

On Digital Optimization of Analog Self-Interference Cancellation for Full-Duplex Wireless Systems

Niklas Knaepper, Gerald Enzner, Aleksej Chinaev

Department of Medical Physics and Acoustics, Division of Speech Technology and Hearing Aids,

Carl von Ossietzky Universität Oldenburg

26111 Oldenburg, Germany

{niklas.knaepper, gerald.enzner, aleksej.chinaev}@uni-oldenburg.de

Abstract—Wireless systems with an inband full-duplex (IBFD) transceiver typically require multiple lines of defense against harsh self-interference, specifically, to avoid saturation of the analog-to-digital converter (ADC) in the receiver. We may unite the typical tandem operation of successive analog and digital self-interference cancellation (SIC) stages by means of digitally-assisted analog SIC. In this case, the ADC in the receive path requires considerable attention due its possibly overloaded operation outside the intended range. Using neural-network-based architectures of the transmitter nonlinearity, we therefore describe and compare four system options for SIC model optimization with different treatment of the receiver ADC in the learning process. We find that omitting the ADC in the backwards path via a so-called straight-through estimation approximation barely impedes model learning, thus providing an efficient alternative to the classical approach of automatic gain control.

Index Terms—Digitally-assisted SIC with A/D conversion

I. INTRODUCTION

Self-interference cancellation (SIC) is an integral part of inband full-duplex communication, where the same frequency band is used for simultaneous transmitting and receiving. Apart from SIC achieved in propagation domain, e.g., via physical separation of antennas, established SIC components are frequently classified as analog or digital, i.e., using analog filter structures or digital signal processing, respectively [1]–[3]. While digital techniques bear more flexibility, higher-order filters and easier implementation, they cannot usually be deployed without help of preceding analog SIC for avoiding saturation of the analog-to-digital converter (ADC) [4].

Digitally-assisted SIC schemes [5]–[7] aim to combine the advantages of digital filtering with those of analog cancellation by modeling the self-interference (SI) channel in digital domain, while realizing the actual cancellation via RF circuitry, i.e., before ADC. Although various publications prove the feasibility of this approach in principle, few explicitly discuss the role of the receiver ADC as part of the signal path during optimization. Thus, it is still not obvious how this ADC affects optimization, or, if and how it needs to be accounted for. The latter is especially true at the beginning of optimization, i.e., when still lacking sufficient SIC to avoid heavy saturation.

A straightforward workaround consists in bypassing the low-noise amplifier (LNA) during training [5], [6]. However, this may not always be enough to completely avoid saturation,

or could even lead to under-utilization of the ADC. Although not explicitly addressed as a concern, the authors in [7] seem to avoid saturation during optimization by careful data design. Similarly, authors in [8] choose to reduce the transmission power during training, which, however, potentially changes the nonlinear behavior of the power amplifier. The effects of quantization noise and saturation on SIC are discussed in [9] under the assumption of a known SIC model. Authors in [10] provide an elaborate investigation of the effects of ADC resolution on optimization, but in a context of a different digitally-assisted SIC scheme, which does not rely on conventional digital filters. Furthermore, [9] and [10] mention the possibility to apply an automatic gain control (AGC) as a remedy for ADC saturation, however, missing implementation details, study and discussion in the context of SIC model optimization.

This work investigates a generalized ADC treatment for digitally-assisted SIC when using (while not critically relying on) neural network models of the nonlinear SI path [11], [12]. Unfortunately, ADC simulation as a component of a backpropagation model is not immediately a productive approach due to ill-conditioned differentiation properties of its step-function. We hence propose model representation of the ADC within the dynamic range, or even including the saturation range, according to the recent straight-through estimation (STE) theory for quantized networks [13], [14]. It is demonstrated that STE allows for SIC learning despite heavy initial ADC saturation and it performs similarly to an optimization strategy that additionally relies on AGC. We also compare to a system trained offline on quantized signals with low LNA gain. Our experimental comparison includes evaluation of SIC and end-to-end performance in terms of bit error rates (BERs).

The remainder of this paper is organized as follows. Sec. II provides an overview of digitally-assisted SIC. Sec. III then describes the approach for system simulation and introduces the different ADC-related learning schemes. Evaluation results are presented in Sec. IV and Sec. V concludes.

