GCMAC-Based Predistortion for Digital Modulations

Francisco-Javier González-Serrano, *Member, IEEE*, Juan José Murillo-Fuentes, *Student Member, IEEE*, and Antonio Artés-Rodríguez, *Member, IEEE*

***Abstract*—The subject of this paper is the compensation for nonlinearities in digital communication systems by means of predistortion. In this work, we apply the Generalized Cerebellar Model Articulation Controller (GCMAC) to simplify and accelerate the predistorter convergence. The range of analyzed predistorters includes: 1) a symbol-rate data predistorter that, for a given time span, achieves a similar level of compensation provided by present techniques, but with faster convergence; 2) a fractionally spaced data predistorter that controls, at the same time, the signal constellation and the transmitted spectrum; 3) a decision-feedback scheme that compensates for remote nonlinearities; and 4) a digital signal data predistorter. The performance of the proposed data and signal predistorters is evaluated using typical linear and nonlinear modulated transmitted signals such as QAM and GMSK.**

***Index Terms*—Amplifier distortion, cerebellar model arithmetic computer, continuous phase modulation, nonlinear distortion, quadrature amplitude modulation.**

I. Introduction

PRACTICAL digital radio systems employing spectrally efficient modulation schemes, such as M-ary passband pulse amplitude modulation[1] (PAM) or continuous phase modulation (CPM), require a compromise between power efficiency and linearity of the high-power amplifier (HPA). If the operating point of the HPA is far (backed-off) from its saturation point, the HPA operates in a quasi-linear region, but the transmitted signal power is reduced. On the contrary, if the HPA is working near its saturation point, a better use of the available HPA power is achieved, but the signal is severely distorted due to its envelope fluctuation.

The need for some compensation technique was recognized a long time ago, and some early proposed solutions are described in [1] and [2]. Basically, nonlinear compensators consist of controlling either the signal before it is sent (TX-techniques) or the noisy received signal (RX-techniques). The absence of noise makes the TX-techniques perform better than the RX ones for a given complexity (a comparison between them can be found in [3]). In this paper, we focus on the TX-techniques, also referred to as predistortion techniques.

The predistorters can be grouped into three main classes. The first one is formed by those techniques that try to reduce the envelope fluctuations in the transmitted signal and, therefore, the spectral spreading of the signal bandwidth. The second one includes the symbol-rate data predistorters, which consist of driving the transmitter by a modified input signal constellation in such a way that the original symbols appear in the correct positions after modulation and amplification in the HPA. Finally, the last group is formed by signal predistorters, which transform the transmitted signal without knowledge of the underlying data sequence.

The reduction of the envelope fluctuation can be achieved by limiting the symbol transitions (as occurs in FQPSK [4], [5]) or by modifying the pulse shape (see [6]). These techniques are usually applied to modulations with few symbols ($M \leq 4$). However, for modulations with dense constellations, such as 16-, 64- or 256-QAM, data predistortion is more attractive.

The simplest data predistorter, proposed by Saleh and Salz [7], consists of a Look-Up Table (LUT) with M memory cells addressed by the current PAM symbol. Later, Karam and Sari [8] generalized this technique using a LUT addressed by P input symbols (P-length LUT). In this way, the nonlinear intersymbol interference (ISI), due to the placement of the static nonlinearity between linear filters, can be reduced. For a given time span, the generalized LUT achieves the best performance as compared with other schemes such as the Volterra predistorter [9]. However, the large memory required (1 Gbyte for a 3-length LUT addressed by a 256-QAM constellation [8]) and the extremely slow convergence rate limit the practical application of this predistorter. Another alternative is the multi-layer perceptron (MLP) network [10], [11], although its problematic training limits its applicability to the predistortion problem.

It is possible to combine the advantages of the above described data predistortion (hardware simplicity and adaptability) and pulse shaping (reduction of side-lobes) tecnhiques simply by increasing the sampling rate. This data predistorter is called the fractional-spaced data predistorter [12].

Data predistorters are useful for linear modulations, such as PAM (16- and 64-QAM). However, for modulations such as CPM, or OFDM, where the transmitted signal amplitude does not depend linearly on data symbols, signal predistortion is more convenient. Typical microwave systems employ IF or RF signal predistorters that invert the memoryless input–output characteristic of the HPA. They can be implemented with analog technology [13], although its limited accuracy restricts its use, or with current digital hardware for signals with reduced bandwidth [14].

In this paper, we present and analyze different data and signal predistorters based on the Generalized Cerebellar Model Arithmetic Computer (GCMAC) [15], a neural network that uses

Paper approved by S. L. Ariyavisitakul, the Editor for Wireless Techniques and Fading of the IEEE Communications Society. Manuscript received June 15, 2000; revised December 4, 2000.

The authors are with the Dpto. Teoría de la Señal y Comunicaciones, Universidad Carlos III, 28911 Madrid, Spain (e-mail: fran@tsc.uc3m.es; antonio@tsc.uc3m.es; murillo@tsc.uc3m.es).

Publisher Item Identifier S 0090-6778(01)08165-X.

[1]Notice that the commonly used phase-shift keying (PSK), amplitude and phase modulation (AM–PM), and quadrature amplitude modulation (QAM) are special cases of passband PAM.

local basis functions. The selection of the GCMAC network is based on the following reason.

1) It has digital inputs. In data predistortion, the symbol sequence is a discrete-amplitude sequence. Therefore, the use of an architecture that accepts analog-valued input signals results in an unnecessary waste of resources. Moreover, the GCMAC network can be driven directly by the input bit sequence. In this way, the coding of the input bit stream, the predistortion of the PAM symbols, and the pulse shaping can be performed simultaneously.
2) It was designed for control applications. Predistortion can be seen as a nonlinear plant (linear filter + static nonlinearity) control problem.
3) It is model-independent. When the inputs are digital, the GCMAC is a universal approximator of finite complexity [16].
4) Its training is simple and fast. When the HPA is fixed or changing very slowly, predistortion can be implemented off-line. However, in some cases adaptive approaches are mandatory. Examples are the CDMA transmitters, which use adaptive transmitter power control (TPC) to solve the near–far problem and to increase the system capacity, and cellular systems, where a base station (BS) failure is compensated for by increasing the transmitted power in the neighboring BS. In conclusion, the GCMAC network results in an attractive solution when adaptive predistortion is required.

The paper is organized as follows. In Section II, we analyze some of the problems caused by nonlinear distortions in communication channels. Present predistortion strategies are included in Section III. After a brief description, the convenience of using the GCMAC network as a data predistorter is justified in Section IV. Section V is devoted to the compensation for local nonlinearities (placed at the transmitter) at the symbol rate; new structures for data predistortion are presented and analyzed. The compensation for local nonlinearities at multiples of the symbol rate is presented in Section VI. Section VII deals with signal predistortion. The performance of the GCMAC network in this application is evaluated and analyzed. The compensation of remote nonlinearities is addressed in Section VIII. The final performance evaluation and comparison of the proposed architectures and the discussion of the obtained results are presented in Section IX.

II. Problem Statement

In this section, we analyze some of the problems caused by nonlinear distortions in some digital communication channels. In particular, we consider two classes of passband systems. The first one uses linear modulations (PAM) with a high number of symbols ($M \geq 16$), and the second one uses multicarrier angular modulations. In both cases, the transmitted signals have important envelope fluctuations that can be seriously affected by a nonlinear amplifier. However, the effects of nonlinear distortion on both modulations are quite different, and, therefore, the predistortion architectures differ in the two cases.

A. Input Signals

In a digital PAM transmitter, the input bit stream is mapped into a set of M complex symbols. For the purposes of this paper, we will assume that this mapping does not introduce redundancy; hence, the input symbols $\{s[k]\}$ are assumed to be independent and identically distributed (i.i.d.), forming a white discrete-time random process. The complex symbols pass through a pulse-shaping filter, $g_{PAM}(t)$, that limits the bandwidth of the transmitted spectrum. The resulting baseband complex signal can be written as

$$x(t) = \sum_{k=-\infty}^{\infty} s[k] g_{PAM}(t - kT_s) \tag{1}$$

where T_s is the symbol period. The signal $x(t)$ drives an amplitude modulator which generates an RF signal $y(t)$ that is amplified by the HPA.

