MULTI-STAGE ADAPTIVE PREDISTORTION OF HPA SATURATION EFFECTS FOR DIGITAL TELEVISION TRANSMISSION

J.T. Stonick and V L. Stonick
Electrical & Computer Engineering
Oregon State University
Corvallis, OR 97331-3211
email: jts@ece.orst.edu

J.M.F. Moura
Electrical & Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213-3890

ABSTRACT

This paper presents a new structure for adaptive predistortion of the memoryless, nonlinear saturation effects caused by High Power Amplifiers (HPA). Timely compensation for HPA distortions is critical for cost-effective prevention of the cliff effect during terrestrial transmission of digital broadcast television. The new structure results from a two-stage approach: The forward model is identified first from measured data, and then the inverse to the forward model is computed. Replacing the analog system by the HPA forward model in the second stage eliminates measurement noise and analog system delays, resulting in faster adaptation and less solution bias. Additionally, block processing reduces noise in the forward modeling stage. In the inverse modeling stage, the use of synthetic data and a closed form expression for the gradient result in more efficient convergence and more accurate solutions. Simulation results using measured HPA data demonstrate an average 5 dB improvement in SINAD for standard SNR operating ranges.

1. INTRODUCTION

In terrestrial transmission of broadcast television, the cost of the high power amplifier (HPA) in the transmitter dominates the overall system cost. For a given power and noise rating, HPA cost increases with the range of input signal levels over which amplification is linear. Predistortion is the mapping of the input signal through an approximation to the inverse of the HPA distortion prior to amplification, thereby extending the range over which HPA amplification appears linear.

For analog systems, the predistorters used in terrestrial television transmitters are low-cost, manually-tunable circuits in the low power, intermediate frequency (IF) stage. These predistorters work well for analog television, where picture quality degrades gracefully as the nonlinear characteristic of the HPA changes with temperature and aging, and thus require only periodic retuning. However, digital television suffers from the cliff effect, a sudden and complete loss of picture from one of nearly perfect quality. The cliff effect occurs when the end-to-end distortions overload the error correction capabilities of the receiver. The multi-level signaling needed to achieve the high data rates required for digital television transmission results in high sensitivity to HPA distortions. Timely adaptive correction of HPA effects is critical to prevent the cliff effect, particularly in the presence of time-varying and unknown channel distortions.

A number of adaptive predistortion structures and methodologies have been proposed [2, 3, 4, 5, 6]. Like the method presented here, the majority provide off-line adaptation and perform the on-line digital pre-distortion on baseband data using Look-Up-Tables (LUT). However, all prior methods adapt the optimal predistorter directly, resulting in an inverse modeling problem. These methods suffer from the slow convergence, solution bias [1] and sensitivity to measurement error that is typical of inverse problems. In addition, each adaptive iteration has a delay corresponding to the time taken for the transmitted symbols to propagate through analog modulation, amplification and demodulation stages, a per-update delay on the order of 40 sample periods.

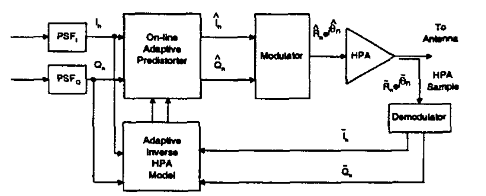

Figure 1: Prior Predistortion Methods

This paper presents a new structure for adaptive HPA predistortion that provides improved end-to-end system performance over existing methods. By identifying the forward model for the HPA distortion first and then computing the inverse to this forward model (instead of directly finding the HPA inverse model), faster convergence to an unbiased solution results, even in the presence of measurement noise.

2. PREDISTORTION PROBLEM

A typical structure for an adaptive predistorter in a digital television transmitter is shown in Fig. 1. The input information is separated into two channels, representing multi-level in-phase (I) and quadrature (Q) data for Quadrature Amplitude Modulation (QAM) systems [1]. The input data is passed through over-sampled, Nyquist pulse-shaping filters to create the input to the on-line adaptive predistorter. Oversampling facilitates pre-correction of out-of-band emissions (to meet the FCC mask) as well as in-band distortions, both created by the amplitude and phase distortions of the HPA. The predistorted data then undergoes digital-to-analog (D/A) conversion, and is modulated, amplified and transmitted. The output of the HPA also is demodulated and sampled for use in predistorter adaptation.

