CONDITIONAL ENTROPY BASED COARSE TIME DELAY ESTIMATION IN THE PRESENCE OF AM/AM AND AM/PM NON-LINEARITIES

Josep Sala

Dpt. of Signal Theory and Communications. Universitat Politècnica de Catalunya
c/Jordi Girona 1-3,Campus Nord UPC D5, 08034 Barcelona, Spain
E-mail: {alvarez}@gps.tsc.upc.es

ABSTRACT

Pre-distortion algorithms for the compensation of AM/AM and AM/PM non-linearities of High Power Amplifiers (HPA) require a previous estimation of the delay introduced by the analog chains at the input and output of the amplifier. This paper presents a coarse time delay estimation (TDE) algorithm to within one-sample resolution based on a conditional entropy principle. The delay estimate is computed via one dimensional search on the proposed TDE spectrum.

1. INTRODUCTION

The advent of non constant amplitude modulations for broadcasting applications such as OFDM in Terrestrial Digital Video Broadcasting (DVB-T) has placed stringent requirements on the linearity of High Power Amplifiers (HPA). Pre-correction systems[1] have been extensively studied in the literature ([2],[3],[4]) to reduce the intermodulation distortion to acceptable levels according to specifications set by the standards. The block diagram of such a system may be observed in the figure, where pre-distortion coefficients are usually determined via optimization of an input/output cost function. Processing is performed in base-band by observing the HPA output.

The analog chains for up- and down-conversion introduce delays which may well be in the order of several tens of samples at the sampling frequency of the pre-correcction system. The main contribution to this delay corresponds to the group delay of the analog filters in the frequency shifting blocks. Precise compensation of the amplifier non-linearity requires that the input base-band signal to the amplifier and its correponding base-band output be time aligned, as the error signal to drive the adaptive algorithm is to be derived from these two signals. Although in principle this time-alignment is not necessary for the compensation of AM/AM distortion, it is mandatory for the compensation of AM/PM distortion.

This paper will primarily consider robust coarse time delay estimation (TDE) of the analog input-output delay of an amplification chain, regardless of the non-linear characteristic of the amplifier. Mutual Entropy (or Conditional Entropy) based criteria are considered for this estimation. Fine time alignment is performed jointly with the adaptation of the pre-distortion coefficients in a later stage.

[1]This work has been supported by projects TIC2000-1025, TIC99-0849 (Spanish Science and Technology Commission),
CIRIT 2000SGR-00083 (Catalan Government) and FIT-070000-2000-649 (Spanish National Programme on Information and Communication Technologies).

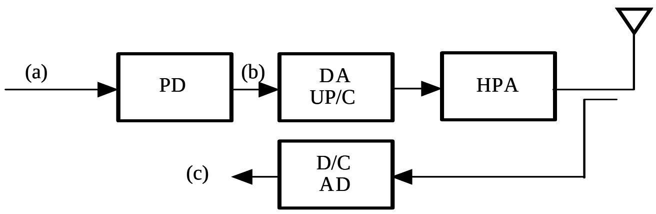

Figure 1: Block diagram of HPA pre-distortion. The input baseband signal (a) is passed through the adaptive predistorter PD, its output is digital to analog converted and frequency shifted to the HPA input. The HPA output is converted back to a digital baseband signal (c). The three signals (a) to (c) are used for adaptive pre-distortion and coarse dealy estimation between (b) and (c). The filters in DA+UP/C and D/C+AD are mostly responsible for the time misalignment.

