Low-Rate Identification of Memory Polynomials

Nikolaus Hammler*, Yonina C. Eldar†, and Boris Murmann*
* Department of Electrical Engineering, Stanford University, CA, USA
† Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

***Abstract*—We propose a new approach for the low-rate identification of memory polynomials (MPs), which are frequently used to model RF power amplifiers (PAs). Based on ideas from the finite rate of innovation framework, we find the coefficients of the MP in the frequency domain, which requires a relatively small number of measurements (samples), commensurate with the degrees of freedom in the model. By choosing a random set of frequency components, the stability of the identification is ensured. We show that the method can be used directly for special input signals, such as one used for orthogonal frequency division multiplexing, and extend the idea for arbitrary inputs. Experiments using measured data from a class-AB PA demonstrate the effectiveness of the approach. The sampling rate for identifying this PA is reduced by a factor of 1024 (from $107.52\,\mathrm{MHz}$ to $105\,\mathrm{kHz}$).**

I. Introduction

Digital predistortion is a widely used approach for enhancing the linearity and efficiency of state-of-the-art power amplifiers (PAs) [1]. In a typical system, the weakly nonlinear PA is preceded by a digital predistorter (DPD) that is continuously adapted to approximate the PA's inverse (see Fig. 1). While there are several options for identifying the inverse (direct learning, indirect learning, etc.) [2], a common issue with all existing techniques is the relatively large and growing bandwidth that must be digitized by the analog-to-digital converter (ADC) in the system's observation path. For example, with a signal bandwidth of $100\,\mathrm{MHz}$ in Long Term Evolution (LTE) Advanced, the output spectrum of the PAs spans several hundred megahertz due to spectral regrowth. Digitizing the full output spectrum without aliasing requires expensive and power-hungry ADCs that are beginning to become a roadblock in the design of cost-efficient systems.

Interestingly, it has been recognized that aliasing is permissible for identifying the system, and that the sampling rate f_s of the ADC can be chosen in accordance with the input signal bandwidth, rather than the output bandwidth of the PA [3], [4], [5], [6]. However, this method still calls for an acquisition bandwidth that includes the spectral re-growth to obtain accurate time domain samples, and thus $f_{\mathrm{sig}} \gg f_s/2$ at the input of the ADC (see Fig. 1). Since the cost and power efficiency of an ADC is related to both the acquisition bandwidth and sampling rate (which is in any case just reduced by a relatively small factor), this approach does not lead to a sustainable solution for future systems.

In this paper, we present a new approach to nonlinear system identification in the frequency domain, where the output signal can be sampled at a substantially lower rate, and a smaller number of measurements is needed to find the model coefficients. The method does not rely on aliasing and therefore the ADC's acquisition bandwidth can be reduced in accordance with the low sampling rate. The technique is built around a system setup where one first identifies a behavioral model of the PA, and then determines the DPD function in subsequent steps [1]. We limit the scope of this paper to the former aspect, i.e. the identification of the PA.

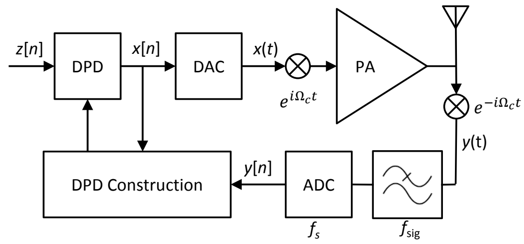

Fig. 1. RF power amplifier with digital predistortion.

The rest of this paper is organized as follows. In Section II we show that it is possible to identify the PA by extracting a set of Fourier coefficients. Section III then explains how these coefficients are obtained. In Section IV, we evaluate the technique using experiments with one synthetic and one real hardware PA. Section V concludes this paper.

II. Identifying the Nonlinear System

A popular choice for modeling nonlinear dynamical systems is the Volterra series. However, due to the high model complexity and overmodeling issues, truncated variants are used in practice, where the memory polynomial (MP) provides a good tradeoff between complexity and performance [1]. In this model,

$$y[n] = \sum_{p=1}^{P}\sum_{q=0}^{Q} c_{pq} \cdot z_p[n-q], \quad z_p[n] = x^p[n]. \tag{1}$$

Here, $x[n]$ is the input signal, $y[n]$ is the output signal that results from sampling $y(t)$ at the Nyquist rate f_{NYQ} of the model, P is the nonlinearity order, Q is the memory depth and c_{pq} are the model coefficients.

