On Achievable Rate and Ergodic Capacity of NAF Multi-Relay Networks with CSI

Tuyen X. Tran, *Student Member, IEEE*, Nghi H. Tran, *Member, IEEE*, Hamid Reza Bahrami, *Member, IEEE*, and Shivakumar Sastry, *Senior Member, IEEE*

***Abstract*—This paper investigates the achievable rate and ergodic capacity of a non-orthogonal amplify-and-forward (NAF) half-duplex multi-relay network where multiple relays exploit channel state information (CSI) to cooperate with a pair of source and destination. In the first step, for a given input covariance matrix at the source, we derive an optimal power allocation scheme among the relays via optimal instantaneous power amplification coefficients to maximize the achievable rate. Given the nature of broadcasting and receiving collisions in NAF, the considered problem in this step is non-convex. To overcome this drawback, we propose a novel method by evaluating the achievable rate in different sub-domains of the vector channels. It is then demonstrated that the globally optimal solution can be derived in closed-form. In the next step, we establish the ergodic channel capacity by jointly optimizing the input covariance matrix at the source and the power allocation among the relays. We show that this is a bi-level non-convex problem and solve it using Tammer decomposition method. This approach allows us to transform the original optimization problem into an equivalent master problem and a set of sub-problems having closed-form solutions derived in the first step. The channel capacity is then obtained using an iterative water-filling-based algorithm. Finally, we analyze the capacity-achieving input covariance matrix at the source in high and low signal-to-noise ratio (SNR) regimes. At sufficiently high SNRs, it is shown that the transmit power at the source should be equally distributed in all broadcasting and cooperative phases. On the other hand, in low SNR regions, the source should spend all its power in the broadcasting phase associated with a relay having the strongest cascaded source-relay and relay-destination channels.**

***Index Terms*—Achievable rate, channel state information, distributed water-filling, ergodic capacity, multiple relays, non-orthogonal amplify-and-forward, power adaptation, relay channel.**

I. Introduction

DUE to channel impairments such as fading, interference, and noise, the wireless channels fluctuate randomly over time. This makes it challenging to design a wireless communications system with guaranteed performance. One effective way to enhance the reliability and data rate of wireless systems is is to employ intermediate network nodes that assist each other by relaying transmissions [2]–[5]. In general, relaying strategies can be categorized as decode-and-forward (DF) [6], [7], compress-and-forward (CF) [8], [9], and amplify-and-forward (AF) [5], [10]–[14]. While DF and CF relaying usually require high decoding complexity, the AF scheme is very simple to implement, since the relay only needs to transmit a scaled version of the signal. As such, AF relaying is of practical interest and in fact, it has been adopted in major wireless standards [15].

While early research on AF relaying has mainly devoted to dual-hop and multi-hop AF relaying systems where the direct link between the source and destination is not taken into account (see for example [16]–[21] and references within), there has been a growing interest in cooperative schemes that make use of the direct link. It is because by considering the source to destination link, cooperative AF networks can further exploit the degrees of freedom of the channel and provide rate as well as diversity advantages [3], [5], [10]–[14]. For such schemes, the intermediate nodes (relays) receive signal from the source during the broadcasting phases and then transmit the noisy amplified version of that signal to the destination during the cooperative phases. Different AF relaying protocols with direct link can be classified into the orthogonal AF (OAF) protocol and the non-orthogonal AF (NAF) protocol [3], [4]. In OAF, the source and relays use separate time slots or orthogonal channels to transmit signals, i.e., the source only transmits in the broadcasting phase and remains silent in the cooperative phase. On the other hand, in NAF, the source-destination link is exploited in both broadcasting and cooperative phases which maximizes the degrees of broadcasting and receiving collisions. Specifically, in the cooperative phase, the source attempts to transmit a new symbol when the relays forward the received symbols in the previous phase. In fact, the NAF protocol is considered to be favorable over other AF schemes, not only from an information-theoretic point of view [10]–[12], [14], but also in terms of diversity performance [22], [23]. However, due to the complexity of broadcasting and receiving collisions, the analysis and optimization of non-orthogonal networks are more challenging and have received less attention than the OAF counterpart [10], [14].

In AF relaying, depending on the availability of the channel state information (CSI), a relay simply amplifies the noisy version of the received signal using a suitable amplification coefficient. For example, by assuming that a relay can only obtain the channel distribution information (CDI) of the source-relay channel, the received signal at the relay is scaled by

Manuscript received July 31, 2013; revised December 6, 2014. The editor coordinating the review of this paper and approving it for publication is D. Gunduz.

The authors are with the Department of Electrical & Computer Engineering, the University of Akron, Akron, OH, USA (e-mail: txt2@zips.uakron.edu, {nghi.tran, hrb, ssastry}@uakron.edu).

This work was supported in part by The University of Akron under a Start-up Grant and a Firestone Fellowship Award. A part of this work was presented at the IEEE International Conference on Communications (ICC), Budapest, Hungary, June 2013 [1].

Digital Object Identifier 10.1109/TCOMM.2014.032314.130591

a constant-gain coefficient before being transmitted to the destination [3], [14], [22]. When the relay has an instantaneous knowledge of the source-relay channel, the channel inversion (CI) technique can be used to amplify the received signal at a desired power level [4], [10]. Recently, motivated by the benefit offered by the cooperation between the relay and destination, references [24] and [25] have examined NAF and OAF single-relay systems in which the relay acquires complete CSI of the source-destination and source-relay links via a feedback channel from the destination. By assuming a fixed covariance matrix at the source, it was shown in [24] and [25] that with CSI, the achievable rate can be significantly improved over what could be achieved using CDI and CI. The idea has been recently extended to AF multi-relay networks in [26] for further improvements, but only for the OAF protocol.

Extending the idea of using CSI at the relays with a suitable power allocation at the relays and the source to enhance the performance of an NAF multi-relay system presents many challenges. First, different from a single-relay system, we need to take into account power budget constrained on the nodes globally and locally. Second, the broadcasting and receiving nature of NAF makes it difficult to analyze and optimize the fundamental limits. For example, for a fixed input covariance matrix, obtaining an optimal power sharing scheme among the relays and the corresponding achievable rate for the NAF networks in closed-form are already challenging. Finding a capacity-achieving strategy is even more difficult, since one needs to simultaneously optimize the input covariance matrix at the source and the power sharing strategy among the relays.

Motivated by the above discussions, this paper addresses the achievable rate and ergodic channel capacity for NAF multi-relay systems with full CSI available at the relays. We adopt the multi-relay protocol originally proposed in [10], [27] where the relays take turns to transmit. This multi-relay protocol has been widely considered in the literature [10], [14], [27], [28]. We shall consider both a *Total Average Power Constraint* (TAPC) and an *Individual Average Power Constraint* (IAPC) imposed on the relay nodes. First, for a given input covariance matrix, we identify the optimal power sharing scheme among the relays by means of the optimal instantaneous power amplification coefficients to maximize the achievable rate. Due to the nature of the NAF protocol, the considered problem is not convex. Our method is to modify the problem by evaluating the achievable rate in different sub-domains of the vector channels. It is then shown that the obtained closed-form solution is indeed the globally optimal solution of the original problem, which can be understood as an extended water-filling scheme with the vessels having both the bottoms and lids. Furthermore, we investigate the capacity limit of the considered NAF system by jointly examining the optimal power allocation scheme at the relays and the optimal input covariance matrix at the source. The optimization problem in this step is a bi-level non-convex problem for which it is not feasible to derive a closed-form solution. As an alternative, we effectively utilize the Tammer decomposition method [29] to transform the original problem to an equivalent master problem and a set of sub-problems. It is then demonstrated that the sub-problems have the same closed-form solution as obtained in the first step. The ergodic capacity can then be obtained via an iterative water-filling-based algorithm. Finally, we analyze the capacity-achieving input covariance matrix at the source in high and low signal-to-noise ratio (SNR) regimes. Specifically, we show that at sufficiently high SNRs, the transmit power at the source should be equally distributed in all broadcasting phases to achieve the capacity. On the other hand, in low SNR regimes, it is demonstrated that the source spends all its power in the broadcasting phase associated with a relay having the strongest cascaded source-relay and relay-destination channels.

