

On Segmented Predistortion for Linearization of RF Power Amplifiers

Genevieve BAUDOIN

ESYCOM, UMR CNRS 9007, Gustave Eiffel University, ESIEE Paris, 93162 Noisy-le-Grand, France

genevieve.baudoin@esiee.fr

Submitted February 6, 2020 / Accepted February 28, 2020

Abstract. *This paper presents a general survey of digital predistortion (DPD) techniques with segmentation. A comparison of global DPD with two segmented approaches namely Vector-Switched DPD and Decomposed Vector Rotation DPD is presented with the support of experimentation on a strongly non-linear 3 ways Doherty PA. It shows the interest of both segmented approaches in terms of linearization performance, complexity and ease of implementation compared to the global DPD. The paper starts with some mathematical generalities on interpolation and splines. It focuses on segmented models derived from Volterra series even if the presented principles can also be applied to neural networks.*

Keywords

Digital predistortion, piecewise models, splines, vector-switched DPD, decomposed vector rotation DPD, power amplifiers

1. Introduction

Digital predistortion is an efficient technique to linearize power amplifiers (PA) in wireless transmitters. It is widely used in base stations of cellular communication and broadcast systems. Power amplifiers are critical elements of radiocommunication systems because their power efficiency conditions the autonomy and cost of equipments and their linearity influences communication performance. New waveforms proposed in order to improve spectral occupancy exhibit very high crest factors and are very sensitive to PA nonlinearities. But to achieve a good power efficiency, it is necessary to operate the PA in a nonlinear region. Linearization techniques are therefore necessary to limit in-band signal distortion and out of band spectral regrowth.

The principle of digital predistortion consists in processing the baseband complex envelop of the PA input signal by a predistorter (DPD) so that the cascade of the DPD and PA becomes linear up to a certain amplitude value. the DPD should have inverse characteristics of those of the PA to be

able to pre-compensate for nonlinear distortion and dynamic behavior (memory effects) of the PA. Many models of nonlinear dynamic system have been proposed for the DPD [1], [2]. They are discrete-time baseband models with complex input and output signals. Most of them are derived from truncated Volterra series (Memory Polynomial models (MP) [3], Generalized Memory Polynomial models (GMP) [4], [5]) or dynamic Volterra series (Dynamic Deviation Reduction-Based models (DDR) [6]) with limitation to finite memory lengths and nonlinearity orders. Neural networks are also potential candidates but they require nonlinear optimization techniques to update their coefficients which is not very appropriate for real-time adaptive DPD.

Volterra series present two interesting properties for the modeling of nonlinear dynamic systems: generality and linearity of the model in function of their coefficients which simplifies their identification. But their number of coefficients increases dramatically with memory depth and order of nonlinearity. Moreover they are generally built with non-orthogonal basis based on monomials which leads to bad numerical properties for their identification especially for high-order nonlinearity.

Models derived from Volterra series such as GMP or DDR have proven their effectiveness for numerous applications using mildly nonlinear PA such as class AB PA. But advanced architectures of power transmitters with good efficiency such as Doherty PA, envelop tracking, switched or out-phasing PA exhibit strongly nonlinear dynamic behavior more difficult to model. Also, new communication systems allow for very high data rate by using very wide bandwidth multidimensional signals (e.g. 4G and 5G systems with carrier aggregation and MIMO). It represents new challenges for DPD in terms of bandwidth, nonlinearity and dynamic behavior. It becomes difficult for a global DPD model to achieve an accurate representation of the system with good numerical properties and low computational complexity. This has moved research interest towards different local modeling approaches in which the global operating space is split into several subspaces represented by local models well suited to each sub-space. These local models have to be joined in some way to cover the global space. One of the motivations is to decompose a complicated problem into several simpler ones.

Another one is to use the locality to obtain quasi-orthogonal basis. The decomposition (segmentation) can be applied on temporal signal magnitude and phase, on signal spectrum, on the system. This segmented or piecewise DPD approach raises different questions such as:

- how to partition the original space,
- how to determine good models for the different segments of the partition,
- how to handle DPD operators with complex inputs,
- how to handle the dynamic aspects,
- how to estimate the coefficients of these local models and how to represent them with sparsity,
- how to join the local models.

Piecewise approximation is not a new idea. There has been a lot of works in particular on piecewise interpolation or approximation of nonlinear real-valued functions, e.g. using splines. The application of these theories to the case of DPD is not straightforward because of two main reasons:

- the DPD is not a simple function. It is a dynamic nonlinear system;
- the input and output of the DPD are complex signals. The DPD is a complex-valued operator of multiple complex variables.

This paper presents a general survey of DPD techniques with segmentation in the temporal domain. It also presents an experimental comparison of different approaches. It focuses on segmented models derived from Volterra series even if the presented principles can also be applied to neural networks. It starts with some mathematical generalities on interpolation, approximation and splines (Sec. 2). In Sec. 3 it gives some generalities on modeling and training of DPD. Section 4 focuses on segmented DPD with functions of a single real-valued variable. Section 5 is dedicated to segmented DPD that can manage nonlinearity and memory domains. Section 6 briefly presents some advanced segmented multidimensional DPD for multiband or MIMO applications. Section 7 is devoted to an experimental comparison of two of the most promising segmented DPD techniques. Section 8 is the conclusion.

2. Some Mathematical Considerations

2.1 Polynomial Interpolation/Approximation

For a function $y = f(x) : [a, b] \rightarrow \mathbb{R}$ and a set of $N + 1$ points $\{(x_0, y_0), (x_1, y_1), \dots, (x_N, y_N)\}$, there is a unique polynomial interpolator of degree N $p(x)$ such that $p(x_i) = y_i$. This polynomial can be expressed in different ways, e.g. using the Newton's divided difference formula or the Lagrange interpolation formula. The interpolation polynomial of f ,

can be written using the basis made of Lagrange polynomials $\mathcal{L}_{N,i}(x)$, as:

$$\mathcal{L}_N f(x) = \sum_{i=0}^N f(x_i) \mathcal{L}_{N,i}(x),$$

with

$$\mathcal{L}_{N,i}(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}, \quad i = 0, 1, \dots, N.$$

The approximation error is $e(x) = f(x) - p(x)$ and for $f(x)$ $(N + 1)$ times differentiable:

$$\begin{aligned} \forall x \in [a, b] \exists \xi \in (\min(x_i, x), \max(x_i, x)), \\ e(x) = f(x) - p(x) = (x - x_0) \cdots (x - x_N) \frac{f^{N+1}(\xi)}{(N+1)!}, \\ \max_{x \in [a, b]} |e(x)| \leq \frac{(b-a)^{N+1}}{(N+1)!} \max_{x \in [a, b]} |f^{N+1}(x)|. \end{aligned}$$

Weierstrass has shown that f can be uniformly approximated by a polynomial but this is not the interpolation polynomial. Indeed, even if the interval between points x_i is reduced and the degree of $p(x)$ is increased, $p(x)$ does not converge towards $f(x)$. This is called Runge's phenomena.

The approximation error depends on the product $v(x) = (x - x_0) \cdots (x - x_N)$ whose value is related to the segmentation points x_0, x_1, \dots, x_N . Tchebychev has studied the segmentation that minimizes the maximum of $|v(x)|$ on the interval $[a, b]$. Tchebychev's alternance theorem states that the monic polynomial $u(x)$ of degree n that minimizes $L = \max |u(x)|$ (on a given interval) takes alternatively the values $\pm L$, $n + 1$ times. The best solution is $u(x) = 2^{-n} T_n(x)$ where T_n is the Tchebychev polynomial of degree n . On the interval $[-1, +1]$, it is defined by:

$$T_n(x) = \cos(n\phi) \text{ with } x = \cos(\phi) \text{ for } x \in [-1, +1].$$

The polynomial $T_n(x)$ has n roots x_k , $T_n(x_k) = 0$.

We can deduce that the segmentation x_0, x_1, \dots, x_N that minimizes the maximum of $|(x - x_0) \cdots (x - x_N)|$ on the interval $[-1, +1]$ corresponds to the roots of the Tchebychev's polynomial of degree $N + 1$. They are given by:

$$x_k = \cos\left(\frac{2(k+1)\pi}{2N+2}\right), \quad k = 0, 1, \dots, N.$$

The solution for the interval $[a, b]$ is obtained as:

$$x_k = \frac{a+b}{2} + \frac{b-a}{2} \cos\left(\frac{2(k+1)\pi}{2N+2}\right), \quad k = 0, 1, \dots, N.$$

Using this segmentation, the approximation error using the Tchebychev segmentation is bounded by:

$$\max_{x \in [a, b]} |f(x) - p(x)| \leq \frac{2}{(N+1)!} \left(\frac{b-a}{4}\right)^{N+1} \max_{x \in [a, b]} |f^{N+1}(x)|.$$

This bound is $4^{N+1}/2$ times smaller than that of the general case. This highlights the importance of the segmentation to

minimize the approximation error. For example, Fig. 1 shows the interpolation by a polynomial of degree $N = 10$ of the function $y = 1/(1 + x^2)$ for $x \in [-5, +5]$ using a uniform segmentation or a Tchebychev segmentation.

The Lagrange polynomials $\mathcal{L}_{N,i}(x)$ using Tchebychev segmentation are interesting because their supports are well localized around points x_i . So they are quasi-orthogonal which is not at all true for uniform segmentation. Figure 2 shows the Lagrange polynomials of degree 10 with a uniform or a Tchebychev segmentation.

In the case of the approximation of a data sequence of M samples (x_k, y_k) , $k = 0, \dots, M - 1$ with $x_k \in [a, b]$ by a polynomial $p(x_k)$ of degree n , $p(x_k)$ can be expressed using different basis corresponding to different segmentations of $[a, b]$. If we call respectively $\mathcal{L}_{i,\text{Tch}}(x_k)$ and $\mathcal{L}_{i,\text{Uni}}(x_k)$, $i = 0, \dots, n$, the Lagrange basis with Tchebychev or uniform segmentation, the approximation polynomial $p(x_k)$ is:

$$p(x_k) = \sum_{i=0}^n c_i \mathcal{L}_{i,\text{Tch}}(x_k) = \sum_{i=0}^n d_i \mathcal{L}_{i,\text{Uni}}(x_k). \quad (1)$$

If a least-square criterion is used, minimizing $\sum_{k=1}^M (p(x_k) - y_k)^2$, the vectors \mathbf{c} and \mathbf{d} containing coefficients c_i or d_i are obtained as solutions of :

$$\mathbf{R}\mathbf{c} = \Phi^T \mathbf{y} \text{ with } \mathbf{R} = \Phi^T \Phi. \quad (2)$$

The matrix \mathbf{y} is the $M \times 1$ column vector of samples y_k , $k = 0, \dots, M - 1$, Φ is the matrix $M \times (n + 1)$ of the basis functions $\Phi(k, i) = \mathcal{L}_i(x_k)$ and Φ^T is the transpose of Φ . In theory, whatever the chosen basis, the optimal solution for p should be the same. But the condition number of the correlation matrix \mathbf{R} strongly depends on the basis. For example, for a polynomial of degree $n = 35$, the condition number of the correlation matrix \mathbf{R} is greater than 10^{16} for the uniform segmentation while it is smaller than 30 for the Tchebychev segmentation. Therefore due to those bad numerical properties, the quality of the approximation obtained with uniform segmentation can be degraded specially when computation is done in single precision.

