



Title	Optimized Low-Complexity Implementation of Least Squares-Based Model Extraction for Digital Predistortion of RF Power Amplifiers
Authors(s)	Guan, Lei, Zhu, Anding
Publication date	2012-03
Publication information	Guan, Lei, and Anding Zhu. "Optimized Low-Complexity Implementation of Least Squares-Based Model Extraction for Digital Predistortion of RF Power Amplifiers." IEEE, March 2012. https://doi.org/10.1109/TMTT.2011.2182656 .
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/8646
Publisher's statement	© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works
Publisher's version (DOI)	10.1109/TMTT.2011.2182656

Downloaded 2026-05-02 00:26:42

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

Optimized Low-Complexity Implementation of Least Squares-Based Model Extraction for Digital Predistortion of RF Power Amplifiers

Lei Guan, *Student Member, IEEE*, and Anding Zhu, *Member, IEEE*

Abstract— Least squares estimation is widely used in model extraction of digital predistortion for RF power amplifiers. In order to reduce computational complexity and implementation cost, it is desirable to use a small number of training samples in the model parameter estimation. However, due to strong correlations between data samples in a real transmit signal, the ill-conditioning problem becomes severe in standard least squares, which often leads to large errors occurring in model extraction. Using a short training sequence can also cause mismatch between the statistical properties of the training data and the actual signal that the amplifier transmits, which could degrade the linearization performance of the digital predistorter. In this work, we propose first to use a 1-bit ridge regression algorithm to eliminate the ill-conditioning problem in the least squares estimation and then use root mean squares-based coefficients weighting and averaging approach to reduce the errors caused by the statistical mismatch. Experimental results show that the proposed approach can produce excellent model extraction accuracy with only a very small number of training samples, which dramatically reduces the computational complexity and the system implementation cost.

Index Terms— digital predistortion, least squares, model extraction, power amplifier.

I. INTRODUCTION

THE increasing demands of higher data rates and wider bandwidths impose severe challenges in designing high efficiency and high linearity radio frequency (RF) power amplifiers (PAs) in wireless transmitters. Digital predistortion (DPD) is proposed to use digital signal processing techniques to compensate for the nonlinear distortion in the RF PA, thereby allowing it to be operated at higher drive levels for higher efficiency [1]. The attraction of this approach is that the nonlinear PA can be linearized by an add-on digital block, freeing vendors from the burden and complexity of manufacturing complex analog/RF circuits. DPD has become one of the most popular linearization techniques in modern

wireless communication systems, especially in high power base stations. As price pressures become more intense, in order to make the system economically competitive, low complexity and low cost solutions are always desirable. The implementation of a DPD mainly includes two parts: (i) the predistortion unit, which nonlinearly processes the input signal before it enters the PA in the transmit chain; and (ii) the model parameter extraction unit, which is used to initialize and update the DPD coefficients. In [2], we proposed an efficient hardware implementation methodology for the predistortion unit of a simplified Volterra series based DPD. In this paper, we focus on the implementation of the model extraction part.

Most DPD models employed today are derived from either polynomials or the Volterra series [3]-[13]. One important observation is that the output of these models is linear with respect to their coefficients. This means that it is possible to extract the nonlinear DPD models in a direct way by using linear system identification algorithms. Because of its fast convergence and high accuracy, the least squares (LS) estimation is widely used for model extraction in DPD systems [12]-[15]. The LS algorithm is a statistical regression-based data fitting method that finds a solution by minimizing the sum of the squares of deviation between the expected and observed data. The best solution can be obtained when the errors among each set of input/output are uncorrelated. In practice, the training samples, which are used for extracting parameters of the DPD, are not independent from each other, and the observation errors are thus not uncorrelated. To reduce the influence of the errors in the observations, the LS estimator often needs a large number of training sequences to derive the best approximation of the DPD coefficients. In other words, the number of samples used to build the LS matrices is much larger than that of unknown parameters. For example, a couple of thousand data samples are normally used [12]-[15]. Since matrix-matrix multiplication and inversion are very complex to operate, least squares with large-size matrices are very expensive to implement. Furthermore, large matrix operations are also time-consuming [16], which may cause problems in some applications such as a system that requires real-time or fast coefficients adaptation.

To avoid large matrices in order to reduce the computational complexity, it is desirable to use only a small number of data

This work was supported by the Science Foundation Ireland under the Principal Investigator Award scheme.

The authors are with the School of Electrical, Electronic and Communications Engineering, University College Dublin, Dublin 4, Ireland (e-mail: lei.guan@ucd.ie; anding.zhu@ucd.ie).

samples in DPD model extraction. Unfortunately, using short sequences leads to two consequences. First, the ill-conditioning problem becomes more severe in the LS estimation because reduced number of training samples can result in rank deficiency of LS matrices that consequently increases model extraction errors. Secondly, model extraction from a short training sequence also suffers from a statistical mismatch problem because the short training sequence often cannot fully represent the statistical property of the real signal that the PA transmits. This mismatch can significantly affect the accuracy of the DPD and thus degrade its linearization performance.

In order to deal with these challenges, in this work, we first introduce a 1-bit ridge regression approach to solve the ill-conditioning problem in the LS-based model extraction using short training sequences, and then propose a root mean square (RMS) based coefficients weighting and an averaging method to reduce the model extraction errors caused by unmatched statistics between the training data and the real transmit signal. Combining these two solutions, we can use very short training sequences to accurately extract the DPD coefficients which will maintain high linearization performance of the DPD while the computational complexity and implementation cost of the DPD are dramatically reduced.

The paper is organized as follows. After briefly reviewing the typical least squares-based model extraction of digital predistortion in Section II, the 1-bit ridge regression algorithm is proposed in Section III. The RMS-weighted coefficients averaging solution is introduced in Section IV and the detailed model extraction procedures and system complexity analysis are given in Section V. The experimental results are presented in Section VI, followed by a conclusion in Section VII.

II. DIGITAL PREDISTORTION REVIEW

The principle of digital predistortion is that a nonlinear distortion function is built up within the digital domain that is the inverse of the distortion function exhibited by the power amplifier. Most DPDs to date are implemented in the digital baseband domain. In these systems, the baseband I/Q (in-phase/quadrature) signal is first pre-processed by the DPD unit, then modulated and up-converted to the RF frequency, and finally sent to the PA. In order to extract and update the coefficients of the DPD, a small fraction of the transmit signal is fed back and transferred back to baseband. The model extraction unit compares the captured input and output data to extract the coefficients for the DPD. The block diagram of a digital predistortion system is illustrated in Fig. 1.

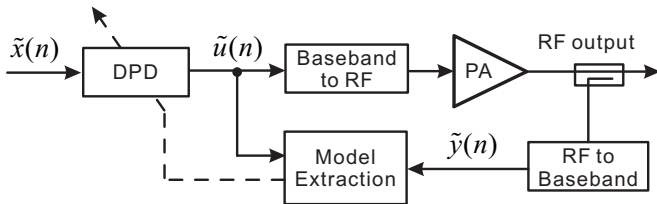


Fig. 1. Block diagram of a digital predistortion system.