The rest of this paper is organized as follows. Section II introduces the considered NAF system. Optimal power sharing schemes at the relays for a given input covariance matrix are derived in Section III. In Section IV, we address the optimal power sharing at the relays and the optimal input covariance matrix at the source simultaneously to achieve the channel capacity. Section V analyzes the capacity-achieving input covariance matrix at the source and the corresponding capacity at high and low SNRs. Numerical results are provided in Section VI to illustrate the gains of the proposed power adaptation schemes. Finally, Section VII concludes the paper.

II. System Model

A. NAF Multi-Relay Protocol

The half-duplex NAF multi-relay system shown in Fig. 1(a) consists of N relay nodes, R_1,...R_N, that assist the transmissions between a source (S) and a destination (D) via multiple cooperation frames as proposed in [10], [27]. The instantaneous complex channel gains of the S-D, S-R_n and R_n-D links are denoted, respectively, as h_{SD}, h_{SR_n} and h_{RD_n}. Signals are transmitted from the source to the destination via a sequence of cooperative frames, each consisting of $2N$ time-slots as illustrated in Fig. 1(b). In each cooperative frame, a relay will only be active in two consecutive time-slots to assist the transmission from S to D while other relays remain silent or participate in other transmissions. While the adopted protocol might not be the best multi-relay strategy, it has been demonstrated in [10], [28] that this protocol provides superior performance over other AF protocols from the diversity-multiplexing tradeoff (DMT) perspective and has the potential to achieve full diversity. Also, by using this protocol, interference among the relays can be avoided within the considered systems. Furthermore, in practice, a relay node can engage in other transmissions when its relaying function is inactive [10], [27]. Such benefits make this multi-relay protocol fundamentally different from other multi-relay protocols that allow the relays to transmit simultaneously [23], [30].

For the considered protocol, each cooperative frame is divided into two N-time-slot phases, the broadcasting phases and the cooperative phases. Each broadcasting phase is carried out in the $(2n-1)$th time-slots, $1 \leq n \leq N$. For a given $(2n-1)$th time slot, only the relay R_n stays active and the source S sends the signal $x_{n,1}$ to both D and R_n. The received signals at R_n and D can be written, respectively, as $r_n = \sqrt{E_s} h_{SR_n} x_{n,1} + w_n$, and $y_{2n-1} = \sqrt{E_s} h_{SD} x_{n,1} + v_{n,1}$, where E_S is a constant related to the transmitted power; $w_n, v_{n,1}$ are the i.i.d zero-mean circularly symmetric Gaussian noises with variance N_0, denoted as $\mathcal{CN}(0, N_0)$.

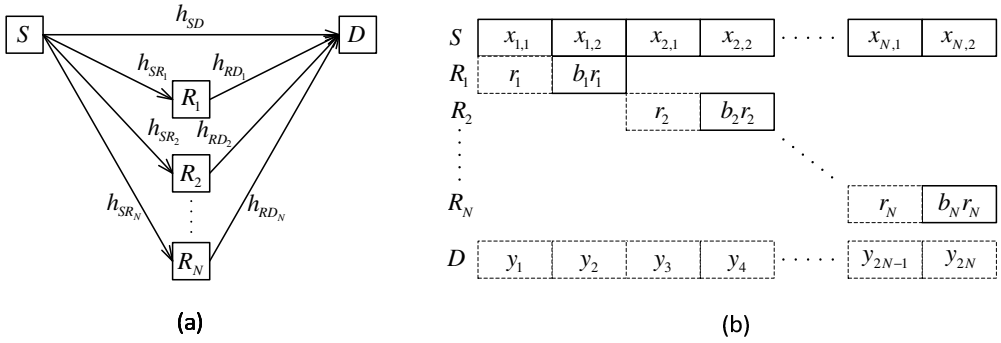

Fig. 1. (a) The half-duplex NAF multi-relay system. (b) The frame structure of the NAF multi-relay protocol; the solid boxes denote the transmitted signals and the dashed boxes denote the received signals.

The cooperative phase happens in the $2n$th time-slots, $1 \leq n \leq N$ [3]. In particular, the source S transmits a new signal $x_{n,2}$ to D, while R_n forwards the amplified version of r_n to D. The received signal at D in the $2n$th time-slot is given by

$$y_{2n} = \sqrt{E_s} h_{SD} x_{n,2} + h_{RD_n} b_n (\sqrt{E_s} h_{SR_n} x_{n,1} + w_n) + v_{n,2},$$

where $v_{n,2} \sim \mathcal{CN}(0, N_0)$ and b_n is the amplification coefficient at the relay R_n. Whitening the noise component, the NAF system can be written as $\mathbf{y} = \sqrt{E_S}\mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{x} = [x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}, ..., x_{N,1}, x_{N,2}]^T$, $\mathbf{y} = [y_1, y_2, ..., y_{2N}]^T$, the equivalent noise $\mathbf{n} \sim \mathcal{CN}(0, N_0)$ and $\mathbf{H}$ is a $2N \times 2N$ block diagonal channel matrix given by $\mathbf{H} = diag(\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_N)$, with $\mathbf{H}_n = \begin{bmatrix} h_{SD} & 0 \\ \alpha_n b_n h_{RD_n} h_{SR_n} & \alpha_n h_{SD} \end{bmatrix}$, and $\alpha_n = \frac{1}{\sqrt{1+b_n^2 |h_{RD_n}|^2}}$ is the noise whitening factor.

B. CSI and Power Constraints

The channels between different nodes in the system are assumed to be constant during at least one cooperative frame, and their magnitudes are independently Rayleigh distributed, i.e., $h_{SD} \sim \mathcal{CN}(0, \phi_{SD})$, $h_{SR_n} \sim \mathcal{CN}(0, \phi_{SR_n})$ and $h_{RD_n} \sim \mathcal{CN}(0, \phi_{RD_n})$, where ϕ_{SD}, ϕ_{SR_n} and ϕ_{RD_n} are the channel variances. Following references [5], [10]–[14], we also assume that D has perfect knowledge of the gains of all links. In addition, as recently considered in [24], such complete CSI can be made available at each relay using a feedback channel from the destination.

