Two or three iterations were enough to optimize the DPD operation.

A thirteenth-order B-VP DPD with a block size $L = 1024$ sa and rank $R = 1$ was tested to validate the proposed structure. In order to ease the convergence of the CLMS identification algorithm, the initial set of equalizing filters, $\mathbf{W}_{n,0}^{(r)}$, were derived from a memoryless structure of the same order, identified by performing a LS estimate from the training datasets. The running complexity of the proposed model was evaluated according to (3.18), yielding 100 FLOPs/sa.

To compare the results obtained with the proposed approach, a thirteenth-order GMP predistorter was designed and tested. The memory of the GMP DPD was configured so that the linearization results were similar to those obtained with the B-VP DPD. According to [18], a model with no even-order terms, no leading-envelope terms, diagonal terms, and non-diagonal terms arranged with $\mathscr{L}_b = \{0\}$ and $\mathscr{M}_b = \{1,\ldots,Q\}$ for the signal and lagging envelope index arrays, respectively, was chosen. A maximum delay of $Q = 15$ samples was introduced for orders one to seven in the diagonal part, and for orders one to five in the non-diagonal part. The rest of the kernels were assumed memoryless. The running complexity of this GMP-DPD implementation was adapted from (16) to (18) in [7], yielding 828 FLOPs/sa.

The AM/AM and AM/PM characteristics of the uncompensated and linearized PA are plotted in figure 7.1. It is revealed that the slope of the AM/AM curve is reduced under DPD operation, what is related to the gain compression referred above. On the other hand, there is an evident reduction of dispersion in the low input-level range of the AM/PM characteristic compared to the uncompensated PA, which is a sign of mitigation of memory effects in the predistorted PA.

The efficacy of both DPDs to remove spectral regrowth is shown in figure 7.2. The uncompensated case corresponds to a test signal driving the PA with the same average output level than the linearized PA. The ACPR data associated with these spectra are summarized in Table 7.1, which reveal that the B-VP DPD performs quite similar than GMP DPD, with the latter gaining a reduction of about 2 dB in the first adjacent channels. However, this improvement is accomplished at the cost of a computational complexity that, in this case, is increased by eight times.

The same conclusions can be drawn for in-band linearization performance, which can be evaluated from figure 7.2 with the spectra of the error signal. Notice that there is only a slight difference between the B-VP and GMP cases at the edges of the channel band. Accordingly, this difference does not affect to the constellation diagram, as shown in figure 7.3, as the scatter plots for both DPDs are indistinguishable. This result is also supported by the negligible difference of EVM (see Table 7.1), with both methods accomplishing a reduction of the EVM from the uncompensated value of 3.9% to a linearized value of 0.4%. In terms of the NMSE between the output of the predistorted PA and the reference signal, the values reported in Table 7.1 reveal that the B-VP DPD is

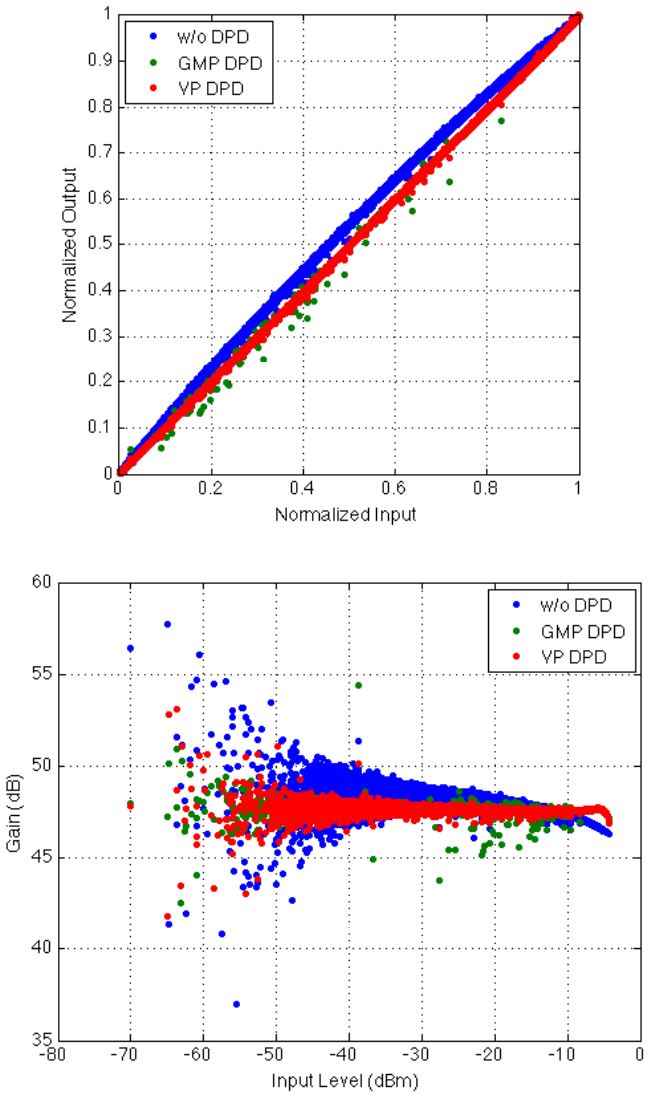

Figure 7.1 AM/AM (a) and AM/PM (b) characteristics of the unpredistorted and linearized GaN PA driven by a 15-MHz LTE signal.

only 0.6 dB above the GMP DPD, which is also related to the fact that the NMSE is more influenced by the in-band rather than the out-of-band performance.

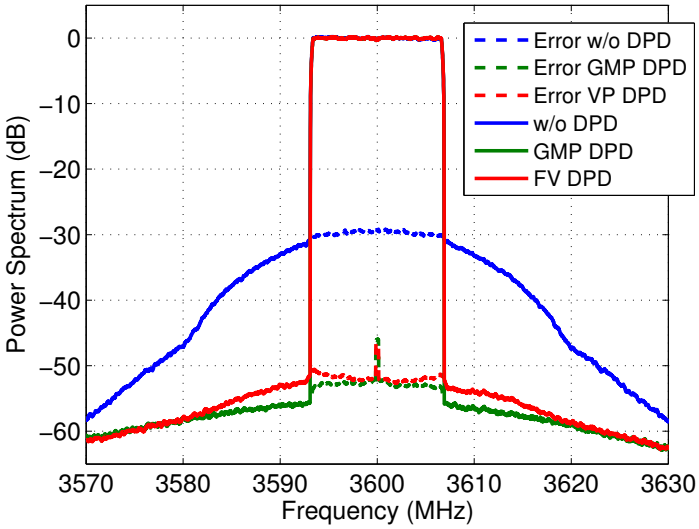

Figure 7.2 Normalized power spectral density of the GaN PA at an average output level of +30 dBm without DPD (blue), employing the GMP-based DPD (green) and the proposed B-VP DPD (red). Dashed lines correspond to the error spectrum.

Table 7.1 DPDs performance comparison in terms of out-of-band emissions in the first and second adjacent channels with the GaN PA at $P_{out} = +30$ dBm

Cases under study	ACPR (dBc)			
	-30 MHz	-15 MHz	+15 MHz	+30 MHz
W/o DPD	-55.60	-36.41	-36.55	-55.80
GMP DPD	-59.80	-56.99	-57.59	-60.60
VP DPD	-59.90	-55.08	-55.68	-60.60

7.1.2 Measured Performance for the FD-VP DPD

In this case, the predistorter was configured with seventh-order, rank $R = 1$, and 1024-points IFFTs. Signal generation was arranged starting with 16-QAM symbols mapped onto 900 active subcarriers to build the 15-MHz LTE signal. A raised-cosine shaping filter was applied in the linear branch of the DPD with a roll-off factor of 0.1 and a delay of 96 sa, while the length of anti-aliasing FIR filter for the rest of the branches was $M = 10$. The signal level was configured to drive the PA in a nonlinear operating condition characterized by an output level of +26.2 dBm. The predistorter signal generation was performed off-line, and relied on a behavioral model of the PA, a seventh-order GMP model with a maximum delay of $Q = 15$ sa, which was identified by applying standard LS to the input and output LTE waveforms at the operating frequency. The coefficients of the equalizing

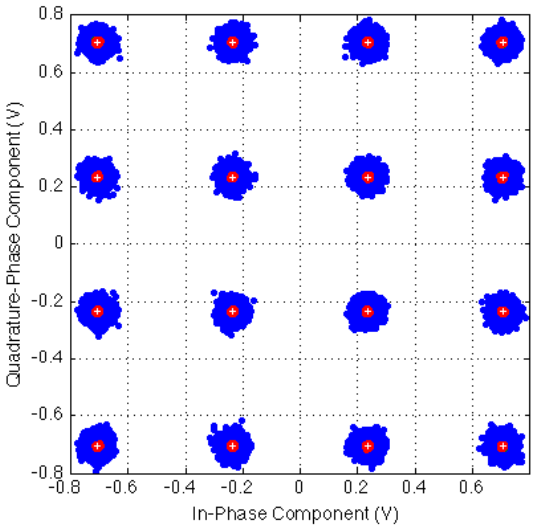

Figure 7.3 Transmitted constellation of the GaN PA associated with the 15-MHz LTE signal at an average output level of +30 dBm without DPD (blue), employing the GMP-based DPD (green) and the proposed B-VP DPD (red). The + marks correspond to the reference constellation for the transmitted symbols.