In the past decade, extensive research has been carried out in digital predistortion and many behavioral models have been developed, such as, memory polynomial [3], generalized memory polynomial [4], and various simplified formats of Volterra series based models [8]-[13]. In this paper, we use the simplified 2nd-order *dynamic deviation reduction*-based Volterra series model proposed in [12], which can be expressed as follows:

$$\begin{aligned} \tilde{u}(n) = & \sum_{p=0}^{\frac{P-1}{2}} \sum_{m=0}^M \tilde{g}_{2p+1,1}(m) |\tilde{x}(n)|^{2p} \tilde{x}(n-m) \\ & + \sum_{p=1}^{\frac{P-1}{2}} \sum_{m=1}^M \tilde{g}_{2p+1,2}(m) |\tilde{x}(n)|^{2(p-1)} \tilde{x}^2(n) \tilde{x}^*(n-m) \\ & + \sum_{p=1}^{\frac{P-1}{2}} \sum_{m=1}^M \tilde{g}_{2p+1,3}(m) |\tilde{x}(n)|^{2(p-1)} \tilde{x}(n) |\tilde{x}(n-m)|^2 \\ & + \sum_{p=1}^{\frac{P-1}{2}} \sum_{m=1}^M \tilde{g}_{2p+1,4}(m) |\tilde{x}(n)|^{2(p-1)} \tilde{x}^*(n) \tilde{x}^2(n-m) \end{aligned} \quad (1)$$

where $\tilde{x}(n)$ and $\tilde{u}(n)$ are the input and the output complex I/Q envelopes, respectively. $\tilde{g}_{2p+1,j}$ ($j = 1, 2, 3, 4$) are the complex Volterra kernels of the DPD, $(\cdot)^*$ represents the complex conjugate operation and $|\cdot|$ returns the magnitude. P is the order of nonlinearity (P is an odd number) and M represents the memory length.

In a compact form, (1) can be rewritten in a matrix format as,

$$\mathbf{U}_{N \times 1} = \mathbf{X}_{N \times K} \mathbf{C}_{K \times 1} \quad (2)$$

where the matrix \mathbf{X} contains all of the linear and product terms $\tilde{x}(n)$, $\tilde{x}(n-1)$, ..., $|\tilde{x}(n)|^2 \tilde{x}(n)$, ..., appearing in the input of the model, for $n=1, 2, \dots, N$, and \mathbf{C} represents the parameter vector containing all of the unknown coefficients $\tilde{g}_{2p+1,j}(\cdot)$. The vector \mathbf{U} represents the DPD output vector. Subscripts indicate the size of the matrix. N represents the number of input/output data samples used and K is the number of coefficients involved.

To extract the coefficients, the p th-order post-inverse [13] or the in-direct learning [17] can be employed, where the feedback signal, i.e., the output of the PA, $\tilde{y}(n)$, is used as the input of the model, while the predistorted output signal, $\tilde{u}(n)$, is used as the expected output. By employing the standard least squares algorithm, the coefficients vector \mathbf{C} can be estimated from,

$$\hat{\mathbf{C}}_{K \times 1} = [(\mathbf{Y}^H)_{K \times N} \mathbf{Y}_{N \times K}]^{-1} (\mathbf{Y}^H)_{K \times N} \mathbf{U}_{N \times 1} \quad (3)$$

where \mathbf{Y} is the PA output matrix in a similar form to the matrix \mathbf{X} , and $(\cdot)^H$ represents Hermitian transpose.

In (3), there are three matrix-matrix multiplications and one matrix inversion. The computational complexity of each such matrix operation depends on the sizes of the matrices involved. For instance, multiplying a $K \times N$ matrix by a $N \times K$ matrix requires $K \times N \times K$ complex multiplications and $(N-1) \times K \times K$

complex additions. The inversion of a $K \times K$ matrix is approximately equivalent to K^3 number of multiplication operations. The total number of complex multiplication operations to be conducted in (3) is

$$O_{N,K}^{\otimes} = N \times K^2 \times 2 + N \times K + K^3 \quad (4)$$

and the total number of complex addition operations is

$$O_{N,K}^{\oplus} = (N-1) \times K^2 + N \times (K-1) \times K + (N-1) \times K \quad (5)$$

In digital signal processing, the multiplication operation is much more complex and resource consuming than that of addition. For example, in the Xilinx Virtex IV FPGA (field programmable gate array), a single 16-bit complex multiplier requires 1,168 4-input LUTs, 594 unit slices and 598 flip-flops, while an adder only needs 32 LUTs, 18 slices and 34 flip-flops. Therefore, $O_{N,K}^{\oplus}$ consumes much less hardware resources than $O_{N,K}^{\otimes}$, and the overall computational complexity of (3) thus mainly depends on $O_{N,K}^{\otimes}$, the total number of complex multiplications involved, namely,

$$O_{N,K} = O_{N,K}^{\otimes} + O_{N,K}^{\oplus} \approx O_{N,K}^{\otimes} \quad (6)$$

Furthermore, in LS estimation, in order to reduce the influence of errors in the observations, the number of data samples used is normally much greater than that of unknown parameters in the model. In other words, N is normally much greater than K . This leads to the conclusion that the computational complexity of model extraction actually mainly depends on the number of the training samples used.

III. 1-BIT RIDGE REGRESSION

Since the complex multiplier is one of the most complex and expensive components in digital circuits, e.g., FPGA hardware [18], in order to reduce the implementation cost in a real system, it is always desirable to use a smaller number of data samples, i.e., a shorter training sequence, to extract model parameters. However, a reduced number of training samples can introduce severe ill-conditioning in the matrix inversion of least squares, which consequently increases the errors of model extraction and thus degrades the DPD performance.

The numerical condition of a matrix is normally measured by its condition number χ ,

$$\chi = \frac{\lambda_{(\max)}}{\lambda_{(\min)}} \quad (7)$$

which is defined as the ratio of the largest eigenvalue $\lambda_{(\max)}$ to the smallest eigenvalue $\lambda_{(\min)}$ of the matrix. In LS estimation, the accuracy of the solution mainly depends on the numerical

condition of the Hessian matrix $\mathbf{Y}^H \mathbf{Y}$. Ideally, if the data samples are uncorrelated from each other, each row of the matrix \mathbf{Y} will be independent of the others. The condition number χ will then be small and the system is well conditioned, so that the best fit can be obtained. However, in a real system, the training samples are normally directly acquired from the transmit signal in the transmitter. It is very difficult to satisfy the statistically independent condition. This situation becomes worse when a smaller number of training sequences are used. The result is that the rank of $\mathbf{Y}^H \mathbf{Y}$ often becomes deficient, which causes $\lambda_{(\min)}$ to tend towards zero and χ becomes infinite. The poor numerical condition will result in large errors in matrix inversion and thus cause the estimated coefficients to deviate considerably from their true values.