Suppose that the source S uses Gaussian input signal and has no CSI knowledge; it has been shown in [14] that the optimal input covariance matrix is a diagonal matrix given as $\mathbf{Q} = diag(\mathbf{Q}_1, \mathbf{Q}_2, ...\mathbf{Q}_n)$, where $\mathbf{Q}_n = \begin{bmatrix} q_{n1} & 0 \\ 0 & q_{n2} \end{bmatrix}$, and $q_{n1} = \mathbb{E}\left[|x_{n,1}|^2\right]$, $q_{n2} = \mathbb{E}\left[|x_{n,2}|^2\right]$. Hence, $q_{n1}E_S$ and $q_{n2}E_S$, $1 \leq n \leq N$, can be interpreted as the average transmit power of the source in the broadcasting phase and the cooperative phase, respectively. The average transmit power constraint at the source can then be expressed as

$$tr(\mathbf{Q}) = \sum_{n=1}^{N} tr(\mathbf{Q}_n) = \sum_{n=1}^{N} (q_{n1} + q_{n2}) \leq P_S. \quad (1)$$

Furthermore, let

$$\mathbf{h} = [h_{SD}, h_{SR_1}, h_{SR_2}, ..., h_{SR_N}, h_{RD_1}, h_{RD_2}, ..., h_{RD_N}]^T$$

be the vector channels and $\mu_n E_S$ be the *instantaneous* power allocated to the relay R_n in one cooperative frame. Using the full CSI knowledge, the relays can adjust their instantaneous transmit power in order to adapt to the current channel conditions. Thus, the instantaneous value of μ_n depends on $\mathbf{h}$ and without loss of generality, we can write $\mu_n = \mu_n(\mathbf{h})$. Such a variation is reflected in the use of an instantaneous power amplification coefficient b_n at each relay R_n, which is given as

$$b_n = \sqrt{\frac{\mu_n(\mathbf{h})\rho}{|h_{SR_n}|^2 q_{n1}\rho + 1}}, \quad (2)$$

with $\rho = E_S/N_0$ being the normalized SNR.

We consider two types of power constraint, the TAPC on all the relays and the IAPC on each relay, which can be written, respectively, as

$$\text{TAPC}: \quad \sum_{n=1}^{N} \int_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_R, \quad (3)$$

$$\text{IAPC}: \quad \int_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_n, n = 1, ...N, \quad (4)$$

where $f(\mathbf{h})$ is the probability density function of the channel state $\mathbf{h}$, P_R is the total average power available for all relays, and P_n is the average power limit on each relay R_n, $n = 1, ...N$. The assumption of total power constraint is feasible, due to the fact that the relays and the destination can be managed directly by the service provider. As such, the relays and the destination can cooperate to share the total power constraint. In addition, a network with such a constraint can serve as a system benchmark to any system with only individual power constraints due to the constraint on each individual RF chain.

Finally, it should be noted that when the relays only have the knowledge of the CDI of the S-R_n link, the amplification coefficient at relay R_n is [3], [14], [22]:

$$b_n^{(CDI)} = \sqrt{\frac{\rho P_R}{N\left(\mathbb{E}[|h_{SR_n}|^2] q_{n1}\rho + 1\right)}}, \quad (5)$$

On the other hand, when only the CSI of the S-R_n link is available at the relays, each relay R_n can use the channel inversion (CI) scheme with the following amplification coefficient [4], [10]:

$$b_n^{(CI)} = \sqrt{\frac{\rho P_R}{N\left(|h_{SR_n}|^2 q_{n1}\rho + 1\right)}}. \tag{6}$$

C. Achievable Rate and Ergodic Channel Capacity

Given the CSI knowledge at the destination, the source-destination instantaneous mutual information can be expressed as

$$I(\mathbf{x},\mathbf{y}|\mathbf{h}) \quad = \quad \log\det[\mathbf{I}_{2N} + \rho\mathbf{H}^{\dagger}\mathbf{H}\mathbf{Q}], \tag{7}$$

where $\dagger$ denotes the Hermitian operator and $log(.)$ is the base-2 logarithm. Thus, for a given input covariance matrix $\mathbf{Q}$, the maximum achievable rate (in bits per channel use) of the NAF system can be defined as

$$R\left(\mathbf{Q}\right) = \frac{1}{2}\max_{\mu_n(\mathbf{h}),n=\overline{1,N}} \int_{\mathbf{h}} I\left(\mathbf{x},\mathbf{y}|\mathbf{h}\right) f\left(\mathbf{h}\right) d\mathbf{h}. \tag{8}$$

Note that the normalization factor of $1/2$ is used in (8) to account for the fact that each transmission from the source to destination associated with an active relay requires two channel uses.

By further optimizing the input covariance matrix at the source and the power allocation at the relays simultaneously, one can achieve the ergodic capacity of the system, which is expressed as

$$C = \frac{1}{2}\max_{\substack{\mathbf{Q}\geq 0, tr(\mathbf{Q})\leq P_S,\\ \mu_n(\mathbf{h}),n=\overline{1,N}}} \int_{\mathbf{h}} I\left(\mathbf{x},\mathbf{y}|\mathbf{h}\right) f\left(\mathbf{h}\right) d\mathbf{h}, \tag{9}$$

In the following, we shall present the solutions to the achievable rate $R\left(\mathbf{Q}\right)$ in (8), and the ergodic capacity C in (9). In particular, by first considering a fixed input covariance matrix $\mathbf{Q}$, we derive the closed-form optimal power allocation solutions at the relays to maximize the achievable rate in (8). These closed-form solutions will be used in Section IV to obtain the ergodic capacity.

III. Achievable Rate and Optimal Power Allocation at the Relays

In this section, we shall develop the optimal power sharing schemes among the relays to maximize the achievable rate $R(\mathbf{Q})$. For the sake of convenience, let $x_n = |h_{RD_n}|^2$, $y = |h_{SD}|^2$, $z_n = |h_{SR_n}|^2$. Using (7) we obtain

$$I(\mathbf{x},\mathbf{y}|\mathbf{h}) = \sum_{n=1}^{N} \log\det\left[\mathbf{I}_2 + \rho\mathbf{H}_n^{\dagger}\mathbf{H}_n\mathbf{Q}_n\right] = \sum_{n=1}^{N}\log\mathbf{\Gamma}_n, \tag{10}$$

where

$$\mathbf{\Gamma}_n = \frac{\rho^2 y^2 q_{n1}q_{n2}}{1+b_n^2 x_n} + \rho\left(q_{n1}y + \frac{q_{n1}b_n^2 x_n z_n + q_{n2}y}{1+b_n^2 x_n}\right) + 1. \tag{11}$$

From (2) and (11), we get $\mathbf{\Gamma}_n = \mathbf{\Gamma}_n[\mu_n(\mathbf{h})] = \frac{\Gamma_{n1}[\mu_n(\mathbf{h})]}{\Gamma_{n2}[\mu_n(\mathbf{h})]}$, with

$$\begin{aligned}\Gamma_{n1}[\mu_n(\mathbf{h})] &= \mu_n(\mathbf{h})\left(\rho q_{n1}y + \rho q_{n1}z_n + 1\right)\rho x_n \\ &+(\rho^3 q_{n1}^2 q_{n2} z_n y^2 + \rho^2 q_{n1}q_{n2}y^2 + \rho^2 q_{n1}^2 z_n y + \rho^2 q_{n1}q_{n2}z_n y \\ &+\rho q_{n1}z_n + \rho q_{n1}y + \rho q_{n2}y + 1),\end{aligned} \tag{12}$$

and $$\Gamma_{n2}[\mu_n(\mathbf{h})] = \rho\mu_n(\mathbf{h})x_n + \left(\rho q_{n1}z_n + 1\right). \tag{13}$$