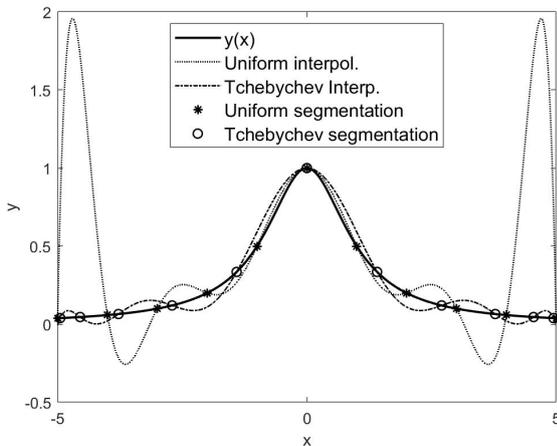


Fig. 1. Interpolation of $y = 1/(1 + x^2)$ with an order 10 polynomial with uniform or Tchebychev segmentation.

The good locality of Lagrange polynomials with Tchebychev segmentation has been exploited for DPD modeling by Barradas et al. in [7–9]. Details will be given later.

Another interesting point to remind is the influence of noisy data y_i on the approximation results. It is important for example to better understand the influence of the measurement noise on the PA output when training the DPD using indirect learning approach (see Sec. 3). It is well known that the noise will introduce a bias on the coefficients of the DPD. Suppose that the data y_i are corrupted by some noise ϵ_i , with $|\epsilon_i| < \epsilon$. The corrupted data are noted $\hat{y}_i = y_i + \epsilon_i$ and the interpolation polynomial \hat{p} . The error $\hat{p}(x) - p(x)$ can be bounded with Lebesgue constant that depends on the segmentation nodes:

$$|\hat{p}(x) - p(x)| \leq \epsilon \sum_{i=0}^n \max_{x \in [a,b]} |\mathcal{L}_{i,n}(x)|.$$

The Lebesgue constant $\max_{x \in [a,b]} |\mathcal{L}_{i,n}(x)|$ increases with the polynomial degree and is much smaller for Tchebychev segmentation than for uniform segmentation.

Lagrange interpolation only requires that $p(x_i) = f(x_i)$ for $i = 0, \dots, n$. Other polynomial interpolations request additional conditions on the derivatives of f . For example, Hermite interpolation requires that the 1st derivatives of f and p be equal at points x_i . For these $2(n + 1)$ conditions, there is a unique polynomial solution $p(x)$ of minimum degree $2n + 1$. The approximation error $e(x) = f(x) - p(x)$, for $f(x)$ $(2n + 2)$ times differentiable, is such that:

$$\forall x \in [a, b] \exists \xi \in (\min(x_i, x), \max(x_i, x)),$$

$$e(x) = (x - x_0)^2 \dots (x - x_n)^2 \frac{f^{2n+2}(\xi)}{(2n + 2)!}.$$

Hermite interpolation does not guarantee a uniform approximation of f when n increases.

But Bernstein polynomials associated to a continuous function f and noted $B_n f(x)$, converge uniformly towards f when the degree n increases. They correspond to a uniform segmentation and are expressed, for $x \in [a, b]$ as:

$$B_n f(x) = \sum_{k=0}^n \binom{n}{k} f(a + (b - a)k/n) (x - a)^k (x - b)^{n-k}.$$

The Bernstein basis polynomials of degree n are defined by $B_{k,n}(x) = \binom{n}{k} (x - a)^k (x - b)^{n-k}$, $k = 0, \dots, n$. Figure 3 shows Bernstein basis polynomials for $n = 10$ on the interval $[-5, 5]$. It can be seen that they are quite well localized around points x_i . Also the sum $\sum_{k=0}^n B_{k,n}(x) = 1 \forall x \in [a, b]$. The Bernstein approximation is obtained by convex linear combination which leads to good numerical properties.

Figure 4 shows the approximation by a polynomial of degree $n = 10$ of the function $y = 1/(1 + x^2)$ for $x \in [-5, +5]$ using a Bernstein polynomial (uniform segmentation) or a Lagrange polynomial with Tchebychev segmentation.

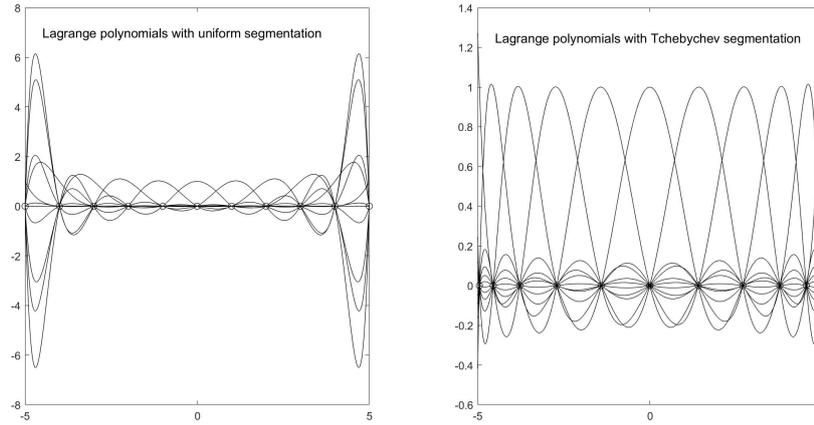


Fig. 2. Lagrange polynomials of degree 10 with uniform (left figure) or Tchebychev (right figure) segmentation.

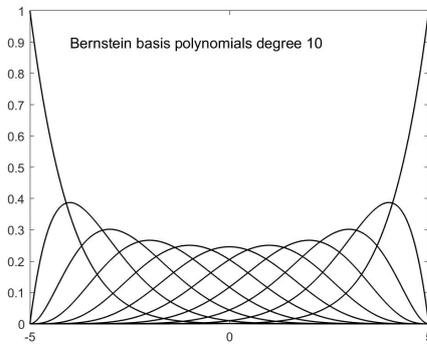


Fig. 3. Bernstein polynomials of degree 10.

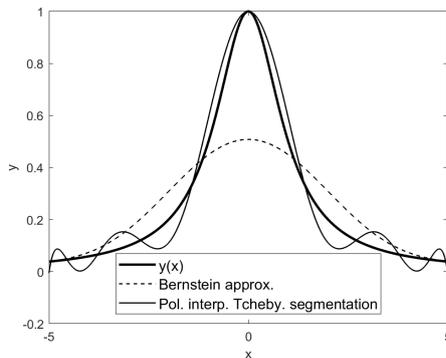


Fig. 4. Approximation of $y = 1/(1+x^2)$ with order 10 Bernstein or Lagrange (Tchebychev segmentation) polynomial.

On this example, it can be seen that the approximation error obtained with the Bernstein polynomial is larger than that obtained with Lagrange interpolation and a Tchebychev segmentation. But there is no Runge's phenomenon in the Bernstein approximation and, if $f(x)$ is continuous, it converges uniformly (even if rather slowly) towards $f(x)$ which is not true for Lagrange interpolation. An interesting way to obtain a faster convergence of the approximation is to use piecewise approximation.

2.2 Piecewise Polynomial Approximation

Approximation of functions by a single polynomial has some limitations, such as Runge's phenomenon for Lagrange uniform interpolation, necessity of high polynomial degree polynomials leading to high computation complexity and potential numerical problems. Piecewise approximation is a way to overcome some of these limitations. It allows to take advantage of the regularity of the function in a limited region.

To approximate a function $y = f(x)$, $x \in [a, b]$, the basic idea of piecewise polynomial approximation is to partition the interval $[a, b]$ in N segments over which f is approximated by a polynomial. As the segments are shorter than the length of $[a, b]$, f should have a more regular shape on each local segment than on the global interval and it should be possible to use polynomials of smaller degree than for the global polynomial approximation. Without additional constraints, the piecewise approximation may be discontinuous. Generally, some regularity is required for the approximation and a popular approach is the approximation by spline functions. There are two different ways to present splines. The first one consists in considering piecewise polynomials with continuity constraints on the function and its first derivatives at the borders of the segments. In this first approach the regularity constraints can be different for the different nodes. The second approach is based on a trade-off between the accuracy of the approximation and its regularity. In the second approach, the natural splines solution is obtained for a set of N data $(x_i, y_i)_{i=1}^N$ by approximating f by a function h optimizing the criterion:

$$\min_h \frac{1}{N} \sum_{i=1}^N (y_i - h(x_i))^2 + \lambda \int_a^b (h''(x))^2 dx$$

where $\lambda > 0$ is a smoothing constant. The solution of this problem is a piecewise function made of polynomials of degree $2m - 1$ in segments defined by boundaries x_i and satisfying continuity at the boundaries for the function and its $2m - 2$ first derivatives. With $m = 2$, this gives cubic splines.

A function $h(x)$ is a polynomial spline of order $d + 1$ in $[a, b]$ if $h \in C^{(d-1)}$ ($C^{(d-1)}$ space of functions with $d - 1$ continuous derivatives) and for a given set \mathbf{s} of non-decreasing segmentation values (knots) ($a = x_0, x_1, \dots, x_{n-1}, x_n = b$), $h(x)$ is a polynomial $p_i(x)$ of degree d on each segment $[x_i, x_{i+1}[$. We note $S_{d,\mathbf{s}}$ the set of all spline functions of order $d + 1$ for the set of knots \mathbf{s} . A cubic spline corresponds to the case $d = 3$. For a cubic spline, $p_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$, $x \in [x_i, x_{i+1}[$ and $p_i'(x) \in C^2[x_0, x_n]$, with the conditions $p_{i-1}(x_i) = p_i(x_i)$, $p_{i-1}'(x_i) = p_i'(x_i)$, $p_{i-1}''(x_i) = p_i''(x_i)$, $i = 1, 2, \dots, n - 1$. Cardinal splines correspond to an infinite number of knots with unit spacing.

With cubic interpolation, for a given set $(x_i, y_i)_{i=0}^n$, there are $4n$ coefficients $(a_i, b_i, c_i, d_i)_{i=0}^{n-1}$ to determine and a set of $n + 1$ linear equations $h(x_i) = y_i$ plus $3(n - 1)$ continuity equations which gives a total of $4n - 2$ linear equations. So there are 2 possible degrees of freedom that correspond to different types of splines. For example, natural splines are defined by setting $h''(x_0) = 0$ and $h''(x_n) = 0$. Natural splines minimize the integral of the squared second derivative of h .

$S_{d,\mathbf{s}}$ is a vector-space of dimension $n + d$. An interesting basis is that of B-splines (Basis-spline) [10]. B-splines are spline functions with minimum support $[x_i, x_{i+d+1}[$. They can be defined by a recurrence relation for a given degree of polynomials d and a partition \mathbf{s} of the interval $[a, b]$ with $n + 1$ knots and $n > d + 2$. The recurrence relation defining the i^{th} B-spline of degree d , noted $B_{i,d,\mathbf{s}}$ ($\forall x \in \mathcal{R}$) is:

$$B_{i,d,\mathbf{s}}(x) = \frac{x - x_i}{x_{i+d} - x_i} B_{i,d-1,\mathbf{s}}(x) + \frac{x_{i+1+d} - x}{x_{i+1+d} - x_{i+1}} B_{i+1,d-1,\mathbf{s}}(x).$$

$$B_{i,0,\mathbf{s}}(x) = X_i(x) = \begin{cases} 1, & \text{if } x_i \leq x < x_{i+1} \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{And } \sum_i B_{i,0,\mathbf{s}}(x) = 1 \quad \forall x.$$

The B-spline $B_{i,d,\mathbf{s}}$ can be expressed as a concatenation of $d + 1$ successive polynomials pieces of degree d . It is equal to 0 outside the segment $[x_i, x_{i+d+1}[$. And $B_{i,d,\mathbf{s}} = 0$ if $x_i = x_{i+d+1}$. On any interval $[x_i, x_{i+d+1}[$, at most $d + 1$ B-splines are different from 0: $(B_{i-d,d}, \dots, B_{i,d})$. For a finite set of knots in a finite interval $[a, b]$, there are $n - d$ B-splines whose support is completely included in $[a, b]$. To obtain a full basis of B-splines functions on $[a, b]$, we must had $2d$ basis-functions whose support is partly inside $[a, b]$, with usually d nodes equal to a and d nodes equal to b . A knot has a multiplicity m if it is repeated m times in the sequence.