Table 7.2 DPDs performance comparison for LTE-downlink signals with the GaN PA at $P_{out} = +30$ dBm

Cases under study	EVM (%)	NMSE (dB)	Computational cost (FLOPs/sa)
W/o DPD	3.89	-27.68	–
GMP DPD	0.36	-48.30	828
VP DPD	0.40	-47.69	100

filters $\mathbf{W}_{n,k}^{(r)}$ were set by running the adaptive block-CLMS algorithm (14) with a block length $L = 1024$. To compute the frequency-domain error (15), a compressed value of the target gain G_c was chosen to ensure that the PA model range was not exceeded when the DPD is driven by the input signal. These settings were applied to generate the predistorted data corresponding to a validation LTE test signal comprising 56 OFDM symbols, which were sequentially passed through the structure depicted in figure 3.3 and applied to the PA input.

As in the B-VP type, the proposed FD-VP DPD has been compared with a conventional time-domain DPD, in this case a seventh-order GMP predistorter selected to optimize the error and configured with maximum delay of $Q = 15$ sa in a way similar to that compared to the B-VP DPD. Table 7.3 illustrates that the proposed FD-VP DPD achieves

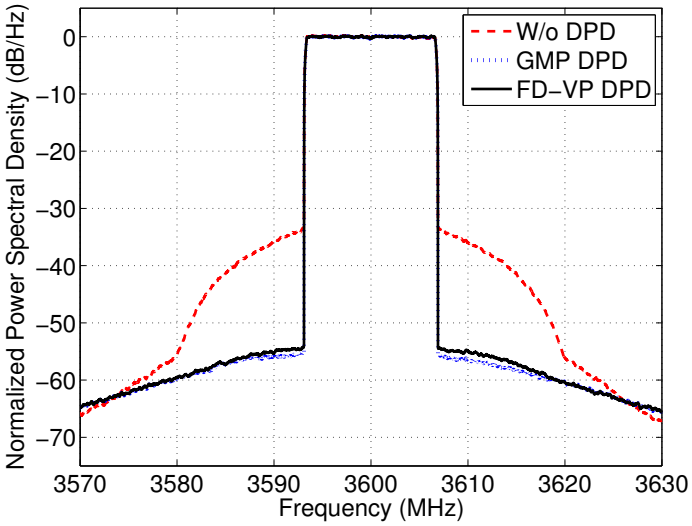

Figure 7.4 Measured power spectral density of the test PA with different DPD linearizers. The unpredistorted PA was driven at an input level to yield the same output power than the linearized PA (+26.2 dBm).

a similar performance to the DPD based on the GMP model, while significantly reducing the computational cost per sample associated to the generation of the predistorted signal. The results of the ACPR confirm that both the FD-VP DPD and the GMP DPD provide an improvement over 17 dB with respect to the nonlinear signal at the same average output power level. Despite the fact that the spectral regrowth of the un-predistorted PA signal failed to comply with the requirements of the LTE-downlink standard, the ACPR requirements were fulfilled after applying both DPDs. Figure 7.4 shows the power spectral density of the LTE signal with and without DPD, demonstrating the reduction of the spectral regrowth and the similarity between the results obtained by both DPDs. It is worth mentioning that these results were achieved by deriving the DPD settings from the aforementioned adaptive algorithm, but using a single estimate of PA model. Although both approaches led to similar values, there is room for improvement in the reduction of the ACPR by using more accurate estimates of the PA model in an iterative fashion.

In-band distortion mitigation of the DPDs can be assessed in terms of the EVM. Table 7.3 reveals that the proposed DPD reduces the EVM from 3.0% to 0.6%, while the GMP DPD produces a higher dispersion of the received symbols around the reference constellation points, with an associated EVM of 1.4%. The more notable mitigation of the in-band distortion for the proposed DPD can be observed in figure 7.5, where one symbol of the 16-QAM constellation of the received LTE signal has been zoomed in. Finally, the NMSE between the sampled signal in the receiver and the theoretical reference is calculated to evaluate the overall linearization achieved. As it can be observed in Table 7.3, the residual error of the LTE signal significantly decreases with the proposed DPD, reducing the NMSE

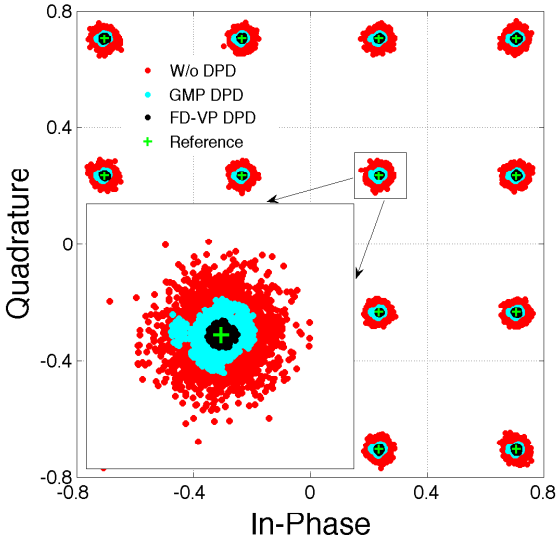

Figure 7.5 16-QAM constellation of a received 15-MHz LTE-downlink signal without DPD (red dots), with a GMP DPD (blue dots), and with the proposed FD-VP DPD (black dots). The reference constellation point is plotted with a green '+' mark. One symbol has been zoomed in.

Table 7.3 DPD performance comparison for a 15-MHz LTE-downlink signal (P_{out} = +26.2 dBm)

Cases under study	NMSE (dB)	ACPR (dBc)		EVM (%)	Computational cost (FLOPs/sa)
		Lower	Upper		
W/o DPD	−30.2	−39.5	−39.6	3.0	–
GMP DPD	−41.7	−57.2	−58.0	1.4	939
FD-VP DPD	−44.5	−56.6	−56.9	0.6	255

of the nonlinear signal by more than 14 dB.

Regarding the complexity of the DPD data generation, the cost per sample associated with (3.19) is 255 FLOPs under the aforementioned settings. This figure can be fairly compared with the particular implementation of this GMP DPD, as shown in Table 7.3: it would require 939 FLOPs per sample.

7.2 Complex-valued Volterra Series Model

An experimental study based on the testbench introduced in section 2.4.1 has been conducted. The VSG served as an I/Q modulator, providing a flexible interface to introduce

Table 7.4 Modeling performance in terms of the NMSE and ACEPR for the test PA with impairments in the I/Q modulator

Model	ACEPR (dBc)					NMSE	#
	−30 MHz	−15 MHz	In-Band	+15 MHz	+30 MHz	(dB)	Coeff.
FV-CS	−62.8	−54.7	−31.7	−54.6	−66.4	−31.6	16
WNL-CS	−68.8	−65.2	−54.5	−65.1	−68.2	−52.7	129

I/Q impairments. The VSG output was fed into a commercial PA constructed with the CGH40010 GaN HEMT transistor, from Cree Inc. The test signal was a 15-MHz orthogonal OFDM signal generated at 3.6 GHz according to the Long Term Evolution (LTE) downlink standard. In the receiver side, the RF signal was measured at the VSA with a sampling rate of 92.16 MS/s providing over 360000 samples. The test signal was composed of 56 OFDM symbols and model identification was performed with only one symbol.

Following the usual methodology for behavioral modeling of transmitter architectures, model parameters were identified by applying a conventional LS procedure to the input-output measurement datasets, providing the NMSE between the modeled and measured signal as a quality metric, in addition to the Adjacent Channel Error Power Ratio (ACEPR), the latter evaluated for both the upper and lower adjacent channels.

Modeling performance of the proposed WNL approach was compared to the FV, given its demonstrated accuracy in the context of PAs. Considering the PA operation point, both models were configured with thirteenth-order, a memory length $Q = 3$ for orders one to five, and memoryless kernels for higher orders. Due to the relatively large number of coefficients in the general proposal (248 for the FV model and 1370 for the WNL model), the compressed-sensing algorithm [37] was applied to select the appropriate kernels of both models.

According to Table 7.4, under I/Q modulator impairments, the FV compressed-sensing (FV-CS) model reduces the number of coefficients to 16 and degrades the NMSE to −31.6 dB, while the pruned WNL-CS model consists of 129 coefficients and presents an NMSE of −52.7 dB that outperforms the former in about 20 dB. While the number of the WNL-CS coefficients can be high, an increase of the FV-CS coefficients leads to overfitting because the appropriate regressors are missing. This behavior is confirmed by the predicted spectra plotted in figure 7.6, showing that the model tightly matches the measured signal, while the FV-CS model fails in the description of spectral regrowth in the adjacent channels. These outcomes are even more evident in figure 7.7, where the spectra of the error are depicted. As a reference we include in this figure the results for the test PA without I/Q modulator impairments (dash-dot line), showing a very low spectrum error inside the signal band for the FV-CS model. Let us clarify that impairments were removed from the I/Q modulator output by using a precompensator based on [32], and therefore this case resembles a nonlinear system for which the baseband PA model holds. However, when I/Q impairments are present, the FV model cannot describe accurately the output and the in-band error degrades by almost 30 dB. On the contrary, the WNL preserves the performance in that case. The spectrum of the error predicted by the WNL-CS under compensated I/Q impairments, not shown, is almost the same as the FV-CS model.