Under both TAPC and IAPC, the set of optimal instantaneous power allocated at the relays $\{\mu_n^*(\mathbf{h})\}$ to maximize the achievable rate $R(\mathbf{Q})$ can be obtained by solving the following optimization problem

$$\begin{aligned}&\max_{\substack{\mu_n(\mathbf{h})\geq 0\\ n=\overline{1,N}}} \int_{\mathbf{h}} I\left(\mathbf{x},\mathbf{y}|\mathbf{h}\right) f\left(\mathbf{h}\right) d\mathbf{h} \\ &s.t.\begin{cases}\sum\limits_{n=1}^{N}\int\limits_{\mathbf{h}}\mu_n\left(\mathbf{h}\right) f\left(\mathbf{h}\right) d\mathbf{h} \leq P_R, \\ \int\limits_{\mathbf{h}}\mu_n\left(\mathbf{h}\right) f\left(\mathbf{h}\right) d\mathbf{h} \leq P_n.\end{cases}\end{aligned} \tag{14}$$

Let $I(\mathbf{x},\mathbf{y}|\mathbf{h}) = I(\boldsymbol{\mu}(\mathbf{h}))$ where $\boldsymbol{\mu}(\mathbf{h}) = [\mu_1(\mathbf{h}),\ldots,\mu_{\mathbf{N}}(\mathbf{h})]$. Differentiating the objective function in (14) with respect to $\mu_n(\mathbf{h})$ we have

$$\begin{aligned}&\frac{\partial}{\partial\mu_n(\mathbf{h})}\int_{\mathbf{h}} I\left(\mu(\mathbf{h})\right) f\left(\mathbf{h}\right) d\mathbf{h} \\ &= \frac{\partial}{\partial\mu_n(\mathbf{h})}\int_{\mathbf{h}}\left(\sum_{n=1}^{N}\log\left(\mathbf{\Gamma}_n[\mu_n(\mathbf{h})]\right)\right) f\left(\mathbf{h}\right) d\mathbf{h} \\ &= \frac{\partial}{\partial\mu_n(\mathbf{h})}\int_{\mathbf{h}}\log\left(\mathbf{\Gamma}_n[\mu_n(\mathbf{h})]\right) f\left(\mathbf{h}\right) d\mathbf{h}.\end{aligned} \tag{15}$$

Since (15) is a functional derivative of the inner product of $\log\left(\mathbf{\Gamma}_n[\mu_n(\mathbf{h})]\right)$ and $f\left(\mathbf{h}\right)$, it follows from [31] that:

$$\frac{\partial}{\partial\mu_n(\mathbf{h})}\int_{\mathbf{h}} I\left(\mu(\mathbf{h})\right) f\left(\mathbf{h}\right) d\mathbf{h} = \frac{\partial\left(\log\mathbf{\Gamma}_n[\mu_n(\mathbf{h})]\right)}{\partial\mu_n(\mathbf{h})} f\left(\mathbf{h}\right) \tag{16}$$

$$= \frac{-A_n}{\ln(2)[C_n\mu_n^2(\mathbf{h}) + D_n\mu_n(\mathbf{h}) + E_n]} f\left(\mathbf{h}\right), \tag{17}$$

where

$$\begin{aligned}A_n &= \rho^2 x_n(\rho q_{n1}z_n + 1)(\rho q_{n1}q_{n2}y^2 + q_{n2}y - q_{n1}z_n), \\ C_n &= \rho^2 x_n^2(\rho q_{n1}y + \rho q_{n1}z_n + 1), \\ D_n &= \rho x_n(\rho q_{n1}z_n + 1)(2\rho q_{n1}y + \rho q_{n2}y + \rho q_{n1}z_n + \\ &\rho^2 q_{n1}q_{n2}y^2 + 2), \\ E_n &= (\rho q_{n1}y + 1)(\rho q_{n2}y + 1)(\rho q_{n1}z_n + 1)^2.\end{aligned} \tag{18}$$

Similarly, the second derivatives can be calculated as:

$$\begin{aligned}&\tfrac{\partial^2}{\partial\mu_n^2(\mathbf{h})}\int_{\mathbf{h}} I\left(\mu(\mathbf{h})\right) f\left(\mathbf{h}\right) d\mathbf{h} \\ &\qquad = \tfrac{A_n[2C_n\mu_n(\mathbf{h})+D_n]}{\ln(2)[C_n\mu_n^2(\mathbf{h})+D_n\mu_n(\mathbf{h})+E_n]^2} f\left(\mathbf{h}\right), \\ &\text{and} \quad \tfrac{\partial^2}{\partial\mu_m(\mathbf{h})\partial\mu_n(\mathbf{h})}\int_{\mathbf{h}} I\left(\mu(\mathbf{h})\right) f\left(\mathbf{h}\right) d\mathbf{h} = 0\,, \quad (m\neq n).\end{aligned} \tag{19}$$

It can be verified that while C_n, D_n and E_n are always positive, A_n might be positive, negative or zero depending on the SNR, power allocation at S, and the instantaneous channel gains. Hence, the problem in (14) is not concave.

As a consequence, the Karush-Kuhn-Tucker (KKT) conditions cannot be applied directly. To overcome this drawback, in the following, we propose a novel approach by evaluating the achievable rate in different sub-domains of the vector channels and show that the globally optimal solution can be obtained. To this end, let define the feasible set for the problem in (14) as

$$\mathcal{F} = \{\boldsymbol{\mu}(\mathbf{h}) \in \mathbb{R}^N | \mu_n(\mathbf{h}) \geq 0, \sum_{n=1}^{N} \int_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_R\}.$$

Furthermore, let $\gamma_n = \rho q_{n1} q_{n2} y^2 + q_{n2} y - q_{n1} z_n$. For each relay R_n, we divide the domain of $\mathbf{h}$ into two disjoint regions, $\mathbf{h}_\mathbf{P}^{(n)} = \{\mathbf{h} | \gamma_n \geq 0\}$ and $\mathbf{h}_\mathbf{N}^{(n)} = \{\mathbf{h} | \gamma_n < 0\}$. In the domain $\mathbf{h}_\mathbf{P}^{(n)}$ we have $A_n \geq 0$, making the first derivative in (17) non-positive, thus, $I(\boldsymbol{\mu}(\mathbf{h}))$ is a decreasing function of $\mu_n(\mathbf{h})$. Furthermore, let $\boldsymbol{\mu}^{(n)}(\mathbf{h})$ be a vector obtained by replacing the element $\mu_n(\mathbf{h})$ in $\boldsymbol{\mu}(\mathbf{h})$ by zero. It is straightforward to see that

$$I(\boldsymbol{\mu}(\mathbf{h})) \leq I(\boldsymbol{\mu}^{(n)}(\mathbf{h})), \ \ \forall \mathbf{h} \in \mathbf{h}_\mathbf{P}^{(n)}.$$

In addition, for any $\boldsymbol{\mu} \in \mathcal{F}$ and $n = 1, 2, ..., N$, we have

$$\begin{aligned} &\int_{\mathbf{h}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} \\ &= \int_{\mathbf{h}_\mathbf{N}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} + \int_{\mathbf{h}_\mathbf{P}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} \\ &\leq \int_{\mathbf{h}_\mathbf{N}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} + \int_{\mathbf{h}_\mathbf{P}^{(n)}} I\left(\boldsymbol{\mu}^{(n)}(\mathbf{h})\right) f(\mathbf{h})\, d\mathbf{h}. \end{aligned} \tag{20}$$

