Digital Predistortion with Compressed Observations for Cloud-based Learning

Arne Fischer-Bühner[1], Emil Matus[2], Manil Dev Gomony[1], Lauri Anttila[3], Gerhard Fettweis[2] and Mikko Valkama[3]

[1]Nokia Bell Labs, Antwerpen, Belgium
[2]Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, Germany
[3]Department of Electrical Engineering, Tampere University, Finland

***Abstract*—This paper presents a novel system architecture for digital predistortion (DPD) of power amplifiers, where the training of the DPD model is done in a remote compute infrastructure i.e. cloud or a distributed unit (DU). In beyond-5G systems it is no longer feasible to perform computationally intensive tasks such as DPD training locally in the low-power radio unit front-end. Thus, we propose to split DPD system and perform the compute-intensive DPD training in the DU where more processing resources are available. To enable the distant training, the observed PA output must be available, however, sending the data intensive observation signal to the DU adds additional cost to the system. A low-complexity compression method is proposed to reduce the bit-resolution of the observation signal by removing the known linear part in the observation to use fewer bits to represent the remaining information. Our numerical simulations show a reduction from 8 to 4 bits/sample for the accurate training of the DPD model to compensate for the distortions of a strongly driven power amplifier operated at 28 GHz with a 200 MHz wide OFDM signal.**

***Index Terms*—5G, 6G, digital predistortion, nonlinear distortion, power-efficiency, power amplifier, cloud-based learning, observation compression**

I. Introduction

Beyond-5G radio architectures need to operate waveforms with a high spectral efficiency to meet capacity targets. However, these waveforms have a high peak-to-average power ratio (PAPR), that reduces the efficiency of the power amplifiers (PA) used. In turn, highly optimized architectures are employed, and PAs are operated at the limits of their capabilities [1]. These efficiency improvements are mostly at the expense of the transmitter system's linearity. Maintaining the linearity of the transmitter is highly necessary since interference with other transmit-channels or users due to spectral regrowth must be generally avoided. Moreover, low in-band signal distortion is a prerequisite for the higher-order symbol constellation mappings enabling the high data rates in beyond-5G systems. The most capable and also widely adopted technique for preserving the transmitter linearity is digital predistortion (DPD) [2].

The general principle of DPD is straightforward: Applying the inverse characteristic of the PA prior to the PA should result in a perfectly linear characteristic in total. However, the required predistortion function is usually unknown and moreover subject to change due to the PA's operating conditions such as temperature variations, aging impact and changes in the transmit signal [3]. Moreover, in multi antenna systems, the nonlinear distortions are discovered beam-dependent [4]. Therefore, the DPD models are identified adaptively by observing the nonlinear distortions at the PA output as illustrated in Fig. 1.

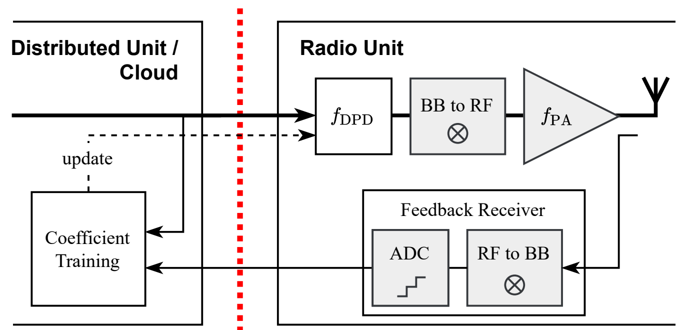


Fig. 1. System level view of the split DPD system where coefficient training is performed in the DU / cloud

Beyond-5G systems will be deployed as split architectures, where a highly integrated radio unit (RU) is located close to the antenna array and needs to be lightweight which is why the power budget is limited. Thus, it is not feasible to perform computationally demanding tasks in the RU. Instead, these tasks are shifted to a distant distributed unit (DU) that has the necessary computational resources available. The split of the low-power/low-complexity RU and a DU for extensive processing and upper layer tasks is in line with the specification defined by the O-RAN Alliance [5]. While DPD clearly is a low-level task that must sit in the transmit front-end close to the PA, the training of the DPD model coefficients is a compute intensive task. Especially in view of a growing number of model coefficients to compensate strong memory distortion stimulated by wide-band signals, DPD coefficient training must be performed on powerful compute hardware [1]. The use of MIMO techniques further adds to the training complexity as multiple PAs and antennas need to be linearized in dependence to the beam [4]. To keep processing complexity off the RU, we propose to shift the DPD coefficient training to the DU or cloud as shown in the block diagram in Fig. 1.

