Differential Deep Detection in Massive MIMO With One-Bit ADC

Don-Roberts Emenonye†, Carl Dietrich†, and R. Michael Buehrer†

Abstract—This article presents a differential detection scheme for the uplink of a massive MIMO system that employs one-bit quantizers on each receive antenna. We focus on the detection of differential amplitude and phase shift keying symbols and use the Bussgang theorem to express the quantized received signal in terms of quantized signals received during previous channel uses. Subsequently, we derive the maximum likelihood detector for the differentially encoded amplitude and phase information symbols. We note that while the one-bit detector can decode the differentially encoded phase information symbols, it fails to decode the differentially encoded amplitude information. To decode the amplitude information, we present a one-bit variable quantization level (VQL) system and train a deep neural network to perform two-symbol differential amplitude detection. Through Monte-Carlo simulations, we empirically validate the performance of the proposed amplitude and phase detectors. The presented numerical results show that the spectral efficiency attained in one-bit differential systems is better than the spectral efficiency attained in one-bit coherent systems.

I. INTRODUCTION

The advantage of massive multi-input-multi-output (MIMO) lies in its ability to provide large multiplexing and diversity gains [1]. However, to achieve these gains, some challenges need to be resolved. One of these challenges lies in the prohibitively high circuit power consumed by a large number of radio-frequency (RF) chains at the base station (BS). A point of power inefficiency in the RF chains occurs at the analog-to-digital converters (ADCs). In particular, it has been shown that an increase in the resolution of these converters causes an exponential increase in their power consumption [2]. To tackle this problem, several authors have investigated the operation of massive MIMO systems in the low-resolution regime.

In [3], a quantized distributed reception scheme is investigated for use in the downlink of a multi-antenna system that is serving a large number of internet-of-things (IoT) devices. In [4], the problem is reformulated for the uplink and a near maximum likelihood detector is presented. In [5], the problem is extended to MIMO systems with orthogonal frequency-division multiplexing (MIMO-OFDM) and a maximum a-posterior (MAP) algorithm for symbol detection is investigated. In [6], a downlink massive MIMO system with a relay is investigated. The relay system comprises one-bit analog-to-digital and one-bit digital-to-analog converters. In that work, the Bussgang decomposition [7] is employed to develop a channel estimation technique. Although the power consumption is reduced by the application of one-bit ADCs, this perceived gain comes at the cost of a more expensive channel estimation phase. In particular, channel estimation in the low-resolution regime employ longer pilot sequences [3], [4], [8]. Hence, the channel is required to stay constant for a longer period of time. Such long pilot sequences negatively impact the spectral efficiency of coherent one-bit system.

To solve this challenge, we present an analysis of the uplink of a system with differential modulation employed at a single user transmitter and one-bit ADCs at the base station. In general, differential/non-coherent communication is a well developed area [9]–[12] that has received considerable interest in recent times. In [12], [13], authors develop a look-up table based differential modulation technique. The table is constructed by minimizing the non-coherent distance between distinguishable codewords. In [14], an autocorrelation-based decision-feedback scheme is adopted for differential detection. In [15], [16], the minimum non-coherent distance is analyzed and used to design constellations for non-coherent modulation. In all these prior works, the impact of low-resolution ADCs on differential modulation is never considered, hence, in this article, we aim to develop detectors for differential detection in the low-resolution regime. While we employ an approximate maximum likelihood detector for differential phase detection, we employ a neural network (NN) for amplitude detection.

In the existing literature, the plethora of data available either through simulation or real-world measurements has enabled researchers to explore the possibility of a completely data-driven communication model. This method involves approximating the submodular blocks at the transmitter and at the receiver with two disjoint functions. Neural networks, as general and well-known function approximators, make this a viable research option. Leveraging this, a novel end-to-end deep learning-based communication system was developed with both the transmitters and the receivers replaced by the encoder and decoder of an autoencoder [17]. In that work, the parameters of the transmitter and the receiver are jointly updated by optimizing a loss function during training. In [18], the deep learning-based communication system was extended for multi-antenna transmission. More recently, deep learning has been used to enable energy-based detection in non-coherent systems. In [19], an end-to-end non-coherent learning-based system is developed.

