Adaptive Predistortion With Direct Learning Based on Piecewise Linear Approximation of Amplifier Nonlinearity

Sungho Choi, *Student Member, IEEE*, Eui-Rim Jeong, *Member, IEEE*, and Yong H. Lee, *Senior Member, IEEE*

***Abstract*—We propose an efficient Wiener model for a power amplifier (PA) and develop a direct learning predistorter (PD) based on the model. The Wiener model is formed by a linear filter and a memoryless nonlinearity in which AM/AM and AM/PM characteristics are approximated as piecewise linear and piecewise constant functions, respectively. A two-step identification scheme, wherein the linear portion is estimated first and the nonlinear portion is then identified, is developed. The PD is modeled by a polynomial and its coefficients are directly updated using a recursive least squares (RLS) algorithm. To avoid implementing the inverse of the PA's linear portion, the cost function for the RLS algorithm is defined as the sum of differences between the output of the PA's linear portion and the inverse of the PA's nonlinear portion. The proposed direct learning scheme, which is referred to as the piecewise RLS (PWRLS) algorithm, is simpler to implement, yet exhibits comparable performance, as compared with existing direct learning schemes.**

***Index Terms*—Direct learning, piecewise linear, polynomial, power amplifier (PA), predistortion, Wiener model.**

I. Introduction

DIGITAL baseband predistortion has been recognized as a cost effective technique for linearizing PAs. In this scheme, the PA input signal is distorted by a predistorter (PD) whose characteristics are the inverse of those of the amplifier. Both PAs and PDs are often modeled by using a polynomial, and PD coefficients are adaptively adjusted [1]–[10]. There are two types of learning techniques for PD parameter adaptation, direct learning [1], [2] and indirect learning methods [3]–[10]. The former adaptively identifies PA model parameters and directly updates PD coefficients based on the PD input and the PA output, while the latter employs an additional device called an adaptive post-inverse filter in the feedback path to identify inverse characteristics of the PA, and coefficients of the post-inverse filter are copied by the PD. Although the direct learning approach can perform better than indirect learning [2], the former is less popular because the PA model parameters need to be identified.

In this paper, we develop a direct learning adaptive PD based on a simple Wiener PA model [12]–[14] whose parameters can be efficiently identified. The proposed Wiener model consists of a linear time invariant (LTI) system followed by a memoryless nonlinearity in which AM/AM and AM/PM characteristics are approximated as piecewise *linear* and piecewise *constant* functions, respectively. A two-step scheme is applied to PA identification: in the first step, LTI system parameters are identified using input samples located in the linear region and in the second step, given estimates of the linear parameters, the memoryless nonlinearity parameters are identified for each segment of the piecewise approximation.[1] Due to the approximation of AM/AM and AM/PM characteristics, the number of parameters to be estimated for each segment is *two*. At each iteration, during adaptive identification, only those two parameters of the segment in which the input sample is located are updated.

The proposed PA model simplifies implementation of the adaptive PD that is modeled by a polynomial. In contrast to the direct learning methods in [1] and [2], which require a linear time-varying prefilter[2] to yield a reference input signal for adaptation, the proposed PD can directly update its parameters without any prefiltering once PA model parameters are given.

The RLS algorithm developed for updating PD parameters evaluates the difference between the output of the PA's linear portion and the inverse of the PA's nonlinear portion which can be implemented using a simple look-up table listing the nonlinear model parameters. In this manner, the proposed scheme avoids implementing the inverse of the PA's linear portion and any linear time-varying filter. It will be shown that the performance of the proposed PD is comparable to that of the existing methods in [2], and the former is considerably simpler to implement.

The organization of this paper is as follows. Section II describes the system model with direct learning. The two-step process for identifying PA model parameters is presented in Section III. In Section IV, the proposed algorithm is derived and its complexity is compared with those of existing algorithms.

Manuscript received May 29, 2008; revised February 10, 2009. Current version published May 15, 2009. This work was supported by the IT R&D program of MKE/IITA in Korea (2008-F-004-01). This paper was presented in part at the IEEE Vehicular Technology ConferenceDublin, Ireland, April 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ali Sayed.

S. Choi and Y. H. Lee are with the School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: shchoi@stein.kaist.ac.kr; yohlee@ee.kaist.ac.kr).

E.-R. Jeong is with the Division of Information Communication and Computer Engineering, Hanbat National University, Daejeon 305-719, Korea (e-mail: erjeong@hanbat.ac.kr).

Digital Object Identifier 10.1109/JSTSP.2009.2020265

[1]A two-step scheme for nonlinear Wiener model identification is introduced in [13] and outlined in [14]. The Wiener model considered there was formed by a linear filter and a memoryless nonlinearity with known shape.

[2]Such a filter is called an instantaneous equivalent linear (IEL) filter in [2] and its coefficients are obtained from the PA identification block.

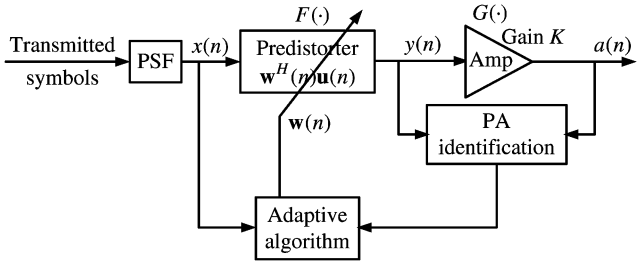

Fig. 1. Transmitter employing direct learning PD.

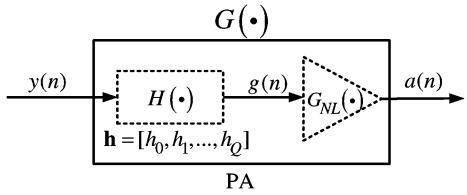

Fig. 2. Wiener system model for a PA with memory.

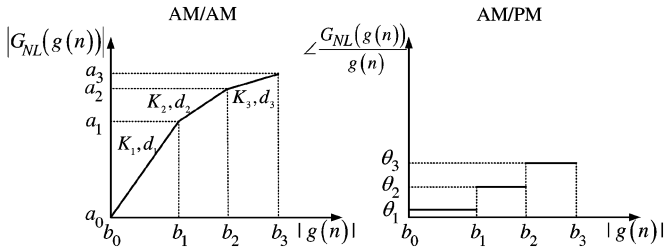

Fig. 3. Piecewise linear model when $\mathrm{M} = 3$.