For the training of the DPD model, the PA output signal must be available in the DU. The PA output observed by the feedback receiver needs to be fed back through the interface separating the RU and the DU. Usually, the link between RU and DU should be especially lightweight as the connecting channel can be long and is therefore sensitive to cost. The observation signal is however very data-intensive due to the high oversampling ratios needed to also capture the out-of-band distortions resulting in a high feedback sampling rate. These sampling rate requirements will further increase in view of very broadband signals expected in 5G and beyond systems. Thus, passing the PA observation signal to the DU will become a bottleneck in future systems and techniques will be needed to reduce the data-rate and amount of data passed back to the DU. Various works focus on reducing the observed bandwidth to allow for a lower analog-to-digital conversion (ADC) rate in the feedback receiver. Methods include undersampling of the observation signal in combination with the direct learning architecture [6], [7], the band-limited Volterra series which allows for a reduced observation bandwidth [8], or spectral extrapolation of a bandwidth constrained observation [9]. While these methods effectively reduce the acquisition bandwidth and thus the rate at which the observation is sampled, they do not address that the individual samples are recorded with a high bit-resolution.

Typically, the ADC in the observation path uses 14-bit to sample the signal [10]. A DPD with reduced ADC dynamic range and lower bit resolution is presented in [11]. Linear cancellation is performed in the analog domain using the pre-distorted signal, which is digitally filtered first to remove the out-of-band parts in the signal. An additional digital-to-analog converter (DAC) and further analog circuitry are needed for the cancellation. The authors show that 8-bit sampling is sufficient to achieve the full DPD performance. A related technique is presented in [12], where the original transmit signal is used for analog cancellation. With that, 1-bit sampling employing sign-based Gauss-Newton learning is proposed. While this approach is very capable, a second DAC stage is needed and a custom ADC design based on comparators is required, which takes considerable development effort. Moreover, the analog signals need to be precisely aligned in amplitude and time which may be a hurdle with higher sampling rates.

A different approach is followed in [13], where the observation is captured in a frequency-domain representation directly from the analog signal. This is achieved by applying a narrow-band filter to the analog signal, that sweeps across the frequency range of interest. The ADC rate is decoupled from the transmit signal bandwidth and fewer samples are needed for the DPD coefficient training. However, custom analog components such as a frequency-tuneable oscillator and a linear integrator are needed for the frequency-domain acquisition, which also makes this approach an expensive solution in actual product development.

In contrast to the aforementioned techniques which mainly aim to relax the ADC requirements, our approach is to mitigate the interface load by applying digital compression to the high-resolution observation signal. The lossy compression method proposed takes advantage of the fact that the principal component, namely the linear part of the transmit signal, is known to both the DU and RU. With that, the resolution required to represent the observation signal, and the total number of bits to pass through the interface in effect, can be significantly reduced.

The main contributions in this paper can be summarized as follows. 1) We present a novel DPD architecture for cloud-based DPD coefficient training that integrates with the split architecture of beyond-5G systems and relies on commercially available ADCs in the feedback receiver. 2) We propose a low-complexity compression method to mitigate the data rate of the observation signal passed back from the RU to the DU. Our compression method is very lightweight and thus the hardware overhead is low. 3) In numerical simulations we evaluated the proposed solution and show that a significant reduction of the bit-resolution is possible without degrading the DPD linearization performance. The rest of the paper is organized as follows. The proposed compression technique and the DPD algorithm are explained in Section II. Subsequently, numerical simulations and results are presented in Section III. Results and limitations are discussed in Section IV. Finally, Section V concludes this work.