II. SYSTEM ARRANGEMENT AND SIMULATION

A block diagram of the transceiver system, including the digitally-assisted SIC mechanism, is shown in Fig. 1 (a), and Fig. 1 (b) illustrates power levels of the involved signals. The analog transmission signal $z(t)$ with power $P_z = 20$ dBm results from the digital transmit signal $s[k]$ through digital-to-analog conversion (DAC) and subsequent power amplification

This work is funded by the German Research Foundation (DFG) under Grant EN 869/4-1 with Project ID 449601577.

(PA) with nonlinearity. Note that this work does not consider issues related to the transmitter DAC, since these are typically not of concern. The transmit signal is received as

$$y(t) = h_{SI}(t) * z(t) + x(t) \quad (1)$$

over the linear SI channel $h_{SI}(t)$, where $*$ denotes linear convolution, and $x(t)$ is the typically smaller signal of interest (SOI). The nonlinear self-interference path from the transmit signal $s[k]$ to the received signal $y(t)$ was labeled "Hammerstein" in [11]. We assume SOI-SNR in the range of [0, 40] dB with respect to a receiver noise floor around $P_n = -77$ dBm not explicitly shown in Eq. (1) and Fig. 1 (a).

A neural network models the nonlinear SI path from signal $s[k]$ to $y(t)$. It uses $s[k]$ to predict, after auxiliary DAC, the received SI signal $y(t)$, such that SIC can be achieved via subtraction. Our network architecture, suitable for the SI path, conforms to that of the "global Hammerstein" model described in [11], i.e., comprising a small multi-layer perceptron (MLP) and a subsequent linear layer. Our network output is scaled to account for the SI signal power of $P_y = -15$ dBm (due to passive isolation) such that the network internally operates around 0 dBm level to ease training conditions. Following the SIC, the residual $r(t)$ passes through the receiver LNA with gain α and finally the receiver ADC. It becomes available as $r[k]$, the digital residual driving model optimization.

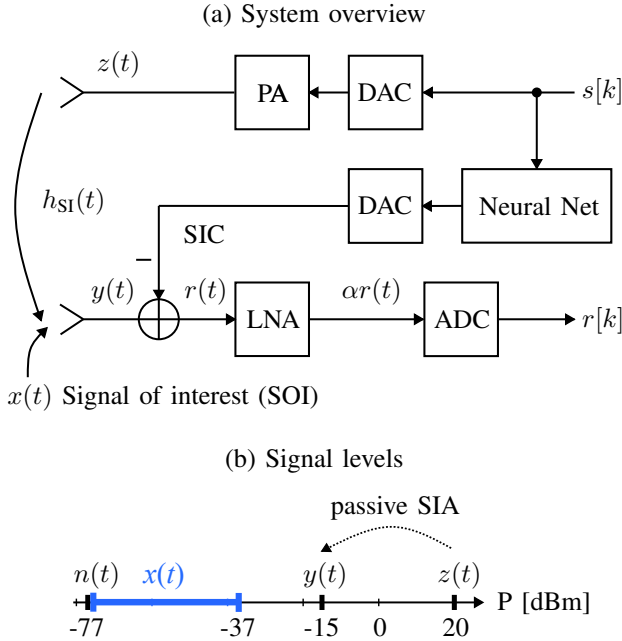


Fig. 1: IBFD transceiver model with digitally-assisted SIC.

For generating data, we employ the procedure for simulating a time-invariant SI system detailed in [11]. In particular, $s[k]$ simulates orthogonal frequency division multiplexing (OFDM) transmission according to the WLAN High-Throughput (HT) standard IEEE-802.11 [15] in complex-baseband representation with 20MHz bandwidth, 16-QAM, and coderate 3/4.

Operation of the presented transceiver system, including analog SIC, is simulated in software. To this end, the analog

quantities $z(t)$, $h_{SI}(t)$, $y(t)$ and $r(t)$ are approximated digitally as $z_a[k]$, $h_{a,SI}[k]$, $y_a[k]$ and $r_a[k]$, respectively, using 64-bit floating-point representations. In contrast, the receiver ADC, i.e., the primary concern in this work, uses a maximum resolution of $B = 12$ bit, thus enabling meaningful simulation of quantization errors. Once the signal has passed the ADC, we return to 64 bit word length for further processing.