† Don-Roberts Emenonye, Carl Dietrich, and R. Michael Buehrer are with Wireless@VT, Bradley Department of ECE, Virginia Tech, Blacksburg, VA, 24061. (Emails: {donroberts, cdietric, rbuehrer}@vt.edu)
II. SYSTEM MODEL

In this article, we consider a single user employing a differential amplitude phase shifting keying (DAPSK) system. The encoding scheme comprises of two concentric circles with different radius (amplitude). We define the amplitude of the inner circle as $\psi_0$, while $\psi_1$ represents the amplitude of the outer circle. The ring ratio is defined as $\alpha = \frac{\psi_1}{\psi_0}$. A unit power constraint is enforced by $\psi_0^2 + \psi_1^2 = 2$. The collection of all points on the inner and outer circles form two phase shift constellations defined as $S_1 = \{x^{(1)}, x^{(2)}, \ldots , x^{(M)}\}$, and $S_2 = \{x^{(M+1)}, x^{(M+2)}, \ldots , x^{(2M)}\}$ respectively. Both circles define the following constellation set, $\mathcal{S} = \{S_1, S_2\}$. $M$ represents the number of points on a particular circle. At the $v$th channel use, the transmitted symbol $x[v]$ depends on a block of $N_b$ bits defined as $b[v] = [b_1[v], b_2[v], b_3[v], \ldots , b_{N_b}[v]]$. The last $N_b-1$ bits in the block specify the phase of the transmitted symbol, while the first bit $b_1[v]$ specifies its amplitude. If we define $Y$ as a function that generates phase shift keying symbols from a block of bits, then the phase information symbol can be generated as $s[v] = Y([b_2[v], b_3[v], \ldots , b_{N_b}[v]])$.

\[
a[0] = \psi_0, \quad c[0] = a[0], \\
c[v] = c[v-1]s[v], \\
x[v] = a[v]c[v], \\
a[v] = \begin{cases} 1, & \text{if } b_1[v] = 0, \\ \frac{\psi_0}{\psi_1}, & \text{if } b_1[v] = 1 \text{ and } \tilde{x}[v-1] = \psi_0, \\ \frac{\psi_1}{\psi_0}, & \text{if } b_1[v] = 1 \text{ and } \tilde{x}[v-1] = \psi_1, \end{cases}
\]

where $a[v] \in \mathcal{A}\{1, \frac{\psi_0}{\psi_1}, \frac{\psi_1}{\psi_0}\}$, $\tilde{x}[v] = |x[v]|$, and $\tilde{x}[v] \in \mathcal{X}\{\psi_0, \psi_1\}$. Clearly, the transmitted symbol switches between $S_1$ and $S_2$, if $b_1[v] = 1$. During the $v$-th channel use, the signal received at the base station can be represented as

\[
y_u[v] = h[v]x[v] + z[v],
\]

where $y[v] = [y_1[v], y_2[v], \ldots , y_U[v]]$, $h[v] = [h_1[v], h_2[v], \ldots , h_U[v]]$, and $n[v]$ is the additive noise vector with $z[v] \sim \mathcal{CN}(0, \sigma^2 I)$. The frequency-selective channel at the $v$-th channel use can be modeled as a combination of $L$ parallel frequency-flat subchannels expressed as

\[
h_u[v] = \sum_{l=0}^{L-1} p[l]g_u[l]e^{-j2\pi lv/N}.
\]

where $p[l]$ is the power delay profile of the channel $l$-th tap with $\sum_{l=0}^{L-1} p[l] = 1$, $N$ is the number of channel uses, $g_u \sim \mathcal{CN}(0, 1)$, is the complex channel gain from the transmitter to the $u$-th receive antenna of the $l$-th path.

To allow for analysis, we assume $h[v] \approx h[v-1]$, hence the received signal can be written as

\[
y_u[v] = \left(\frac{a[v]}{a[v-1]}\right)y_u[v-1]s[v] + z_u'[v], \\
y_u[v] = a'[v]y_u[v-1]s[v] + z_u'[v],
\]

where $z_u'[v] = z_u[v] - a'[v]s[v]z_u[v-1]$. Note that $z_u'[v] \sim \mathcal{CN}(0, \varphi_z)$, where $\varphi_z = 2\sigma_z^2$ if $b_1[v] = 0$. Likewise, if $b_1[v] = 1$, then $\varphi_z = \sigma_z^2(1 + \frac{\psi_0^2}{\psi_1^2})$, or $\varphi_z = \sigma_z^2(1 + \frac{\psi_1^2}{\psi_0^2})$. The signal-to-noise ratio translates to $\rho = \frac{1}{\varphi_z}$, and we define the amplitude ratio as $a'[v] = \frac{a[v]}{a[v-1]}$. The received signal across all antennas can be written in the vectorized form

\[
y[v] = a'[v]y[v-1]s[v] + z'[v].
\]

