This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

Multi-Channel Nonlinearity Mitigation Using Machine Learning Algorithms

Haotian Zhao, Julian Camilo Gomez Diaz, *Member, IEEE* and Sebastian Hoyos, *Senior Member, IEEE*

Abstract—This paper investigates multi-channel machine learning (ML) techniques in the presence of receiver nonlinearities and noise, and compares the results with the single-channel receiver architecture. It is known that the multi-channel architecture relaxes the sampling speed requirement of analog to digital conversion and provides significant robustness to clock jitter and front-end noise due to the bandwidth-splitting property inherent in these receivers. However, when a high-voltage swing signal is used in a wireline communication link, the received signal suffers from third-order harmonic distortions and inter-modulation products caused by the nonlinearity profile of the analog front-end (AFE). To this end, this paper proposes the channel decision passing (CDP) algorithm in combination with nonlinear feedback cancellation as a low-complexity candidate for nonlinearity mitigation and compares the performance of this solution with other well-known ML algorithms. Simulation results show significant improvement in a multi-channel receiver architecture equipped with nonlinear feedback cancellation and CDP in comparison with its single-channel counterpart under practical nonlinearity profiles and noise conditions.

Index Terms—Multi-channel receiver, Nonlinearities, Machine learning, Supervised learning, Unsupervised learning, Reinforcement learning.

1 Introduction

The exponentially growing demand for high-rate data communication drives wireless and wireline transceivers to operate with a wider bandwidth. Multi-channel receiver architectures are becoming increasingly popular among receiver designs considering the improvement on spectral efficiency by dividing the available channel bandwidth into a number of subchannels with fractional bandwidth, such that the high data rate can be achieved by configuring multiple data modulation formats with subcarriers in a multi-tone or multi-band transmission [1]. Moreover, the sampling speed requirement of analog-to-digital converters (ADCs) can be relaxed by the number of parallel channels compared to the traditional ADC-based baseband receiver. Therefore, the impact of inter-symbol interference (ISI) can be greatly reduced due to the extension of the symbol time, which in turn simplifies the equalization design [2].

While scaling of CMOS technology improves the speed and power consumption of integrated circuits, it leads to the degradation of the linearity due to the reduced power supplies [3]. The harmonic distortions and the inter-modulation products caused by the nonlinearity characteristic of the analog front end (AFE) lead to the degradation of the link bit-error-rate (BER) performance. As higher-order modulation formats are pursued to increase communication rate, nonlinearities will eventually become the bottleneck of the transmission capacity [4].

Considering the challenges of improving the linearity by analog means including cost, area footprint, and power consumption [5], [6], digital nonlinearity mitigation techniques offer great flexibility and adaptability. Previous digital post-processing algorithms can be divided into two basic categories. Those that aim to invert the nonlinearity effects and those that try to cancel them out. In [7], [8], [9], the adaptive feedforward nonlinearity cancellation structure is proposed. This technique uses a linear reference receiver with least mean square (LMS) filters to adaptively reproduce the nonlinearity distortions, which are then subtracted from the main receiver. However, the requirement of a more linear reference path than the main path limits its application to multi-channel receivers, especially when considering the increasing number of channels. A post-compensation algorithm inverting the nonlinearity effect is presented in [10], [11], [12]. This approach firstly identifies the nonlinear system using either Hammerstein's model [13], [14], [15] or Volterra series [16], [17], and then performs the p-th inverse calculation [18], [19] or Richardson iteration [20]. However, this method uses part of the nonlinearity distortion from the high-frequency bandpass filters, thus inherently giving only approximate distortion estimates, while the in-band distortions still degrade the quality of the received signal.

Machine Learning (ML) and Deep Learning (DL) have been applied in various areas as powerful tools for pattern recognition and modeling of complex systems. Recently, nonlinearity mitigation techniques based on ML&DL algorithms have been widely used in single-channel communication systems for channel estimation, and symbol detection [21], [22], [23], [24], [25], [26]. In general, machine learning algorithms fall into three broad categories. Supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the machine is provided with training data. The supervised learning algorithms analyze the training data and produce correct output from the associated labels. Supervised learning algorithms applied to nonlinearity mitigation include k-nearest neighbors (KNN) [27], [28], [29] and support vector machine (SVM) [30], [31]. In contrast

- *Haotian Zhao was with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843. E-mail: haotian.zhao@tamu.edu.*
- *Julian Camilo Gomez Diaz was with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843. E-mail: juliancgd@tamu.edu.*
- *Sebastian Hoyos was with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843. E-mail: hoyos@tamu.edu.*

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

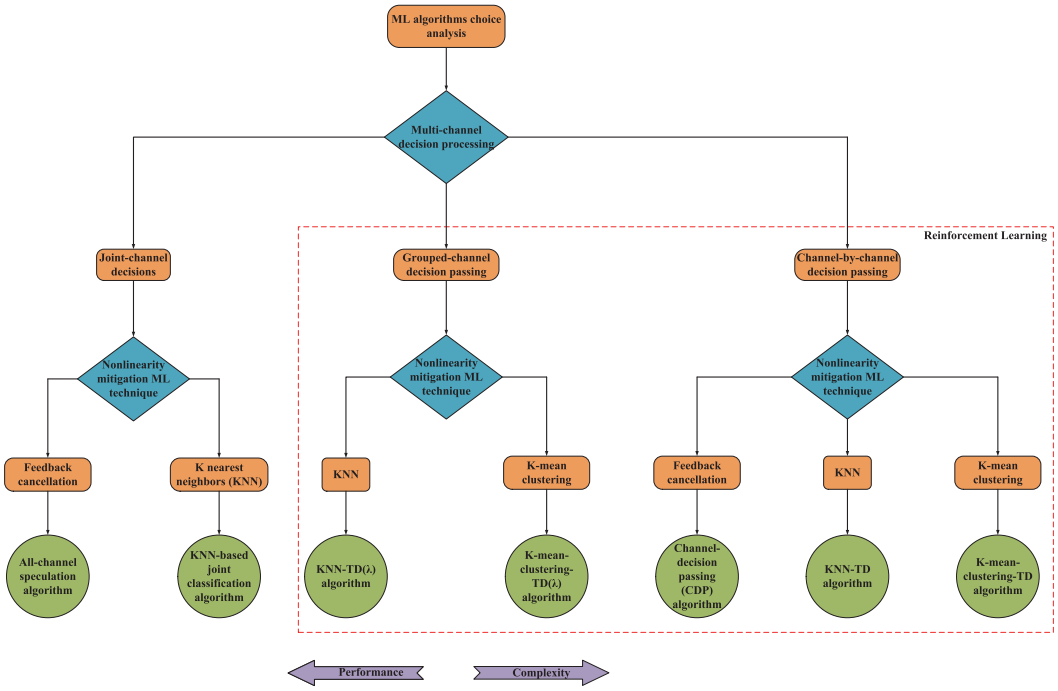


Fig. 1: Chart of ML-based multi-channel nonlinearity mitigation algorithms arranged by trade-off ranking of performance versus complexity.

with supervised learning, unsupervised learning [32] algorithms act on unsorted data according to its similarities, patterns, and differences of it without the involvement of training data. Typical unsupervised learning algorithms include k-mean clustering and hierarchical clustering. Although overlooked for a long time, reinforcement learning (RL) is starting to gain attention after multiple successful applications such as Google's DeepMind [33]. In reinforcement learning, the machine receives feedback by interacting with the environment and learns to choose actions that maximize the reward. Traditional reinforcement learning algorithms include Monte-Carlo learning [34], and Temporal-Difference (TD) learning [35]. Deep Learning is a specific subset of Machine Learning, which works on the principle of back-propagation to find the gradient loss across the layers of a neural network. Based on the implementation style of the neural network, DL can be supervised learning, unsupervised learning, or reinforcement learning. Multiple neural network structures have been applied for alleviating the nonlinearity effect of the receiver including multi-layer perceptron (MLP) [36], convolutional neural network (CNN) [37] and recurrent neural network (RNN) [38], [39].

This paper investigates the applicability of machine learning algorithms to the nonlinearity mitigation of ADC-based multi-channel receivers, which has not been widely investigated in the research field. A two-step consideration for possible combinations of techniques is shown in Fig. 1. Firstly, channel-decision processing is taken into account, which can be divided into joint-channel decisions, grouped-channel decision passing, and channel-by-channel decision passing. For joint-channel decisions, the decisions of the transmitted symbols from all channels are made simultaneously. For grouped-channel passing, the channels are split into several groups. The decision for a group of channels will be made, and the information will be passed on to the following groups to help them make their decisions. Similarly, for channel-by-channel decision passing, the decision of the transmitted symbol will be made channel by channel and passed to the following channel. It is self-evident that from channel-by-channel decision passing to joint-channel decisions, the performance will increase since joint-channel decisions aim for an exhaustive search of optimal solutions at the expense of increased complexity. Besides, note that the nonlinearity mitigation process implemented in a decision-passing fashion can be properly modeled by the Markov Decision Process (MDP) [40], [41]. Thus, the nonlinearity mitigation algorithms developed with a decision-passing technique need to comply with the RL paradigm.

For each channel-decision operation method, the nonlinearity mitigation can be implemented with different ML-based strategies. KNN and K-mean clustering algorithms are considered candidates for their simplicity as they do not require knowledge of the nonlinearity model. As an unsupervised learning algorithm, K-mean clustering can directly operate on the received data without the guidance of the training sequence for classification. However, the K-mean algorithm BER performance suffers when the received samples interfuse with other constellation points caused by nonlinearities. Such performance degradation can be improved by using a training sequence like in the KNN algorithm. Both KNN and K-mean clustering do not have any involvement of nonlinearity cancellation, which can bring performance limitations. To stress this, with a known nonlinearity model, the nonlinearities of the received samples can be effectively removed through a feedback cancellation scheme to generate soft decisions.

Based on the principle of this design flow, we extend on an algorithm previously proposed by the authors [4] named the channel speculation passing (CSP) algorithm. The CSP algorithm is modified to make the decisions not only using speculation based on the smallest Euclidian distance but instead using ML-based classification. We call this modified algorithm channel decision passing (CDP). The CDP algorithm iteratively makes decisions on the output symbols of each channel, which are then used to

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

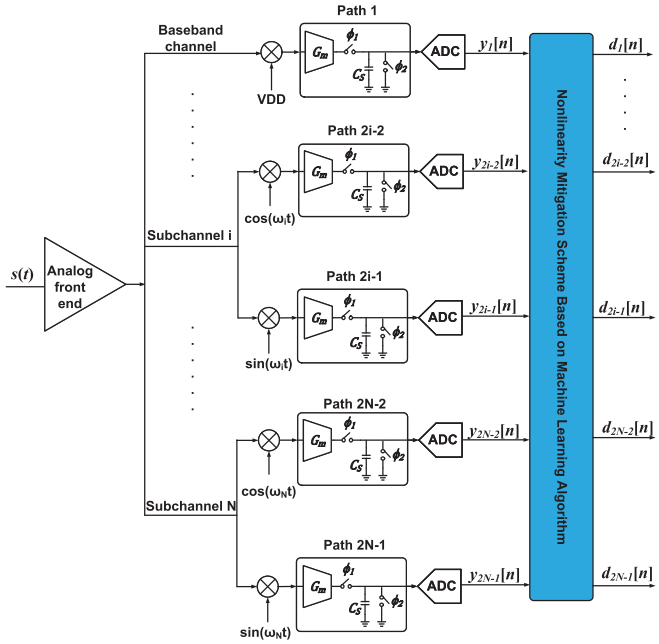


Fig. 2: Block diagram of the multi-channel receiver.

cancel out the nonlinearities of the following channels until the outputs from all channels converge to a stable state. A detailed comparison regarding BER performance and time complexity is made between KNN, K-mean clustering, and CDP algorithms. Fig. 1 organizes in a chart the possible multi-channel nonlinearity mitigation methods and their expected trade-off ranking in terms of performance versus complexity. A combination of KNN/K-mean clustering with TD(λ) learning is considered to be able to mitigate nonlinearities without knowing the nonlinearity model, where λ is a parameter controlling if the channel-by-channel decision passing ($\lambda = 1$) or grouped-channel decision passing ($\lambda \in (0,1)$) is used.

