A Distributed MIMO Relay Scheme Inspired by Backpropagation Algorithm

Rui Wang*, Yi Jiang*, Wei Zhang†
*School of Information Science and Technology, Fudan University, Shanghai, China
†School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia
Email: {ruiwang18, yijiang}@fudan.edu.cn, w.zhang@unsw.edu.au

GLOBECOM 2021 - 2021 IEEE Global Communications Conference | 978-1-7281-8104-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GLOBECOM46510.2021.9685608

***Abstract*—This paper studies a distributed scheme for a multi-input multi-output (MIMO) relay network, where the transmit nodes are subject to the nonlinear instantaneous power constraints. We introduce a novel perspective of regarding a relay network as a so-termed *quasi-neural network* by drawing its striking analogies with a (four-layer) artificial neural network (ANN). We propose a nonlinear amplify-and-forward (NAF) scheme inspired by the back-propagation (BP) algorithm, namely the NAF-BP, to optimize the transceivers to maximize the output signal-to-interference-plus-noise ratio (SINR) of the data streams. The NAF-BP algorithm can be implemented in a distributed manner with no channel state information (CSI) and no data exchange between the relay nodes. The NAF-BP can also coordinate the distributed relay nodes to form a virtual array to suppress interferences from unknown directions. Extensive simulations verify the effectiveness of the proposed scheme.**

***Index Terms*—relay network; quasi-neural network; interference suppression; back-propagation algorithm; nonlinear amplify-and-forward**

I. Introduction

In order to meet the ever-increasing demand for higher network capacity, various network architectures have been proposed/adopted in 5G/B5G communications, including the macrocells, microcells, small cells, and relay networks [1]. Network density and communication reliability can be improved via relaying and multi-hop communications, which are important components in 5G/B5G mobile communications [2]. Three types of relay schemes have been proposed: the compress-and-forward (CF) scheme [3], the decode-and-forward (DF) scheme [4], and the amplify-and-forward (AF) scheme [5], [6], among which the AF scheme is the most popular owing to its simplicity and decent performance.

While most existing AF schemes ignore the nonlinearity of the power amplifiers (PA) and only assume an *average* transmit power constraint [6]–[8], some papers [9]–[11] consider PA's *instantaneous* amplitude limitation, which is more realistic than the average power constraint. As the instantaneous limitation is nonlinear, the corresponding AF relay schemes are referred to as nonlinear amplify and forward (NAF) schemes. The authors in [9]–[11] apply the NAF scheme to an orthogonal frequency division multiplexing (OFDM) system and use Bussgang's theorem for optimization. But these work are limited to the single-antenna scenario.

Work in this paper was supported by National Natural Science Foundation of China Grant No. 61771005.

In this paper, we design a relay scheme from a novel perspective by drawing the interesting similarities between a relay network and an artificial neural network (ANN) as follows:

- i) the source, the destination, and the relay nodes are like the neurons of different layers of an ANN;
- ii) the data transmission from the source node to the relay nodes and then to the destination is like the propagation of the training data between the layers in the ANN;
- iii) the beamforming weights of the source and the relays, and the equalizer of the destination are like the connecting weights between the neurons in the ANN;
- iv) the nonlinear PAs of the relays and the source are like the neurons' activation functions.

Given these striking analogies, we propose to regard a relay network as a so-termed *quasi-neural network*.

Now that a back-propagation (BP) algorithm [12] is widely used to train a neural network, we can use it – with modifications of course – to optimize the processing functions of the destination, the relay nodes, and the source. We propose the so-termed NAF-BP scheme to minimize the mean squared error (MSE) of the data streams, through back propagating some derivatives from the destination to the relay nodes and then from the relays to the source.

More important, the proposed relay scheme can be implemented in a distributed manner with no channel state information (CSI), nor information exchange between the relay nodes, except for a set of training sequences. In contrast, most existing work on the relay network assumes that a global CSI is known to the relay nodes, precisely or imprecisely [6], [13].

Furthermore, the NAF-BP can coordinate relay nodes to form a virtual array to achieve interference suppression [14], a problem inadequately investigated in the relay network literature, despite its growing importance with the increasingly crowded frequency spectrum. In [7], the authors consider a relay network interfered by a jammer, assuming that a processing center can aggregate the received signals of the relay and the destination. In [15], the authors consider the inter-cell interference for the cell-edge users adjacent to the relay and proposed a coordinate ascent algorithm, assuming that the relay and the destination share perfect CSI. In contrast, the proposed NAF-BP requires no CSI nor a central unit that aggregates data. The effectiveness of the proposed scheme is verified by the extensive simulations.

Notations: $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ stand for conjugate, transpose, and conjugate transpose, respectively. $\mathbb{Z}$ is the set of integers. $\sigma\circ\mathbf{V}$ stands for a composite function of $\sigma(\cdot)$ and $\mathbf{V}$. $\mathbb{C}^{N\times K}$ is the set of $N\times K$ complex matrices. $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix with vector $\mathbf{a}$ being its diagonal and $\mathrm{vec}(\cdot)$ denotes a vectorization operation of stacking the columns of the matrix into a long column-vector. $|\mathbf{a}|$ stands for taking absolute value of $\mathbf{a}$ element-wise and $\mathbf{a}\le\mathbf{b}$ stands for $a_i\le b_i$ element-wise.