II. Proposed Compression Method

For the DPD coefficient training, the PA output is down-converted and digitized in the feedback receiver. Sampling of the analog signal yields the observed PA output signal $y(k)$, which is an approximate version of the down-converted PA output in terms of bandwidth and bit-resolution. The bit-resolution of the observation signal must be large enough, to ensure that the nonlinear distortions, which have much lower energy than the signal itself, are not masked by the quantization noise floor. Reduction of the bit-resolution causes a degradation of the linearization performance. The noise introduced by quantization has a variance σ_n^2 proportional to the squared step-size b, as provided in (1). The step size, in turn, is a function of the number of bits N and the dynamic range Δu of the quantization.

$$\sigma_\mathrm{n}^2 \sim b^2 = \left(\frac{\Delta u}{2^N}\right)^2 \tag{1}$$

As it is our goal to compress the observation signal i.e. reduce N, it becomes evident from (1) that the dynamic range Δu must be lowered to mitigate the effects of using fewer bits and thus quantization steps. However, the dynamic range is given by peak values of the observation signal as a lower dynamic range would introduce clipping distortion which adds undesired distortion to the observed PA output.

Fortunately, the undistorted transmit symbols s are known to both sides of the interface, as shown in Fig. 2. On the RU side, these are translated into the time-domain transmit signal using the inverse Fourier transformation (IFFT), typically with additional crest factor reduction applied to limit peak powers in the signal. Performing the same processing steps provides

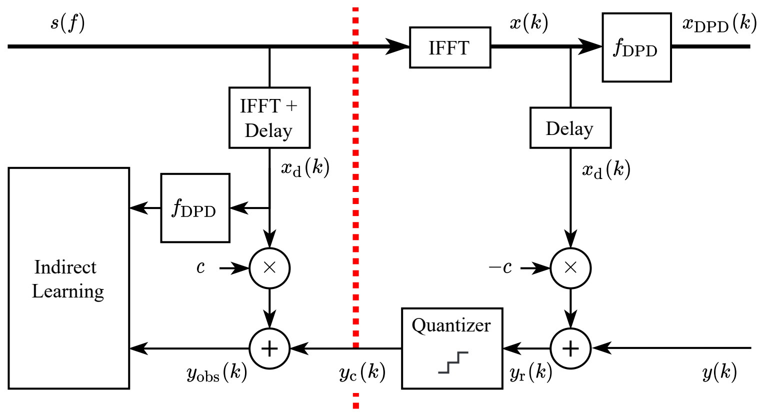


Fig. 2. Proposed architecture with compression applied to y in the feedback path by linear cancellation and quantization

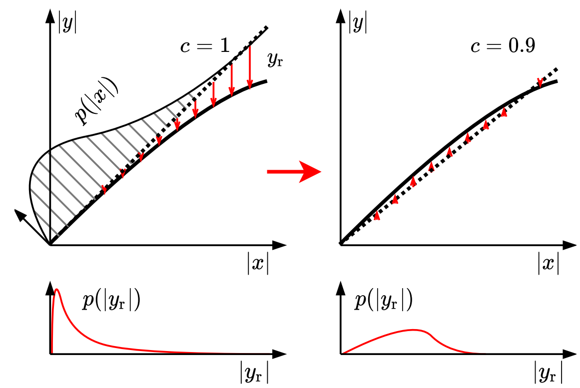


Fig. 3. Impact of scaling the cancellation reference signal with factor c on the distribution of y_{r}

the identical signal on the DU side. It is thus sufficient to pass the information about the unknown nonlinear distortion introduced by the PA. This can be achieved by subtracting a delayed version x_d of the original transmit signal x from the observed PA output y.

$$y_{\text{r}}(k) = y(k) - x_d(k) \tag{2}$$

Subtracting x_d is identical to canceling the linear component in the PA output signal which significantly deducts power from the in-band part of the observation signal. The measured PA output and transmit signals must be aligned in time and amplitude, to avoid impact from the delay and linear gain introduced by the transmission path and the feedback receiver. Alignment in terms of signal mean power is assumed to happen in the feedback receiver. Time alignment is achieved by delaying the reference signals x_{d} with regard to x in order to bridge the delay introduced in the analog transmission path of the front-end and while observing the PA output signal. Delay and amplitude estimation in the digital domain are necessary processing steps for the coefficient training in any way [2]. Some overhead arises in that the delay must be applied on both the RU and DU side. Furthermore, the delay filter on the DU side needs to cover the additional time needed for compression and sending the observation signal from the RU to the DU through the connecting channel.