The remainder of this paper is organized as follows. The system architecture of the multi-channel receiver is presented in section II, and the signal-to-noise and distortion ratio (SNDR) improvement of the multi-channel receiver over baseband single-channel receiver is summarized. The nonlinearity mitigation methods based on multiple machine learning algorithms are introduced in section III, comparisons between machine learning algorithms from the simulation are analyzed in section IV, and finally, section V concludes the paper.

2 Multi-channel Receiver Architecture Advantages

The architecture of the N-channel ADC-based multi-channel receiver (MCRX) is shown in Fig. 2, which is designed to detect a multi-band signal. The baseband channel is a single path configured to detect a pulse amplitude modulation (PAM) signal, whereas the remaining subchannels $i = 2,3,...,N$ have both I&Q paths configured to detect a quadrature amplitude modulation (QAM) signal. The multi-band signal $S(t)$ to the input of the multi-channel receiver can be expressed as:

$$s(t) = a_0(t) + \sum_{i=1}^{N-1} \mathrm{Re}\{c_i(t)e^{-j\omega_i t}\} \tag{1}$$

where $a_0(t)$ is the PAM signal, and $c_i(t)$ are QAM signals that are up-converted and transmitted through the higher frequency sub-bands with center frequency $\omega_i = 2\pi f_i$. The QAM constellation points can be expressed as $c_i(t) = a_i(t) + jb_i(t)$, where $a_i(t)$ and $b_i(t)$ are PAM signals corresponding to the in-phase and quadrature component of the QAM signal. This multi-band signal passes through the analog front-end (AFE), mixer, integrator, and finally the ADC before becoming the input to the ML unit, which is designed to mitigate the nonlinearity distortion introduced by the AFE.

The proposed architecture has multiple advantages over the traditional baseband single-channel receiver (SCRX), specifically considering the effects of noise, clock jitter [42], phase noise, and nonlinearity. Fig. 3 shows the main noise and distortion sources in the signal and clock paths of the multi-channel receiver. The SNDR improvement provided by the multi-channel architecture will be discussed in detail as followed.

2.1 Analog Front-end Noise

At the input of the receiver, the noise generated by the termination resistor and the input-referred noise of the AFE can be modeled as a white power spectral density (PSD) $\overline{v_{n,in}^2} = E[v_{n,in}^2]$, where $E[]$ is the expected value. This noise is shaped by the channel-loss compensation magnitude response, $H_{MC}(f)$, and is added with the mixer's thermal noise with white PSD given by $\overline{v_{mixer}^2}$. This can be expressed as:

$$\overline{v_{n,AFE}^2}(f) = \overline{v_{n,in}^2}|H_{MC}(f)|^2 + \overline{v_{mixer}^2} \tag{2}$$

After down-conversion, the multi-band signal passes through the integrator of each data-transmission path. As shown in Fig. 2, the integrator is implemented by a sinc-type filter with main lobe bandwidth equalling to the inverse of integration time T_s, resulting in a split channel bandwidth $\frac{BW}{2N-1} = \frac{1}{T_s}$, where BW is the original signal bandwidth. Assuming a double-sided power spectral densities, the variance of the total integrated noise of the signal at the output of the baseband channel is:

$$\sigma_{MCRX_1}^2 = \int_{-\infty}^{+\infty} \overline{v_{n,AFE}^2}(f) \left| \frac{G_m}{C_S} T_s \; sinc\left(\frac{\pi f}{f_s}\right) \right|^2 df \tag{3}$$

where $f_s = \frac{1}{T_s}$. G_m is the transconductance of the operational transconductance amplifier (OTA) and C_S is the capacitance of the sampling capacitor as shown in Fig. 2. The other subchannels $i = 2,3,...,N$, have a combined I&Q output total integrated noise variance that can be found by integrating the sinc filtered down-converted double-sided power spectral density in the frequency domain:

$$\sigma_{MCRX_i}^2 = \int_{-\infty}^{+\infty} \overline{v_{n,AFE}^2}(f - f_i) \left| \frac{G_m}{C_S} T_s \; sinc\left(\frac{\pi f}{f_s}\right) \right|^2 df \tag{4}$$

thus, I&Q paths have the same noise variance.

From Eqns. 3 and 4, each channel of the multi-channel receiver sees a reduced bandwidth with respect to the single-channel receiver operating with the same overall transmission bandwidth, then the signal-to-noise ratio (SNR) of the multi-channel receiver will be improved compared to the single-channel receiver, which can be expressed as:

$$\sigma_{SCRX}^2 = \int_{-\infty}^{+\infty} \overline{v_{n,in}^2}|H_{SC}(f)|^2 df \tag{5}$$

where $H_{SC}(f)$ is the channel-loss compensation frequency response of the single-channel receiver analog front-end.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

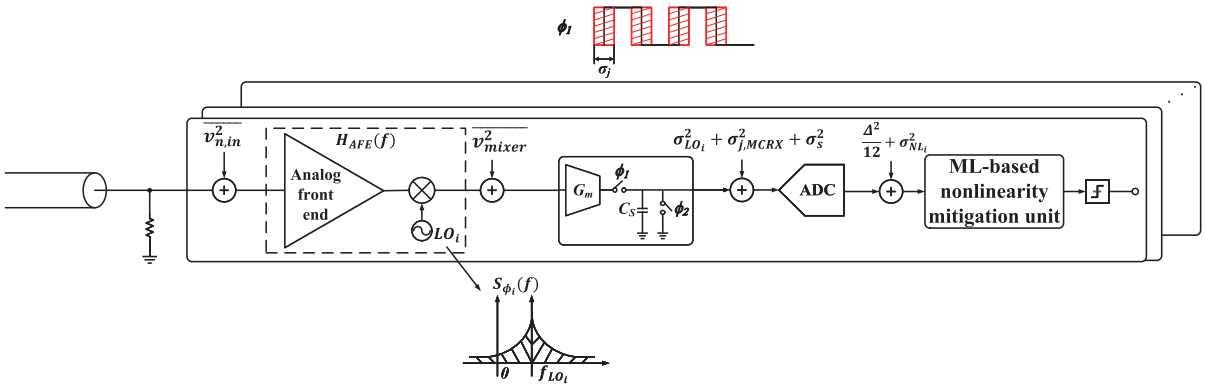


Fig. 3: Noise & distortion sources in a multi-channel receiver.

If we consider the ideal scenario when the ISI channel is lossless, with the input peak-to-peak voltage swing (V_{pp}), the transfer function of the single-channel receiver's analog front-end is required to be $|H_{SC}| = \frac{V_{FS}}{V_{pp}}$ bringing the single-band signal to the full-scale sampling range V_{FS} of the ADC. In order to keep the same V_{pp} for both single-band signal and multi-band signal, the voltage swing on each subband of the multi-band signal needs to be decreased by a certain value. It is worth mentioning, this multi-band signaling remains orthogonal at a minimum channel separation $\delta F = \frac{1}{T_s}$, which requires the center frequency to be a multiple of the subband bandwidth such that $\omega_i = \frac{2\pi i}{T_s} (i = 1, ...N-1)$ in Eqn. 1. Based on this condition, we verify through simulation that the maximum value of the multi-band signal at the transmit side is obtained at instant $t_\Phi = \frac{T_s}{5(N-1)}$. Assuming each carrier has a peak voltage amplitude A, then the multiband signal's peak-to-peak voltage is expressed as:

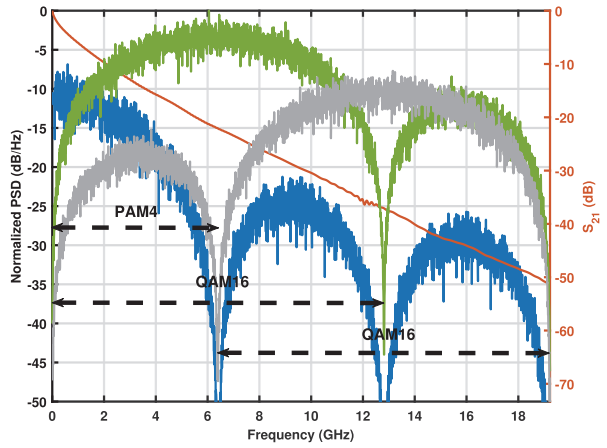
$$V_{pp}(N) = 2A\left[1 + \sum_{i=1}^{N-1}\left(cos\left(\frac{2\pi i}{5(N-1)}\right) + sin\left(\frac{2\pi i}{5(N-1)}\right)\right)\right] \tag{6}$$

Thus, the multi-channel receiver analog front-end requires a transfer function with flat gain $|H_{MC}(N)| = \frac{V_{FS}}{2A}$ that brings each sub-band signal to V_{FS} producing noise amplification. Then, Eqns. 3, 4 and 5, can be simplified, and the signal-to-noise ratio of the single-channel receiver and multi-channel receiver can be expressed as:

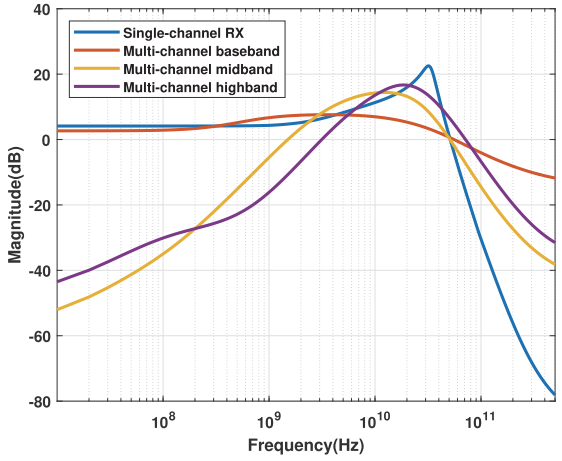
$$\begin{aligned} SNR_{SC} &= \frac{P_s}{\overline{v_{n,in}^2} BW |H_{SC}|^2}, \\ SNR_{MC} &= \frac{P_s}{\overline{v_{n,in}^2} \frac{BW}{2N-1} |H_{MC}(N)|^2}. \end{aligned} \tag{7}$$

Here we neglect the mixer's thermal noise, considering only the input-referred additive white Gaussian noise (AWGN) with bandwidth BW, and assuming $\frac{G_m}{C_S} T_s = 1$ in Eqns. 3 and 4.

By evaluating Eqns. 6 and 7, note that the multi-channel receiver shows an SNR penalty compared to the single-channel receiver for an ideal lossless inter-symbol-interference (ISI) channel. However, in a practical scenario where the AFEs are used to compensate for the ISI channel loss, the input-referred noise will exhibit a frequency-dependent amplification across bands leading to an SNR improvement of the multi-channel receiver with respect to the traditional baseband single-channel receiver. To illustrate this, we run circuit simulations to evaluate a single-channel PAM-4 receiver with 32GHz bandwidth and a 3-channel receiver configured to detect a PAM-4 signal in baseband, and two QAM-16 signals in mid-band and high-band channels. These setups achieve a 64 Gigabit per second (Gbps) transmission



(a)



(b)

Fig. 4: (a) Multi-band signal using multi-carrier signaling and a wireline ISI channel example (30dB loss at 16GHz), (b) Analog-front end frequency response of single-channel RX and multi-channel RX.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

rate [43]. The baseband PAM-4, mid-band QAM-16, and high-band QAM-16 modulation format combination of the 3-channel receiver can provide uniform performance across all channels since an effective PAM-4 signal is processed by each I&Q path after down-conversion. As shown in Fig. 4(a) each path has a bandwidth of 6.4GHz, the mid-band signal is centered at 6.4GHz and the high-band signal is centered at 12.8GHz.