II. System Model and Problem Formulation

A. System Model

We consider the relay network as shown in Fig. 1, which consists of one M_s-antenna source, N M_r-input M_r-output relay nodes, and one M_d-antenna destination. The source wants to transmit N_s data streams $\{\mathbf{s}(i)\in\mathbb{C}^{N_s},\ i\in\mathbb{Z}\}$ via the relay nodes to the destination.

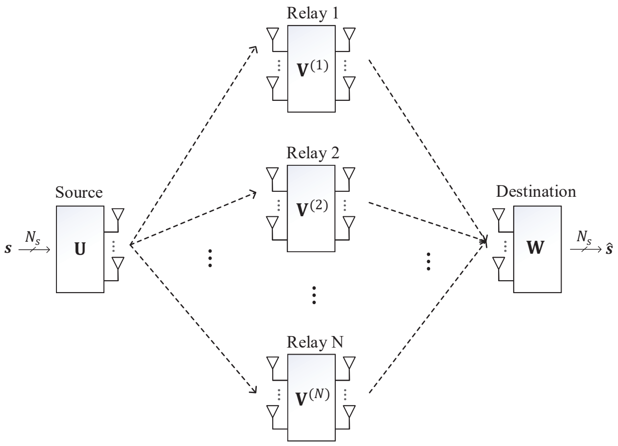

Fig. 1: A relay network.

Before transmitting signal $\mathbf{s}$, the source processes it as

$$\mathbf{x}=f_s(\mathbf{s}),\tag{1}$$

where $f_s:\mathbb{C}^{N_s}\to\mathbb{C}^{M_s}$ is the processing function of the source. Given N relay nodes, of which the n-th receives

$$\mathbf{r}^{(n)}=\mathbf{H}_r^{(n)}\mathbf{x}+\boldsymbol{\eta}_r^{(n)},\ n=1,\cdots,N,\tag{2}$$

where the superscript (n) refers to the n-th relay, i.e., $\mathbf{H}_r^{(n)}\in\mathbb{C}^{M_r\times M_s}$ represents the channel between the source and the n-th relay, which is frequency flat, and $\boldsymbol{\eta}_r^{(n)}\sim\mathcal{CN}(0,\sigma_r^2\mathbf{I})$ is the noise.

To avoid self-interference, the relay nodes work in the frequency division duplex (FDD) mode, i.e., to receive and transmit on two different frequencies. Using $f_r:\mathbb{C}^{M_r}\to\mathbb{C}^{M_r}$ as the processing function, the n-th relay transmits

$$\mathbf{a}^{(n)}=f_r(\mathbf{r}^{(n)}).\tag{3}$$

Note that $f_r(\cdot)$ is as general as $f_s(\cdot)$. Denote

$$\tilde{\mathbf{a}}\triangleq\mathrm{vec}\left(\left[\mathbf{a}^{(1)},\mathbf{a}^{(2)},\cdots,\mathbf{a}^{(N)}\right]\right)\in\mathbb{C}^{NM_r}.\tag{4}$$

The destination receives

$$\mathbf{y}=\mathbf{H}_d\tilde{\mathbf{a}}+\boldsymbol{\eta}_d,\tag{5}$$

where $\mathbf{H}_d\in\mathbb{C}^{M_d\times NM_r}$ is the relay-to-destination channel and $\boldsymbol{\eta}_d\sim\mathcal{CN}(0,\sigma_d^2\mathbf{I})$ is the noise. Here we assume that the clocks of the relay nodes are synchronized and thus $\mathbf{a}^{(i)}$'s are time-aligned [16].

Applying the processing function $f_d:\mathbb{C}^{M_d}\to\mathbb{C}^{N_s}$ to the received signal, the destination yields

$$\hat{\mathbf{s}}=f_d(\mathbf{y}).\tag{6}$$

These processing functions $f_s(\cdot),f_r(\cdot)$, and $f_d(\cdot)$ are the transmit and receive functionalities of the nodes; thus we also refer to them as the transceivers of the relay network.

B. Problem Formulation

This paper focuses on optimizing the transceivers by the minimum mean squared error (MMSE) criterion subject to the *instantaneous* amplitude/power constraint *per antenna*, i.e.,

$$\begin{aligned}\min_{f_s(\cdot),f_r(\cdot),f_d(\cdot)}\quad &\mathbb{E}\left\|\hat{\mathbf{s}}-\mathbf{s}\right\|^2\\ s.t.\quad &|\mathbf{x}|\le\mathbf{1},|\tilde{\mathbf{a}}|\le\mathbf{1},\end{aligned}\tag{7}$$

where $\mathbf{x}$ and $\tilde{\mathbf{a}}$ are given in (1) and (4), respectively, and to set the amplitude limit be 1 entails no loss of generality.