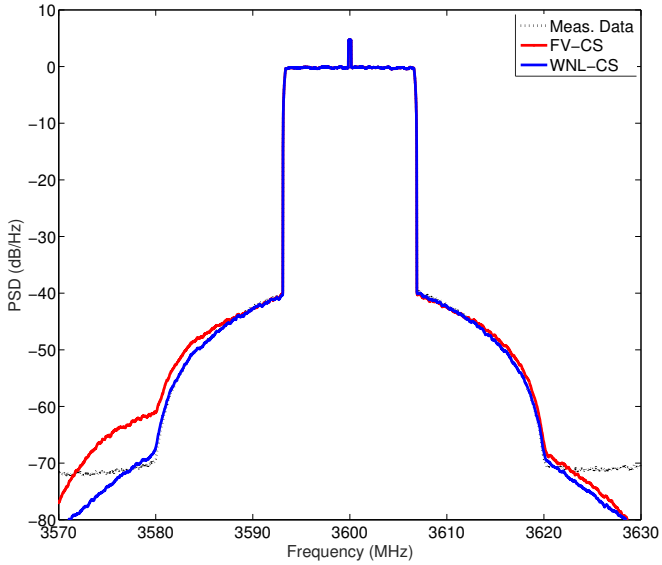

Figure 7.6 Normalized power spectral density of the measured (dotted line) and modeled (solid lines) output of the PA driven at 3.6 GHz with an impaired I/Q modulator.

7.3 Structured Volterra Series Model

A set of measurements were acquired in order to validate the predistortion capabilities of the algorithm. The weblab introduced in section 2.4.2 was used. The measurements were taken at an output power of +35.2 dBm. The test signal is compound of 56 OFDM symbols of a 15-MHz OFDM signal generated from 16-QAM symbols modulated onto 900 subcarriers and filtered by a raised-cosine with roll-off factor of 0.1, according to the LTE-downlink standard. This signal exhibits a PAPR of about 11 dB and a hard clipping to the 7 samples with the highest absolute value was applied to reduce the PAPR to 10 dB. An oversampling of 1 to 6 applied to the original signal results in a sampling frequency of 92.16 MHz which was also used to record the measurements. The test signal contains over 360000 samples.

A GMP model was selected to test the algorithm. This model has the structure shown in (7.1), where a configuration of seventeenth order and a maximum distance from the diagonal of 10 was set. This corresponds to $K_a = K_b = K_c = 16$, $L_a = 10$ and $L_b = M_b =$

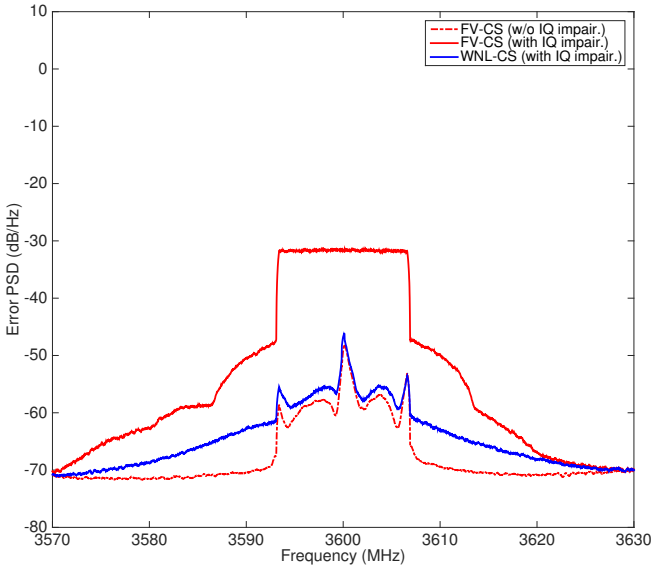

Figure 7.7 Spectrum of the error signal between the measured and modeled output of the PA driven at 3.6 GHz with an impaired I/Q modulator (solid lines) and an I/Q modulator with precompensated impairments (dot-dashed line).

$L_c = M_c = 5$. The resulting Volterra model contains 1087 components.

$$\begin{aligned} y(k) = & \sum_{p=0}^{K_a} \sum_{l=0}^{L_a} a_{pl} x(k-l)|x(k-l)|^p + \\ & + \sum_{p=1}^{K_b} \sum_{l=0}^{L_b} \sum_{m=1}^{M_b} b_{plm} x(k-l)|x(k-l-m)|^p + \\ & + \sum_{p=1}^{K_c} \sum_{l=0}^{L_c} \sum_{m=1}^{M_c} c_{plm} x(k-l)|x(k-l+m)|^p . \end{aligned} \tag{7.1}$$

The thresholding function designed for this experiment was $f(l,m,p) = |l \mp m| + (p+1)$ according to the form of the GMP regressors of $x(k-l)|x(k-l \pm m)|^p$. This function assigns a higher score —which corresponds to less priority— to high orders and lags, considering a memory fading behavior [57]. The number of maximum components to be considered $n_{\max}$ was set to 200.

The predistorter was calculated through the indirect learning scheme [10]. The 30% of the signal with the highest maximum value at the output of the PA was used for modeling and the span value from 0 to 1 with an increment of 0.2 was swept. Once the predistorter was identified, the validation was carried out with the complete signal and the predistorted signal was obtained placing the previous PA input signal at the input of the DPD. The

output of the DPD was then sent to the web platform and the performance parameters of NMSE, ACPR and EVM were measured in the returned signal.

Table 7.5 Performance results of DPD in a sweep of span α values.

case		ACPR-2 (dBc)	ACPR-1 (dBc)	ACPR+1 (dBc)	ACPR+2 (dBc)	EVM (%)	NMSE (dB)	# Coeff.
w/o DPD		-38.04	-32.82	-32.93	-38.00	5.00	-25.52	-
α	0.00	-39.76	-36.83	-36.82	-39.80	2.82	-30.46	200
	0.20	-42.57	-40.46	-40.51	-42.75	1.89	-33.90	198
	0.40	-44.15	-41.35	-41.29	-44.17	1.78	-34.49	190
	0.60	-44.65	-42.92	-42.91	-44.74	1.46	-36.18	148
	0.80	-50.04	-48.58	-48.71	-50.24	0.80	-41.52	17
	1.00	-55.09	-53.88	-54.13	-55.37	0.52	-45.62	194

The performance parameters of the DPD are presented in Table 7.5. The ACPR, EVM and NMSE values experience a decrease with the span, which indicates that the introduction of the span enhances the predistortion capabilities of the model. The signal without predistortion is characterized by a first lower and upper ACPRs of -32.82 dBc and -32.93 dBc. The predistortion enhances this values from about 4 dB for $\alpha = 0$ to 21 dB in the case of $\alpha = 1$, where the minimum values of -53.88 dBc and -54.13 dBc are reached. Similar improvements are achieved for the second lower and upper ACPRs, where the minimum is also obtained in the same case and take the values of -55.09 dBc and -55.37 dBc, corresponding to an improvement of about 17 dB with respect to the case without DPD. These values are in accordance with figure 7.8, where the power spectral density (PSD) of the predistorted signal for the best case is compared to the case without DPD. The reduction of the spectral regrowth achieved by the DPD agrees with the PSD of the error signal, which shows a low in-band and out-of-band error densities. The EVM is decreased from a 5 % to a 2.8 % for $\alpha = 0$ and a 0.52 % in the case of $\alpha = 1$. This enhancement is clear in the constellation shown in figure 7.9, where the predistorted case shows a reduction in dispersion compared to the signal without predistorsion. The NMSE is reduced from -25.52 dBm without predistortion to -30.45 dB for the lowest value of the span and to -45.62 dB for $\alpha = 1$. Finally, the number of components in the pruned model results in roughly the same value for all the range of the span, which indicates that the pruning capabilities of the algorithm remains at the same performance level.

7.4 Transmitter Linearization Adaptable to Power-Varying Operation

To illustrate the proposed method, the joint compensation and linearization of two transmitters, with adaptable capability in a wide dynamic range, is demonstrated.

7.4.1 Linearization of a Basic Transmitter

The first case of study is the basic transmitter modeled in [50], referred to as TUT-1. The TUT-1 was arranged with the I/Q modulator integrated in the commercial generator

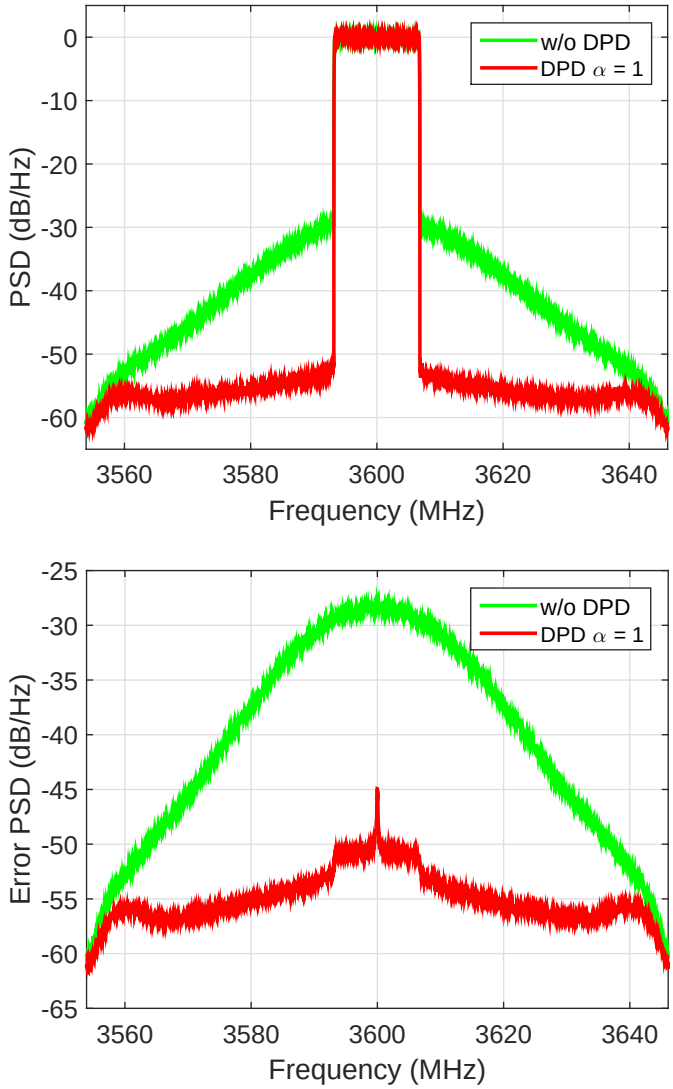

Figure 7.8 Spectral densities for one realization of the experiment for the cases without DPD and with DPD in the case of $\alpha = 1$.

