Selection of Compressed Training Data for RF Power Amplifier Behavioral Modeling

Ziming Wang, John Dooley, Keith Finnerty, Ronan Farrell

Department of Electronic Engineering,
National University of Ireland Maynooth

Abstract—In this paper, we present an algorithm which uses the probability information of the input signal to inform the selection of a compressed training dataset for RF PA behavioural model extraction. The proposed algorithm can dramatically reduce the number of training samples. The accuracy of this algorithm is validated by extraction of behavioural models using a large dataset of consecutive samples and a reduced training dataset determined using the proposed algorithm. A noticeable reduction in computational complexity and faster execution time is achieved with the new approach.

Index Terms—Digital Predistortion, Behavioural Modelling, Least Square, Probability Distribution

I. INTRODUCTION

Characteristic behavior of nonlinear power amplifiers can be efficiently and accurately represented using behavioral models. The two main procedures which are carried out using behavioral models are the extraction of model parameters and the calculation of model outputs. It is generally accepted that a model with good generalization capabilities can be extracted for common mobile telecommunication standards using the discrete complex envelope representations of the input and output signals. Calculating the model parameters for discrete signals in some digital signal processing hardware often requires that a very large number of consecutive samples are recorded to improve the generalisation capabilities. Further to this, with the proliferation of high efficiency PA architectures such as Doherty and outphasing PAs for multi-band, multistandard, even more samples are needed to extract a model which will encompass all characteristics of the PA.

One strategy commonly employed to reduce the computational complexity of large matrix computations required to train a behavioral model, is to reduce the number of model parameters. This is an effective means of reducing both the extraction time and the calculation of model outputs. However it does not guarantee the generalisation properties of the model. It is still required to have a good training signal which encompasses the characteristics of the PA to train the reduced dimension model.

In the case of RF power amplifiers the typical selection process for behavioral model training data involves the selection of consecutive data samples which include key samples such as the maximum magnitude and maximum rate of change. Since modern communication standard signals are not truly periodic, it is not possible to simply select a short section of consecutive data to represent the whole dataset. Guan *et al* proposed an algorithm to use multiple short training data to build multiple small matrices for behavioural modeling[1]. However, this method does not reduce the total number of training samples and cannot guarantee the robustness for a nonlinear system with memory. In [2] we see that a subset of samples to train a digital pre-distorter for a given power amplifier can be selected based on the probability distribution function of the signals used.

In this paper, the information carried by training samples in relation to the different groups of model parameters is explored. We identify the limitations when consecutive training samples are applied to a Least Squares algorithm to extract the model parameter values. In Sec. II, we introduce the theoretical relationship between the model parameters and the training data. The traditional training sample selection process is outlined in Sec. III-A and a new training sample selection method is introduced in Sec. III-B. The experimentally measured results to validate the method are given in Sec. IV.

II. INFORMATION OF TRAINING SAMPLES

We use a general memory polynomial model $\phi(\cdot)$ to represent all polynomial-based models. This general model indicates the relationship between the model terms.

$$
\phi\big(
\begin{aligned}
&x(n), x(n)^2, ..., x(n)^{k_0} \\
&x(n-1), x(n-1)^2, ..., x(n-1)^{k_1} \\
&\vdots \\
&x(n-m), x(n-m)^2, ..., x(n-m)^{k_m}
\end{aligned}
\big) \tag{1}
$$

, where $x(n)$ is the input signal, m is the memory depth and $k_0 \cdots k_m$ is the highest order in different memory terms.

It has been well documented that coefficient estimation for such models can be performed by solving an over-determined system of equations using Least Square(LS) techniques.

$$
\phi^{\mathrm{T}}\phi \cdot \hat{\omega} = \phi^{\mathrm{T}}y + e \tag{2}
$$

where ω are the coefficients of the behavioural model and y is the output signal of nonlinear PA and e is the difference between an observed value and the fitted value of the model.

In (2), $\phi^{\mathrm{T}}\phi$ and $\phi^{\mathrm{T}}y$ are the auto-covariance and cross-covariance matrices. Since the modulated signal can be treated as standardized random variables, the covariance matrix is equal to the correlation matrix. As a result, the coefficient

calculation can be performed using (3).

$$\hat{\omega} = R_{\phi\phi}^{-1} \cdot R_{\phi y} \tag{3}$$

where $R_{\phi\phi}$ is auto-correlation matrix and $R_{\phi y}$ is cross-correlation matrix.