III. RECEIVE PROCESSING

At the base station, each of the receive antennas is equipped with a low-resolution ADC which is specified as a quantizing function. The quantizing function is defined as $q = \mathcal{Q}(s)$, where $\mathcal{Q} : \mathbb{C} \rightarrow \mathcal{A}$, and $\mathcal{A}$ is the set of quantization alphabets. The quantizer independently compares the real and imaginary part of the received signal to a set of predefined thresholds. The thresholds are defined as:

\[-\infty = \xi_0 < \xi_1 < \cdots < \xi_Q < \xi_{Q+1} = +\infty.\]

The label $q$ is unique across different quantization bins. To concisely represent the information from the quantization function we need $2|\log_2 Q|$ number of bits.

A. One-Bit Receive Processing

In this section, we consider the one-bit quantizer defined with the $\text{sgn}$ function. The quantized version of the received signal at the $u$-th antenna can be defined as

\[
q_u[v] = \text{sgn}(\Re(y_u[v])) + j \text{sgn}(\Im(y_u[v])),
\]

The quantized symbols across all antennas at the $v$-th channel use can be compactly written as

\[
q[v] = [q_1[v], q_2[v], \ldots , q_U[v]].
\]

To find an approximation, we use the Bussgang Theorem \cite{7} to decompose the received quantized signal to the unquantized signal and an uncorrelated noise

\[
q[v] = \eta_v y[v] + \epsilon_v,
\]

where the $\eta$ is the quantization scaling factor and $\epsilon \sim \mathcal{CN}(0, \sigma^2)$ is the quantization noise component which is uncorrelated to $\eta$. With some algebraic manipulation, we write \cite{5} as

\[
\frac{1}{\eta_v}q_u[v] - \frac{\epsilon_v}{\eta_v} = a'[v]s[v] \left[ \frac{1}{\eta_{v-1}}q_u[v-1] - \frac{\epsilon_{v-1}}{\eta_{v-1}} \right] + z_u'[v].
\]

Assuming that the quantization effects remain constant during the entire transmission such that $\eta_v = \eta_{v-1} = \eta$ and $\epsilon_v = \epsilon_{v-1} = \epsilon$, we write

\[
q_u[v] = a'[v]s[v]q_u[v-1] - a'[v]s[v]\epsilon + \epsilon + \eta z_u'[v], \\
q_u[v'] = a'[v]s[v]q_u[v-1] + w_u[v],
\]

where $w_u[v] = \eta z_u'[v] - a'[v]s[v]\epsilon + \epsilon$ is the combined effect of the thermal noise and the quantization noise. Note that $z_u'[v] \sim \mathcal{CN}(0, \varphi_z)$, where $\varphi_z = \eta^2 \varphi_z + 2\sigma_z^2$ if $b_1[v] = 0$. Likewise,
if \( b_1[v] = 1 \), then \( q_{z,c} = \eta^2 q_z + \sigma^2 (1 + \frac{\psi^2}{\psi^2}) \), or \( q_{z,c} = \eta^2 q_z + \sigma^2 (1 + \frac{\psi^2}{\psi^2}) \). The signal-to-noise ratio translates to \( \rho = \frac{1}{q_{z,c}} \).

For ease of analysis, we transform the system from the complex domain to the real domain. First, the quantized and unquantized received signal can be written as

\[
q_{R,u}[v] = \begin{bmatrix} q_{R,u,1}[v] \\ q_{R,u,2}[v] \end{bmatrix} = \begin{bmatrix} \Re(q_u[v]) \\ \Im(q_u[v]) \end{bmatrix},
\]

(11)

\[
y_{R,u}[v] = \begin{bmatrix} y_{R,u,1}[v] \\ y_{R,u,2}[v] \end{bmatrix} = \begin{bmatrix} \Re(y_u[v]) \\ \Im(y_u[v]) \end{bmatrix},
\]

(12)

where \( q_{R,u}[v] \) and \( y_{R,u}[v] \) are the quantized and unquantized received signal, respectively, during the \( v \)-th channel use.