Usually, c_{pq} are found in the time domain, either on a sample-by-sample basis using algorithms like least mean squares (LMS) [7] or least squares (LS) [8]. Motivated by results from the finite rate of innovation (FRI) framework [9], [10], we propose solving the system in the frequency domain. In the continuous time domain, the p-th order kernel of (1) can be regarded as an FRI signal with fixed time delays:

978-1-4799-3432-4/14/$31.00 ©2014 IEEE

$$h_p(t) = \sum_{q=0}^{Q} c_{pq} \cdot \delta(t - t_q) \tag{2}$$

where $t_q = q/f_{\mathrm{NYQ}}$. Since (2) is defined by no more than $Q+1$ degrees of freedom (DoF), we expect that (1) can be identified by a minimum of $P \cdot (Q+1)$ samples. Hence, our goal is to reduce the sampling rate as much as possible, independent of the model's Nyquist rate. A solution for identifying an impulse response obeying the FRI model was presented in [11]. This approach works for arbitrary input signals but relies on an approximation of the discrete-time Fourier transform (DTFT) of the impulse response by the discrete Fourier transform (DFT), which is only met under certain conditions. Furthermore, the described method is unstable when the taps of the impulse response are clustered around the origin, which is usually the case for a PA model.

In the following two subsections, we first outline a solution for the case of periodic input signals and then generalize to arbitrary signals.

A. Periodic Input Signals

We start by applying the DTFT to (1):

$$Y(e^{j\omega}) = \sum_{p=1}^{P} \sum_{q=0}^{Q} c_{pq} \cdot Z_p(e^{j\omega}) \cdot e^{-j\omega q} \tag{3}$$

where the DTFT is defined as

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] \cdot e^{-j\omega n}. \tag{4}$$

The problem with (3) is that each value for $Y(e^{j\omega})$ requires an integration of the analog signal $y(t)$ from $-\infty$ to ∞. However, if we assume a periodic input $x[n]$ of length N, the output signal is also periodic. In this case, the spectra $Y(e^{j\omega})$ and $Z_p(e^{j\omega})$ can be sampled at $\omega = \frac{2\pi}{N}k$ to obtain

$$Y[k] = \sum_{p=1}^{P} \sum_{q=0}^{Q} c_{pq} \cdot Z_p[k] \cdot e^{-j\frac{2\pi}{N}kq} \tag{5}$$

where $Y[k]$ and $Z_p[k]$ denote the DFT of $y[n]$ and $z_p[n]$, respectively, with the DFT defined as:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}kn}. \tag{6}$$

Note that the DFT corresponds to the Fourier coefficients of the analog signal $y(t)$ using N samples at $t = nT_{\mathrm{NYQ}}$. More precisely, the PA performs an aperiodic convolution while the DFT assumes cyclic convolution. Let us now assume that the input signal is split into N-length blocks, and denote the l-th block as $x_l[n], 0 \leq n \leq N-1$. The discrepancy is then captured by two factors (illustrated in Fig. 2(b)):

1) Parts of the convolution for $x_{l-1}[n]$ leak into $x_l[n]$.
2) The convolution of the end of $x_l[n]$ leaks into $x_{l+1}[n]$. But this part must wrap around to the beginning of $x_l[n]$.

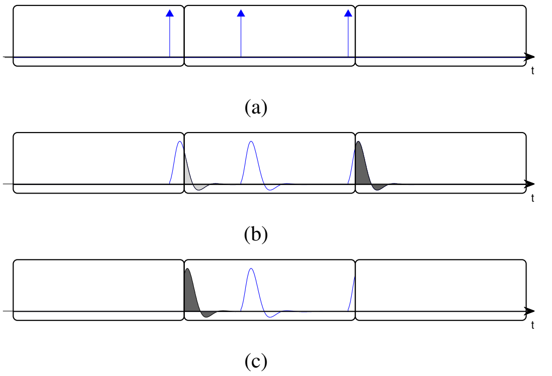

Fig. 2. Correction of the DFT (a) 3-block input signal (b) aperiodic convolution. Signal leaking from previous block (light gray). Signal leaking into next block (dark gray) (c) corrected block by removing light contribution and wrapping dark contribution as done by (15).

Apart from periodic signals, the equivalence between cyclic and aperiodic convolution is maintained for finite-length signals or for signals with a cyclic prefix (CP), as for instance in orthogonal frequency division multiplexed (OFDM) communication. The only requirements are that the impulse response is shorter than the CP and that the block length N matches the useful symbol duration of the OFDM signal. The coefficients are then found from (5) by obtaining at least $P(Q+1)$ Fourier coefficients.