For the ideal scenario, where the channel loss is not present, the 3-channel receiver shows a 4.4dB SNR penalty from simulation which is consitent with the derivations in Eqns. 6-7. While considering an ISI channel with 30dB loss at the Nyquist frequency of 16GHz, the AFEs are designed to partially equalize the channel loss showed in Fig. 4(b) resulting in a 3.7dB SNR improvement of the multi-channel receiver compared to the single-channel receiver. A closed-form expression of this SNR improvement is provided at the end of this section.

2.2 Phase Noise in the Down-conversion LOs

The impact of phase noise in the local oscillators (LOs) in a frequency-interleaved (FI) ADC has been studied in [44]. Although this work focuses on evaluating the SNDR of the ADC as a function of the input frequency, with a special interest in the periodicity caused by the different sub-channels of the FI-ADC, the findings in this work are applicable to the proposed receiver. While the multi-channel architecture used in this paper uses sinc filters allowing the derivation of useful closed-form expressions. Additionally, using multi-channel signaling has the advantage that reconstruction of the multiband signal is not needed, instead, each subband is equalized separately to later infer the bits modulated on this subband. In terms of additional noise contributions in the multi-channel receiver, the phase noise of the LOs is translated to baseband and the noise contribution from the LOs is delimited by the integrator bandwidth.

To analyze the impact of phase noise from the LOs, the analysis will be initially carried out for an input tone at frequency f_c and later on is generalized for a multi-tone input. Thus, we have the multiplication of the input tone and the ith channel LO with phase noise $\phi_i(t)$ and denote the output of the mixer voltage swing as A.

$$\begin{aligned} &v_{mixer_i}(t) = Acos(2\pi f_c t)cos(2\pi f_{LO_i} t + \phi_i(t)) = \\ &\frac{1}{2}A\left(cos(2\pi(f_c - f_{LO_i})t - \phi_i(t)) + cos(2\pi(f_c + f_{LO_i})t + \phi_i(t))\right) \\ &\approx \frac{1}{2}Acos(2\pi(f_c - f_{LO_i})t - \phi_i(t)) \end{aligned} \tag{8}$$

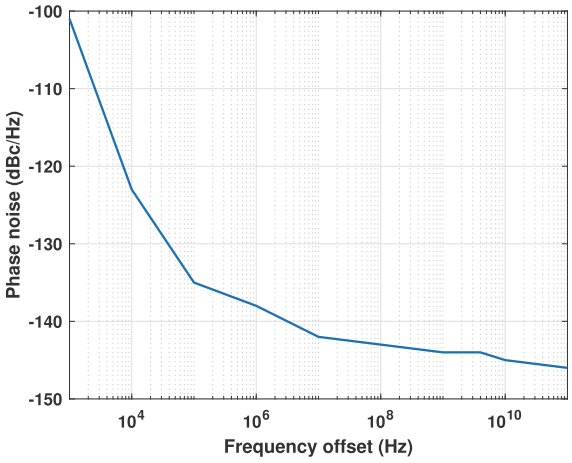
The term at the sums of the frequencies $f_c + f_{LO_i}$ is intentionally discarded as it will be filtered out by the sinc filter. Denoting $f_c - f_{LO_i}$ as f_{IF_i}, we have that the output of the sinc-filter is:

$$\begin{aligned} &v_{mixer_i}(t) \approx \frac{1}{2}Acos(2\pi f_{IF_i} t - \phi_i(t)) \\ &= \frac{1}{2}A\left[cos(2\pi f_{IF_i} t)cos(\phi_i(t)) + sin(2\pi f_{IF_i} t)sin(\phi_i(t))\right] \end{aligned} \tag{9}$$

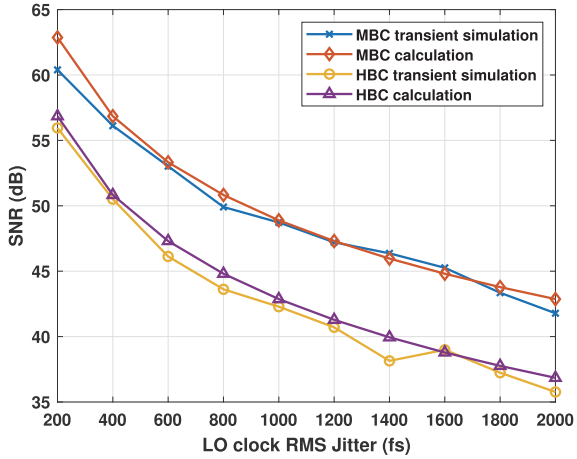
Furthermore, for small phase variations $\phi_i(t) \ll 1$ rad, we can use the small angle approximations to rewrite Eqn. 9.

$$v_{mixer_i}(t) \approx \frac{1}{2}Acos(2\pi f_{IF_i} t) + \frac{1}{2}A\phi_i(t)sin(2\pi f_{IF_i} t) \tag{10}$$

The first term in Eqn. 10 corresponds to the desired downconverted tone with a clean LO whereas the second term is the noise introduced by the phase noise of the LO that is induced at the output of the mixer. This noise is in essence the phase noise of the LO modulated to the IF frequency in the ith channel.



(a)



(b)

Fig. 5: (a) Phase noise profile example, (b) Comparison between theoretical and simulation results of LO-induced phase noise.

$$v_i(t) = \frac{1}{2}A\phi_i(t)sin(2\pi f_{IF_i} t) \tag{11}$$

Since the noise at the output of the mixer will be filtered out by the sinc filter with impulse response $h(t)$ and sampled at time $t = T_s$. At the filter's output, we can express the noise y_i as:

$$y_i = v_i(t) * h(t)|_{t=T_s} \tag{12}$$

where $*$ denotes convolution and the noise variance σ^2_{LO} is the variance of y_i. The random variable y_i is assumed zero-mean and has a PSD given by $S_{y_i}(f)$, therefore the variance of the total integrated noise induced by the LO phase noise is calculated as:

$$\begin{aligned} &\sigma^2_{LO_i} = E[|y_i|^2] \\ &= \int_{-\infty}^{\infty} S_{y_i}(f)df = \int_{-\infty}^{\infty} S_{v_i}(f)\left|\frac{G_m}{C_S}T_s\ sinc\left(\frac{\pi f}{f_s}\right)\right|^2 df \end{aligned} \tag{13}$$

where $S_{v_i}(f)$ is the PSD of the noise $v_i(t)$, which can be derived from Eqn. 11 and expressed as:

$$\begin{aligned} S_{v_i}(f) &= S_{\phi_i}(f) * \left[\frac{A^2}{4}\left(\frac{1}{4}\delta(f-f_{IF_i})+\frac{1}{4}\delta(f+f_{IF_i})\right)\right] \\ &= \frac{A^2}{16}\left(S_{\phi_i}(f-f_{IF_i})+S_{\phi_i}(f+f_{IF_i})\right) \end{aligned} \tag{14}$$

here $S_{\phi_i}(f)$ is the phase noise power spectral density used to characterize the oscillator phase noise.

Now we can generalize our analysis for a multi-tone signal. In this case, the PSD of the noise S_{v_i} at the output of *ith* channel mixer can be expressed as:

$$S_{v_i}(f) = \frac{A^2}{16}\sum_{k=1}^{N}\left(S_{\phi_i}(f-f_{IF_{k,i}})+S_{\phi_i}(f+f_{IF_{k,i}})\right) \tag{15}$$

where $f_{IF_{k,i}} = f_{c_k} - f_{LO_i}$, f_{c_k} denotes the frequency of the *kth* tone.

Considering the frequency response of the sinc filter, the noise variance of the *ith* channel of the multi-channel receiver is:

$$\sigma^2_{LO_i} = \frac{A^2}{16}\int_{-\infty}^{\infty} S_{v_i}(f)\left|\frac{G_m}{C_S}T_s\ sinc\left(\frac{\pi f}{f_s}\right)\right|^2 df \tag{16}$$

Finally, it is worth mentioning that in a practical implementation, the clock used for the LOs, and the sampling clock of the ADC are going to be derived from the same clock source with frequency f_{source} using clock dividers. Therefore, if we assume that noiseless clock dividers are available, the jitter standard deviation of the master clock source will be the same as that of the LOs and also the sampling clock, namely σ_j. Therefore, we have the following relationship [45], [46], [47]:

$$\int_{-\infty}^{\infty} S_{\phi_i}(f)df = (2\pi\sigma_j f_{LO_i})^2 = \left(2\pi\sigma_j\frac{f_{source}}{M_i}\right)^2 \tag{17}$$

where M_i is the division constant that generates f_{LO_i} from f_{source}.

Fig. 5(a) shows the phase noise profile of an oscillator. To verify the above-mentioned phase noise analysis, a transient simulation is run by passing a QAM-16 signal through the mid-band channel (MBC) and high-band channel (HBC) of a 3-channel receiver, respectively. As shown in Fig. 5(b), the simulation results are well-aligned with the derived expressions.

2.3 Jitter-induced Noise at the Sampler

The jitter-induced noise is the major limitation of the baseband single-channel receiver since the ADC needs to sample at the full signal bandwidth (BW) [48]. Thus, for a single-channel receiver sampled by the ADC with full-scale range V_{FS}, the jitter-induced noise is

$$\sigma^2_{j,SCRX} = \left(\frac{V_{FS}}{2\sqrt{2}}2\pi\sigma_j BW\right)^2 \tag{18}$$

In comparison, the ADC on each path of the multi-channel receiver samples the signal with an ideal reduced bandwidth of $\frac{BW}{2N-1}$. The jitter-induced noise variance in each path of the multi-channel receiver is

$$\sigma^2_{j,MCRX} \geq \left(\frac{V_{FS}}{2\sqrt{2}}2\pi\sigma_j\frac{BW}{2N-1}\right)^2 \tag{19}$$

In Eqn. 19, although equality is achieved with ideal low-pass filters, it provides a good approximation that is used in this paper's derivations.

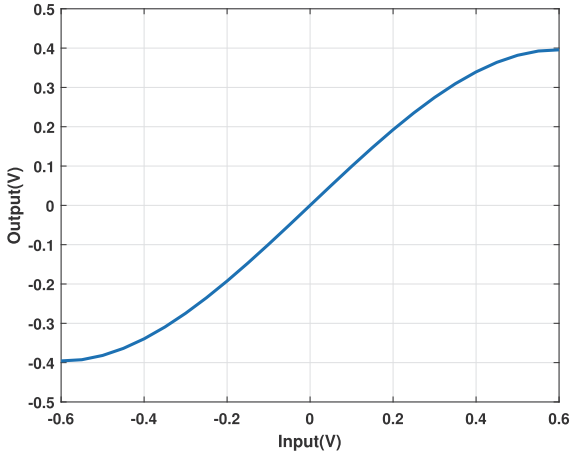
2.4 Sampling Noise

As shown in Fig. 2, the sinc-filters in the multi-channel receiver are realized by the resettable Gm-C filters. The double-sided noise power spectral density produced by the transconductance amplifier given by $2kTG_m$ will be shaped by the transfer function of the sinc filter during the integration phase, and an extra $\frac{kT}{C_S}$ noise coming with the capacitor C_S from the previous reset phase will also contribute to the total sampling noise. Here, k is the Boltzmann constant, and T is the temperature. Then the total integrated sampling noise variance is given by:

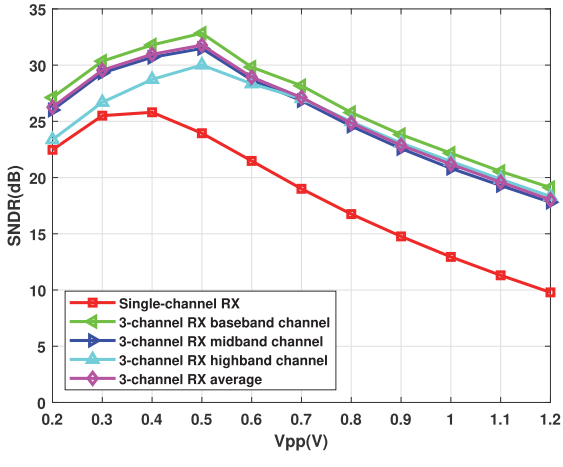
$$\begin{aligned} \sigma_s^2 &= \frac{kT}{C_S} + \int_{-\infty}^{\infty} 2kTG_m\left|\frac{T_s}{C_S}sinc\left(\frac{\pi f}{f_s}\right)\right|^2 df \\ &= \left(1+\frac{2G_mT_s}{C_S}\right)\frac{kT}{C_S} \end{aligned} \tag{20}$$

2.5 Nonlinearity

Compared to the traditional baseband single-channel receiver, the multi-channel architecture also shows robustness to the nonlinearity distortion. Fig. 6(a) shows a static nonlinearity profile of the AFE captured by running a transient simulation, which is used for



(a)



(b)

Fig. 6: (a) Nonlinearity profile, (b) SNDR vs. Vpp comparison between single-channel RX and multi-channel RX.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

both the single-channel receiver and the multi-channel receiver. For a given V_{pp}, the amplitude of the received signal to the input of the multi-channel receiver needs to be scaled down to limit the maximum peak-to-peak voltage. Thus the amplitude of each sub-channel signal of the multi-channel receiver is less than the amplitude of the single-channel receiver. This condition leads to that the multi-channel receiver suffering less from the nonlinearity distortion overall compared to the single-channel receiver.