By analysing the general memory polynomial model fro system with and without memory, we can establish the differences in the coefficient and input sample combinations which ultimately lead to spreading effect in AM/AM plot for a system with memory. By better understanding the probability of occurrence of the samples within this range of possible values, a more compact set of training samples can be chosen to train an accurate model.

A. Memoryless System

Firstly we consider a memoryless nonlinear system, then the general polynomial ϕ becomes

$$\phi\begin{pmatrix} x(n), \; x(n)^2, \; ..., \; x(n)^k \end{pmatrix}$$

According to (3), we know the coefficients are related to the auto-correlation and cross-correlation matrix. So we have

$$R_{\phi\phi} = E[\phi\bar{\phi}] = \int \phi \cdot \rho(\phi) \, d\phi \tag{4}$$

$$R_{\phi y} = E[\phi(\cdot)\bar{y}_n] = \iint \phi y_n \cdot \rho(\phi, y_n) \, d\phi dy \tag{5}$$

$$\rho(\phi, y_n) = \rho(\; x_n, \; x_n^2, \; ..., \; x_n^k, y_n \;) \tag{6}$$

, where $\rho(\phi)$ and $\rho(\phi, y_n)$ are the probability distribution function(PDF) and joint probability distribution respectively.

Extending (6) to the chain rule of probability, we obtain:

$$\begin{aligned} &\rho(x_n, \; x_n^2, \; ..., \; x_n^k, y_n) \\ &= \rho(x_n) \\ &\quad \cdot \rho(y_n|x_n) \cdot \rho(x_n^2|x_n, y_n) \cdot \rho(x_n^3|x_n, x_n^2, y_n) \\ &\quad \cdots \rho(x_n^k|x_n, x_n^2, \cdots, x_n^{k-1}, y_n) \end{aligned} \tag{7}$$

From this analysis, it can be seen that for the memoryless system, the output sample y_n are only dependent on the input sample x_n.

$$\rho(y_n|x_n) = 1 \tag{8}$$

Furthermore, it is easy to prove that $x_n^2 \cdots x_n^k$ are fixed for a given x_n.

$$\rho(x_n^2|x_n, y_n) = \cdots = \rho(x_n^k|x_n, x_n^2, \cdots, x_n^{k-1}, y_n) = 1 \tag{9}$$

Hence for a memoryless system, we can rewrite (7) as

$$\begin{aligned} &\rho(x_n, \; x_n^2, \; ..., \; x_n^k, y_n) \\ &= \rho(x_n) \end{aligned} \tag{10}$$

Considering (3), (4), (5) and (10), it can be proven that, in a memoryless system, the coefficients estimated using least squares are directly related to the marginal probability of input signal $\rho(x_n)$.

B. System with Memory

Let us consider the joint probability of memory polynomial and output signal directly,

$$\begin{aligned} &\rho(x_n, x_n^2, ..., x_n^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n) \\ &= \rho(x_n) \cdot \rho(x_n^2|x_n) \cdots \rho(x_n^k|x_n, x_n^2, \cdots, x_n^{k-1}) \\ &\quad \cdot \rho(x_{n-1}|x_n, \cdots) \cdots \rho(x_{n-1}^{k1}|x_{n-1}, \cdots) \cdots \\ &\quad \cdot \rho(x_{n-m}|x_n, x_{n-1}, ..., x_{n-m+1}, \cdots) \cdots \rho(x_{n-m}^{k_m}|x_{n-m}, ...) \\ &\quad \cdot \rho(y_n|x_n, x_{n-1}, x_{n-2}, ..., x_{n-m}, ...) \end{aligned} \tag{11}$$

Because (9) still holds for a system with memory, (11) can be rewritten as follows:

$$\begin{aligned} &\rho(x_n, x_n^2, ..., x_n^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n) \\ &= \rho(x_n) \end{aligned} \tag{12a}$$

$$\begin{aligned} &\cdot \rho(x_{n-1}|x_n) \cdot \rho(x_{n-2}|x_n, x_{n-1}) \\ &\quad ...\rho(x_{n-m}|x_n, x_{n-1}, ..., x_{n-m+1}) \end{aligned} \tag{12b}$$

$$\cdot \rho(y_n|x_n, x_{n-1}, x_{n-2}, ..., x_{n-m}) \tag{12c}$$

One should notice that (8) is not possible in the case of a system with memory since the output at time t is not only determined by the input signal at time t but at time $t-1$ to $t-m$. Using this insight we rearrange the order of the chain rule of probability. So the term (12a) indicates the nonlinearity as same as the (10) in the memoryless system. Then terms (12b) depend on the bandwidth of input signal. and terms (12c) can be treated as the degree of the memory-effect in a nonlinear system with memory. Furthermore, in the view of behaviour modeling, (12c) can also represent the generalisation error which is related to the accuracy of behaviour models.

