Low-Complexity Feedback Data Compression for Closed-Loop Digital Predistortion

Arne Fischer-Bühner[$#], Lauri Anttila[#], Vishnu Unnikrishnan[#], Manil Dev Gomony[$], Mikko Valkama[#]
[$]Nokia Bell Labs, Belgium
[#]Department of Electrical Engineering, Tampere University, Finland
arne.fischer@nokia.com

***Abstract*— This paper proposes sample combining as a low-complex and effective feedback data compression technique that allows to significantly reduce the computational effort and buffering needs for parameter adaptation in a closed-loop digital predistortion (DPD) system. Compression is achieved by applying an integrate & dump operation to an undersampled feedback signal. The proposed method is experimentally validated for RF measurement based behavioral modeling as well as closed-loop DPD of a 3.5 GHz GaN Doherty PA, taking also quantization effects of the feedback path into account. Our results demonstrate that the proposed technique is as capable as state-of-the-art histogram-based sample selection, however, at a much lower complexity.**

***Keywords*— Behavioral modeling, closed-loop adaptation, digital predistortion (DPD), low-complexity, power amplifier (PA).**

I. Introduction

Power-efficient transmission (TX) is of key importance to any modern wireless communication system. Therein, the power amplifier (PA) is the main contributor to TX power consumption as it suffers from a well-known trade-off between amplification linearity and power efficiency. Digital predistortion (DPD), amongst other techniques, is commonly applied to mitigate nonlinear as well as dynamic distortion due to the amplification, enabling linear transmission at higher power and thus greater efficiency [1]. As part of a DPD system as shown in Fig. 1, the PA output is observed using a dedicated feedback receiver in order to track changes in the PA behavior. The feedback is then used in conjunction with the intended TX to update a DPD model suitable to linearize the PA.

Two obstacles are inherent to the feedback based model estimation. Firstly, in-order to fully capture the distortion effect on wideband TX signals, high oversampling ratios are needed, e.g. $5\times$ of the signal's effective linear bandwidth. Furthermore, due to the large peak to average power ratio (PAPR) of spectrally efficient waveforms, a high bit-resolution is required to represent the feedback with a sufficiently low quantization noise floor, resulting in high analog-to-digital conversion (ADC) energy consumption as well as high data-rates for the signal processing. To alleviate the sampling-rate problem, an undersampled feedback signal can be acquired instead, which is sufficient to extract the DPD model coefficients [2]. Elimination of the linear signal component for data reduction has been discussed as a way to reduce the required bit-resolution [3]. A second issue arises with increasing model size, where the coefficient estimation becomes computationally complex and poses a substantial contribution towards the overall DPD system's complexity and power consumption. Histogram-based sample selection methods have been studied to address the computational cost of the estimation, which rely on selecting a representative subset of feedback samples [4], [5], [6]. While these approaches allow model identification using significantly less samples, the presented methods require to first analyze a large TX sequence or capture a longer feedback sequence in order to extract statistical properties of the signals before a condensed sample set is derived. Alternatively, the statistics can be precomputed, which however renders the proposed methods incompatible with variation of the PA behavior over time and in particular with changes in the transmit signals.

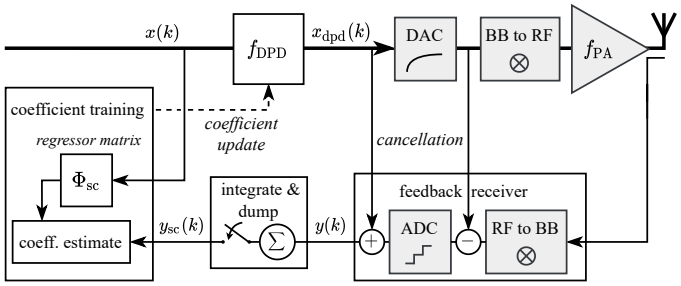

Fig. 1. Illustration of the adaptive DPD system with sample combining realized as an integrate & dump filter in the feedback path.