Note that minimizing the MSE amounts to maximizing the output signal-to-noise ratio (SNR), because [17]

$$\mathsf{SNR}=\frac{1}{\mathsf{MSE}}-1.\tag{8}$$

To guarantee the instantaneous power constraint, after applying the linear weights to obtain

$$\mathbf{z}\triangleq\mathbf{U}\mathbf{s}\in\mathbb{C}^{M_s},\tag{9}$$

and

$$\mathbf{b}^{(n)}\triangleq\mathbf{V}^{(n)H}\mathbf{r}^{(n)}\in\mathbb{C}^{M_r},\ n=1,\cdots,N,\tag{10}$$

we use the *nonlinear* soft envelope limiter (SEL) [18] [1]

$$\sigma(x)\triangleq\begin{cases}x & |x|\le1\\ e^{j\angle(x)} & |x|>1,\end{cases}\tag{11}$$

to model the nonlinear PA; here $\angle(\cdot)$ stands for taking the angle of a variable. Although one may use a PA model more sophisticated than (11), such as [19, Chapter 3.5]

$$\sigma(x)=\frac{|x|e^{j\angle(x)}}{(1+|x|^{2p})^{\frac{1}{2p}}},\tag{12}$$

it will cause no fundamental difference to the proposed algorithm.

Hence the transmitted signals from the source and the relay nodes are

$$\mathbf{x}=\sigma(\mathbf{z})=\sigma(\mathbf{U}\mathbf{s}),\tag{13}$$

and

$$\mathbf{a}^{(n)}=\sigma(\mathbf{b}^{(n)})=\sigma(\mathbf{V}^{(n)H}\mathbf{r}^{(n)}),n=1,\cdots,N,\tag{14}$$

respectively, where the clipping $\sigma(\cdot)$ is applied element-wise to $\mathbf{z}$ and $\mathbf{b}^{(n)}$. As the source node's $f_s(\mathbf{s})=\sigma\circ\mathbf{U}\mathbf{s}$ and the relay

[1] Although it is an abuse of notation to use σ at the risk of causing confusion with the noise power, we adopt it to emphasize its connection to the activation function in ANN.

nodes' $f_r(\mathbf{r}^{(n)}) = \sigma \circ \mathbf{V}^{(n)}\mathbf{r}^{(n)}$ are both nonlinear composite functions, we refer to such a relay method as a NAF scheme.

In the end, the destination node applies a linear beamformer to obtain

$$\hat{\mathbf{s}} = \mathbf{W}^H \mathbf{y} \in \mathbb{C}^{N_s}. \tag{15}$$

Now (7) becomes an unconstrained optimization problem

$$\min_{\mathbf{U},\mathbf{V}^{(n)},\mathbf{W}} \mathbb{E}\left\| \mathbf{W}^H \left\{ \mathbf{H}_d \begin{bmatrix} \sigma\left(\mathbf{V}^{(1)H}\left(\mathbf{H}_r^{(1)}\sigma(\mathbf{U}\mathbf{s}) + \boldsymbol{\eta}_r^{(1)}\right)\right) \\ \vdots \\ \sigma\left(\mathbf{V}^{(N)H}\left(\mathbf{H}_r^{(N)}\sigma(\mathbf{U}\mathbf{s}) + \boldsymbol{\eta}_r^{(N)}\right)\right) \end{bmatrix} \right.\right.$$
$$\left.\left. + \boldsymbol{\eta}_d \right\} - \mathbf{s} \right\|^2. \tag{16}$$

Although (16) appears challenging, we introduce a novel perspective of regarding a relay network as a *quasi-neural network* and hence can borrow the idea of the BP algorithm to optimize the nodes of the relay network.

C. The Analogies Between A Relay Network and An ANN

We present the diagram of the NAF relay network in Fig. 2, and observe its striking analogies to the ANN shown in Fig. 3, which are as follows:

i) the data streams, the transmit antennas of the source and the relays, and the output of the destination are analogous to the different layers of the four-layer ANN, respectively;

ii) the operations $\mathbf{U}$, $\{\mathbf{V}^{(n)} \circ \mathbf{H}_r^{(n)}, n = 1, \cdots, N\}$, and $\mathbf{W} \circ \mathbf{H}_d$ in the relay network are analogous to the connection weights $\boldsymbol{\omega}^{(l)}, l = 1, 2, 3$ in the four-layer ANN, respectively;

iii) the SEL operation $\sigma(\cdot)$ of the source and the relays is analogous to the activation function of the first and the second hidden layers in ANN, although the latter typically uses a Rectified Linear Unit (ReLU) or a Sigmoid function.

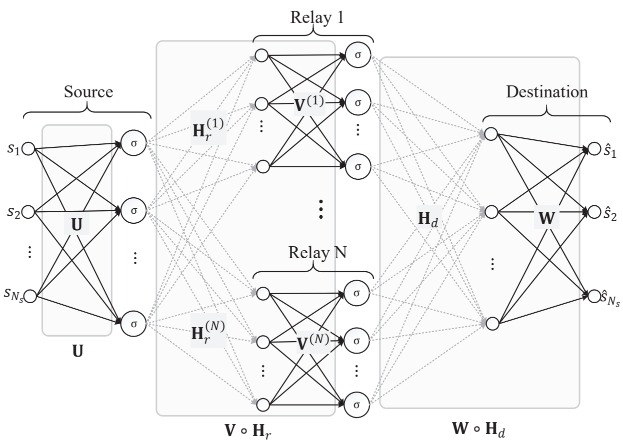

Fig. 2: The diagram of the NAF relay network.