Quantization and saturation of the receiver ADC are applied individually to the I/Q components of the complex baseband signal using the mid-rise step-function

$$\text{ADC}\{\alpha r_{a,IQ}[k]\} = \Delta \cdot \text{round}\left(\frac{g(\alpha r_{a,IQ}[k]) + \Lambda}{\Delta}\right) - \Lambda, \quad (2)$$

with Λ the absolute saturation threshold, $\Delta = 2\Lambda/(N - 1)$ the quantization stepsize, $N = 2^B$ the number of quantization steps for B bits, and $g(\cdot)$ a clipping function

$$g(x) = \begin{cases} -\Lambda, & x \leq -\Lambda \\ x, & -\Lambda < x < \Lambda \\ \Lambda, & x \geq \Lambda. \end{cases} \quad (3)$$

Note that in (2), the shift by Λ and $-\Lambda$ before and after rounding, respectively, is used to express a mid-rise as opposed to a mid-tread quantization step-function.

III. SIMULATED SIC MODEL OPTIMIZATION WITH ADC

We now turn our attention to the optimization of the neural network part in the context of the entire system in Fig 1. Due to our all digital simulation of the system, backpropagation would theoretically be able to traverse the simulated ADC layer with its digital input and output. In order to still account for a realistic scenario, i.e., with the analog ADC input and

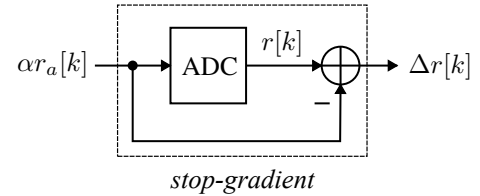


Fig. 2: "ADC Noise" block by "stop-gradient" encapsulation.

the ADC layer inaccessible to the backpropagation algorithm, we employ *Tensorflow's* [16] *stop-gradient* (SG) function, as illustrated in Fig. 2. Specifically, we encapsulate computation of the digitization error as

$$\Delta r[k] = \text{SG}(\text{ADC}\{\alpha r_a[k]\} - \alpha r_a[k]), \quad (4)$$

comprising both quantization noise and saturation error. Our digital residual after ADC is then simulated additively as

$$r[k] = \alpha r_a[k] + \Delta r[k]. \quad (5)$$

In this way, the ADC layer is deliberately isolated from the backwards pass, whereas its distortion effects are indeed provided in the digital residual $r[k]$, thereby realistically influencing the evaluation of gradients during backpropagation. With this technique as a foundation, we consider different strategies to cope with ADC effects during optimization.

A. Backpropagation through ADC (BPAD)

Looking at Fig. 1, backpropagation for minimization of $r[k]$ would intuitively consider the ADC as a nonlinear layer that acts upon the analog residual signal $r(t)$. In practice, this is not possible, since (i) the gradient of the ADC step-function is zero almost everywhere and (ii) the input to the ADC layer, namely $r_a[k]$, is not observable and hence not known.

We may consider partial ADC simulation to at least account for saturation effects, i.e., according to (3), which leads to the arrangement in Fig. 3. The LNA output is passed through a saturation layer $g(\cdot)$, which represents the functional part of the ADC that is within the dynamic range well conditioned for backpropagation. The remaining quantization noise is then added via the previously detailed "ADC Noise" block fed with already saturated input. In this configuration, we can exploit the simplicity of $g(\cdot)$ by designating edge values $r[k] = \pm\Delta$ to represent all ADC inputs outside the dynamic range and, thus, do not require knowledge about $r_a[k]$ to evaluate gradients in the backwards pass. Slight approximation occurs when $\alpha r_a[k]$ resides very closely below the saturation threshold, which, however, is negligible with sufficiently fine quantization.

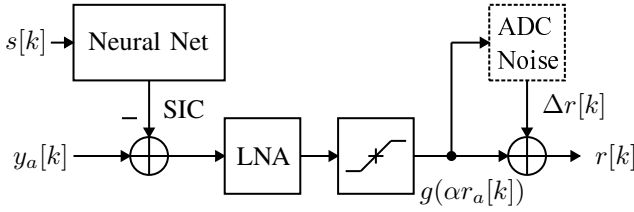


Fig. 3: Setup for simulated training with partial ADC model.

During backpropagation, this approach hence disregards any samples identified to fall outside the dynamic range, as they do not contribute to the overall gradient. This may slow down the neural network learning, specifically, in the beginning, the stronger the saturation by the SI input.

With partial ADC modeling using $g(\cdot)$, a linear approximation of the ADC's step function takes place within the dynamic range. Strictly speaking, this resembles an approximation according to the straight-through-estimation (STE) principle [14] in the field of quantized neural networks. STE approximates a quantized activation function with a continuous proxy function to compute meaningful gradients during backpropagation.