Next, the quantized signal at the previous symbol interval received at the \( u \)-th base station antenna is converted from the complex to the real domain

\[
F_{R,u}[v] = \begin{bmatrix} \Re(q_u[v - 1]) \\ \Im(q_u[v - 1]) \end{bmatrix} \in \mathbb{R}^{2 \times 2}.
\]

(13)

To ensure a compact derivation of the likelihood of the received signal, we enforce a sign change on both the quantized received signal \( f_u[v] \) as

\[
\tilde{F}_{R,u}[v] = \begin{bmatrix} \tilde{F}_{R,u,1}[v] \\ \tilde{F}_{R,u,2}[v] \end{bmatrix} = \begin{bmatrix} f_{R,u,1}^T[v] \\ f_{R,u,2}^T[v] \end{bmatrix} \in \mathbb{R}^{2 \times 2}.
\]

(14)

where \( \tilde{F}_{R,u}[v] \) is defined as

\[
\tilde{F}_{R,u,i}[v] = q_u[v]f_{R,u,i}^T[v].
\]

(16)

The noise and the transmit signal can be written as

\[
w_{R,u}[v] = \begin{bmatrix} \Re(w_u[v]) \\ \Im(w_u[v]) \end{bmatrix} \in \mathbb{R}^{2 \times 1},
\]

and the quantized received signal can be written as

\[
q_{R,u}[v] = a'[v]\tilde{F}_{R,u}[v]s_R[v] + w_{R,u}[v],
\]

(17)

and based on the value of \( q_{R,u,i}[v] \), we define two sets of indices, \( \mathcal{P} \) and \( \mathcal{N} \)

\[
\mathcal{P} = \{(i, u) : q_{R,u,i}[v] \geq 0\}, \quad \mathcal{N} = \{(i, u) : q_{R,u,i}[v] < 0\}.
\]

Finally, with these definitions, we can write the likelihood function as

\[
L(a'[v]s_R[v]|\rho) = \Pr\left(a'[v]\sqrt{\rho}f_{R,u,i}^T[v]s_R[v] \geq -w_{R,u,i}[v] \mid (i, u) \in \mathcal{P}\right)
\]

\[
\Pr\left(a'[v]\sqrt{\rho}f_{R,u,i}^T[v]s_R[v] \geq w_{R,u,i}[v] \mid (i, u) \in \mathcal{N}\right)
\]

(18)

Using (18), the maximum likelihood detector can be written as

\[
b_1[v] = \begin{cases} 0, & \text{if } \|\tilde{a}'[v]s_R[v]\| = 1, \\ 1, & \text{if } \|\tilde{a}'[v]s_R[v]\| \neq 1, \end{cases}
\]

(19)

and with an abuse of notation the remaining bits can be recovered from the phase information by \( \hat{b}_2[v], \hat{b}_3[v], \cdots, \hat{b}_N[v] \) = \( \Upsilon^{-1}\left(\tilde{a}'[v]s_R[v]/\|\tilde{a}'[v]s_R[v]\|\right) \). Empirical results indicate that the detector suffers from a substantial error floor, which can be attributed to the performance of the amplitude recovery part of the detector. This is intuitive because the one-bit quantizer only represents one level of amplitude (i.e., 0 or 1).

B. Deep Differential Detection With One-Bit And Variable Quantization Levels

To enable detection with one-bit, we propose to group the base station antennas such that each group has the same quantization levels and receive antennas in different groups employ different quantization levels. In this article, the following quantization groups are used \( \mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3 \) such that \( |\mathcal{U}_1| + |\mathcal{U}_2| + |\mathcal{U}_3| = U \). The quantizer operation in the \( j \)-th group can be defined as

\[
q_{R,u,j,i}[v] = \mathcal{Q}(y_{R,u,j,i}[v]),
\]

(21)

where \( u_j \in \mathcal{U}_j \), and \( y_{R,u,j,i}[v] \in \{\zeta_{1,j}, \zeta_{2,j}\} \). More specifically,

\[
q_{R,u,j,i}[v] = \begin{cases} \zeta_{3,j}, & \text{if } y_{R,u,j,i}[v] > \zeta_{2,j}, \\ \zeta_{1,j}, & \text{if } y_{R,u,j,i}[v] < \zeta_{2,j} \end{cases}
\]

Again we can decompose the received signal through the Bussgang theorem

\[
q[v] = \eta h[v]x[v] + \eta z[v] + \epsilon[v],
\]

