Competitive Linearity for Envelope Tracking: Dual-Band Crest Factor Reduction and 2D-Vector-Switched Digital Predistortion

Harald Enzinger, Karl Freiberger, and Christian Vogel

Abstract—This is the preprint version of the paper describing the winning solution of the 2017 IMS student design competition "PA linearization through DPD", published in IEEE Microwave Magazine, volume 19, issue 1, pages 69-77, Jan.-Feb. 2018.

As wireless communication standards evolve to support ever-higher data rates, the required linearity and bandwidth must increase, which leads to higher energy consumption [1]. The problem of high energy consumption has become more serious due to the wide-spread adoption of orthogonal frequency division multiplexing (OFDM), which is used in LTE and WiFi applications [2]. The low energy efficiency of OFDM-based systems results from the statistical distribution of OFDM signals, combined with the efficiency characteristics of radio frequency (RF) power amplifiers (PAs). OFDM signals spend most of their time at low magnitudes, where the energy efficiency of PAs is low; however, they also exhibit peaks that must be linearly amplified as well. A detailed analysis of the joint linearity-efficiency behavior of RF PAs, excited by single-carrier and OFDM signals can be found in [3].

To reduce the energy consumption of wireless transmitters, several highly efficient PA architectures have been developed [4]. For base stations, the Doherty PA dominates [5], while, for handset devices, envelope tracking is now widely used [6]. The idea behind envelope tracking is to use a time-varying supply that tracks the instantaneous envelope of the communication signal. An important advantage of envelope tracking is that the supply modulator can be designed independently of the carrier frequency [6]. This makes envelope tracking very attractive for multi-band applications, where signals at different frequencies are sent through the same PA.

A common characteristic of many high-efficiency PAs is their reduced linearity with respect to conventional PAs. Consequently, linearization is often required, with digital predistortion (DPD) being one of the most used techniques. The importance of DPD as a microwave technique is illustrated by the fourth edition of the student design competition "PA Linearization through DPD", sponsored by IEEE Microwave Theory and Techniques Society (MTT-S) Technical Coordinating Committees 9 and 11, which took place at the 2017 IEEE MTT-S International Microwave Symposium (IMS) in Honolulu, Hawaii, last June.

Harald Enzinger (enzinger@tugraz.at) and Karl Freiberger (freiberger@tugraz.at) are with the Signal Processing and Speech Communication Laboratory at Graz University of Technology, Austria. Christian Vogel (c.vogel@ieee.org) is with FH JOANNEUM, University of Applied Sciences, Austria and the Signal Processing and Speech Communication Laboratory at Graz University of Technology, Austria.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Score</th>
<th>Target</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ACPR target is achieved.</td>
<td></td>
<td>ACPR</td>
<td>—</td>
</tr>
<tr>
<td>(2) NMSE target is achieved.</td>
<td></td>
<td>NMSE</td>
<td>(1)</td>
</tr>
<tr>
<td>(3) Power difference between bands is within ±0.5 dBm.</td>
<td></td>
<td>Power</td>
<td>(2) (3)</td>
</tr>
<tr>
<td>20 dBm</td>
<td>10 points/dBm (+)</td>
<td>Efficiency</td>
<td>—</td>
</tr>
</tbody>
</table>

TABLE I: An overview of the score computation.
Fig. 1: The measurement setup of the 2017 DPD competition. See also www.dpdcompetition.com.

Fig. 2: The dual-band OFDM signal: (a) power spectral density (b) time domain waveforms.

Scoring was based on four criteria: adjacent channel power ratio (ACPR), normalized mean square error (NMSE), output power, and drain efficiency. As Table I shows, output power had a strong influence on the score. The goal was clearly to maximize the output power, while achieving the targets for ACPR and NMSE. Failing in either ACPR or NMSE immediately leads to a score approaching zero.

**SYSTEM ARCHITECTURE**

Since the task was to linearize the PA in concurrent dual-band operation, we used a concurrent dual-band linearization architecture [7]. In such an architecture, the dual-band signal is split into two single-band signals, which are processed jointly at a lower sampling rate. This renders the processing independent of the center frequencies of the two bands.

To divide the dual-band signal into two single-band signals, we transformed the dual-band signal to the frequency domain, applied frequency translation, filtering and decimation, and converted the individual bands back to the time domain. To recombine the single-band signals, we used a similar method. The sampling rate of the single-band signals was 80 MHz.

An overview of our system architecture is shown in Fig. 3. We follow the common practice of separating the signal processing before the PA into crest factor reduction (CFR) and DPD. The purpose of CFR is to limit the peaks of the source signal, such that the operating range of the DPD is clearly defined. Within this range, the DPD ideally implements the inverse PA transfer characteristic, such that the PA output signal is a scaled replica of the CFR signal.