In CPM, the transmitted signal at baseband is as follows:

$$x(t) = e^{j\phi(t)} \tag{2}$$

where $\phi(t)$, the information-carrying phase, is

$$\phi(t) = 2\pi h \sum_{k=-\infty}^{\infty} s[k] g_{CPM}(t - kT_s). \tag{3}$$

The pulse shape $g_{CPM}(t)$ determines the spectral behavior of the modulation scheme, while h is the modulation index and $s[k]$ are the data symbols. The resulting signal, $x(t)$, is converted to an RF signal, $y(t)$.

B. Amplifier Model

For the channels of interest, the output of the HPA is described, at baseband, in terms of a memoryless nonlinear complex function $\Gamma(\cdot)$ of the input amplitude [17]. If the input signal is of the form

$$y(t) = r(t)\cos(w_c t + \phi(t)) \tag{4}$$

then the output of the model yields

$$z(t) = \Gamma(y(t)) = A[r(t)]\cos(w_c t + \phi(t) + \Psi[r(t)]) \tag{5}$$

where

$$A[r(t)] = \frac{\alpha_a r}{1 + \beta_a r^2}$$
$$\Psi[r(t)] = \frac{\alpha_\varphi r^2}{1 + \beta_\varphi r^3} \tag{6}$$

and parameters α_a, α_φ, β_a, and β_φ must be adjusted to approximate the features of the real amplifier.

To achieve linear amplification, the average output power of the HPA may be reduced. This reduction is called output back-off (OBO), and it is defined as the ratio of the saturation (maximum) to the actual output power.

C. Analysis of Nonlinearly Amplified M-ary PAM

The nonlinear behavior of HPA has three main unwanted effects. First, the widening of the transmitted pulse restores the

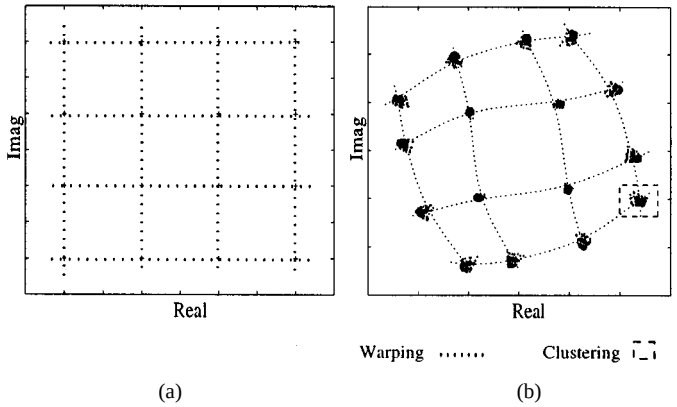

Fig. 1. (a) Original symbol constellation. (b) Received symbol constellation after nonlinear distortion ($OBO = 0$ dB).

side-lobes which might cause severe adjacent channel interference (ACI). Second, due to the memoryless nonlinear behavior of HPA, the received constellation is no longer lying on the original lattice ("warping effect"). Finally, the inclusion of the HPA between linear transmission and receiving filters leads to nonlinear ISI that produces the spreading of the received constellation point in small clusters ("clustering effect"). The last two effects, sketched in Fig. 1, can be described in terms of the following discrete-time dynamic equation:

$$r[k] = F(\ldots, s[k-1], s[k], s[k+1], \ldots) \tag{7}$$

where

$r[k]$ received symbol;
$\{s[k]\}$ original sequence of PAM symbols;
$F(\bullet)$ multivariable nonlinear mapping which describes the behavior of the channel.[2]

For the scope of this paper, the length of the channel memory is assumed to be finite.

D. Analysis of Nonlinearly Amplified Multi-Carrier CPM

If the outputs of baseband modulating branches $x_i(t)$ are combined, the output $x(t)$ results in a nonconstant envelope signal. If $x(t)$ drives, after frequency conversion, a power-efficient amplifier, intermodulation (IM) components appear at the output, limiting the capacity of the system. We apply predistortion to multicarrier CPM to cancel harmonics within the transmitted bandwidth.

III. Present Predistortion Techniques

The aim of amplifier predistortion is to invert the nonlinear function of the HPA so that the response of the predistorter $H(\cdot)$ plus the HPA is linear. Since the HPA behaves linearly for a large range of input amplitudes, in practical implementations the predistortion function is modeled in the form

$$H(x(t)) = x(t) + \hat{H}(x(t)) \tag{8}$$

where $x(t)$ is a linear term containing the signal to be amplified, and $\hat{H}(\cdot)$ represents the deviation from linearity. This way, the predistorter just compensates for the residual nonlinear distortion.

For amplitude-modulated signals, e.g., M-PAM, the nonlinear behavior of the power amplifier directly affects the sequence of transmitted digital symbols, and vice versa. For this reason, the appropriate modification of both the shape of the transmitted signal (signal predistortion) or the underlying data sequence (data predistortion) could produce a noticeable reduction of the nonlinear distortion. Usually, data predistorters work at symbol rate and only the information within the channel bandwidth is predistorted. Digital signal predistorters have higher sampling rates, making possible a better spectrum control at the expense of a higher computational cost.

In single-carrier angular modulated carriers, e.g., CPM, the nonlinear distortion within its bandwidth is usually negligible. Moreover, harmonics are cancelled by post-amplifier passive filtering. However, if we face a multicarrier transmission, IM products appear within the transmission band, increasing the adjacent channel interference (ACI). Since the multicarrier waveform does not depend linearly on the underlying data sequence, signal predistortion is more convenient.

It may be concluded that the type of predistortion to use depends on the signals involved. In this sense, PAM transmitters may be compensated with either data or signal predistorters. However, only signal predistortion is efficient when the signal is multicarrier and phase modulated.

A. Data Predistortion

The data predistorter transforms a finite sequence of P correlative PAM symbols and produces one predistorted symbol $b[k]$, defined as

$$\begin{aligned} b[k] &= H(s[k-Q], \ldots, s[k], \ldots, s[k+R]) \\ &= H(\boldsymbol{s}_P[k]) \end{aligned} \tag{9}$$

where $P = Q + R + 1$ is the memory length of the predistorter. The data predistorter $H(\cdot)$ has to be designed in such a way that, after linear filtering and nonlinear processing in the link, the constellation of the average samples at the detector would match (or approximate) the desired M-ary PAM signal constellation. As the predistorter is placed before the premodulating filter [see Fig. 2(a)], memory is needed to learn its response.

The necessary components for predistorter control are the transmitted signal $s[k]$ and the received symbol $r[k]$. When the major source of nonlinear distortion is placed at the transmitter (local nonlinearity), predistorter Fig. 2(a) includes a local receiver that feedbacks the transmitted signal and provides an approximation of the real received signal. On the contrary, when the output of the nonlinear source is not accessible, for instance in digital satellite communications with the on-board amplifier working at saturation, the previous scheme is no longer valid. However, in some cases, as will be further explained in Section VIII, it is possible to acquire some knowledge about the nature of the remote nonlinearity present in the channel.

B. Signal Predistortion

Fig. 3(a) illustrates the hardware configuration for an adaptive signal predistorter. The data modem generates the desired com-

[2]We consider the channel including all the elements and devices between the modulator and the detector.