While the above cancellation significantly reduces the dynamic range of the signal, it also affects the value distribution of the signal. The proportional relationship of σ_{n}^2 to the square of the step-size b^2 given by (1) only holds true if the value distribution of the quantized signal is flat with respect to the step-size of the quantization [14]. Applying the proposed cancellation in (2) can lead to an uneven value distribution of y_{r}. Particularly with strong non-linear distortion present in the first iteration of the DPD training, the condition for (1) is no longer fulfilled. The impact that the cancellation has on the value distribution is illustrated in Figure 3. Large deviations from the linear behavior usually occur at peak powers of the signal, where the PA operates in its compression region. Thus, the high PAPR of the OFDM signal x is further increased by applying (2) and the resulting strongly non-uniform value distribution of y_{r} prevents the desired reduction of the quantizer bit-resolution. To resolve this issue, we propose to adapt the cancellation in the first iteration to keep parts of the linear component in the residual signal. This is achieved by applying linear scaling to the reference signal with a scaling factor c as

$$y_{\text{r}}(k) = y(k) - c \cdot x_{\text{d}}(k). \tag{3}$$

The effect of scaling x_{d} with c is thereby subject to the specific PA non-linearity and saturation behavior, and a general optimal value cannot be provided. Instead, c should be chosen specific to the PA at hand to minimize the resulting dynamic range and, secondly, the PAPR of the y_{r}. In this work, the best choice for c is done based on the numerical simulations described in Section III-B. This consideration is only relevant for the first of multiple training iterations. In all of the following iterations, the PA nonlinearity is largely compensated by DPD already and c is best chosen as $c = 1$. Subsequently to the cancellation the bit-resolution of the signal is reduced yielding the compressed signal y_{c} by applying mid-thread quantization to y_{r} with a step-size b and sufficient dynamic range Δu. For the training on the DU side, the PA observation signal y_{obs} is recovered from the compressed signal y_{c} by adding the identical cancellation signal again.

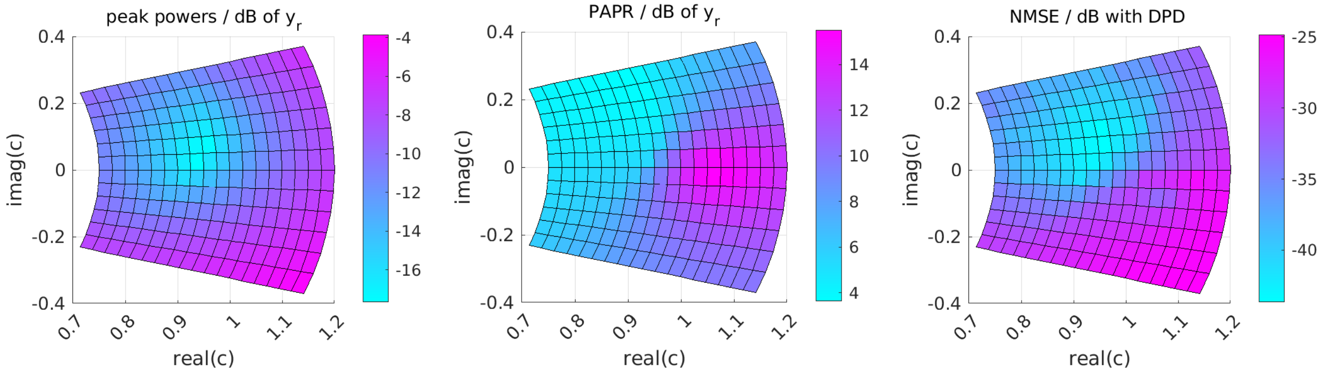
$$y_{\text{obs}}(k) = y_{\text{c}}(k) + c \cdot x_{\text{d}}(k) \tag{4}$$