In our system architecture, the supply signal is derived from the CFR signal. We chose the CFR signal as the basis for the supply signal because the supply signal should track the instantaneous envelope of the PA output signal, which, after the DPD is operating, is very similar to the CFR signal.
Initially, we also experimented with a three-dimensional (3D) DPD architecture, where the supply signal is used as a third input to the DPD [8]. However, in the present setup, the 3D-DPD did not improve the linearization performance.

**SUPPLY SIGNAL GENERATION**

The supply signal generation was specified by the competition organizers and involved a two-step process. First, a slow envelope of the dual-band signal is computed from the magnitudes of the single-band signals by

\[ u_{env}[n] = |u_1[n]| + |u_2[n]|. \]  

(1)

Fig. 2 (b) shows that this slow envelope follows the peaks of the dual-band envelope. Second, the supply signal is derived from the slow envelope by the memoryless shaping function

\[ z[n] = \left( \frac{s_{env}^6}{z_{\text{min}}^6 + s_{\text{env,norm}}^6[n]} \right)^{1/6}, \]  

(2)

where \( z_{\text{min}} \) is the desired minimum value of the supply signal and \( s_{\text{env,norm}}[n] \) is the normalized slow envelope, obtained by dividing \( u_{env}[n] \) by its maximum. As Fig. 2 (b) shows, the supply signal tracks the slow envelope, but remains above \( z_{\text{min}} = 0.2 \) at all times. Other methods for generating the dual-band envelope tracking supply signal are summarized in [9].

An important detail concerning the supply signal is that it reaches the PA over a different path than the RF input signal. Any delay mismatch between these paths deteriorates the linearity. To equalize the delay mismatch, we time-shifted the supply signal such that the ACPR was optimized.

**CREST FACTOR REDUCTION**

The many available CFR methods [10] can be divided into two types: 1) methods that add a certain amount of tolerable distortion and 2) methods that do not add distortion, but modify the signal generation. For the DPD competition, the signal generation was fixed, so we had to use a CFR method of the first type. Most CFR methods of this type use some sort of clipping and filtering to trade-off more inband distortion for less out-of-band distortion. In the variant proposed in [11], which was also used by previous winners of the DPD competition [12], [13], the filtering is applied on the error signal rather than on the clipped signal. In this way the desired signal is not affected by the filtering, and, consequently, we had more flexibility for designing the error-shaping filter.

For the DPD competition we adapted the method from [11] to the dual-band case [14], which lead us to the CFR architecture in Fig. 4. We apply clipping on the single-band signals, but with the objective of reducing the crest factor of the dual-band signal. For this purpose, we estimate the peak envelope of the dual-band signal from the single-band signals by

\[ s_{env}[n] = |s_1[n]| + |s_2[n]|. \]  

(3)

If the estimated peak envelope \( s_{env}[n] \) is larger than the desired maximum \( u_{\text{max}} \), the magnitude of the single-band signals must be reduced. This magnitude reduction is implemented by the multipliers in Fig. 4, which apply the time-varying gain

\[ g[n] = \begin{cases} 1, & s_{env}[n] > u_{\text{max}}, \\ u_{\text{max}} / s_{env}[n], & \text{otherwise}. \end{cases} \]  

(4)

The resulting clipping distortion is very broadband and affects the inband and out-of-band performance nearly equally. However, practical targets for the out-of-band performance, measured by ACPR, are typically much stricter than those for the in-band performance, measured by NMSE, error vector magnitude [15], noise power ratio [16], or error power ratio [17]. To account for this asymmetry, we use an error shaping filter to trade-off more inband distortion for less out-of-band distortion. In the error shaping filter, the clipping error is computed by subtracting the clipped signal from the original signal. Afterwards, the clipping error is filtered by a lowpass filter to reduce its effect on the ACPR. Finally, the output signals are obtained by subtracting the filtered error signals from delayed versions of the original signals.

Due to the error shaping filter, a certain amount of peak-regrowth occurs. To obtain a dual-band signal with a clearly defined maximum magnitude, we added another dual-band clipper after the error shaping filter. With the clipping threshold of the second clipper being equal to the first one, the distortion produced by the second clipper could be tolerated without filtering.