Any spline function $h_{d,\mathbf{s}}$ can be written as a linear combination of B-splines:

$$h_{d,\mathbf{s}}(s) = \sum_i c_i B_{i,d,\mathbf{s}}(x).$$

Figure 5 shows an example of cubic B-splines defined for $x \in [0, 1]$ with a uniform segmentation and a sequence of 9 knots. The basis functions are well localized around the knots and are therefore quasi-orthogonal. It also shows another spline basis proposed in [11] that keeps a very short support when d increases.

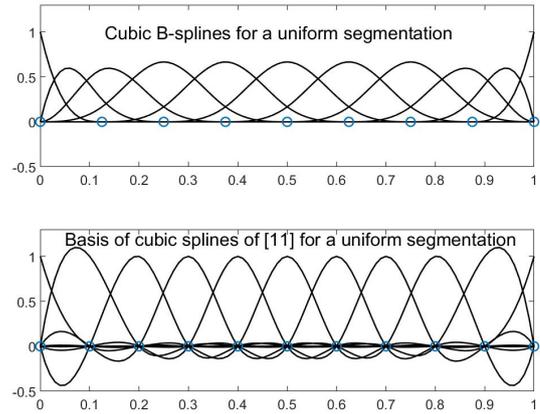


Fig. 5. Cubic B-splines (top with 9 knots) or spline basis of [11] (bottom) with uniform segmentation.

Splines functions can be used to approximate functions generally defined by a set of data points. The approximation can be based on interpolation or on minimizing a given criterion. A commonly used approximation criterion in the field of DPD is the least-square criterion. The approximation by a spline function with LS criterion can be achieved as in (1) and (2). in the case of the approximation of a data sequence of M samples (x_k, y_k) , $k = 0, \dots, M - 1$ with $x_k \in [a, b]$ by a spline function $h(x)$ of degree d and segmentation \mathbf{s} , $h(x)$ can be expressed as a linear combination of B-splines with coefficients c_i :

$$h(x) = \sum_i c_i B_{i,d,\mathbf{s}}(x).$$

For LS criteria $\min \sum_{k=1}^M (h(x_k) - y_k)^2$, the vector \mathbf{c} of coefficients c_i is obtained as the solution of (3), where \mathbf{y} is the $M \times 1$ column vector of samples y_k , $k = 0, \dots, M - 1$ and Φ is the matrix $M \times (n + 1)$ of the B-spline functions.

$$\mathbf{R}\mathbf{c} = \Phi^T \mathbf{y} \quad \text{with } \mathbf{R} = \Phi^T \Phi. \quad (3)$$

3. Generalities on DPD

A general presentation of digital predistortion can be found in [1]. Figure 6 shows a basic transmitter architecture using adaptive DPD.

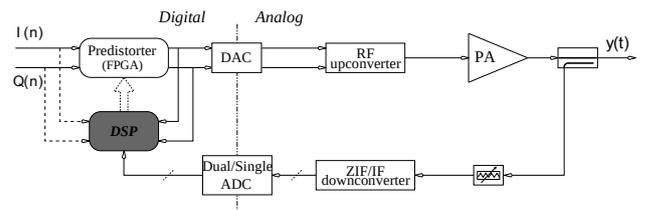


Fig. 6. Transmitter architecture with DPD.

Volterra based models are very common for DPD. Most of them can be expressed as a linear combination of regressors that are derived from the input signal $z(n)$. For example, for the generalized memory polynomial model (GMP), the output $y(n)$ is expressed as:

$$y(n) = \sum_{k=0}^{K_a-1} \sum_{l=0}^{L_a-1} a_{k,l} z(n-l) |z(n-l)|^k + \sum_{k=1}^{K_b} \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} b_{k,l,m} z(n-l) |z(n-l-m)|^k + \sum_{k=1}^{K_c} \sum_{l=0}^{L_c-1} \sum_{m=1}^{M_c} c_{k,l,m} z(n-l) |z(n-l+m)|^k. \quad (4)$$

It can be written as a dot product between the $N_c \times 1$ vector \mathbf{d} of all the coefficients ($a_{k,l}$, $b_{k,l,m}$, $c_{k,l,m}$) and the $N_c \times 1$ regressor vector $\boldsymbol{\phi}_z(n)$, with $y(n) = \boldsymbol{\phi}_z^T(n) \mathbf{d}$. The regressor vector $\boldsymbol{\phi}_z(n)$, built on z , contains the different basis functions ($z(n-l) |z(n-l)|^k$, $z(n-l) |z(n-l-m)|^k$, $z(n-l) |z(n-l+m)|^k$). For a set of N signal samples, the column vector $\mathbf{y} = (y(n), \dots, y(n-N+1))^T$ is equal to:

$$\mathbf{y} = \boldsymbol{\Phi}_z \mathbf{d}, \text{ with}$$

$$\boldsymbol{\Phi}_z = (\boldsymbol{\phi}_z(n), \boldsymbol{\phi}_z(n-1), \dots, \boldsymbol{\phi}_z(n-N+1))^T,$$

$$\boldsymbol{\phi}_z(n) = (\phi_1(n), \phi_2(n), \dots, \phi_{N_c}(n))^T.$$

where $\boldsymbol{\Phi}_z$ is the $N \times N_c$ matrix of regressors, $\phi_i(n)$ is the i^{th} basis function, N_c is the number of coefficients.

There are two general approaches for the identification of coefficients, namely the direct (DLA) and the indirect learning (ILA) approaches.

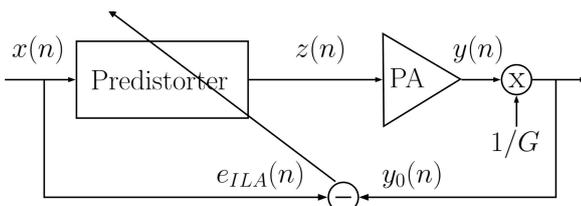


Fig. 7. Principle of DLA direct learning approach.

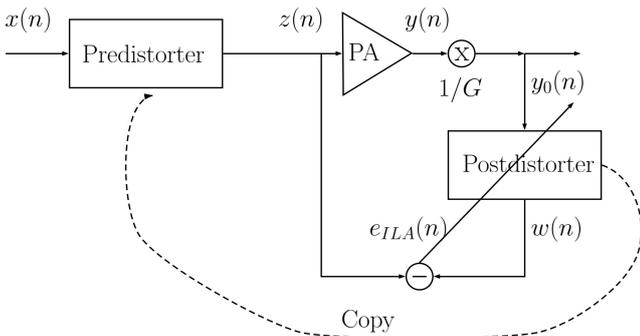


Fig. 8. Principle of ILA direct learning approach.

DLA approach tries to minimize a criterion based on the error $e_{DLA}(n) = x(n) - y_0(n)$ where x is the DPD input and $y_0(n) = y(n)/G$ is the normalized output with a reference gain G (Fig. 7). Many of DLA algorithms are based on first identifying the PA model and then inverting it or use it to train the DPD.

ILA approach (Fig. 8) is based on first solving a post-distortion problem and then using the postdistorter as a pre-distorter. The postdistorter is a fictive block placed after the PA that corrects the normalized PA output in order to minimize a criterion based on error $e_{ILA} = (z(n) - w(n))$ where z in the PA input and $w(n)$ is the postdistorter output (Fig. 8). Using a LS criterion $J = \sum_{i=n-N+1}^n |e_{ILA}(n)|^2$ on N observation samples corresponding to the PA output and input $(y_0(i), z(i))_{i=n-N+1}^n$, the coefficients of the postdistorter that minimize J are solutions of a linear set of equations. The optimum coefficients vector is obtained with the pseudo-inverse of the regression matrix $\boldsymbol{\Phi}_{y_0}$ built on y_0 (5).

$$\boldsymbol{\Phi}_{y_0}^H \boldsymbol{\Phi}_{y_0} \mathbf{d} = \boldsymbol{\Phi}_{y_0}^H \mathbf{z},$$

$$\mathbf{d} = \left(\boldsymbol{\Phi}_{y_0}^H \boldsymbol{\Phi}_{y_0} \right)^{-1} \boldsymbol{\Phi}_{y_0}^H \mathbf{z} \quad (5)$$

where $\boldsymbol{\Phi}_{y_0}^H$ is the hermitian transpose of $\boldsymbol{\Phi}_{y_0}$.

ILA approach is popular because the coefficient identification is a linear optimization problem. However, the measurement noise at the output of the PA introduces a bias on the solution. The DLA approach does not suffer from this drawback but it leads to a nonlinear optimization problem.

Both approaches depends on matrix $\mathbf{R} = \boldsymbol{\Phi}_{y_0}^H \boldsymbol{\Phi}_{y_0}$. Unfortunately, the successive signal samples are usually correlated and therefore matrix \mathbf{R} is badly conditioned, especially when the nonlinearity orders are high. To manage this problem, different methods can be used: regularization techniques such as L_2 norm Tikhonov (ridge regression), orthogonal polynomial basis, orthogonalization (e.g. Gram-Schmidt technique), dimension reduction by suppressing less-significant basis functions (e.g. orthogonal matching pursuit algorithm (OMP)). Another possible approach is to segment the problem into several problems of smaller dimensions. Different types of segmented DPD have been proposed. They are presented in the following sections.

4. Segmented DPD with Functions of a Single Real-Valued Variable

We have seen in Sec. 2 some results about interpolation or approximation of functions by polynomials or piecewise polynomials in the case of functions of a single real-valued variable. But in general, DPD are not such simple functions. First the input signal is complex-valued and secondly the DPD has to take into account the memory effects of the PA. It is ruled by nonlinear differential equations. It can be

simplified by considering that the memory length is finite. The DPD can then be represented by a multivariate function of complex-valued variables. The complexity is much higher than for a function of a single real-value input variable.

Fortunately there are several cases, where simplifications can be done allowing to represent the nonlinear aspects of the DPD by a simple single-variable function of a real variable. This is object of the next two sections.

4.1 Piecewise Modeling in Quasi-Memoryless Models

In this section we will address the cases of quasi-memoryless models and of block oriented models separating nonlinearity from memory effects. These are special cases where single-variable real-valued functions can be used for the DPD model.