B. Straight-Through Estimation (STE)

The aforementioned STE proxy is frequently chosen as a simple neutral linear function. Therefore, we may also extend the linear STE approximation of just the dynamic range of the ADC to additionally include the saturation range. As shown by Fig. 4, this approach translates to omitting the presence of the ADC entirely during backpropagation, which naturally does not entail any additional implementation effort.

This simplicity, however, comes at the cost of potentially large discrepancy between the actual ADC layer nonlinearity in the forward pass and the simplified proxy in the backward pass, specifically, in the undermodeled saturation range. Yet,

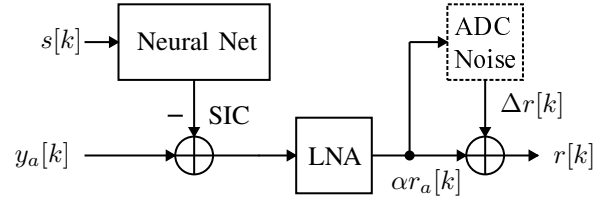


Fig. 4: Setup for simulating STE training.

stable convergence of SIC model learning can be foreseen, since the saturated residual $r[k]$ of the forward pass is only ever equal to or smaller than the $\alpha r_a[k]$ input to the ADC and, hence, this type of STE generally yields gradients smaller but never larger than without approximation. We therefore expect this STE approach to cope well with initial saturation of the ADC and to not critically impede the desired SIC learning. Moreover, with optimization progress and successful SIC, the STE approximation in the saturation range will diminish.

C. Automatic Gain Control (AGC) based strategy

If a transceiver is equipped with AGC in place of the plain LNA, as shown by Fig. 5, that alone might be enough to avoid initial saturation. Moreover, if the AGC continuously holds an appropriate signal level throughout optimization, even as the residual $r_a[k]$ becomes very small via SIC, quantization errors are mitigated, potentially allowing for better performance compared to a system without AGC.

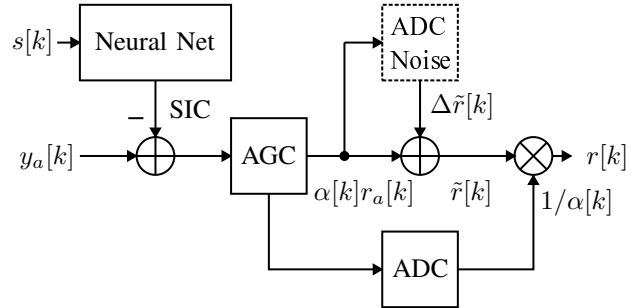


Fig. 5: Setup for simulating training with AGC.

When operating an AGC during model learning, however, backpropagation needs to be informed of the time variable gain $\alpha[k]$ at all times in order to employ a gain-compensated $r[k] = \tilde{r}[k]/\alpha[k]$ instead of the observed residual $\tilde{r}[k]$, such that the gain-compensated residual $r[k]$ suitably diminishes over the course of learning. While AGC systems commonly provide this gain information, usually referred to as received signal strength indicator (RSSI) $\alpha[k]$, an additional ADC is required to observe the quantity for further use on the digital side. For realistic simulation, a similar technique as used for the "ADC Noise" block is thus used for AGC gain quantization, however, not precisely shown in Fig. 5 for the sake of simplicity.

D. Digital-Training Approach (DTA)

As another baseline configuration for avoiding ADC saturation during SIC learning, both LNA and AGC circuitry also can be omitted [5] or the transmission power can be reduced

during SI acquisition (with potential change of PA nonlinearity as compared to actual transmission) [8]. For reference, we here adopt the strategy of using recordings $y[k]$ of $y_a[k]$ without LNA gain to train a model for digital SIC. In a second step the learned model is employed in the digitally-assisted SIC scheme of Fig. 1. The core idea of this procedure is that a model trained on the digital $y[k]$ may still manage to predict the analog $y_a[k]$ reasonably well, perhaps better than the $y[k]$, since there is limited possibility for unintentional modeling of the fine-grained nonlinear distortion of the quantization step-function with our small model capacity.

Fig. 6 depicts the training arrangement, where, different to before, SIC is executed in digital domain. As a consequence, utilization of $r[k]$ during backpropagation is now relieved of any concerns regarding ADC induced distortions. Note that during the recording of training data $y[k]$, the analog SIC in Fig. 1 must be suspended. This also implies that training has to be carried out offline, which may be undesirable.