Since the connection weights in an ANN can be effectively optimized using the BP algorithm, one can imagine that the relay network may be similarly optimized, as explained in the next.

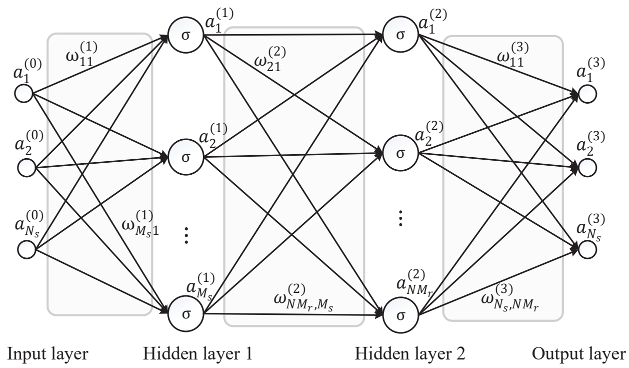

Fig. 3: A four-layer ANN.

III. The Distributed NAF-BP Scheme

Assuming that the CSI is static but unknown, we propose a NAF scheme that can solve (16) based on pilot sequences by borrowing the idea of the BP algorithm.

Considering

$$J \triangleq \|\hat{\mathbf{s}} - \mathbf{s}\|^2 \tag{17}$$

as a single realization of the cost function in (16), we attempt to minimize J with respect to the weights of the source, the relays, and the destination. To this end, we derive the gradients $\frac{\partial J}{\partial \mathbf{U}^*}$, $\{\frac{\partial J}{\partial \mathbf{V}^{(n)*}}\}_{n=1}^N$, and $\frac{\partial J}{\partial \mathbf{W}^*}$ based on a single sample of the pilot $\mathbf{s}(i)$ as follows.

According to the chain rule, the derivative of J with respect to the weights of the destination is

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{W}^*} &= \left(\frac{\partial J}{\partial \mathbf{w}_1^*}, \frac{\partial J}{\partial \mathbf{w}_2^*}, \cdots, \frac{\partial J}{\partial \mathbf{w}_{N_s}^*}\right) \\ &= \left(\frac{\partial J}{\partial \hat{s}_1}\frac{\partial \hat{s}_1}{\partial \mathbf{w}_1^*}, \frac{\partial J}{\partial \hat{s}_2}\frac{\partial \hat{s}_2}{\partial \mathbf{w}_2^*}, \cdots, \frac{\partial J}{\partial \hat{s}_1}\frac{\partial \hat{s}_{N_s}}{\partial \mathbf{w}_{N_s}^*}\right) \\ &\overset{(a)}{=} \mathbf{y}\left(\frac{\partial J}{\partial \hat{\mathbf{s}}^*}\right)^H, \end{aligned} \tag{18}$$

where $\overset{(a)}{=}$ holds because $\frac{\partial \hat{s}_m}{\partial \mathbf{w}_m^*} = \mathbf{y}, m = 1, \cdots, N_s$ [cf. (15)], and

$$\frac{\partial J}{\partial \hat{\mathbf{s}}^*} = \hat{\mathbf{s}} - \mathbf{s} \tag{19}$$

follows immediately from (17).

Similar to (18), the derivative with respect to the weights of the n-th relay is

$$\frac{\partial J}{\partial \mathbf{V}^{(n)*}} = \mathbf{r}^{(n)}\left(\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right)^H, n = 1, \cdots, N, \tag{20}$$

where we have used $\frac{\partial b_q^{(n)}}{\partial \mathbf{v}_q^{(n)*}} = \mathbf{r}^{(n)}$, $q = 1, \ldots, M_r$ [cf. (10)]. Note that

$$\frac{\partial J}{\partial \mathbf{b}^{(n)*}} = \frac{\partial \mathbf{a}^{(n)*}}{\partial \mathbf{b}^{(n)*}}\frac{\partial J}{\partial \mathbf{a}^{(n)*}} + \frac{\partial \mathbf{a}^{(n)}}{\partial \mathbf{b}^{(n)*}}\left(\frac{\partial J}{\partial \mathbf{a}^{(n)*}}\right)^*, \tag{21}$$

where $\frac{\partial J}{\partial \mathbf{a}^{(n)*}}$ can be extracted from $\frac{\partial J}{\partial \tilde{\mathbf{a}}^*}$ [cf. (4)] and

$$\frac{\partial J}{\partial \tilde{\mathbf{a}}^*} = \frac{\partial \hat{\mathbf{s}}^H}{\partial \tilde{\mathbf{a}}^*}\frac{\partial J}{\partial \hat{\mathbf{s}}^*}. \tag{22}$$

Combining (5) and (15) leads to

$$\hat{s} = \mathbf{W}^H \mathbf{H}_d \tilde{\mathbf{a}} + \mathbf{W}^H \boldsymbol{\eta}_d, \tag{23}$$

from which we obtain

$$\frac{\partial \hat{\mathbf{s}}^H}{\partial \tilde{\mathbf{a}}^*} = \mathbf{H}_d^H \mathbf{W}. \tag{24}$$