Hence, the objective function in (14) can then be upper-bounded as

$$\begin{aligned} &\int_{\mathbf{h}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} \\ &\leq \tfrac{1}{N} \sum_{n=1}^{N} \left(\int_{\mathbf{h}_\mathbf{N}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} + \int_{\mathbf{h}_\mathbf{P}^{(n)}} I\left(\boldsymbol{\mu}^{(n)}(\mathbf{h})\right) f(\mathbf{h})\, d\mathbf{h} \right), \end{aligned} \tag{21}$$

and one achieves the equality when $\mu_n(\mathbf{h}) = 0 \forall \mathbf{h} \in \mathbf{h}_\mathbf{P}^{(n)}, n = 1, ..., N$. Using (21), we then have the following modified optimization problem:

$$\begin{aligned} \max_{\substack{\mu_n(\mathbf{h}) \geq 0 \\ n = \overline{1,N}}} \ & \tfrac{1}{N} \sum_{n=1}^{N} \left(\int_{\mathbf{h}_\mathbf{N}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} \right. \\ & \left. + \int_{\mathbf{h}_\mathbf{P}^{(n)}} I\left(\boldsymbol{\mu}^{(n)}(\mathbf{h})\right) f(\mathbf{h})\, d\mathbf{h} \right) \\ s.t. \ & \sum_{n=1}^{N} \int_{\mathbf{h}_\mathbf{N}^{(n)}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_R. \end{aligned} \tag{22}$$

Given that the second term of the objective function in (22) does not depend on $\mu_n(\mathbf{h})$ and the fact that $A_n < 0$ in the sub-domain $\mathbf{h}_\mathbf{N}^{(n)}$, the Hessian matrix of the objective function in (22) is diagonal with strictly negative elements. Thus, (22) is now a concave optimization problem and any solution in the form of $\mu_n(\mathbf{h}) = \mu_\mathbf{n}^*(\mathbf{h})$ is *globally optimal*. The problem in (22) can then be solved using the KKT conditions. Specifically, the Lagrangian can be written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\mu}(\mathbf{h}), \lambda) = & -\tfrac{1}{N} \sum_{n=1}^{N} \left(\int_{\mathbf{h}_\mathbf{N}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h} \right. \\ & \left. + \int_{\mathbf{h}_\mathbf{P}^{(n)}} I\left(\boldsymbol{\mu}^{(n)}(\mathbf{h})\right) f(\mathbf{h})\, d\mathbf{h} \right) \\ & -\lambda \left(P_R - \sum_{n=1}^{N} \int_{\mathbf{h}_\mathbf{N}^{(n)}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \right) \\ & - \sum_{n=1}^{N} \lambda_n \left(P_n - \int_{\mathbf{h}_\mathbf{N}^{(n)}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \right). \end{aligned} \tag{23}$$

Differentiating the Lagrangian in (23) we obtain

$$\begin{aligned} &\frac{\partial \mathcal{L}(\boldsymbol{\mu}(\mathbf{h}), \lambda)}{\partial \mu_n(\mathbf{h})} = \\ &\begin{cases} 0, & \mathbf{h} \in \mathbf{h}_\mathbf{P}^{(n)} \\ (\lambda + \lambda_n) f(\mathbf{h}) - \frac{1}{N} \frac{\partial}{\partial \mu_n(\mathbf{h})} \int_{\mathbf{h}_\mathbf{N}^{(n)}} I(\boldsymbol{\mu}(\mathbf{h})) f(\mathbf{h})\, d\mathbf{h}, & \mathbf{h} \in \mathbf{h}_\mathbf{N}^{(n)}, \end{cases} \end{aligned} \tag{24}$$

By equating the gradient of the above Lagrangian to zero and solving for $\mu_n(\mathbf{h})$, the globally optimal instantaneous power allocation $\mu_n^*(\mathbf{h})$ for the NAF system under both TAPC and IAPC is finally expressed as

$$\begin{aligned} &\mu_n^*(\mathbf{h}) = \\ &\begin{cases} 0, & A_n \geq 0 \\ \left(\frac{-D_n + \sqrt{D_n^2 - 4C_n \left(E_n + \frac{A_n}{\ln(2) N (\lambda^* + \lambda_n^*)} \right)}}{2C_n} \right)^+, & A_n < 0 \end{cases} \end{aligned} \tag{25}$$

where we have ignored the other root since it violates the non-negativity of $\mu_n(\mathbf{h})$. In (25), we have $D_n^2 - 4C_n \left(E_n + \frac{A_n}{\ln(2) N (\lambda^* + \lambda_n^*)} \right) > 0$. Therefore, the values of $\mu_n^*(\mathbf{h})$ are always real. In addition, $\lambda^* \geq 0$ and $\lambda_n^* \geq 0$ are the constants satisfying

$$\begin{cases} \sum_{n=1}^{N} \int_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_R, \\ \lambda^* \left(P_R - \sum_{n=1}^{N} \int_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \right) = 0, \\ \int_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_n, \\ \lambda_n^* \left(P_n - \int_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \right) = 0, \\ \lambda^* + \lambda_n^* > 0. \end{cases} \tag{26}$$

The two parameters λ^* and λ_n^* depend on the channel distribution $f(\mathbf{h})$, the SNR ρ, and the average power constraints P_R and $\{P_n\}$, and they can be found numerically using bisection. The solution in (25) is an extended water-filling based scheme over time and space with the vessels having both bottoms and lids. It can also be observed that the average power allocated to a particular relay not only depends on its related links but also depends on all the other links involved in the system. Hereafter, we shall refer to these schemes as OPA-R schemes.

From (14), it is easy to see that as long as $\sum_{n=1}^{N} P_n < P_R$, TAPC is always satisfied and one needs to consider IAPC only. If it is the case, the conditions on λ^* and λ_n^* can be simplified to:

$$\begin{cases} \lambda^* = 0, \\ \int\limits_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \le P_n, \\ \lambda_n^* \left(P_n - \int\limits_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \right) = 0. \end{cases} \tag{27}$$

On the other hand, when $\sum_{n=1}^{N} P_n \ge P_R$, one needs to take into account TAPC and IAPC simultaneously, which is certainly of more interest.

Finally, it is noted that if only TAPC is assumed, the optimal solution $\mu_n^*(\mathbf{h})$ is simplified to

$$\mu_{n,TAPC}^*(\mathbf{h}) = \begin{cases} 0, & A_n \ge 0, \\ \left(\frac{-D_n + \sqrt{D_n^2 - 4C_n\left(E_n + \frac{A_n}{\ln(2) N \lambda^*}\right)}}{2C_n} \right)^+, & A_n < 0, \end{cases} \tag{28}$$

where $\lambda^* > 0$ is a constant that satisfies

$$\sum_{n=1}^{N} \int\limits_{\mathbf{h}} \mu_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} = P_R.$$

IV. Channel Capacity

In the preceding section, given the diagonal input covariance matrix $\mathbf{Q}$, we derived the maximum achievable rate $R(\mathbf{Q})$ of the NAF systems by optimizing the instantaneous power allocation at the relays. In this section, we investigate the ergodic capacity of the system by simultaneously optimizing the diagonal input covariance matrix at the source and the instantaneous power allocations at the relays. Unless otherwise stated, we will focus on the case of using both the TAPC and IAPC where $\sum_{n=1}^{N} P_n \ge P_R$, which implies that TAPC and IAPC need to be taken into account simultaneously.