In this article, we propose a very lightweight, yet effective, compression technique that builds upon the undersampling approach and adds sample combining, as to further condense the sample set and make identification more robust. We show that in the presence of quantization noise, sample combining is as effective as state-of-the-art sample selection methods, but with lower, almost negligible, complexity. Our method lends itself to real-time, throughput-oriented implementation, obviating the necessity to first extract statistics from a large set of samples. Thus, in addition to reducing model estimation complexity, sample buffering needs and chip data-rates are reduced. To assess the feedback reduction trade-offs and implications on the modeling performance, we provide results based on RF measurements for both direct PA modeling and linearization, considering a closed-loop adaptive DPD on a 2-stage Doherty GaN PA running a 100 MHz waveform at 3.5 GHz. The joint impact of feedback quantization and sample count on PA modeling accuracy and DPD performance is reviewed, seeking to minimize the overall data volume required

to identify model parameters and perform DPD adaptation.

The paper is organized as follows. Sec. II details the proposed sample combining method for PA modeling as well as the corresponding closed-loop DPD system. In Sec. III, PA modeling results are shown with respect to quantization and parameter count. In Sec. IV, closed-loop DPD measurement results are presented and compared with a reference sample selection method. Sec. V provides concluding remarks.

II. Proposed Methods

The assumed DPD system with feedback path is depicted in Fig. 1. Therein, $x(k)$ are the complex-valued baseband samples of the desired TX signal, sampled at a frequency f_s, and $y(k)$ are complex-valued samples of PA feedback, with k denoting the respective sample index.

A. General PA Modeling

Disregarding the predistorter at first, the nonlinear PA input to output behavior can be described on a per-sample basis by means of e.g. linear-in-parameter models with

$$\hat{y}(k) = \mathbf{g}^\mathrm{T}\,\phi(x(k), x(k-1), ..., x(k-M)), \tag{1}$$

where $\hat{y}$ is the model output, $\mathbf{g} \in \mathbb{C}^{J\times 1}$ are complex-valued coefficients, and $\phi(k) \in \mathbb{C}^{J\times 1}$ is a vector containing J nonlinear regressors which are dependent on the input samples up to a maximum history M. In this paper, we employ the well-known generalised memory polynomial (GMP) as a regression basis [7]. It is noted that any other linear-in-parameter model could be chosen instead. The coefficients $\mathbf{g}$ can be identified by seeking the least-squares solution of an over-determined system given a sequence $\mathbf{y} = [y(1), y(2), ..., y(L)]^\mathrm{T}$ and respective input regressors $\Phi = [\phi(1), \phi(2), ..., \phi(L)]^\mathrm{T}$, e.g. by applying the Moore-Penrose inverse as

$$\mathbf{g} = (\Phi^\mathrm{H}\Phi)^{-1}\Phi^\mathrm{H}\,\mathbf{y}, \tag{2}$$

where $(\cdot)^\mathrm{H}$ denotes the Hermitian transpose. The length L of the identification sequence is of high importance as a large L will dominate the complexity of computing (2). However, a too short sequence may not be statistically representative e.g. due to correlation amongst samples or quantization noise, and the system may end up under-determined. This relationship is revisited in the experiments in Sec. III.

B. Undersampling and Sample Selection

From (2), it becomes apparent that for determining the coefficients, the temporal relationship of samples in $\mathbf{y}$ is of no further relevance. This is exploited when undersampling the feedback, where only every δ-th sample in $\mathbf{y}$ is used. The sampling rate at the feedback ADC is consequently reduced to $f_\mathrm{us} = f_\mathrm{s}/\delta$. It however remains a requirement that the acquired samples support the bandwidth of the input signal, i.e. the sample and hold times of the ADC must support full sampling rate f_s. Equation (2) can then be used with $\mathbf{y}_\mathrm{us} = [y(\delta), y(2\delta), ..., y(L\delta)]^\mathrm{T}$, and matching $\Phi_\mathrm{us} = [\phi(\delta), \phi(2\delta), ..., \phi(L\delta)]^\mathrm{T}$. Further, undersampling improves the parameter identification. Since consecutive samples typically exhibit strong correlation, selecting only distant samples yields a statistically richer sample set. Consequently, a shorter length L can be chosen, which reduces the complexity of (2). Sample selection, or Mesh selection (MeS) techniques in [4], [5], [6] build upon the same principle, aiming to reduce L by selecting the samples in $\mathbf{y}$ such that statistics of the signals, the model excitation, or the AM-AM characteristic are properly represented in the reduced set.