To set the clipping threshold, we simulated CFR with 100 realizations of the input signal and different clipping thresholds. These simulations showed that the ACPR and NMSE targets of TABLE I can be achieved with sufficient margin by using a clipping threshold that produces a crest factor of 8.6 dB.
**Structure of the Digital Predistorter**

Linearity with respect to the coefficients is an important property of many DPD structures. This property allows the use of efficient training algorithms, which are based on solving systems of linear equations. For concurrent dual-band systems, a model that is linear in the coefficients is given by

\[
x_1[n] = \sum_{i=1}^{I_1} c_{1,i} \Phi_{1,i} \{ u_1[n], u_2[n] \},
\]

\[
x_2[n] = \sum_{i=1}^{I_2} c_{2,i} \Phi_{2,i} \{ u_1[n], u_2[n] \},
\]

where \(c_{1,i}\) and \(c_{2,i}\) are coefficients, \(\Phi_{1,i}\) and \(\Phi_{2,i}\) are basis functionals, \(i\) enumerates the basis functionals, and \(I_1\) and \(I_2\) are the numbers of coefficients and basis functionals per frequency band. For the DPD competition, we used basis functionals of the form

\[
\Phi_{b,i} \{ u_1[n], u_2[n] \} = u_b[n - m_0] |u_1[n - m_1]|^{p_1} |u_2[n - m_2]|^{p_2},
\]

which are inspired by the generalized memory polynomial (GMP) [18]. With this type of basis functionals, the top-level DPD structure can be drawn as in Fig. 5 (a), where "abs" represents the absolute value operation.

To further specify our model, we divided the basis functionals into four groups: linear basis functionals where both \(p_1\) and \(p_2\) are zero, intra-band and cross-band basis functionals where either \(p_1\) or \(p_2\) is zero, and mixed-band basis functionals where both \(p_1\) and \(p_2\) are larger than zero. Using this classification, the model for band one can be drawn like in Fig. 5 (b). In the following discussion, we specify the behavior of each of the blocks in Fig. 5 (b) in terms of equations. The model for band two has the same structure as the model for band one and is obtained by swapping the band indices of all variables.

The linear block in Fig. 5 (b) contains a linear finite impulse response filter given by

\[
x_1^{(\text{linear})}[n] = \sum_{m=0}^{M} c_m u_1[n - m].
\]
two-dimensional (2-D) memory polynomial [19], given by

\[ x_{m}^{(\text{mixed})}[n] = \sum_{m=0}^{M} \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} c_{m,p_1,p_2} \times u_1[n-m] \times |u_1[n-m]|^{p_1} \times |u_2[n-m]|^{p_2}. \]

Again, we use summation limits that depend on the indices of previous sums to heavily prune the model.

In all models, we use both odd-order and even-order terms. Many authors exclude even-order terms from baseband models [20] because conventional even-order terms in passband models produce no output near the carrier frequency [21]. However, in [22], [23] we have analytically shown that even-order terms in baseband models correspond to modified, odd-symmetric, even-order terms in passband models that do produce output near the carrier frequency. Including even-order terms in baseband models helps to keep the nonlinear order low, which improves the numerical properties of the model [23]. Results obtained by automatic pruning also indicate that odd-order and even-order terms contribute approximately equally to the modeling accuracy [24].

To further increase the modeling accuracy, several authors have proposed piecewise models [25]–[29]. These models circumvent the problems associated with high-order basis functionals by using low-order basis functionals with higher locality. Most of them are based on splines, which enforce a certain level of continuity between the segments. An exception is the vector-switched model [26] that does not enforce continuity. Dropping the continuity constraint has two advantages: 1) it simplifies the model formulation since any model can be used for the individual segments, and 2) it improves the numerical properties since the support of each basis functional is confined to the modeling region of the respective segment. Potential problems due to discontinuities between the segments have not been observed when the model is trained by least squares fitting with long signal vectors [26].

For the DPD competition, we adapted the vector-switched technique to the dual-band case, which lead us to the 2D-vector-switched model in Fig. 7. Each of the regional models in Fig. 7 (a) contains the dual-band DPD structure discussed above. The switching between the regional models is based on the magnitude of the current sample in each of the frequency bands. For mapping sample magnitudes to modeling regions, we used a vector quantizer with manually selected centroids as shown in Fig. 7 (b). In the overall model, we included eight identical regional models, each having 72 coefficients per frequency band. By using identical regional models, the vector-switched model can be implemented efficiently by a single regional model with switched coefficients. The identification complexity grows linearly with the number of regional models. Like in any other polynomial or Volterra-based model, Horner’s method can be used to efficiently compute the basis signals required for filtering and training [30].