SMU200A of Rohde & Schwarz, and a PA based on Cree's board for the evaluation of the power GaN HEMT CGH40010, operated at a carrier frequency of 3.6 GHz. Later on a DPD linearizer for a more realistic transmitter (TUT-2) will be tested.

The probing signal was designed with an OFDM format and 15-MHz bandwidth, according to the LTE-downlink standard. The input vector $\mathbf{x}$ containing over 300000 samples of the complex envelope was uploaded into the generator and transmitted with

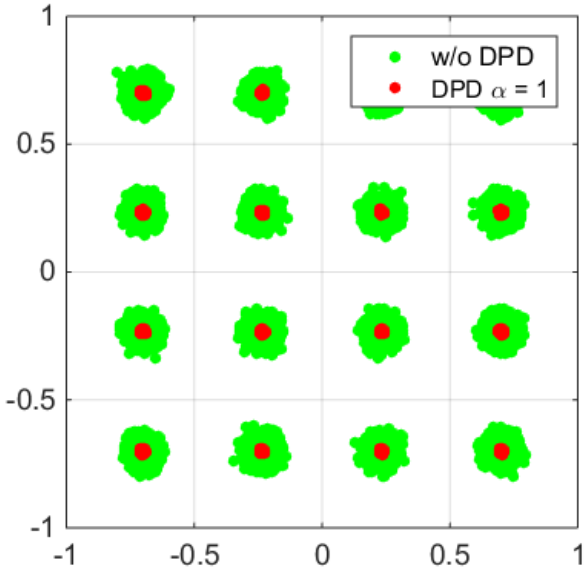

Figure 7.9 15-MHz bandwidth signal constellation for the original signal without DPD and with DPD in the case of $\alpha = 1$.

a peak power of about +30 dBm, corresponding to a measured output average level of +19 dBm. Since the peak-to-average power ratio (PAPR) level is 11 dB, the upper limit to the transmitter power is fixed by the maximum modulator peak level, about +15 dBm (+6 dBm of average power). The samples of the output signal were gathered by averaging 300 acquisitions in the vector signal analyzer. Driven with this signal, the TUT-1 serves as a first proof of concept of the proposed linearization approach. To demonstrate the DPD capability to compensate the nonlinerities as well as the I/Q impairments, the settings of the modulator were configured with a quadrature error of 1 deg. Whereas commercial I/Q modulators show quadrature errors in the range of 1–3 deg [58]–[59], the choice of 1 deg was intended to show to which extent a reduced error can have a significant impact on the linearization capability of a DPD. The acquired baseband samples were subsequently off-line postprocessed with Matlab and used in a conventional indirect learning architecture to obtain the DPD coefficients at a maximum average output power of $+20.8$ dBm ($+31.8$ dBm of peak output power).

The model parameters were chosen to reduce the error between $y(k)/G$, the measured output scaled with the target gain of the linearized TUT-1, and $x(k)$. The underlying CVS structure was configured with thirteenth order, a maximum delay $Q = 3$ for orders 1 to 5, and memoryless (ML) for the higher orders, comprising 1370 coefficients. The OMP algorithm was executed to provide the ordering of 200 coefficients per segment. The synthesized DPD was composed of 22 normalized coefficients above a selected -68 dB

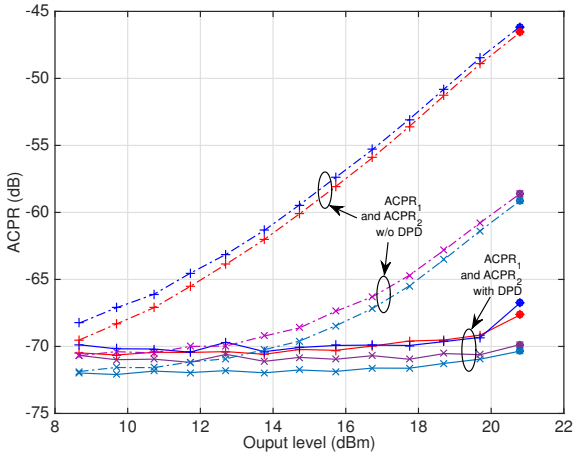

Figure 7.10 Adjacent channel power ratios of the output spectrum with (solid lines) and without (dashed lines) DPD. First adjacent channel (+ marks) and second adjacent channel (× marks). Basic transmitter TUT-1.

threshold. The results of the linearization produced by the DPD at the output power of $+20.8$ dBm are plotted with filled circles in figure 7.10, where an ACPR of about -66 dB in the first adjacent channel is satisfactorily compared to the value of -46 dB obtained with the non-linearized transmitter. A reduction of about 10 dB in ACPR is also observed in the second adjacent channel. Likewise, about 6 dB of NMSE reduction was achieved with the TUT-1 driven at the maximum level, as it is shown with filled marks in figure 7.11.

The normalized coefficients of the DPD were extended at other power levels following the explained procedure, and those below the threshold were discarded. A second series of measurements was then performed with the new predistorted signals calculated according to the model adjusted at the different output levels, yielding the ACPR displayed with solid lines in figure 7.10. To have a reference of the DPD performance, the ACPRs of the transmitter without DPD are also shown (dashed lines), demonstrating a reduction over 21 dB in the first adjacent channel at $P_o = +20.8$ dBm. The DPD performs in all the dynamic range, demonstrating values of ACPR_1 and ACPR_2 better than -65 dB and -70 dB, respectively, without any further adjustment of the coefficients. This capability is also observed in the results of the NMSE displayed in figure 7.11 for the transmitter without DPD, in dashed line, and with DPD, in solid line. NMSE levels of about -41 dB in a wide range of operating powers demonstrate a flexible functioning of the linearizer.

The results of the DPD performance for the TUT-1 were encouraging and led to apply the procedure to a new transmitter where the power capability of the output device was exploited thoroughly.

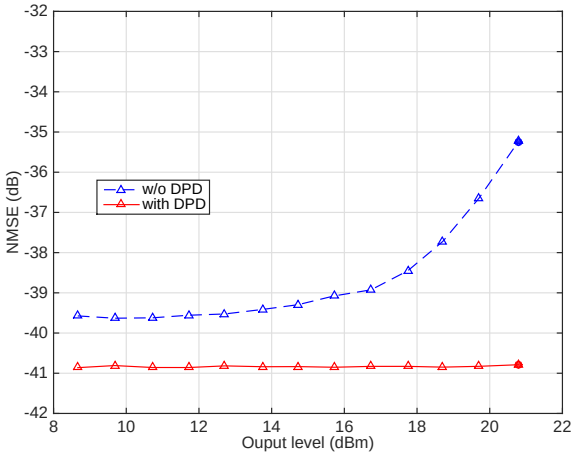

Figure 7.11 NMSE of the output signal without DPD (dashed line) and with DPD (solid line). Basic transmitter TUT-1.

7.4.2 Linearization of a Realistic Transmitter

The objective of this subsection is the joint compensation and linearization of a transmitter operating near its maximum level of +40 dBm. In this transmitter, referred to as TUT-2, a ZHL42W preamplifier of MiniCircuits is connected at the input of the Cree's evaluation board to drive it to a higher nonlinear operating point and delivering a maximum average output power of +26 dBm (+37 dBm of peak power), with a small-signal chain gain of about 46.7 dB. Again, 1 deg of quadrature error was considered for the I/Q modulator. The AM/AM and AM/PM characteristics are plotted in figure 7.12, revealing a gain compression of about 1.6 dB. The output spectrum is shown in figure 7.13, where we observe an important spectral regrowth that fails to comply with the standard ACPR of −45 dB. Let us also remark that not only the PA is contributing to the nonlinear distortion, but the I/Q modulator is also driving the preamplifier beyond its linear operation level. Again, a thirteenth-order CVS structure with 1370 coefficients was pruned. After repeating the procedure described in section 7.4.1, the corresponding DPD was designed. The linearization capability is also shown in the same figure, demonstrating a spectral regrowth reduction of about 20 dB in the first adjacent channel. Although a second iteration will usually provide better results, this initial design is adequate enough for the objectives of the present experiment.

The normalized magnitudes of the 23 most relevant coefficients at $P_o = +26$ dBm ($P_i = -20$ dBm) are plotted with large marks in figure 7.14 and, in agreement with our previous discussion, the values adapted to other levels are displayed with straight lines. Notice that all the values are referred to the ML linear coefficient and that the number of parameters can be reduced as the PA is operated at lower levels. Since the precision requirements allow neglecting the normalized coefficients below a −50 dB threshold, the model at a PA input of $P_i = -20$ dBm is composed of only 14 coefficients, a value that

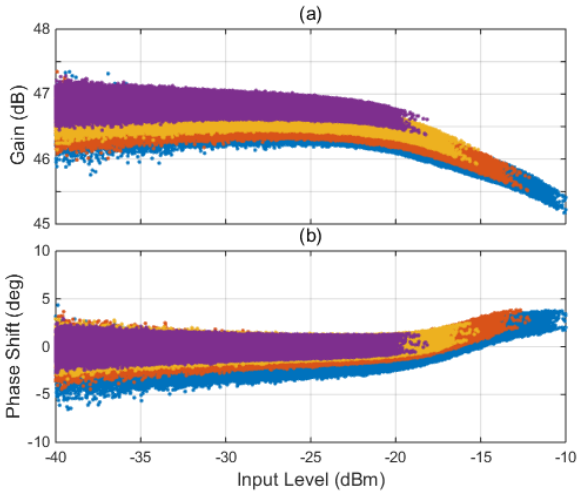

Figure 7.12 Compression curves of the TUT-2 transmitter for average output levels of +26 dBm (blue), +23.4 dBm (red), 20.6 dBm (orange), and +17.8 dBm (magenta). Gain compression (a) and AM/PM (b) characteristics.