III. Numerical Results

A. Simulation Conditions

In order to evaluate the compression method, we pursued simulations using MATLAB. For the PA we used the measured 28 GHz mmWave PA models provided in [15]. The PA models are provided as coefficients of a common memory polynomial [16], as

$$f_{\text{PA}}(x(k)) = \sum_{i=1}^{K} \sum_{j=0}^{M} g_{i,j}\; x(k-j)\, |x(k-j)|^{i-1} \tag{5}$$



(a) Peak powers of y_r (b) PAPR of y_r (c) Performance after 1st DPD iteration, $N = 3\,$bit

Fig. 4. Impact of different values of the scaling factor c on properties of y_r and the NMSE DPD performance

with weights $\boldsymbol{g} = [g_{1,1}, \dots, g_{1,M}, g_{2,1}, \dots, g_{K,M}]$ where the polynomial order is $K = 11$ and the maximum memory history is $M = 2$. For the predistorter nonlinear function f_{DPD} a memory polynomial similar to the one in (5) is used, but with coefficients $\boldsymbol{a} = [a_{1,1}, \dots, a_{1,M}, a_{2,1}, \dots, a_{K,M}]$, a nonlinear order $K = 7$, and a maximum memory depth of $M = 4$.

The DPD model is trained in multiple iterations using the indirect learning scheme. In each iteration, a post-distorter is estimated from y_{obs} and x_{DPD}, by modeling the known PA input signal x_{DPD} as a function f_{DPD} of the observed PA output y_{obs}. The identified model coefficients of the post-distorter are then copied and used as the identical predistorter function. Thus, for the DPD training, the latest predistortion function is applied to x_{d} on the DU side to generate a time-aligned version of the PA input signal x_{DPD}. Then, the method of least squares is applied to solve

$$\boldsymbol{a} = \arg\min_{\hat{\boldsymbol{a}}} \left(x_{\mathrm{DPD}}(k) - f_{\mathrm{DPD}}(y_{\mathrm{obs}}(k) \,|\, \hat{\boldsymbol{a}})\right), \tag{6}$$

yielding

$$\boldsymbol{a} = (\mathrm{U}^H \mathrm{U})^{-1} \mathrm{U}^H \boldsymbol{x}_{\mathrm{DPD}}, \tag{7}$$

where $(\cdot)^H$ denotes the Hermitian (complex) transpose,

$$\boldsymbol{x}_{\mathrm{DPD}} = [x_{\mathrm{DPD}}(1), x_{\mathrm{DPD}}(2), ..., x_{\mathrm{DPD}}(n)]^{\mathrm{T}}$$

and

$$\mathrm{U} = \begin{bmatrix} \varPhi_{1,0}(1) \cdots \varPhi_{1,M}(1) & \cdots & \varPhi_{K,M}(1) \\ \varPhi_{1,0}(2) \cdots \varPhi_{1,M}(2) & \cdots & \varPhi_{K,M}(2) \\ \vdots \qquad\qquad \vdots & \ddots & \vdots \\ \varPhi_{1,0}(n) \cdots \varPhi_{1,M}(n) & \cdots & \varPhi_{K,M}(n) \end{bmatrix}$$

with

$$\varPhi_{i,j}(k) = y_{\mathrm{obs}}(k-j)\,|y_{\mathrm{obs}}(k-j)|^{i-1}.$$

n is the number of samples used for training.

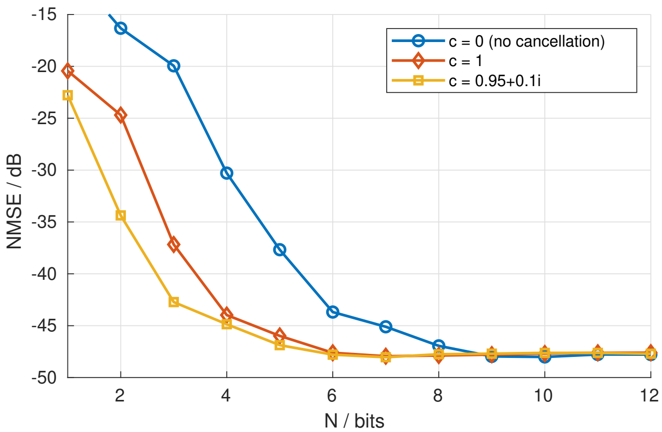
In each training iteration, the model coefficients are fully retrained. The signals used for training and evaluation are OFDM signals with a 200 MHz bandwidth, including a 10 % guard band on both sides, and sampled with five times oversampling. The sub-carrier modulation is 64 QAM, while the sub-carrier spacing is 60 kHz. Windowed overlap and add (WOLA) is used for spectral shaping, and additional crest factor reduction using peak windowing is employed to limit the PAPR of the transmit signal to 7 dB [17]. We use different signals for the training iterations and the evaluation of the DPD performance. The length of each training sequence is 2k samples. For the performance evaluation we use a much longer signal with seven consecutive OFDM symbols.