The ergodic capacity of the NAF system can be expressed as

$$C = \frac{1}{2} \max_{\substack{\mathbf{Q} \ge 0, tr(\mathbf{Q}) \le P_S, \\ \mu_n(\mathbf{h}), n=\overline{1,N}}} \int\limits_{\mathbf{h}} I(\mathbf{x}, \mathbf{y}|\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \quad s.t. \quad \begin{cases} \int\limits_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \le P_n, \\ \sum\limits_{n=1}^{N} \int\limits_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \le P_R, \\ \sum\limits_{n=1}^{N} (q_{n1} + q_{n2}) \le P_S. \end{cases} \tag{29}$$

In general, it is not possible to obtain the closed-form solution of (29) due to the complexity and non-convexity of the problem. However, we observe that this problem belongs to the class of bi-level non-convex optimization problems with separable objective functions. In particular, we can consider each $\mu_n(\mathbf{h})$ as a local variable that appears only in one block and $\mathbf{Q}$ as the global variable that appears in all the blocks. Exploiting this property, the Tammer decomposition method proposed in [29] can be used to effectively address this problem. The main idea of Tammer decomposition method is to transform an original optimization problem with high complexity into an equivalent master problem and a set of sub-problems with significantly lower complexity. In our case, if we temporarily fix the global variable $\mathbf{Q}$, the problem in (29) can be separated into multiple sub-problems whose solutions can be verified to be the same as that of the maximizing the achievable rate problem derived in the previous section. The capacity can then be obtained by solving the master problem via an iterative water-filling-based algorithm. The decomposition process is described in detail in the following.

First, for convenience, let

$$\begin{aligned} \Lambda_n(\mathbf{Q}_n, \mu_n(\mathbf{h})) &= \int\limits_{\mathbf{h}} \log\det\left[\mathbf{I}_2 + \rho \mathbf{H}_n^\dagger \mathbf{H}_n \mathbf{Q}_n\right] f(\mathbf{h})\, d\mathbf{h} \\ &= \int\limits_{\mathbf{h}} \log\left(\mathbf{\Gamma}_n[\mu_n(\mathbf{h})]\right) f(\mathbf{h})\, d\mathbf{h}. \end{aligned} \tag{30}$$

Then from (10), we can rewrite the objective function in (29) as follows

$$\int\limits_{\mathbf{h}} I(\mathbf{x}, \mathbf{y}|\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} = \sum_{n=1}^{N} \Lambda_n(\mathbf{Q}_n, \mu_n(\mathbf{h})). \tag{31}$$

Following Tammer decomposition method, solving the problem in (29) is equivalent to solving the following master problem

$$\max_{\mathbf{Q}_n, n=\overline{1,N}} \sum_{n=1}^{N} \Lambda_n^*(\mathbf{Q}_n) \;\; s.t. \sum_{n=1}^{N} tr(\mathbf{Q}_n) \le P_S, \tag{32}$$

where $\Lambda_n^*(\mathbf{Q}_n)$ is the optimal-value function corresponding to the nth sub-problem, and it can be written as

$$\Lambda_n^*(\mathbf{Q}_n) = \max_{\mu_n(\mathbf{h}) \ge 0} \Lambda_n(\mathbf{Q}_n, \mu_n(\mathbf{h})) \quad s.t. \quad \begin{cases} \int\limits_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \le P_n, \\ \sum\limits_{n=1}^{N} \int\limits_{\mathbf{h}} \mu_n(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \le P_R. \end{cases} \tag{33}$$

The sub-problems in (33) are obtained by setting $\mathbf{Q}$ to a fixed value and they are subject to a global constraint and multiple individual constraints. Therefore, instead of solving the sub-problems separately as in [29], we need to solve them simultaneously. It is straightforward to verify that each sub-problem in (33) has the same solution as the problem in (14); the solution to this problem is shown in (25). Our approach is to solve the sub-problems and then use the information obtained from their optimal solution $\{\mu_n^*(\mathbf{h})\}$ to determine the optimal value of $\mathbf{Q}$.

To solve the master problem, our method is to apply the line search technique using a penalty function M_k [32] to convert the constrained master problem in (32) into an unconstrained problem as follows

$$\max_{\mathbf{Q}_n, n=\overline{1,N}} \left[\sum_{n=1}^{N} \Lambda_n^*(\mathbf{Q}_n) - M_k \left(P_S - \sum_{n=1}^{N} tr(\mathbf{Q}_n) \right)^2 \right]. \tag{34}$$

Let

$$\mathbf{\Lambda}^*(\mathbf{Q}) = \sum_{n=1}^{N} \Lambda_n^*(\mathbf{Q}_n) - M_k \left(P_S - \sum_{n=1}^{N} tr(\mathbf{Q}_n) \right)^2, \quad (35)$$

in which the optimal-value function $\Lambda_n^*(\mathbf{Q}_n)$ in each iteration is obtained by solving the sub-problems with the closed-form solutions given in (25). All the sub-problems in each iteration are solved and the solutions are used to determine the new search direction. The value of $\mathbf{Q}$ will also be adjusted accordingly. Note that the sub-problem in (33) is feasible for all values of $\mathbf{Q}_n$ satisfying $\mathbf{Q}_n \geq 0$ and $\sum_{n=1}^{N} tr(\mathbf{Q}_n) \leq P_S$. The iterative algorithm, which is referred to as master solver (MASSOL), to solve the master problem can then be summarized as follows:

MASSOL Algorithm

(1) **(Initialization)** Initialize the penalty parameter $M_k := M_0$ and the optimality tolerance $\epsilon := \epsilon_0$.
(2) **(Starting point)** Set $\mathbf{Q} := \bar{\mathbf{Q}}$ with $\bar{\mathbf{Q}}$ being an initial point. Solve all the sub-problems in (33) to obtain $\mathbf{\Lambda}^*(\mathbf{Q})$ and the gradient $\nabla\mathbf{\Lambda}^*(\mathbf{Q})$.
(3) **(Line search) while** $(\|\nabla\mathbf{\Lambda}^*(\mathbf{Q})\| > \varepsilon_0)$:
- Generate the descent direction vector $\mathbf{p} = -\nabla\mathbf{\Lambda}^*(\mathbf{Q})$.
- Set the initial step length $\tau := 1$ and $\hat{\mathbf{Q}} = diag(\hat{q}_{11}, \hat{q}_{12}, ..., \hat{q}_{n1}, \hat{q}_{n2})$ where $\hat{q}_{n1} = q_{n1} + \tau\mathbf{p}_{2n-1}$ and $\hat{q}_{n2} = q_{n2} + \tau\mathbf{p}_{2n}$, with $\mathbf{p}_{2n-1}$ and $\mathbf{p}_{2n}$ being the $(2n-1)$th and $2n$th elements of $\mathbf{p}$, respectively.
- Solve all the sub-problems in (33), verify the sufficient decrease condition (SDC) and adjust the step length τ until we find $\hat{\mathbf{Q}}$ that satisfies the SDC.

(4) **(Check convergence condition and update penalty function)**

if $\sum_{n=1}^{N} \left(|\hat{q}_{n1} - q_{n1}|^2 + |\hat{q}_{n2} - q_{n2}|^2 \right) > \epsilon_0$

- Update $M_k := M_{k+1}$. Set $\mathbf{Q} = \hat{\mathbf{Q}}$.
- Solve all the sub-problems in (33) to obtain $\mathbf{\Lambda}^*(\mathbf{Q})$ and $\nabla\mathbf{\Lambda}^*(\mathbf{Q})$.
- Go to Step 3.

else Stop.