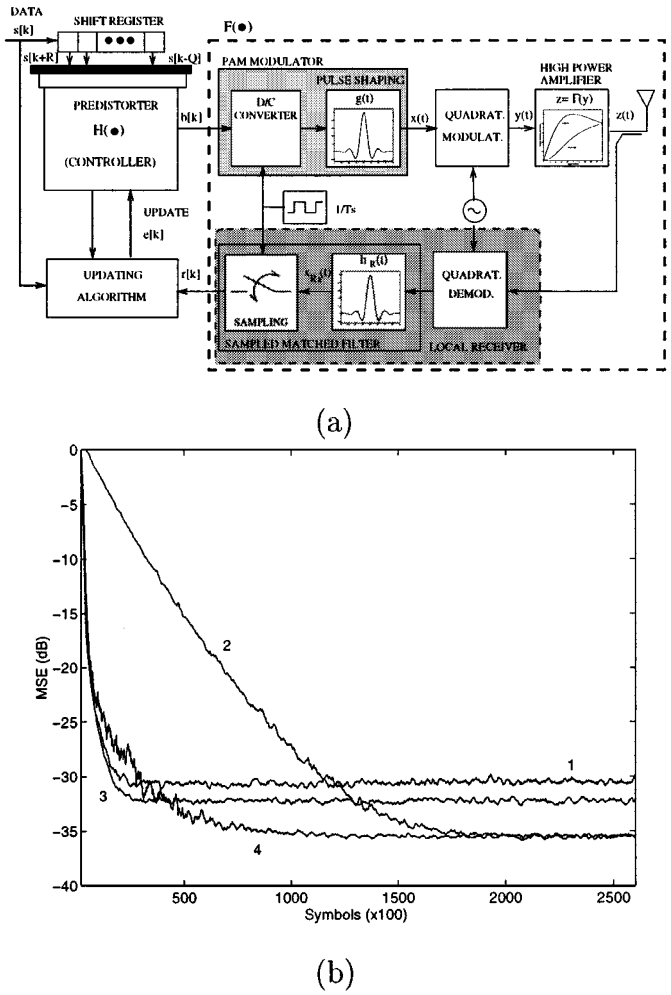

Fig. 2. (a) Model of the transmission system with data predistortion. (b) Convergence curves of 3-length GCMAC data predistorters. Curve 1: CMAC ($\rho_i = \rho = 15$); curve 2: LUT ($\rho_i = \rho = 1$); curve 3: GCMAC ($\boldsymbol{\rho} = [15, 8, 15]$); curve 4: GCMAC ($\boldsymbol{\rho} = [15, 1, 15]$).

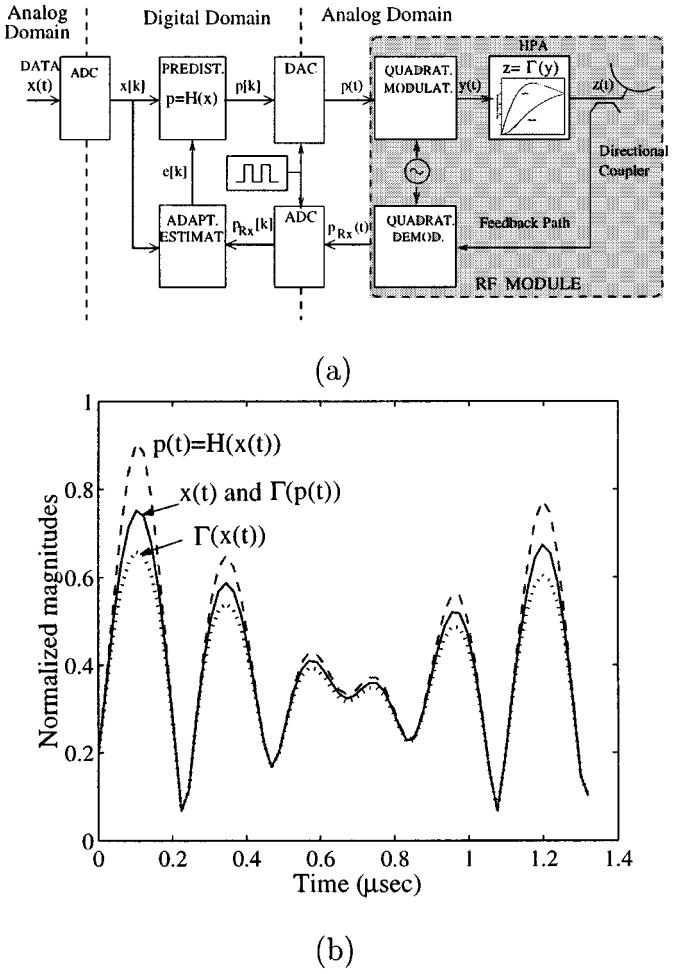

Fig. 3. (a) Digital signal predistorter at the baseband. (b) Time-plot of a normalized multicarrier GSM signal (saturation level = 1).

plex signal $x(t)$. The predistorter generates a complex signal $p(t) = H(x(t))$ that corrects for the nonlinearities introduced by the RF module. The complex function $H(\cdot)$ can be approximated using adaptive methods driven by the difference between the instantaneous complex modulation envelope at the demodulator output and the desired modulation envelope. Unlike data predistortion, signal compensation is placed after the premodulating filter. The function to invert is of the form in (5) and no memory is introduced.

In Fig. 3(b), we represent the normalized magnitude of a multicarrier CPM signal $x(t)$ along with its amplified version $\Gamma(x(t))$ and the predistorted signal $p(t) = H(x(t))$. Notice that the normalized amplified predistorted signal is nearly equal to $x(t)$. In this figure, it can be observed how the predistorter linearizes the HPA by increasing the high level points near saturation.

C. Predistortion as a Function Approximation

Predistortion can be interpreted as a function approximation. The smoothness property of the ideal predistortion functions is the key element to design simple, but effective, predistorters. For this reason, architectures with a few parameters, such as memoryless polynomic networks or LUT's, are often used, especially in signal predistortion. In relation to the predistortion function, we present some features to be taken into account.

First, the ideal predistortion function has a multidimensional input space. This implies that the predistorter may include memory to cope with the nonlinear time dependences. One approach is the generalized LUT [8]. This technique is very simple and robust and there are no constraints on the function $H(\cdot)$ [18]. Unfortunately, the number of adjustable parameters increases as M^P, where P is the memory length, which makes this structure only practical for nonlinearities with short memory and modulations with few symbols[3]. As stated before, one reasonable alternative is to use architectures with generalization abilities, such as the MLP (that provides a global generalization in the whole input space) or the local-basis functions (LBFs) networks.

The second issue is that the input variables of the ideal predistortion function are complex symbols.

In data predistortion, an issue to be explored is the relevance of the warping and clustering effects. From Fig. 1(b), it can be

[3]In M-ary QAM transmitters, the number of adjustable parameters of a LUT-based predistorter can be divided by a factor of 4 by exploiting the quadrant symmetries of the symbol constellation [8].

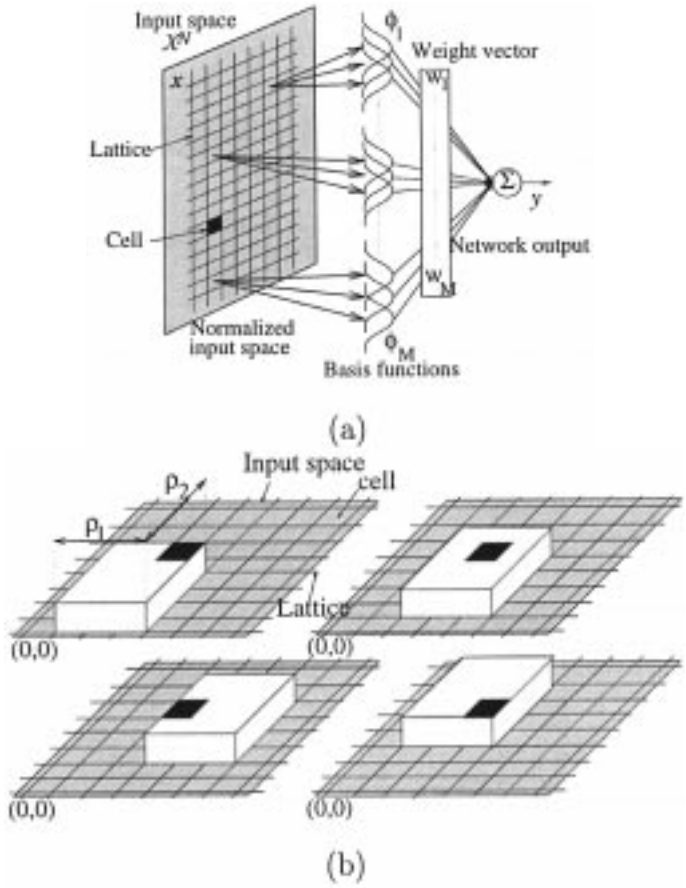

Fig. 4. The GCMAC network. (a) Three-layer architecture. (b) Constant LBFs.

asserted that compensation for the warping effect is more essential than compensation for the clustering effect for reducing the bit error rate (BER). In other words, without disregarding the clustering effect, more attention should be paid to the warping effect.

IV. The GCMAC Network

The GCMAC network can take advantage of the particular features of the predistortion functions to achieve fast and effective compensation for typical nonlinear distortions.