The performance of the proposed method is examined through a computer simulation in Section V. Finally, Section VI presents conclusions.

II. System Model

Fig. 1 shows the baseband equivalent model of a direct-learning PD. Transmitted symbols are band-limited by a pulse shaping filter (PSF) to produce $x(n)$, which is then predistorted to yield $y(n)$ and amplified by the PA of desired gain K. The input/output relations of the PA and PD are denoted as $a(n) = G(y(n))$ and $y(n) = F(x(n))$, where $G(\cdot)$ and $F(\cdot)$ represent PA and PD characteristics, respectively. Because an ideal PD satisfies

$$a(n) = G\left(F\left(x(n)\right)\right) = Kx(n) \tag{1}$$

our objective is to find $F(\cdot)$ that satisfies (1). Next, we introduce models for PAs and PDs.

A. Modeling a Power Amplifier and Its Inverse

A PA with memory can be modeled as a linear time invariant (LTI) system followed by a memoryless nonlinearity (Fig. 2), which is referred to as the Wiener system model. Let $y(n)$ and $g(n)$ denote the input to the PA and the output of the PA's linear portion, respectively. $g(n)$ is then written as

$$\begin{aligned} g(n) &= \sum_{q=0}^{Q} h_q^* y(n-q) \\ &= \mathbf{h}^H \mathbf{y}(n) \end{aligned} \tag{2}$$

where $\mathbf{h} = [h_0, h_1, \ldots, h_Q]^T$ denotes the impulse response of the LTI system and $\mathbf{y}(n) = [y(n), y(n-1), \ldots, y(n-Q)]^T$. We assume that

$$\sum_{q=0}^{Q} |h_q| = 1 \tag{3}$$

and make the following observation.

Observation 1: If $|y(n-q)| \in [0, b_1)$ for all q, $0 \le q \le Q$, then $|g(n)| \in [0, b_1)$, where b_1 is a positive constant.

This is a direct consequence of (3). Observation 1 will be useful for identifying the LTI system, and the assumption in (3) will be justified after introducing the nonlinearity model. The nonlinearity is represented as

$$a(n) = G_{\mathrm{NL}}\left(g(n)\right)$$

where $g(n)$ and $a(n)$ denote the input and the output of the PA's nonlinear portion, respectively, and $G_{\mathrm{NL}}(\cdot)$ is a nonlinear function whose magnitude $|G_{\mathrm{NL}}(g(n))|$ is a piecewise linear, continuous and monotonically increasing function of $|g(n)|$, and the phase $\angle(G_{\mathrm{NL}}(g(n))/g(n))$ is piecewise constant (Fig. 3). To be more specific, we define sets of intervals $I_b = \{[b_0, b_1), [b_1, b_2), \ldots, [b_{M-1}, b_M)\}$ for the input magnitude $|g(n)|$ and $I_a = \{[a_0, a_1), [a_1, a_2), \ldots, [a_{M-1}, a_M)\}$ for the output magnitude $|a(n)|$, where $0 = b_0 < b_1 < \ldots < b_M$, $0 = a_0 < a_1 < \ldots < a_M$, and $a_m = |G_{\mathrm{NL}}(b_m)|$ for all m. Suppose that $|g(n)| \in [b_{m-1}, b_m)$ at time n, $1 \le m \le M$, then the PA output is written as $|a(n)| = K_m|g(n)| + d_m$ (AM/AM) and $\angle(a(n)/g(n)) = \theta_m$ (AM/PM), where K_m and d_m are the slope and intercept, respectively, and θ_m is the phase distortion in the mth segment. Combining these, the output can be written as

$$\begin{aligned} a(n) &= G_{\mathrm{NL}}\left(g(n)\right) = \left(K_m\,|g(n)| + d_m\right) e^{j\left(\angle g(n) + \theta_m\right)} \\ &\triangleq K_{c,m} g(n) + d_{c,m} e^{j\angle g(n)} \end{aligned} \tag{4}$$

where $K_{c,m} = K_m e^{j\theta_m}$ and $d_{c,m} = d_m e^{j\theta_m}$. It is assumed that $K_1 = K$ and $d_1 = \theta_1 = 0$. This assumption guarantees a PA exhibiting ideal performance when $|g(n)| \in [b_0, b_1)$. The model in (4) can be thought of as a special case of the block-based model in [11], where the set of intervals I_b is defined and polynomial modeling is performed at each interval. Note that the overall characteristic of a PA with memory is denoted by $a(n) = G(y(n))$ and the characteristics of a memoryless nonlinearity are denoted as $a(n) = G_{\mathrm{NL}}(g(n))$. For a memoryless PA, $g(n) = y(n)$ and $G(\cdot) = G_{\mathrm{NL}}(\cdot)$. Next, we present an observation which justifies the assumption in (3).

Observation 2: Consider the Wiener model $G(\cdot)$ which is a cascade of $H(\cdot)$ and $G_{\mathrm{NL}}(\cdot)$ in (4). Let $H'(\cdot) = \alpha H(\cdot)$, where α is a real-valued constant ($\alpha > 0$), and $G'_{\mathrm{NL}}(\cdot)$ be a piecewise nonlinearity whose parameters $\{K'_m, d'_m, \theta'_m\}$ are defined over the sets of intervals I'_a and I'_b, where $I'_a = I_a$ and

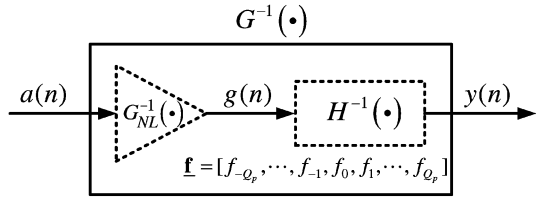

Fig. 4. Inverse of the Wiener model where $H^{-1}(\cdot)$ is approximated as a noncausal FIR equalizer with impulse response $\underline{\mathbf{f}}$.

$I_b' = \{[0, \alpha b_1), [\alpha b_1, \alpha b_2), \cdots, [\alpha b_{M-1}, \alpha b_M)\}$. Then the cascade of $H'(\cdot)$ and $G_{\mathrm{NL}}'(\cdot)$, denoted by $G'(\cdot)$, is identical to $G(\cdot)$, if $K_m' = (1/\alpha)K_m$, $d_m' = d_m$, $\theta_m' = \theta_m$.