Fig. 6(b) shows the signal-to-noise and distortion ratio (SNDR) of the single-channel receiver and the multi-channel receiver as a function of input swing. In this simulation, we consider the impacts of $500nV^2/Hz$ input-referred noise, which is shaped by the AFEs' frequency response, and the nonlinearities. With lower input swing, where the SNDR performance is dominated by noise, the multi-channel receiver outperforms the single-channel receiver as discussed in section **2.1**. Moreover, as the input swing increases, the overall linearity of the signal deteriorates, and the SNDR of the single-channel receiver drops faster compared to the multi-channel receiver, resulting in increasing SNDR improvement of the multi-channel receiver.

Now let's use $\sigma^2_{NL,SC}$ to represent the distortion variance of the single channel receiver caused by the nonlinearities, and $\sigma^2_{NL_i}$ to represent the distortion variance caused by the nonlinearities of the ith channel of the multi-channel receiver [49]. From the simulation result shown in Fig. 6(b), we can conclude that $\sigma^2_{NL,SC} > \sigma^2_{NL_i}(i = 1,2,...N)$. Thus, the nonlinearity distortion variance of the single-channel receiver is greater than that of the multi-channel receiver. In the multi-channel receiver, the high-frequency components of nonlinearity distortion falling beyond the baseband bandwidth are partially filtered out by the sinc filter response created by the windowed Gm-C filters, and the remaining components alias back into the Nyquist band after sampling. So, $\sigma^2_{NL_i}$ indicates the in-band nonlinearity distortion variance.

2.6 Sampled-data SNDR Improvement

Now we can derive the overall sampled-data SNDR expression considering the impact of front-end noise, the sampling noise from the Gm-C filter, sampling clock jitter, phase noise of the LO in the ith sub-channel ($i = 1,2,...,N$), the quantization noise from the ADC and the nonlinearity distortion introduced by the AFE:

$$SNDR_{MCRX_i} = \begin{cases} \dfrac{P_{signal}}{\sigma^2_{MCRX_i} + \left(\frac{V_{FS}}{2\sqrt{2}} 2\pi\sigma_j \frac{BW}{2N-1}\right)^2 + \sigma^2_s + \frac{\Delta^2}{12} + \sigma^2_{NL_i}}, & i = 1 \\ \dfrac{P_{signal}}{\sigma^2_{MCRX_i} + \sigma^2_{LO_i} + \left(\frac{V_{FS}}{2\sqrt{2}} 2\pi\sigma_j \frac{BW}{2N-1}\right)^2 + \sigma^2_s + \frac{\Delta^2}{12} + \sigma^2_{NL_i}}, & i \geq 2 \end{cases} \tag{21}$$

where Δ is the quantization interval of the ADCs

By comparison, the single-channel receiver has an SNR with no LO noise contribution but with full-bandwidth jitter-induced noise given by,

$$SNDR_{SCRX} = \frac{P_{signal}}{\sigma^2_{SCRX} + (\frac{V_{FS}}{2\sqrt{2}} 2\pi\sigma_j BW)^2 + \frac{kT}{C_S} + \frac{\Delta^2}{12} + \sigma^2_{NL,SC}} \tag{22}$$

where it has been assumed that the single channel receiver uses an RC sampling network with $\frac{kT}{C_S}$ sampling noise.

By evaluating Eqns. 21 and 22, the sampled-data SNDR improvement given by the multi-channel architecture in comparison with the baseband single-channel receiver sampling at the full bandwidth can be expressed as:

$$SNDR_{improvement} = \frac{\frac{1}{2N-1} \sum_{i=1}^{N} C_i \dfrac{P_{signal}}{\sigma^2_{MCRX_i} + \sigma^2_{LO_i} + \left(\frac{V_{FS}}{2\sqrt{2}} 2\pi\sigma_j \frac{BW}{2N-1}\right)^2 + \sigma^2_s + \frac{\Delta^2}{12} + \sigma^2_{NL_i}}}{\dfrac{P_{signal}}{\sigma^2_{SCRX} + \left(\frac{V_{FS}}{2\sqrt{2}} 2\pi\sigma_j BW\right)^2 + \frac{kT}{C_S} + \frac{\Delta^2}{12} + \sigma^2_{NL,SC}}} \tag{23}$$

where $C_i = 1$ for $i = 1$ and $C_i = 2$ for $i \geq 2$ to correctly account in the average for the I&Q paths of the higher frequency channels. Also, $\sigma^2_{LO_1} = 0$ for the baseband channel of the multi-channel receiver.

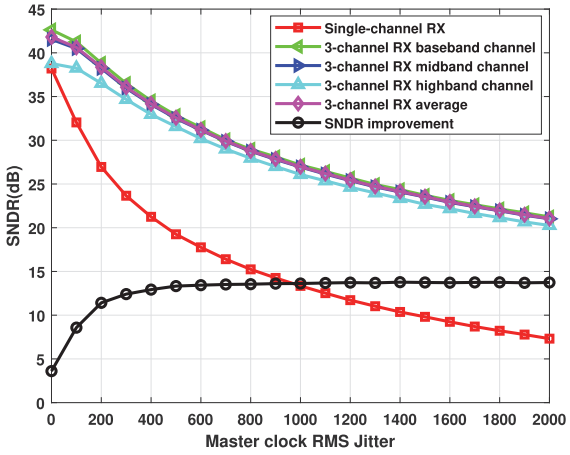


Fig. 7: SNDR vs. Jitter comparison between single-channel RX and multi-channel RX with $500nV^2/Hz$ input-referred noise and linear AFE.

Fig. 7 shows the overall SNDR of a baseband single-channel receiver and multi-channel receiver as the function of the master clock RMS jitter. In this simulation, we assume the analog front-ends are linear. Besides, it is assumed the input-referred noise is shaped by the AFE's frequency response shown in Fig. 4(b) considering the AFEs are used for channel loss compensation. Vpp is set to 1.2V for both the single-channel receiver and the multi-channel receiver and the PSD of the input-referred noise is set to $500nV^2/Hz$. In the absence of jitter-induced noise, the average SNDR of the 3-channel receiver offers 3.7dB improvement over that of the single-channel baseband receiver. Compared to the baseband channel (BBC), the MBC and HBC SNR performance of the multi-channel receiver degrades by the phase noise of the LOs according to Eqn. 17, where the LO frequency of the ith channel f_{LO_i} is $(i-1)\frac{BW}{2N-1}(i = 2,...N)$. As the master clock RMS jitter is increased, the impact of the LOs phase noise is limited by the sinc filter, and the SNDR of each subchannel is dominated by the jitter-induced noise at the sampler, leading to similar SNDRs between subchannels. With the master clock jitter of 2ps, the SNDR improvement provided by the multi-channel architecture is 13.5dB compared with the single-channel baseband receiver.

2.7 Communications System SNDR and SNR Derivations

The SNDR derived so far is defined as the sampled signal power divided by the sum of all the contributions of total integrated noise and distortion variance under the assumption that these

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

distortion sources are uncorrelated to each other. A related signal-to-noise ratio commonly used for bit-error-rate (BER) calculation in wireless/wireline communications systems is $\frac{E_b}{N_o}$. E_b represents the input signal energy per information bit, and N_o is twice the noise power spectral density. Considering the nonlinearity distortion variance σ^2_{NL}, and taking $N_o/2$ as the input referred noise density, the SNDR at the ith channel sampler output for a single tone per channel can be written as:

$$SNDR_i = \frac{P_s}{P_{n_i} + \sigma^2_{NL_i}} \leq \frac{P_s}{P_{n_i}} = \frac{V^2_{FS}/8}{\int_{-\infty}^{+\infty} \frac{No}{2} |H_{MC_i}(f)|^2 df} \tag{24}$$

where $H_{MC_i}(f)$ is the transfer function from the receiver input to the ith channel sampler output that brings the input signal with energy per bit E_b to a full-scale baseband signal with peak-to-peak amplitude V_{FS}. Furthermore, the fraction $\frac{P_s}{P_{n_i}}$ is the sampled-data SNR and can be upper bounded as:

$$\frac{P_s}{P_{n_i}} \leq \frac{\frac{bE_b}{T_s}}{\frac{No}{2} f_{max}} = \left(\frac{2b}{T_s f_{max}}\right) \frac{E_b}{N_o} \tag{25}$$

where b is the number of bits per symbol, T_s is the symbol period, and f_{max} is the maximum signal bandwidth. For most signals, we have $T_s f_{max} = 1$.

Now, we derive the communications signal-to-noise and distortion ratio ($SNDR_{COMMS}$) which represents the SNDR that results after the signal passes through the digital baseband nonlinearity mitigation algorithm, which can be expressed as:

$$SNDR_{COMMS} = \frac{P_s}{P_n + \sigma^2_{NL_{res}}} \tag{26}$$

where $\sigma^2_{NL_{res}}$ represents the residual nonlinearity variance that the nonlinearity mitigation algorithm was not able to remove. Such residual nonlinearity can be modeled as additive noise, allowing to write an expression for the communications SNR as:

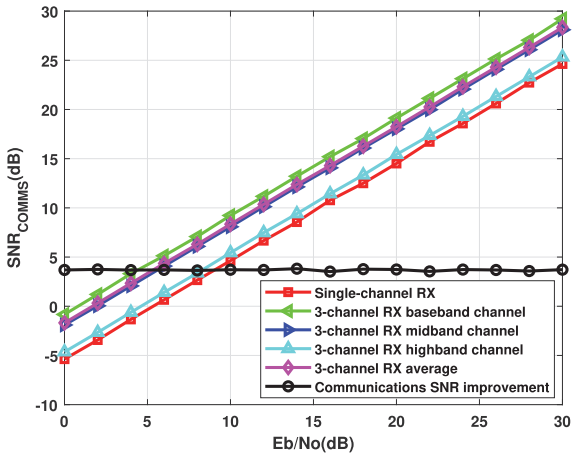
$$SNR_{COMMS} = \left(\frac{1}{2b}\right) SNDR_{COMMS} \leq \frac{E_b}{N_o} \tag{27}$$

Furthermore, the inequality in Eqn. 24 allows writing the average SNR improvement ($\frac{SNR_{MC}}{SNR_{SC}}$) as a function of the corresponding nonlinearities in each case, and assuming the same input $\frac{No}{2}$ and the same full-scale sampling signal power for both receivers as a function of the receivers' transfer functions from the input to the samplers' outputs:

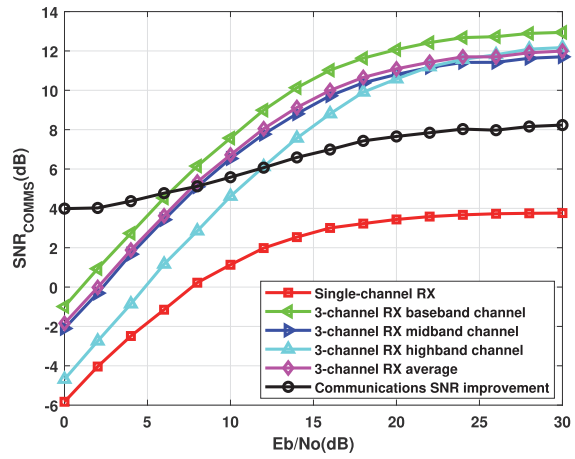
$$SNR_{improvement} = \frac{SNR_{MC}}{SNR_{SC}} = \frac{\int_{-\infty}^{+\infty} \frac{No}{2} |H_{SC}(f)|^2 df + \sigma^2_{NL_{SC}}}{\frac{1}{(2N-1)} \sum_{i=1}^{N} C_i \left[\int_{-\infty}^{+\infty} \frac{No}{2} |H_{MC_i}(f)|^2 df + \sigma^2_{NL_i}\right]} \tag{28}$$

where $C_i = 1$ for $i = 1$ and $C_i = 2$ for $i \geq 2$ as explained for Eqn. 23.