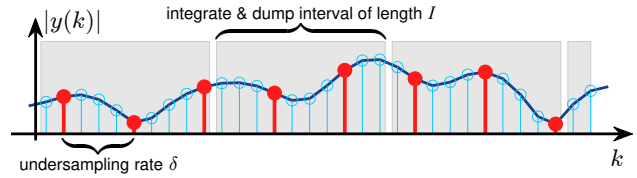

Fig. 2. Illustration of the proposed sample combining scheme. Under-sampled samples within a integration interval are combined through accumulation.

C. Proposed Sample Combining

Extending the undersampling idea, any linear operation on the feedback data is permitted, given that it does not systematically suppress relevant information (e.g. a low-pass), and it can be applied similarly to the respective entries in Φ. To further improve the representativeness of $\mathbf{y}$, we thus propose to combine several undersampled consecutive samples by means of an integrate & dump operation. Thus, undersampled feedback samples and their respective regression entries in Φ are accumulated within a given integration period of length I, as illustrated in Fig. 2. The resulting sample vector $\mathbf{y}_\mathrm{sc}$ and regression matrix Φ_sc are then given as

$$\mathbf{y}_\mathrm{sc} = \begin{bmatrix} \sum\limits_{l\delta\in(0,I]} y(l\delta) \\ \sum\limits_{l\delta\in(I,2I]} y(l\delta) \\ \cdots \\ \sum\limits_{l\delta\in((L-1)I,LI]} y(l\delta) \end{bmatrix}, \quad \mathbf{\Phi}_\mathrm{sc} = \begin{bmatrix} \sum\limits_{l\delta\in(0,I]} \phi(l\delta)^\mathrm{T} \\ \sum\limits_{l\delta\in(I,2I]} \phi(l\delta)^\mathrm{T} \\ \cdots \\ \sum\limits_{l\delta\in((L-1)I,LI]} \phi(l\delta)^\mathrm{T} \end{bmatrix}, \tag{3}$$

with $l, \delta \in \mathbb{N}$. The accumulation yields a condensed signal with more diverse excitation being incorporated per sample. A loss of information is largely avoided due to a high correlation of the original samples. Instead, the information needed for identification is encoded in a different form, allowing more reliable estimation of model parameters with a reduced set size L. As shown in Fig. 1, the proposed compression can be efficiently implemented by an integrate & dump filter that converts the undersampled signal from the ADC to the compressed signal y_sc at a downsampled rate $f_\mathrm{sc} = f_\mathrm{s}/I$. The complexity of this operation is almost negligible as at most $\lceil I/\delta \rceil$ additions are required for compressing y into y_sc. For the coefficient estimation, the same amount of $\lceil I/\delta \rceil$ regression vectors $\phi(k)$ need to be summed.

D. Closed-loop DPD and Feedback Compression

Similar to the behavioral model in (1), the predistorter f_dpd of Fig. 1 may be expressed with

$$x_\mathrm{dpd}(k) = \mathbf{a}^\mathrm{T}\,\phi(x(k), x(k-1), ..., x(k-M)), \tag{4}$$