A similar result was recently published in [12]. However, this solution relies on probing the system with $P \cdot (Q+1)$ sinusoidal input signals and only makes use of the p-th harmonic in the output signal.

B. Arbitrary Input Signals

Based on the approach described in [13], we incorporate appropriate boundary conditions to ensure the equivalence between $Y[k]$ of an arbitrary block and the right side of (5). To simplify notation, we define a windowed DTFT as follows:

$$X^{a,b}(e^{j\omega}) = \sum_{n=a}^{b} x[n] e^{-j\omega n}. \tag{7}$$

The DTFT used in (3) sums $y[n]$ from $-\infty$ to ∞. However, the objective is to find an expression for $Y^{0,N-1}(e^{j\omega})$ as this would correspond to the definition of the DFT over a block from $n = 0, \ldots, N-1$. In the first step, we let $n = 0, \ldots, \infty$. For the result to be valid, the sum over $z_p[n]$ must be adjusted to incorporate values down to $n = -Q$. This results in

$$Y^{0,\infty}(e^{j\omega}) = \sum_{p=1}^{P} \sum_{q=0}^{Q} c_{pq} \cdot e^{-j\omega q} \cdot Z_p^{-q,\infty}(e^{j\omega}). \tag{8}$$

The term for Z_p can be split as follows:

$$Z_p^{-q,\infty}(e^{j\omega}) = Z_p^{-q,-1}(e^{j\omega}) + Z_p^{0,\infty}(e^{j\omega}) \tag{9}$$

$$= Z_p^{0,\infty}(e^{j\omega}) + \sum_{n=1}^{q} z_p[-n] e^{j\omega n}. \tag{10}$$

In the second step, we let $n = N, \dots, \infty$. Again, for the result to be valid, the summation for $z_p[n]$ must start q samples earlier:

$$Y^{N,\infty}(e^{j\omega}) = \sum_{p=1}^{P}\sum_{q=0}^{Q} c_{pq} \cdot e^{-j\omega q} \cdot Z_p^{N-q,\infty}(e^{j\omega}), \tag{11}$$

with

$$Z_p^{N-q,\infty}(e^{j\omega}) = Z_p^{N-q,N-1}(e^{j\omega}) + Z_p^{N,\infty}(e^{j\omega}). \tag{12}$$

The term $Z_p^{N-q,N-1}$ can be simplified as follows:

$$Z_p^{N-q,N-1}(e^{j\omega}) = \sum_{n=-q}^{-1} z_p[n+N]e^{-j\omega(n+N)} \tag{13}$$

$$= e^{-j\omega N}\sum_{n=1}^{q} z_p[N-n]e^{j\omega n}. \tag{14}$$

In the final step, (14) is subtracted from (8), which yields:

$$Y^{0,N-1}(e^{j\omega}) = \sum_{p=1}^{P}\sum_{q=0}^{Q} c_{pq} \cdot e^{-j\omega q}\Bigg(Z_p^{0,N-1}(e^{j\omega}) + \\ + \sum_{n=1}^{q}\left(z_p[-n] - z_p[N-n]\right)e^{j\omega(n-q)}\Bigg).$$

This result equals the DFT for $\omega = \frac{2\pi}{N}k$ and resembles (5) very closely. The extra sum corresponds to the boundary conditions for the leakage discussed previously. This is illustrated in Fig. 2. This term equals zero for finite signals with a finite length, periodic signals or signals with an appropriate CP.

To summarize, the result obtained by using the DFT of the non-periodic output signal $y[n]$ can be corrected by adding the following correction term to $Z_p[k]$ in (5):

$$\sum_{n=1}^{q}\left(z_{p,-1}[N-n] - z_{p,0}[N-n]\right)e^{j\frac{2\pi}{N}k(n-q)}. \tag{15}$$

The method can be generalized to the full Volterra model [2], which results in:

$$Y[k] = \sum_{p=1}^{P}\sum_{q_1=0}^{Q_1}\cdots\sum_{q_p=0}^{Q_p} c_{p,q_1,\cdots,q_p}\mathrm{DFT}\left\{\prod_{r=1}^{p} x[n-q_r]\right\}. \tag{16}$$

Unfortunately, due to the cross terms, the result cannot be written in terms of $Z_p[k]$ because the result is a p-fold convolution of $X[k]e^{-j2\pi kq_r/N}$. However, this approach allows us to extend the MP with cross terms. Although a DFT is required for each additional cross term, it is reasonable when the number of these coefficients is small.