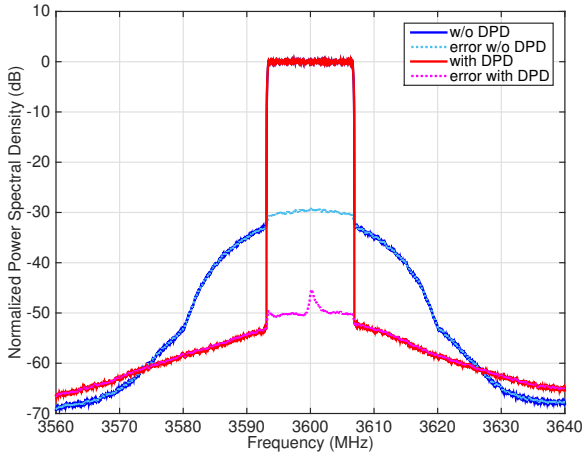

Figure 7.13 Normalized output spectrum and spectrum of the error for an average output level of $P_o = +26$ dBm (+37 dBm, peak), with and without DPD. Realistic transmitter TUT-2.

is reduced to 8 coefficients at $P_i = -30$ dBm. Next, a set of predistorted signals were calculated according to the extended parameters.

In order to evaluate the performance of the DPD, the ACPR of the linearized signals are displayed in figure 7.15 with + and × marks for the first and second adjacent channels,

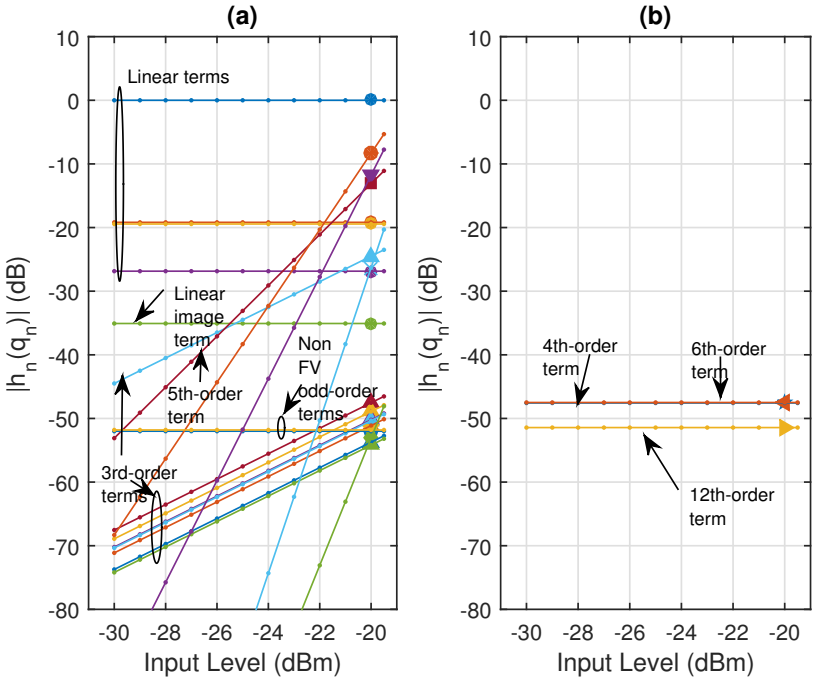

Figure 7.14 Measured normalized coefficients at $P_o = +26$ dBm (large marks) and computed by using (6.15) (straight lines). (a) Odd-order and (b) even-order coefficients. The PA input power is represented in the abscissas axis. Realistic transmitter TUT-2.

respectively. The ACPRs of the transmitter without DPD are also shown, demonstrating a reduction of 17 dB in the first adjacent channel at $P_o = +26$ dBm. The DPD linearization makes the ACPR_1 better than -55 dB and the ACPR_2 better than -60 dB, keeping the flexibility of compensating the linear and nonlinear impairments in the range from $P_o = +16$ dBm to $+26$ dBm without any further modification of the coefficients. This additional capability is also observed in the results of the NMSE displayed in figure 7.16 for the transmitter without DPD (dashed line) and with DPD (solid line). NMSE levels of about -50 dB in a wide range of operating powers demonstrate a maintained performance of the linearizer. Lastly, the normalized output spectra of the linearized transmitter are represented in figure 7.17 for average power levels ranging from $+17.8$ dBm to $+26.6$ dBm ($+28.8$ dBm to $+37.6$ dBm of peak power) in 3 dB steps.

7.4.3 Power Adaptability of ACC and CVS Linearizers

For TUT-2, we also implemented the DPD for joint mitigation of the modulator and the PA using the ACC model for SISO transmitters (6.16). The corresponding DPD was designed with equivalent settings (i.e, thirteenth order and $Q = 3$ for orders 1-5), yielding

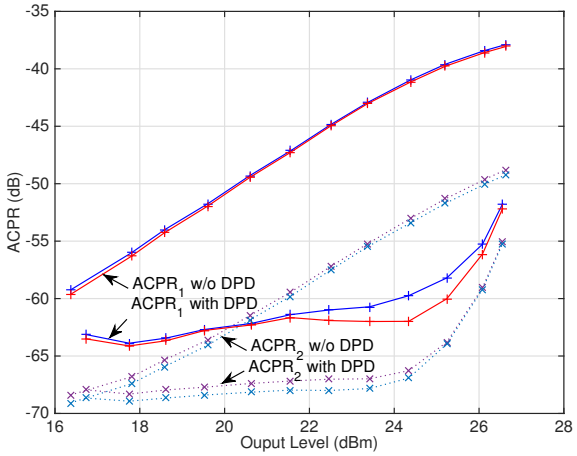

Figure 7.15 Adjacent channel power ratios of the output spectrum with and without DPD. First adjacent channel (solid lines) and second adjacent channel (dotted lines). Realistic transmitter TUT-2.

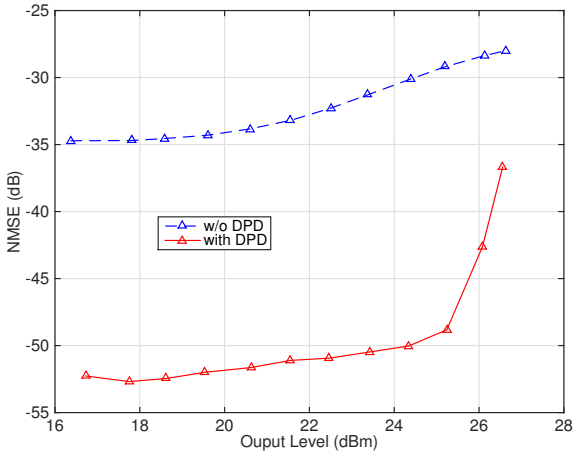

Figure 7.16 NMSE of the output signal without DPD (dashed line) and with DPD (solid line). Realistic transmitter TUT-2.

32 coefficients that were estimated following the conventional LS algorithm. For the PA delivering $P_o = +26$ dBm, the attained ACPR_1 shown with filled circles in figure 7.18 a) has improved from -38.6 dB without DPD (see figure 7.15) to -55 dB. This value is very similar to that obtained with the present proposal, represented again here for comparison (triangles). These results indicate a good optimization of the performance at the operational

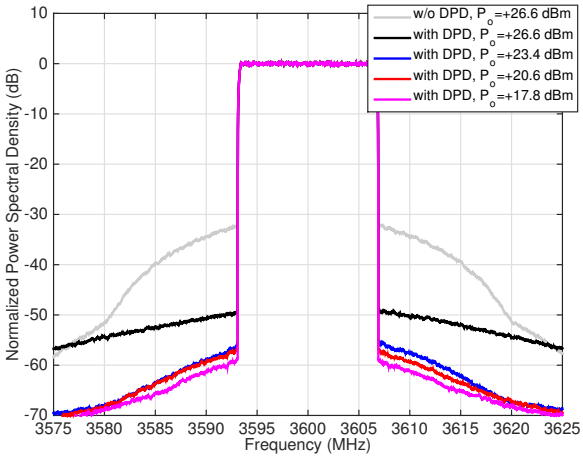

Figure 7.17 Normalized spectral density of the linearized output for power levels ranging from +17.8 dBm to +26.6 dBm. As a reference, the trace of the output spectrum without DPD is also shown. Realistic transmitter TUT-2.

point where both models were estimated (filled marks). In terms of model-order reduction, notice that the CVS DPD needs 14 coefficients, which compares favourably with the 32 coefficients necessary for a similar performance of the ACC DPD. When the ACC DPD is extended to other drive levels, its ACPR_1 deteriorates despite the PA is entering into a weakly nonlinear regime. On the contrary, the ACPR_1 of the CVS DPD shows a progressive reduction with a maximum improvement of about 9 dB with respect to the ACC DPD, demonstrating a better adaptability to output power variations. The NMSE results plotted in figure 7.18 b) repeat the performance deterioration of the ACC DPD and the adaptability of the CVS DPD under power level variations.