Proof: Let $g'(n) = H'(y(n))$, and assume that $|g(n)| \in [b_{m-1}, b_m)$. Then $g'(n) = \alpha g(n)$ and $|g'(n)| \in [\alpha b_{m-1}, \alpha b_m)$. From (4), $a'(n) = G'(y(n)) = G_{\mathrm{NL}}'(g'(n)) = (K_m'|g'(n)| + d_m')e^{j(\angle(g'(n))+\theta_m')} = (K_m|g(n)| + d_m)e^{j\angle g(n)+\theta_m} = G_{\mathrm{NL}}(g(n)) = G(y(n)) = a(n)$ where the fourth equality holds when $K_m' = (1/\alpha)K_m$, $d_m' = d_m$, and $\theta_m' = \theta_m$, and α is real. Therefore, $G_{\mathrm{NL}}'(\cdot)$ defined over I_a and I_b' is identical to $G_{\mathrm{NL}}(\cdot)$ and $G'(\cdot) = G(\cdot)$. ■

This observation indicates that it is always possible to model the PA's linear portion as $H(\cdot)$ satisfying (3).

The inverse of the PA model $G^{-1}(\cdot)$, which is useful for deriving a PD, consists of $G_{\mathrm{NL}}^{-1}(\cdot)$ followed by $H^{-1}(\cdot)$ (Fig. 4). To derive the inverse of $|G_{\mathrm{NL}}(g(n))|$, we write $|G_{\mathrm{NL}}(g(n))| = |a(n)| \triangleq M_G(|g(n)|)$. Note that the inverse $M_G^{-1}(|a(n)|)$ can be determined uniquely for $|a(n)| \in [a_{m-1}, a_m)$, $1 \le m \le M$ because $M_G(|g(n)|)$ is continuous and monotonically increasing. From (4), $M_G^{-1}(|g(n)|)$ is given by

$$M_G^{-1}(|a(n)|) = |g(n)| = \frac{1}{K_m}(|a(n)| - d_m) \tag{5}$$

when $|a(n)| \in [a_{m-1}, a_m)$, $1 \le m \le M$. Then $G_{\mathrm{NL}}^{-1}(a(n))$ can be written as

$$G_{\mathrm{NL}}^{-1}(a(n)) = g(n) = \frac{1}{K_m}(|a(n)| - d_m)\, e^{j(\angle a(n) - \theta_m)}. \tag{6}$$

Since the PA's linear portion $H(\cdot)$ is modeled by a finite-impulse response (FIR) filter, the inverse $H^{-1}(\cdot)$ is an all-pole infinite-impulse response (IIR) filter which becomes noncausal when some zeros of $H(\cdot)$ are located outside of the unit circle in the z-transform domain. To avoid the difficulty in implementing a noncausal IIR filter, $H^{-1}(\cdot)$ is approximated by a noncausal FIR *equalizer* of span $2Q_p + 1$ [17] whose impulse response is denoted as $\underline{\mathbf{f}} = [f_{-Q_p}, \ldots, f_{-1}, f_0, f_1, \ldots, f_{Q_p}]$. The output of the inverse system $y(n)$ is given by

$$\begin{aligned} y(n) &\cong \sum_{i=-Q_p}^{Q_p} f_i^* g(n-i) \\ &= \underline{\mathbf{f}}^H \underline{\mathbf{g}}(n) \end{aligned} \tag{7}$$

where $\underline{\mathbf{g}}(n) = [g(n-Q_p), \ldots, g(-1), g(0), g(1), \ldots, g(n+Q_p)]^T$. Using (6) in (7), $G^{-1}(a(n))$ is written as

$$\begin{aligned} G^{-1}(a(n)) &= y(n) \\ &\cong \sum_{i=-Q_p}^{Q_p} f_i^* \left(\frac{1}{K_{m(i)}} \left(|a(n-i)| - d_{m(i)} \right) \right. \\ &\qquad \left. \times e^{j\left(\angle a(n-i) - \theta_{m(i)}\right)} \right) \end{aligned} \tag{8}$$

where $m(i)$ is an integer, $1 \le m(i) \le M$, and $\{K_{m(i)}, d_{m(i)}, \theta_{m(i)}\}$ represent the nonlinearity parameters corresponding to $a(n-i)$. For example, if $|a(n-i)| \in [0, a_1)$ then $m(i) = 1$ and $\{K_{m(i)}, d_{m(i)}, \theta_{m(i)}\} = \{K_1, d_1, \theta_1\}$.

B. Modeling Predistorters

We model the PD by a polynomial of order $2P - 1$. When the PA is assumed to be memoryless, a memoryless polynomial can be used for the PD and $F(\cdot)$ in (1) can be represented as

$$\begin{aligned} F(x(n)) &= \sum_{p=0}^{P-1} w_p^*(n) x(n) |x(n)|^{2p} \\ &= \mathbf{w}^H(n)\mathbf{u}(n) \end{aligned} \tag{9}$$

where

$$\begin{aligned} \mathbf{w}(n) &= [w_0(n), w_1(n), \ldots, w_{P-1}(n)]^T \\ \mathbf{u}(n) &= x(n)\left[1, |x(n)|^2, \ldots, |x(n)|^{2(P-1)}\right]^T. \end{aligned}$$

The weight vector $\mathbf{w}(n)$ is determined so that the PD based on (9) approximates $G_{\mathrm{NL}}^{-1}(\cdot)$.

When memory effects are presented in a PA, a polynomial with memory depth $2Q_p + 1$ is used

$$\begin{aligned} F(x(n)) &= \sum_{p=0}^{P-1} \sum_{q=-Q_p}^{Q_p} w_{p,q}^*(n) x(n-q) |x(n-q)|^{2p} \\ &= \underline{\mathbf{w}}^H(n)\underline{\mathbf{u}}(n) \end{aligned} \tag{10}$$

where

$$\begin{aligned} \underline{\mathbf{w}}(n) &= \left[\mathbf{w}_{-Q_p}^T(n), \ldots, \mathbf{w}_{-1}^T(n), \right. \\ &\qquad \left. \mathbf{w}_0^T(n), \mathbf{w}_1^T(n), \ldots, \mathbf{w}_{Q_p}^T(n)\right]^T \\ \mathbf{w}_q(n) &= [w_{0,q}(n), w_{1,q}(n), \ldots, w_{P-1,q}(n)]^T \\ \underline{\mathbf{u}}(n) &= \left[\mathbf{u}_{-Q_p}^T(n), \ldots, \mathbf{u}_{-1}^T(n), \right. \\ &\qquad \left. \mathbf{u}_0^T(n), \mathbf{u}_1^T(n), \ldots, \mathbf{u}_{Q_p}^T(n)\right]^T \\ \mathbf{u}_q(n) &= x(n-q)\left[1, |x(n-q)|^2, \ldots, |x(n-q)|^{2(P-1)}\right]^T. \end{aligned}$$

Note that the polynomial in (10) is noncausal and its memory depth is identical to the span of the FIR equalizer in (7). This is because the polynomial should approximate $G^{-1}(\cdot)$ which is a cascade of $G_{\mathrm{NL}}^{-1}(\cdot)$ and $H^{-1}(\cdot)$.