2. SIGNAL MODEL

AM/AM and AM/PM distortion of an HPA are defined in terms of a non-linear gain on the input signal. Let u_x and u_y be the modulus of the equivalent base-band signals b_x and b_y corresponding to the input and output of the HPA, respectively. Non-linear distortion is expressed in terms of,

$$b_y(t) = g(u_x(t))e^{j\Psi(u_x(t))}b_x(t)$$

with $g(u_x)$ the real non-linear gain associated with AM/AM distortion and $\Psi(u_x)$ the modulus dependent phase shift associated with AM/PM distortion. When the time delay is included in the model and assuming residual aliasing in the discrete domain (non-linear distortion is much more important an effect), we can define,

$$b_y(nT_s + \Delta) \quad = \quad b_y(nT_s - (-\Delta))$$

$$\begin{aligned} b_y^{-\Delta}(nT_s) &= g(u_x(nT_s))e^{j\Psi(u_x(nT_s))}b_x(nT_s) \\ &= \gamma(u_x(nT_s))b_x(nT_s) \end{aligned} \quad (1)$$

As all signals in (1) have undergone no propagation, the usual assumption in pre-distortion schemes is to deem noise negligible as compared with non-linear effects. Hence, our purpose is the derivation of a coarse estimate for Δ down to sampling period resolution. The inverse non-linear characteristic is expressed in its turn as,

$$\begin{aligned} b_x(nT_s - \Delta) &= \\ b_x^{\Delta}(nT_s) &= g'(u_y(nT_s))e^{j\Psi'(u_y(nT_s))}b_y(nT_s) \\ &= \gamma'(u_y(nT_s))b_y(nT_s) \end{aligned} \quad (2)$$

3. CONDITIONAL ENTROPY TDE

The objective of Conditional Entropy TDE is to find that relative delay for which the mutual information between input and output is maximized, or for which the conditional entropy between input and output is minimized. We will propose a simplified version of these principles for reasons of computational complexity. From the model previously exposed (2) we obtain,

$$\begin{aligned} b_y^*(nT_s)b_x^{\Delta}(nT_s) &= \gamma'(u_y(nT_s))\,|b_y(nT_s)|^2 \\ &= u_y^2(nT_s)\gamma'(u_y(nT_s)) \end{aligned} \quad (3)$$

But the inverse non-linear gain is unknown and the estimate of Δ, $\widehat{\Delta}$, must be obtained without aprioristical knowledge of γ'. Nevertherless, it is clear from the previous expression that for the true delay Δ, the product $b_y^*(nT_s)b_x^{\Delta}(nT_s)$ is solely dependent on u_y. Hence, we can construct an optimization criterion based on this functional dependence in the following way: let $\xi(.)$ be a suitable function of a complex random variable, where suitability is to be defined later. Then, we derive the estimate as the maximization of the mutual information between $\xi(b_y^*b_x^{\Delta})$ and u_y,

$$\widehat{\Delta}_\xi = \arg\max_{\widehat{\Delta}} \mathrm{I}\left(\xi(b_y^*b_x^{\widehat{\Delta}})\ ;\ u_y\right)$$

so that whenever $\widehat{\Delta}$ does not coincide with the true delay Δ, randomness extraneous to the random variable u_y is equivalent to a loss of mutual information. For this to be a valid criterion, we require the following condition on the function $\xi(.)$,

$$\mathrm{I}\left(\xi(b_y^*b_x^{\widehat{\Delta}})\ ;\ u_y\right) \le \mathrm{I}\left(\xi(b_y^*b_x^{\Delta})\ ;\ u_y\right) \quad (4)$$

which is guaranteed when the correspondence between u_y and $\xi(b_y^*b_x^{\Delta})$ is one-to-one. The choice of the function $\xi(\cdot)$ will solely depend on performance versus complexity trade-offs in the estimation of the delay. Observe also that the function $u_y^2\gamma'(u_y)$ in (3) is invertible as $\gamma'(u_y)$ is, too. Hence, the following two criteria are equivalent,

$$\widehat{\Delta}_1 = \arg\max_{\widehat{\Delta}} \mathrm{I}(\xi(b_y^*b_x^{\widehat{\Delta}}); u_y) \quad (5)$$

$$\widehat{\Delta}_2 = \arg\min_{\widehat{\Delta}} \mathrm{H}(u_y|\xi(b_y^*b_x^{\widehat{\Delta}})) \quad (6)$$