The first works on DPD were dedicated to signals with narrow bandwidths for which PA memory effects could be neglected. In that case, the quasi-memoryless PA can be modeled with its AM/AM and AM/PM characteristics as functions of the magnitude (or power) of the input signal $z(n)$. The baseband equivalent of the PA output can be expressed as $y(n) = G_{PA}(|z(n)|)z(n)$ where the magnitude and the phase of the gain G_{PA} respectively represent the AM/AM and AM/PM of the PA. A major breakthrough was realized by Cavers [12] who proposed to model the DPD as a simple complex gain G_{PD} depending only on the magnitude of the input signals and allowing to compensate for the AM/AM and AM/PM distortion of the PA. So, the output $z(n)$ of the DPD, for an input signal $x(n)$ can be written as:

$$\begin{aligned} z(n) &= G_{PD}(|x(n)|)x(n), \text{ with } x(n) = |x(n)| \exp(j\phi_x(n)), \\ G_{PD}(|x(n)|) &= |G_{PD}(|x(n)|)| e^{j\theta_{PD}(|x(n)|)}, \\ z(n) &= |x(n)| |G_{PD}(|x(n)|)| e^{j(\phi_x(n) + \theta_{PD}(|x(n)|))}. \end{aligned}$$

The DPD complex gain should pre-compensate the PA gain so that the PA output be proportional to the input signal x with a reference gain G , which means:

$$\begin{aligned} y(n) &= G_{PA}(|z(n)|) z(n) = G_{PA}(|z(n)|) G_{PD}(|x(n)|) x(n) \\ &= G x(n) \quad \forall x(n). \end{aligned}$$

So: $G_{PA}(|x(n)| G_{PD}(|x(n)|)) G_{PD}(|x(n)|) = G, \forall x(n)$.

The DPD corrective complex gain can be implemented by a Look-up-Table (LUT) [12] or by a polynomial function [13]:

$$G_{DPD}(|x(n)|) = \sum_{k=1}^K a_k |x(n)|^{k-1}. \quad (6)$$

The content of the LUT or the coefficients of the polynomial are adaptively updated by DLA or ILA. In practice, either the DPD is directly updated or the PA gain is first estimated and then inverted to obtain the DPD.

For polynomial DPD, the DPD output is directly obtained by (6). For the LUT case, the LUT is addressed by the

quantized value of some companding function of the input. If we note $(x_k)_{k=0}^{N-1}$ the different possible quantized values, the LUT contains the corresponding gain $G_{PD,k} = G_{PD}(x_k)$. For an input value x , the corrective gain is obtained by some interpolation or approximation of the function $G_{PD}(x)$ from the data $(x_k, G_{PD}(x_k))$.

In order to cope with complicated PA characteristics, it is possible to increase the polynomial order or the LUT size but at the price of introducing ripples in the function (Runge phenomenon), increasing training convergence time, complexity and numerical problems. Therefore, different segmented approaches have been proposed: piecewise linear regression of the AM/AM and AM/PM of the PA gain [14], piecewise polynomial modeling of AM/AM and AM/PM in two regions (with a high order polynomial for the saturation region and a low order one for the linear region) [15], cubic spline interpolation [16], [17], piecewise interpolation by arc of circles [18], piecewise bilinear rational function [19] of the PA gain with piecewise inverse of the PA model to obtain the DPD model. An advantage of piecewise linear and piecewise bilinear function is that they can be easily inverted. For example, in the case of a bilinear function, if the PA model, for an input z and an output y is defined by:

$$\begin{aligned} |y| &= \frac{a_i |z| + b_i}{c_i |z| + d_i}, \quad \arg(y) = \arg(z) + \frac{\phi_i |z| + \psi_i}{\theta_i |z| + \eta_i} \\ |z| &\in R_i, \quad i \in \{1, 2, \dots, N\} \text{ where } R_i \text{ is the segment } N^{\circ}i. \end{aligned}$$

the inverse model used for the DPD (with input x and output z) is given by:

$$|z| = \frac{d_i |x| - b_i}{-c_i |x| + a_i}, \quad \arg(z) = \arg(x) - \frac{\phi_i |z| + \psi_i}{\theta_i |z| + \eta_i}.$$

In most of those studies, the segmentation (number of segments, position of the knots) and degrees of polynomial segments are determined in an empirical way. In [20], authors use AM/AM derivatives to segment it into linear, nonlinear and saturation regions.

Cavers worked on the optimal LUT-spacing [21]. He derived the optimum companding function of input magnitude for table indexing. It depends on the signal statistics and on PA characteristics. He showed that for a class AB PA, equispacing by amplitude is closed to the optimum. In [22], authors propose a non-uniform LUT indexing function that allows for a signal to quantization noise that does not depend on input signal statistics or power backoff of the PA. In [23] a segmented approach is used with more LUT entries in the strongly nonlinear segments than in the linear ones. In [24] the LUT-spacing is dynamically optimized in function of online estimated PA characteristics and the input signal statistics using histograms to approximate the signal statistics. In [25] and in [26], the authors theoretically study respectively the optimal spacing of piecewise linear LUT DPD and of a quadratically interpolated LUT DPD.

4.2 Piecewise Modeling in Block-Oriented Models

In order to take into account PA memory effects, one possible approach is to use block oriented nonlinear models (BONL) separating nonlinearity from memory effects such as Wiener, Hammerstein, Wiener-Hammerstein models. These models associate in cascade or in parallel several linear time-invariant filters (LTI) that represent the dynamics of the system and static nonlinear (SNL) blocks. A Wiener model is made of a cascade of an LTI followed by an SNL block and an Hammerstein model of an SNL block followed by a linear filter. One drawback of BONL models is that the identification of their parameters has to be done by nonlinear optimisation technique. But an interesting point is that SNL blocks can be represented as complex gains that are single-variable functions of their input signal magnitude. And therefore, all the piecewise techniques presented in 4.1 can be applied to the SNL blocks of BONL models.

In [27], the PA is modeled by a Wiener model where the SNL block is represented by a simplicial canonical piecewise linear (SCPWL) function [28]. Two SCPWL functions are used respectively for AM/AM and AM/PM characteristics. The DPD is an Hammerstein model. The inverse of the SNL-block of the PA model is also a piecewise linear function that can be easily obtained by inverse coordinate mapping. In [29], a piecewise linear predistorter is also proposed. the parameters of the SNL block are estimated by particle swarm optimization (PSO).

In [30], the authors apply a Wiener model to the PA. The SNL block of the PA Wiener model comprises an AM/AM characteristic that is modeled by a piecewise linear continuous and monotonically increasing function and an AM/PM characteristic represented by a piecewise constant function. The segmentation is determined empirically. The DPD is modeled by a memory polynomial model. A direct learning adaptive DPD is proposed. First the PA model is identified using RLS (recursive least square) algorithm. There are two parameters to identify for each segment of the SNL block. The inverse of the PA model is easy to obtain. It is an Hammerstein model with a piecewise SNL block. The DPD parameters are obtained by a piecewise RLS (PRLS) algorithm. The error to minimize is calculated thanks to the inverse function of the PA. This segmented approach leads to a direct learning adaptive algorithm that is less complex than common direct learning algorithms but it offers the same level of performance.

In [31], two Hammerstein models with Catmull-Rom cubic spline static nonlinearity are used for the DPD. One for correcting AM/AM distortion and the other for correcting AM/PM distortion. The DPD coefficients are identified using ILA with a separable nonlinear least squares (SNLS) optimization [32]. The segmentation is optimized empirically. The approach allows to compensate for high-degree nonlinearity with a limited set of coefficients.

In [33], we proposed a BONL DPD called "Filtered LUT" or FLUT. The FLUT DPD is made of an SNL block followed by a linear filter. But at the difference of Hammerstein models, the filter coefficients vary with the magnitude of the DPD input signal. The SNL block is implemented by a linear piecewise LUT (gain-LUT). A codebook stores the coefficients of the different filters. Both the gain-LUT and the codebook of filters are indexed by the magnitude of the input signal with a uniform companding function.

4.3 Piecewise Modeling in Volterra Based Models

Models derived from Volterra or from dynamic Volterra series can be reformulated (or generalized) with nonlinear single-variable functions of the input-signal magnitude.

For example, an MP DPD model with input x and output z defined as:

$$z(n) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} a_{k,l} x(n-l) |x(n-l)|^k,$$

can be reformulated with L single-variable nonlinear function $f_{NL,l}$ as:

$$\begin{aligned} z(n) &= \sum_{l=0}^{L-1} x(n-l) \left(\sum_{k=0}^{K-1} a_{k,l} |x(n-l)|^k \right) \\ &= \sum_{l=0}^{L-1} a_{k,l} x(n-l) f_{NL,l}(|x(n-l)|). \end{aligned}$$

These single-variable nonlinear functions $f_{NL,l}$ can be approximated by different piecewise functions; A DPD model where the $f_{NL,l}$ are modeled by complex-valued cubic splines is proposed by Safari et al. in [34] and compared with MP model. The obtained piecewise model is linear with respect to its coefficients and can be identified using LS criterion and by ILA approach with similar equations as (3) and (5). It shows better results than MP model for a smaller number of coefficients. The same kind of approach is proposed in [35] using 2nd order nonlinearity piecewise MP and a single knot.

The same method can be applied to reformulate GMP, DDR or any other Volterra based models. The principle is to first do the summation on the nonlinearity orders. In a GMP model, each of the 3 terms of (4) can be reformulated in the same way, e.g., for the second term:

$$\begin{aligned} &\sum_{k=1}^{K_b} \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} b_{k,l,m} u(n-l) |u(n-l-m)|^k \\ &= \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} x(n-l) \left(\sum_{k=1}^{K_b} b_{k,l,m} |u(n-l-m)|^k \right) \\ &= \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} x(n-l) f_{NL,l,m}(|x(n-l-m)|). \end{aligned}$$

There are ML single-variable nonlinear functions $f_{NL,l,m}$.

Barradas et al. in [9] show how this approach can be applied to any Volterra based model. They approximate the $f_{NL,l,m}$ by cubic splines functions constructed with B-splines. They suggest to use Tchebychev nodes as splines knots. They show that the numerical stability (matrix conditioning) of ILA identification with this approach is much better than with GMP model for similar performance and complexity. In [7] they expound that the high locality (limited support) of nonlinear splines basis makes them quasi-orthogonal which explains their better numerical properties. They develop their analysis by a theoretical comparison of polynomial and LUTs in PA modeling.

This approach is applied in [36] with $f_{NL,l,m}$ functions approximated by piecewise Lagrange (APL) basis functions.

Authors of [37] consider the determination of DPD as a multivariate regression problem and use the fact that any multivariate function can be approximated by a sum of separable functions to express a very general form of DPD with single-variable real-valued nonlinear functions. These functions can then be piecewise approximated by splines.

In [38], Zhu et al. propose a different approach to cope with envelop tracking PA, the behavior of which changes significantly in functions of power region. The technique is called "Decomposed Piecewise Volterra Series". It is based on vector threshold decomposition of the input signal $x(n) = |x(n)|e^{j\phi}$. It was initially proposed by Heredia for real-valued signals in [39]. For a given set of real positive increasing thresholds $(\lambda_i)_{i=1}^N$ and $\lambda_0 = 0$, N sub-signals x_i are obtained, with:

$$x_i(n) = \begin{cases} 0, & |x(n)| \leq \lambda_{i-1} \\ (|x(n)| - \lambda_{i-1})e^{j\phi}, & \lambda_{i-1} < |x(n)| \leq \lambda_i \\ (\lambda_i - \lambda_{i-1})e^{j\phi}, & |x(n)| > \lambda_i \end{cases}$$

$$\text{with } x(n) = \sum_{i=1}^N x_i(n).$$

Every sub-signal is processed by a specific sub-DPD model (here DDR models are used). And the outputs of all these DPD are summed to obtain the global predistorted signal. In [40], the authors apply GMP models for the sub-DPD and the thresholds are determined using the slope and the rate of slope of the AM/AM characteristic. In [41], the same technique is used with a learning algorithm that decorrelates the DDR polynomial basis functions and is applied on each sub-DPD independently.