The ACPR_1 and NMSE behavior displayed in the figure is an example of overfitting in the case of the ACC DPD. The local character of the optimization is originated by the *a priori* assumption of the DPD structure. Therefore, the values estimated by the LS algorithm for the unneeded coefficients are 'coupled' to the indispensable ones and a variation of operational conditions disturbs the well functioning of the DPD at $P_o = +26$ dBm. The flexible performance exhibited by the CVS DPD is understood recalling that the method presented here relies on a complete set of non-orthogonal regressors, and on the OMP, a recursive algorithm to compute representations of the system output with respect to these non-orthogonal regressors. After a selection of the coefficients that surpass a given threshold, the resultant CVS DPD structure retains only the indispensable regressors.

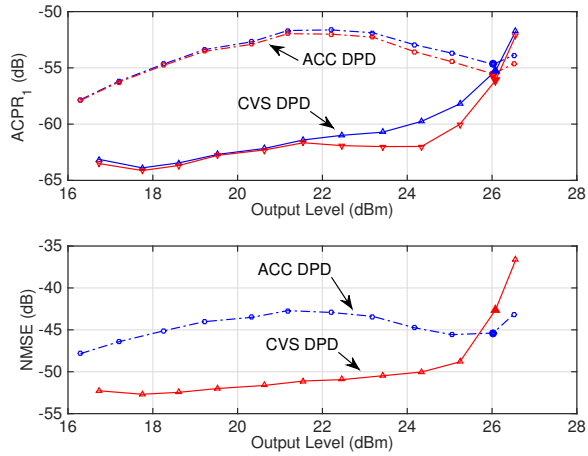

Figure 7.18 Comparison of a) lower and upper ACPR_1 and b) NMSE for the SISO ACC model [47] (circles) and the present proposal with the CVS model (triangles). The filled marks correspond to performance at the identification level ($P_o = +26$ dBm). Realistic transmitter TUT-2.

8 Conclusions and future work

8.1 Conclusions

In this section we summarize the principal results that were reached along these last years of research in the nonlinear mitigation of current wireless communications systems, being the main contributions of this Thesis.

First, chapter 3 addresses the doubly symmetric Parafac decomposition of Volterra kernels. Two novel DPDs that take advantage of block processing and a frequency domain representation to operate under a reduced complexity have been shown. Starting with a general VP structure, a block-oriented version of the CLMS adaptive algorithm in the frequency domain has been derived for this model to proceed with parameter identification, yielding a DPD that we have referred to as the B-VP. This structure has been modified to make it specially suitable for OFDM signals, taking advantage of the natural representation of M-QAM symbols in the frequency domain in an OFDM signal. In this case, a significant part of the OFDM modulator can be embedded into the DPD architecture, which has been denoted as FD-VP. While the B-VP DPD is intended to operate with time-domain to time-domain input and output waveform block samples, the FD-VP delivers time-domain output block samples from frequency-domain input symbols, with a block length corresponding to the number of subcarriers of the OFDM signal. Expressions for the running complexity of both approaches have been derived, allowing an analysis of the impact of the model parameters (e.g, order, rank and block length) in the complexity. Both DPD approaches have been tested experimentally with a commercial PA, and compared with a high-performance time-domain DPD. The PA was driven with a 15-MHz LTE-downlink signal operating in a working condition that violates the limit of unwanted emissions for the LTE standard when no predistorsion is implemented. Measurement results show that the linearization metrics obtained with the proposed methods are as good as those of the reference DPD, both in terms of spectral regrowth reduction and in-band linearization. Additionally, this performance is accomplished with a significantly lower running complexity, what makes the proposed methods suitable for a real-time implementation.

Next, in chapter 4, a formal deduction of a Volterra model for complex-valued nonlinear systems has been analyzed. It has also been demonstrated that the input-output relationships derived *ad-hoc* for several communication systems –a power amplifier, a beamformer, an I/Q modulator– are particular cases of the CVS model. To manage the increment of the parameters, the kernel sparsity in this approach has been exploited and the structure complexity with the proposed WNL-CS model has been reduced. Through experimental validation, it has been also shown that this model is able to accurately represent a diverse range of distortion sources in communication subsystems, including impairments at the I/Q modulator combined with the nonlinear distortion associated to PAs. Although conventional behavioral models are accurate for the description of nonlinearities in PAs, they show a degraded performance when the input signal is impaired by a quadrature modulator. On the contrary, the richness of its regressors makes the proposed model a robust representation of a general transmitter-receiver architecture.

In chapter 5, an improved method for the sparse recovery of Volterra series models has been presented. It adds structural information of the PA model in the selection process to achieve a reduction in the optimum number of coefficients while maintaining the fidelity. The benefits of the introduction of the span parameter has been experimentally demonstrated. This method is susceptible to be applied to any kind of Volterra series behavioral model. Predistortion with models given by the algorithm have been performed showing a high reduction of the complexity of the model with the same level of performance that the ones given by the complete model before pruning.

Finally, in chapter 6, a reliable procedure to identify the model coefficients of a power-scalable DPD for the joint compensation of wireless communications transmitters has been reported. It has been applied to a DPD based on the CVS model and has been approvingly compared with other recent DPD proposals that have been advanced to overcome the limitations of conventional models under the presence of I/Q modulator impairments. The proposed DPD design uses a regressor search algorithm and a threshold to limit the number of parameters, allowing a significant reduction guaranteed by discarding negligible coefficients. Several causes are behind the model-order reduction: irrelevant kernel types, truncated nonlinear order, truncated memory depth, sparse memory delays, etc. The coefficients were identified at a power level where the PA is near saturation, and the invariance of the denormalized kernels in the measured dynamic range was exploited to estimate the values at other operating conditions without the need of interpolation or a new series of measurements. Tailoring of the number of parameters at other levels is performed with a simple rejection of regressors with normalized coefficients below the predetermined threshold. In consequence, the number of coefficients is further reduced as the transmitter enters into the less nonlinear operating range. In terms of model-order reduction and precision, a better performance with respect to other alternative approaches has been demonstrated in a wide range of measured power levels below the DPD upper operating point.

8.2 Future Work

The work presented in this Thesis provides solutions to some of the practical problems in PA and transmitter modeling and linearization under sparsity hypothesis and can be extended in any of the following ways.

The use of iterative signal processing techniques for the sparse regression of Volterra models has a natural continuity with the use of machine learning and deep learning techniques. Although it has been proved that compressive-sensing techniques allow to recover a sparse regression that works with the same level of performance than those attained by the LS regression, the use of artificial intelligence techniques as those afore-mentioned are promising in this field. The adaptation of multi-layered recognition algorithms to the unsupervised recovering of DPDs is an interesting field still to be discovered.

8.3 Contributions

This Thesis is based on the following list of contributions:

[2017a] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and **J. A. Becerra**, "Formal deduction of a Volterra series model for complex-valued systems," *Signal Processing*, vol. 131, no. Supplement C, pp. 245 – 248, 2017.

[2017b] Crespo-Cadenas, C., M. J. Madero-Ayora, Reina-Tosina, J., and **J. A. Becerra**, "Transmitter linearization adaptable to power-varying operation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 10, pp. 3624–3632, Oct 2017.

[2017c] **J. A. Becerra**, M. J. Madero-Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Digital predistortion of power amplifiers using structured compressed-sensing Volterra series," *Electronics Letters*, vol. 53, no. 2, pp. 89–91, 2017.

[2017d] M. J. Madero-Ayora, C. Crespo-Cadenas, J. Reina-Tosina, and **J. A. Becerra**, "Effect of an impaired I/Q modulator in the structure of a behavioral model for power amplifiers," in *2017 IEEE 18th Wireless and Microwave Technology Conference (WAMICON)*, April 2017, pp. 1–4.

[2017e] C. Crespo-Cadenas, J. Reina-Tosina, M. J. Madero-Ayora, and **J. A. Becerra**, "Identification of volterra model parameters in wireless systems," in *2017 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, Jan 2017, pp. 96–99.

[2017f] M. J. Madero-Ayora, C. Crespo-Cadenas, J. Reina-Tosina, and **J. A. Becerra**, "Relationship Between a Pruned Volterra Model Structure and Impairments in an I/Q Modulator," in *47th European Microwave Confecence*, Oct 2017.

[2016a] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and **J. A. Becerra**, "Joint compensation of modulator and power amplifier nonlinearities based on a complex-valued Volterra model," in *2016 46th European Microwave Conference (EuMC)*, Oct 2016, pp. 1067–1070.

[2016b] **J. A. Becerra**, M. J. Madero-Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Podado de modelos de Volterra con técnicas de muestro compresivo." in *XXXI Simposium Nacional de la Unión Científica Internacional de Radio (XXXI Spanish National Symposium of the International Scientific Union of Radio)*, September 2016.

[2015a] **J. A. Becerra**, Madero-Ayora, M. J., Reina-Tosina, J., and Crespo-Cadenas, C, "Modelo ampliamente no lineal para la compensación de moduladores I/Q," in *XXX Simposium Nacional de la Unión Científica Internacional de Radio (XXX Spanish National Symposium of the International Union of Radio Science)*, September 2015.

[2015b] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and **J. A. Becerra**, "A widely nonlinear approach to compensate impairments in I/Q modulators," in *2015 European Microwave Conference (EuMC)*, Sept 2015, pp. 506–509.

[2015c] M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, and **J. A. Becerra**, "On the robustness of a widely nonlinear approach to model impairments in I/Q modulators," in *2015 Integrated Nonlinear Microwave and Millimetre-wave Circuits Workshop (INMMiC)*, Oct 2015, pp. 1–3.