The input to the predistorter comprises complex baseband symbols, denoted as

$$I_n + jQ_n = R_n \cos(\theta_n) + jR_n \sin(\theta_n) = R_n e^{j\theta_n} \quad (1)$$

[1]For Vestigal Sideband (VSB) standard systems, the two channels represent the data and its Hilbert Transform

Supported in part by Ben Franklin Foundation Grant RD10052 and NSF Grant MIP-9157221. Data measurements with support from ITS Corporation in McMurray, PA. Patent Pending, 08/852,944

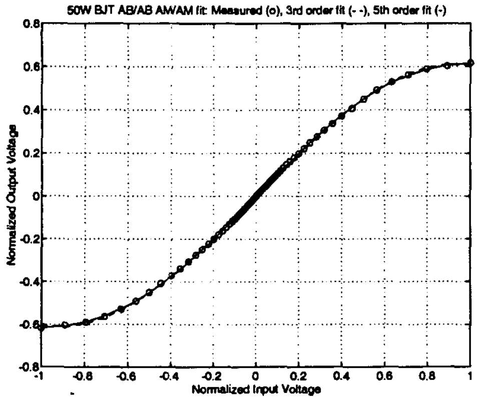

Figure 2: BJT class AB AM Char.

where $R_n = \sqrt{I_n^2 + Q_n^2}$ represents the amplitude and $\theta_n = -\arctan(Q_n/I_n)$ the phase for each symbol. The symbols out of the predistorter are denoted by

$$\hat{R}_n e^{j\hat{\theta_n}} = \hat{I}_n + j\hat{Q}_n \tag{2}$$

where $\hat{R}_n$ and $\hat{\theta}_n$ denote the predistorted amplitude and phase, respectively, of the signal to be amplified and transmitted. Note that (2) also is baseband-equivalent input to the HPA since modulation does not affect *baseband* symbols. The demodulated output of the HPA is

$$\tilde{A}_n e^{j\tilde{\theta}_n} = \tilde{I}_n + j\tilde{Q}_n, \tag{3}$$

which [1-5] contains amplitude-dependent amplitude $A_n\left(\hat{R}_n\right)$ (AM/AM) and phase $\phi_n(\hat{R}_n)$ (AM/PM) distortions, *i.e.*,

$$\tilde{A}_n e^{j\tilde{\theta}_n} = A_n\left(\hat{R}_n\right) e^{j(\hat{\theta_n} + \phi_n(\hat{R}_n))} \tag{4}$$

The AM/AM and AM/PM distortions are smooth and memoryless nonlinearties, known to be well-modeled by odd and even power polynomials respectively[7]. Fig. 2 shows the measured AM/AM distortion, along with polynomial models of different degrees, for a typical HPA (For AM/PM distortion measurements and models, see [1]). Since the amplitude and phase distortion depend only on the amplitude of the input to the HPA, the distortions can be corrected for separately.

The goal of predistortion is to generate warped values, $\hat{I}_n$ and $\hat{Q}_n$, such that the baseband symbols out of the HPA $\tilde{I}_n$ and $\tilde{Q}_n$ are equal to the original input symbols I_n and Q_n. This goal requires solving two independent problems: First, the magnitude of the data must be warped such that

$$A_n\left(\hat{R}_n\right) = R_n. \tag{5}$$

Second, the input phase must be warped such that

$$\hat{\theta}_n + \phi_n\left(\hat{R}_n\right) = \theta_n. \tag{6}$$

After identifying the AM/PM distortion, $\phi_n(\hat{R}_n)$, the optimal correction requires only subtraction,

$$\hat{\theta}_n = \theta_n - \phi_n\left(\hat{R}_n\right) \tag{7}$$

The phase identification procedure is discussed in [1] and will not be treated further in this paper. Correcting for the AM/AM distortion – which is functional, not additive - is more difficult and is the focus of this paper.

Previous approaches, as in Fig. 1, use the predistorter input $\{I_n, Q_n\}$ and HPA demodulated output $\{\tilde{I}_n, \tilde{Q}_n\}$ as data to adaptively update the off-line predistorter, resulting in an inverse modeling problem. The measured output $\{\tilde{I}_n, \tilde{Q}_n\}$ contains both system distortion and measurement noise. Coefficients for the off-line adaptive predistorter are coefficients of a polynomial model for the inverse HPA, which must be several degrees higher than the forward model.

Typical predistorter operation is as follows: After the input signal $\{I_n, Q_n\}$ propagates through the predistorter, modulator, HPA and demodulator, the result $\{\tilde{I}_n, \tilde{Q}_n\}$ is used to compute the error used in the adaptive update of the model parameters. The delay through the analog sections typically is on the order of 20 - 100 μs, where 20 μs corresponds to about a 40 symbol period delay for broadcast digital television transmission.