5. Segmented DPD for Nonlinearity and Memory Domains

In Sec. 4, the segmentation is applied on the nonlinearities only. But for some types of PA, e.g. Doherty PA, the memory effects are different at different power levels. In such cases, it may be useful to also segment the memory domain. In this section, we present different approaches that are not limited to the piecewise approximation of single-variable

functions for nonlinearities. These approaches partition the global space of the input signals and fits a piecewise DPD to each of the regions of the partition. In this section, we present three techniques: Vector-Switched (VS) DPD, Continuous Piecewise Linear (CPWL) DPD and Decomposed Vector Rotation (DVR).

5.1 Switched DPD

Switched DPD allows to derive piecewise DPD taking into account nonlinearity and memory effects and applying the segmentation in both domains. The principle consists in switching several DPD models, each of them being well suited to a specific segment (or region) of the input signal. The continuity between the different models is a delicate notion. There are at least two questions: how should this continuity be defined and is it really necessary. Indeed, we can consider continuity with time and continuity with magnitude. When the models include some kind of memory, the final condition of one model can be used as the initial conditions of the next model, which ensures some kind of temporal continuity.

Switched DPD allows to use in each region a DPD model of smaller complexity (nonlinearity orders, memory lengths) than would be necessary with a global model. It has the ability to represent hard nonlinearities and to identify the region-DPD models with good numerical stability.

In [42], Afsardoost et al. propose the vector-switched (VS) DPD model. the VS model is a set of DPD models that can be switched and applied to the input according to some switching function based on the region of the input signal. The space of input signals $x(n)$ is partitioned into N regions using vector quantization (VQ). VQ is applied to vectors of Q successive complex input samples $\mathbf{X}^{(Q)}(n) = \{x(n), x(n-1), \dots, x(n-Q+1)\}$. For an input sample $x(n)$, the class of $\mathbf{X}^{(Q)}(n)$ is determined and the corresponding model is chosen for $x(n)$. Authors note that it is generally sufficient to use $Q = 2$ and to apply VQ on the input magnitude only.

Authors of [41] apply this approach with a set of GMP models and a learning algorithm that decorrelates the GMP polynomial basis functions. This training is applied on each DPD model independently using the input samples of the corresponding region. The VQ segmentation is achieved on the magnitude of the input signal with $Q = 1$.

5.2 From PWL to Memory-SCPWL DPD

Piecewise linear functions and in particular simplicial canonical piecewise linear (SCPWL) functions were first applied to approximate functions of a single real-variable variable such as those presented in Sec. 4, in static cases to model AM/AM and AM/PM characteristics of quasi-memoryless PA [43], [44] or in dynamic cases to represent nonlinearities functions in Volterra based models [45]. Then they were also used to model functions of complex input [46] with the

name memory-SCPWL functions. A PWL representation has several interests. In particular, thanks to its linear affine property, it can be inverted very easily. A PWL function can be described segment per segment, which may require a large number of coefficients. The number of coefficients can be reduced by using a global representation called Canonical PWL (CPWL) function [47] or an even more compact form called Simplicial Canonical PWL (SCPWL) function [28]. A SCPWL function f of a single real-variable x , is given by:

$$f(x) = c_0 + \sum_{i=1}^{\sigma-1} c_i \lambda_i(x), \quad x \in \mathcal{R}$$

where σ is the number of segment breakpoints, c_i are coefficients, and $\lambda_i(x)$ are basis functions defined with a set of increasing breakpoints values $(\beta_i)_{i=1}^{\sigma}$ by:

$$\lambda_i(x) = \begin{cases} \frac{1}{2}(x - \beta_i + |x - \beta_i|), & x < \beta_\sigma \\ \frac{1}{2}(\beta_\sigma - \beta_i + |\beta_\sigma - \beta_i|), & x \geq \beta_\sigma. \end{cases}$$

The coefficients c_i can be complex-valued.

Cheong et al. in [46] modified that expression to make it suitable for modeling nonlinearities and memory effects. For a given memory length L , considering the last L input samples $(x(n-l))_{l=0}^L$, the new form is:

$$f(x) = \sum_{l=0}^L \left[c_{l,0} + \sum_{i=1}^{\sigma-1} c_{l,i} \lambda_i(|x(n-l)|) e^{j \arg(x(n-l))} \right]. \quad (7)$$

This new form is linear with respect to its coefficients $c_{l,i}$ which simplifies its identification. Replacing $|x - \beta_k|$ where x is real-valued by $||x| - \beta_k| e^{j \arg(x)}$ where $x = |x| e^{j \arg(x)}$ is complex-valued, is called Decomposed Vector Rotation (DVR). Authors of [46] compared the memory-SCPWL DPD with MP and GPM DPD. They showed that memory-SCPWL DPD offers better modeling accuracy for sharp nonlinearities and that it is less sensible to noise at PA output thus reducing the potential bias on coefficients for ILA identification.

Of course, CPWL functions can also be used with the generalized form of Volterra DPD using function of single real-valued variables [48].

5.3 Decomposed Vector Rotation (DVR) Models for DPD

Zhu [49] has extended the memory-SCPWL model given by (7) with a more general model called Decomposed Vector Rotation DVR model. It starts from the representation by CPWL of a finite memory nonlinear system with real-valued input and output signals z and y . CPWL achieves a partition of the input signal space into K polyhedral regions separated by hyperplanes whose boundaries are defined with thresholds $(\beta_k)_{k=1}^K$. The input-output relation is expressed as:

$$y(n) = \sum_{l=0}^L a_l x(n-l) + b + \sum_{k=1}^K c_k \left| \sum_{l=0}^L a_{k,l} x(n-l) - \beta_k \right|$$

where L is the memory length. The hyperlanes are defined by $\sum_{l=0}^L a_{k,l} x(n-l) - \beta_k = 0$.

CPWL can approximate a wide range of continuous nonlinear function with a very good accuracy [50]. Unfortunately, this model is not linear with respect to its coefficients (a_l, b, c_k) and it is not directly usable for complex-valued input signal x . To overcome these two limitations, Zhu proposed a new formulation:

$$y(n) = \sum_{l=0}^L a_l x(n-l) + b + \sum_{k=1}^K \sum_{l=0}^L c_{k,l} ||x(n-l)| - \beta_k| e^{j \arg(x(n-l))}.$$

In order to introduce interaction between signals at different time instants, some other terms can be added to the model. Depending on these added terms, the model is more or less complex. One possible expression is:

$$y(n) = \sum_{l=0}^L a_l x(n-l) + b + \sum_{k=1}^K \sum_{l=0}^L c_{k,l,1} ||x(n-l)| - \beta_k| e^{j \arg(x(n-l))} + \sum_{k=1}^K \sum_{l=0}^L c_{k,l,21} ||x(n-l)| - \beta_k| e^{j \arg(x(n-l))} |x(n)| + \sum_{k=1}^K \sum_{l=1}^L c_{k,l,22} ||x(n-l)| - \beta_k| x(n) + \sum_{k=1}^K \sum_{l=1}^L c_{k,l,23} ||x(n-l)| - \beta_k| x(n-l) + \sum_{k=1}^K \sum_{l=1}^L c_{k,l,24} ||x(n-l)| - \beta_k| x(n-l) + \dots \quad (8)$$

Experimental tests on envelop tracking and doherty PA have shown the very good performance of DVR models. In those tests the threshold β_k were fixed and uniformly spaced. But an optimization of their values should improve the results.

In [51], Zhu discusses the respective interests of Volterra series and CPWL functions for DPD.

Authors of [52] propose a modification of the DVR model avoiding the calculation of absolute values. And the same kind of modified model is used in [53] for linearization of radio over fiber link.

6. Advanced Segmented DPD for Multidimensional DPD

In this section, we briefly discuss some more advanced points such as application of segmented DPD to systems using carrier aggregation with concurrent multi-band transmit-

ters or MIMO techniques. Both cases correspond to multiple-inputs DPD (or multidimensional DPD).

In systems using carrier aggregation with concurrent multi-band transmitters, the input signal includes several frequency bands that can be widely spaced in the non-contiguous intra-band or inter-band aggregation. It is possible to consider the multi-band signal as a unique one-dimensional (1D) single-band signal and to design the DPD accordingly but it will result in the use of very high sampling frequency and lead to very high complexity. Another better approach consists in splitting the 1D signal into N sub-signals corresponding to each of the N band. This is a segmentation in the frequency domain. The sub-signals have a much narrower bandwidth than the 1D original signal and can be separately digitized at a reasonable sampling rate. Separate feedback paths are used for the different sub-signals. Most often, $N = 2$ or 3 . The N -dimension DPD must take into account the different sub-signals to construct the predistorted signals corresponding to each carrier and suppress the cross-distortion generated between the sub-signals [54]. For example, for a 2D-DPD with two input sub-signals x_1 and x_2 and two output predistorted sub-signals z_1 and z_2 , a possible DPD model is:

$$z_1(n) = \sum_{m=0}^{M-1} \sum_{k=0}^N \sum_{j=0}^k c_{k,j,m}^{(1)} x_1(n-m) |x_1(n-m)|^{k-j} |x_2(n-m)|^j$$

$$z_2(n) = \sum_{m=0}^{M-1} \sum_{k=0}^N \sum_{j=0}^k c_{k,j,m}^{(2)} x_2(n-m) |x_2(n-m)|^{k-j} |x_1(n-m)|^j.$$

This multi-dimensional MP model have a large number of coefficients. Therefore it can be interesting to use piecewise models to decrease the number of necessary coefficients. Naraharisetti et al. [11], [55] have reformulated the model in a more general form that is expressed for z_i with $i = 1$ or 2 as:

$$z_i(n) = \sum_{m=0}^{M-1} G_m^{(i)} \left(|x_1(n-m)|^2, |x_2(n-m)|^2 \right) x_i(n-m).$$

The nonlinear gain functions $G_m^{(i)}$ are functions of the two real-valued signals $|x_1|$ and $|x_2|$. They can be approximated by different piecewise functions and in particular using spline basis functions.

2D cubic splines basis are used in [11], [55] and extended to 3D cubic spline basis in [56]. The coefficients of the cubic spline basis are obtained by LS fitting of the measured data. Using the cubic spline basis functions $\phi_{j,k}$, the gain functions are approximated by:

$$G_m^{(i)} \left(|x_1|^2, |x_2|^2 \right) = \sum_{j=0}^{N_{s1}} \sum_{k=0}^{N_{s2}} c_{ijk}^{(m)} \phi_{j,k}(|x_1|^2, |x_2|^2).$$

Naraharisetti proposed the following formulation for the splines basis [11], [55]. For a given set of 2D-knots

$(|x_{1,u}|^2, |x_{2,v}|^2)$, the basis functions $\phi_{j,k}(|x_1|^2, |x_2|^2)$ are created with 2D interpolation cubic spline such that:

$$\begin{aligned} \phi_{j,k}(|x_{1,u}|^2, |x_{2,v}|^2) &= \delta_{j,u} \delta_{k,v} \\ &= \begin{cases} 1 & \forall (j,k) = (u,v) \\ 0 & \forall \text{ other knots.} \end{cases} \end{aligned}$$

The 2D cubic spline basis are built from 1D Cubic spline basis by tensor product. The number of coefficients is equal to $MN_{s1}N_{s2}$.

In [57], instead of segmenting the model using single input magnitude, the partitioning is done using vector quantization on combined input magnitudes. In [58], [59], 2D CPWL models are proposed.