Inserting (24) and (19) into (22) yields $\frac{\partial J}{\partial \tilde{\mathbf{a}}^*} = \mathbf{H}_d^H \mathbf{W} \frac{\partial J}{\partial \hat{\mathbf{s}}^*}$; thus,

$$\frac{\partial J}{\partial \mathbf{a}^{(n)*}} = \mathbf{H}_d^{(n)H} \mathbf{W} \frac{\partial J}{\partial \hat{\mathbf{s}}^*}, \tag{25}$$

where $\mathbf{H}_d^{(n)} \in \mathbb{C}^{M_d \times M_r}$ is the channel between the n-th relay and the destination;

Since $\mathbf{a}^{(n)} = \sigma(\mathbf{b}^{(n)})$ [cf. (14)] is an element-wise function of $\mathbf{b}^{(n)}$, $\frac{\partial \mathbf{a}^{(n)*}}{\partial \mathbf{b}^{(n)*}}$ and $\frac{\partial \mathbf{a}^{(n)}}{\partial \mathbf{b}^{(n)*}}$ are diagonal as

$$\frac{\partial \mathbf{a}^{(n)*}}{\partial \mathbf{b}^{(n)*}} = \operatorname{diag}\left(\frac{\partial a_1^{(n)*}}{\partial b_1^{(n)*}}, \cdots, \frac{\partial a_{M_r}^{(n)*}}{\partial b_{M_r}^{(n)*}} \right), \tag{26}$$

and

$$\frac{\partial \mathbf{a}^{(n)}}{\partial \mathbf{b}^{(n)*}} = \operatorname{diag}\left(\frac{\partial a_1^{(n)}}{\partial b_1^{(n)*}}, \cdots, \frac{\partial a_{M_r}^{(n)}}{\partial b_{M_r}^{(n)*}} \right), \tag{27}$$

with

$$\frac{\partial a_q^{(n)*}}{\partial b_q^{(n)*}} = \begin{cases} 1 & \left| b_q^{(n)} \right| \le 1 \\ \frac{1}{2\left| b_q^{(n)} \right|} & \left| b_q^{(n)} \right| > 1 \end{cases}, \; q = 1, \cdots, M_r, \tag{28}$$

and

$$\frac{\partial a_q^{(n)}}{\partial b_q^{(n)*}} = \begin{cases} 0 & \left| b_q^{(n)} \right| \le 1 \\ -\frac{1}{2\left| b_q^{(n)} \right|} e^{j2\angle b_q^{(n)}} & \left| b_q^{(n)} \right| > 1 \end{cases}, \; q = 1, \cdots, M_r. \tag{29}$$

Inserting (25)-(27) into (21) leads to $\frac{\partial J}{\partial \mathbf{b}^{(n)*}}$; and inserting it in (20) yields the gradient $\frac{\partial J}{\partial \mathbf{V}^{(n)*}}$.

In a way similar to the above derivations, we obtain the derivative with respect to the weights of the source as

$$\frac{\partial J}{\partial \mathbf{U}^*} = \frac{\partial J}{\partial \mathbf{z}^*} \mathbf{s}^H, \tag{30}$$

where

$$\frac{\partial J}{\partial \mathbf{z}^*} = \frac{\partial \mathbf{x}^*}{\partial \mathbf{z}^*} \frac{\partial J}{\partial \mathbf{x}^*} + \frac{\partial \mathbf{x}}{\partial \mathbf{z}^*} \left(\frac{\partial J}{\partial \mathbf{x}^*} \right)^*, \tag{31}$$

$$\frac{\partial J}{\partial \mathbf{x}^*} = \sum_{n=1}^{N} \mathbf{H}_r^{(n)H} \mathbf{V}^{(n)} \frac{\partial J}{\partial \mathbf{b}^{(n)*}}, \tag{32}$$

$$\frac{\partial \mathbf{x}^*}{\partial \mathbf{z}^*} = \operatorname{diag}\left(\frac{\partial x_1^*}{\partial z_1^*}, \cdots, \frac{\partial x_{M_s}^*}{\partial z_{M_s}^*} \right), \tag{33}$$

and

$$\frac{\partial \mathbf{x}}{\partial \mathbf{z}^*} = \operatorname{diag}\left(\frac{\partial x_1}{\partial z_1^*}, \cdots, \frac{\partial x_{M_s}}{\partial z_{M_s}^*} \right), \tag{34}$$

with

$$\frac{\partial x_p^*}{\partial z_p^*} = \begin{cases} 1 & |z_p| \le 1 \\ \frac{1}{2|z_p|} & |z_p| > 1 \end{cases}, \; p = 1, \cdots, M_s, \tag{35}$$

and

$$\frac{\partial x_p}{\partial z_p^*} = \begin{cases} 0 & |z_p| \le 1 \\ -\frac{1}{2|z_p|} e^{j2\angle z_p} & |z_p| > 1 \end{cases}, \; p = 1, \cdots, M_s. \tag{36}$$

Now all the distributed nodes can update their processing weights based on the derivatives (18), (20) and (30). Indeed, the destination node can update $\mathbf{W}$ according to (18) using $\mathbf{y}$ and $\frac{\partial J}{\partial \hat{\mathbf{s}}^*}$ given in (19), of which both are locally available given known pilot $\mathbf{s}$. Therefore, the destination can update its weight without the CSI.