These direct inversion methods have a number of drawbacks:

1. Inverse modeling is known to be more sensitive to noise than forward modeling, particularly when inverse models correspond to higher-order nonlinearities, as is the case here. Noise sensitivity results in slower convergence and increased solution bias, particularly for coefficients of higher-order terms, which are critical in inverting HPA saturation effects.[?]
2. High amplitude symbols, which are subject to the most HPA distortion and thus more critical for determining the best predistorter, occur less during normal operation. Adaptation rates for critical higher-order coefficients are limited by the frequency of occurrence of these symbols.
3. The gradient term required in the adaptation depends upon the gradient of the forward model, which is not available. Errors in estimates of the gradient result in slow adaptation and increased error in the resulting estimates, even in the absence of measurement noise [1].
4. The analog system delay causes an approximately 40 sample delay between successive updates due to the data acquisition needed for adaptation

3. NEW STRUCTURE

The new predistorter, as it would appear in a QAM transmitter for digital television, is shown in Fig. 3. The on-line structure is identical to prior methods. The primary differences are in the method of off-line adaptation. Rather than using the predistorter input as data to determine the optimal predistortion directly, the predistorter output (HPA input) is used with the HPA output data to identify the HPA AM/AM and AM/PM distortions first, resulting in a multi-stage approach. In the first stage, the difference between $(\hat{I}_n, \hat{Q}_n)$ and $(\tilde{I}_n, \tilde{Q}_n)$ is used to adapt the parameters, $\{a_n\}$ and $\{p_n\}$, of polynomial models for the HPA AM/AM and AM/PM distortions respectively. The model parameters then are passed to a second stage in which the inverse to the forward HPA model is computed, where the inverse models are polynomials of higher-order. The multi-stage procedure can be summarized as follows: After a single analog system delay, HPA I/O data is collected in a 32K ROM, and Recursive Least Squares (RLS) is used to adaptively compute the forward model parameters. Concurrently, new data is being collected, resulting in a pipelined system and a one-time, rather than per-update, delay. Block processing facilitates a noise reduction procedure. The HPA forward modeling is described in Sect. 3.1. In the second stage, the parameters of the inverse to the HPA model are adaptively computed using PLMS [1]. Since the HPA forward model is used in place of the analog system, no measurement noise and no analog system delays exist..

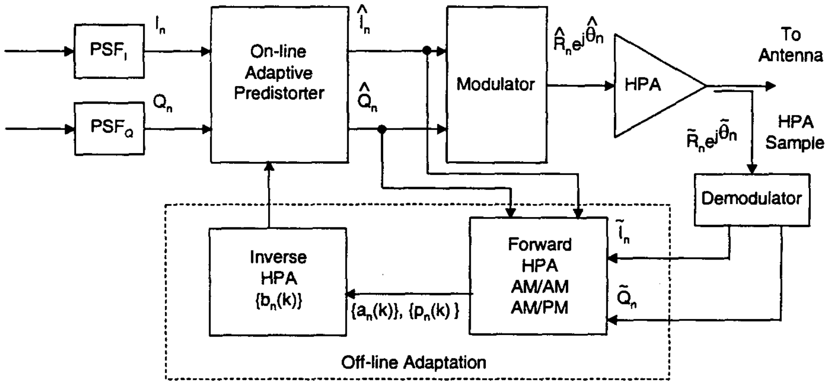

Figure 3: New Adaptive Predistorter

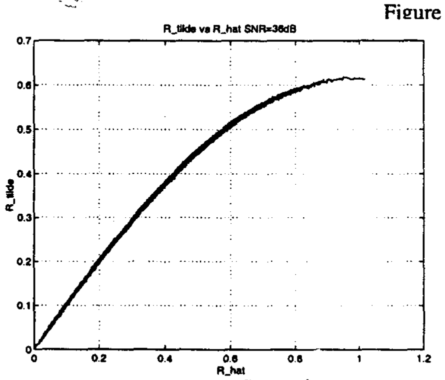

Figure 4: $\tilde{R}_n$ vs $\hat{R}_n$

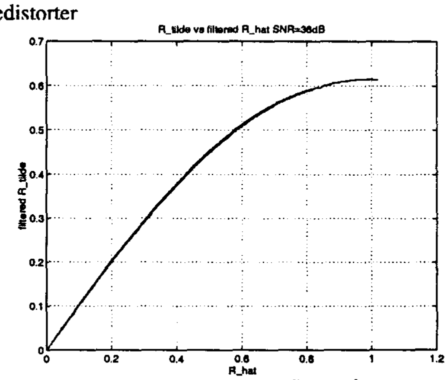

Figure 5: Filtered $\tilde{R}_n$ vs $\hat{R}_n$

Since adaptation is off-line, synthetic data sets - constructed to provide efficient adaptation - are used. The inverse modeling process is described in more detail in Sect. 3.2. The updated inverse model is used to recompute all elements of the LUT simultaneously. for downloading to the on-line predistorter.