Moreover, the ill-conditioning problem also often causes large variations among the extracted coefficient values, namely, some coefficients can be very small and the others may reach very large values. A consequence is that higher resolution digital circuits must be utilized to process and store the coefficients. This is because, in digital hardware, e.g., FPGAs, the DPD coefficients are normally stored in binary form. The number of binary bits required for representing the values of the coefficients is decided by the dynamic range of the coefficients, or the ratio of the maximum value to the smallest resolution of the system. For example, the number of binary bits, Ψ , is normally calculated from,

$$\Psi \geq \log_2 \left(\frac{\max(g)}{\Delta} \right) \quad \Psi : \text{integer} \quad (8)$$

where $\max(g)$ is the largest amplitude of the real part (or the imaginary part) of the coefficients, and Δ is the smallest resolution. We can see that, for a given resolution, larger variations in coefficients, a larger number of binary bits are required in the storage.

Although there are many algorithms being developed to resolve the ill-conditioning problem, such as QR factorization and singular value decomposition (SVD) [19], these algorithms often involve complicated signal processing, which is not feasible in a low-cost DPD system. One simple way to ease the ill-conditioning problem is to inject a diagonal matrix $\delta \mathbf{I}$ ($\delta > 0$, \mathbf{I} is the identity matrix) into the Hessian matrix $\mathbf{Y}^H \mathbf{Y}$ in (3), which is called *ridge regression* [20]. The model extraction equation (3) can be rewritten as,

$$\hat{\mathbf{C}}_{K \times 1, \text{new}} = \left[(\mathbf{Y}^H)_{K \times N} \mathbf{Y}_{N \times K} + \delta \mathbf{I}_{K \times K} \right]^{-1} (\mathbf{Y}^H)_{K \times N} \mathbf{U}_{N \times 1} \quad (9)$$

If $\lambda_{(\max)} \gg \delta \geq \lambda_{(\min)}$, the new condition number can be represented as follow,

$$\chi_{\text{new}} = \frac{\lambda_{\text{new}(\max)}}{\lambda_{\text{new}(\min)}} = \frac{\lambda_{(\max)} + \delta}{\lambda_{(\min)} + \delta} \approx \frac{\lambda_{(\max)}}{\delta} \quad (10)$$

Due to the injection of $\delta \mathbf{I}$, the smallest eigenvalue will no longer approach zero and thus χ_{new} will not be infinite anymore;

therefore the ill-conditioning problem can be eased.

In conventional ridge regression, it is often difficult to find an optimum value for δ to maintain high computational stability without losing accuracy. However, in digital hardware implementation, the baseband I/Q and DPD signals are normally represented in finite-bit binary format, e.g., as 16-bit binary numbers. This introduces some “quantization” errors in the signal processing. The smallest error equals the 1-bit binary resolution in the hardware process. For example, if the maximum amplitude of the signal is normalized to one, the 1-bit residue will be $1/2^{16} \approx 1.53e-5$ if 16-bit binary numbers are used. Because this residue error occurs in the system anyway, introducing this small value into the model extraction matrix does not significantly increase the system error. Therefore, in practice, instead of conducting a complicated optimization process to find the proper value for δ , in this work, we propose simply to assign the value of this 1-bit least residue error to δ in the ridge regression in (9). We call this approach as *1-bit ridge regression* (1-bit RR). One may argue that this 1-bit RR may not be the optimal choice, but considering the implementation complexity and system accuracy, this selection is actually indeed one of the best options. This is verified through the following tests.

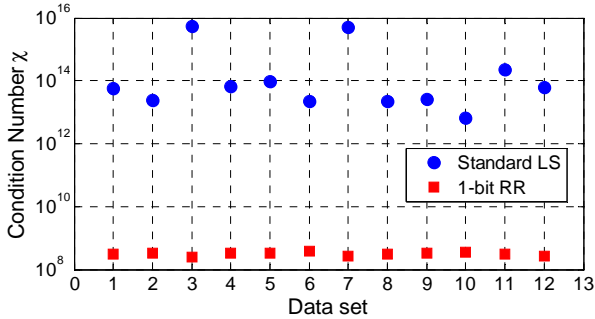


Fig. 2. Condition numbers with standard LS and with 1-bit RR.

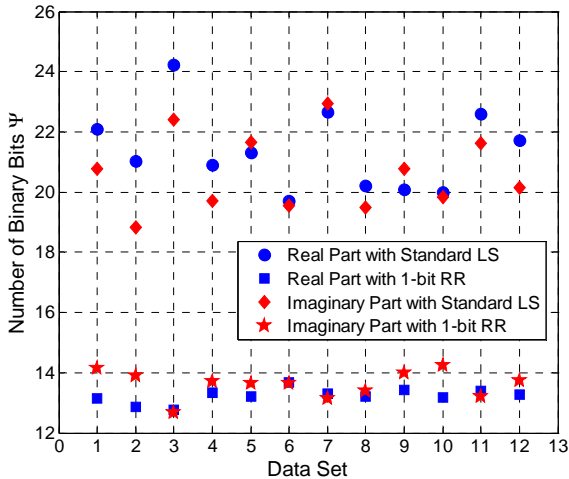


Fig. 3. Required number of binary bits with standard LS and with 1-bit RR, ($\delta=2.5 \times 10^{-4}$).

To evaluate the performance improvement by the proposed 1-bit RR based model extraction, 12 sets of baseband sampling data are captured from the input and output of a high power Doherty amplifier excited with an 8-carrier WCDMA signal.

Each data set includes 512 training samples. For comparison, standard least squares in (3) and 1-bit ridge regression in (9) are employed to extract the model coefficients respectively, from each set of data separately. The condition numbers calculated from these data are shown in Fig. 2, where we can see that, in the standard LS, the condition number is normally very large, at 10^{13} to 10^{16} level, while when the 1-bit RR is employed, it is dramatically reduced, to below 10^9 . The required numbers of binary bits for representing the coefficients are shown in Fig. 3. Without using 1-bit RR, 19 to 24 bits are required to represent the coefficients while with 1-bit RR, only 14 bits are needed. This indicates that the variations among the coefficients are dramatically “smoothed” when the condition number of the matrix is reduced.

To evaluate the model extraction accuracy, we captured another set of an independent data sequence, which includes 8,000 samples. We calculated the output from the input by multiplying with the coefficients extracted with 1-bit RR and standard LS, respectively, and compared with the actual output. The NMSE (normalized mean square error) [9] results are shown in Fig. 4, where we can find that large errors occurred with the standard LS while the modeling errors were dramatically reduced by employing the proposed approach. The conclusion is that, although a small bias is introduced in the model extraction equation, the model accuracy is not degraded but actually significantly improved because the ill-conditioning problem is eliminated.