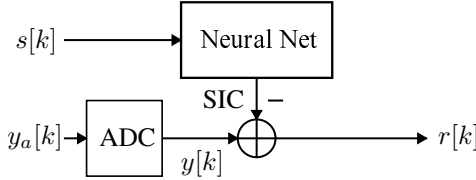


Fig. 6: Setup for digital SIC training.

IV. EXPERIMENTAL EVALUATION

All system configurations are optimized for 6000 epochs using a training corpus generated as described in [11], consisting of 10 signal sequences of length 4096 each, where the transmit signal $s[k]$ is normalized to unit variance. The size of the training corpus is deemed appropriate with respect to the relatively small model capacity. We choose the Adam [17] optimizer with MSE loss and learning rate 0.03. The ADC is simulated with $B = 12$ bits resolution and $\Lambda = 1$ saturation.

A. Learning Behavior and Performance of SIC

The SOI is absent during SIC learning, i.e., $x(t) = 0$ in Fig. 1. For BPAD and STE, the LNA will be considered with different gains $\{30, 40, 50\}$ dB to demonstrate scenarios with different severity of ADC saturation. We employ a frame-wise implementation of the AGC system with frame length $N_{\text{AGC}} = 64$, a square-law detector [18] and a set-point determined experimentally to provide an appropriate signal level that mostly avoids saturation. The range of available AGC gains is configured to $[-20, 40]$ dB to enable proper attenuation of initial SI, while providing for sufficient gain in the later phase of SIC optimization. The ADC used to observe $\alpha[k]$ is configured identically to the primary ADC of the receive path, but with a dynamic range that suits the gain range. For DTA, the LNA is bypassed for avoiding the ADC saturation while recording the training data.

Fig. 7 (a) firstly compares the development of the power of the SIC residual $r_a[k]$ over the course of neural network learning, where BPAD and STE were configured with LNA

gain of 30 dB. The dashed line at -35 dBm indicates the approximate signal power where ADC saturation is just avoided. In reality, as discussed before, $r_a[k]$ is not observable during learning, but the simulation allows its inspection. We observe similar performance of all learning strategies. In all cases, the SIC residual starts around -15 dBm (i.e., the SI level) and falls to around -63 dBm, corresponding to nearly 50 dB digital SI attenuation. Notably, there seems to be no significant advantage of AGC efforts in this SIC analysis.

To better illustrate different sensitivities of BPAD and STE, Fig. 7 (b) further compares the power of $r_a[k]$ over the course of learning for LNA gains $\{30, 40, 50\}$ dB and thus additionally more pronounced effects of (initial) ADC saturation. With LNA=40dB, BPAD firstly indicates a somewhat decelerated learning process and it fails to converge with LNA gain of 50dB, where a critical number of $r_a[k]$ samples fall outside the modeled dynamic range. Meanwhile, STE is seemingly unimpressed by even the strongest saturation at LNA=50dB, i.e., exhibiting only minor differences across all three gains.

B. Demodulation Performance in Terms of Bit-Error Rate

In order to evaluate the BER, $r_a[k]$ signals, obtained with learned SIC models and applied to signals from a separate test dataset, are mixed with SOI $x_a[k]$ at different signal-to-noise ratios $\text{SNR} = P_{x,a}/P_{n,a}$. The mixed signals are then demodulated using Matlab's WLAN Toolbox¹ to recover the original bit sequences. Where packet detection fails, we assign $\text{BER} = 0.5$. Performance in terms of an averaged BER versus SNR is shown in Fig. 8. The underlying signal-to-inference-and-noise ratios $\text{SINR} = P_{x,a}/P_{r,a}$ after SIC is depicted on the right y-axis. It refers to the SOI power w.r.t. residual SI and noise disturbances before demodulation.

Fig. 8 (a) compares the performance of all system options, where LNA gain of 30 dB is used for BPAD and STE. In accordance with results from Fig. 7 (a), the BER performance turns out to be similar across the different systems. BER values start to fall around $\text{SNR} = 20$ dB based on the underlying $\text{SINR} \geq 10$ dB. The desired $\text{BER} \approx 0$ is achieved with all systems in the operating range of $\text{SNR} = [30, 45]$ dB. Performances of BPAD, STE and DTA systems then quickly degrade again for $\text{SNR} > 45$ dB due to SOI saturation by the ADC, while the AGC system continuous to maintain low BER by adjusting the signal level properly. Note that this advantage is now provided by the AGC mainly operating on the SOI, i.e., not on the SIC as in the former model learning process.