3.1. Forward HPA Modeling

In the computation of the forward model, noise power is reduced by using block processing. Two sets of data are collected in FIFOs (described here in polar form even though the collected data are in rectangular form): The predistorted data $(\hat{R}_n, \hat{\theta}_n)$, and the recovered data $(\tilde{R}_n, \tilde{\theta}_n)$, for $n = 0, 1, 2, \ldots, N$. The data pairs are then re-ordered in ascending order based R_n. The result of this reordering for the amplitude data is demonstrated in Fig. 4. The data shown here was contaminated by a -36 dB additive noise distortion. While this re-ordering highly correlates the signal, it is perfectly legitimate since the distortion is memoryless.

To reduce the noise before solving the forward identification problem, the ordered sequences $(\tilde{R}_n, \tilde{\theta}_n)$ are treated as highly oversampled time waveforms and filtered to reject the out-of-band noise. The noise reduction in the AM/AM data is observed by comparing $\breve{R}_n$, the filtered data of Fig. 5, with the unfiltered data of Fig. 4. The data is then returned to its original order, and RLS is used to compute the optimal polynomial model $\{a_n(k)\}$.

$$\min_{a_n(k)} \sum_{n=0}^{N-1} \left(\sum_{m=0}^{M-1} a(m)\hat{R}_n^{2m+1} - \breve{R}_n \right)^2 \tag{8}$$

The phase identification follows the same procedure. For the above example (36dB SNR), prefiltering the data resulted in a 6 dB decrease in converged parameter variance in addition to faster convergence and reduced bias.

3.2. Inverse HPA Modeling

The inverse of the HPA AM/AM distortion is computed, not from the noisy measured data, but rather as the inverse of the model identified Sect. 3.1. The procedure is illustrated in Fig. 6. Coefficients of the polynomial inverse model are updated using the PLMS algorithm [1].

$$b_{n+1}(i) = b_n(i) - \alpha e_n R_n^i \nabla_{\breve{R}_n} A_n\left(\breve{R}_n\right) \tag{9}$$

One advantage over previous approaches is that the gradient of the AM/AM distortion (forward model) - required in (9) - is available in closed form:

$$\nabla_{\breve{R}_n} A_n\left(\breve{R}_n\right) = \sum_{m=0}^{M-1} (2m+1)a(m)\breve{R}_n^{2m} \tag{10}$$

Inclusion of this gradient was shown in [1] to significantly increase convergence speed and decrease solution bias.

The second major advantage of this approach is that the off-line implementation allows the use of synthetic data to drive the system of Fig. 6. By using noiseless data that is uniformly distributed over the full range of input signal values, the entire inverse is characterized accurately and efficiently. In contrast, direct inversion approaches have the least data at the largest input amplitude values where it is needed most. While this problem arises during the first stage, it is less problematic due to the faster convergence and lower noise sensitivity of forward relative to the inverse problems. The overall effectiveness of this new multi-stage structure is illustrated next.

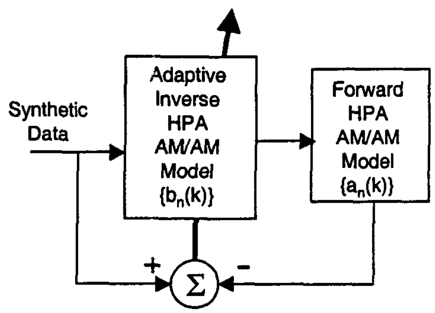

Figure 6: AM/AM Inverse Modeling

4. SIMULATION RESULTS

This section presents simulation results demonstrating the increased effectiveness of this new multi-stage structure when compared to structures that directly compute the inverse. The results consist of simulating the system of Fig. 3, which computes the predistorter as the inverse of the HPA forward model, (new) and that of Fig. 1, in which the predistorter is computed using direct inverse modeling, (old). The input test signal comprises a 64 QAM signal with a maximum envelop value of 0.72, which is slightly above the maximum saturated voltage of the HPA, amplified with an HPA having the AM/AM characteristic shown in Fig. 2 and the AM/PM characteristic shown in [1]. Measurement noise is added to generate signals having SNR values ranging from 20 to 100 dB. After running the simulations for 10000 samples, the resulting output SINAD (signal to noise and distortion) values are computed. SINAD basically measures the total error, as would be seen by the error-correcting device at the receiver. Results shown in Fig. 7 are the average of the values obtained from two sets of data.