B. Determining the scaling factor c

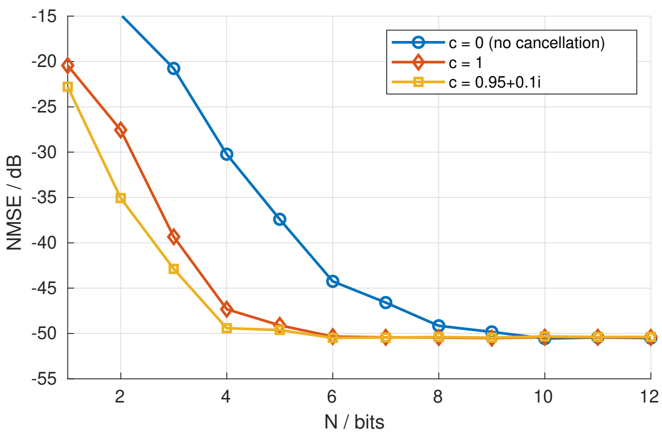
The optimum value for the scaling factor c is essential for proper coefficient training and can be derived numerically. Fig. 4 shows properties of the residual signal y_r after cancellation and the DPD performance after the first iteration for different values of c in the complex plane. Fig. 4a shows that with the simulated PA, the minimal peak values are found at a scaling factor $c < 1$ and slightly rotated in the complex plane. An important observation, shown in Fig. 4b, is the impact that the scaling factor c has on the PAPR of the y_{r}. With the evaluated PA, a scaling $c = 1$ results in a high PAPR of approximately 12 dB. Instead, choosing a lower and complex rotated scaling factor c gives a significantly lower PAPR of 6 dB or better. The absolute value for c is related to the curvature of the PA's amplitude characteristic, while the phase of c is specific to phase characteristic of the PA. When applying the proposed compression with a low resolution to represent y_{r}, such as 3 bits, the resulting DPD performance is subject to both the dynamic range and PAPR of y_{r}. The simulated performance shown in Fig. 4c visibly coincides with the dynamic range y_r. However, a high PAPR of y_{r} impairs the compression and causes a weak NMSE performance. The impact of different scaling factors c on the simulated DPD performance in terms of EVM and ACLR largely coincides with the NMSE behavior.

C. Performance Evaluation

The NMSE performance in relation to different quantizer resolutions N is presented in Fig. 5. Although the previously discussed scaling with c applies to the first training iteration only, it also affects the performance of following iterations. In Fig. 5a, the performance after the first training iteration is depicted for the different cases, with and without cancellation, showing a clear improvement for low bit-resolutions $N < 8$



(a) 1st training iteration



(b) 3rd training iteration

Fig. 5. DPD performance measured as NMSE as a function of the quantization bit-resolution N for samples in y_c

if compression is applied. Also, the impact of a specifically chosen scaling factor c is visible for small bit resolutions. The NMSE performance further improves after three training iterations, by $\approx 3\,\text{dB}$ equally for all bit-resolutions as shown in Fig. 5b. However, the initial benefit of using a specific scaling factor c remains visible after three training iteration. The compression achieved by the proposed method can be quantified as a reduction from 8 bits to 4 bits for a comparable NMSE performance.