III. Training Data Compression

A. Traditional Training Samples Selection

In most cases, a modulated signal can be approximated to a wide-sense cyclostationary process[3]. Thus, the distribution for any given set of samples will be time-invariant over a time shift nT as described.

$$[X_1, ..., X_T] \sim [X_{1+nT}, ... X_{T+nT}] \qquad \forall n \tag{13}$$

In the case of a memoryless system with modulated signal input, providing there are sufficient consecutive samples in a reduced dataset to satisfy (13), the PDF of this reduced consecutive training dataset will be similar to that of the full dataset. As a result, selecting a consecutive modulated training dataset is able to extract the accurate model in a memoryless system. Fig. 1 shows the distribution of a 4-carrier LTE signal with different number of consecutive samples. The fundamental shape of all PDF are consistent even with different numbers of samples.

However in terms of nonlinearity, the efficiency of using a selected set of consecutive training samples in modelling highly depends on how close the statistical properties of the given signal are to a wide-sense stationary process. It leads to an increase in training samples required for multi-standard signal. On the other hand, in terms of a system with memory,
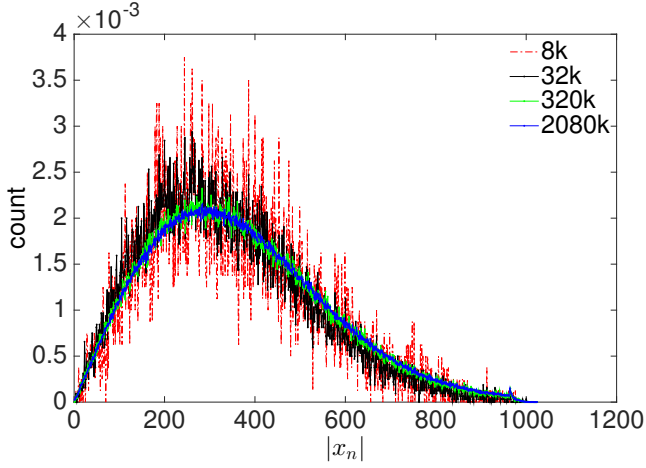
Fig. 1. Normolized Probability Distribution of LTE Signal with Different Number of Consecutive Samples
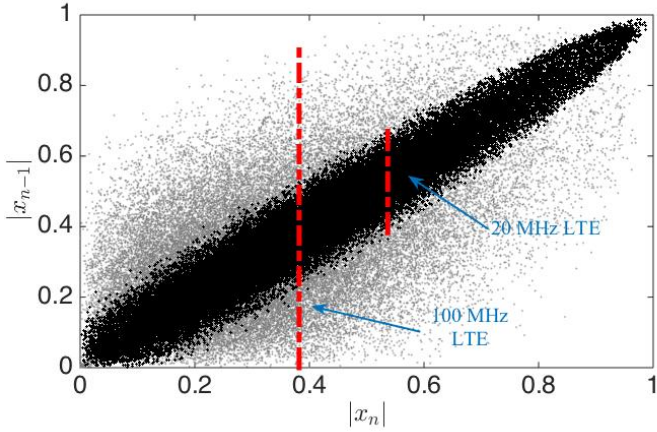