For the n-th relay node, $\mathbf{r}^{(n)}$ is locally available in (20). It only needs to obtain $\frac{\partial J}{\partial \mathbf{b}^{(n)*}}$ as shown in (21), where $\frac{\partial \mathbf{a}^{(n)}}{\partial \mathbf{b}^{(n)*}}$ and $\frac{\partial \mathbf{a}^{(n)*}}{\partial \mathbf{b}^{(n)*}}$ are also locally available. To obtain $\frac{\partial J}{\partial \mathbf{a}^{(n)*}}$ as shown in (25), we let the destination broadcast the beamformed derivative $\left(\mathbf{W}\frac{\partial J}{\partial \hat{\mathbf{s}}^*}\right)^*$ to the relay nodes through the reverse channel. Owing to the channel reciprocity, the n-th relay will receive the signal $\mathbf{H}_d^{(n)T}\left(\mathbf{W}\frac{\partial J}{\partial \hat{\mathbf{s}}^*}\right)^* = \left(\frac{\partial J}{\partial \mathbf{a}^{(n)*}}\right)^*$. Hence the n-th relay obtains the derivative (20) without knowing the CSI and without data exchange between other relays.

The source node can obtain the derivative in (30) in a way similar to the relay node's obtaining (20). The relay nodes transmit $\left[\mathbf{V}^{(n)}\left(\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right)\right]^*, n = 1, \ldots, N$ in a time-aligned manner, thus the source will obtain the superimposed term $\sum_{n=1}^{N} \mathbf{H}_r^{(n)H} \mathbf{V}^{(n)} \left(\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right)$, i.e., (32), and obtain the derivative $\frac{\partial J}{\partial \mathbf{U}^*}$.

Now we see that the distributed NAF scheme is conducted *over-the-air* and obtains derivatives through two phases as follows: i) the forward propagation of the signal, in which the source transmits the pilot sequences processed by $\sigma \circ \mathbf{U}$ and the relays forward the received samples being processed by $\sigma \circ \mathbf{V}^{(n)}, n = 1, \cdots, N$; ii) the backpropagation of the derivatives, in which the destination and the relays broadcast the beamformed derivatives, i.e., $\left(\mathbf{W}\frac{\partial J}{\partial \hat{\mathbf{s}}^*}\right)^*$ and $\left(\mathbf{V}^{(n)}\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right)^*, n = 1, \cdots, N$ in the back-propagation channel, respectively.

In the simulation, we consider the realistic constraint that the feedback derivatives need to be clipped by the PAs, i.e., the destination transmits $\sigma\left(\left(\mathbf{W}\frac{\partial J}{\partial \hat{\mathbf{s}}^*}\right)^*\right)$ rather than $\left(\mathbf{W}\frac{\partial J}{\partial \hat{\mathbf{s}}^*}\right)^*$, and each relay transmits $\sigma\left(\left(\mathbf{V}^{(n)}\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right)^*\right)$ rather than $\left(\mathbf{V}^{(n)}\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right)^*$. But even with this constraint, the NAF-BP algorithm works, because the derivatives are typically quite small especially when near to the convergence.

The aforementioned derivatives are based on a one-sample pilot. For the general case of L-length pilot sequence, we can average the L derivatives to be

$$\overline{\mathbf{d}}_{\mathbf{u}} = \frac{1}{L} \sum_{i=1}^{L} \frac{\partial J}{\partial \mathbf{U}^*}(i), \tag{37}$$

$$\overline{\mathbf{d}}_{\mathbf{v}^{(n)}} = \frac{1}{L} \sum_{i=1}^{L} \frac{\partial J}{\partial \mathbf{V}^{(n)*}}(i), \tag{38}$$

and

$$\overline{\mathbf{d}}_{\mathbf{w}} = \frac{1}{L}\sum_{i=1}^{L}\frac{\partial J}{\partial \mathbf{W}^*}(i). \tag{39}$$

Furthermore, with multiple sets of pilot sequences, we can use the momentum method [12] to update the derivatives as

$$\mathbf{d}_{\mathbf{u}}(t) = \lambda \mathbf{d}_{\mathbf{u}}(t-1) + (1-\lambda)\overline{\mathbf{d}}_{\mathbf{u}}(t), \tag{40}$$

$$\mathbf{d}_{\mathbf{v}^{(n)}}(t) = \lambda \mathbf{d}_{\mathbf{v}^{(n)}}(t-1) + (1-\lambda)\overline{\mathbf{d}}_{\mathbf{v}^{(n)}}(t), \tag{41}$$

and

$$\mathbf{d}_{\mathbf{w}}(t) = \lambda \mathbf{d}_{\mathbf{w}}(t-1) + (1-\lambda)\overline{\mathbf{d}}_{\mathbf{w}}(t), \tag{42}$$