The case of DPD for MIMO transmitters is quite similar to that of multi-band transmitters. nonlinear crosstalk effects due to coupling at the inputs of the elementary PAs create cross-modulation of the elementary signals. In [60], multivariate polynomial models are used. To reduce the complexity of such DPD, different piecewise approaches are studied. For example in [61], [62] Dual-Input Canonical Piecewise-Linear DPD are presented for MIMO applications with two antennas.

A new promising approach for multidimensional DPD with Multiple input multiple outputs, called Tensor-Network-Based DPD, is proposed in [63]. This new development should facilitate the identification of Volterra series with very high memory lengths and memory orders. They are based on the tensor-network based MIMO Volterra system framework of [64].

7. Experimental Comparisons

In this section, we give some experimental comparisons of three types of DPD: a non-segmented GMP model, a vector-switched DPD and a DVR DPD.

The experiments have been done with a strongly nonlinear 3-way Doherty PA based on three LDMOS transistors (BLF7G22LS-130 from Ampleon). Its maximum peak output power is 57 dBm. Its linear gain is 16 dB. The PA is preceded by a driver with a gain of 31.5 dB. The used carrier frequency is 2.14 GHz. The input signal is an LTE with a bandwidth of 20 MHz and a PAPR of 8 dB. The tests were achieved at an average PA output power equal to 47.3 dBm. For the presented results, training has been done with ILA.

7.1 Non-Segmented GMP-DPD

The structure of the GMP model (values of the eight parameters $(K_a, L_a, K_b, L_b, M_b, K_c, L_c, M_c)$) is determined by an hill-climbing (HC) optimization algorithm that we proposed in [65]. This algorithm optimizes a trade-off between the number of coefficients and the accuracy of the model. The obtained GMP structure is given by:

($K_a = 11, L_a = 2, K_b = 4, L_b = 1, M_b = 1, K_c = 3, L_c = 4, M_c = 1$). It has 38 coefficients. The obtained normalized mean square error (NMSE) is equal to -32.7 dB. NMSE is defined as the ratio between the power of the error and the power of the output signal. The AM/AM and AM/PM characteristics with and without the GMP DPD are given in Fig. 9.

To improve the results, we can try to increase the number of coefficients to the detriment of complexity. But the HC algorithm shows that the results cannot be significantly improved by increasing the number of coefficients. For example with 80 coefficients, the NMSE is only equal to -33.9 dB (only 1 dB better than with 38 coefficients).

By inspection of the AM/AM characteristic, we can clearly distinguish different areas with different slopes (gains). The segmented approaches may be good candidates to improve performance without significantly increasing the complexity.

7.2 Vector-Switched DPD (VS-DPD)

To apply Switched-Vector DPD we have to determine the good segmentation and the structures of models in each segment. We have chosen GMP models for each segment (or VQ class) with 14 coefficients and a maximum order of nonlinearity equal to 5 (instead of 38 coefficients and a maximum nonlinearity of 11 in the global GMP model). We used the same structure for all of the segment-models because the FPGA implementation will be sized by the most complex of those structures.

We have compared several types of segmentation:

- scalar uniform quantization, scalar non-uniform Lloyd-Max segmentation, vector-quantization (VQ) with different vector dimensions,
- training of the segmentation codebook on the PA input or output signal,
- segmentation determined by the signals or by the PA characteristics.

For each type of segmentation, we have tested different values for the number of segments

We observed that segmentation determined by the signals gives better results than segmentation driven by PA characteristics (slope variation in the AM/AM characteristic). So we focussed on segmentation determined by the signals. Learning the VQ codebook with the PA normalized output signal y_0 is slightly better than with the original input signal x . But it is easier to train with the input signal so we trained VQ on input signal. Moreover, quantizing complex signals does not improve results compared to quantizing signal magnitude. So we applied quantization on signal magnitude.

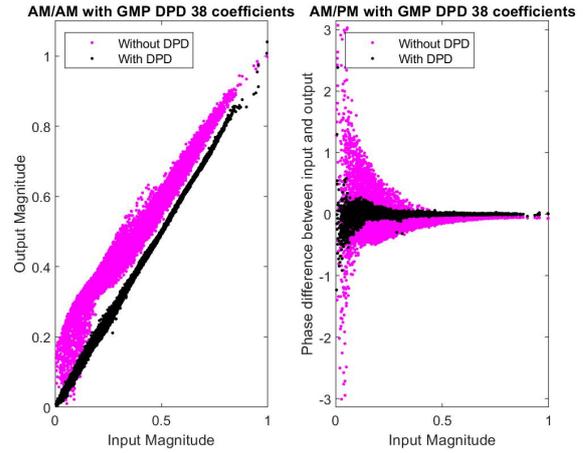


Fig. 9. AM/AM and AM/PM characteristics with and without the GMP DPD (38 coefficients).

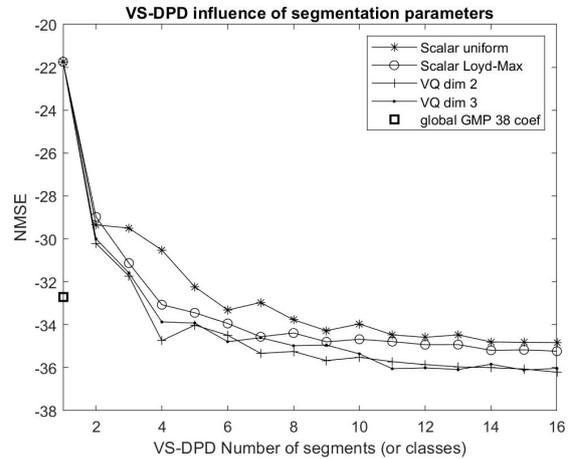


Fig. 10. Influence of the number of segments on NMSE.

Figure 10 illustrates how varies the NMSE in function of the number of segments N_S for different types of quantization and VQ vector dimension. It can be seen that for scalar quantization and $N_S > 10$, using Lloyd-Max quantizer instead of uniform segmentation only slightly improves the results. But using VQ segmentation with a vector size equal to 2 or 3 instead of scalar segmentation, improves the NMSE by approximately 1 dB. A vector size equal to 2 is sufficient. The results improve slowly when increasing the number of classes but this number must remain small enough in order that in each training buffer of N samples, the population size of each class is large enough for a good identification of the class-model. On the same Fig. 10, we have added the result obtained by the global GMP-DPD with 38 coefficients. We see that it is possible to achieve a similar NMSE with VS-DPD using 4 segments and VQ Segmentation or Lloyd-Max scalar segmentation each segment corresponding to a model of 14 coefficients.

Figures 11 and 12 show the result of the segmentation of AM/AM characteristic into respectively 12 and 4 classes with VQ (vector size = 2) or scalar Lloyd-Max quantization.

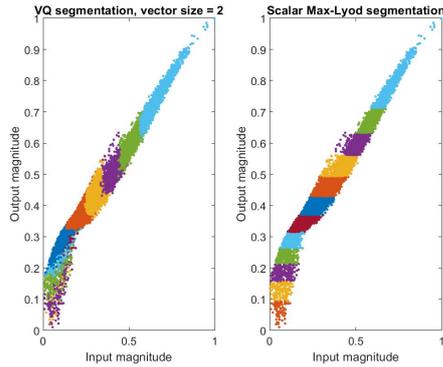


Fig. 11. Segmentation of the AM/AM characteristic into 12 classes. Left: 2D VQ, right: scalar Max-Lyod.

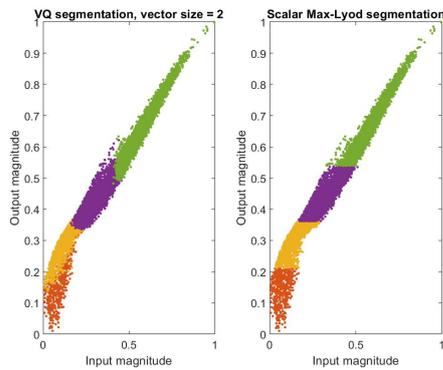


Fig. 12. Segmentation of the AM/AM characteristic into 4 classes. Left: 2D VQ, right: scalar Max-Lyod.

Concerning the implementations, for VS-DPD with 12 segments we have to store 168 (12×14) coefficients instead of 38 coefficients for global GMP-DPD but this is not a problem because it remains very small compared to the memory size of common FPGA. But VS-DPD has many advantages compared to global GMP-DPD. First, the real-time computation complexity of the DPD is much reduced for VS-DPD (14 coefficients instead of 38). Secondly, the identification of coefficients is greatly facilitated, since the covariance matrix \mathbf{R} of (5) has much smaller dimensions and is better conditioned (there is a ratio of $\approx 10^5$ between the two condition numbers). Thirdly, the dynamic of coefficients is strongly reduced. For VS-DPD, the ratio between the magnitudes of the largest and the smallest coefficients is smaller than 500 for all the models while it is around $4e5$ for global GMP-DPD. This last point is important for fixed point implementation of DPD. The only small drawback of SV-DPD compared to global GMP-DPD is that each input sample has to be quantified in order to determine its class and the coefficients of the DPD have to be modified for each signal sample.

7.3 Decomposed Vector Rotation DVR-DPD

For DVR-DPD we have to determine the DPD structure: number of segments (or number of thresholds), memory depths and terms that we keep in the model given by (8). Indeed in (8) there are 6 elementary types of basis functions,

but may be some of them are not usefull and some other could be added. In [66] we have proposed an algorithm based on hill-climbing heuristic for the sizing of DVR models. This algorithm searches for a structure that optimizes a trade-off between the number of coefficients of the model and its modeling accuracy. The thresholds are equi-spaced but it would be interesting to optimize their values.

Figure 13 shows the influence of the number of segments K and of the memory depth M on the NMSE. We can observe that the curves are not monotonically decreasing when the number of segments increases. May be an optimization of the threshold values would made the curves more regular.

Figure 14 shows the number of DVR coefficients in function of K and M . The number of coefficients N_C is equal to $M + 1 + 2K(M + 1) + 3KM = 5KM + M + 2K + 1$.

We can notice that the obtained NMSE values are quite similar to those of VS-DPD for the same number of segments. For example, for $N_S = 12$ segments, the NMSE are close to -36 dB in both cases with $M = 3$ for DVR. For SV-DPD each model has 14 coefficients and we have to store 12×14 coefficients in memory. For DVR, the model is the same for each sample. The DVR model has 208 coefficients. The condition number of that matrix is 10^5 times greater than that of the global GMP-model with 38 coefficients. The dynamic of magnitude of coefficients is more important for DVR (≈ 1800) than for VS-DPD (≈ 500) but it much smaller than that of global GMP-DPD. For DVR, the identification of coefficients requires to deal with a 208×208 matrix \mathbf{R} while for SV-DPD there are 12 identifications to do (one per segment), each of them with a 14×14 matrix \mathbf{R} . Therefore identification step is less complex for VS-DPD.

Compared to the global GMP-DPD with 38 coefficients, the NMSE is improved by 3 dB with SV-DPD (12 segments and 14 coefficients per segment) and by 3.5 dB with DVR (12 segments and memory length = 3).

Figure 15 shows the normalized power spectral densities obtained with the different DPD (the sampling frequency is equal to 200 MHz) with 12 segments for VS or DVR DPD.