[2014a] C. Crespo-Cadenas, J. Reina-Tosina, M. J. Madero-Ayora, and **J. A. Becerra**, "Volterra-parafac digital predistorter for OFDM wireless systems," in *2014 44th European Microwave Conference*, Oct 2014, pp. 829–832.

[2014b] C. Crespo-Cadenas, P. Aguilera-Bonet, **J. A. Becerra**, and S. Cruces, "On nonlinear amplifier modeling and identification using baseband Volterra-Parafac models," *Signal Processing*, vol. 96, no. Part B, pp. 401 – 405, 2014.

[2014c] **J. A. Becerra**, M. J. Madero-Ayora, Reina-Tosina, J., and Crespo-Cadenas, C, "Predistorsionador adaptativo por bloques basado en un modelo Volterra-Parafac," in *XXIX Simposium Nacional de la Unión Científica Internacional de Radio (XXIX Spanish National Symposium of the International Union of Radio Science)*, September 2014.

The author has been involved in the following publications that are not part of the thesis.

[2018a] **J. A. Becerra**, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. García-Frías, and G. Arce, "A doubly orthogonal matching pursuit algorithm for sparse predistortion of power amplifiers," *IEEE Microwave and Wireless Components Letters*, DOI: 10.1109/LMWC.2018.2845947. 1–3, 2018.

[2018b] **J. A. Becerra**, D. Herrera, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Sparse model selection of digital predistorters using Subspace Pursuit," in *13th European Microwave Integrated Circuits Conference (EuMIC)* , September 2018.

[2018c] A. Pérez-Hernández, **J. A. Becerra**, M. J. Madero-Ayora, and C. Crespo-Cadenas, "A Comprehensive Performance Benchmark of Component Selection Techniques for Volterra Behavioral Models," in *XXXIII Simposium Nacional de la Unión Científica Internacional de Radio (XXXIII Spanish National Symposium of the International Union of Radio Science)*, September 2018.

[2018d] **J. A. Becerra**, M. J. Madero-Ayora, A. Pérez-Hernández, and C. Crespo-Cadenas, " Algoritmo Subspace Pursuit para la Selección de Parámetros de Modelos de Volterra," in *XXXIII Simposium Nacional de la Unión Científica Internacional de Radio (XXXIII Spanish National Symposium of the International Union of Radio Science)*, September 2018.

[2018e] **J. A. Becerra**, A. Pérez-Hernández, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Efficient linearization of a RF Transmitter under 5G waveforms through Iterated Ridge Regression," in *91st ARFTG Microwave Measurement Conference*, June 2018.

[2017c] A. Pérez-Hernández, D. Herrera, M. J. Madero-Ayora, and **J. A. Becerra**, "Modelado Sin Memoria de Amplificadores de Potencia con Curvas de Bézier Cúbicas," in *XXXII Simposium Nacional de la Unión Científica Internacional de Radio (XXXII Spanish National Symposium of the International Union of Radio Science)*, September 2017.

[2017c] M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, and **J. A. Becerra**, "Revisión de Modelos de Volterra para Sistemas no Lineales con Señales Complejas," in *XXXII Simposium Nacional de la Unión Científica Internacional de Radio (XXXII Spanish National Symposium of the International Union of Radio Science)*, September 2017.

[2016] **J. A. Becerra**, M. J. Madero-Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Structured compressed-sensing for Volterra series models," in *2016 IEEE 17th Annual Wireless and Microwave Technology Conference (WAMICON)*, April 2016, pp. 1–4.

[2012] **J. A. Becerra**, J. Reina-Tosina, E. A. Ojeda-García, and C. Crespo-Cadenas, "Modelo simple con características electrotérmicas para dispositivos HEMT de GaN," in *XXVII Simposium Nacional de la Unión Científica Internacional de Radio (XXVII Spanish National Symposium of the International Union of Radio Science)*, September 2012.

List of Figures

List of Tables

Bibliography

[1] I. E. Agency, "More data, less energy. making network standby more efficient in billions of connected devices," International Energy Agency, Tech. Rep., 2014.

[2] T. C. Group, "Smart 2020: Enabling the low carbon economy in the information age," The Climate Group, Tech. Rep., 2008.

[3] K. M. Gharaibeh, *Nonlinear distortion in wireless systems: Modeling and simulation with MATLAB*. John Wiley & Sons, 2011.

[4] V. Volterra, *Sopra le funzioni che dipendono da altre funzioni*. Tip. della R. Accademia dei Lincei, 1887.

[5] N. Wiener, "Response of a non-linear device to noise," Massachusetts Inst. of Tech. Cambridge Radiation Lab, Tech. Rep., 1942.

[6] V. J. Mathews and G. Sicuranza, *Polynomial signal processing*. John Wiley & Sons, Inc., 2000.

[7] A. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1510–1520, June 2010.

[8] P. Landin, M. Isaksson, and P. Handel, "Comparison of evaluation criteria for power amplifier behavioral modeling," in *Proc. IEEE MTT-S Int. Microwave Symp. Digest*, Jun. 2008, pp. 1441–1444.

[9] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*. Artech House, London, 2014.

[10] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223–227, Jan. 1997.

[11] A. Legendre, "New methods for the determination of the orbits of the comets," *F. Didot, Paris*, 1805.

[12] 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, Nov 2001.

[13] R. Tibshirani, M. Wainwright, and T. Hastie, *Statistical Learning with Sparsity: The LASSO and Generalizations*. Chapman and Hall/CRC, 2015.

[14] B. Widrow, "Thinking about thinking: the discovery of the LMS algorithm," *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 100–106, Jan. 2005.

[15] R. Schwartz, *R&S SMU200A Vector Signal Generator Operating Manual*, 2016.

[16] K. Technologies, *Keysight X-Series Signal Analyzer: N9030A Real-Time Spectrum Analyzer User's & Programmer's Reference*, 2017.

[17] P. Landin, S. Gustafsson, C. Fager, and T. Eriksson, "Weblab: A web-based setup for PA digital predistortion and characterization [application notes]," *IEEE Microw. Mag.*, vol. 16, no. 1, pp. 138–140, Feb. 2015.

[18] D. 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, oct 2006.

[19] 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, Jul. 2008.

[20] 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 Microw. Wireless Compon. Lett.*, vol. 19, no. 8, pp. 530–532, Aug. 2009.

[21] G. Favier, A. Y. Kibangou, and T. Bouilloc, "Nonlinear system modeling and identification using volterra-PARAFAC models," *Int. J. Adapt. Control Signal Process.*, vol. 26, no. 1, pp. 30–53, sep 2011.

[22] T. Bouilloc and G. Favier, "Nonlinear channel modeling and identification using baseband volterra–parafac models," *Signal Process.*, vol. 92, no. 6, pp. 1492–1498, jun 2012.

[23] C. Crespo-Cadenas, P. Aguilera-Bonet, J. Becerra-González, and S. Cruces, "On nonlinear amplifier modeling and identification using baseband volterra-parafac models," *Signal Process.*, vol. 96, pp. 401–405, mar 2014.

[24] C. Crespo-Cadenas, J. Reina-Tosina, M. J. Madero-Ayora, and J. A. Becerra-González, "Volterra-parafac digital predistorter for OFDM wireless systems," in *Proc. 44th European Microwave Conf. (EuMC)*, Oct. 2014, pp. 829–832.

[25] S. Haykin, *Adaptive filter theory*. Prentice Hall, 2002.

[26] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[27] S. Benedetto, E. Biglieri, and R. Daffara, "Modeling and performance evaluation of nonlinear satellite links-a volterra series approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, no. 4, pp. 494–507, July 1979.

[28] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[29] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[30] P. Chevalier, P. Duvaut, and B. Picinbono, "Complex transversal Volterra filters optimal for detection and estimation," in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP) 1991*, Apr. 1991, pp. 3537–3540 vol.5.

[31] P. Chevalier, A. Oukaci, and J. P. Delmas, "Third order widely non linear Volterra mvdr beamforming," in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP) 2011*, May 2011, pp. 2648–2651.

[32] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and J. A. Becerra-González, "A widely nonlinear approach to compensate impairments in I/Q modulators," in *Proc. European Microwave Conf. (EuMC)*, Sep. 2015, pp. 506–509.

[33] P. Schreier and L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[34] S. O. Rice, "Volterra systems with more than one input port — distortion in a frequency converter," *The Bell System Technical Journal*, vol. 52, no. 8, pp. 1255–1270, Oct. 1973.

[35] R. Raich and G. T. Zhou, "On the modeling of memory nonlinear effects of power amplifiers for communication applications," in *Proc. IEEE 10th Digital Signal Processing Workshop and the 2nd Signal Processing Education Workshop*, Oct. 2002, pp. 7–10.

[36] H. Li and T. Adal, "Complex-valued adaptive signal processing using nonlinear functions," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, feb 2008.

[37] J. Reina-Tosina, M. Allegue-Martínez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 745–753, Feb. 2015.

[38] E. Lima, T. Cunha, H. Teixeira, M. Pirola, and J. Pedro, "Base-band derived Volterra series for power amplifier modeling," in *Proc. 2009 IEEE MTT-S Int. Microw. Symp. Digest (IMS 2009)*, June. 2009, pp. 1361–1364.

[39] A. Abdelhafiz, A. Kwan, O. Hammi, and F. Ghannouchi, "Digital predistortion of LTE-A power amplifiers using compressed-sampling-based unstructured pruning of Volterra series," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2583–2593, Nov. 2014.