where $t \in \{1, 2, ...T\}$ is the set index and $\lambda \in (0, 1)$ is the momentum parameter. The processing coefficients are then updated by

$$\mathbf{U}(t) = \mathbf{U}(t-1) - \alpha \mathbf{d}_{\mathbf{u}}(t), \tag{43}$$

$$\mathbf{V}^{(n)}(t) = \mathbf{V}^{(n)}(t-1) - \alpha \mathbf{d}_{\mathbf{v}^{(n)}}(t),\ n = 1, \cdots, N, \tag{44}$$

and

$$\mathbf{W}(t) = \mathbf{W}(t-1) - \alpha \mathbf{d}_{\mathbf{w}}(t), \tag{45}$$

where $\alpha \in (0, 1)$ is the learning rate.

Therefore, the distributed NAF-BP algorithm can optimize the relay network using no explicit channel information, nor information exchange between the distributed relay nodes. Moreover, the NAF-BP algorithm can achieve the MMSE solution in the presence of interferences even without knowing the directions of the interferences.

IV. Numerical Simulations

In this section, we verify the effectiveness of the proposed NAF scheme via numerical simulations. The source-to-relay channel and the relay-to-destination channel are assumed to be frequency-flat Rayleigh fading and remain static in the simulated time duration. The pilot sequence length $L = 100$. The nonlinear PA is simulated as the clipping function $\sigma(\cdot)$ in (11).

We use the output SNR/SINR of each stream of the destination according to (8) as the metric to evaluate the performance of the system, based on the average of 500 Monte Carlo trials. $\rho_{\text{relay}}(=\frac{M_s}{\sigma_r^2})$ and $\rho_{\text{dest}}(=\frac{NM_r}{\sigma_d^2})$ are the SNRs of the source-to-relay and the relay-to-destination channels, respectively. The momentum parameters used in (40)−(45) are $\alpha = 0.3$ and $\lambda = 0.9$. Other parameters are given on the top of the figures.

In the first example, we simulate the distributed NAF-BP scheme to see its convergence behavior under different number of the relay nodes ($N = 2$ or 4) and different number of the relay antennas ($M_r = 2$ or 4), where the pilot is the QPSK signal and the legends 'stream #1' or 'stream #2' represent the first or the second stream. Fig. 4 shows that the output SNR improves as the number of relays and relay antennas increases, which is not surprising. We also see that most gain is achieved in the first $T = 200$ iterations, where one iteration represents one round of forward and backward training sessions, so the NAF-BP is based on 200 iterations for the subsequent simulations.

We then simulate the scenario under interferences. Fig. 5 shows that a relay network with 4 single-antenna relays running the distributed NAF-BP algorithm can suppress up to 3 interferences owing to the inter-relay coordination. In other words, the relay nodes are being coordinated by the destination node to form a virtual array for interference suppression but with no information exchange among themselves.

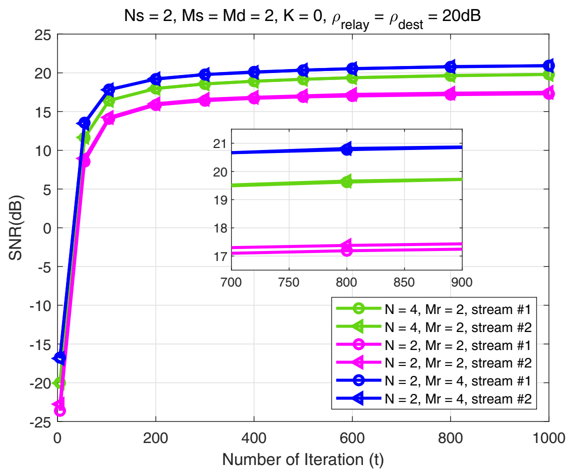

Fig. 4: Output SNR of the relay network achieved by the NAF-BP as the number of iterations with respect to different number of the relay nodes and the relay antennas.

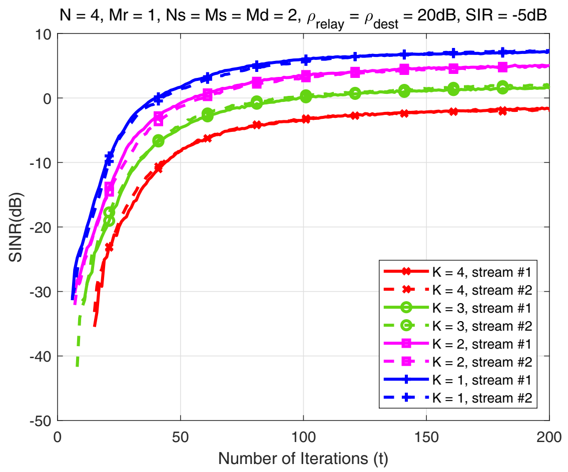

Fig. 5: Output SINR of the destination with respect to different number of the interferences.