Fig. 2. the Range of x_{n-1} given x_n. The grey point is 100-MHz LTE signal and the black point is 20-MHz LTE signal.
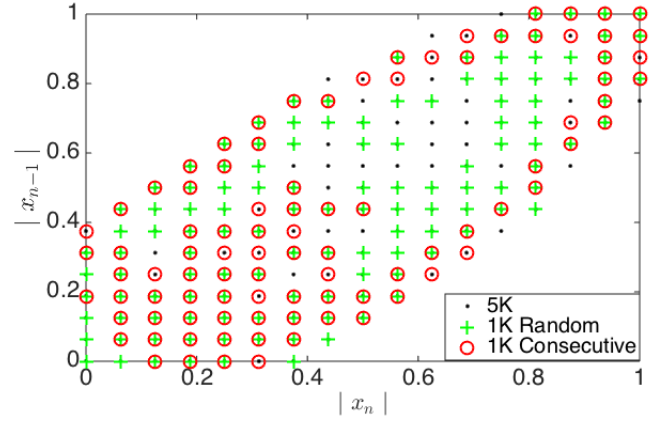

Fig. 3. The distribution of x_{n-1} for a given x_n in 5-bit resolution. The grey point is the whole dataset of 5k multi-tone samples. The red ring is 1K consecutive samples including maximum magnitude. And the green cross is 1K selected samples from random dataset.

the training samples have to satisfy not only (10), but also (12) which includes the probability of nonlinear characteristics, characteristics due to the signal bandwidth and characteristics due to memory effects. So the more complicated memory-effect mechanisms present in new high efficiency PA architectures, such as Doherty, LINC, ET, EER, will require more consecutive samples to guarantee (12c). Similarly for wider bandwidth signals, (12b) becomes harder to train for. As shown in Fig. 2, the red lines indicates the range of $\rho(x_{n-1}|x_n)$ in 100MHz and 20MHz LTE respectively.

B. Compression of Training Data

Firstly we assume there are an over-determined equations with 2 million 14-bit LTE samples and 100 14-bit coefficients of memory polynomials. Then by using LS algorithm, the information of the over-determined system is compressed into the correlation space. And it is worth noticing that no matter how many training samples are used in the optimisation algorithm, (2) is always a 100 by 100 determined equation, which contain 29-bit compressed information. To compress the training process, we design a compressive training algorithm consisting of probability sampling to obtain the same auto-correlation and cross-correlation matrix in (3) by involving the minimal samples, rather than roughly using a large section of consecutive samples.

Procedure:
1: Choose a resolution for the histogram;
2: Generate histogram from the whole input signal dataset based on $|x_n|$;
3: Determine the factor to reduce the length of the dataset by;
4: Obtain a target histogram by proportionally reducing each histogram bin size according to the factor;
5: i) Select a random sample x_j and identify the $|x_j|$ bin based on its magnitude;
   ii) If the count in $|x_j|$ bin is less than the count in $|x_j|$ bin of the target histogram, keep the sample. Otherwise, disregard the sample;
6: Repeat step.5 until all target histogram bins are full;
7: For each selected sample x_j, we also keep the previous samples $(x_j, x_{j-1}, \cdots, x_{j-m}, y_j)$;
8: Use the selected training samples to calculate the coefficients.

The random selection of samples ensures that the terms (12b) and (12c) can be covered correctly since the various distributions of the whole dataset can be conserved by random selection. Fig. 3 shows that the selected samples follow the same $\rho(x_{n-1}|x_n)$ distribution of the whole dataset. Meanwhile because the target histogram is the scaling down histogram of the whole dataset, the information is fully reserved, as shown in Fig. 4.

Because the information of (12) is reserved, $R_{\phi\phi}$ and $R_{\phi y}$ in (3) are guaranteed to yield a set of determined equations. As a result, by means of this method, the type of signal is not restricted to satisfy the cyclostationary process. And the number of training samples, which carry the necessary information to obtain the optimised solution, can be dramatically reduced.
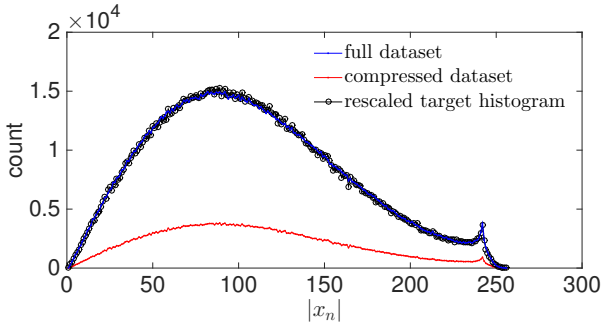
3