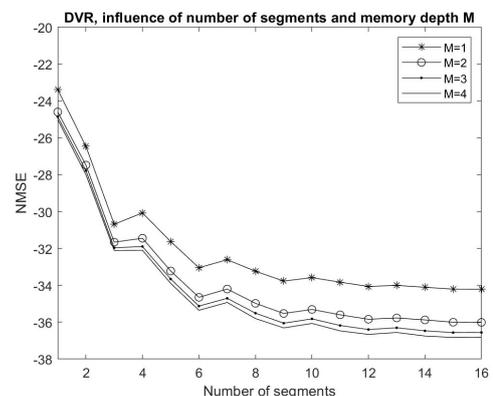


Fig. 13. Influence of the number of segments on NMSE.

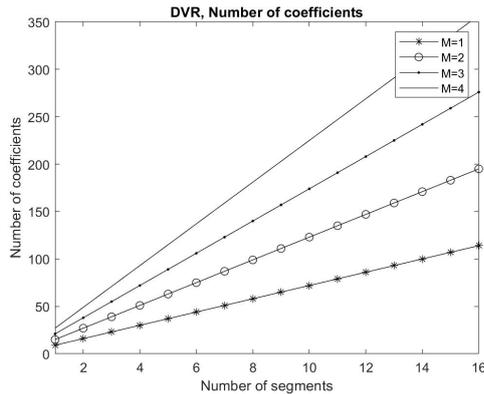


Fig. 14. Number of coefficients in function of the number of segments and of the memory length.

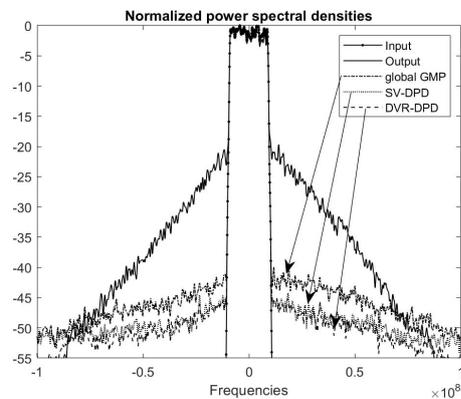


Fig. 15. Normalized power spectral densities, 12 segments.

8. Conclusion

This paper is a survey of predistortion techniques using segmentation.

A comparison of global GMP DPD with two segmented approaches namely Vector-Switched DPD (VS-DPD) and Decomposed Vector Rotation DPD (DVR-DPD) is presented with the support of experimentation on a strongly nonlinear 3 ways Doherty PA. It shows the interest of both segmented approaches in terms of linearization performance, complexity and ease of implementation compared to the global GMP-DPD. VS-DPD and DVR-DPD improves the NMSE by more than 3 dB. The VS-DPD shows very good numerical properties both for fixed-point real-time implementation and for identification of coefficients. The dynamic of the coefficients of the segmented SV and DVR DPD is very small compared to that of the global GMP DPD. The numerical conditioning of SV-DPD is very good.

There remain several paths to explore in the area of segmented DPD, such as optimizing the segmentation, using tensor algorithms for multivariate DPD, studying the potentialities of local model networks, association with neural nets.

References

- [1] BAUDOIN, G., VENARD, O., PHAM D. G. Digital predistortion. In *Digitally Enhanced Mixed Signal Systems, Materials, Circuits & Devices*, pages 65–123. Institution of Engineering and Technology, 2019. DOI: 10.1049/PBCS040E_ch3
- [2] WOOD J. *Behavioral Modeling and Linearization of RF Power Amplifiers*. United States: Artech House Publishers, 2014. ISBN: 978-1-60807-120-3
- [3] DING, L., ZHOU, G. T., MORGAN, D. R., et al. A robust digital baseband predistorter constructed using memory polynomials. *IEEE Transactions on Communications*, 2004, vol. 52, no. 1, p. 159–165. DOI: 10.1109/TCOMM.2003.822188
- [4] MORGAN, D.R., MA, Z., KIM, J., et al. A generalized memory polynomial model for digital predistortion of RF power amplifiers. *IEEE Transactions on Signal Processing*, 2006, vol. 54, no. 10, p. 3852–3860. DOI: 10.1109/TSP.2006.879264
- [5] GOTTHANS, T., BAUDOIN, G., MBAYE, A. Comparison of modeling techniques for power amplifiers. In *Proceedings of the 23rd International Conference Radioelektronika (RADIOELEKTRONIKA)*. Pardubice (Czech Republic), 2013, p. 232–235. DOI: 10.1109/RadioElek.2013.6530922
- [6] ZHU, A., PEDRO, J. C., BRAZIL, T. J. Dynamic deviation reduction-based volterra behavioral modeling of RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 2006, vol. 54, no. 12, p. 4323–4332. DOI: 10.1109/TMTT.2006.883243
- [7] BARRADAS, F. M., CUNHA, T. R., LAVRADOR, P. M., et al. Higher locality non-linear basis functions of Volterra series based models to improve extraction conditioning. In *IEEE MTT-S International Microwave Symposium (IMS2014)*. Tampa (USA), 2014, p. 1–4. DOI: 10.1109/MWSYM.2014.6848447
- [8] BARRADAS, F. M., CUNHA, T. R., LAVRADOR, P. M., et al. Polynomials and LUTs in PA behavioral modeling: A fair theoretical comparison. *IEEE Transactions on Microwave Theory and Techniques*, 2014, vol. 62, no. 12, p. 3274–3285. DOI: 10.1109/TMTT.2014.2365188
- [9] BARRADAS, F. M., CUNHA, T. R., LAVRADOR, P. M., et al. Using spline basis functions in Volterra series based models. In *2014 International Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMiC)*. Leuven (Belgium), 2014, p. 1–3. DOI: 10.1109/INMMiC.2014.6815079
- [10] BOOR, DE C. *B(asic)-Spline Basics*. MRC Technical Summary Report #2952. Mathematics Research Center, University of Wisconsin-Madison, 1986. Available at: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a172773.pdf>
- [11] NARAHARISSETTI, N., ROBLIN, P., QUINDROIT, C., et al. Efficient least-squares 2-d-cubic spline for concurrent dual-band systems. *IEEE Transactions on Microwave Theory and Techniques*, 2015, vol. 63, no. 7, p. 2199–2210. DOI: 10.1109/TMTT.2015.2435731
- [12] CAVERS, J. K. Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements. *IEEE Transactions on Vehicular Technology*, 1990, vol. 39, no. 4, p. 374–382. DOI: 10.1109/25.61359
- [13] BAUDOIN, G. JARDIN, P. Adaptive polynomial pre-distortion for linearization of power amplifiers in wireless communications and WLAN. In *International Conference on Trends in Communications. (EUROCON)*. Bratislava (Slovakia), 2001, p. 157–160 DOI: 10.1109/EURCON.2001.937787
- [14] SAFARI, N., TANEM, J. P., ROSTE, T. Block based predistortion for amplifier linearization in burst type mobile satellite communications. In *The European Conference on Wireless Technology (ECWT)*. Paris (France), 2005, p.325–328. DOI: 10.1109/ECWT.2005.1617723

- [15] JIN, M., SHIN, S. K., OH, D. A novel predistorter for power amplifier. In *3rd International Conference on Microwave and Millimeter Wave Technology (ICMMT)*. Beijing (China), 2002, p. 1129–1133. DOI: 10.1109/ICMMT.2002.1187906
- [16] LOHTIA, A., GOUD, P. A., ENGLEFIELD, C. G. Power amplifier linearization using cubic spline interpolation. In *IEEE 43rd Vehicular Technology Conference (VETEC)*. Secaucus (USA), 1993, p. 676–679. DOI: 10.1109/VETEC.1993.508782
- [17] ACCIARI, G., GIANNINI, F., LIMITI, E., et al. Baseband predistortion lineariser using direct spline computation. In *32nd European Microwave Conference (EuMC)*. Milan (Italy), 2002, p. 1–4. DOI: 10.1109/EUMA.2002.339229
- [18] HUSSEIN, M. A., WANG, Y., FEUVRIE, B., et al. Piecewise complex circular approximation of the inverse characteristics of power amplifiers for digital predistortion techniques. In *International Conference on Digital Telecommunications (ICDT)*. Bucharest (Romania), 2008, p. 59–63. DOI: 10.1109/ICDT.2008.33
- [19] YANG, W., ZHOU, S., ZHU, W. Adaptive predistortion technology based on piecewise nonlinear function. In *IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions (ICCASCAS)*. Chengdu (China), 2002, p. 478–482. DOI: 10.1109/ICCASCAS.2002.1180663
- [20] FISHER, P. O., AL-SARAWI, S. F. Memoryless AM/AM behavioral model for RF power amplifiers. In *3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*. Nadi (Fiji), 2016, p. 131–138. DOI: 10.1109/APWC-on-CSE.2016.031
- [21] CAVERS, J. K. Optimum table spacing in predistorting amplifier linearizers. *IEEE Transactions on Vehicular Technology*, 1999, vol. 48, no. 5, p. 1699–1705. DOI: 10.1109/25.790551
- [22] BOUMAIZA, S., LI, J., JAIDANE-SAIDANE, M., et al. Adaptive digital/RF predistortion using a nonuniform LUT indexing function with built-in dependence on the amplifier nonlinearity. *IEEE Transactions on Microwave Theory and Techniques*, 2004, vol. 52, no. 12, p. 2670–2677. DOI: 10.1109/TMTT.2004.837313
- [23] HASSANI, J. Y., KAMAREI, M. A flexible method of LUT indexing in digital predistortion linearization of RF power amplifiers. In *IEEE International Symposium on Circuits and Systems (ISCAS)*. Sydney (Australia), 2001, vol. 1, p. 53–56. DOI: 10.1109/ISCAS.2001.921786
- [24] LIN, C.-H., CHEN, H.-H., WANG, Y.-Y., et al. Dynamically optimum lookup-table spacing for power amplifier predistortion linearization. *IEEE Transactions on Microwave Theory and Techniques*, 2006, vol. 54, no. 5, p. 2118–2127. DOI: 10.1109/TMTT.2006.872808
- [25] BA, S. N., WAHEED, K., ZHOU, G. T. Optimal spacing of a linearly interpolated complex-gain LUT predistorter. *IEEE Transactions on Vehicular Technology*, 2010, vol. 59, no. 2, p. 673–681. DOI: 10.1109/TVT.2009.2034749
- [26] SELVADURAI, D., SIDEK, R. M., VARAHRAM, P., et al. A robust non-uniform indexation of a quadratically interpolated LUT predistorter for RF power amplifiers. In *IEEE 12th Malaysia International Conference on Communications (MICC)*. Kuching (Malaysia), 2015, p. 329–332. DOI: 10.1109/MICC.2015.7725457
- [27] CHEONG, M. Y., WERNER, S., LAAKSO, T. I., et al. Predistorter design employing parallel piecewise linear structure and inverse coordinate mapping for broadband communications. In *14th European Signal Processing Conference (EUSIPCO)*. Florence (Italy), 2006, p. 1–5.
- [28] JULIAN, P., DESAGES, A., AGAMENNONI, O. High-level canonical piecewise linear representation using a simplicial partition. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 1999, vol. 46, no. 4, p. 463–480. DOI: 10.1109/81.754847
- [29] LIU, H., LIU, X. Adaptive piecewise linear predistorter based on PSO and indirect learning architecture. In *2nd International Congress on Image and Signal Processing (CISP)*. Tianjin (China), 2009, p. 1–3. DOI: 10.1109/CISP.2009.5304678
- [30] CHOI, S., JEONG, E., LEE, Y. H. Adaptive predistortion with direct learning based on piecewise linear approximation of amplifier nonlinearity. *IEEE Journal of Selected Topics in Signal Processing*, 2009, vol. 3, no. 3, p. 397–404. DOI: 10.1109/JSTSP.2009.2020265
- [31] WU, X., ZHENG, N., YANG, X., et al. A spline-based Hammerstein predistortion for 3g power amplifiers with hard nonlinearities. In *2nd International Conference on Future Computer and Communication (ICFCC)*. Wuhu (China), 2010, p. V3-741–745. DOI: 10.1109/ICFCC.2010.5497423
- [32] DEMPSEY, E. J., WESTWICK, D. T. Identification of Hammerstein models with cubic spline nonlinearities. *IEEE Transactions on Biomedical Engineering*, 2004, vol. 51, no. 2, p. 237–245. DOI: 10.1109/TBME.2003.820384
- [33] JARDIN, P., BAUDOIN, G. Filter lookup table method for power amplifier linearization. *IEEE Transactions on Vehicular Technology*, 2007, vol. 56, no. 3, p. 1076–1087. DOI: 10.1109/TVT.2007.895566
- [34] SAFARI, N., FEDORENKO, P., KENNEY, J. S., et al. Spline-based model for digital predistortion of wide-band signals for high power amplifier linearization. In *IEEE/MTT-S International Microwave Symposium*. Honolulu (USA), 2007, p. 1441–1444. DOI: 10.1109/MWSYM.2007.380504
- [35] GIDONI, T., SOCHER, E., COHEN, E. Digital predistortion using piecewise memory polynomial for 802.11 wifi applications. In *IEEE International Conference on the Science of Electrical Engineering (ICSEE)*. Eilat (Israel), 2016, p. 1–3. DOI: 10.1109/ICSEE.2016.7806082
- [36] ZHANG, L., FENG, Y. An improved digital predistortion in wide-band wireless transmitters using an under-sampled feedback loop. *IEEE Communications Letters*, 2016, vol. 20, no. 5, p. 910–913. DOI: 10.1109/LCOMM.2016.2546257
- [37] JIANG, H., WILFORD, P. A. Digital predistortion for power amplifiers using separable functions. *IEEE Transactions on Signal Processing*, 2010, vol. 58, no. 8, p. 4121–4130. DOI: 10.1109/TSP.2010.2049742
- [38] ZHU, A., DRAXLER, P. J., HSIA, C., et al. Digital predistortion for envelope-tracking power amplifiers using decomposed piecewise Volterra series. *IEEE Transactions on Microwave Theory and Techniques*, 2008, vol. 56, no. 10, p. 2237–2247. DOI: 10.1109/TMTT.2008.2003529
- [39] HEREDIA, E. A., ARCE, G. R. Piecewise Volterra filters based on the threshold decomposition operator. In *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference (ICASSP)*. Atlanta (USA), 1996, p. 1593–1596. DOI: 10.1109/ICASSP.1996.544107
- [40] ZHANG, C., WANG, J., WU, W. A piecewise generalized memory polynomial model for envelope tracking power amplifiers. In *Asia-Pacific Microwave Conference (APMC)*. Nanjing (China), 2015, p. 1–3. DOI: 10.1109/APMC.2015.7413559
- [41] ABDELAZIZ, M., ANTILA, L., BRIHUEGA, A., et al. Decorrelation-based piecewise digital predistortion: Operating principle and RF measurements. In *16th International Symposium on Wireless Communication Systems (ISWCS)*. Oulu (Finland), 2019, p. 340–344. DOI: 10.1109/ISWCS.2019.8877236
- [42] AFSARDOOST, S., ERIKSSON, T., FAGER, C. Digital predistortion using a vector-switched model. *IEEE Transactions on Microwave Theory and Techniques*, 2012, vol. 60, no. 4, p. 1166–1174. DOI: 10.1109/TMTT.2012.2184295