Fig. 8 shows the communication SNR comparison between the single-channel receiver and the multi-channel receiver as a function of Eb/No. In this simulation, we set the sampling clock jitter to zero, and we consider both the linear and nonlinear AFEs where the nonlinear profile is the same one as in Fig. 6(a), which can be approximated by the 3rd order intercept point (IIP3) of 14.5dBm. The frequency response of single-channel receiver and multi-channel receiver's AFEs is shown in Fig. 4(b) to compensate for 30dB loss at the 16GHz Nyquist rate. When the AFEs are assumed to be linear, the input signal is not distorted by the nonlinearities, referring to Eqn. 28, the 3-channel receiver provides 3.7dB communications SNR improvement over the single-channel receiver as shown in Fig. 8(a). While practically, when the AFEs are assumed to be nonlinear, the communications SNR of the receivers are limited by the nonlinearities as shown in Fig. 8(b). The robustness to nonlinearities of the multi-channel receiver as discussed in section **2.5** leads to further communications SNR improvement over the single-channel receiver.



(a)



(b)

Fig. 8: Communications SNR vs Eb/No comparison between single-channel RX and 3-channel multi-channel RX considering 30dB loss at 16GHz Nyquist rate and no jitter in the sampling clock with (a) linear AFE, (b) nonlinear AFE.

The analysis and simulation in this section disregard the impact of residual ISI-channel loss that was not compensated by the continuous-time equalizer (CTLE) added to the AFE. Although any residual ISI can be removed with digital equalization, a noise amplification penalty is incurred that is not included in Eqn. 28. The rest of this paper will focus on the investigation of ML-based nonlinearity mitigation algorithms for multi-channel receivers.

3 Nonlinearity Mitigation Methods

3.1 Nonlinearity Model with Memory Effect

Although the lumped nonlinearities of these building blocks can be modeled by a static profile as shown in Fig. 6(a), such a model does not account for memory effects of earlier inputs induced on

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

current symbols. A practical model that includes memory effects is the Hammerstein model,

$$y_i[n] = \sum_{k=1}^{K} \sum_{q=0}^{Q-1} \alpha_{k,q} s_i[(n-q)T] \, |s_i[(n-q)T]|^{k-1} \tag{29}$$

where the transmitted signal after down-conversion through each path is represented by $s_i[nT]$, Q is the memory depth, K is the order of the nonlinearity polynomial, and $\alpha_{k,q}$ represents the nonlinearity parameter. Considering the even order nonlinearities vanish when fully-differential structures are used, Eqn. (29) can be simplified to:

$$y_i[n] = \sum_{k=1}^{K-1} \sum_{q=0}^{Q-1} \alpha_{2k-1,q} s_i[(n-q)T]^{2k-1} \tag{30}$$

Fig. 6(a) shows the static nonlinearity profile of the system can be closely approximated by using only the 3rd-order nonlinearities. Moreover, considering the limited input swing and the memory effect decays as the memory order increases, we set $K = 3$ and $Q = 2$ in Eqn. 30. As discussed in Section **2.5**, the residual nonlinearity distortion falling in the baseband will be mitigated by the nonlinearity mitigation scheme based on machine learning algorithms.

Fig. 9 shows the histogram of a received PAM-4 signal distorted by noise and nonlinearities. Each distorted symbol corresponds to one of the PAM-4 constellation points, thus the received symbols can be categorized into 4 classes. It can be seen from Fig. 9, due to the noise and nonlinearities, the symbols belonging to the adjacent classes can be interfused with each other. Thus, only setting up the thresholds between classes cannot provide satisfying BER performance. Now, it's clear that the nonlinearity mitigation process of the receiver can be modeled as a classification problem, and ML-based classification algorithms like the KNN algorithm and the K-mean clustering algorithm can be applied to improve the classification accuracy. On the other hand, by partially removing the nonlinearity distortion, the communications SNR of the received symbols can be improved. As a result, the receiver's BER performance can be boosted by passing those symbols with limited residual nonlinearities through the slicer.

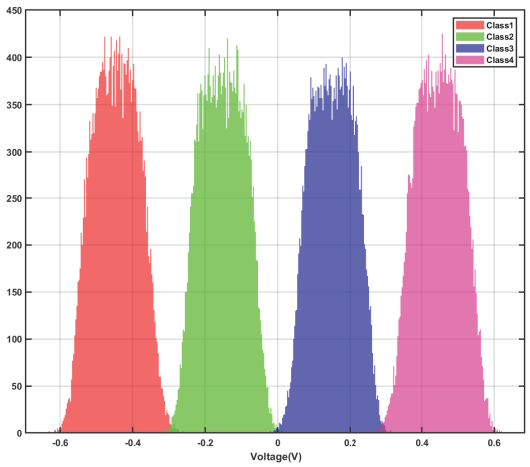


Fig. 9: Histogram of received symbols distorted by noise and nonlinearities.

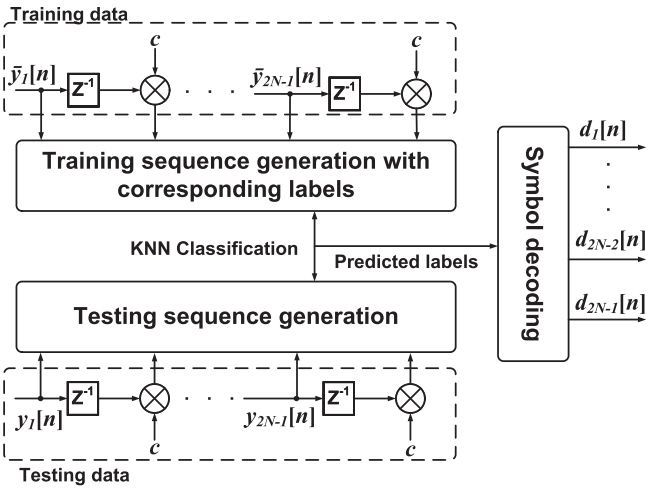


Fig. 10: Scheme flow diagram of KNN-based joint classification.

The implementation detail of the multi-channel nonlinearity mitigation based on the KNN algorithm, the K-mean clustering algorithm, and the channel-decision-passing algorithm will be introduced in the rest of this section.

3.2 K-Nearest Neighbors (KNN) Algorithm

The KNN algorithm is one of the most frequently used ML algorithms in the category of supervised learning. In the KNN algorithm, the received data is divided into a training sequence and a testing sequence. Compared to other machine learning algorithms, for example, SVM and ANNs, which require the training process to obtain the parameters of the model, the classification process of KNN is based on the training sequence only and is independent of any learned parameters, so the testing data can be directly classified. The training sequence $\{(Y_1, l_1), ...(Y_M, l_M)\}$ consists of M instances with the assigned labels $l_j \in L$. An instance in the training sequence Y_j can be described by a $D-$dimensional feature vector: $[a_1(Y_j), ..., a_D(Y_j)]$, where $a_r(Y_j)$ denotes the value of the rth feature of the instance. Given a query instance Y_q in the testing sequence to be classified by the KNN algorithm, the distance between Y_q and an arbitrary instance Y_j in the training sequence can be expressed as:

$$d(Y_j, Y_q) = \sqrt{\sum_{r=1}^{D} (a_r(Y_j) - a_r(Y_q))^2} \tag{31}$$

let $N_k(Y_q) = [Y_1, ..., Y_k]$ denote the set containing k instances that are nearest to Y_q from the training sequence, then the class of Y_q can be predicted by:

$$\hat{f}(Y_q) = \arg\max_{l \in L} \sum_{j \in N_k(Y_q)} \delta(l, l_j) \tag{32}$$

where,

$$\delta(l, l_j) = \begin{cases} 1, & l = l_j \\ 0, & l \neq l_j \end{cases} \tag{33}$$

thus, the majority of labels in $N_k(Y_q)$ will be the predicted label of Y_q.

Referring to Fig. 10, a joint classification over the received symbols through all channels is required when applying the KNN algorithm for nonlinearity mitigation of the multi-channel

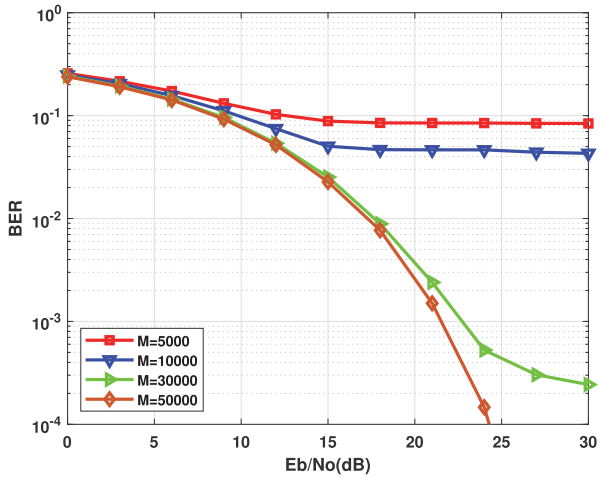


Fig. 11: BER performance of the KNN algorithm as a function of Eb/No for a different number of instances in the training sequence (M) with nonlinear AFE of 14.5dBm IIP3.

receiver. An instance in the training sequence is expressed as (Y_j, l_j), where $Y_j = [\bar{y}_1[n], ..., \bar{y}_{2N-1}[n], c\bar{y}_1[n-1], ..., c\bar{y}_{2N-1}[n-1]]$. $\bar{y}_i[n] (i = 1, 2, ..., 2N-1)$ are the currently-received samples in the training sequence, $\bar{y}_i[n-1]$ are the previously-received samples in the training sequence weighted by c, which is a coefficient used to account for the memory effect according to Eqn. 29. Since the memory effect decays as the memory depth increases, the value of this coefficient c is in the range from 0 to 1. Its optimal value can be found by sweeping from 0 to 1 with a 0.1 step size and picking the one leading to the best BER performance given by the KNN algorithm. $l_j \in L$ is the corresponding label of the instance in the training sequence. Assuming a PAM-4 signal is transmitted in the baseband subband, and QAM-16 signals are transmitted in the other $(N-1)$ passband subbands, then $L = \{1, 2, ..., 4^{2N-1}\}$. The label of an instance in the testing sequence $Y_q = [y_1[n], ..., y_{2N-1}[n], cy_1[n-1], ..., cy_{2N-1}[n-1]]$ is predicted following Eqn. 31 and Eqn. 32. Once the output label is generated, the predicted transmitted symbols of each channel can be decoded through label mapping and used for BER calculation.