A. Network Structure

The structure of the Generalized CMAC network is illustrated in Fig. 4. The GCMAC divides the input space into cells using a lattice.[4] This division normalizes the input space [see Fig. 4(a)]. In the GCMAC approach, the LBFs are supported on rectangular domains that are evenly distributed on the normalized input space in such a way that exactly $\rho_{\max}$ LBFs cover each cell, where $\rho_{\max} = \max(\rho_1, \ldots, \rho_P)$, and ρ_i is the size of the LBF along the ith axis[5] $(1 \leq \rho_i < M)$. Fig. 4(b) shows the distribution of constant LBFs for $\rho_1 = 3$ and $\rho_2 = 4$. As can be observed, the generalization is influenced by the geometry of the local domains and, for this reason, the vector $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_P]^{\mathrm{T}}$ is called the generalization vector. When $\rho_i = \rho \; \forall \, i$, the GCMAC becomes the Albus' CMAC [19]. For $\rho_i = 1 \; \forall \, i$, there is no generalization (LUT); the larger ρ_i are, the more generalization is obtained.

The GCMAC input–output function can be decomposed into two consecutive mappings. The first one produces an N-dimensional addressing vector $\boldsymbol{a}$ given by

$$\boldsymbol{x} \rightarrow \boldsymbol{a}(\boldsymbol{x}) = [\Phi_1(\boldsymbol{x}), \ldots, \Phi_N(\boldsymbol{x})]^{\mathrm{T}} \tag{10}$$

where $\{\Phi_1(\boldsymbol{x}), \ldots, \Phi_N(\boldsymbol{x})\}$ is the set of basis functions.[6]. The addressing vector $\boldsymbol{a}$ only has $\rho_{\max}$ nonzero elements and, generally, the relationship $P < \rho_{\max} \ll N$ holds. In addition, the transformed inputs are normalized in order to take into account the different number of points assigned to the basis functions located at the border of the input space. The addressing vector lies in a higher dimensional space where the desired function can be approximately linear. For this reason, the second map consists of the projection of the transformed input vectors $\boldsymbol{a}$ onto a vector of weights $\boldsymbol{w}$, which produces the output of the network

$$y = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{a}(\boldsymbol{x}) = \sum_{j=1}^{N} \boldsymbol{w}_j \Phi_j(\boldsymbol{x}). \tag{11}$$

Hence, the approximation used by the GCMAC network is linear in the unknown coefficients $\boldsymbol{w}$ and, therefore, simple instantaneous learning laws can be used, for which convergence can be established subject to well-understood restrictions. For instance, the minimization of the cost function

$$J[k] = \epsilon^2[k] = |e[k]|^2 \tag{12}$$

where $e[k] = y_d(\boldsymbol{x}[k]) - y[k] = y_d(\boldsymbol{x}[k]) - \boldsymbol{w}^{\mathrm{T}}[k]\boldsymbol{a}(\boldsymbol{x}[k])$ is the instantaneous error produced in the approximation of the ideal function $y_d(\boldsymbol{x}[k])$ at instant k, yields the well-known LMS algorithm

$$\boldsymbol{w}[k+1] = \boldsymbol{w}[k] + \beta e[k] \boldsymbol{a}(\boldsymbol{x}[k]) \tag{13}$$

that in the CMAC-related literature is called the Albus' rule [19].

B. Computational Requirements

The number of nonoverlapped basis functions along the ith axis, N_i, is bounded by

$$\frac{L_i}{\rho_i} \leq N_i \leq \frac{L_i}{\rho_i} + 1 \tag{14}$$

where L_i is the number of discrete levels ($L_i = M \; \forall \; i$ in data predistortion of M-PAM symbols) and ρ_i is the length of the support of LBFs along the ith axis. Since there are $\rho_{\max}$ nonoverlapped sets of LBFs covering each cell, the total number of LBF's (weights), N, can be bounded by

$$\rho_{\max} \min_i \left(\frac{M}{\rho_i}\right)^P \leq N \leq \rho_{\max} \max_i \left(\frac{M}{\rho_i} + 1\right)^P. \tag{15}$$

The GCMAC network is trained by using a modified version of (13) [15]. Our algorithm requires one float point complex subtraction to compute the actual error, $\rho_{\max}$ complex additions and

[4]This preliminary quantification is not required when the input space is digital (data predistortion).

[5]Each axis is related to one of the correlative input symbols.

[6]In order to remain consistent with the notation developed for the CMAC network, the input vector is represented by $\boldsymbol{x}$

one float-point inversion to compute the gains of the weights, and $\rho_{\max}$ scaling operations and $\rho_{\max}$ additions to update the weights. The network's output is computed after $2\rho_{\max}-1$ complex multiplications and accumulations.

V. Symbol-Rate Data Predistortion

In this section, we analyze two new compensation methods for nonlinear sources that are placed at the transmitter (TX-techniques). Thus, the output of the nonlinear source is accessible for implementing a feedback branch that provides the control signal to the predistorter. To simplify the comparison among the proposed predistortion schemes, we have applied them to a 16-QAM system with root-raised cosine pulse-shaping filter ($\alpha = 0.5$), an HPA operating at 2 dB of output back-off, and the model of the amplifier in (5) and (6) with parameters $\alpha_a = 2$, $\alpha_\varphi = 1$, $\beta_a = 1$, and $\beta_\varphi = 1$.

A. LUT + CMAC Data Predistorter

Recalling Fig. 1, it is observed that most of the energy of error produced by the mismatch between the original and the distorted constellations results from the warping effect. For this reason, we suggest to distribute the identification of the predistortion surface into two steps. In the first one, the influence of the actual symbol $s[k]$ on the surface is removed. Then, the remaining clustering effect is reduced. Thus, the predistortion surface can be divided into the form

$$H(\boldsymbol{s}_P[k]) = H_1(s[k]) + H_2(\boldsymbol{s}_P[k]) \tag{16}$$

where $H_1(\cdot)$ models the influence of the actual symbol (warping effect) and $H_2(\cdot)$ the remaining nonlinear ISI (clustering effect).

The previous decomposition allows the use of combined simple networks to approximate the ideal predistortion function. Our approach is based on the linear combination of two nonlinear networks with low complexity: the memoryless LUT and the CMAC network [see Fig. 5(a)]. The operation of this scheme is as follows. First, the warping effect $H_1(s[k])$ is reduced by means of the memoryless LUT networks, ignoring the contribution of the nonlinear ISI (clustering effect). Then, after convergence down to an acceptable noise floor has been reached, the CMAC network attempts to reduce the remaining unknown nonlinearity $H_2(\boldsymbol{s}_P[k])$ (clustering effect). The CMAC uses information from the first network and allows subsequent convergence from the nonlinear noise floor to the final compensation depth.

The performance of the LUT + CMAC scheme is depicted in Fig. 5(b). It is interesting to note that either the memoryless LUT or the CMAC network performs clearly worse than the ideal 3-length LUT ($Q = R = 1$, $P = 3$). However, when they are combined in the LUT + CMAC scheme, two desirable properties are achieved: fast training and low error. Indeed, the LUT + CMAC clearly outperforms the memoryless LUT predistorter or the CMAC predistorter by 10 dB and 5 dB respectively; in addition, the achieved MSE is almost equal to that obtained with the (much more complex) 3-length LUT.

The main drawback of this architecture stems from the rigid model used to factorize the ideal predistortion function (sum of nonlinear functions). Besides, the sequential training might cause problems in transmissions with time-variant power levels.