When there is significant measurement noise, corresponding to SNR values of less than 40 dB, the impact of the HPA distortion is low relative to that of the noise. In this case, the two methods show comparable performance, with the multi-stage method demonstrating a slight advantage ($\leq$ 1 dB) due to the noise reduction used in the forward modeling process. A marked improvement occurs as the SNR is increased through the range of SNR values (40-60 dB) corresponding to standard operating conditions, with a peak improvement of about 5 dB at an input SNR of 60 dB. This 5 dB improvement provides a significant buffer, allowing an additional 5 dB of channel distortion without occurrence of the cliff effect. As the noise goes to zero, performance differences again diminish, with the new method again showing a slight advantage ($\approx$ 1 dB) as a result of using PLMS instead of LMS to update the parameters of the inverse model.

It is important to note that the same number of points was used in simulating both old and new systems. However, the new approach converges 10 to 20 times faster than traditional methods since the HPA model replaces the analog system, eliminating the impact of analog system delay on adaptation rates.

5. CONCLUSION

This paper presents a new structure for adaptive predistortion of memoryless HPA nonlinearities to be used in digital television transmitters. This structure provides a two-stage approach where the forward model is identified first from measured data, and then the inverse to the forward model is computed. By replacing the analog system by the HPA forward model in the second stage, measurement noise and analog system delays are eliminated. As a result, the solution bias and slow convergence associated with noise in the inverse modeling process is eliminated, and analog system delays do not impact adaptation rates. Additional gains include improved robustness to noise, faster convergence rates. and reduced bias in solutions. These advantages result from using block noise reduction in the forward modeling stage and PLMS [1] to adapt the inverse model. In contrast to direct inversion methods, this new multi-stage structure requires only a single pipeline. rather than an inter-update, delay on the order of 20 - 100 μs. Simulation results demonstrate up to a 5 dB improvement in SINAD for operation within the standard SNR range of 40-60 dB, which significantly extends the cliff effect threshold.

6. REFERENCES

[1] J.T. Stonick, V.L. Stonick, J.M.F Moura, R.S. Zborowski, "Memoryless Polynomial Adaptive Predistortion", ICASSP95 , pp 981-4, May 9-12,1995.

[2] E. Biglieri, *et al*, "Analysis and Compensation of Nonlinearities in Digital Transmission Systems", IEEE JSAC, vol 6, no 1, Jan 1988, 42-51.

[3] G. Karam and H. Sari, "Generalized Data Predistortion Using Intersymbol Interpolation", Philips J. of Research, vol 46, no 1, 1991, pp 1-22.

[4] A. Saleh and J. Salz, "Adaptive Linearization of Power Amplifiers in Digital Radio Systems", BSTJ, April 1983, pp 1019-1033.

[5] S. Stapleton and F. Costescu, "An Adaptive Predistorter for a Power Amplifier Based on Adjacent Channel Emissions", IEEE Trans on VT, vol 41, no 1, Feb 1992, pp 49-56.

[6] R.C. Davis and W. Boyd, "Adaptive Predistortion Technique for Linearizing a Power Amplifier for Digital Data Systems", U.S. Patent 4,291,277, issued 22 September 1982.

[7] W. Bösch and G. Gatti, "Measurement and Simulation of Memory Effects in Predistortion Linearizers", IEEE Trans on MTT, vol 37, no 12, Dec 89, pp 1885-1890.

[8] V.J. Mathews, "Adaptive Polynomial Filters", IEEE Signal Proc. Mag., July 1991, pp 10-26.

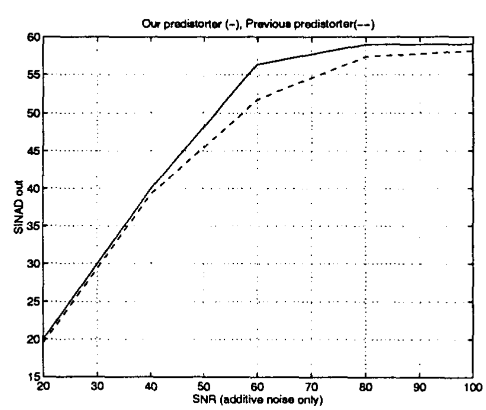

Figure 7: Simulation Results New(-)vs Old(- -)