(22)
where \( e'[v] = \eta z[v] + e[v] \) is Gaussian, i.e. \( e'[v] \sim \mathcal{CN}(0, \sigma^2 = \eta^2 \sigma_z^2 + \sigma_e^2) \) and considering a single antenna, \( u \) in the real domain, we have

\[
q_{R,u,i}[v] = \eta \mathbf{h}_{R,u,i}^T[v] \mathbf{x}_R[v] + e'_{R,u}[v],
\]

where

\[
e'_{R,u}[v] = \begin{bmatrix} (e'_{R,u,1}[v]) \cr (e'_{R,u,2}[v]) \end{bmatrix} = \begin{bmatrix} \Re(e'_{R,u}[v]) \\
\Im(e'_{R,u}[v]) \end{bmatrix},
\]

(24)

The second order statistics of the quantized received signal at an antenna in the \( j \)-th group can be approximated as

\[
\Lambda_i[v] = \frac{1}{U} \sum_{u=1}^U |q_{R,u,i}[v]|^2 = \tilde{x}[v]^2 \sum_{u=1}^U \eta^2 h_{R,u,i}^H[v] h_{R,u,i}[v] \]

\[
+ \frac{1}{U} \sum_{u=1}^U \epsilon'_{R,u,i}[v] h_{R,u,i}^H[v] \epsilon_{R,u,i}[v].
\]

(25)

Assuming quantization gains and quantization noise variance are equal across all antenna groups, the quantized received signal can be written as

\[
\Lambda_i[v] = \frac{1}{U} \sum_{u=1}^U |q_{R,u,i}[v]|^2 = \tilde{x}[v]^2 \eta^2 \alpha_i^2 + \tilde{\sigma}_e^2.
\]

(26)

Note due to channel hardening [20], as \( U \to \infty, \sum_{u=1}^U h_{R,u,i}^H[v] h_{R,u,i}[v] \) converges to a constant, \( \alpha_i \). A maximum likelihood detection approach based on the observation \( \Lambda \) is used to test the hypothesis that the symbol amplitude remains constant across adjacent symbols i.e., \( \tilde{x}[v] = \tilde{x}[v-1] \). More specifically, a hypothesis testing rule can be used to determine if \( b_1[v] = 1 \) or if \( b_1[v] = 0 \). This hypothesis is defined as \( H_1 \) and is confirmed if

\[
\Omega(\Lambda | H_1) > \Omega(\Lambda | H_0),
\]

(27)

where \( \Omega(\Lambda | H_1) \) is the conditional pdf of \( \Lambda[v] \). Hence, the hypothesis test in (27) is used to develop an energy detection threshold between two neighbouring DAPSK concentric circles.

Assuming that the channel amplitude, the quantization effect, and composite noise variance are known, the conditional pdf of \( \Lambda \) follows a non-central chi-square distribution and can be written as

\[
\Omega(\Lambda | \alpha, \tilde{x}[v], \eta, \tilde{\sigma}_e^2) = \frac{U}{\tilde{\sigma}_e^2} \left( \frac{\alpha^2 \tilde{x}^2[v] \eta^2}{\alpha^2 \tilde{x}^2[v] \eta^2} \right)^{U-1} e^{-\frac{U}{\tilde{\sigma}_e^2} (\Lambda + \alpha^2 \tilde{x}^2[v] \eta^2)} I_{U-1} \left( \frac{2U}{\tilde{\sigma}_e^2} \sqrt{\Lambda \alpha^2 \tilde{x}^2[v] \eta^2} \right),
\]

(28)

where \( U > 0 \) and \( I_{U-1} \) is the modified Bessel function of the first kind. While the distribution is dependent on the amplitude of the transmitted symbol, \( \tilde{x}[v] \in \{ \psi_0, \psi_1 \} \), the respective distributions conditioned on either \( \psi_0 \) or \( \psi_1 \) are not symmetric, therefore, no closed form solution for hypothesis testing can be developed. Hence, we employ a neural network to select the correct hypothesis.