Note that in the MASSOL algorithm, the gradient $\nabla\mathbf{\Lambda}^*(\mathbf{Q})$ is calculated as

$$\nabla\mathbf{\Lambda}^*(\mathbf{Q}) = \nabla\mathbf{\Lambda}^*(q_{11}, q_{12}, ..., q_{n1}, q_{n2}) \quad (36)$$

$$= \left(\frac{\partial\nabla\mathbf{\Lambda}^*(\mathbf{Q})}{\partial q_{11}}, \frac{\partial\nabla\mathbf{\Lambda}^*(\mathbf{Q})}{\partial q_{12}}, ..., \frac{\partial\nabla\mathbf{\Lambda}^*(\mathbf{Q})}{\partial q_{n1}}, \frac{\partial\nabla\mathbf{\Lambda}^*(\mathbf{Q})}{\partial q_{n2}} \right)^T, \quad (37)$$

where

$$\frac{\partial\nabla\mathbf{\Lambda}^*(\mathbf{Q})}{\partial q_{nj}} = \frac{\mathbf{\Lambda}^*[q_{11}, q_{12}, .., q_{nj} + \theta, ...q_{n1}, q_{n2}] - \mathbf{\Lambda}^*(\mathbf{Q})}{\theta}, \quad (38)$$

for $j = 1, 2$ and θ is an incremental step set to 10^{-3}. The penalty function M_k is determined such that the sequence $\{M_k\}$, $k = 0, 1, ..,$ tends to infinity when k increases and $M_{k+1} > M_k > 0$ [33]. In addition, while the starting point M_0 should be relatively small to avoid numerical computation difficulties, the sequence $\{M_k\}$ needs to grow quickly to a large value in order to provide a converged solution [33]. In our implementation, via extensive numerical results, we observe that using $\epsilon_0 = 10^{-3}$, $M_k = 10^k$, and $\theta = 10^{-3}$ yields a good convergence.

For the MASSOL algorithm, within each iteration, the solutions of the sub-problems are obtained from (25) and they are globally optimal. Therefore, during the line search, the feasible condition of the sub-problems is always satisfied. As a result, the proposed iterative algorithm is guaranteed to converge to a local minimizer [29]. In addition, as we show in Appendix A, $\Lambda_n(\mathbf{Q}_n, \mu_n(\mathbf{h}))$ in (30) is a concave function of $\mathbf{Q_n}$. Since the optimal-value function $\Lambda_n^*(\mathbf{Q}_n)$ in (33) is the point-wise maximum of $\Lambda_n(\mathbf{Q}_n, \mu_n(\mathbf{h}))$, $\Lambda_n^*(\mathbf{Q}_n)$ is also a concave function of $\mathbf{Q}_n$ [34]. It is further noted that the constraint in (32) is convex in the space of positive semi-definite matrices [35]. Therefore, the master problem in (32) is a concave optimization problem of $\{\mathbf{Q}_n\}$ and the local solution is a globally optimal solution. As a result, the MASSOL algorithm leads to the globally optimal solution. It should be noted that this property holds for any initial point $\bar{\mathbf{Q}} \geq 0$.

Before closing this section, it is worth mentioning about the complexity of the proposed iterative water-filling algorithm. It can be seen that in each iteration, the solution of the sub-problem given in (25) requires the calculation of λ^* and λ_n^*, $n = 1, ...N$, using bisection method. The complexity therefore grows linearly with the number of relays N [36]. To solve the master problem, the iterative process in the algorithm increases the running time but not the complexity of the algorithm [37]. Therefore, we can conclude that the complexity of the proposed iterative water-filling algorithm increases linearly with N. This property is important for a system with a large number of relays.

V. Capacity and Capacity-Achieving Input Covariance at Low and High SNRs

As presented in previous section, the ergodic capacity of the multi-relay NAF system can be calculated effectively using iterative water-filling algorithms. In this section, under the assumptions of high and low SNRs, we further analyze the capacity-achieving input covariance matrices and the behavior of the considered system, and obtain the capacity in closed-form.

A. High SNR Regime

At sufficiently high SNRs, i.e., when $\rho = E_S/N_0$ is sufficiently large, the function $\mathbf{\Gamma}_n$ in (11) is dominated by the term ρ^2. Thus, we obtain the following approximation

$$\log\mathbf{\Gamma}_n \approx \log\left(\frac{\rho^2 y^2 q_{n1} q_{n2}}{1 + b_n^2 x_n} \right)$$

$$= \log\left(\rho^2 y^2 q_{n1} q_{n2}\right) - \log\left(1 + \frac{\rho\mu_n(\mathbf{h}) x_n}{1 + \rho q_{n1} z_n} \right) \quad (39)$$

$$\approx \log\left(\rho^2 y^2 q_{n1} q_{n2}\right) - \log\left(1 + \frac{\mu_n(\mathbf{h})\, x_n}{q_{n1} z_n}\right). \tag{40}$$

Recall that $A_n = \rho^2 x_n(\rho q_{n1} z_n + 1)(\rho q_{n1} q_{n2} y^2 + q_{n2} y - q_{n1} z_n)$. Therefore, when ρ is sufficiently large, we have

$$P\left(\mu_n(\mathbf{h}) = 0\right) = P\left(A_n > 0\right) \approx 1, \tag{41}$$

It then follows that

$$\mathbb{E}_{\mathbf{h}}\left[\log\left(1 + \frac{\mu_n(\mathbf{h})\, x_n}{q_{n1} z_n}\right)\right] \approx 0. \tag{42}$$

Then by combining (10), (40), (42), we have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{h}}\left[I\left(\mathbf{x}, \mathbf{y}|\mathbf{h}\right)\right] \approx \mathbb{E}_{\mathbf{h}}\left[\sum_{n=1}^{N} \log\left(\rho^2 y^2 q_{n1} q_{n2}\right)\right] \\
&-\mathbb{E}_{\mathbf{h}}\left[\sum_{n=1}^{N} \log\left(1 + \tfrac{\mu_n(\mathbf{h}) x_n}{q_{n1} z_n}\right)\right] \\
&\approx \mathbb{E}_{\mathbf{h}}\left[\sum_{n=1}^{N} \log\left(\rho^2 y^2 q_{n1} q_{n2}\right)\right] \\
&\leq \mathbb{E}_{\mathbf{h}}\left[2N \log \frac{\rho y \sum_{n=1}^{N} (q_{n1} + q_{n2})}{2N}\right] \\
&= 2N \log\left(\tfrac{\rho \phi_{SD} P_S}{2N}\right).
\end{aligned} \tag{43}$$

In (43), the second inequality is obtained by applying Cauchy-Schwarz inequality and the equality is achieved when $q_{n1} = q_{n2} = \frac{P_S}{2N}$, $n = 1, ...N$. In addition, the third equality comes from the fact that $\sum_{n=1}^{N} (q_{n1} + q_{n2}) = P_S$ and $\mathbb{E}_{\mathbf{h}}[y] = \mathbb{E}_{\mathbf{h}}\left[|h_{SD}|^2\right] = \phi_{SD}$. Therefore, the equal power allocation at the source over all the transmission phases, referred to as the EPA-S scheme, is optimal. For convenience, let $q_s = \frac{P_S}{2N}$. The optimal input covariance matrix is simply a scaled identity matrix, and can be expressed as $\mathbf{Q} = \mathbf{Q}_{EPA-S} = q_s \mathbf{I}_{2N}$. The channel capacity can then be simplified to

$$C = R\left(\mathbf{Q}_{EPA-S}\right) = \frac{1}{2} \max_{\substack{\mu_n(\mathbf{h}), n=\overline{1,N} \\ \mathbf{Q} = q_s \mathbf{I}_{2N}}} \int_{\mathbf{h}} I\left(\mathbf{x}, \mathbf{y}|\mathbf{h}\right) f\left(\mathbf{h}\right) d\mathbf{h}. \tag{44}$$