Fig. 8 (b) finally summarizes end-to-end BER performance for a larger 40 dB LNA gain, where system parameters for AGC and DTA are identical to Fig. 8 (a). We observe marginally improved but fundamentally similar performance for BPAD and STE up until $\text{SNR} = 30$ dB. The larger LNA gain, however, causes earlier SOI saturation by the ADC, thus narrowing the operational SNR range with low BER to a smaller window merely around 30 dB. Note that high BER at large SNR is avoided by the additional effort with AGC.

¹Demodulation functions from <https://de.mathworks.com/help/wlan/ug/802-11n-packet-error-rate-simulation-for-2x2-tgn-channel.html>

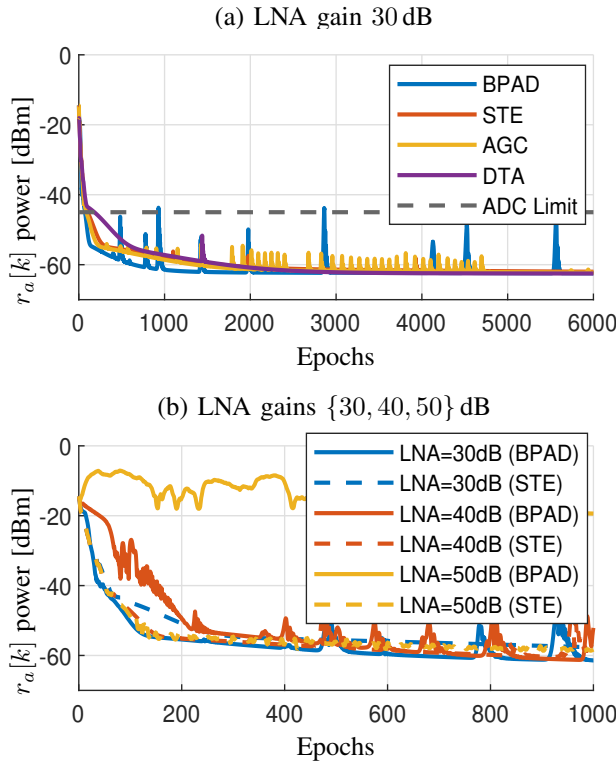


Fig. 7: Learning behavior of SIC for various LNA gains.

V. CONCLUSIONS

The learning of a digital system function for analog (i.e., digitally-assisted) SIC rests upon the residual in the digital domain and is complicated by the ADC function in between. This study presented a range of viable system options regarding the modeling of the ADC with different amenities and limitations: an intuitive yet fragile backpropagation-through-ADC (BPAD) approach, a surprisingly simple and robust straight-through-estimation (STE) technique, a more costly AGC-based system, and an offline digital-training-based approach (DTA). For low system complexity and online adaptivity of SIC systems, the STE approach turns out favorable in our simulation.

REFERENCES

- [1] M. Heino *et al.*, “Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays,” *IEEE Comm. Mag.*, vol. 53, no. 5, pp. 91–101, 2015.
- [2] K. Kolodziej, B. Perry, and J. Herd, “In-band full-duplex technology: Techniques and systems survey,” *IEEE Trans. Microwave Theory Techn.*, vol. 67, no. 7, pp. 3025–3041, 2019.
- [3] B. Smida, A. Sabharwal, G. Fodor, G. Alexandropoulos, H. Suraweera, and C. Chae, “Full-duplex wireless for 6G: Progress brings new opportunities and challenges,” *IEEE Jnl. Selected Areas Commu.*, vol. 41, no. 9, pp. 2729–2750, 2023.
- [4] D. Korpi, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, “Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3821–3836, 2014.
- [5] A. Kiayani, L. Anttila, and M. Valkama, “Active RF cancellation of nonlinear TX leakage in FDD transceivers,” in *IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, 2016, pp. 689–693.
- [6] A. Kiayani *et al.*, “Adaptive nonlinear RF cancellation for improved isolation in simultaneous transmit–receive systems,” *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 5, pp. 2299–2312, 2018.