The classification accuracy of the KNN algorithm depends on the number of neighbors k and the number of instances M in the training sequence. Based on the simulation results, the variation of k only slightly changes the BER in the regime dominated by noise, whereas M can significantly impact the BER performance as shown in Fig. 11. However, the time complexity of the KNN algorithm is in the order of $O(kDlog(M))$ even by using the KD-tree method, which organizes the training sequence as a binary-tree data structure and allows faster search for nearest neighbors compared to brute force searching method. Here, D is the dimension of the feature vector. This performance-complexity trade-off can preclude the implementation of the KNN algorithm as a fully-integrated nonlinearity mitigation method.

3.3 K-mean Clustering

As a representative unsupervised learning algorithm, the K-mean clustering algorithm can partition the unlabelled data into clusters without requiring the training sequence. Fig. 12 shows the operation process of the multi-channel nonlinearity mitigation applying

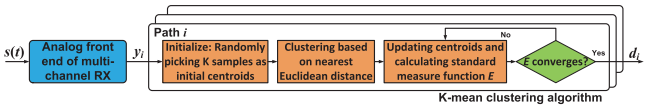


Fig. 12: Scheme flow diagram of K-mean clustering algorithm.

the K-mean clustering algorithm. The output sequence y_i of each path of the multi-channel receiver is past through a K-mean clustering block. Initially, K samples are randomly selected from y_i as the centroids assuming a PAM-K signal is transmitted through the path. Secondly, based on the nearest Euclidean distance principle, each sample is assigned to the corresponding cluster represented by the centroid. Thirdly, the centroids are updated by calculating the mean value of the clusters. Finally, the second and third steps will be iteratively operated until the standard measure function converges, which is expressed by:

$$E = \sum_{k=1}^{K} \sum_{i=1}^{I_k} d(x_i^k, c_k)^2 \tag{34}$$

where I_k is the number of samples in kth cluster, c_k is the centroid of kth cluster, and $d(x_i^k, c_k)$ is the Euclidean distance of the ith sample in the kth cluster to the corresponding centroid.

With a low value of K, the K-mean clustering shows fast convergence speed and avoids falling into local optimums for the randomly-selected initial centroids. Such attribute is well aligned for the multi-channel receiver, since a relatively low-order modulation format like PAM-4 can be configured for each data path, which allows mitigating the nonlinearity distortions of the received samples in an efficient way.

3.4 Channel Decision Passing (CDP) Algorithm

The block diagram of the CDP algorithm with feedback nonlinearity cancellation scheme is shown in Fig. 13. Considering

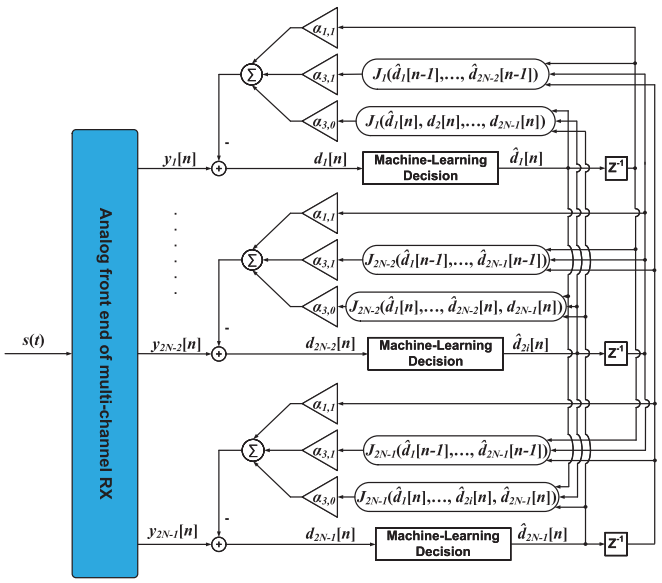


Fig. 13: Block diagram of feedback cancellation nonlinearity mitigation scheme in combination with the proposed CDP algorithm.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

the SNR improvement based on the bandwidth splitting property of multi-channel receivers, the event that symbols transmitted through multiple channels are simultaneously in error has a low probability of occurrence. Based on this, implementing the nonlinearity mitigation on the multi-channel receivers in a channel-by-channel fashion will be desirable to increase computational efficiency. Such a process can actually be modeled as the model-free Markov Decision Process (MDP), calling for the application of reinforcement learning (RL) algorithms on the nonlinearity mitigation of multi-channel receivers. Here, "model-free" means the transition matrix and reward of MDP are unknown to the system and it can only take actions by a trial-and-error process through a reward mechanism.

The state-action interaction process of the CDP algorithm is illustrated in Fig. 14. Since the nonlinearity mitigation will be processed channel-by-channel with subchannels processed path-by-path, the initial state starting from the single path of the baseband channel, P_1, can be expressed as $s_0 = [d_1[n], d_2[n], ..., d_{2N-1}[n], P_1]$, $d_i[n](i = 1, 2, ...2N-1)$ represents the detected symbol of each data path without nonlinearity mitigation. $A_i = [\hat{d}_i^{(1)}, ..., \hat{d}_i^{(V)}]$ is the action space where the transmitted symbol from the corresponding path P_i will be speculated from V constellation symbols. Thus, A_i varies from the path to path. The detected outputs are replaced by the speculated decisions, then an intermediate state is: $s_i = [\hat{d}_1[n], \hat{d}_2[n], ...\hat{d}_i[n], d_{i+1}[n]..., d_{2N-1}[n], P_{i+1}]$, P_{i+1} represents the path of the current state.

The reward r, which represents the error between one of the constellation points $\hat{d}_i^{(v)}(v = 1, 2, ..., V)$ and the nonlinearity cancellation result using this symbol associated with outputs from other channels, can be calculated by:

$$r^{(v)} = -|y_i[n] - \alpha_{3,0} J_i(\hat{d}_1[n], ...\hat{d}_i^{(v)}, d_{i+1}[n], ...d_{2N-1}[n]) - \alpha_{3,1} J_i(\hat{d}_1[n-1], ...\hat{d}_{2N-1}[n-1]) - \alpha_{1,1}\hat{d}_i[n-1] - \hat{d}_i^{(v)}| \quad (35)$$

Based on greedy policy [50], the constellation symbol $\hat{d}_i^{(m)}$ corresponding to the maximum reward $r^{(m)}$ will be the taken action from A_i, which is regarded as the speculated decision $\hat{d}_i[n]$ on the currently received symbol of path P_i and thus updates the decision vector of the state representing the combination of estimated symbols for all paths.

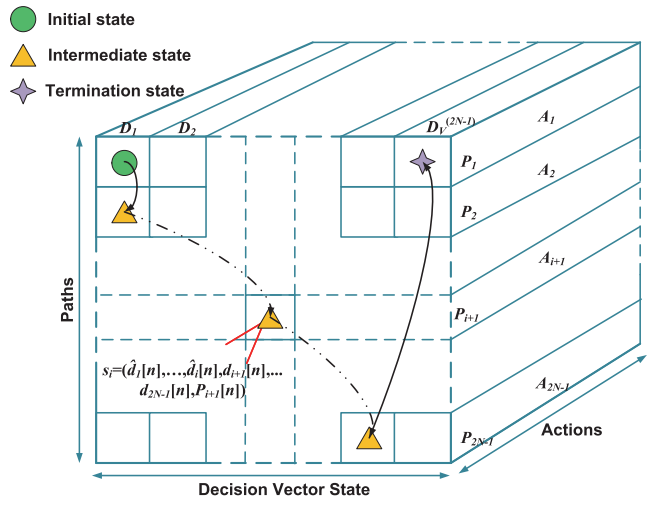


Fig. 14: State-action interaction process of CDP algorithm.

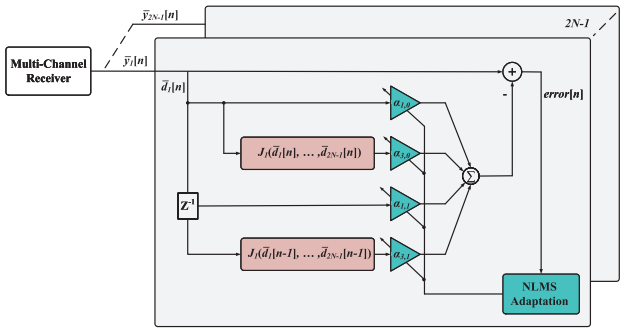


Fig. 15: Nonlinearity parameters adaption based on NLMS algorithm.

In Eqn. 35, $y_i[n](i = 1, 2, ...2N-1)$ is the distorted sequence sampled by the ADC, J_i is the function including all the inband nonlinearity terms of a path and $\hat{d}_i[n-1]$ are the estimation of the previous symbols used for addressing the memory effect. $\alpha_{1,1}$, $\alpha_{3,0}$ and $\alpha_{3,1}$ are nonlinearity parameters that are obtained using the normalized least mean square (NLMS) algorithm with a training sequence as shown in Fig. 15, where $\bar{y}_i[n](i = 1, 2, ...2N-1)$ represents the received samples of the training sequence, and $\bar{d}_i[n]$ represents the known symbol associated with $\bar{y}_i[n]$.

In the CDP algorithm, an episode is terminated once the detected symbols from all paths are replaced by the speculative decisions $\hat{d}_i[n]$. So the termination state is: $s_T = [\hat{d}_1[n], ..., \hat{d}_{2N-1}[n], P_1]$. The learning process will stop when the termination state is stable from episode to episode. The proposed CDP algorithm shows a fast convergence speed, which will be discussed in the following section.

4 Simulation Results

In this section, we will compare the nonlinearity mitigation methods based on selected machine learning algorithms in terms of time complexity and BER performance. As stated in IV-A, both the time complexity and BER performance of the KNN algorithm are significantly impacted by the size of the training sequence. While in order to improve the classification accuracy in the low SNR regime, the number of neighbors k is set to 21 for the rest of the simulations, and the size of training sequence M is set to 50000.

The time complexity comparison result between the ML-based multi-channel nonlinearity mitigation methods is presented in Table 1. Here, we use the elapsed time and the number of floating point operations (FLOPs) as the indication of time complexity. In the simulation, a 3-channel receiver is configured to detect a PAM-4 signal through the baseband channel and QAM-16 signals through 2 higher frequency subchannels. Each ML algorithm operates under the setup of input swing Vpp = 1.2V, Eb/No =

ML Algorithms	Elapsed time	Flops	BER
KNN (M=50000)	3.045 s	6556×10^5	9.44×10^{-6}
K-mean clustering	0.052 s	180×10^5	1.66×10^{-4}
CDP	0.145 s	322.4×10^5	$< 1 \times 10^{-6}$

TABLE 1: Time complexity comparison between different ML-based nonlinearity mitigation algorithms with Eb/No=20dB and IIP3=14.5dBm.

20dB and IIP3 = 14.5dBm. A testing sequence containing 1E5 symbols is passed through the system for simulation.

Through the simulations, the number of iterations required by K-mean clustering is 3. For CDP, the average number of episode processing one set of symbols through all paths is 1.04. Based on these results, the FLOPs for each algorithm are calculated by counting the real-number addition/subtraction, the real-number multiplication/division, and the comparison between real numbers as one FLOP respectively. The elapsed time is collected by running the algorithm in MATLAB.

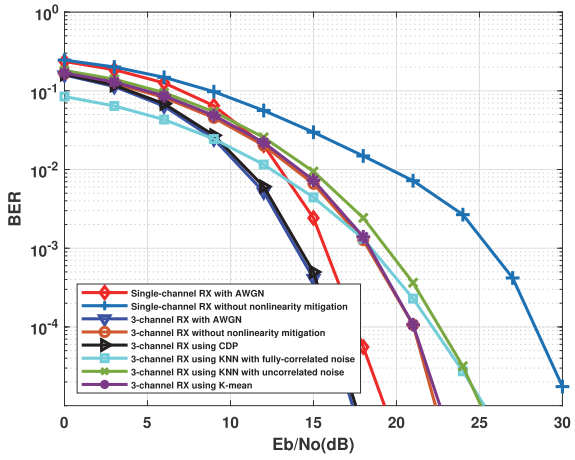
It can be learned from Table 1, the K-mean clustering algorithm shows the lowest time complexity since this algorithm does not require the assistance of training sequence nor the nonlinearities feedback cancellation, but instead directly partitions the received samples into clusters. However, for a system with low IIP3, the received samples can be severely distorted by the nonlinearities and therefore become interfused with other constellation points. Those samples cannot be correctly classified with a clustering operation. Thus, the K-mean clustering provides the worst BER performance among these three algorithms.