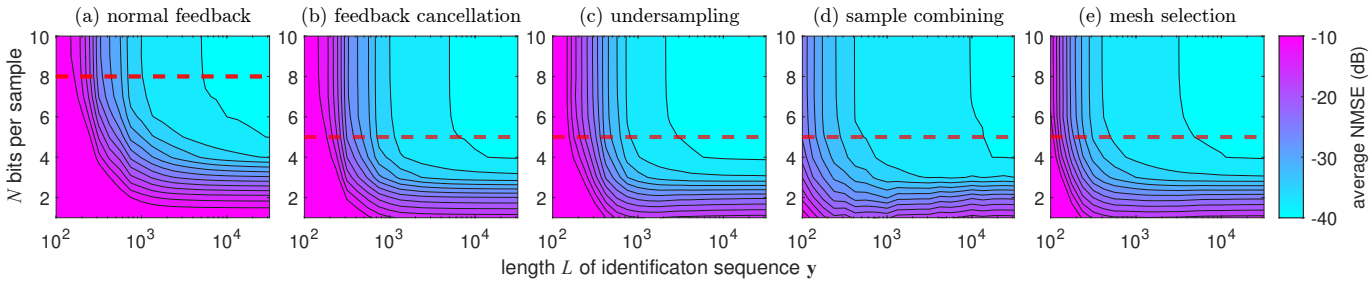

Fig. 3. PA behavioral modeling results at 3.5 GHz with respect to the feedback sequence length L and the bit-resolution N with different methods applied.

with model coefficients $\mathbf{a} \in \mathbb{C}^{J\times 1}$ and $x_{\mathrm{dpd}}(k) \in \mathbb{C}$ as predistorted input to the PA. Different than in direct modeling, the model coefficients are derived by iterative adaptation, for which we utilize the damped Gauss-Newton learning rule

$$\mathbf{a}^{(i+1)} = \mathbf{a}^{(i)} - \mu_i(\Phi_i^{\mathrm{H}}\Phi_i)^{-1}\Phi_i^{\mathrm{H}}\mathbf{e}_i. \tag{5}$$

Above, μ_i is the learning rate parameter and $\mathbf{e}_i = (\mathbf{y}_i - G\mathbf{x}_i)$ is the difference between the intended TX and measured PA output samples in the current iteration i. G denotes the desired real-valued gain of the PA, DPD and feedback channel. Φ_i contains the respective regressors. The discussed sample reduction methods can be directly applied to the provided learning rule, modifying $\mathbf{e}_i$ (instead of $\mathbf{y}_i$), and Φ_i.

To assess the feedback data volumes, we also consider uniform quantization on the feedback signal with N bits to mimic the impact of a low-resolution ADC. The quantization is applied as mid-thread, covering the full dynamic range of the signal. A way to reduce this dynamic range is by cancelling the known linear part of the feedback and only quantize the remaining error signal, as depicted in Fig. 1. We assume that for cancellation the feedback and forward signal are perfectly matched in time and amplitude. It is noted that in a real setup, the delay and power mismatch between $x_{\mathrm{dpd}}(k)$ and the feedback $y(k)$ needs to be compensated.

III. PA Direct Modeling Results

We apply the described feedback reduction techniques in the context of behavioral modeling of a gallium-nitride (GaN) 2-stage Doherty PA (*QPA3503*), operating at 3.5 GHz center frequency with +36.5 dBm output power. A 100 MHz-wide 256 QAM modulated OFDM waveform is applied, with a reduced peak-to-average power ratio (PAPR) of 7.5 dB. The *NI PXIe-5840* vector signal transceiver (VST) is used for analog signal generation and RF up-conversion, and as receiver for the amplified signal. A GMP with total of $J = 90$ coefficients is used. Fig. 3 depicts the modeling accuracy in terms of normalized mean square error (NMSE) with respect to a swept identification sequence length L, and a varied number of bits N per sample. Each point in the graphs is an average of 20 trials, each with different TX sequences, with the same data presented to each method.

We compare five cases. In the depicted case in Fig. 3(a), coefficient estimation is based on L consecutive, quantized samples $\mathbf{y}$. A lower bound for L as well as for N can be observed, with a transition zone where more samples allow