- [43] ASCHBACHER, E., CHEONG M. Y., BRUNMAYR, P., et al. Prototype implementation of two efficient low-complexity digital predistortion algorithms. *EURASIP Journal on Advances in Signal Processing*, 2007, vol. 2008, no. 1, p. 1–15. DOI: 10.1155/2008/473182
- [44] SEO, M., JEON, S., IM, S. Compensation for nonlinear distortion in OFDM systems using a digital predistorter based on the SCPWL model. In *6th International Conference on Wireless and Mobile Communications (ICWMC)*. Valencia (Spain), 2010, p. 27–32. DOI: 10.1109/ICWMC.2010.24
- [45] ZHAI, J., ZHANG, L., ZHOU, J., et al. A nonlinear filter-based vVolterra model with low complexity for wideband power amplifiers. *IEEE Microwave and Wireless Components Letters*, 2014, vol. 24, no. 3, p. 203–205. DOI: 10.1109/LMWC.2013.2293657
- [46] CHEONG, M. Y., WERNER, S., BRUNO, M. J., et al. Adaptive piecewise linear predistorters for nonlinear power amplifiers with memory. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2012, vol. 59, no. 7, p. 1519–1532. DOI: 10.1109/TCSI.2011.2177007
- [47] CHUA, L. O. KANG, S. M. Section-wise piecewise-linear functions: Canonical representation, properties, and applications. *Proceedings of the IEEE*, 1977, vol. 65, no. 6, p. 915–929. DOI: 10.1109/PROC.1977.10589
- [48] YU, Z., WANG, W., SHANG, G. A generalized model based on canonical piecewise linear functions for digital predistortion. In *Asia-Pacific Microwave Conference (APMC)*. New Delhi (India), 2016, p. 1–4. DOI: 10.1109/APMC.2016.7931346
- [49] ZHU, A. Decomposed vector rotation-based behavioral modeling for digital predistortion of rf power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 2015, vol. 63, no. 2, p. 737–744. DOI: 10.1109/TMTT.2014.2387853
- [50] JULIAN, P., JORDAN, M., DESAGES, A. Canonical piecewise-linear approximation of smooth functions. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 1998, vol. 45, no. 5, p. 567–571. DOI: 10.1109/81.668868
- [51] ZHU, A. Behavioral modeling for digital predistortion of RF power amplifiers: from Volterra series to CPWL functions. In *IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR)*. Austin (USA), 2016, p. 1–4. DOI: 10.1109/PAWR.2016.7440126
- [52] ZHAI, J., ZHANG, L., YU, Z., et al. A modified canonical piecewise-linear function-based behavioral model for wideband power amplifiers. *IEEE Microwave and Wireless Components Letters*, 2016, vol. 26, no. 3, p. 195–197. DOI: 10.1109/LMWC.2016.2524512
- [53] MATEO, C., CARRO, P. L., GARCIA-DUCAR, P., et al. Digital predistortion based on B-spline CPWL models in a RoF LTE mobile fronthaul. In *47th European Microwave Conference (EuMC)*. Nuremberg (Germany), 2017, p. 1136–1139. DOI: 10.23919/EuMC.2017.8231048
- [54] BASSAM, S. A., HELAOUI, M., GHANNOUCHI, F. M. 2-D digital predistortion (2-D-DPD) architecture for concurrent dual-band transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 2011, vol. 59, no. 10, p. 2547–2553. DOI: 10.1109/TMTT.2011.2163802
- [55] NARAHARISSETTI, N., QUINDROIT, C., ROBLIN, P., et al. 2D cubic spline implementation for concurrent dual-band system. In *IEEE MTT-S International Microwave Symposium Digest (MTT)*. Seattle (USA), 2013, p. 1–4. DOI: 10.1109/MWSYM.2013.6697739
- [56] LIU, Y., LI, C., QUAN, X., et al. Multiband linearization technique for broadband signal with multiple closely spaced bands. *IEEE Transactions on Microwave Theory and Techniques*, 2019, vol. 67, no. 3, p. 1115–1129. DOI: 10.1109/TMTT.2018.2884413
- [57] YOUNES, M., KWAN, A., AKBARPOUR, M., et al. Two-dimensional piecewise behavioral model for highly nonlinear dual-band transmitters. *IEEE Transactions on Industrial Electronics*, 2017, vol. 64, no. 11, p. 8666–8675. DOI: 10.1109/TIE.2017.2703683
- [58] YU, Z., WANG, W. Concurrent dual-band digital predistortion based on canonical piecewise linear functions. In *47th European Microwave Conference (EuMC)*. Nuremberg (Germany), 2017, p. 1058–1059. DOI: 10.23919/EuMC.2017.8231028
- [59] ZHAI, J., WU, S., ZHANG, L., et al. A 2-D-canonical piecewise linear function-based behavioral model for concurrent dual-band power amplifiers. *IEEE Microwave and Wireless Components Letters*, 2018, vol. 28, no. 11, p. 1050–1052. DOI: 10.1109/LMWC.2018.2873191
- [60] SAFFAR, D., BOULEJFEN, N., GHANNOUCHI, F., et al. Behavioral modeling of MIMO transmitters exhibiting nonlinear distortion and hardware impairments. In *6th European Microwave Integrated Circuit Conference (EuMIC)*. Manchester (UK), 2011, p. 486–489.
- [61] LUO, Q., YU, C., ZHU, X. A dual-input canonical piecewise-linear function-based model for digital predistortion of multi-antenna transmitters. In *IEEE/MTT-S International Microwave Symposium (IMS)*. Philadelphia (USA), 2018, p. 559–562. DOI: 10.1109/MWSYM.2018.8439236
- [62] NIU, J., ZHAI, J., YU, Z., et al. Low-complexity digital predistorter for MIMO transmitters with nonlinear crosstalk. In *2019 IEEE MTT-S International Wireless Symposium (IWS)*. Guangzhou (China), 2019, p. 1–3. DOI: 10.1109/IEEE-IWS.2019.8804139
- [63] CHEN, C., BATSELIER, K., TELESCU, M., et al. Tensor-network-based predistorter design for multiple-input multiple-output nonlinear systems. In *IEEE 12th International Conference on ASIC (ASICON)*. Guiyang (China), 2017, p. 1117–1120. DOI: 10.1109/ASICON.2017.8252676
- [64] BATSELIER, K., CHEN, Z., WONG, N. Tensor network alternating linear scheme for MIMO Volterra system identification. *Automatica*, 2017, vol. 84, no. 07, p. 26–35. DOI: 10.1016/j.automatica.2017.06.033.
- [65] WANG, S., HUSSEIN, M. A., VENARD, O., et al. A novel algorithm for determining the structure of digital predistortion models. *IEEE Transactions on Vehicular Technology*, 2018, vol. 67, no. 8, p. 7326–7340. DOI: 10.1109/TVT.2018.2833283
- [66] KANTANA, C., VENARD, O., BAUDOIN, G. Decomposed vector rotation model sizing by hill-climbing heuristic for digital predistortion of RF power amplifiers. In *Radio & Wireless Week (RWW)*. San Antonio (USA), 2020, p. 232–235.

About the Authors . . .

Geneviève BAUDOIN received graduate engineering degree from Ecole Nationale Supérieure des Télécommunications, Paris, France in 1977 and the habilitation for Ph.D. direction from the University of Marne La Vallée, Marne La Vallée, France, in 2000. She was a lecturer at the University of Paris-Ouest before she joined the Philips Research laboratory in France as a research-engineer and worked on ultrasound signal processing for medical applications. She is currently Professor at ESIEE Paris, France. She is a member of the ESYCOM research laboratory, Université Paris-Est Marne-la-Vallée. Her research and teaching activities include digital signal processing, wireless communications, wireless transmitter linearization, transceiver architectures and algorithms for wireless communications and localization systems.