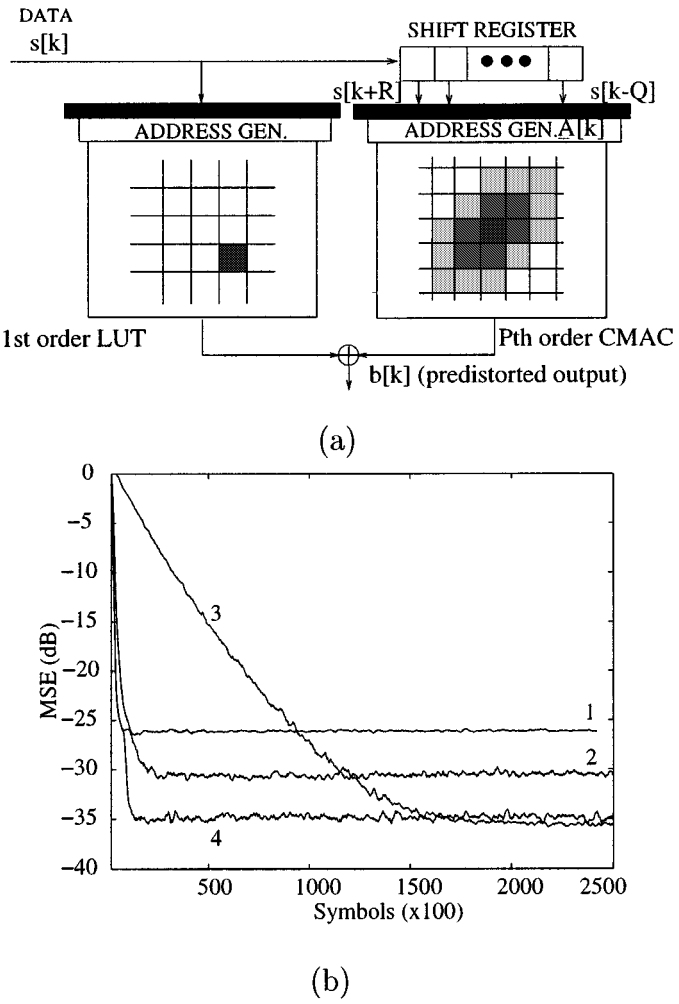

Fig. 5. (a) LUT + CMAC predistorter. (b) Convergence curves. Curve 1: 1-length LUT; curve 2: 3-length CMAC ($\rho = 15$); curve 3: 3-length LUT; curve 4: LUT + CMAC ($P = 3$ and $\rho = 15$).

B. The Generalized CMAC Network

The GCMAC network is a better approach to predistortion. Its architecture, defined by the generalization vector $\boldsymbol{\rho}$, is suitable for approximating functions with strong dependences along certain coordinates (directions). Notice that in predistortion there is strong influence of the actual symbol (warping effect) and weak respect to the future and past ones (clustering effect). Fig. 5(b) compares the evolution of four 3-length GCMAC networks. It is observed that reducing the element of the generalization vector related to the actual symbol ($\rho_2 \downarrow$) yields a substantial decrease in final error without increasing the speed of convergence. The main conclusion is that the GCMAC is able to yield the same compensation capability that the LUT does with a smaller number of parameters and with much faster convergence.

VI. Fractionally Spaced Data Predistortion

Although symbol-rate predistorters are effective solutions to compensate for the warping and clustering effects, they have a

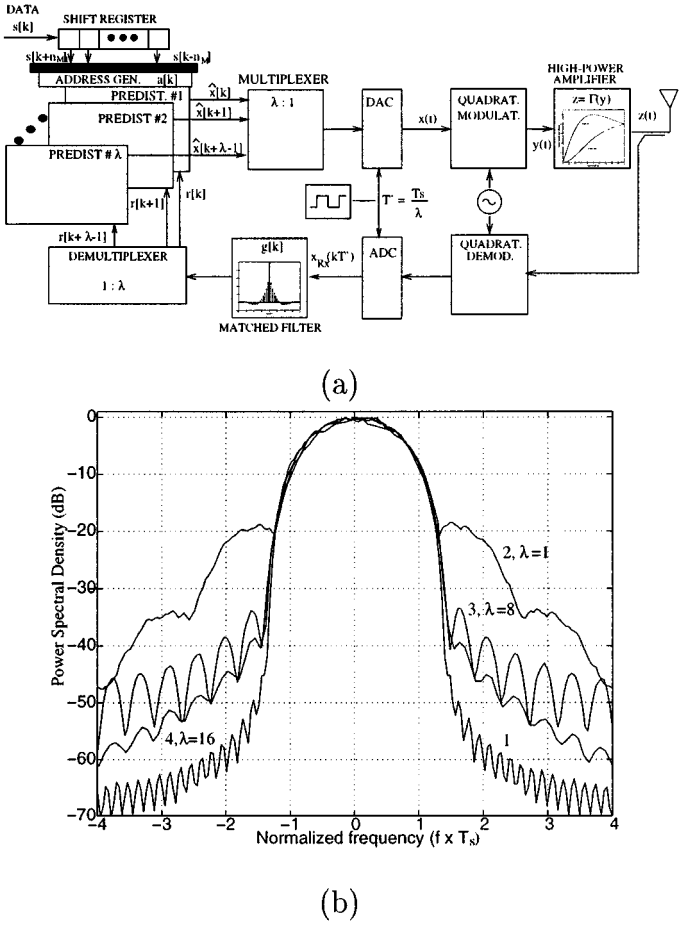

Fig. 6. (a) Model of the transmission system with fractional-spaced data predistortion. (b) Transmitted power spectral density (PSD). Curve 1: desired root-raised cosine characteristic with roll-off factor $\alpha = 0.5$; curve 2: conventional data predistortion ($\lambda = 1$); curve 3: predistortion with $\lambda = 8$; curve 4: predistortion with $\lambda = 16$.

limited influence on the transmitted signal spectrum. To achieve control over the transmitted spectrum, it is necessary to treat jointly the data predistorter and the shaping filter, but with a low increase of the computational cost.

The system we are proposing is based on predistorting the transmitted symbols so that the received PAM signal approximates the ideal one, not only at the maximum eye-opening instants nT_s, but also at the intermediate instants $kT_s/\lambda = kT'$, where λ is an integer that satisfies $\lambda \geq 2$ [12]. The suggested system, which provides a distortion-free communication at all kT_s/λ instants, is called the fractionally spaced data predistorter (FSDP).

Our scheme, depicted in Fig. 6(a), consists of λ GCMAC-based predistorters working in parallel. Each one is trained to approximate the desired sequence at one of the λ intermediate instants. Since all the intermediate values between two consecutive symbol instants depends on the same sequence of M-ary PAM symbols, the predistorters can be operated at symbol rate. In this way, the weight addressing algorithm is shared by the λ predistorters, reducing the computational burden of the method.

We have made use of a GCMAC-based FSDP ($\boldsymbol{\rho} = [15,\ 1,\ 15]^{\mathrm{T}}$) to perform the predistortion of the 16-QAM system described in Section V at instants T_s/λ, where $\lambda = 1, 8, 16$. Fig. 6(b) shows the transmitted power spectral density (PSD) obtained with these predistorters. It can be clearly seen that the proposed predistorters reduce the spectral spreading: the secondary lobes at $fT_s = 3$ (adjacent channel) have been reduced 9 dB ($\lambda = 8$, curve 3) and 18 dB ($\lambda = 16$, curve 4) with respect to conventional symbol-rate predistortion ($\lambda = 1$, curve 2).

Higher reduction of side-lobes can be achieved by increasing both the sampling factor λ and the predistortion order P. Moreover, when the roll-off factor α is too low, the reduction of side-lobes gets stronger by increasing the order P than the oversampling factor λ. As an example, in a 4-QAM system with roll-off factor $\alpha = 0.22$ (radio interface of UMTS), the side-lobe reduction provided by a fifth-order GCMAC FSDP with $\lambda = 8$ is the same (10 dB) as that provided by a third-order GCMAC FSDP with $\lambda = 16$ [20]. It is obvious that increasing the order P implies increasing the complexity. The key point is that the GCMAC, in contrast to other nonlinear networks, does not suffer the so-called "curse of dimensionality" (the complexity grows exponentially with P). Indeed, the number of floating-point operations required by the GCMAC network to predistort a given sample does not depend on the order P, but on the maximum component of vector $\boldsymbol{\rho}$ (see Section IV).

To summarize, the GCMAC-based FSDP reduces the interference between adjacent channels (ACI) using digital processing at baseband, making feasible the simplification of the analog transmission filters after the HPA.