III. Obtaining the Fourier Coefficients

While (5) can be solved using a low pass approximation, it was proposed to sample the spectrum at random locations [14], which greatly improves the reconstruction stability. For this reason, we select a set of $M \geq P \cdot (Q+1)$ randomly chosen frequency bins. The effect of the frequency selection on the reconstruction stability is currently being investigated. Based on practical considerations, we decided to use sequential demodulation: In each block of length $T_{\text{blk}} = 1/f_{\text{blk}}$, the input signal is demodulated, integrated for T_{blk} and sampled at rate f_{blk} (see Fig. 3). The demodulation approach works if the mixing sequence is guaranteed to run over a whole block. Since we need to allow for settling of the modulation sequences, it takes two blocks to demodulate one coefficient (see Fig. 4). Furthermore, following the developments presented in [10], it is possible to extend the architecture to multiple branches so as to recover additional coefficients. Per $2L$ demodulation branches, L coefficients per block can be obtained. T_{blk} is chosen given the tradeoff between the undersampling ratio, identification time, and number of branches. During identification, the system is assumed to be time-invariant, which is a reasonable assumption for PAs [15]. The values of $Z_p[k]$ in (5) are calculated in the time domain from $x^p[n]$. Since only LP bins per block are required, they can be calculated efficiently using the Goertzel algorithm.

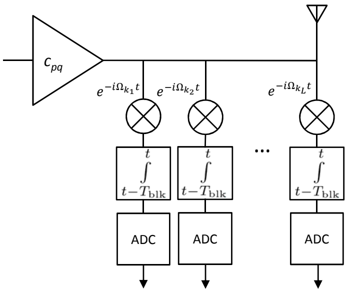

Fig. 3. Sequential demodulation with L branches.

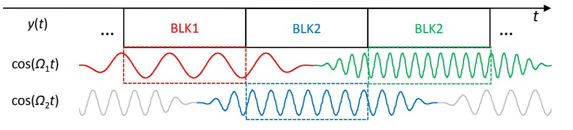

Fig. 4. Demodulation of one Fourier coefficients per block. The correct result is obtained when the mixing sequences are guaranteed to run over a whole block.

IV. Numerical Results

To verify the functionality of the method, we use the PA model from example 2 of [8], which was extracted from a real class-AB PA. A 20 MHz LTE signal is generated, upsampled to 307.2 MHz and fed through the model. Using a block size of $N = 4096$, the 15 coefficients are recovered up to machine precision (MSE $= 3.8 \cdot 10^{-13}$) after 15 blocks (critical sampling).

In our second experiment, we use data from a GaN class-AB PA carrying a 5 MHz Wideband Code Division Multiple Access (WCDMA) signal. The data consists of two sets of input and output signals of the PA, each $N = 32768$ samples long and sampled at 107.52 MHz. The first set is used to identify the model (i.e., the MP), whereas the second set is used to assess the quality of the model. An MP with nonlinearity

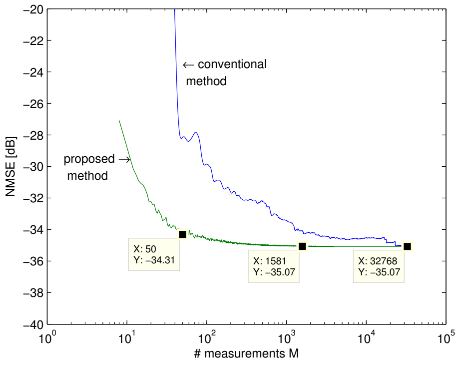

Fig. 5. NMSE vs. number of measurements. For the proposed method, the NMSE is averaged over 50 trials with randomly chosen frequency locations and a fixed input signal.

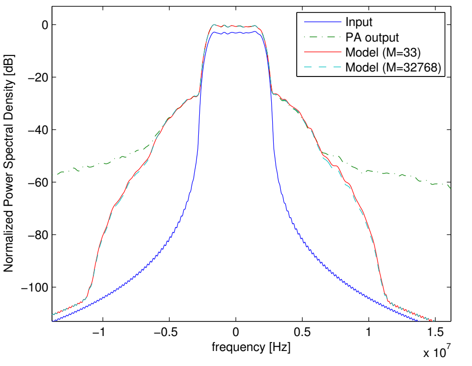

Fig. 6. Output spectrum of a PA identified with $M = 33$ samples (proposed, NMSE: -34.31 dB) and with $M = 32768$ (conventional, NMSE: -35.07 dB).

orders $p = \{1, 3, 5\}$ and a memory depth of 2 taps ($Q = 1$) is assumed.