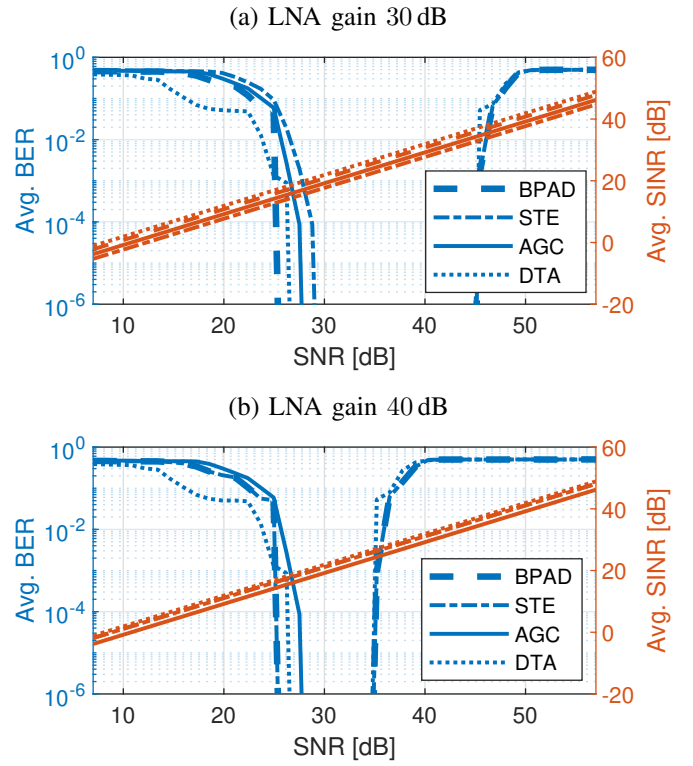


Fig. 8: Demodulation performance for various LNA gains.

- [7] Y. Liu, X. Quan, W. Pan, and Y. Tang, “Digitally assisted analog interference cancellation for in-band full-duplex radios,” *IEEE Communications Letters*, vol. 21, no. 5, pp. 1079–1082, 2017.
- [8] J. W. Kwak, M. S. Sim, I.-W. Kang, J. S. Park, J. Park, and C.-B. Chae, “A comparative study of analog/digital self-interference cancellation for full duplex radios,” in *IEEE Asilomar Conf. on Signals, Systems, and Computers*, 2019, pp. 1114–1119.
- [9] T. Riihonen and R. Wichman, “Analog and digital self-interference cancellation in full-duplex MIMO-OFDM transceivers with limited resolution in A/D conversion,” in *IEEE Asilomar Conf. on Signals, Systems and Computers*, 2012, pp. 45–49.
- [10] J. Xing, S. Ge, Y. Liu, Z. Cui, and J. Meng, “Comprehensive analysis of quantization effects on digital-controlled adaptive self-interference cancellation system,” *IEEE Access*, vol. 8, pp. 75 772–75 784, 2020.
- [11] G. Enzner, A. Chinaev, and A. Sezgin, “On neural-network representation of wireless self-interference for inband full-duplex systems,” in *European Signal Processing Conference*, 2025.
- [12] A. Kristensen, A. Burg, and A. Balatsoukas-Stimming, “Advanced machine learning techniques for self-interference cancellation in full-duplex radios,” in *Asilomar Conf. Signals, Systems, and Computers*, 2019, pp. 1149–1153.
- [13] Y. Bengio, N. Léonard, and A. Courville, “Estimating or propagating gradients through stochastic neurons for conditional computation,” *arXiv preprint arXiv:1308.3432*, 2013.
- [14] P. Yin, J. Lyu, S. Zhang, S. Osher, Y. Qi, and J. Xin, “Understanding straight-through estimator in training activation quantized neural nets,” *arXiv preprint arXiv:1903.05662*, 2019.
- [15] I. Qureshi and S. Asghar, “A systematic review of the IEEE-802.11 standard’s enhancements and limitations,” *Wireless Personal Comm.*, vol. 131, no. 4, pp. 2539–2572, 2023.
- [16] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, “Tensorflow: a system for large-scale machine learning,” in *USENIX Symposium on Operating Systems Design and Implementation*, 2016, pp. 265–283.
- [17] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *Intl. Conf. Learning Representations*, vol. abs/1412.6980, 2014.
- [18] D. Whitlow, “Design and operation of automatic gain control loops for receivers in modern communications systems,” *Microwave Journal*, vol. 46, no. 5, pp. 254–256, 2003.