With a training sequence, the BER performance of KNN is improved compared to K-mean clustering. However, due to the large size of the training sequence, KNN shows the worst time complexity. On the other hand, the noise and nonlinearities present in the training samples with known labels can be interfused with other categories which will lead to BER degradation. In comparison, the CDP algorithm using the nonlinearities feedback cancellation outperforms both KNN and K-mean clustering algorithms with reasonable time complexity.

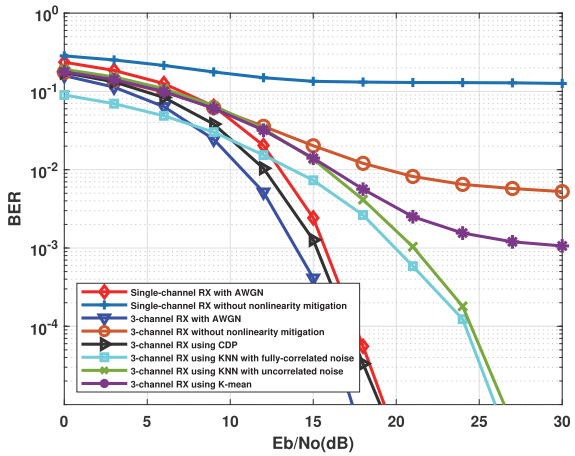
Fig. 16 shows the simulation results of BER performance for the ML-based nonlinearity mitigation methods under the same nonlinearity profile and data rate conditions. In this simulation, the swing of the input signal through the AFE is set to 0.5V and 1.2V respectively. The IIP3 of the system is set to 14.5dBm, and the memory effect is modeled by setting $\frac{\alpha_{1,1}}{\alpha_{1,0}} = 0.16$ referring to Eqn. 30. The comparison between the single-channel receiver and the multi-channel receiver is made assuming the single-channel receiver is configured to detect a baseband PAM-4 signal sampled at 32GS/s, while a 3-channel receiver is configured to detect a combination of baseband PAM-4 signal and mid-band QAM-16 signal centered at 6.4GHz and high-band QAM-16 signal centered at 12.8GHz with ADCs sampling at 6.4GS/s.

It can be seen from Fig. 16, for both Vpp=0.5V and Vpp=1.2V cases, the multi-channel receiver shows improved BER performance compared to the single-channel receiver when the nonlinearity mitigation methods are not applied. Such improvement dues to the SNR improvement following Eqn. 28 and the inherent nonlinearity robustness of the multi-channel receiver over the single-channel receiver caused by the carrier's amplitude reduction in the signal bands since the combined multi-band signal has the same overall amplitude limit as the single-band counterpart.

Fig. 16(a) also shows the BER performance of multi-channel receivers using different ML-based nonlinearity mitigation algorithms when the nonlinearity distortion of the received samples is limited with $Vpp = 0.5V$. Under this condition, KNN and K-mean clustering cannot provide noticeable BER improvement compared to the multi-channel receiver without a nonlinearity mitigation scheme since the 3rd-order harmonic distortions and the intermodulation products are negligible. While the CDP algorithm can still offer improved BER performance since the residual nonlinearities caused by the memory effect can be efficiently



(a)



(b)

Fig. 16: BER performance comparison between the single-channel receiver and 3-channel receiver with & without ML-based nonlinearity mitigation algorithms from nonlinear AFE with 14.5dBm IIP3 and memory effect quantified by $\frac{\alpha_{1,1}}{\alpha_{1,0}} = 0.16$ and (a) Vpp = 0.5V, (b) Vpp = 1.2V.

removed by the feedback cancellation scheme. Fig. 16(b) shows the BER plots as the counterpart of those in Fig. 16(a) when the received samples are severely distorted by the nonlinearities with $Vpp = 1.2V$. Compared to the multi-channel receiver without any nonlinearity mitigation method, all the ML-based algorithms significantly improve the BER in the high Eb/No regime, where the SNDR of the received samples is dominated by nonlinearities. While, as mentioned above, with the implementation of the nonlinearity feedback cancellation when the system interacts with the environment, the CDP algorithm provides better BER performance compared to both KNN and K-mean clustering algorithms. Overall, the proposed CDP algorithm can provide consistent nonlinearity mitigation when the peak-to-peak swing of the input signal increases from 0.5V to 1.2V. In practice, the CDP algorithm can benefit from improved SNR of the received samples by simply increasing Vpp of the input signal, while the nonlinearity distortion can still be effectively removed by the CDP algorithm with acceptable complexity.

It needs to mention here, through our simulation, the KNN

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3259880

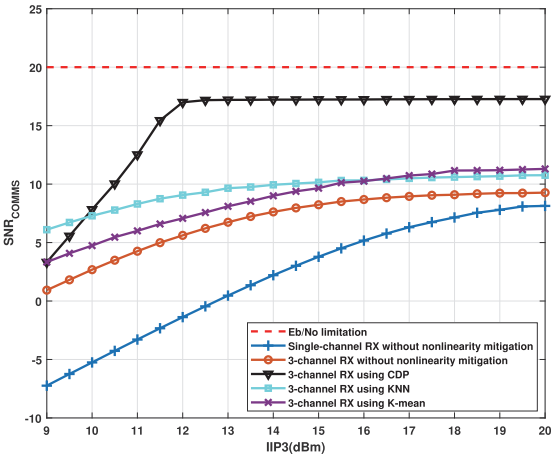


Fig. 17: Communication SNR vs IIP3 comparison between the single-channel receiver and 3-channel receiver with 20dB Eb/No and memory effect quantified by $\frac{\alpha_{1,1}}{\alpha_{1,0}} = 0.16$.

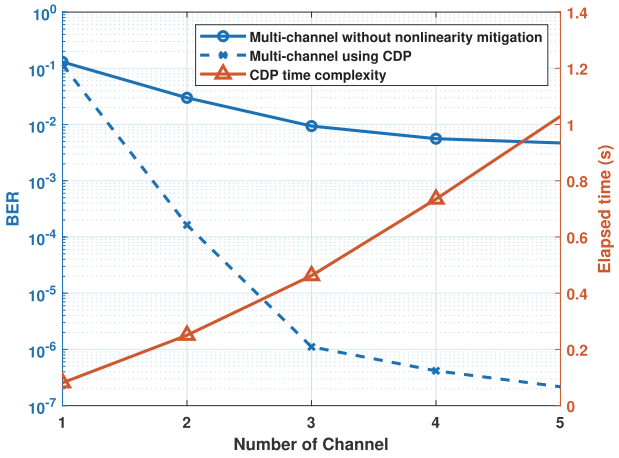


Fig. 18: BER & Time complexity of CDP algorithm vs the number of channels from simulation set up of 20dB Eb/No and 14.5dBm IIP3 with memory effect quantified by $\frac{\alpha_{1,1}}{\alpha_{1,0}} = 0.16$.

algorithm shows the capability of mitigating correlated noise considering KNN's jointly-classification operation over the symbols from all channels. Fig. 16 shows that the KNN algorithm with fully-correlated noise improves the BER performance in the low SNR regime compared to the KNN plot with uncorrelated noise. In comparison, the K-mean clustering and the CDP algorithms make the symbol's decision independently in each channel, the correlation among channel noise realizations does not have an impact on the BER performance. Practically, the noise between channels will show some degree of correlation when the sampling clock of ADCs has a high common-mode clock jitter that dominates the SNR of the receiver. This correlation, however, becomes weak for an arbitrary input signal, and in general obtaining correlated noise across channels that lead to improved performance in the low SNR regime remains an open problem.

Fig. 17 shows the communications SNR comparison result between nonlinearity mitigation algorithms as the function of IIP3 with 20dB Eb/No, which is a practical setup for the wireline receiver and provides an upper bound for the single-channel receiver and the 3-channel receiver referring to Eqn. 27. The SNR_{COMMS} of the multi-channel receiver using the CDP algorithm can be calculated from the soft decisions after the feedback cancellation with the converged decisions from all channels. It can be seen when the IIP3 of the AFE is 9dBm, the received samples of the multi-channel receiver are severely distorted by the nonlinearities when the nonlinearity mitigation method is not applied. In this case, the CDP algorithm can provide only 3dB communications SNR improvement compared to the multi-channel receiver without a nonlinearity mitigation scheme. While, as the IIP3 of the AFE increases, the communications SNR of the multi-channel receiver without nonlinearity mitigation gets improved and the CDP algorithm provides increasing communications SNR boost. When the IIP3 of the AFE is greater than 12dBm, the communications SNR of the multi-channel receiver is greater than 5dB, which leads to the BER without nonlinearity mitigation of less than 10^{-2}. In this case, the nonlinearity distortion can be significantly removed by the CDP algorithm. Thus, the communications SNR provided by the multi-channel receiver with the CDP algorithm approaches the performance upper bound.

Although the KNN algorithm and K-mean clustering algorithm can only generate the hard decisions since they are both classification algorithms, their equivalent SNR_{COMMS} can be obtained from the resulting BER. As shown in Fig. 17, the nonlinearity distortion can be partially removed by the KNN or K-mean clustering algorithms. However, there's a noticeable SNR_{COMMS} gap when compared to the CDP algorithm as IIP3 increases. This performance gap is caused by the memory effect which leads to the unremovable residual nonlinearity distortion by the KNN or K-mean clustering algorithm. In summary, the linearity requirement of the analog front end can be effectively relaxed by the proposed CDP algorithm.

The BER performance and the time complexity of the CDP algorithm are investigated as shown in Fig. 18. In this simulation, the Eb/No is set to 20dB, and the IIP3 of the AFE is set to 14.5dBm with the associated memory effect. The channel loss is assumed to be not presented in this simulation. Thus, referring to Eqns. 6 and 7, there's an SNR penalty as $N > 1$, where N is the number of channels. However, due to the amplitude reduction of each sub-band signal, the multi-channel receiver suffers less from the nonlinearity distortion as the channels increase, leading to improved communications SNR as the result. It can be seen from Fig. 18 when no nonlinearity mitigation is applied to the multi-channel receiver, the BER performance is improved with the increasing number of channels, which leads to a higher BER when the CDP algorithm is applied. Moreover, as N increases, the amplitude of each subband signal will decrease. As the result, the memory effect will eventually become the performance limiter, which leads to converged communications SNR of the received sample and thus similar BER performance with the increasing number of channels. Fig. 18 also shows the time complexity of the CDP algorithm linearly increases with the number of channels. This simulation runs on MATLAB with an Intel I9-12900H processor and 32GB memory. The result proves that the complexity of the CDP algorithm is in the order of $O(N)$, thus showing the applicability of multi-channel receivers by stacking more channels to improve the transmission bandwidth without the risk of complexity issues.

5 Conclusion

This paper quantifies the performance advantages of the multi-channel receiver over the single-channel receiver in terms of front-end noise, phase noise, clock jitter, and nonlinearities with expression and simulation results. The analysis and simulations illustrate the need to achieve further nonlinearity mitigation. Thus, multiple ML-based algorithms are investigated for nonlinearity mitigation in the multi-channel receiver architecture including the KNN-based joint classification algorithm, K-mean clustering algorithm, and RL-based CDP algorithm. Time complexity, BER performance, and communications SNR comparisons are made between these algorithms. The proposed CDP algorithm shows fast convergence speed and significantly improved BER performance, making this method a practical solution for fully-integrated nonlinearity mitigation in multi-channel receivers.