Fig. 5 shows the normalized mean square error (NMSE) for conventional approaches [7], [8] which use all samples at the Nyquist rate and the proposed approach for different numbers of samples M. We choose a block size of $N_{\text{blk}} = 1024$, corresponding to a sampling rate of $105\,\text{kHz}$ (a 1024-fold reduction). It can be seen that the proposed approach requires far fewer samples than the conventional approach for a certain model accuracy. For example, our method achieves the same NMSE as the conventional approach with $M = 32768$ samples using only $M = 1581$ samples. With the low sampling rate of $105\,\text{kHz}$, the identification time is increased by a factor of 49 compared to the conventional approach. The identification time can be reduced by increasing the number of branches or by sacrificing model accuracy. For example, an NMSE of -34.31 dB (<1 dB penalty relative to the minimum in Fig. 5) requires only $M = 33$ samples. In this case, the identification time is roughly 56% longer than the conventional approach with $M = 32768$. Fig. 6 shows the power spectrum of the input and output signals along with the model outputs.

Although the proposed method suggests an increased computational load due to additional calculations of frequency bins, it requires far fewer samples and results in a smaller LS system. Potentially, this results in a lower computational cost.

V. Conclusion

We have presented a new approach for the low-rate identification of MPs and explored the idea in the context of PA system identification. Compared to conventional schemes, this method avoids sampling the output signal at the Nyquist rate and does not require the acquisition of the full spectrum. The model coefficients are obtained by measuring a random set of components using a frequency domain model. The technique offers a flexible trade-off between accuracy, complexity and identification time through adjusting the block size and the number of branches.

Acknowledgment

This work was funded by the Stanford Initiative for Rethinking Analog Design (RAD). We thank Anding Zhu for providing the test data used in this paper.

References

[1] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE microwave magazine*, vol. 10, no. 7, pp. 52–64, 2009.

[2] A. Zhu, "Digital predistortion and its combination with crest factor reduction," in *Digital Front-End in Wireless Communication and Broadcasting*. Cambridge University Press, 2011.

[3] Y.-M. Zhu, "Generalized sampling theorem," *IEEE Trans. on Circuit and Systems II*, vol. 39, no. 8, pp. 587–588, 1992.

[4] W. Frank, "Sampling requirements for volterra system identification," *IEEE Signal Processing Letters*, vol. 3, no. 9, pp. 266–268, 1996.

[5] H. Koeppl and P. Singerl, "An efficient scheme for nonlinear modeling and predistortion in mixed-signal systems," *IEEE Trans. on Circuit and Systems II: Express Briefs*, vol. 53, no. 12, pp. 1368–1372, 2006.

[6] L. Ding, F. Mujica, and Z. Yang, "Digital predistortion using direct learning with reduced bandwidth feedback," in *2013 IEEE MTT-S International Microwave Symposium Digest*, Jun 2013, pp. 1–3.

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

[8] L. Ding, G. Zhou, D. Morgan, M. Zhengxiang, J. Kenney, K. Jaehyeong, and C. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. on Comm*, vol. 52, pp. 159–165, Jan 2004.

[9] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, Jun 2002.

[10] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Trans. on Signal Processing*, vol. 59, pp. 1491–1504, 2011.

[11] M. McCormick, Y. M. Lu, and M. Vetterli, "Learning sparse systems at sub-nyquist rates: A frequency-domain approach," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010.

[12] A. Bolstad and B. A. Miller, "Sparse volterra systems: Theory and practice," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 5740–5744.

[13] R. Pintelon, J. Schoukens, and G. Vandersteen, "Frequency domain system identification using arbitrary signals," *IEEE Trans. on Automatic Control*, vol. 42, no. 12, pp. 1717–1720, Dec 1997.

[14] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. on Signal Processing*, vol. 60, no. 9, pp. 4643–4657, Sep 2012.

[15] L. Ding, "Digital predistortion of power amplifiers for wireless applications," Ph.D. dissertation, Georgia Institute of Technology, 2004.