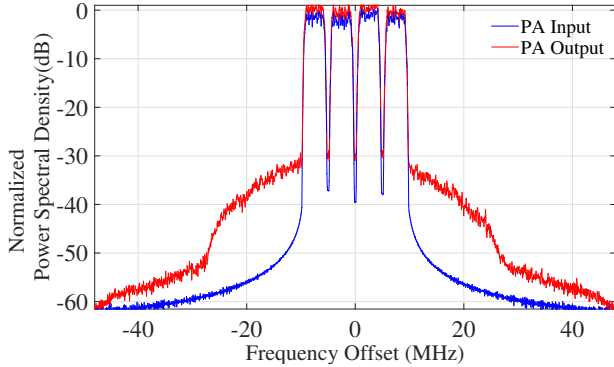Fig. 4. Histogram for full dataset and compressed training dataset

Fig. 5. PA input and output spectra with a 4-carrier WCDMA signal

Fig. 6. Doherty PA Test with 16K complete vector

TABLE I
BEHAVIOURAL MODELING PERFORMANCE

	Standard LS	Compressed Training
Number of Training Samples	8192	681
NMSE(dB)	-45.3	-45.3
Complex Multiplication	26,606,080	2,270,440
Execution Time(second)	11.473	0.092

IV. EXPERIMENTALLY MEASURED RESULTS

To validate the performance of the proposed compressed training signal dataset, an experimental testbench was assembled. The hardware testbench comprised of a Rohde & Schwarz SMU200A, a 40W Doherty power amplifier and a Rohde & Schwarz FSQ. The frequency spectrum of the 4-carrier WCDMA input and PA output are shown in Fig. 5. The 14-bit 16K datasets recorded by the FSQ were used to extract the GMP model[4] in (14) with 40 coefficients. Then the coefficients of the model are calculated by means of standard fixed-point Recursive least squares (RLS) algorithm with consecutive training samples or selected samples, respectively. The results of NMSE versus the number of training samples are shown in Fig. 6 for both the traditional and proposed method. The slightly improvement of compressed algorithm using an increased number of training samples is due to averaging of measurement noise, a function of the LS algorithm. The result is a more general model of the power amplifier. The multiplication operations according to the formula in [1] and execution time by a 2.4GHz Core i5 CPU with 8GB DDR to calculate the coefficients are shown in Table. I.

$$\sum_{k \in K} \sum_{p \in P} \sum_{q \in Q} \omega_{kpq} x(n-p) \mid x(n-q) \mid^k \qquad (14)$$

V. CONCLUSION

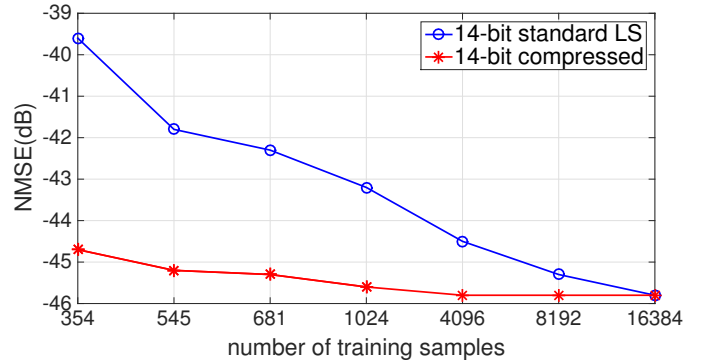The proposed compressed training algorithm is a method to identify samples which carry the necessary information to model a nonlinear dynamic system. This novel algorithm is highly efficient since there is no additional multiplication operation involved. The accuracy of the algorithm is guaranteed by the theory study in Sec. II and is verified using experimentally measured data.

REFERENCES

[1] L. Guan and A. Zhu, "Optimized low-complexity implementation of least squares based model extraction for digital predistortion of rf power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 60, no. 3, pp. 594–603, 2012.

[2] R. B. Benedict, "Selecting samples for amplifier digital predistortion estimation," Jan. 20 2009. US Patent 7,479,828.

[3] J. Goerlich, D. Bruckner, A. Richter, O. Strama, R. Thoma, and U. Trautwein, "Signal analysis using spectral correlation measurement," in *Instrumentation and Measurement Technology Conference, 1998. IMTC/98. Conference Proceedings. IEEE*, vol. 2, pp. 1313–1318, IEEE, 1998.

[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," *Signal Processing, IEEE Transactions on*, vol. 54, no. 10, pp. 3852–3860, 2006.