Because \( h[v] = h[v-1] \), the second order statistics at the current and previous channel uses denoted by \( \lambda \) are correlated. More specifically, the elements of the following vector are correlated \( \lambda = \{ \lambda_1[v], \lambda_2[v-1] \} \), \( i \in \{ 1, 2 \} \). This vector will serve as one of the inputs to the neural network based amplitude detector. To detect the phase information with the 1-bit detector presented in (19), we allocate a particular group to use the signum quantization function. For instance, if the \( j \)-th group is reserved for phase detection, then \( q_{R,u,i}[v] \in \{-1, 1\} \), and \( \zeta_{2,j} = 0 \). Hence, with an amplitude ratio \( a'[v] = 1 \), the maximum likelihood phase detector can be written as

\[
\hat{s}_{R,VQL} = \arg \max U \prod_{l=1}^L \Phi \left( \sqrt{\rho} f_{R,u,i}^T[v] s_{R}^i[v] \right).
\]

(29)

C. Structure of Neural Network Based Amplitude Detector

The NN-based amplitude detector is represented by a DNN with \( L \) fully-connected (FC) layers. The operation of the \( l \)-th layer of the neural network can be described as

\[
w_l = f_{\theta_l}(w_l-1) = \varphi'(W^l w_l-1 + b^l),
\]

(30)

where \( W^l \) and \( b^l \) describes the weights and biases terms of the \( l \)-th layer, \( w_l-1 \) describes the output of the previous layer, \( \varphi' \) denotes the activation function of the \( l \)-th layer, and \( \theta = \{ W^l, b^l \} \) denotes the parameters of the \( l \)-th layer. The NN amplitude detector can be described as \( \theta = \{ \theta_1, \theta_2, \ldots, \theta_L \} \). Note that at the 1st layer, \( w_0 \) denotes the input to the NN-based decoder. This input consists of the second order statistics represented by \( \lambda \) concatenated with a one-hot representation of the measured SNR. The one-hot representation of the SNR is denoted as \( \varphi \). Hence, \( w_0 = \{ \lambda, \varphi \} \). The first \( L - 1 \) layers are equipped with a Relu activation function [21], while the \( L \)-th layer in the neural based decoder employs the Softmax function [22]. The output of the Softmax layer denoted by \( w_L \) provides two pseudo-probabilities, each denoting the likelihood that the decoded label is either zero or one. Hence, the operation of bit prediction can be described as

\[
\hat{b}_1[v] = \begin{cases} 0, & \text{if } \arg \max s_{R}^i[v] = 1 \ , \\ 1, & \text{else if } \arg \max s_{R}^i[v] = 2. \end{cases}
\]

(31)

In summary, the NN-based amplitude detector operates by comparing two signals received during adjacent channel uses and determines whether the transmitted symbols are both from the inner constellation or from the outer constellation. If the neural network determines that the transmitted symbols are from the same circle, the prediction is \( \hat{b}_1[v] = 0 \), otherwise the prediction is \( \hat{b}_1[v] = 1 \).

D. Training Procedure of Neural Network Based Amplitude Detector

The NN-based amplitude detector is trained offline using a randomly generated dataset. The dataset consists of a collection of blocks of bits, \( \{ b[v] \}_{v=1}^V \), and a corresponding
Table I: Training Parameters

| Parameter                  | Value   |
|----------------------------|---------|
| Learning Rate, $\alpha$    | 0.001   |
| Iterations (epochs)        | 250     |
| Batch size, $V$            | 1000    |

collection of differential modulated symbols, $\{x[v]\}_{v=1}^{V}$. The left-most bit in each block of bits is converted to one-hot encoded vectors with two elements denoted as $\{\omega[v]\}_{v=1}^{V}$. At each channel use, fading channel vectors and random Gaussian noise vectors are generated. The measured SNR is converted to one-hot encoded vectors. We adopt the binary cross entropy loss function for training

$$\mathcal{L}(\theta) = -\frac{1}{V} \sum_{v=1}^{V} \omega^T[v] \log(w^L),$$

(32)
equipped with this loss function and a learning rate, $\alpha$, the stochastic gradient descent algorithm is used to update the neural network parameters

$$\theta := \theta - \alpha \nabla \mathcal{L}(\theta).$$

(33)

In this work, an advanced version of the stochastic gradient descent algorithm - the Adam optimizer [23] is used to update the neural network parameters. The training parameters are presented in Table I. The test dataset is similarly generated. The instantaneous values of test dataset is different from the training dataset, but have the same statistics.