**TRAINING OF THE DIGITAL PREDISTRIBUTOR**

Two architectures are commonly used to train digital predistorters: indirect learning and direct learning [31]. Both originate from earlier work on adaptive control [32]. To give an overview of these architectures, we introduce the signal vectors \( \mathbf{u}, \mathbf{x}, \) and \( \mathbf{y} \), containing the CFR signal, the PA input signal, and the delay-, phase-, and gain-corrected PA output signal of one frequency band. Similarly, we introduce the matrices \( \mathbf{U}, \mathbf{X}, \) and \( \mathbf{Y} \), which contain columns of basis signals obtained by sending the respective signals of both frequency bands through the basis functionals of one frequency band model.

The training is applied for each frequency band and for each modeling region of the 2D-vector-switched model. For training the regional models, only the samples belonging to the respective region are used. In the following, we discuss the training for one frequency band and one modeling region.

A model of the PA is given by the system of linear equations

\[ \mathbf{y} = \mathbf{X} \mathbf{c}_{PA}, \]

where \( \mathbf{c}_{PA} \) contains the PA coefficients. An postinverse model of the PA is given by

\[ \mathbf{x} = \mathbf{Y} \mathbf{c}_{DPD}, \]

where \( \mathbf{c}_{DPD} \) contains the postinverse PA coefficients.
In the indirect learning architecture, (12) is solved for \( c_{\text{DPD}} \) and the result is used in the DPD, which implements

\[
x_{\text{DPD}} = U c_{\text{DPD}}.
\]

A serious drawback of indirect learning is that solving (12) finds the optimal postdistorter, which is in general different from the optimal predistorter. The goal of direct learning is to find a better solution by directly optimizing the predistorter. This can be done by inverting the PA model in (11) using iterative methods [33], or by closed-loop adaptation based on repeated measurements [34]. Direct learning is in general more difficult than indirect learning since the cascade of the DPD and the PA is not linear in the DPD coefficients. The objective for direct learning can be either to minimize the error signal

\[
e = u - y
\]

or to optimize standard-relevant performance metrics [35]. Adaptive algorithms for minimizing (14) either include a nonlinear model of the PA [36], or they approximate the PA by a complex gain [37]. The latter leads to a rather simple algorithm which practically converges to better solutions than indirect learning [38]. This algorithm can be formulated as follows. Estimate the error in the DPD coefficients by solving

\[
e = U c_{\text{error}}
\]

for \( c_{\text{error}} \) and update the DPD coefficients by

\[
c^{(i+1)}_{\text{DPD}} = c^{(i)}_{\text{DPD}} + \mu c^{(i)}_{\text{error}},
\]

where \( i \) is the iteration index and \( \mu \) is the step size.

For the DPD competition, we initialized the predistorter with indirect learning, followed by several iterations of the simple direct learning algorithm described above. For the delay and gain correction, we used the methods described in [39]. By using the root-mean-square gain estimate from the measurement without the DPD in all iterations, the output power with the DPD converged to the same value as in the first measurement without the DPD. The phase correction was based on a least squares estimate of the complex gain.

**Performance Evaluation**

At the competition, each team had 20 minutes for training. After this period, a new realization of the input signal was used to evaluate the score. Our solution achieved a score of 71.8 points, corresponding to an ACPR of -49.2 dB, an NMSE of -35.7 dB, an output power of 24.4 dBm, and a drain efficiency of 22.3%. Higher drain efficiency, like in [8], can be reached with a dynamic bias, as discussed in [40].

Plots of the linearization performance are shown in Fig. 8. From Fig. 8 (a), we see that after ten iterations of training, the DPD reduces the out-of-band distortion to levels close to the CFR-induced distortion. The scatter plot in Fig. 8 (b) shows the amplitude modulation to amplitude modulation (AM-AM) and amplitude modulation to phase modulation (AM-PM) for the lower frequency band, before and after linearization.

**Conclusions**

The DPD competition is an excellent opportunity for young researchers to apply their digital signal processing skills on a realistic measurement setup. The direct comparison of solutions also facilitates a benchmarking of research results, which is otherwise very hard to achieve. The main ingredients for winning the 2017 DPD competition were:

- a dual-band CFR based on clipping and filtering
- a dual-band DPD based on a modified GMP
- the 2D-vector-switched technique
- direct learning with several iterations.

These ingredients combine several state-of-the-art methods for CFR and DPD with new extensions we have developed specifically for the 2017 DPD competition.

**Acknowledgments**

We thank Prof. Pere L. Gilabert, Prof. Gabriel Montoro and Mr. David López-Bueno from Universitat Politècnica de Catalunya (UPC), and Hermann Boss from Rohde & Schwarz for the excellent organization of the 2017 DPD competition. The research leading to these results has received funding from the Austrian Research Promotion Agency FFG under the project numbers 835187 and 4718971.
REFERENCES