III. PA Identification

The Wiener system model in Fig. 2 is identified by a two-step scheme that estimates the linear portion first and then identifies the nonlinear portion. Recursive least squares (RLS) algorithms [16] are used for both the linear and nonlinear system identifications.

A. Step 1: LTI System Identification

For a given PA, we assume the knowledge of typical AM/AM and AM/PM characteristic curves. From these curves, we intuitively determine the interval $[0, b_1)$ over which the PA exhibits linear characteristics. Let n_i, $i = 1, 2, \ldots$, denote time indices indicating the occurrence of the event "$|y(n_i - q)| \in [0, b_1)$ for all q, $0 \le q \le Q$" (magnitudes of $Q+1$ consecutive samples $\{y(n_i - Q), \ldots, y(n_i)\}$ are located in the first interval $[0, b_1)$). Then, from Observation 1, $|g(n_i)| \in [0, b_1)$ and the nonlinearity in Fig. 2 can be ignored whenever $\{g(n_i)\}$ are entered, because $K_1 = K$ and $d_1 = \theta_1 = 0$. Therefore, the identification problem reduces to a linear estimation problem at time indices $\{n_i\}$. Let $\mathbf{h}' = K\mathbf{h}$. The cost function for estimating $\mathbf{h}'$ can be written as

$$\mathcal{E}_{\mathbf{h}'}(k) = \sum_{i=1}^{k} \lambda^{k-i} \left| e_{\mathbf{h}'}(i) \right|^2$$

where $e_{\mathbf{h}'}(i) = a(n_i) - \hat{\mathbf{h}'}^H(n_k)\mathbf{y}(n_i)$, $\hat{\mathbf{h}'}(n_k)$ is an estimate of $\mathbf{h}'$, $\mathbf{y}(n_i) = [y(n_i), y(n_i - 1), \ldots, y(n_i - Q)]^T$ and λ is a forgetting factor, $0 < \lambda \le 1$. The RLS algorithm for obtaining the optimal $\hat{\mathbf{h}'}(n_k)$ minimizing $\mathcal{E}_{\mathbf{h}'}(k)$ can be derived as

$$\boldsymbol{\beta}(n_k) = \frac{\lambda^{-1}\mathbf{L}(n_{k-1})\mathbf{y}(n_k)}{1 + \lambda^{-1}\mathbf{y}^H(n_k)\mathbf{L}(n_{k-1})\mathbf{y}(n_k)} \tag{11a}$$

$$\hat{\mathbf{h}'}^H(n_k) = \hat{\mathbf{h}'}^H(n_{k-1}) + \boldsymbol{\beta}(n_k) \times \left(a(n_k) - \hat{\mathbf{h}'}^H(n_{k-1})\mathbf{y}(n_k) \right) \tag{11b}$$

$$\mathbf{L}(n_k) = \lambda^{-1}\mathbf{L}(n_{k-1}) - \lambda^{-1} \times \boldsymbol{\beta}(n_k)\mathbf{y}^H(n_k)\mathbf{L}(n_{k-1}) \tag{11c}$$

where $\boldsymbol{\beta}(n_k) \in \mathbb{C}^{Q+1}$ is a gain vector and $\mathbf{L}(n_k) \in \mathbb{C}^{(Q+1)\times(Q+1)}$ is the inverse correlation matrix. Because the RLS algorithm is convergent in the mean value after $2(Q+1)$ iterations [16], the parameters in (11a)–(11c) tend to converge when n_k becomes $n_{2(Q+1)}$. The time index $n_{2(Q+1)}$ indicates that the event "$|y(n_i - q)| \in [0, b_1)$ for all $0 \le q \le Q$" occurs $2(Q+1)$ times over the set of samples $\{y(0), y(1), \ldots, y(n_{Q+1})\}$. In practice, $n_{2(Q+1)}$ would not be large, because usually Q is small and the interval $[0, b_1)$ is wider than the other intervals $[b_{m-1}, b_m)$, $2 \le m \le M$. After the convergence, $\hat{\mathbf{h}'}$ is normalized by $\sum_{q=0}^{Q} |h'_q|$ to yield the desired estimate $\hat{\mathbf{h}}$, where $\{h'_q\}$ are the elements of $\hat{\mathbf{h}'}$. Note that $\sum_{q=0}^{Q} |h'_q|$ is an estimate of K, and $\hat{\mathbf{h}}$ satisfies the assumption in (3).

B. Step 2: Identification of Nonlinearity

In this step, the estimate $\hat{\mathbf{h}}$ obtained in Step 1 is fixed and the parameters $\{K_{c,m}\}$ and $\{d_{c,m}\}$ in (4) are estimated. Due to Observation 1, we keep the first interval $[0, b_1)$ the same as in Step 1. For simplicity, the other intervals $\{[b_1, b_2), [b_2, b_3), \cdots [b_{M-1}, b_M)\}$ are assumed to be equispaced. For each m, $2 \le m \le M$, let n_i^m, $i = 1, 2, \ldots$ denote time indices indicating the occurrence of the event "$b_{m-1} < |\hat{g}(n_i^m)| \le b_m$" where $\hat{g}(n_i^m) = \hat{\mathbf{h}}^H\mathbf{y}(n_i^m)$. The cost function for estimating $\{K_{c,m}, d_{c,m}\}$ of the mth interval is given by