VII. Signal Predistortion

Unlike data predistorters, signal predistorters generate a baseband, or IF, signal that compensates for the nonlinearities introduced by the RF module without accessing the underlying original data symbol sequence. In particular, if the nonlinearities introduced by the up-converter circuits are neglected, the RF module has a memoryless nonlinear behavior caused exclusively by the HPA. In practice, the exact inverse is not realizable and can only be approximated. Thus, in some practical systems, the HPA response is first approximated by a truncated power series, usually of third order, and later a polynomial is computed so as to cancel the distortion in the combined response of the predistorter and the HPA [21]. This strategy, extremely simple, provides poor results, principally due to the inaccurate modeling of the nonlinear amplifier and the unstability problems of the adaption of the polynomial coefficients [14].

On the contrary, the GCMAC we propose as predistorter does not require a model of the nonlinear source. The training only requires pairs of input/output samples across the composite system consisting of the inverse (predistorter) and direct system (HPA).

The digital implementation of the GCMAC network requires the use of analog-to-digital converters [see Fig. 3(a)]. The oversampling rate must be selected according to both the bandwidth of the signals to be processed and generated and the power consumption in the digital processor. The number of levels of the quantizer is selected to keep the performance of the predistorter

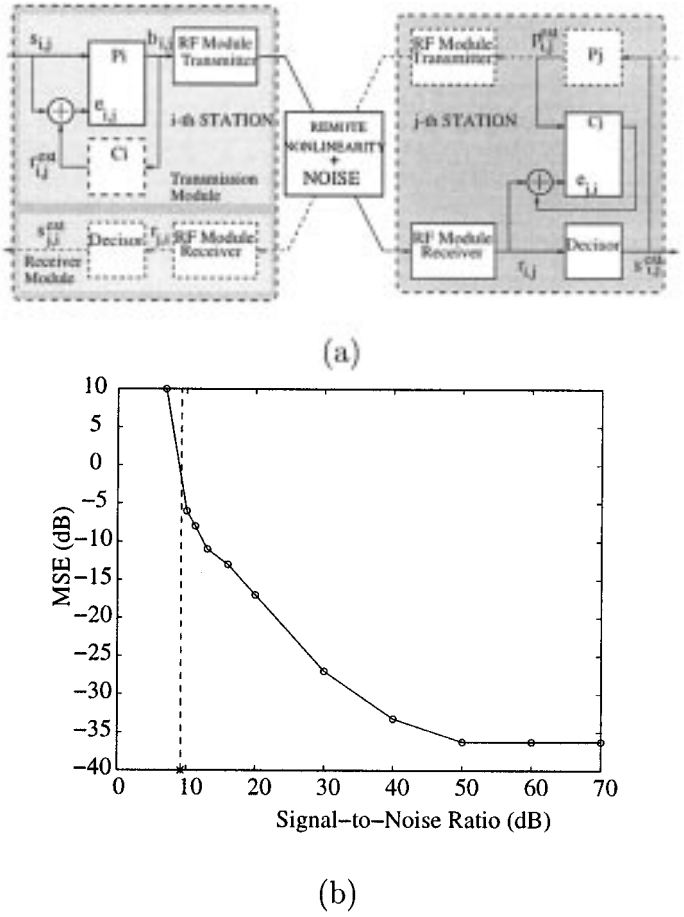

Fig. 7. (a) Model of a communication system with compensators for remote nonlinearities. (b) Analysis of the mean square error (MSE) versus SNR at the down-link for a satellite communication system.

between the desired margins of quality. It may be used with nonuniform quantization to compress areas with larger signal variations and enlarge those with smaller variations.

VIII. Compensation for Remote Nonlinearities

For digital communication systems where the output of the nonlinearity cannot be fed back into the predistorter (for instance, satellite systems), we propose the decision-feedback predistorter (DFP) shown in Fig. 7(a). Assuming a symmetric satellite link between two earth stations, each station can identify, and, therefore, invert, the nonlinear channel (composed of the nonlinear amplifiers of both the earth station and the satellite transponder) by only using the information carried by the received signal.

Each earth station consists of two transmission and receiver modules. The transmission module of the ith earth station is formed of a predistorter P_i and a network C_i. The first one maps the input data symbols $s_{i,j}[k]$ to be transmitted to the jth earth station onto a predistorted sequence $b_{i,j}[k]$. The second one is trained to approximate the nonlinear behavior of the satellite channel. When the ith station is transmitting, the predistorter P_i corrects its weights by using the estimated received symbol $r_{i,j}^{\text{est}}[k]$, obtained after passing the predistorted symbol $b_{i,j}[k]$ through the network C_i. Obviously, the operation of the system strongly depends on the appropriate model of the channel.

When the jth station is currently receiving data, the model C_j is adapted. However, the training of the model presents an important problem: to know the transmitted signal, $b_{i,j}[k]$. This problem can be solved by observing that the local predistorter P_j operating near the optimum performance can provide an estimation of the transmitted symbol $b_{i,j}^{\text{est}}[k]$. Indeed, if the detected symbol $s_{i,j}^{\text{est}}[k]$ is passed through the predistorter, an estimate of the current transmitted symbol is produced. Just as in the analysis of DFEs, it is required that the slicer be suitably adjusted to make the right decisions. In this way, the model of the channel C_j can be successfully updated. Therefore, when the transmission module is active, the channel network C_j remains with their current coefficients and the predistorter network P_j is updated. On the other hand, when the receiving module is working, the model of the channel is appropriately updated and the weights of the corresponding predistorter network are left at their actual values.

The performance of the proposed scheme was evaluated by simulating a point-to-point, symmetric digital communications system. The transmitter considered is the 16-QAM described in Section V. The nonlinear distortion is caused by the HPA of the transmitters of both the earth station and the satellite transponder. Noise was only considered at the downlink. The predistorter[7] P_j and the channel networks C_j were implemented by using 3-length GCMAC networks with generalization vectors $\boldsymbol{\rho}_{P_j} = \boldsymbol{\rho}_{C_j} = [15,\ 1,\ 15]^{\mathrm{T}}$. To test the robustness of the system, we have computed the mean square error (MSE), obtained after comparing the original constellation with the received one, as a function of the signal-to-noise ratio (SNR) measured at the downlink. Results, plotted in Fig. 7(b), reveal that there are three different regions. For SNRs greater than 50 dB, the predistortion noise, due to the incomplete compensation for the nonlinearity, prevails over the noise at the downlink. For SNRs ranging from 13 dB to 50 dB, the MSE is inversely proportional to the SNR. Finally, for SNRs less than 10 dB, the weights of the networks begin to diverge, and the performance of the system gets worse.

IX. Performance Results

Additional results on the performance in data and signal predistortion are included within this section. We first predistort a digital M-QAM modulated system. Secondly, we propose a signal predistortion scheme for a GSM booster.

A. M-QAM Modulated Systems

In this section, we quantify the validity of the proposed predistorters using the equivalent SNR degradation caused by the residual nonlinear distortion at a specified BER. For this purpose, the channel is assumed to have a flat frequency response with additive, white, circularly symmetric, Gaussian noise. If $[\text{SNR}]_C$ is the SNR, expressed in decibels, required by the compensated system to obtain the specified BER at a given output

[7]In this section, only symbol-rate predistortion is considered. If transmission spectrum control is required, the FSDP should be used.

back-off, and $[\mathrm{SNR}]_G$ is the required SNR to obtain the same BER on the Gaussian channel, then the total degradation is defined as

$$[TD] = [\mathrm{SNR}]_C - [\mathrm{SNR}]_G + [OBO]_C. \tag{17}$$

The total degradation results in a convex function of the output back-off, taking the minimum value at the optimum output back-off (OBO_{opt}); this function can be obtained by following the quasi-analytical procedure described in [3].

We have simulated a M-QAM transmitter with roll-off factor $\alpha = 0.5$ and $\alpha = 0.25$. The desired BER is 10^{-4}. In the comparison, we have included the following compensation systems (schemes are listed below in the order they are labeled in figures and table):

1) automatic control gain (no predistortion);
2) memoryless data predistortion: 1-length LUT ($Q = R = 0$, $P = 1$);
3) data predistortion: 3-length LUT ($Q = R = 1$, $P = 3$);
4) standard CMAC: $P = 3$, $\rho = M - 1$;
5) LUT + CMAC: $P = 3$, $\rho = M - 1$;
6) complex signal predistortion; 6-bit ADC; sampling frequency $= 8 \times 1/T_s$. GCMAC: $P = 2$, $\boldsymbol{\rho} = [32, 32]$;
7) GCMAC: $P = 3$, $\boldsymbol{\rho} = [M-1, 1, M-1]^{\mathrm{T}}$;
8) FSDP: $\lambda = 4$; GCMAC: $P = 3$, $\boldsymbol{\rho} = [M-1,\ 1,\ M-1]^{\mathrm{T}}$.