[40] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Structured compressed-sensing for Volterra series models," in *Proc. 2016 IEEE 17th Ann. Wireless Microw. Techn. Conf.*, Apr. 2016, pp. 1–4.

[41] L. Anttila, P. Handel, and M. Valkama, "Joint mitigation of power amplifier and i/q modulator impairments in broadband direct-conversion transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 730–739, apr 2010.

[42] M. Younes and F. M. Ghannouchi, "Generalised twin-box model for compensation of transmitters radio frequency impairments," *IET Communications*, vol. 8, no. 4, pp. 413–418, Mar. 2014.

[43] H. Qian, S. Yao, H. Huang, X. Yang, and W. Feng, "Low complexity coefficient estimation for concurrent dual-band digital predistortion," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 10, pp. 3153–3163, Oct. 2015.

[44] O. Hammi, A. Kwan, and F. M. Ghannouchi, "Bandwidth and power scalable digital predistorter for compensating dynamic distortions in RF power amplifiers," *IEEE Trans. Broadcast.*, vol. 59, no. 3, pp. 520–527, Sep. 2013.

[45] Y. Guo, C. Yu, and A. Zhu, "Power adaptive digital predistortion for wideband RF power amplifiers with dynamic power transmission," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 11, pp. 3595–3607, Nov. 2015.

[46] B. Schubert, A. Gökceoglu, L. Anttila, and M. Valkama, "Augmented Volterra predistortion for the joint mitigation of power amplifier and I/Q modulator impairments in wideband flexible radio," in *Proc. IEEE Global Conf. Signal and Information Processing*, Dec. 2013, pp. 1162–1165.

[47] Z. A. Khan, E. Zenteno, P. Händel, and M. Isaksson, "Digital predistortion for joint mitigation of I/Q imbalance and MIMO power amplifier distortion," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 1, pp. 322–333, Jan. 2017.

[48] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and J. Becerra-González, "Formal deduction of a Volterra series model for complex-valued systems," *Signal Process.*, vol. 131, pp. 245–248, Feb. 2017.

[49] J. Reina-Tosina, M. Allegue-Martínez, M. J. Madero-Ayora, C. Crespo-Cadenas, and S. Cruces, "Digital predistortion based on a compressed-sensing approach," in *Proc. European Microwave Conf. (EuMC)*, Oct. 2013, pp. 408–411.

[50] C. Crespo-Cadenas, J. Reina-Tosina, M. J. Madero-Ayora, and J. A. Becerra, "Identification of Volterra model parameters in wireless systems," in *Proc. IEEE Topical Conf. RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, Jan. 2017, pp. 96–99.

[51] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals Systems and Computers*, Nov. 1993, pp. 40–44 vol.1.

[52] M. Amiri, S. Bassam, M. Helaoui, and F. Ghannouchi, "New order selection technique using information criteria applied to SISO and MIMO systems predistortion," *Int. J. Microwave Wireless Technolog.*, vol. 5, no. 2, pp. 123–131, Mar. 2013.

[53] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[54] S. M. Kay, *Fundamentals of Statistical Signal Processing. Detection Theory*. Prentice Hall, Englewood Cliffs, New Jersey, 1998.

[55] D. Wisell, "Identification and measurement of transmitter non-linearities," in *IEEE ARFTG Conference Digest-Fall, 56th*, vol. 38, 2000, pp. 1–6.

[56] H. Cao, A. S. Tehrani, C. Fager, T. Eriksson, and H. Zirath, "I/q imbalance compensation using a nonlinear modeling approach," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 3, pp. 513–518, 2009.

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

[58] J. K. Cavers and M. W. Liao, "Adaptive compensation for imbalance and offset losses in direct conversion transceivers," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 581–588, Nov. 1993.

[59] *Operation and performance of the ISL5239 predistortion linearizer*, Intersil application note AN1022, Jul. 2002.

[60] J. H. de M. Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Tensor CP decomposition with structured factor matrices: Algorithms and performance," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 757–769, Jun. 2016.

[61] W. Demenitroux, C. Maziere, E. Gatard, S. Dellier, C. Saboureau, M. Campovecchio, and R. Quere, "A new multi-harmonic Volterra model dedicated to GaN packaged transistor or sspa for pulse application," in *Proc. IEEE MTT-S Int. Microwave Symp*, Jun. 2011, pp. 1–4.

[62] H. Zhou, G. Wan, and L. Chen, "A nonlinear memory power amplifier behavior modeling and identification based on memory polynomial model in soft-defined shortwave transmitter," in *Proc. 6th Int. Conf. Wireless Communications Networking and Mobile Computing (WiCOM)*, Sep. 2010, pp. 1–4.

[63] G. T. Zhou and G. B. Giannakis, "Nonlinear channel identification and performance analysis with PSK inputs," in *Proc. First IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communications*, Apr. 1997, pp. 337–340.

[64] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[65] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[66] A. Montanari, *Graphical models concepts in compressed sensing*. Cambridge University Press, 2012, p. 394–438.

[67] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[68] G. Shi, D. Liu, D. Gao, Z. Liu, J. Lin, and L. Wang, "Advances in theory and application of compressed sensing," *Acta Electronica Sinica*, vol. 37, no. 5, pp. 1070–1081, 2009.

[69] M. Davies, "Compressed sensing: Theory and applications."

[70] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[71] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013, vol. 1, no. 3.

[72] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, "Compressive sensing: From theory to applications, a survey," *J. Commun. Networks*, vol. 15, no. 5, pp. 443–456, 2013.

[73] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

[74] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 143–168, 2008.

[75] L. Gan, "Block compressed sensing of natural images," in *15th International Conference on Digital Signal Processing*. IEEE, 2007, pp. 403–406.

[76] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[77] G. Kutyniok, "Theory and applications of compressed sensing," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 79–101, 2013.

[78] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[79] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Comput. Networks*, vol. 45, no. 3, pp. 245–259, 2004.

[80] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for ℓ_1-regularized minimization with applications to compressed sensing," *Computational and Applied Mathematics (CAAM) Technical Report, Rice University*, vol. 43, p. 44, 2007.

[81] J. K. Cavers, "The effect of quadrature modulator and demodulator errors on adaptive digital predistorters for amplifier linearization," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 456–466, May 1997.

[82] T. Blumensath and M. E. Davies, "Stagewise weak gradient pursuits," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4333–4346, Nov. 2009.

[83] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *CoRR*, vol. abs/1108.0477, 2011. [Online]. Available: http://arxiv.org/abs/1108.0477

[84] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce, "Approximate message passing in coded aperture snapshot spectral imaging," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015, pp. 609–613.

[85] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, Jan. 2010, pp. 1–5.

[86] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: II. analysis and validation," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, Jan. 2010, pp. 1–5.

[87] M. Younes, O. Hammi, A. Kwan, and F. Ghannouchi, "An accurate complexity-reduced plume model for behavioral modeling and digital predistortion of RF power amplifiers," *IEEE Trans. Ind. Electron.*, vol. 58, no. 4, pp. 1397–1405, Apr. 2011.

[88] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley, 1980.

[89] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *IEEE International Symposium on Information Theory (ISIT)*, June 2014, pp. 236–240.

[90] A. Montanari, "Graphical models concepts in compressed sensing," *CoRR*, vol. abs/1011.4328, 2010. [Online]. Available: http://arxiv.org/abs/1011.4328

[91] V. Kekatos and G. Giannakis, "Sparse Volterra and polynomial regression models: Recoverability and estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5907–5920, Dec. 2011.

[92] V. Kekatos and G. Giannakis, "Compressed sensing for Volterra and polynomial regression models," in *Signal Processing Conference, 2011 19th European*, Aug. 2011, pp. 373–377.

[93] M. Khalifa, A. Abdelhafiz, A. Kwan, F. Ghannouchi, and O. Hammi, "Behavioral modeling of envelope tracking power amplifier using Volterra series model and compressed sampling," in *EEE InternationalWireless Symposium (IWS),*, Mar. 2015, pp. 1–4.

[94] A. Bolstad and B. Miller, "Sparse Volterra systems: Theory and practice," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May. 2013, pp. 5740–5744.

[95] A. Zhu, J. Pedro, and T. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of rf power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[96] Z. Yu and E. Zhu, "A comparative study of learning architecture for digital predistortion," in *Proc. Asia-Pacific Microwave Conf. (APMC)*, vol. 1, Dec. 2015, pp. 1–3.

[97] T. Magesacher and P. Singerl, "Benchmarking of learning architectures for digital predistortion," in *Proc. Systems and Computers 2016 50th Asilomar Conf. Signals*, Nov. 2016, pp. 648–651.

[98] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *Proc. IEEE Int. Symp. Wireless Communication Systems*, Oct. 2008, pp. 309–313.

[99] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. 23, no. 5, pp. 285–291, May 1976.

[100] M. Ergen, *Mobile Broadband - Including WiMAX and LTE*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[101] D. H. Holma and D. A. Toskala, *LTE for UMTS - OFDMA and SC-FDMA Based Radio Access*. Wiley Publishing, 2009.

[102] M. Allegue-Martínez, "Modeling and compensation of non-linear effects in wireless communications systems," Ph.D. dissertation, Universidad de Sevilla, 2012.

[103] A. S. Tehrani, "Behavioral modeling of wireless transmitters for distortion mitigation," Ph.D. dissertation, Chalmers University of Technology, 2012.

[104] C. Crespo-Cadenas, M. Madero-Ayora, J. Reina-Tosina, and J. Becerra-Gonzalez, "A widely nonlinear approach to compensate impairments in I/Q modulators," in *2015 European Microwave Conf. (EuMC)*, Sept 2015, pp. 506–509.