$$\mathcal{E}_m(k) = \sum_{i=1}^{k} \lambda^{n-i} \left| e_m(i) \right|^2$$

where $e_m(i) = a(n_i^m) - \hat{\mathbf{k}}_m^H(n_k^m)\hat{\mathbf{g}}(n_i^m)$, $\hat{\mathbf{k}}_m(n_i^m) = [K_{c,m}(n_i^m), d_{c,m}(n_i^m)]^T$, and $\hat{\mathbf{g}}_m(n_i^m) = [\hat{g}(n_i^m), e^{j\angle\hat{g}(n_i^m)}]^T$. The RLS algorithm for obtaining the optimal $\hat{\mathbf{k}}_m(n_k^m)$ minimizing $\mathcal{E}_m(k)$ can be derived as

$$\boldsymbol{\alpha}_m(n_k^m) = \frac{\lambda^{-1}\mathbf{P}_m\left(n_{k-1}^m\right)\hat{\mathbf{g}}_m\left(n_k^m\right)}{1 + \lambda^{-1}\hat{\mathbf{g}}_m^H\left(n_k^m\right)\mathbf{P}_m\left(n_{k-1}^m\right)\hat{\mathbf{g}}_m\left(n_k^m\right)} \tag{12a}$$

$$\hat{\mathbf{k}}_m^H(n_k^m) = \hat{\mathbf{k}}_m^H\left(n_{k-1}^m\right) + \boldsymbol{\alpha}_m\left(n_k^m\right) \times \left(a\left(n_k^m\right) - \hat{\mathbf{k}}_m^H\left(n_{k-1}^m\right)\hat{\mathbf{g}}_m\left(n_k^m\right) \right) \tag{12b}$$

$$\mathbf{P}_m(n_k^m) = \lambda^{-1}\mathbf{P}_m\left(n_{k-1}^m\right) - \lambda^{-1}\boldsymbol{\alpha}_m\left(n_k^m\right)\hat{\mathbf{g}}_m^H\left(n_k^m\right)\mathbf{P}_m\left(n_{k-1}^m\right) \tag{12c}$$

where $\boldsymbol{\alpha}_m(n_k^m) \in \mathbb{C}^2$ is a gain vector and $\mathbf{P}_m(n_k^m) \in \mathbb{C}^{2\times 2}$ is an inverse correlation matrix. Because the number of parameters to be estimated in $\hat{\mathbf{k}}_m$ is two, the parameters in (12a)–(12c) are convergent in the mean value when n_k^m becomes n_4^m (after four iterations). As m approaches M, the value of n_4^m tends to increase, because the probability that the event "$b_{m-1} < |\hat{g}(n_i)| \le b_m$" occurs tends to decrease. Using (12a)–(12c), $\{K_{m,c}, d_{m,c}\}$ are estimated for all m, $2 \le m \le M$, and converted into $\{K_m, d_m, \theta_m\}$. The results are saved in a look-up table. Although $|G_{\mathrm{NL}}(g(n))|$ associated with the estimates of $\{K_m, d_m, \theta_m\}$ is not necessarily continuous, the estimates can be used for evaluating $G_{\mathrm{NL}}^{-1}(\cdot)$ in (6). It is recommended that the PD is inactive during the initial identification process. After the initialization, parameters of $H(\cdot)$ and $G_{\mathrm{NL}}(\cdot)$ can be updated, whenever necessary, while the PD is active.

IV. Development of the Proposed PD Algorithm

RLS algorithms for determining PD parameters are described, assuming that the parameters $\mathbf{h}$ and $\{K_c, d_c, \theta_m\}$ are given. The proposed scheme, which is referred to as the *piecewise* RLS (PWRLS), is developed for a memoryless PA first and extended for a PA with memory.

A. Memoryless PA Case

Referring to (1) and (9), our objective is to make the output $\mathbf{w}^H(n)\mathbf{u}(n)$ of the PD close to the ideal output $G_{\mathrm{NL}}^{-1}(Kx(n))$. To this end, the following least squares (LS) cost function is considered:

$$\mathcal{E}(n) = \sum_{i=1}^{n} \lambda^{n-i} \left| e(i) \right|^2$$

where

$$e(i) = G_{\mathrm{NL}}^{-1}\left(Kx(i)\right) - \mathbf{w}^H(n)\mathbf{u}(i) \tag{13}$$

TABLE I
NUMBER OF COMPLEX MULTIPLICATIONS PER ITERATION

		PWRLS	NARLS [2]
PD adaptation	PD Filtering	$P(2Q_p+1)+P$	$P(2Q_p+1)+P$
	PD weights update	$3\left(P(2Q_p+1)\right)^2+3\left(P(2Q_p+1)\right)$ $+P(Q+1)+3$	$3\left(P(2Q_p+1)\right)^2+4\left(P(2Q_p+1)\right)$
	IEL filtering	–	$(3P_a+Q)(Q+1)$

and λ is a forgetting factor $(0<\lambda\leq 1)$. In the right-hand side (RHS) of (13), $\mathbf{w}^H(n)\mathbf{u}(i)$ denotes the output, at time i, of the PD with weight $\mathbf{w}^H(n)$. The RLS algorithm for obtaining the optimal $\mathbf{w}(n)$ minimizing $\mathcal{E}(n)$ can be derived as follows [16]:

$$\boldsymbol{\kappa}(n)=\frac{\lambda^{-1}\boldsymbol{\Gamma}(n-1)\mathbf{u}(n)}{1+\lambda^{-1}\mathbf{u}^H(n)\boldsymbol{\Gamma}(n-1)\mathbf{u}(n)} \tag{14a}$$

$$\xi(n)=G_{\mathrm{NL}}^{-1}\left(Kx(n)\right)-\mathbf{w}^H(n-1)\mathbf{u}(n) \tag{14b}$$

$$\mathbf{w}^H(n)=\mathbf{w}^H(n-1)+\boldsymbol{\kappa}(n)\xi^*(n) \tag{14c}$$

$$\boldsymbol{\Gamma}(n)=\lambda^{-1}\boldsymbol{\Gamma}(n-1)-\lambda^{-1}\boldsymbol{\kappa}(n)\mathbf{u}^H(n)\boldsymbol{\Gamma}(n-1) \tag{14d}$$

where $\boldsymbol{\kappa}(n)\in\mathbb{C}^{P\times 1}$ is a gain vector and $\boldsymbol{\Gamma}(n)\in\mathbb{C}^{P\times P}$ is the inverse correlation matrix. From (6) the *a priori error* $\xi(n)$ in (14b) can be written as

$$\xi(n)=\frac{1}{K_m}\left(\left(|Kx(n)|-d_m\right)e^{j\left(\angle Kx(n)-\theta_m\right)}\right)-\mathbf{w}^H(n-1)\mathbf{u}(n) \tag{15}$$

when $|Kx(n)|\in[a_{m-1},a_m), 1\leq m\leq M$. In the PWRLS, the weights $\mathbf{w}(n)$ are updated using (14a), (15), (14c), and (14d).