Results for an M-QAM transmitter are shown in Fig. 8. We make the following comments.

1) The standard CMAC has serious limitations in predistortion applications.
2) The GCMAC signal predistorter achieves good results but at the expense of oversampling the signal several times above the symbol rate. The high sampling rate might be impractical when the predistorter only can be inserted at IF. Results slightly improve when the quantization levels are higher than 128.
3) The GCMAC-based symbol-rate data predistorter performs as well as the LUT. In a 16-QAM system, their gains[8] are about 3 dB compared with the automatic control gain (ACG). As the number of constellation symbols increases ($M \uparrow$), the gains achieved by the LUT and GCMAC with respect to the ACG increase as well. In this way, the achieved gain using a GCMAC predistorter in a 64-QAM transmitter with roll-off factor $\alpha = 0.5$ is about 6 dB. However, regarding the number of symbols needed to achieve a given gain, the GCMAC network clearly outperforms the LUT [see Fig. 5(b)].
4) When the roll-off factor falls from $\alpha = 0.5$ to $\alpha = 0.25$, the length of the effective channel memory increases. As a consequence, for a given length of predistortion P, predistorters achieve worse results.
5) The gain achieved with the LUT + CMAC network ($P = 3$) is about 3 dB for a 16-QAM with $\alpha = 0.5$ (6.5 dB for a 64-QAM) compared with the ACG and up to 0.8 dB with respect to the 3-length CMAC. It is necessary to emphasize that the LUT + CMAC predistorter of a 16-QAM transmitter only requires 136 complex memory positions (568 positions are required to predistort a 64-QAM transmitter), while the 3-length LUT requires 4096 complex memory positions (262 144 positions for a 64-QAM transmitter). In addition, its convergence is significantly faster (about one order of magnitude) than the LUT [see Fig. 5(b)].
6) Finally, even though the FSDP produces degradation similar to the symbol-rate predistorter, due to the flat channel assumption, it seems reasonable to expect a noticeable improvement in bandlimited channels.

[8]The gain is defined as the difference between the values of the total degradation evaluated at the optimum output back-off.

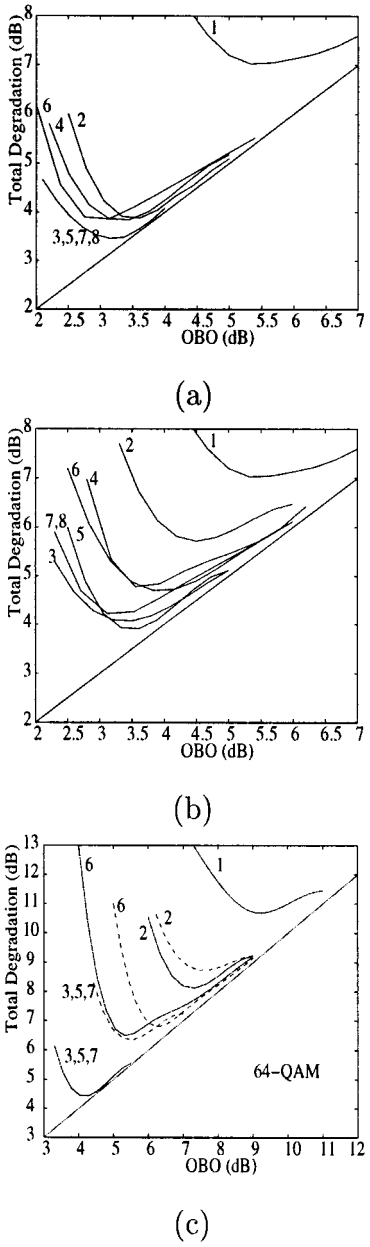

Fig. 8. Total degradation versus output back-off for a M-QAM system for a BER= 10^{-4}. (a) $M = 16$, $\alpha = 0.5$. (b) $M = 16$, $\alpha = 0.25$. (c) $M = 64$ Solid line: $\alpha = 0.5$. Dashed line: $\alpha = 0.25$.

B. GSM Booster

In GSM, carriers are GMSK modulated [22]. Besides, when slow frequency hopping (SFH) is introduced, the carrier fre-

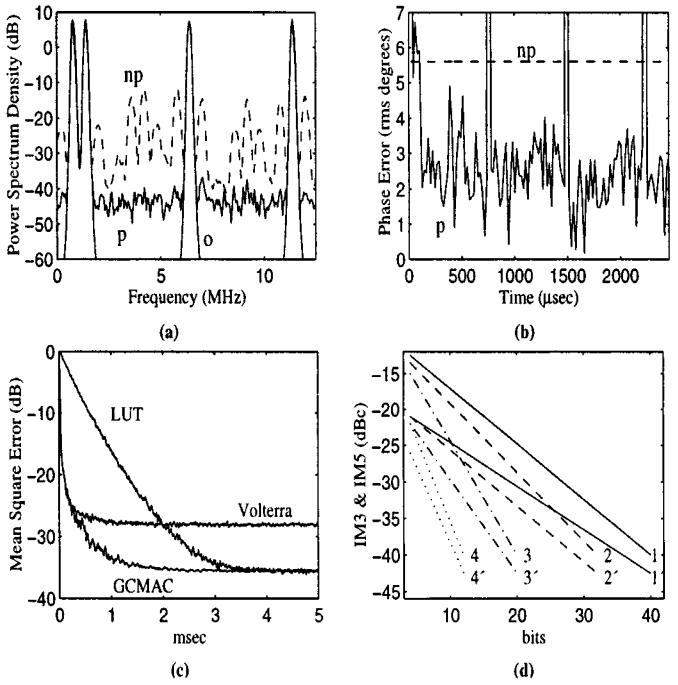

Fig. 9. (a) Multicarrier GMSK power spectrum density, (b) phase error, (c) MSE convergence comparison, and (d) IM3 and IM5 reduction for different OBO.

quencies change every TDMA frame (4.62 ms) [23]. Therefore, efficient predistorters must be used [24]. If only one carrier is amplified, this type of modulation is robust to nonlinear distortion caused by the HPA. We apply predistortion to multicarrier GMSK in a GSM booster to avoid harmonics within the transmitted bandwidth while controlling the phase error within the specifications of the GSM standard [22].

We have simulated a commercial GSM booster following the specifications in [25]. Two blocks define the model. A bandpass filter imposes the linear spectral conditions (amplitude and group delay ripple). In addition, a nonlinear memoryless function (polynomial) of the magnitude produces the output according to IM specifications. Since such a transmitter introduces a negligible time-correlation in the transmitted signal, a 1-length GCMAC signal predistorter has been considered. The input space is bidimensonal (complex signal) and is quantified with 8 bits, the adaptation step is $\beta = 0.01$, and the generalization in both dimensions is $\rho = 128$. Regarding the GSM parameters, we recall that the bit period is $T_b = 3.6923\ \mu$s, and we set the length of the Gaussian premodulating filter to 5 bits and the modulating index to $h = 0.5$. Finally, the system was simulated with a sampling rate of 0.2 ns.

Fig. 9(a) and (b) have been obtained with four carriers comparing the predistortion results, "p", with the nonpredistorted conventional scheme, 'np', along with the original linear case "o". The amplifier working at saturation. Fig. 9(a) includes the PSD. Notice that more than 30-dB difference exists between the "np" line and the "p" one at some frequencies. Fig. 9(b) shows the phase error. The compensation proposed within this work is GSM standard compliant as the limit for this feature is 5° rms. The SFH period has been set to 0.16 times the original in order to appreciate the adaptation features. These signals can be observed in Fig. 3(b). It is interesting to point out that the booster is a simple repeater. Thus, there is no option to introduce other predistortion techniques such as peaking control [26] to improve the efficiency of the amplifier.

The learning rate is studied in Fig. 9(c), comparing GCMAC, LUT, and third-order Volterra solutions for a fixed complexity. The results show that the GCMAC-based architecture outperforms present predistorters. The reduction of the IM along the number of bits is represented in Fig. 9(d) for two carriers. The lines 1-1', 2-2', 3-3', and 4-4' differ in the OBO set to 0, 1, 4, and 10 dB, respectively. In the same manner, the groups 1-2-3-4 and 1'-2'-3'-4' correspond to the cases of third- and fifth-order IM (IM3 and IM5).