where $\widehat{\Delta}_2$ is obtained via minimization of the entropy of u_y conditioned on $\xi(b_y^*b_x^{\widehat{\Delta}})$. The equivalence between both criteria is exemplified by the expression,

$$I(\xi(b_y^*b_x^{\widehat{\Delta}}); u_y) = H(u_y) - H(u_y|\xi(b_y^*b_x^{\widehat{\Delta}})) \quad (7)$$

so that the independence of $H(u_y)$ on Δ establishes the equivalence. Both methods do therefore require some kind of probability density function (pdf) or histogram estimation. The advantage is that only a coarse estimate is necessary so that coarse histogram estimates shall suffice. We will focus on the $\widehat{\Delta}_2$ estimate. For ξ the identity function, a suboptimum approach to the method above consists of performing coarse quantization of the random variables u_y and $b_y^*b_x^{\widehat{\Delta}}$ and evaluating the corresponding discrete histograms. Given that the product $b_y^*b_x^{\widehat{\Delta}}$ is complex, the number of necessary histogram bins is quadratic. We prefer to seek a simpler approach. If this product is expanded, we get,

$$\begin{aligned} b_y^*b_x^{\widehat{\Delta}} &= u_y u_x^{\widehat{\Delta}} \exp[j(\theta_x^{\widehat{\Delta}} - \theta_y)] \\ &= u_y u_x^{\widehat{\Delta}} \exp[j(\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta} - \Psi(u_x^{\Delta})] \\ &= u_y u_x^{\widehat{\Delta}} \exp[-j\Psi(u_x^{\Delta})] \exp[j(\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta})] \end{aligned} \quad (8)$$

But from (2) we know that,

$$u_x^{\Delta} = g'(u_y)u_y \ \rightarrow\ u_x^{\widehat{\Delta}} = g'(u_y^{\widehat{\Delta}-\Delta})u_y^{\widehat{\Delta}-\Delta}$$

and therefore (8) becomes,

$$\begin{aligned} b_y^*b_x^{\widehat{\Delta}} &= \Big(u_y g'(u_y^{\widehat{\Delta}-\Delta})u_y^{\widehat{\Delta}-\Delta} \exp[-j\Psi(g'(u_y)u_y)]\Big) \cdot \\ &\quad \cdot \exp[j(\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta})] \end{aligned} \quad (9)$$

with θ_x^{Δ} and $\theta_x^{\widehat{\Delta}}$ random variables independent of u_x^{Δ} and hence of u_y for the type of modulations we are considering (OFDM and M-QAM). In expression (9), modulus-dependent and phase-dependent variables have been factored apart and constitute independent random variables. One possible choice for the $\xi(b_y^*b_x^{\widehat{\Delta}})$ function is the argument of $b_y^*b_x^{\widehat{\Delta}}$, hence,

$$\xi(b_y^*b_x^{\widehat{\Delta}}) = \arg(b_y^*b_x^{\widehat{\Delta}}) = -\Psi(g'(u_y)u_y) + \left(\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta}\right) \quad (10)$$

with $-\Psi(g'(u_y)u_y)$ a constant term independent of $\widehat{\Delta}$, and $\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta}$ a perturbation term independent of u_x^{Δ}. Hence, when the difference $\Delta - \widehat{\Delta}$ is sufficiently large, the distribution of $\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta}$ is uniform (for OFDM modulations) and as a consequence, so is the distribution of $\arg(b_y^*b_x^{\widehat{\Delta}})$, and independent of u_y. Only when the difference $\Delta - \widehat{\Delta}$ is sufficiently small, we get that $\xi(b_y^*b_x^{\widehat{\Delta}}) \simeq -\Psi(g'(u_y)u_y)$ which translates to statistical dependence between the phase of $b_y^*b_x^{\widehat{\Delta}}$ and the modulus u_y.