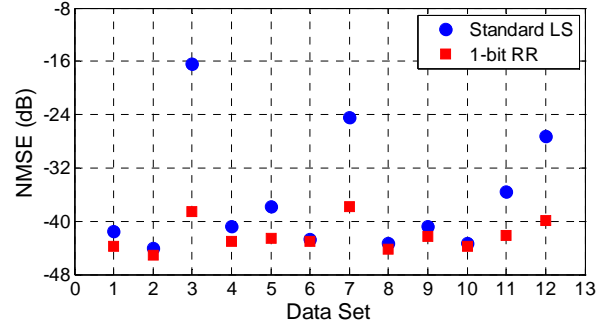


Fig. 4. Model extraction error evaluation in the time domain.

IV. RMS-WEIGHTED COEFFICIENTS AVERAGING

In a manner that is different from a linear system, the transfer function of a nonlinear system, such as a RF power amplifier, not only depends on its internal features but also strongly depends on the characteristics of its input signal. This requires that the statistical distribution of the selected training samples for the DPD model extraction must be very close to that of the real signal that the power amplifier transmits; otherwise the extracted parameters will not be accurate enough for describing the nonlinear behavior of the DPD. Due to limited storage resource in a real system, only a limited number of sampling data points can be captured. And more importantly, in order to reduce computational complexity, only a very small number of samples are used to form the LS matrix in the model extraction. It is very difficult to ensure that the statistical distribution requirements are satisfied. For instance, in a real system, the magnitude distribution of an 8-carrier WCDMA signal should be close to a Rayleigh distribution. This statistical property can

only be fully represented by using a very long sequence, e.g., 16,000 samples, as shown in Fig. 5 (a). If a shorter sequence is used, the statistical distribution will vary significantly. For example, the histograms of three randomly selected sets of 512-sample record data from the same WCDMA signal are shown in Fig. 5 (b) (c) (d), respectively, where we can clearly see that there are significant variations among these data and none of them is particularly close to that of the long sequence.

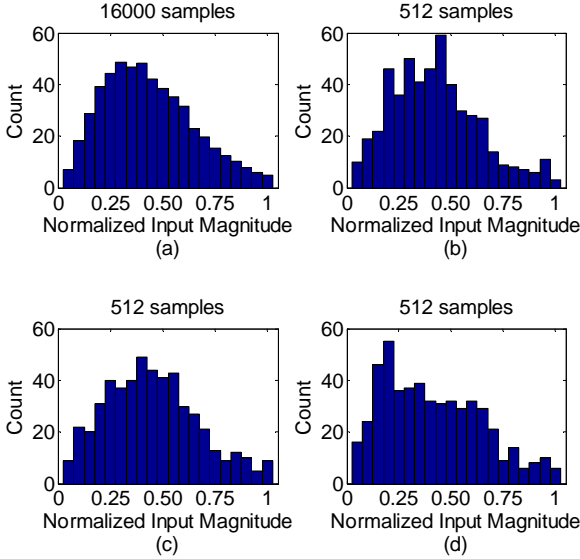


Fig. 5. Magnitude histograms of different data sets.

Furthermore, as discussed in [15], the peak, RMS and PAPR (peak to average power ratio) values of the input/output signals significantly affect DPD model extraction and the overall system performance. Fig. 6 gives peak, RMS and PAPR values for 12 sets of sampled input data captured from an 8-carrier WCDMA signal. We can see there are large variances among these values. These statistical variations mean that the coefficients extracted using these data will differ from each other and it is difficult to know which set of the coefficients is the best approximation that produces optimal performance.

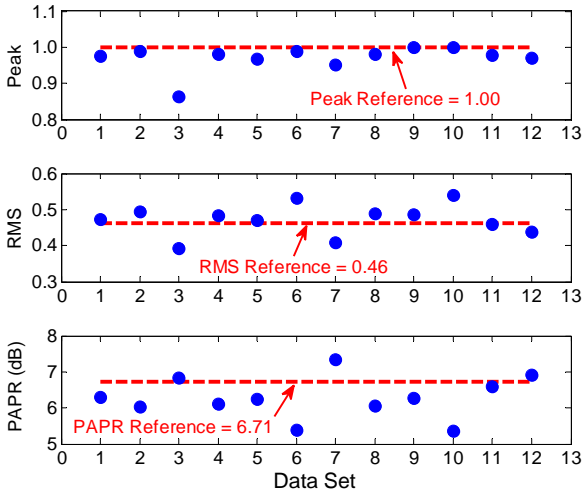


Fig. 6. Peak, RMS and PAPR values of different data sets.

To eliminate the statistical effects caused by using short sequences in model extraction, in this work, we propose an approach based on *RMS-weighted coefficients averaging* (RMS-WA). This method can be conducted in three steps. First, we record multiple short data sets, e.g., G sets of 512 samples from input and output of the PA. These short sequences can be obtained by being acquired at different time spans or selected from a recorded long sequence. We then extract the coefficients using each set of data respectively employing 1-bit ridge regression as described in the previous Section.

Because superposition is not applicable in a nonlinear system, nonlinear scaling must be conducted on the DPD coefficients, if the RMS values are different; otherwise they cannot be used for linearizing the PA operated at the different average power levels. In the second step, we scale the coefficients extracted from different sets of data to that at the same level of RMS value. The scaling weights are calculated as below,

$$\omega_{i,2p+1} = \left(\frac{RMS_0}{RMS_i} \right)^{2p} \quad (11)$$

where RMS_i represents the RMS value of the i th set of input data used for the model extraction, and RMS_0 is that of the actual input signal that the PA transmits, which can be estimated from the long sequence composed of all G sets of data. The i represents the i th set of coefficients ($i \geq 3$) and $2p+1$ corresponds to the $(2p+1)$ th order coefficients, e.g., $\tilde{g}_{2p+1,j}$ in (1). After scaling, the coefficients vector of each set becomes

$$\mathbf{C}_{i,weighted} = \begin{bmatrix} \tilde{g}_{i,1,1}(0) \\ \vdots \\ \omega_{i,3} \times \tilde{g}_{i,3,1}(0) \\ \omega_{i,3} \times \tilde{g}_{i,3,1}(1) \\ \vdots \\ \omega_{i,7} \times \tilde{g}_{i,7,2}(3) \\ \vdots \end{bmatrix} \quad (12)$$

Finally, to smooth the variations among the coefficients, the G sets of coefficients are summed and then averaged to obtain the final set of coefficients as below,

$$\bar{\mathbf{C}} = \frac{1}{G} \sum_{i=1}^G \mathbf{C}_{i,weighted} \quad (13)$$

where $\bar{\mathbf{C}}$ represents the optimized coefficients vector.

Similar to the approach in Section II, we used another set of independent data for cross validation. The error power spectral density (EPSD) [5] performance is shown in Fig. 7, where we see that the model errors are very small, e.g., below -50 dBc and -55 dBc in-band and out-of-band, respectively, when the coefficients are optimized. The time domain NMSE is -45.6 dB. However, if we do not process the coefficients, namely, directly use the coefficients extracted from short sequences,

large errors often occur. For example, Fig. 7 also shows the EPSP results obtained by using randomly selected coefficients, e.g., the coefficients extracted from the third and fifth sets of 512 samples. As expected, due to statistical mismatch, larger errors occur in both cases, e.g., 5 to 10 dB worse in EPSP. And the NMSE only reaches -38.5 dB and -42.1 dB, respectively.