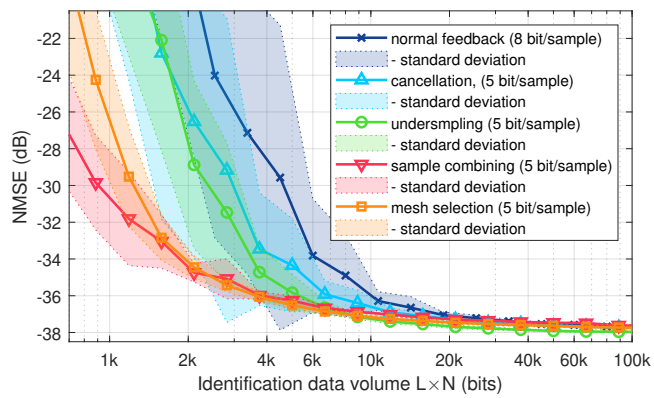

Fig. 4. Modeling accuracies with respect to identification data volume. The sample count L is altered while the bit-resolution N is fixed for each method.

to compensate for the effect of stronger quantization. In (b), as well as in the following cases, feedback cancellation is additionally employed and thus the lower bound for N is shifted towards a lower N. In (c), cancellation is combined with undersampling with $\delta = 15$ which gives slight improvement regarding the required number of samples, but significantly relaxes ADC sampling requirements. In (d), the proposed sample combining method is added which gives a major contribution towards a reduced number of samples. Sample combining is applied with an integration window of $I = 40$ samples and $\delta = 15$ thus accumulating 2-3 samples per iteration. In (e), the performance of the MeS method as described in [4] is shown for reference, providing a similar performance as the proposed sample combining method.

For each of the methods we identified a minimum bit-resolution to minimize the feedback data, indicated by the red dashed lines in Fig. 3. These settings are 8 bits for the case without cancellation, and 5 bits for the cases where cancellation is applied. The resulting modeling NMSE values over the identification data size $L \times N$ are shown in Fig. 4.

IV. Closed-loop DPD Measurement Results

The same PA (*QPA3503*) and equipment are next used for the closed-loop DPD experiments, with the same TX signal specifications (100 MHz bandwidth, 7.5 dB PAPR). For each closed-loop iteration, a different TX sequence was generated. The DPD experiments are repeated for five trials, using new data in each trial. Again, each evaluated method was presented the same TX data to ensure fair comparison. For DPD, a GMP with $J = 90$ basis functions is employed.

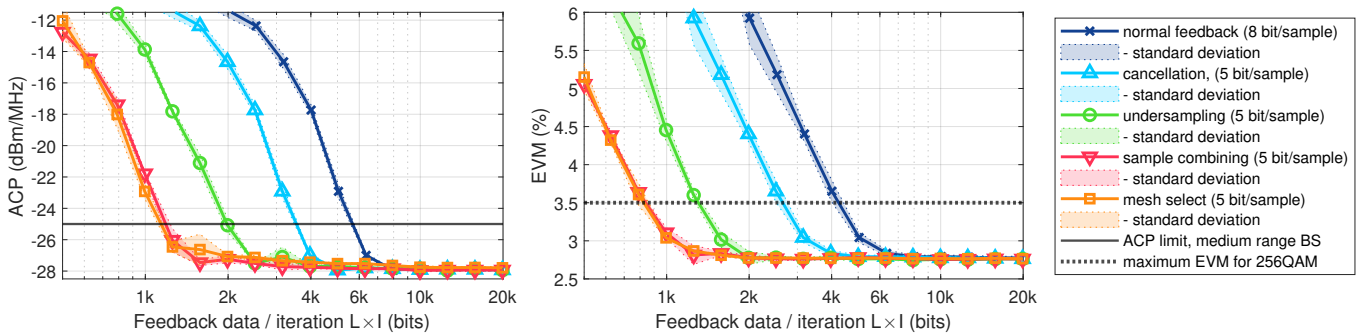

Fig. 5. Closed-loop DPD performances of different feedback compression schemes at 3.5 GHz with respect to the feedback data per iteration. The data volume is varied by changing the sample count, while the number of bits/sample remains fixed for each method as stated in the legend.