The preference of choosing $\xi(\cdot)$ as the argument of $b_y^*b_x^{\widehat{\Delta}}$ instead of its modulus is because u_y appears as a multiplicative factor in (9),

$$|b_y^*b_x^{\widehat{\Delta}}| = u_y g'(u_y^{\widehat{\Delta}-\Delta})u_y^{\widehat{\Delta}-\Delta} \ge 0$$

and statistical dependence between u_y and this modulus is present whatever the difference $\widehat{\Delta} - \Delta$.

The criteria in (5) and (6) fail when the one-to-one correspondence between $\xi(b_y^* b_x^{\widehat{\Delta}})$ and u_y is not fulfiled. Such is the case in the absence of AM/PM distortion, $\Psi(u) = 0$, where equation (10) becomes,

$$\xi(b_y^* b_x^{\widehat{\Delta}}) = -\Psi(g'(u_y)u_y) + \left(\theta_x^{\widehat{\Delta}} - \theta_x^{\Delta}\right) = \theta_x^{\widehat{\Delta}} - \theta_x^{\Delta}$$

Note that in this case, the condition required in (4) does not hold. Then, the alternative conditional entropy criterion to be applied is as follows,

$$\widehat{\Delta}_3 = \arg\min_{\widehat{\Delta}} \mathrm{H}(\xi(b_y^* b_x^{\widehat{\Delta}})|u_y)$$

where $\xi(b_y^* b_x^{\Delta}) = \theta_x^{\Delta} - \theta_x^{\Delta} = 0$ and $\mathrm{H}(0|u_y) = 0$ guarantees that even in the absence of AM/PM distortion, $\widehat{\Delta}_3 = \Delta$.

4. CONDITIONAL ENTROPY ESTIMATION

Conditional entropy is estimated in terms of the data histogram and carried out in two different ways: hard histogram and soft histogram [5]. The latter provides better results for the same number of samples and is usually preferred for short data records. The hard histogram is based on rectangular activation functions $\Pi(\cdot)$, so that the probability that samples of the random variable X are contained in a given bin is estimated as,

$$\widehat{p}_i = \frac{1}{N}\Sigma_{j=1}^{N}\Pi\left(\frac{x_j - c_i}{\delta}\right) \tag{11}$$

with N the number of samples of the data record, c_i the bin centroids in the estimation of the probability and δ the bin width. In soft histogram estimation, the activation functions are triangular and overlapping, so that the ocurrence of a given sample contributes to two bins as follows,

$$\widehat{p}_i = \Sigma_{i=1}^{B}\frac{1}{N}\Sigma_{j=1}^{N}\Lambda\left(\frac{x_j - c_i}{\delta}\right) \tag{12a}$$

with $\Lambda\left(\frac{x}{\delta}\right)$ expressed in terms of convolution as,

$$\Lambda\left(\frac{x}{\delta}\right) = \Pi\left(\frac{x}{\delta/2}\right) * \Pi\left(\frac{x}{\delta/2}\right)$$

Although the summation is taken over the number of bins B, for each sample x_j only two triangular functions are activated. For either histogram and associated bin centers c_i, data are quantized as,

$$q(x) = \Sigma_{i=1}^{B} c_i \Pi\left(\frac{x - c_i}{\delta}\right)$$

Conditional entropy is then estimated in the following way: let us construct the $N_\xi \times N_u$ matrix $\mathbf{A} = [a_{p,q}]$ over a data record of N samples where N_ξ and N_u are, respectively, the number of bins into which the data $\xi(b_y^* b_x^{\widehat{\Delta}})$ and the modulus u_y are quantized. Each component is defined as,

$$a_{p,q} = \widehat{p}(u_p, \xi_q)$$

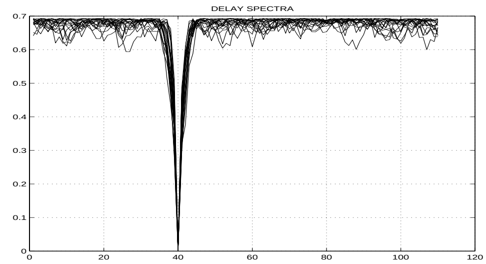

Figure 2: Conditional Entropy TDE Spectrum obtained for a 2×2 grid in the (u_y, ξ) plane. Twenty realizations of 128 samples each are shown. The oversampling factor is 5. Saleh's model has been used.