The power spectral density (PSD) plots of the distorted and linearized PA output signal with and without compression are provided in Fig. 6. Using a 4-bit represented observation signal without compression, the derived DPD model is unable to compensate any of out-of-band distortions in the PA output. With additional compression applied and $c = 1$, DPD is still unable to compensate for the spectral leakage. However, when choosing c appropriately, a 4-bit approximate representation of the residual signal is sufficient to also compensate for the out-of-band distortions. After three training iterations and properly chosen c in the first iteration, the out-of-band distortions are completely removed from the PA output.

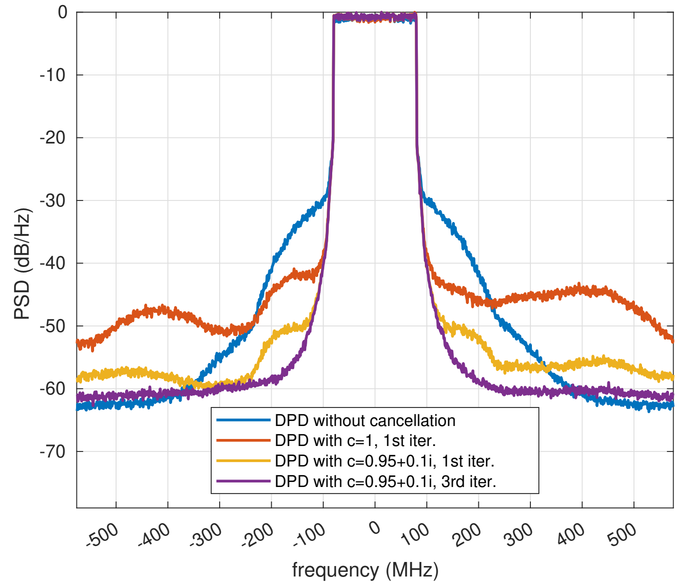


Fig. 6. PSD graphs with and without cancellation using quantization using $N = 4$ bits

IV. Discussion

While the proposed compression method shows good results in our simulations, we expect our method to perform similarly with different PA models. It is however noted that the achieved compression is related to the amount of distortion introduced by the PA. Stronger nonlinear distortion inevitably leads to a larger dynamic range of the cancelled signal and thus requires more bits to represent. Also, the choice of the scaling factor c in the first training iteration is specific to the PA gain and phase characteristic. Frequent re-estimation of the coefficients will make the impact of a specifically chosen scaling factor c for the first training iteration vanish. However, as multiple training iterations imply a greater computational effort, training is repeated only until the linearity targets are reached. With indirect learning, this can be the case after few iterations already so that the benefit from a specifically chosen scaling factor c is notable.

Our proposed compression adds only little processing and hardware overhead. Since the compression is applied in the digital domain, no custom analog hardware is needed. Instead, the proposed architecture relies on commercially available ADCs in the feedback receiver. To apply the original transmit signal for cancellation, the transmit signal must be delayed to be time-aligned with the observed PA output. Hence, an additional delay filter is needed in the RU, in addition to the scaling, addition and quantization. The proposed compression only targets the bit-resolution of the feedback signal. However, combination of the proposed method with the mentioned sampling rate reduction techniques in [6]–[9] is imaginable.

Some processing overhead is introduced by the DPD architecture split. In addition to sending the observation signal from the RU to the DU, the time-domain PA input signal must be regenerated on the DU side. For that, identical processing steps that happen prior to the PA in the RU must be repeated in the DU to retrieve an identical signal for the coefficient estimation.

The available transmit symbols need to be transformed into the time-domain signal and the current predistortion function must be reapplied to enable indirect learning. However, in contrast to the processing complexity to compute the DPD coefficients, the additional overhead is secondary.

V. Conclusions

In this paper, a novel DPD architecture for cloud-based learning was considered. The proposed architecture integrates with the distributed radio architecture expected in 5G and beyond-5G systems. Due to the limited power-budget of the RU front-end, compute intensive tasks such as DPD training are no longer feasible in the RU. Instead, a split DPD architecture was presented, where DPD model training is pursued in a remote DU. Sending the observed PA output from the RU to the DU was identified as the major bottleneck of the proposed DPD architecture. A low-complexity compression method was proposed to reduce the bit-rate of observation signal sent back to the DU. The proposed method exploits the fact that the linear component of the observation signal is already known on the DU side. Thus, it is sufficient to only feed back the information about the nonlinear distortions. Numerical simulations show that the DPD model can be accurately trained on the DU side, using a bit-resolution of 4 bits/sample to represent the compressed observation signal, compared to the 8-bit quantized uncompressed observation signal.