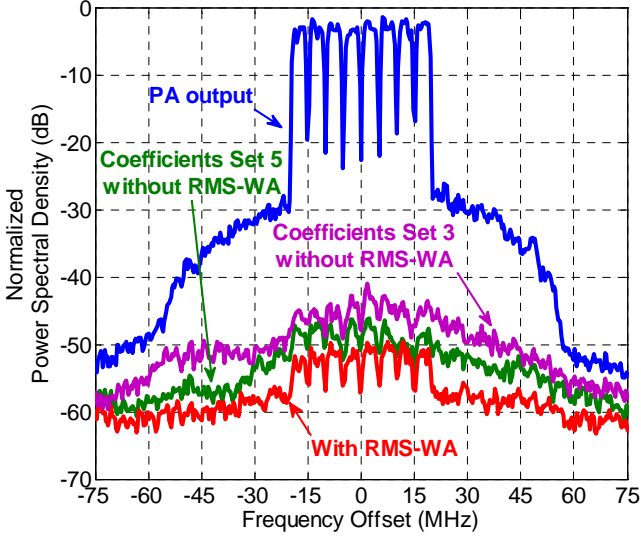


Fig. 7. EPSP performance with and without coefficients optimization.

V. THE PROPOSED LOW COMPLEXITY MODEL EXTRACTION

After introducing two solutions, 1-bit ridge regression and RMS-weighted coefficients averaging, to tackle the problems of ill-conditioning and statistical mismatch that arise from using short training sequences in DPD model extraction, in Section III and IV respectively, we now combine them together to outline the operational procedures for conducting the proposed low complexity model extraction, and give a detailed implementation complexity reduction analysis.

A. Model Extraction Procedures

The block diagram of the model extraction is illustrated in Fig. 8, and the detailed procedures are described below:

i) *Data Acquisition*: Multiple short training sequences are acquired from the input and output of the RF PA. These short data sequences can be obtained by acquiring at different time spans or selected from a recorded long sequence.

ii) *Time Alignment and Data Normalization*: All data are properly time-aligned between the input and the output. The magnitudes of the data are normalized by the maximum magnitude of the long sequence which is composed of all the short sequences. RMS values are then calculated for the long sequence and for each short sequence, respectively. In other words, RMS_0 and RMS_i are obtained.

iii) *1-bit Ridge Regression*: The ridge regression factor δ is calculated from $1/2^\Psi$, where Ψ is the number of binary bits used for representing the I/Q signals. After selecting proper DPD parameters, e.g., nonlinear order P and memory length M , the DPD coefficients C_i are then extracted using (9) for each set of short sequences. These 1-bit RR operations can be conducted in

a time multiplexed manner, i.e., they are operated with the same hardware block during different time periods.

iv) *RMS-weighted Coefficients Averaging*: Each set of coefficients is weighted according to (12). And all sets of the coefficients are then summed and averaged in (13) to obtain the optimized set of coefficients \bar{C} .

v) The optimum coefficients \bar{C} are finally sent to the DPD unit for predistorting the input signal.

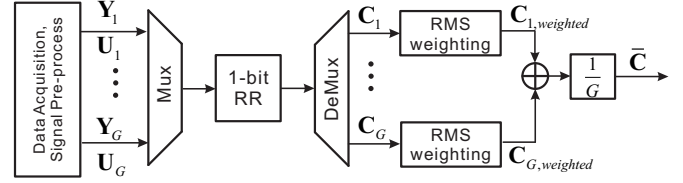


Fig. 8. The proposed model extraction.

B. Implementation Complexity Analysis

Employing 1-bit ridge regression as in (9), we can use a much shorter sequence, e.g., with a data length of L , to extract the DPD coefficients. Although K extra complex addition operations are required in (9) compared to (3), the overall computational complexity can still be approximately estimated by using the number of complex multiplications operations $O_{L,K}^\otimes$ involved as in (4) with N replaced by L . Let's define R as the ratio of N to L ,

$$R = \frac{N}{L} \quad (14)$$

the complexity ratio between two approaches can then be estimated using

$$\frac{O_{L,K}}{O_{N,K}} \approx \frac{O_{L,K}^\otimes}{O_{N,K}^\otimes} = \left(\frac{1}{R} + \frac{R-1}{R} \times \frac{K^2}{2NK + K^2 + N} \right) \times 100\% \quad (15)$$

Fig. 9 illustrates the total complexity ratio verse R for different numbers of coefficients, K . Because $N \gg K$, the second term of (15) approaches zero, the above equation can be approximately rewritten as

$$\frac{O_{L,K}}{O_{N,K}} \approx \frac{1}{R} \times 100\% \quad (16)$$

We thus conclude that the computational complexity of the proposed approach is reduced to roughly $1/R$ of that of the standard approach.

Converting large matrices to multiple smaller ones in the signal processing not only reduces computational complexity but also saves execution time. For example, Table I shows the execution time of the matrix operations in MATLAB R2008b using 3.0 GHz Intel (R) Core (TM) 2Duo CPU with 4GB RAM for $N = 6144$ and $L = 512$ with different K values. We can see that the total execution time for twelve 512-sized operations is

only a fraction of that involved in executing one 6144-sized operation. This indicates that, although multiple model extractions are required in the proposed approach, the total execution time for model extraction is actually much shorter than that using the conventional approach employing a single long sequence. This is crucial in some time-critical systems, such as wireless LAN data transmissions, where the data traffic can change dynamically, which requires the DPD to update its coefficients in a very short time period.

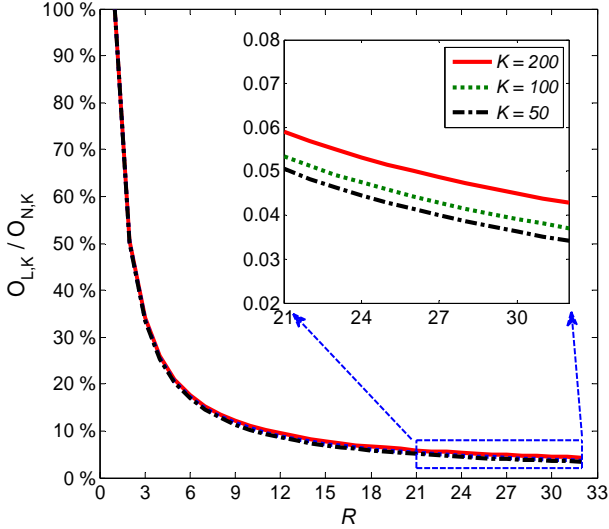


Fig. 9. The computational complexity ratio.