B. Extension for PA With Memory

For a PA with memory, $G^{-1}(\cdot)$ is a cascade of $G_{\mathrm{NL}}^{-1}(\cdot)$ and $H^{-1}(\cdot)$ (Fig. 4). Therefore, direct extension of the RLS algorithm in (14a)–(14d) for a PA with memory needs the knowledge about $H^{-1}(\cdot)$ (or $\underline{\mathbf{f}}$ in (7)) to implement $G^{-1}(Kx(n))$ using (18). However, obtaining the impulse response of the FIR equalizer $\underline{\mathbf{f}}$ is cumbersome, because its span $2Q_p+1$ is usually much larger than $Q+1$ which is the span of $H(\cdot)$. (This is particularly true when some zeros of $H(\cdot)$ are located near the unit circle.) To avoid implementing $H^{-1}(\cdot)$, we propose to employ the following error sequence in the LS cost function in (13)

$$\begin{aligned}e(i)&=G_{\mathrm{NL}}^{-1}\left(Kx(i)\right)-\sum_{q=0}^{Q}h_q\underline{\mathbf{w}}^H(n)\underline{\mathbf{u}}(i-q)\\&=G_{\mathrm{NL}}^{-1}\left(Kx(i)\right)-\underline{\mathbf{w}}^H(n)\sum_{q=0}^{Q}h_q\underline{\mathbf{u}}(i-q)\\&=G_{\mathrm{NL}}^{-1}\left(Kx(i)\right)-\underline{\mathbf{w}}^H(n)\underline{\mathbf{u}}_{\mathbf{h}}(i)\end{aligned} \tag{16}$$

where $\underline{\mathbf{w}}^H(n)\underline{\mathbf{u}}(i-q)$ is the output, at time $i-q$, of the PD with weight $\underline{\mathbf{w}}^H(n)$ and $\underline{\mathbf{u}}_{\mathbf{h}}(i)\triangleq\sum_{q=0}^{Q}h_q\underline{\mathbf{u}}(i-q)$. Note that $e(i)$ in (16) is the difference between $G_{\mathrm{NL}}^{-1}(\cdot)$ and the output of $H(\cdot)$, and the implementation of $H^{-1}(\cdot)$ is avoided by taking advantage of the cascaded structure of the Wiener model in Fig. 2. To minimize the LS cost function employing the error sequence in (16), we can use the RLS algorithm in (14a)–(14d) after replacing $\{\mathbf{w}(n),\mathbf{u}(n)\}$ with the augmented vectors $\{\underline{\mathbf{w}}(n),\underline{\mathbf{u}}_{\mathbf{h}}(n)\}$ and modifying the *a priori* error $\xi(n)$ in (14b) as follows:

$$\xi(n)=G_{\mathrm{NL}}^{-1}\left(Kx(n)\right)-\underline{\mathbf{w}}^H(n-1)\underline{\mathbf{u}}_{\mathbf{h}}(n). \tag{17}$$

Before concluding this section, we present some remarks of interest.

1) The computational load of the proposed PWRLS is compared with that of the nonlinear adjoint RLS (NARLS) in [2]. In this comparison, the NARLS assumes a Wiener PA model whose nonlinear portion is modeled by a polynomial of order $2P_a-1$ [1]. Table I lists the number of complex multiplications required for implementing the two RLS schemes at each iteration. The computational complexities of the RLS algorithms are almost identical with the exception of the term $(3P_a+Q)(Q+1)$ for IEL filtering which is needed to update NARLS coefficients. This term indicates that the IEL filter induces a considerable computational load. The proposed PWRLS replaces the IEL filter with a simple look-up table listing $\{K_m,d_m,\theta_m\}$ and is simpler to implement than the NARLS.
2) A PD may be modeled by the cascade of $G_{\mathrm{NL}}^{-1}(\cdot)$ and $H^{-1}(\cdot)$ instead of a memory polynomial. The cascaded PD model is not considered in this work, because of the following reasons.
   - As mentioned at the beginning of this section, identification of $H^{-1}(\cdot)$ is usually cumbersome.
   - When $G_{\mathrm{NL}}^{-1}(\cdot)$ is directly used for predistortion, the number of piecewise intervals M should be considerably large (for example, $M\geq 64$) [4]. The proposed scheme does not need such a large M, as shown in the following section.
3) The Hammerstein model [12], which is a cascade of $G_{\mathrm{NL}}(\cdot)$ followed by $H(\cdot)$, may be employed to model a PA instead of the Wiener model. The reasons why the Wiener model is employed are described as follows.
   - An adaptive PD based on the Hammerstein model needs to estimate $H^{-1}(\cdot)$.
   - The Wiener model is more popular for modeling a PA [1], [2], [12].

V. SIMULATION RESULTS

The performance of the proposed PD is demonstrated through a computer simulation. The simulation environments are as follows. The transmitted symbols are modulated by 16 quadrature amplitude modulation (QAM) and pulse-shaped by a digital square root raised cosine filter with roll-off factor of 0.22 whose operating (sampling) rate is 10 times higher than the symbol

rate. As in [2], the PA is modeled by an FIR filter with coefficients [0.7692 0.1538 0.0769] followed by a memoryless nonlinear model given by [15]

$$G(y(n)) = \frac{1.1y(n)}{1+0.3|y(n)|^2}\exp\left(j\frac{0.8|y(n)|^2}{1+3|y(n)|^2}\right).$$

Note that the sum of impulse responses of the FIR filter is 0.9999. The zeros of the filter are given by $\{-0.1 \pm j0.3\}$ and located inside the unit circle.[3] The ideal gain of the PA is assumed to be 1 ($K = 1$), and the peak back-off (PBO) is 3 dB.