IV. NUMERICAL RESULTS

We perform Monte-Carlo simulations to evaluate the proposed differential amplitude and differential phase detection schemes. The proposed differential schemes is compared with corresponding coherent schemes presented in [3], [8]. We consider $N = 256$ channel uses. In the coherent scheme, a fraction of channel uses, $\xi$, is used for the transmission of pilot symbols. The differential scheme use no pilot symbols. The variable quantization level setup consist of three groups with equal number of antennas $|U_1| = |U_2| = |U_3| = \frac{U}{3}$. The second group is used for phase detection and each antenna in this group employs the signum function for quantization. The first and third groups employ one-bit quantization with the following thresholds, $\zeta_{2,1} = \frac{\psi_0 (1 + \cos(\pi/4))}{2}$ and $\zeta_{2,3} = -\zeta_{2,1}$. The output of the quantizers used in the first group is described as

$$q_{R,u_1,i}[v] = \begin{cases} 
\psi_1, & \text{if } y_{R,u_1,i}[v] > \zeta_{2,1}, \\
\psi_0, & \text{if } y_{R,u_1,i}[v] < \zeta_{2,1}, 
\end{cases}$$

and the output of the quantizers used in the second group is specified as

$$q_{R,u_3,i}[v] = \begin{cases} 
-\psi_0, & \text{if } y_{R,u_3,i}[v] > \zeta_{2,3}, \\
-\psi_1, & \text{if } y_{R,u_3,i}[v] < \zeta_{2,3}. 
\end{cases}$$

Figure 1: BER for DPSK in the differential system in comparison with the BER for PSK in the coherent scheme. The channel used in this simulation is frequency selective with $L = 31$ channel taps. $\xi$ indicates the fraction of channel uses need for pilot transmission in the coherent system. “Diff” indicates differential modulation.

Figure 2: Spectral efficiency attained in the differential system in comparison with the spectral efficiency attained using PSK in the coherent scheme. The channel used in this simulation is frequency selective with $L = 31$ channel taps. $\xi$ indicates the fraction of channel uses need for pilot transmission in the coherent system. “Diff” indicates differential modulation.
Figure 1 presents the BER incurred in decoding the block of information bits - \( \hat{b}[v] = [b_1[v], b_2[v], b_3[v], \ldots, b_{N_b}[v]] \). The phase information symbol (remaining \( N_b - 1 \) bits) is decoded with the neural network as presented in (31). The phase information symbol (remaining \( N_b - 1 \) bits) is decoded with (29) and \( \{b_2[v], b_3[v], \ldots, b_{N_b}[v]\} = \Upsilon^{-1}\left( \frac{s_{R,V}[v]}{\|s_{R,V}[v]\|} \right) \).

The BER attained for both the differential and coherent detectors decreases with an increase in receive antennas and with an increase in signal-to-noise ratio. The coherent system employs \( \xi = 0.5 \) of the total channel uses for transmitting pilot symbols. The BER performance of coherent 16-PSK is better than the BER performance of the differential 16-DPSK. The performance gap between the differential system and the coherent system is reduced when the modulation is increased from 16 to 32.

Figure 2 presents the spectral efficiency attained in both the differential and the coherent system. This spectral efficiency is calculated with

\[
S.E. = \begin{cases} 
N_b (1 - SER), & \text{if } SER \geq SER_{th}, \\
0, & \text{otherwise},
\end{cases}
\]

where \( SER_{th} \) is a threshold of the symbol error rate which is derived from the block error rate. In this work, this \( SER_{th} \) value is set to \( 5\% \). In the spectral efficiency plot, the advantage of the differential system is apparent. For a modulation order of 16, the coherent system has a better BER, but the spectral efficiency of the differential system is two-times better than the spectral efficiency of the coherent system. This is because the one-bit coherent system uses more than half of the available channel uses for pilot symbols. Note that at high SNR, the spectral efficiency of the differential system is much higher than the spectral efficiency of the coherent system, regardless of modulation order.