	Execution time (second)		
	512 samples	512 samples × 12	6144 samples
K = 30	0.0313	0.3756	7.0219
K = 48	0.0436	0.5323	13.4324
K = 73	0.0733	0.8796	21.6894
K = 96	0.0981	1.1772	27.3907
K = 107	0.1273	1.5276	32.4513

VI. EXPERIMENTAL RESULTS

In order to validate the performance of the proposed model extraction methodology for digital predistortion of RF power amplifiers, we tested a LDMOS Doherty amplifier operated at 2.14 GHz and excited with a 20 MHz signal carrier LTE (long term evolution) signal and a 40 MHz 8-carrier WCDMA signal. The average output power was 40 dBm. The test bench was set up as shown in Fig. 10. A baseband I/Q complex signal source was created in MATLAB in a PC, and fed to an FPGA baseband board for signal processing, e.g., digital predistortion. The output of the FPGA was then sent to an RF board, where the signal was modulated and up-converted to the RF frequency, and finally sent to the PA. In the output, the RF signal was down-converted and demodulated to baseband again for model extraction. The baseband I/Q data sampling rate was 368.64 M samples/second, and the linearization bandwidth in

the RF chain was 180 MHz. Around 16,000 data samples were recorded each time. The DPD coefficients were extracted from the captured data records in MATLAB and then fed to the DPD unit implemented in FPGA. The system performance was then verified independently using separate test signals.

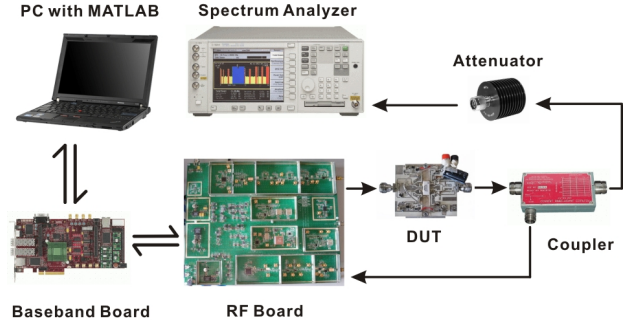


Fig. 10. The DPD experimental test bench.

The modeling accuracy is affected by various factors, such as statistical characteristics of training samples, selection of nonlinear order and memory length in the model, and so on. The main motivation of this work is to propose an alternative model extraction method to replace the standard least squares in order to reduce the computational complexity and implementation cost. For performance evaluation, we shall compare the system complexity between the proposed method and the standard LS based on the same linearization performance. For this reason, in this work, we first conduct digital predistortion for the PA using the standard approach to find out what is the best linearization performance the DPD can achieve. We then use this performance as the target to find out how many data samples are required and how many sets of coefficients need to be averaged to obtain the optimum coefficients in the proposed approach. Finally, we can compare the system complexity.

A. 20MHz single carrier LTE signal

The first test was conducted with a 20 MHz single carrier LTE signal. After various tests, we found that a DPD model with $P=11$ and $M=2$ (48 coefficients in total) produced the best performance. Also, the best model extraction accuracy was achieved by using 8,192 training samples in the standard LS-based coefficients estimation. The AM/AM and AM/PM characteristics are shown in Fig. 11, where we can see that both static nonlinearities and memory effects are almost completely removed after DPD. The frequency spectrum of the PA output after DPD is shown in Fig. 12, marked with “Long Sequence LS”. In addition, the NRMSE (normalized root mean square error) [13] and ACPR (adjacent channel power ratio) values are listed in Table II. After employing DPD, the NRMSE is reduced from 13.17% to 0.61% and the first and the second adjacent ACPR are reduced from -28 dBc and -48 dBc to -61 dBc and -65 dBc, respectively.

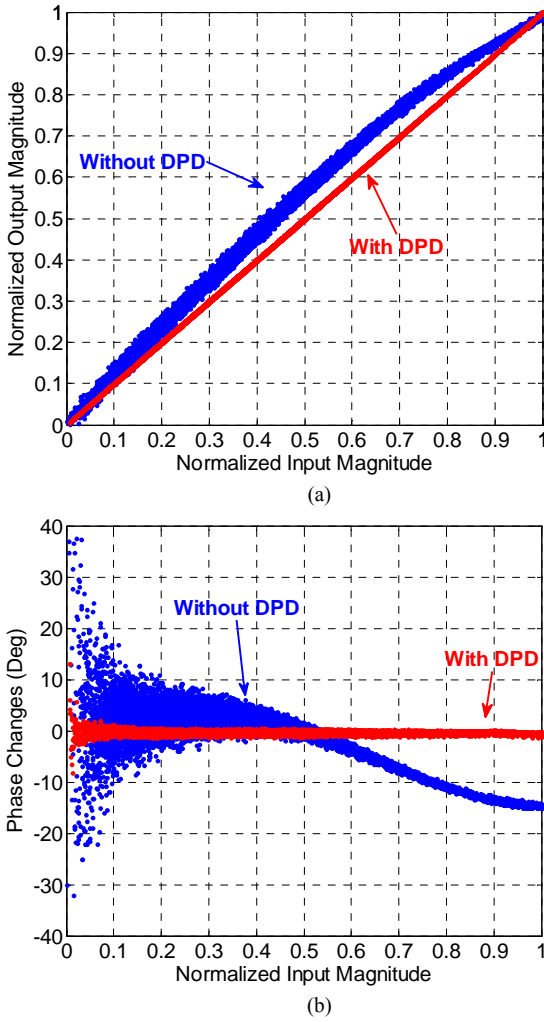


Fig. 11. AM/AM and AM/PM plots with and without DPD for a 20MHz LTE signal.

In the proposed new approach, we only used 1,024 samples in the 1-bit ridge regression, and conducted the model extraction 6 times. The extracted 6 sets of coefficients were then weighted and averaged to obtain the optimum set of coefficients. In order to verify the individual improvements made by the two solutions in the proposed approach, we conducted the validation test in three steps: (i) directly use the standard LS with the short sequence; (ii) employ the 1-bit ridge regression instead of the standard LS; (iii) employ both the 1-bit ridge regression and the RMS-WA. The results are shown in Fig. 12 and Table II. Due to the ill-conditioning problem, the DPD extracted from a short sequence in the standard LS only achieves 11 dB improvements in the first adjacent channel and almost no improvement can be made in the second adjacent channel. Also, the NRMSE only reaches 2.35%. However, when the 1-bit ridge regression is applied, namely, a small diagonal matrix $\delta \mathbf{I}$ is inserted into the model extraction equation, the system performance can be significantly improved. The ACPRs are reduced by 16 dB in both the first and the second adjacent channels and the NRMSE reaches 0.70%. Finally, with the coefficients weighting and averaging,

the ACPRs can be further improved by 4 to 5 dB and the NRMSE reaches 0.65%, which is almost the same performance as the best achieved by the standard LS using a long sequence.