Acknowledgments

This research project was funded by the NSF Award 2148354.

References

[1] S. Hoyos, B. Sadler, and G. Arce, "Broadband multicarrier communication receiver based on analog to digital conversion in the frequency domain," *IEEE Transactions on Wireless Communications*, vol. 5, no. 3, pp. 652–661, 2006.

[2] K. Gharibdoust, A. Tajalli, and Y. Leblebici, "A $\mathbf{4}\times\mathbf{9}$ Gb/s $\mathbf{1}$ pJ/b hybrid NRZ/multi-tone I/O with crosstalk and isi reduction for dense interconnects," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 992–1002, 2016.

[3] K. Lee, I. Nam, I. Kwon, J. Gil, K. Han, S. Park, and B.-I. Seo, "The impact of semiconductor technology scaling on CMOS RF and digital circuits for wireless application," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, pp. 1415–1422, 2005.

[4] H. Zhao, J. C. G. Diaz, and S. Hoyos, "Multi-channel receiver nonlinearity cancellation using channel speculation passing algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021.

[5] B. Razavi, "Design considerations for direct-conversion receivers," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, no. 6, pp. 428–435, 1997.

[6] X. Fan, H. Zhang, and E. SÁnchez-Sinencio, "A noise reduction and linearity improvement technique for a differential cascode LNA," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 3, pp. 588–599, 2008.

[7] E. A. Keehr and A. Hajimiri, "Equalization of third-order intermodulation products in wideband direct conversion receivers," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2853–2867, 2008.

[8] N.-A. Vu, H.-N. Le, T.-H.-T. Tran, and Q.-K. Trinh, "Novel distortion compensation scheme for multichannel direct RF digitization receiver," in *2019 19th International Symposium on Communications and Information Technologies (ISCIT)*, pp. 156–161, 2019.

[9] J. Marttila, M. Allén, M. Kosunen, K. Stadius, J. Ryynänen, and M. Valkama, "Reference receiver enhanced digital linearization of wideband direct-conversion receivers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 2, pp. 607–620, 2017.

[10] R. Vansebrouck, C. Jabbour, O. Jamin, and P. Desgreys, "Fully-digital blind compensation of non-linear distortions in wideband receivers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 8, pp. 2112–2123, 2017.

[11] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind maximum-likelihood identification of wiener systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3017–3029, 2009.

[12] A. Taleb, J. Sole, and C. Jutten, "Quasi-nonparametric blind inversion of wiener systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 917–924, 2001.

[13] A. Naitali and F. Giri, "Wiener–hammerstein system identification–an evolutionary approach," *International Journal of Systems Science*, vol. 47, no. 1, pp. 45–61, 2016.

[14] K.-K. Xu, H.-D. Yang, and C.-J. Zhu, "A novel extreme learning machine-based hammerstein-wiener model for complex nonlinear industrial processes," *Neurocomputing*, vol. 358, pp. 246–254, 2019.

[15] M. A. H. Shaikh and K. Barbé, "Wiener–hammerstein system identification: A fast approach through spearman correlation," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 5, pp. 1628–1636, 2019.

[16] K. Shi and A. Redfern, "Blind volterra system linearization with applications to post compensation of ADC nonlinearities," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3581–3584, 2012.

[17] P. M. S. Burt and J. H. de Morais Goulart, "Efficient computation of bilinear approximations and volterra models of nonlinear systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 804–816, 2017.

[18] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 5, pp. 285–291, 1976.

[19] K. Dogancay, "Blind compensation of nonlinear distortion for bandlimited signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 9, pp. 1872–1882, 2005.

[20] M. Hotz and C. Vogel, "Linearization of time-varying nonlinear systems using a modified linear iterative method," *IEEE Transactions on Signal Processing*, vol. 62, no. 10, pp. 2566–2579, 2014.

[21] M. Li, S. Yu, J. Yang, Z. Chen, Y. Han, and W. Gu, "Nonparameter nonlinear phase noise mitigation by using M-ary support vector machine for coherent optical systems," *IEEE Photonics Journal*, vol. 5, no. 6, pp. 7800312–7800312, 2013.

[22] G. Chen, J. Du, L. Sun, W. Zhang, K. Xu, X. Chen, G. T. Reed, and Z. He, "Nonlinear distortion mitigation by machine learning of SVM classification for PAM-4 and PAM-8 modulated optical interconnection," *Journal of Lightwave Technology*, vol. 36, no. 3, pp. 650–657, 2018.

[23] J. Huo, M. Wang, M. He, X. Zhou, and K. Long, "Joint SVM and KNN module algorithm for 16-QAM signal decision," *Optical Engineering*, vol. 59, no. 12, p. 126105, 2020.

[24] E. Liu, Z. Yu, C. Yin, and K. Xu, "Nonlinear distortions compensation based on artificial neural networks in wideband and multi-carrier systems," *IEEE Journal of Quantum Electronics*, vol. 55, no. 5, pp. 1–5, 2019.

[25] E. Giacoumidis, S. Mhatli, J. Wei, S. T. Le, I. Aldaya, M. F. C. Stephens, M. E. MacCarthy, A. D. Ellis, N. J. Doran, and B. J. Eggleton, "Intra and inter-channel nonlinearity compensation in wdm coherent optical OFDM using artificial neural network based nonlinear equalization," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2017.

[26] S. Liu, Y. M. Alfadhli, S. Shen, M. Xu, H. Tian, and G.-K. Chang, "A novel ANN equalizer to mitigate nonlinear interference in analog-RoF mobile fronthaul," *IEEE Photonics Technology Letters*, vol. 30, no. 19, pp. 1675–1678, 2018.

[27] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient KNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, 2018.

[28] D. Wang, M. Zhang, M. Fu, Z. Cai, Z. Li, H. Han, Y. Cui, and B. Luo, "Nonlinearity mitigation using a machine learning detector based on k-nearest neighbors," *IEEE Photonics Technology Letters*, vol. 28, no. 19, pp. 2102–2105, 2016.

[29] D. Wang, M. Zhang, M. Fu, Z. Cai, Z. Li, Y. Cui, and B. Luo, "KNN-based detector for coherent optical systems in presence of nonlinear phase noise," in *2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS)*, pp. 1–3, 2016.

[30] M. Sanchez-Fernandez, M. de Prado-Cumplido, J. Arenas-Garcia, and F. Perez-Cruz, "Svm multiregression for nonlinear channel estimation in multiple-input multiple-output systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2298–2307, 2004.

[31] Z. Fu, A. Robles-Kelly, and J. Zhou, "Mixing linear svms for nonlinear classification," *IEEE Transactions on Neural Networks*, vol. 21, no. 12, pp. 1963–1975, 2010.

[32] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8595–8598, 2013.

[33] A. Nandy and M. Biswas, "Google's deepmind and the future of reinforcement learning," in *Reinforcement Learning*, pp. 155–163, Springer, 2018.

[34] B. Bouzy and G. Chaslot, "Monte-carlo go reinforcement learning experiments," in *2006 IEEE Symposium on Computational Intelligence and Games*, pp. 187–194, 2006.

[35] A.-H. Tan, N. Lu, and D. Xiao, "Integrating temporal difference methods and self-organizing neural networks for reinforcement learning with delayed evaluative feedback," *IEEE Transactions on Neural Networks*, vol. 19, no. 2, pp. 230–244, 2008.

[36] L. Zou, M. Jiang, C. Zhao, Y. He, D. Zhu, and Q. Huang, "Bldnet: Robust learning-based detection for high-order qam with nonlinear distortion," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 262–266, 2020.
[37] K. Wang, C. Wang, J. Zhang, Y. Chen, and J. Yu, "Mitigation of soa-induced nonlinearity with the aid of deep learning neural networks," *Journal of Lightwave Technology*, vol. 40, no. 4, pp. 979–986, 2022.
[38] J. C. G. Diaz, H. Zhao, Y. Zhu, S. Palermo, and S. Hoyos, "Recurrent neural network equalization for wireline communication systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 4, pp. 2116–2120, 2022.
[39] X. Dai, X. Li, M. Luo, Q. You, and S. Yu, "Lstm networks enabled nonlinear equalization in 50-gb/s pam-4 transmission links," *Applied optics*, vol. 58, no. 22, pp. 6079–6084, 2019.
[40] E. Stevens-Navarro, Y. Lin, and V. W. S. Wong, "An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1243–1254, 2008.
[41] M. Matta, G. C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, A. Nannarelli, M. Re, and S. Spanò, "A reinforcement learning-based QAM/PSK symbol synchronizer," *IEEE Access*, vol. 7, pp. 124147–124157, 2019.
[42] S. Hoyos, S. Pentakota, Z. Yu, E. Sobhy Abdel Ghany, X. Chen, R. Saad, S. Palermo, and J. Silva-Martinez, "Clock-jitter-tolerant wide-band receivers: An optimized multichannel filter-bank approach," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 2, pp. 253–263, 2011.
[43] L. Wang, Y. Fu, M. LaCroix, E. Chong, and A. C. Carusone, "A 64Gb/s PAM-4 transceiver utilizing an adaptive threshold ADC in 16nm FinFET," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pp. 110–112, 2018.
[44] S. W. Callender, *Wideband signal acquisition via frequency-interleaved sampling*. University of California, Berkeley, 2015.
[45] A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, 1998.
[46] T. H. Lee and A. Hajimiri, "Oscillator phase noise: A tutorial," *IEEE journal of solid-state circuits*, vol. 35, no. 3, pp. 326–336, 2000.
[47] W. Kester, "Converting oscillator phase noise to time jitter," *Tutorial MT-008, Analog Devices*, p. 2009, 2009.
[48] J. C. Gomez Diaz, S. Kiran, S. Palermo, and S. Hoyos, "Jitter-robust multicarrier ADC-based serial link receiver architecture : (invited special session paper)," in *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1151–1154, 2019.
[49] W. Namgoong, "Modeling and analysis of nonlinearities and mismatches in ac-coupled direct-conversion receiver," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 163–173, 2005.
[50] S. M. Kakade, "A natural policy gradient," *Advances in neural information processing systems*, vol. 14, 2001.

Haotian Zhao received his M.S. degree in electronic engineering from Stony Brook University, New York, in 2019. He is currently working towards his Ph.D. degree at Texas A&M University, Texas, USA. His research interests include analog and mixed-signal systems and circuit design, especially applying machine learning algorithms on wireless/wireline communication systems.

Julian Camilo Gomez Diaz received his B.S. degree in electrical engineering from Pontificia Universidad Javeriana (PUJ), Bogota, Colombia. He is currently pursuing a Ph.D. degree at Texas A&M University, College Station, TX, USA. He was an analog design intern with Intel Co., Santa Clara, CA, USA in 2019 where he worked in residue amplifiers for high-speed pipelined SAR ADCs. His research interests include high-speed serial links, and analog and mixed-signal circuit design.

Sebastian Hoyos received the B.S. degree in electrical engineering from Pontificia Universidad Javeriana (PUJ), Bogota, Colombia, in 2000, and the M.S. and Ph.D. degrees in electrical engineering from the University of Delaware, Newark, in 2002 and 2004, respectively. He was with Lucent Technologies Inc., Bogota, Colombia, from 1999 to 2000 for the Andean region in South America. Simultaneously, he was a lecturer with PUJ, where he lectured on microelectronics and control theory. During his M.S. and Ph.D. studies, he was with PMC-Sierra Inc., the Delaware Research Partnership Program, and the Army Research Laboratory Collaborative Technology Alliance in Communications and Networks. He was a Postdoctoral Researcher (2004-2006) with the Berkeley Wireless Research Center, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. He joined Texas A&M University, College Station, Texas in 2006 where he is currently an Associate Professor with the Department of Electrical and Computer Engineering. His research interests include telecommunication systems, digital signal processing, analog and mixed-signal processing, and circuit design.