V. CONCLUSION

This article has investigated the uplink of a massive MIMO system with differential amplitude and phase modulation employed at the transmitter and one-bit ADCs employed at each receive antenna at the base station. The Bussgang theorem is used to express the quantized received signal in terms of quantized signals received during previous channel uses. With this expression, we derived the maximum likelihood expression for the differential encoded amplitude and phase information symbols. Because the maximum likelihood detector failed to decode the amplitude information, we developed and trained a neural network-based amplitude detector. We validated the performance of the proposed detectors through Monte-Carlo simulations and provided a comparison with coherent one-bit detectors. Our results indicate that the one-bit differential system outperforms the one-bit coherent system in terms of spectral efficiency.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Communications Magazine, vol. 52, no. 2, pp. 186–195, 2014.
[2] R. H. Walden, “Analog-to-digital converter survey and analysis,” IEEE Journal on Selected Areas in Communications, vol. 17, no. 4, pp. 539–550, 1999.
[3] J. Choi, D. J. Love, D. R. Brown, and M. Boutin, “Quantized Distributed Reception for MIMO Wireless Systems Using Spatial Multiplexing,” IEEE Transactions on Signal Processing, vol. 63, no. 13, pp. 3537–3548, 2015.
[4] J. Choi, J. Mo, and R. W. Heath, “Near Maximum-Likelihood Detector and Channel Estimator for Uplink Multiuser Massive MIMO Systems With One-Bit ADCs,” IEEE Transactions on Communications, vol. 64, no. 5, pp. 2005–2018, 2016.
[5] C. Studer and G. Durisi, “Quantized Massive MU-MIMO-OFDM Up-link,” IEEE Transactions on Communications, vol. 64, no. 6, pp. 2387–2399, 2016.
[6] C. Kong, A. Mezghani, C. Zhong, A. L. Swindlehurst, and Z. Zhang, “Multipair Massive MIMO Relaying Systems With One-Bit ADCs and DACs,” IEEE Transactions on Signal Processing, vol. 66, no. 11, pp. 2984–2997, 2018.
[7] O. T. Demir and E. Björnson, “The Bussgang Decomposition of Nonlinear Systems: Basic Theory and MIMO Extensions [Lecture Notes],” IEEE Signal Processing Magazine, vol. 38, no. 1, pp. 131–136, 2021.
[8] J. Choi, D. J. Love, and D. R. Brown, “Channel estimation techniques for quantized distributed reception in MIMO systems,” in 2014 48th Asilomar Conference on Signals, Systems and Computers, 2014, pp. 1066–1070.
[9] D. Divsalar and M. K. Simon, “Maximum-likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channels/spl minus/metrics and performance,” IEEE Transactions on Communications, vol. 42, no. 1, pp. 76–89, 1994.
[10] Feng-Wen Sun and H. Leib, “Multiple-phase codes for detection without carrier phase reference,” IEEE Transactions on Information Theory, vol. 44, no. 4, pp. 1477–1491, 1998.
[11] R. Knopp and H. Leib, “M-aray phase coding for the noncoherent AWGN channel,” IEEE Transactions on Information Theory, vol. 40, no. 6, pp. 1968–1984, 1994.
[12] R. Wei, Y. Chen, and C. Kuo, “Further Results on Noncoherent Block-Coded MPSK,” in 2006 IEEE International Conference on Communications, vol. 3, 2006, pp. 1231–1236.
[13] R. Wei, “Differential Encoding by a Look-Up Table for Quadrature-Amplitude Modulation,” IEEE Transactions on Communications, vol. 59, no. 1, pp. 84–94, 2011.
[14] A. Schenk and R. F. H. Fischer, “Noncoherent Detection in Massive MIMO Systems,” in WSA 2013: 17th International ITG Workshop on Smart Antennas, 2013, pp. 1–8.
[15] M. Chowdhury, A. Manolakos, and A. J. Goldsmith, “Design and performance of noncoherent massive SIMO systems,” in 2014 48th Annual Conference on Information Sciences and Systems (CISS), 2014, pp. 1–6.
[16] A. Manolakos, M. Chowdhury, and A. J. Goldsmith, “Constellation design in noncoherent massive SIMO systems,” in 2014 IEEE Global Communications Conference, 2014, pp. 3690–3695.
[17] T. O’Shea and J. Hoydis, “An Introduction to Deep Learning for the Physical Layer,” IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, 2017.
[18] T. J. O’Shea, T. Erpek, and T. C. Clancy, “Physical layer deep learning of encodings for the MIMO fading channel,” in 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2017, pp. 76–80.
[19] T. Van Luong, Y. Ko, N. A. Vien, M. Matthaiou, and H. Q. Ngo, “Deep Energy Autencoder for Noncoherent Multicarrier MU-SIMO Systems,” IEEE Transactions on Wireless Communications, vol. 19, no. 6, pp. 3952–3962, 2020.
[20] H. Q. Ngo and E. G. Larsson, “No Downlink Pilots Are Needed in TDD Massive MIMO,” IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 2921–2935, 2017.
[21] A. F. Agarap, “Deep learning using rectified linear units (relu),” arXiv preprint arXiv:1802.08375, 2018.
[22] J. S. Bridle, “Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters,” in Advances in neural information processing systems, 1990, pp. 211–217.
[23] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.