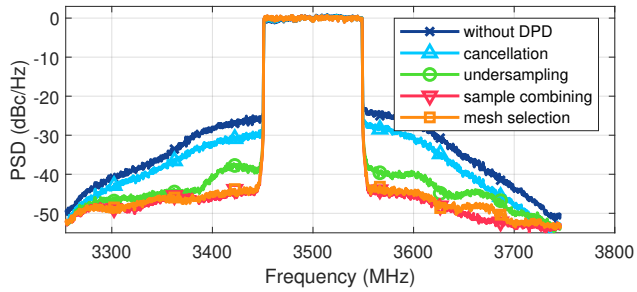

Fig. 6. Spectra after 20 closed-loop iterations with L=300 samples each.

Proper choice of the learning rate μ is crucial for successful closed-loop DPD. A too low μ value will unnecessarily rein the convergence whereas a too high μ may cause instability. Furthermore, the optimal choice for μ has to be made with respect to the statistical relevance of the training sequence, i.e. a larger training sequence allows choosing larger μ. However, stable convergence can also be ensured with very short training batches, given a small μ and many iterations. To select a suitable μ, we identified for each method and setting L the maximum steady-state μ_{ss} which retains the performance on a preset DPD. Based on this study the relationship of the form

$$\log \mu_{\text{ss}} = \alpha \log L + \beta \tag{6}$$

was derived, with β specific to each reduction method, whereas α is identical. The closed-loop DPD experiments are then parameterized by choosing $\mu_1 = 2\mu_{\text{ss}}$ in the first iteration and decay afterwards towards μ_{ss}, to speed-up the convergence. The learning rate is further limited to maximum one.

With these settings, the results in Figs. 5 and 6 are achieved after 20 closed-loop iterations. Solid DPD performances are reached for all methods at a large feedback data volume, whereas the proposed sample combining, and MeS allow achieving a high degree of linearity also with significantly less feedback data. We find that the feedback data volume required to meet the 5G NR maximum adjacent channel power (ACP) specifications for medium range base-stations or comply with the minimum specified symbol error vector magnitude (EVM) for 256 QAM can be reduced by a factor of 5×, when compared to the baseline case, where the feedback is normally sampled. The experiments furthermore demonstrate that the proposed sample combining technique is as capable as the histogram-based sample selection, however, without need to extract signal statistics of the excitation or feedback signal as to select specific samples in the transmit signals and feedback.

V. Conclusion

Sample combining for low-complexity feedback signal compression for DPD model adaptation was proposed in this paper. Despite the simplicity of the method, the presented experiments show that solid performances for PA behavioral modeling or closed-loop DPD can be reached with a significantly reduced data volume with the proposed method, while taking also a low-resolution quantization into account. The method was demonstrated to achieve similar performances as histogram-based sample selection, however with negligible added complexity. It is thus well suited for implementation. At the same time, the ADC rate and bit-width can be reduced, reducing the overall data load of the DPD system.

Acknowledgment

This project was funded by EU Horizon 2020 under the Marie Skłodowska-Curie grant agreement No. 860921, and by Academy of Finland under grants 319994, 338224, 332361.

References

[1] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1880–1890, May 2017.

[2] H. Huang, P. Mitran, and S. Boumaiza, "Digital predistortion function synthesis using undersampled feedback signal," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 10, pp. 855–857, Oct. 2016.

[3] Y. Liu, X. Quan, S. Shao, and Y. Tang, "Digital predistortion architecture with reduced ADC dynamic range," *Electron. Lett.*, vol. 52, no. 6, pp. 435–437, Mar. 2016.

[4] T. Wang, P. L. Gilabert, and G. Montoro, "Under-sampling effects and computational cost reduction in RF power amplifier behavioral modeling," in *10th Europ. Microw. Integr. Circuits Conf. (EuMIC)*, 2015, pp. 57–60.

[5] J. Kral, T. Gotthans, R. Marsalek, M. Harvanek, and M. Rupp, "On feedback sample selection methods allowing lightweight digital predistorter adaptation," *IEEE Trans. Circuits Syst.*, vol. 67, no. 6, pp. 1976–1988, Feb. 2020.

[6] T. Wang and P. L. Gilabert, "Mesh-selecting for computational efficient PA behavioral modeling and DPD linearization," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 1, pp. 37–40, Jan. 2021.

[7] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.