In the previous simulations, we have assumed that the back-propagation from the destination to the relay nodes and from the relays to the source are noise-free. Now we consider a more realistic scenario where the destination's and the relays' broadcasting of the sequence $[\mathbf{w}(\hat{s} - s)]^*$ and $\left[\mathbf{V}^{(n)}\frac{\partial J}{\partial \mathbf{b}^{(n)*}}\right]^*, n = 1, \cdots, N$ are contaminated by the noise $\boldsymbol{\zeta}_d \sim \mathcal{CN}(0, \sigma_{\zeta_d}^2\mathbf{I})$ and $\boldsymbol{\zeta}_r \sim \mathcal{CN}(0, \sigma_{\zeta_r}^2\mathbf{I})$, respectively. The SNRs of the reverse channel are defined as

$$\rho_{\text{back1}} \triangleq \frac{M_d}{\sigma_{\zeta_d}^2}, \text{and } \rho_{\text{back2}} \triangleq \frac{NM_r}{\sigma_{\zeta_r}^2}. \tag{46}$$

Fig. 6 shows that the loss of the output SNR of each stream is lower than 1dB even when the SNR of the reverse channel is only 0dB. It verifies that the distributed NAF-BP algorithm is robust to the noise in the back-propagation channel, owing to the average operation in (37)-(39).

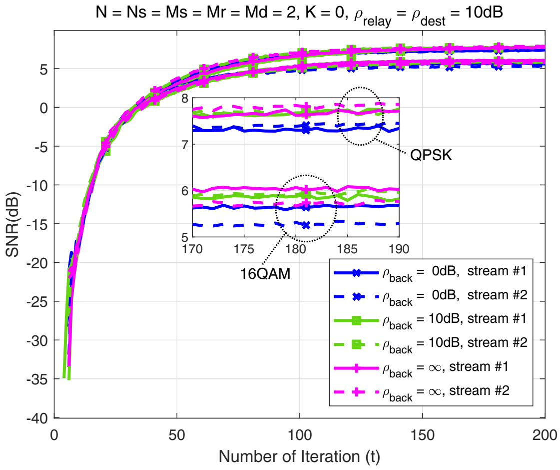

Fig. 6: Output SNR of the destination versus the number of iterations with and without noise in the reverse channel.

V. Conclusions

This paper proposes a novel nonlinear amplify-and-forward (NAF) scheme for a relay network. We model the instantaneous power limitation of the power amplifier (PA) as a soft clipping function. By regarding a relay network as a four-layer quasi-neural network, we propose a back-propagation inspired algorithm for the NAF relay scheme (NAF-BP) to optimize the weights of the destination, the relays, and the source. The nodes in the relay network coordinate via the forward propagation of the signal and the backpropagation of the derivatives, assuming no channel state information (CSI) nor data exchange between the relays. The extensive simulations verify the effectiveness of the proposed scheme in both with interference and interference-free environments. While the PA is modeled as a soft envelope limiter (SEL) in this paper, other more sophisticated and higher-fidelity PA models can be adopted with no fundamental difficulties.

References

[1] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, no. 3, pp. 1206–1232, 2015.

[2] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the metis project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[3] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[4] J. N. Laneman and G. W. Wornell, "Cooperative diversity in wireless networks: algorithms and architectures," *Dissertation Massachusetts Institute of Technology*, 2002.

[5] G. Zheng, K.-K. Wong, A. Paulraj, and B. Ottersten, "Collaborative-relay beamforming with perfect CSI: Optimum and distributed implementation," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 257–260, 2009.

[6] J. Li, A. P. Petropulu, and H. V. Poor, "Cooperative transmission for relay networks based on second-order statistics of channel state information," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1280–1291, 2011.

[7] M. A. M. Sadr, M. A. Attari, and R. Amiri, "Robust relay beamforming against jamming attack," *IEEE Communications Letters*, vol. 22, no. 2, pp. 312–315, 2018.

[8] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4306–4316, 2008.

[9] D. Simmons, D. Halls, and J. P. Coon, "OFDM-based nonlinear fixed-gain amplify-and-forward relay systems: SER optimization and experimental testing," in *2014 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2014.

[10] D. E. Simmons and J. P. Coon, "Two-way OFDM-based nonlinear amplify-and-forward relay systems," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 3808–3812, May 2016.

[11] J. Guerreiro, R. Dinis, and P. Montezuma, "Analytical evaluation of nonlinear amplify-and-forward relay systems for OFDM signals," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, pp. 1–5, Sep. 2014.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[13] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2499–2517, 2009.

[14] R. Wang and Y. Jiang, "An interference-resilient relay beamforming scheme inspired by back-propagation algorithm," in *2020 Information Theory and Applications Workshop (ITA)*, pp. 1–6, 2020.

[15] S. A. Ayoughi and W. Yu, "Interference mitigation via relaying," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 1137–1152, 2019.

[16] R. Yang and Y. Jiang, "Consensus-based clock synchronization for wide area networks," in *IEEE Wireless Communications and Networking Conference.*, pp. 1–6, 2021.

[17] Dongning Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[18] H. E. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *Bell System Technical Journal*, vol. 61, no. 7, pp. 1519–1526, 1982.

[19] E. Perahia and R. Stacey, *Next Generation Wireless LANs: 802.11n and 802.11ac, 2nd edition*. Cambridge University Press, 2013, 2013.