with ξ_q and u_p the centroids corresponding to the quantization of $b_y^* b_x^{\widehat{\Delta}}$ and of u_y. The joint pdf $\widehat{p}(u_p, \xi_q)$ is estimated according to,

$$\widehat{p}(u_p, \xi_q) = \widehat{p}(u_p)\widehat{p}(\xi_q|u_p)$$

where $\widehat{p}(u_p)$ and $\widehat{p}(\xi_q|u_p)$ are obtained from either (11) or (12a). The conditional entropy estimate becomes then,

$$\begin{aligned}\widehat{H}(\xi(b_y^* b_x^{\widehat{\Delta}})|u_y) &= -\Sigma_{p=1,q=1}^{N_u,N_\xi}\widehat{p}(u_p,\xi_q)\log\widehat{p}(\xi_q|u_p) \\ &= -\Sigma_{p=1}^{N_u}\widehat{p}(u_p)\Sigma_{q=1}^{N_\xi}\widehat{p}(\xi_q|u_p)\log\widehat{p}(\xi_q|u_p)\end{aligned}$$

The search for the optimum $\widehat{\Delta}$ will require minimization of $\widehat{H}(\xi(b_y^* b_x^{\widehat{\Delta}})|u_y)$ over $\widehat{\Delta}$. The search time will depend on the bin resolution and the number of samples. Nevertheless, the advantage is that the non-linearity is not time-varying and many samples can be used to obtain a coarse delay estimation. Fine delay estimation is then carried out jointly with the estimation of pre-distortion coefficients. The number of bins need not be very high as only a coarse estimation is needed.

The performance of this scheme is evaluated in terms of the probability that the estimated conditional entropy provides a coarse minimum equal to that of the true entropy. When the data record increases, this probability goes asymptotically to one (law of large numbers).

5. SIMULATIONS

Saleh's model [1] for TWT (Travelling Wave Tube) amplifiers has been used in determining the performance of the algorithm. This model is defined in terms of the following AM/AM and AM/PM distortion curves,

$$\begin{aligned}g(u_x) &= KA_{\mathrm{sat}}^2\frac{u_x}{u_x^2 + A_{\mathrm{sat}}^2} \\ \Psi(u_x) &= \frac{\pi}{3}\frac{u_x^2}{u_x^2 + A_{\mathrm{sat}}^2}\end{aligned}$$

The Conditional Entropy TDE spectrum has been obtained for a number of realizations on the same amplifier. The minimum is always obtained to within a one-sample resolution,

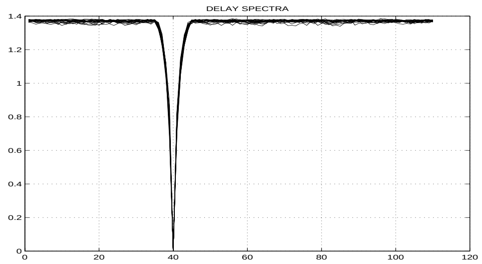

Figure 3: Conditional Entropy TDE Spectrum obtained for a 4×4 grid in the (u_y, ξ) plane. Twenty realizations of 512 samples each are shown. The oversampling factor is 5. Saleh's model has been used. The quality of the notch guarantees that the algorithm will be capable of determining a reliable time delay estimate on a low number of samples.

where the true delay has been set to $\Delta = 40$. The discrimination is quite good, a distinct peak at the minimum always appears. Results are shown for several bin resolutions. The complexity required of this algorithm is not very high as a number of bins as low as a 2x2 grid suffices to obtain very reliable time delay estimations. In order to guarantee that each bin in the (u_y, ξ) grid activates a sufficient number of times, conditional entropy is evaluated on a data record thirty-two times the grid size (assuming equiprobability, each bin would activate 32 times).