Both the proposed PWRLS and the NARLS in [2] are considered. For the NARLS,[4], parameters of the polynomial-based Wiener PA model are estimated by an RLS algorithm minimizing $\mathcal{E}(n) = \sum_{i=1}^{n}\lambda^{n-i}|e(i)|^2$ where $e(i) = a(i) - \hat{a}(i)$ and $\hat{a}(n)$ is the output of the PA model given by

$$\hat{a}(n) = \sum_{p=0}^{P_a-1}\alpha_p\left(\sum_{q=0}^{Q}h_q y(n-q)\right)\left|\sum_{q=0}^{Q}h_q y(n-q)\right|^{2p}. \tag{18}$$

In (18), $\sum_{q=0}^{Q}h_q y(n-q)$ is the output of the linear portion and α_p denotes the coefficients of the polynomial modeling the memoryless nonlinearity (see Fig. 2). The linear and nonlinear model parameters h_q and α_p are estimated alternatively, following the approach in [1]. This method needs $3Q^2 + 11Q + 2P_a + 6$ and $3P_a^2 + 5P_a + Q + 1$ multiplications per iteration for linear and nonlinear portions, respectively. The proposed identification scheme is considerably simpler to implement than the polynomial-based identification based on (18), because the former updates only two parameters for nonlinearity and needs 18 multiplications per iteration. However, cost of this simplicity in implementation is slower convergence speed. This will be shown through simulation results.

All the RLS algorithms for the PA identification and PD start with the initial weight vector being one for the first element and zeros for the others, i.e., $[1, 0, \ldots, 0]^T$, and the initial inverse correlation matrix is a normalized identity matrix whose diagonal values are set at one thousand (1000). The forgetting factor λ is set at 0.995.

A. PA Identification Performance

For both the proposed and polynomial-based Wiener model, Q is set at three ($Q = 3$). Two values of M, $M \in \{5, 15\}$, and two values of P_a, $P_a \in \{3, 4\}$, are considered where M is the number of piecewise segments of the proposed model and $2P_a - 1$ is the polynomial order of the nonlinearity shown in (18). The first segment $[0, b_1)$ is determined so that input magnitudes which are less than 30% of the peak magnitude are located in $[0, b_1)$, and the other segments are equispaced in the remaining 70% region. The linear portion of the proposed PA model is identified during the first 500 sample period and then the nonlinear portion is identified. Fig. 5 shows the learning curves for the MSE, $E[|a(n) - \hat{a}(n)|^2]$, which are obtained by averaging over 500 trials. As expected, the proposed identification scheme converges slower than the polynomial-based scheme, but the former is simpler to implement at each iteration. For the proposed scheme, the speed of convergence decreases as M increases. This is because the probability that "$b_{m-1} < |\hat{g}(n_i)| \leq b_m$" occurs decreases as m increases, $2 \leq m \leq M$. After convergence, the proposed scheme with $M = 5$ and 15 exhibit comparable performance to the polynomial-based schemes with $2P_a - 1 = 5$ and 7, respectively. Of course, the identification methods with larger values of M and P_a perform better than the corresponding schemes with smaller M and P_a.

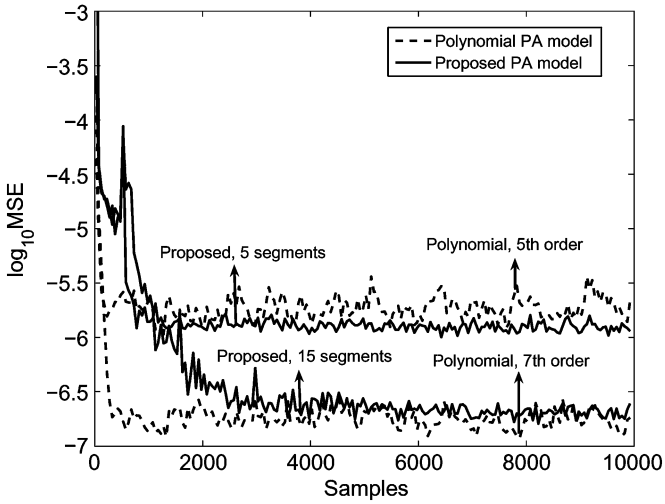

Fig. 5. Learning curves for the proposed and polynomial-based identification schemes. In the proposed scheme, first 500 samples are used for Step 1 and the rest are for Step 2.

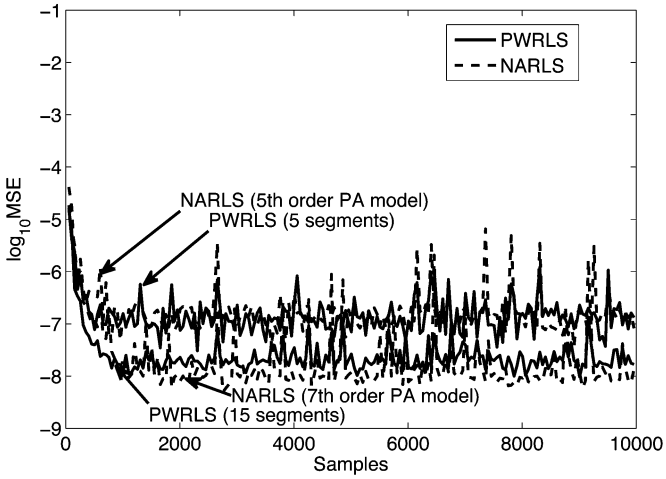

Fig. 6. Learning curves for PWRLS and NARLS in [2].

B. PD Performance

For both the PWRLS and NARLS, the order of PD polynomial is given by $(P, Q_p) = (4, 2)$, and the results of PA identification after 10 000 iterations are used for implementing the adaptive PDs. Fig. 6 shows the learning curves for the MSEs between $x(n)$ (PD input) and $a(n)$ (PA output), $E[|x(n)-a(n)|^2]$, which are obtained by averaging over 500 trials. As expected

[3]An FIR filter with zeros which are located outside the unit circle is also considered in our simulation. The results are not reported here because they lead to observations which are similar to those presented in this section.

[4]In [2], the NARLS is developed under the assumption that PA model coefficients are given.