X. Conclusion

In this paper, we have proposed and analyzed new data and signal predistorters based on a reduced-complexity neural network called Generalized CMAC (GCMAC).

We have proposed ahybrid structure for symbol-rate data predistortion composed of a 1-length LUT and a P-length CMAC. The proposed scheme achieves low error with fast convergence at the expense of a rigid model to approximate the predistortion function. A much more flexible predistortion scheme is based on the GCMAC network. The GCMAC allows different generalization factors along each direction of input space. It makes this network appropriate for modeling the special features of the predistortion function with high accuracy. The GCMAC achieves performance similar to other structures with much more complexity, such as the Volterra filters or the full LUT. Power gains over 3 dB (6 dB in a 64-QAM system) show that the apparent weakness of the QAM techniques, when used with nonlinear amplifiers, can be effectively overcome.

In order to achieve spectrum control at frequencies beyond the inverse of the symbol period, FSDPs have been proposed. In this way, the level of side-lobes are easily reduced up to 15 dB. These results show that either the use of post-HPA filters or the increase of channel spacing can be relaxed.

Even though standard predistortion requires the feedback of the output of the nonlinear source, it is possible to compensate for remote nonlinearities such as satellite transponders. Thus, a new architecture is proposed for this purpose and its robustness against noise was tested.

Finally, the GCMAC network was introduced into the signal predistortion architecture. We applied it to a GSM booster where tests and simulations were performed. The obtained results show that reductions on IM were significant and that the phase error obtained was standard compliant. Current Base Transceivers Stations, reassignment of cells due to a failure in one unit, cells with large coverage, or even terminals may also take advantage of nonlinear compensation.

Acknowledgment

The authors would like to acknowledge the suggestions of the anonymous reviewers.

REFERENCES

[1] I. W. J. Weber, "The use of TWT amplifiers in m-ary amplitude and phase-shift keying systems," in *Int. Conf. on Communications*, vol. 3, June 1975, pp. 17–21.

[2] D. Falconer, "Adaptive equalization of channel nonlinearities in QAM data transmission," *Bell Syst. Technol. J.*, vol. 57, pp. 2589–2611, Sept. 1978.

[3] S. Pupolin and L. J. Greenstein, "Performance analysis of digital radio links with nonlinear transmit amplifiers," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 534–546, Apr. 1987.

[4] P. S. Leung and K. Feher, "A superior modulation technique for mobile and personal communications," *IEEE Trans. Broadcasting*, vol. 39, pp. 288–294, June 1993.

[5] K. Feher, *Wireless Digital Communications: Modulation and Spread Spectrum Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[6] D. Subasinghe and K. Feher, "Baseband pulse shaping techniques for $\pi/4$-DQPSK in nonlinealy amplified land mobile channels," in *Proc. 41ST IEEE Vehicular Technology Conf.*, St. Louis, MO, May 1991, pp. 759–764.

[7] A. Saleh and J. Salz, "Adaptive linearization of power amplifiers in digital radio systems," *Bell Syst. Tech. J.*, vol. BSTJ-62, pp. 1019–1033, Apr. 1983.

[8] G. Karam and H. Sari, "A data predistortion technique with memory for QAM radio systems," *IEEE Trans. Commun.*, vol. 39, pp. 336–344, Feb. 1991.

[9] E. Biglieri, S. Barberis, and M. Catena, "Analysis and compensation of nonlinearities in digital transmission systems," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 42–51, Jan. 1988.

[10] A. Bernardini and S. D. Fina, "A new predistortion technique using neural nets," *Signal Processing*, vol. 34, pp. 231–243, 1993.

[11] N. Benvenuto, F. Piazza, and A. Uncini, "A neural network approach to data predistortion with memory in digital radio systems," in *Proc. IEEE Int. Conf. on Communications*, Geneva, Switzerland, 1993, pp. 232–236.

[12] G. Karam and H. Sari, "Data predistortion techniques using intersymbol interpolation," *IEEE Trans. Commun.*, vol. 38, pp. 1716–1723, Oct. 1990.

[13] M.-G. Di-Benedetto and P. Mandarini, "A new analog predistortion criterion with application to high efficiency digital radio links," *IEEE Trans. Commun.*, vol. 43, pp. 2966–2974, Dec. 1995.

[14] J. K. Cavers, "Adaptation behavior of a feedforward amplifier linearizer," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 31–40, Feb. 1995.

[15] F. J. González-Serrano, A. Artés-Rodríguez, and A. R. Figueiras-Vidal, "Generalizing CMAC architecture and training," *IEEE Trans. Neural Networks*, vol. 9, pp. 1509–1514, Nov. 1998.

[16] F. J. González-Serrano, A. R. Figueiras-Vidal, and A. Artés-Rodríguez, "Fourier analysis of the generalized CMAC neural network," *Neural Networks*, vol. 11, pp. 391–396, 1998.

[17] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. COM-29, pp. 1715–1720, Nov. 1981.

[18] C. F. N. Cowan and P. F. Adams, "Non-linear system modeling: Concept and application," in *Proc. Int. Conf. Account. Speech and Processing*, San Diego, CA, Mar. 1984, pp. 45.6.1–45.6.4.

[19] J. Albus, "A new approach to manipulator control: The cerebellar model articulation controller," *J. Dynamic Syst., Measurements and Control*, vol. 63, pp. 220–227, Sept. 1975.

[20] F. J. González-Serrano and J. J. Murillo-Fuentes, "Adaptive nonlinear compensation for CDMA communication systems," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 34–42, Jan. 2000.

[21] G. Karam and H. Sari, "Analysis of predistortion, equalization and ISI cancellation techniques in digital radio systems with nonlinear transmit amplifiers," *IEEE Trans. Commun.*, vol. 37, pp. 1245–1253, Dec. 1989.

[22] "ETSI/GSM Recommendation 05.01, Modulation," ETSI, 1996.

[23] "ETSI/GSM Recommendation 05.02, Multiplexing and Multiple Access on the Radio Path," ETSI, 1996.

[24] J. Murillo-Fuentes and F. González-Serrano, "Applying GCMAC to predistortion in GSM base stations," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. V, Phoenix, AZ, Mar. 1999, pp. 2575–2578.

[25] *Manual N AE02B-A0170-001 for Models 385 602-3120, 385 602-3220, 385 602-3320*, Andrew.

[26] B. M. Popovic, "Synthesis of power efficient multitone signals with flat amplitude spectrum," *IEEE Trans. Commun.*, vol. 39, p. 1031–1033, July 1991.

Francisco-Javier González-Serrano (M'94) was born in Betanzos, Spain, on October 10, 1968. He received the Telecommunication Engineering and Ph.D. degrees from the University of Vigo, Spain, in 1992 and 1997, respectively.

From 1992 to 1993, he was with the DSP Research Group at the University of Vigo as a Research Associate. From 1993 to 1998, he sucessively was an Assistant and Associate Professor with the University of Vigo. Currently, he is an Associate Professor with the University Carlos III of Madrid, Spain. His research interests are in the area of digital signal processing, wireless communications, and multimedia communications.

Juan José Murillo-Fuentes (S'00) received the Telecommunication Engineering degree from the University of Seville, Spain, in 1996. He is currently working toward the Ph.D. degree at the University Carlos III, Madrid, Spain.

Since 1999, he has been an Assistant-Professor at the same university. His main research interests include neural networks and their applications to communications, array processing with emphasis in algorithms for blind separation of sources and multiuser detection in third-generation systems, and image processing applied to teledetection.

Antonio Artés-Rodríguez (M'89) was born in Alhama de Almería, Spain, in 1963. He received the Ingeniero de Telecomunicación and Doctor Ingeniero de Telecomunicación degrees, both from the Universidad Politécnica de Madrid, Spain, in 1988 and 1992, respectively.

He is now an Associate Professor at the Departamento de Tecnologías de las Comunicaciones, Universidad Carlos III, Madrid, Spain. His research interests include detection, estimation, and statistical learning methods, and their application to signal processing and communication.