Results in figure (5) for high oversampling shows the sensitivity of the algorithm at low (2x2) grid resolutions. Differences can be observed over the 20 realizations.

6. REFERENCES

[1] A.A.M. Saleh, "Frequency-independent and frequency-dependent non-linear models for TWT amplifiers", IEEE Trans. on Communications, vol COM-29, no. 11, pp. 1715-1720, November 1981.

[2] M. Kuipers, R. Prasad, "Pre-distorted amplifiers for OFDM in Wireless Indoor Multimedia Communications", IEEE Vehicular Technology Conference (VTC2000), pp.395-399.

[3] M. Faulkner, M. Johansson,"Adaptive Linearization using Pre-distortion: Experimental Results", IEEE Trans. on Vehicular Technology, VOL. VT-43, no.2, pp. 323-332, May 1994.

[4] J. Sala, H. Durney, "Unconditionally Convergent Pre-Distortion of Non-Linear High Power Amplifiers", IEEE VTS 53rd Vehicular Technology Conference (VTC-Spring 2001), ISBN 0-7803-6730-8 (CD-ROM).

[5] H. Chongfu, "Reliability of Fuzzy Risk Assessment by Information Distribution", 19th International Conference of the North American Fuzzy Information Processing Society, NAFIPS 2000, pp. 278-282, ISBN 0-7803-6274-8.

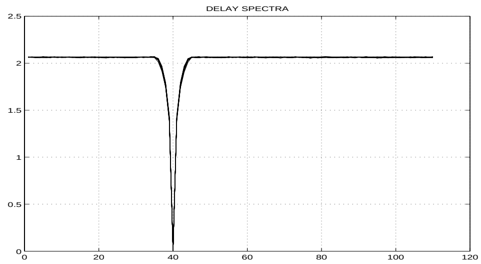

Figure 4: Conditional Entropy TDE Spectrum obtained for a 8×8 grid in the (u_y, ξ) plane. Twenty realizations of 2048 samples each are shown. The oversampling factor is 5. Saleh's model has been used. Observe that for such grid resolutions, the TDE spectrum is practically realization-independent. Discrimination is very good as shown by the steepness of the notch. This feature is preserved for higher oversampling factors (see next figure). Good results are still obtained when the data block size to grid size ratio is decreased below 32. The grid margins must be made dependent on the signal dynamic range for best performance in the quantization of u_y and ξ. Otherwise, the sharpness of the notch is degraded and a flat minimum may appear due to poor quantization of u_y and ξ.

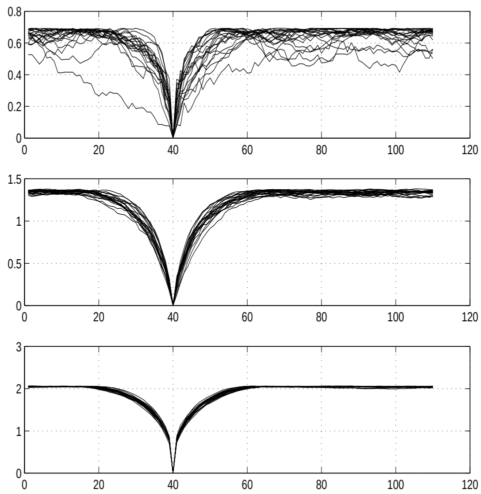

Figure 5: This figures shows the Conditional Entropy TDE spectra for an oversampling of 25 for the following grid resolutions: 2x2 (upper), 4x4 (middle) and 8x8 (lower). It can be appreciated that the notch width is sensitive to the signal autocorrelation (oversampling factor) and that the spectra become less realization-independent for large oversampling factors.