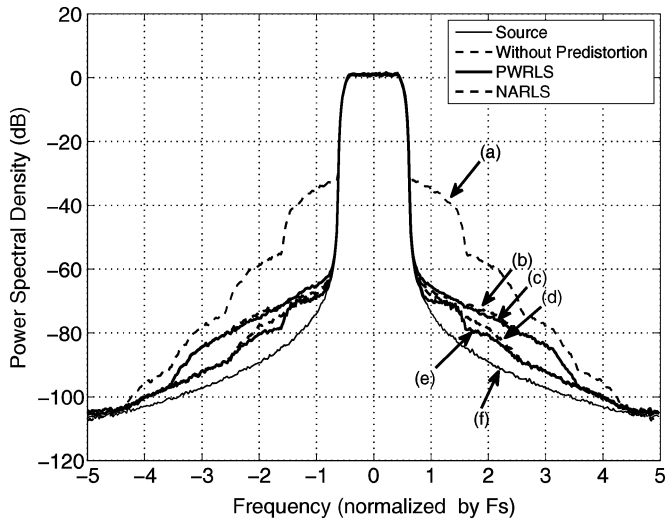

Fig. 7. Comparison of power spectrum density at PA output. (a) Without predistortion. (b) NARLS in [2] with fifth-order polynomial PA model. (c) PWRLS with five segments PA model. (d) NARLS in [2] with seventh-order polynomial PA model. (e) PWRLS with 15 segments PA model. (f) PD input.

from Fig. 5, the PWRLS with $M = 5$ and 15 exhibit comparable performance to the NARLS with $2P_a - 1 = 5$ and 7, respectively, and the adaptive PDs with larger M and P_a perform better than those with smaller M and P_a. Fig. 7 demonstrates the output power spectral density (PSD) performance. The PSDs are in accordance with the MSE performances in Fig. 6. The PWRLS with 15 segments $(M = 15)$ and the NARLS with seventh-order PA model $(2P_a - 1 = 7)$ outperform the others. The PWRLS with 15 segments $(M = 5)$ shows slightly better PSD than the NARLS with 7th order polynomial $(2P_a - 1 = 7)$. These results indicate that the PWRLS can exhibit a comparable linearization performance to the NARLS while the former requires less computational load.

VI. Conclusion

An adaptive scheme, called the PWRLS algorithm for a direct learning PD was developed based on piecewise approximation of PA nonlinearity. The PA is identified via a two step process that sequentially estimates the linear and nonlinear parameters of a PA. Compared with a polynomial-based nonlinear PA model, the piecewise linear approximation reduces the computational load at each iteration, but slows down the rate of convergence. This tradeoff between the simplicity and the rate of convergence would be beneficial for hardware implementation and acceptable in terms of convergence time in practical PDs.

The piecewise linear approximation also simplifies the adaptive PD by replacing the IEL filter, which is needed for the NARLS, by a simple look-up table listing coefficients of the nonlinearity. The simulation results showed that the linearization performance of the PWRLS is comparable to the NARLS. Further work in this direction will include the optimal segmentation for piecewise linear approximation and an analysis on the rate of convergence.

References

[1] H. W. Kang, Y. S. Cho, and D. H. Youn, "On compensating nonlinear distortions of an OFDM system using an efficient adaptive predistorter," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 522–526, Apr. 1999.

[2] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 120–133, Jan. 2007.

[3] Y. Qian and T. Yao, "Structure for adaptive predistortion suitable for efficient adaptive algorithm application," *Electron. Lett.*, vol. 38, no. 21, pp. 1282–1283, Oct. 2002.

[4] J. K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 374–382, Nov. 1990.

[5] R. Marsalek, P. Jardin, and G. Baudoin, "From post-distortion to predistortion for power amplifiers linearization," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 308–310, Jul. 2003.

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

[7] D. R. Morgan, Z. Ma, and L. Ding, "Reducing measurement noise effects in digital predistortion of RF power amplifiers," in *Proc. IEEE ICC*, Seattle, WA, May 2003, pp. 2436–2439.

[8] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[9] L. Ding, Z. Ma, D. R. Morgan, M. Zierdt, and J. Pastalan, "A least-squares/newton method for digital predistortion of wideband signals," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 833–840, May 2006.

[10] Y.-D. Kim, E.-R. Jeong, and Y. H. Lee, "Adaptive compensation for power amplifier nonlinearity in the presence of quadrature modulation/demodulation errors," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4717–4721, Sep. 2007.

[11] N. Safari, J. P. Tanem, and T. Røste, "A block-based predistortion for high power-amplifier linearization," *IEEE Trans. Micorw. Theory Tech.*, vol. 54, no. 6, pp. 2813–2820, Jun. 2006.

[12] R. Raich and G. T. Zhou, "On the modeling of memory nonlinear effects of power amplifiers for communications," in *Proc. Digital Signal Process. Workshop 2nd Signal Process. Education Workshop*, Pine Mountain, GA, Oct. 2002, pp. 7–10.

[13] N. J. Bershad, P. Celka, and J.-M. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for gaussian data and noisy input and output measurements," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 675–689, Mar. 1999.

[14] J. E. Cousseau, J. L. Figueroa, S. Werner, and T. I. Laakso, "Efficient nonlinear wiener model identification using a complex-valued simplicial canonical piecewise linear filter," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1780–1792, May 2007.

[15] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. COM-29, no. 11, pp. 1715–1720, Nov. 1981.

[16] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[17] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2001.

Sungho Choi (S'05) received the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2007, where he is currently pursuing the Ph.D. degree in KAIST.

His research interests are in the area of communication signal processing, which include synchronization, detection, and estimation for wireless communication systems. Currently, he is focusing on the behavioral wideband power amplifier modeling and predistortion for linearization of nonlinear power amplifiers for wireless applications.

Eui-Rim Jeong (M'04) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1995, 1997, and 2001, respectively.

He is currently an Assistant Professor in the Division of Information Communication and Computer Engineering, Hanbat National University, Daejeon, Korea. His research interests include communication signal processing, digital predistortion, and wireless networks.

Yong H. Lee (S'81–M'84–SM'98) was born in Seoul, Korea, in 1955. He received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1978 and 1980, respectively, and the Ph.D. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 1984.

From 1984 to 1988, he was an Assistant Professor with the Department of Electrical and Computer Engineering, State University of New York, Buffalo. Since 1989, he has been with the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, Korea, where he currently is a Professor and the Dean of the College of Information Science and Technology. His research activities are in the areas of physical layer design and communication signal processing, which includes synchronization, interference cancellation, and signal processing for RF devices and MIMO/OFDM systems. He has been actively involved in commercial R&D efforts to design and implement various transceivers. He is an author or a coauthor of more than 160 papers published in international journals and conference proceedings.