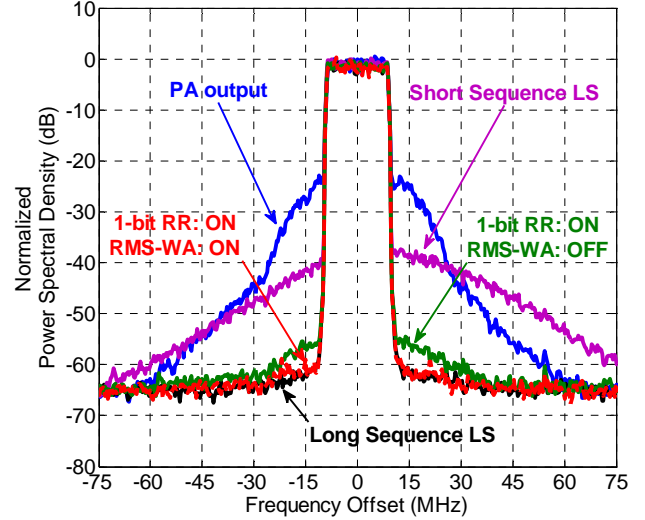


Fig. 12. PA output spectra with and without DPD for a 20MHz LTE signal.

TABLE II
LINEARIZATION PERFORMANCE FOR 20MHz LTE SIGNAL

	NRMSE (%)	ACPR (dBc)		Number of Training Samples
		+/- 20MHz	+/- 40MHz	
Without DPD	13.17	-28.3/-27.6	-48.1/-49.2	
With DPD Standard LS	2.35	-39.4/-42.5	-46.3/-50.4	1024
With DPD 1-bit RR: ON RMS-WA: OFF	0.70	-55.9/-56.8	-62.1/-61.8	1024
With DPD 1-bit RR: ON RMS-WA: ON	0.65	-60.5/60.4	-64.9/-65.0	1024
With DPD Standard LS	0.61	-61.4/-61.3	-65.3/-65.9	8192

TABLE III
NUMBER OF COMPLEX MULTIPLICATIONS INVOLVED

Operation	Number of Complex Multiplication		Ratio
	Standard LS ($N = 8192$)	Proposed Approach ($L = 1024$)	
$S_1 = \mathbf{Y}^H \mathbf{Y}$	$8192 \times 48^1 \times 48$	$1024 \times 48 \times 48$	8:1
$S_2 = (S_1)^{-1}$	$48 \times 48 \times 48$	$48 \times 48 \times 48$	1:1
$S_3 = S_2 \mathbf{Y}^H$	$48 \times 8192 \times 48$	$48 \times 1024 \times 48$	8:1
$S_4 = S_3 \mathbf{U}$	8192×48	1024×48	8:1
Total	38,252,544	4,878,336	7.8:1

The number of parameters $K = 48$.

Although the same linearization performance is achieved, the computational complexity is dramatically reduced in the proposed approach, because only 1,024 samples, one eighth of that in the standard approach, are used. The numbers of complex multiplications involved are given in Table III. For the same number of coefficients, i.e., 48, the standard approach

must conduct much large-sized matrix-matrix multiplication, such as 48×8192 multiplied by 8192×48 , while in the proposed approach, these sizes are dramatically reduced, e.g., to 48×1024 by 1024×48 . The total number of complex multiplication operations is reduced to less than 15%, i.e., $1/7.8$, of that in the standard approach.

B. 40MHz multi-carrier WCDMA signal

The second test used a 40 MHz 8-carrier WCDMA signal as the excitation to the PA. In this case, nonlinear order $P=9$ and memory length $M=4$, the total number of coefficients was 73. Again, we compare the complexity under the same linearization performance. In this test, we found that using 6,144 samples in the standard LS can produce excellent linearization results. For example, the AM/AM and AM/PM characteristics are shown in Fig. 13, which indicates that the distortion induced by the nonlinearity of the PA was almost completely removed after employing the DPD. The frequency spectra are given in Fig. 14 and the NRMSE and ACPR performance are presented in Table IV. One may notice that the memory effects in this case are much stronger than that in the LTE case because of the wider bandwidth of the signal.

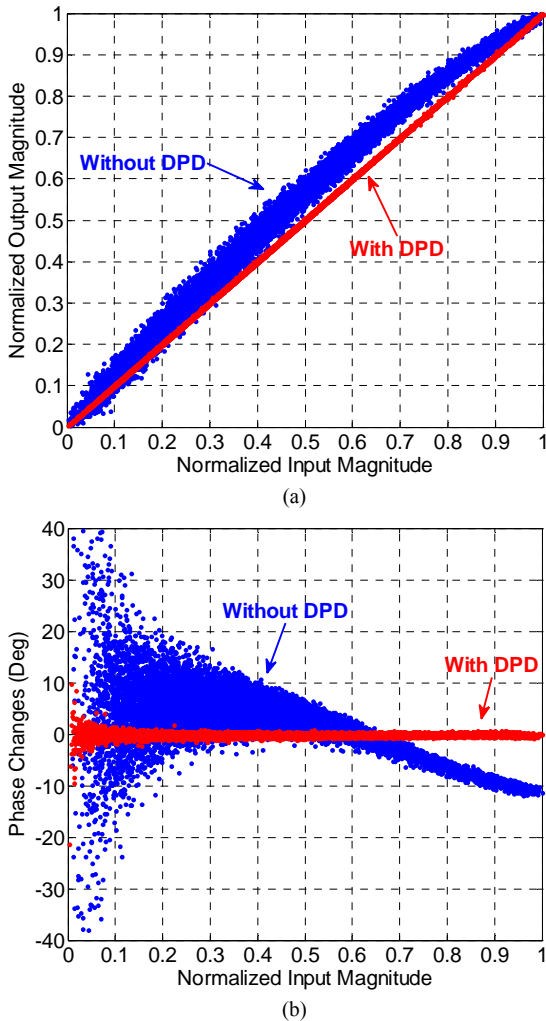


Fig. 13. AM/AM and AM/PM plots with and without DPD for an 8-carrier WCDMA signal.

To achieve the same performance, in the proposed approach,

we only used 8 sets of 512 samples to extract the optimum set of coefficients for the DPD. The output spectrum plot is shown in Fig. 13 and NRMSE and ACPR figures are given in Table IV. It shows that, without using 1-bit RR, the ACPR only can be reduced to -38 dBc while with 1-bit RR, 11 dB more can be achieved, and with the coefficients weighting and averaging, a further 7 dB reduction can be obtained. The NRMSE is reduced from 3.47% to 1.02% and finally reaches 0.64%. Similarly, for the same linearization performance, the total number of complex multiplication operations in this test is reduced to less than 9%, i.e., $1/12$ of that in the conventional approach. This further indicates that by employing the proposed approach, we could use multiple much shorter sequences, instead of a long sequence, as the training data, to accurately extract the DPD coefficients, which leads to a significant reduction in system complexity and implementation cost.