Acknowledgment

This project has been partially funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (Innovative Training Networks) grant agreement No. 860921 and the ECSEL-JU project BEOND5 grant No. 876124. A. Fischer-Bühner would like to express his special thanks to the Vodafone Chair Mobile Communication Systems, Technical University Dresden for the opportunity and guidance during his diploma thesis.

References

[1] P. Pascual Campo, V. Lampu, A. Meirhaeghe, J. Boutellier, L. Anttila, and M. Valkama, "Digital Predistortion for 5G Small Cell: GPU Implementation and RF Measurements," *Journal of Signal Processing Systems*, vol. 92, no. 5, pp. 475–486, may 2020.

[2] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1880–1890, 2017.

[3] R. N. Braithwaite, *General principles and design overview of digital predistortion*. Cambridge University Press, 2011, p. 143–191.

[4] C. Fager, T. Eriksson, F. Barradas, K. Hausmair, T. Cunha, and J. C. Pedro, "Linearity and Efficiency in 5G Transmitters: New Techniques for Analyzing Efficiency, Linearity, and Linearization in a 5G Active Antenna Transmitter Context," *IEEE Microwave Magazine*, vol. 20, no. 5, pp. 35–49, 2019.

[5] O-RAN Alliance, "O-RAN Specification," https://www.o-ran.org/specifications, 2021.

[6] Y. Liu, J. J. Yan, H.-T. Dabag, and P. M. Asbeck, "Novel Technique for Wideband Digital Predistortion of Power Amplifiers With an Under-Sampling ADC," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2604–2617, 2014.

[7] H. Huang, P. Mitran, and S. Boumaiza, "Digital predistortion function synthesis using undersampled feedback signal," *IEEE Microwave and Wireless Components Letters*, vol. 26, no. 10, pp. 855–857, 2016.

[8] Y. Li, X. Wang, and A. Zhu, "Sampling Rate Reduction for Digital Predistortion of Broadband RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 3, pp. 1054–1064, mar 2020.

[9] Y. Ma, Y. Yamao, Y. Akaiwa, and K. Ishibashi, "Wideband Digital Predistortion Using Spectral Extrapolation of Band-Limited Feedback Signal," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 7, pp. 2088–2097, jul 2014.

[10] Texas Instruments Inc., "Using AFE77xx in a Digital Pre-Distortion System," https://www.ti.com/lit/an/sbaa370a/sbaa370a.pdf, 2019.

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

[12] H. Wang, G. Li, C. Zhou, W. Tao, F. Liu, and A. Zhu, "1-bit Observation for Direct-Learning-Based Digital Predistortion of RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 7, pp. 2465–2475, 2017.

[13] N. Hammler, A. Cathelin, P. Cathelin, and B. Murmann, "A spectrum-sensing DPD feedback receiver with 30× Reduction in ADC acquisition bandwidth and sample rate," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 9, pp. 3340–3351, 2019.

[14] Easton, Roger, *A Review on Quantization*. Rochester Institute of Technology, 2004, ch. 14, pp. 281–302. [Online]. Available: https://www.cis.rit.edu/class/simg712-90/notes/14-Quantization.pdf

[15] A. Brihuega, L. Anttila, M. Abdelaziz, T. Eriksson, F. Tufvesson, and M. Valkama, "Digital Predistortion for Multiuser Hybrid MIMO at mmWaves," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3603–3618, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9097919/

[16] L. Ding, G. Zhou, D. Morgan, Z. Ma, J. Kenney, J. Kim, and C. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 159–165, 2004.

[17] O. Vaananen, J. Vankka, and K. Halonen, "Simple algorithm for peak windowing and its application in GSM, EDGE and WCDMA systems," *IEE Proceedings - Communications*, vol. 152, no. 3, p. 357, 2005.