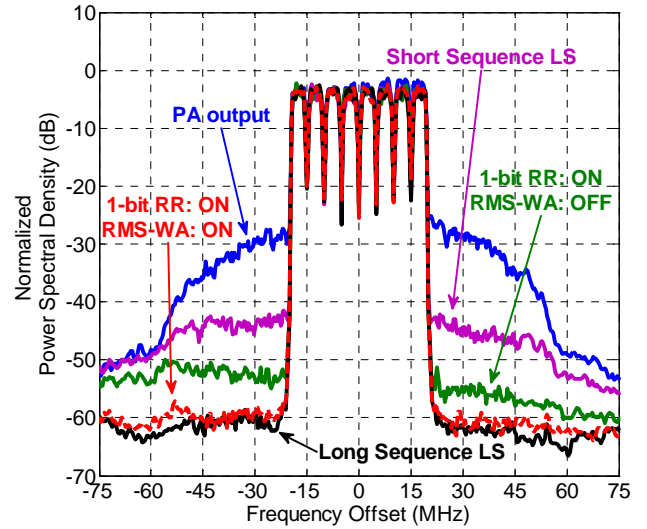


Fig. 14. PA output spectra with and without DPD for an 8-carrier WCDMA signal.

TABLE IV
LINEARIZATION PERFORMANCE FOR 40MHz WCDMA SIGNAL

	NRMSE (%)	ACPR (dBc)		Number of Training Samples
		+/- 5MHz	+/- 10MHz	
Without DPD	11.13	-24.3/-23.2	-24.8/-24.5	
With DPD Standard LS	3.47	-38.5/-39.4	-40.8/-41.3	512
With DPD 1-bit RR: ON RMS-WA: OFF	1.02	-49.8/-52.4	-50.1/-53.4	512
With DPD 1-bit RR: ON RMS-WA: ON	0.64	-56.0/56.3	-56.3/-57.1	512
With DPD Standard LS	0.48	-57.1/-56.2	-58.0/-58.1	6144

VII. CONCLUSION

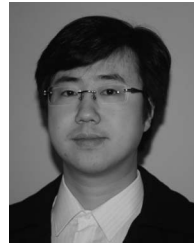
In this paper, we have proposed a 1-bit ridge regression algorithm to eliminate the ill-conditioning problem in least

squares-based DPD coefficients estimation and also an RMS-based coefficients weighting and averaging technique to reduce model extraction errors caused by statistical mismatch between the training data and the real signal that the power amplifier transmits. Experimental results show that, by employing these two algorithms, we are able to use a small number of training samples to accurately extract the DPD coefficients to achieve excellent linearization performance for RF PAs. Compared to the conventional approach, the computational complexity and implementation cost of the model extraction employing these new approaches is dramatically reduced because the sizes of the matrices involved in the parameter estimation are significantly reduced. The execution time of the algorithm is also much shorter than that using the standard least squares employing a long sequence. This is crucial in some real-time systems.

We demonstrated how the proposed approach can be applied to a simplified Volterra DPD model and presented the linearization performance for a high power Doherty amplifier in this paper. The proposed technique is not solely limited to that particular model or PA. The same solution can also be employed for any other DPD models (whose output is linear in relation to their coefficients) to linearize other type of PAs. Since only standard matrix operations are required, the proposed approach can be easily implemented in standard digital circuits, such as FPGAs or standard DSP chips.

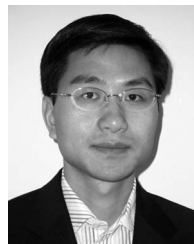
REFERENCES

- [1] P. B. Kennington, *High Linearity RF Amplifier Design*, Norwood, MA: Artech House, 2000.
- [2] L. Guan, and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 4, pp. 866 - 872, Apr. 2010.
- [3] 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
- [4] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp.3852-3860, Oct. 2006.
- [5] S. Hong, Y. Y. Woo, J. Kim, J. Cha, I. Kim, J. Moon, J. Yi, and B. Kim, "Weighted polynomial digital predistortion for low memory effect Doherty power amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp. 925-931, May 2007.
- [6] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Augmented Hammerstein predistorter for linearization of broadband wireless transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1340-1349, Jun.2006.
- [7] J. Kim, Y. Y. Woo, J. Moon and B. Kim, "A new wideband adaptive digital predistortion technique employing feedback linearization," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 2, pp. 385-392, Feb. 2008.
- [8] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223-227, Jan. 1997.
- [9] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4323 - 4332, Dec.2006.
- [10] N. Safari, T. Roste, P. Fedorenko, and J. S. Kenny, "An approximation of Volterra series using delay envelopes, applied to digital predistortion of RF power amplifiers with memory effects," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 2, pp. 115-117, Feb. 2008.
- [11] D. Mirri, G. Iuculano, F. Filicori, G. Pasini, G. Vannini, and G. P.Gualtieri, "A modified Volterra series approach for nonlinear dynamic systems modeling," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 8, pp. 1118-1128, Aug. 2002.
- [12] L. Guan, and A. Zhu, "Simplified Dynamic Deviation Reduction-based Volterra Model for Doherty Power Amplifiers," *IEEE Int. Workshop on Integrated Nonlinear Microwave and Millimeterwave Circuits, INMMIC*, Vienna, Austria, April, 2011.
- [13] A. Zhu, P. J. Draxler, J. J Yan, T. J. Brazil, D. F. Kinball, and P. M. Asbeck, "Open-loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 7, pp. 1524-1534, Jul. 2008.
- [14] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159 - 165, Jan. 2004.
- [15] O. Hammi, S. Carichner, B. Vassilakis and F. M. Ghannouchi, "Synergetic crest factor reduction and baseband digital predistortion for adaptive 3G Doherty power amplifier linearizer Design," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 11, pp. 2602-2608, Nov. 2008.
- [16] L. Zhuo and V. K. Prasanna, "High-performance designs for linear algebra operations on reconfigurable hardware," *IEEE Trans. Computers*, vol. 57, no. 8, pp. 1057-1071, Aug. 2008.
- [17] L. Ding, R. Raich, and G. T. Zhou, "A Hammerstein predistortion linearization design based on the indirect learning architecture," in *Int. Acoust., Speech, Signal Process. Conf.*, May 2002, vol. 3, pp.2689-2692.
- [18] <http://www.xilinx.com>.
- [19] O. Nelles, *Nonlinear System Identification*. Berlin: Springer-Verlag, 2001.
- [20] A. Hoerl and R. Kennard, "Ridge regression biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 3, pp. 55-67, Aug. 1970.



Lei Guan (S'09) received the B.E. degree and the M.E. degree, both in Electronic Engineering from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively. He is currently studying towards the Ph.D. degree in the RF & Microwave Research Group at University College Dublin, Ireland.

His research interests include linearization and system-level modeling of RF/Microwave power amplifiers with emphasis on digital predistortion and FPGA hardware implementation based on Volterra Series. He also has interests in wireless communication system design and digital signal processing.



Anding Zhu (S'00-M'04) received the B.E. degree in telecommunication engineering from North China Electric Power University, Baoding, China, in 1997, and the M.E. degree in computer applications from Beijing University of Posts and Telecommunications, Beijing, China, in 2000, and the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2004.

He is currently a Lecturer with the School of Electrical, Electronic and Communications Engineering, UCD. His research interests include high-frequency nonlinear system modeling and device characterization techniques with a particular emphasis on Volterra-series-based behavioral modeling and linearization for RF PAs. He is also interested in wireless and RF system design, digital signal processing and nonlinear system identification algorithms.