Note that the capacity in (44) can be achieved by incorporating the EPA-S scheme at the source and the OPA-R scheme at the relays. By substituting $q_{n1} = q_{n2} = q_s$ into (25) we obtain the optimal power allocation at the relays to achieve the capacity in (44) as

$$\bar{\mu}_n^*(\mathbf{h}) = \begin{cases} 0, & \bar{A}_n \geq 0, \\ \left(\frac{-\bar{D}_n + \sqrt{\bar{D}_n^2 - 4\bar{C}_n\left(\bar{E}_n + \frac{\bar{A}_n}{\ln(2) N (\bar{\lambda}^* + \bar{\lambda}_n^*)}\right)}}{2\bar{C}_n}\right)^+, & \bar{A}_n < 0, \end{cases} \tag{45}$$

where

$$\begin{aligned}
&\bar{A}_n = \rho^2 x_n q_s (\rho q_s z_n + 1)(\rho q_s y^2 + y - z_n), \\
&\bar{C}_n = \rho^2 x_n^2 (\rho q_s y + \rho q_s z_n + 1), \\
&\bar{D}_n = \rho x_n (\rho q_s z_n + 1)(3\rho q_s y + \rho q_s z_n + \rho^2 q_s^2 y^2 + 2), \\
&\bar{E}_n = (\rho q_s y + 1)^2 (\rho q_s z_n + 1)^2,
\end{aligned} \tag{46}$$

and $\bar{\lambda}^*$ and $\bar{\lambda}_n^*$ satisfying

$$\begin{cases}
\sum_{n=1}^{N} \int_{\mathbf{h}} \bar{\mu}_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_R, \\
\bar{\lambda}^* \left(P_R - \sum_{n=1}^{N} \int_{\mathbf{h}} \bar{\mu}_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h}\right) = 0, \\
\int_{\mathbf{h}} \bar{\mu}_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq P_n, \\
\bar{\lambda}_n^* \left(P_n - \int_{\mathbf{h}} \bar{\mu}_n^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h}\right) = 0, \\
\bar{\lambda}^* \geq 0,\ \bar{\lambda}_n^* \geq 0,\ \left(\bar{\lambda}^* + \bar{\lambda}_n^*\right) > 0.
\end{cases} \tag{47}$$

Given the above results, the capacity for the considered system at high SNRs can then be obtained in closed-form as

$$\begin{aligned}
C = \tfrac{1}{2} \sum_{n=1}^{N} \int_{\mathbf{h}} \log \Big[&\tfrac{\rho^2 y^2 q_s^2}{1 + b_n^2 x_n} + \\
&\rho\left(q_s y + \tfrac{q_s \bar{b}_n^2 x_n z_n + q_s y}{1 + b_n^2 x_n}\right) + 1\Big] f(\mathbf{h})\, d\mathbf{h},
\end{aligned} \tag{48}$$

where

$$\bar{b}_n = \sqrt{\frac{\rho \bar{\mu}_n^*(\mathbf{h})}{z_n q_s \rho + 1}}, \tag{49}$$

with $\bar{\mu}_n^*(\mathbf{h})$ given in (45).

B. Low SNR Regime

In this subsection, we shall investigate the system that operates at low SNR regimes, i.e., when ρ is small. From (10) and (11), for a given input covariance matrix, the unconditional mutual information can be calculated as

$$\begin{aligned}
I\left(\mathbf{x}, \mathbf{y}\right) &= \mathbb{E}_{\mathbf{h}}[I(\mathbf{x}, \mathbf{y}|\mathbf{h})] \\
&= \mathbb{E}_{\mathbf{h}}\Bigg[\log \prod_{n=1}^{N} \Big(1 + \tfrac{\rho^2 y^2 q_{n1} q_{n2}}{1 + b_n^2 x_n} + \\
&\qquad \rho\left(q_{n1} y + \tfrac{q_{n1} b_n^2 x_n z_n + q_{n2} y}{1 + b_n^2 x_n}\right)\Big)\Bigg]
\end{aligned} \tag{50}$$

For convenience, let

$$F_n = \frac{\rho^2 y^2 q_{n1} q_{n2}}{1 + b_n^2 x_n} + \rho\left(q_{n1} y + \frac{q_{n1} b_n^2 x_n z_n + q_{n2} y}{1 + b_n^2 x_n}\right). \tag{51}$$

When ρ is very small, by ignoring the terms containing high orders of ρ, we obtain the following approximation:

$$\prod_{n=1}^{N} (1 + F_n) \approx 1 + \sum_{n=1}^{N} F_n. \tag{52}$$

It then follows from (50) that

$$\begin{aligned}
I\left(\mathbf{x}, \mathbf{y}\right) &\approx \mathbb{E}_{\mathbf{h}}\left[\log\left(1 + \sum_{n=1}^{N} F_n\right)\right] \\
&= \mathbb{E}_{\mathbf{h}}\Bigg[\log\Bigg(1 + \rho y \sum_{n=1}^{N} \left(q_{n1} + \tfrac{q_{n2}}{1 + b_n^2 x_n}\right) \\
&\qquad + \sum_{n=1}^{N} \left(\tfrac{\rho^2 y^2 q_{n1} q_{n2} + \rho q_{n1} b_n^2 x_n z_n}{1 + b_n^2 x_n}\right)\Bigg)\Bigg] \\
&\leq \mathbb{E}_{\mathbf{h}}\Big[\log\left(1 + \rho y P_S + \right. \\
&\qquad \left.\rho^2 \sum_{n=1}^{N} \tfrac{y^2 q_{n1} q_{n2}(1 + \rho q_{n1} z_n) + q_{n1} \mu_n(\mathbf{h}) x_n z_n}{1 + \rho q_{n1} z_n + \rho \mu_n(\mathbf{h}) x_n}\right)\Big].
\end{aligned} \tag{53}$$

The inequality in (53) comes from the fact that $\sum_{n=1}^{N} \left(q_{n1} + \frac{q_{n2}}{1 + b_n^2 x_n}\right) \leq \sum_{n=1}^{N} (q_{n1} + q_{n2}) = P_S$ and we

achieve the equality when either $q_{n2} = 0$ or $\mathbb{E}_{\mathbf{h}}[b_n] = 0$ for each relay R_n, $n = 1, ...N$. Note that $\mathbb{E}_{\mathbf{h}}[b_n] = 0$ means the relay R_n is inactive. All schemes achieving the equality in (53) can then be categorized in two cases as follows.

(i) All the relays are inactive. In this case, the system becomes a direct channel, i.e., we have only a direct transmission from the source to the destination. It then follows that

$$I(\mathbf{x}, \mathbf{y}) \approx \mathbb{E}_{\mathbf{h}}\left[1 + \rho y P_S + \sum_{n=1}^{N} \rho^2 y^2 q_{n1} q_{n2}\right]. \quad (54)$$

It can be easily verified that the unconditional mutual information in (54) is maximized when $q_{n1} = q_{n2} = \frac{P_S}{2N}$, $n = 1, ...N$. Then for this case, the maximum information rate can be expressed as

$$I(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\mathbb{E}_{\mathbf{h}}\left[\log \prod_{n=1}^{N}\left(1 + \rho^2 y^2 \left(\frac{P_S}{2N}\right)^2 + \rho y \frac{P_S}{N}\right)\right]$$
$$= \frac{1}{2}\mathbb{E}_{\mathbf{h}}\left[2N \log\left(1 + \rho y \frac{P_S}{2N}\right)\right] \approx \frac{1}{2}\mathbb{E}_{\mathbf{h}}\left[\rho y P_S \log e\right]. \quad (55)$$

(ii) At least one relay is active. In this case, without loss of generality, let K be the set of the active relays. For these relays, we have

$$\mathbb{E}_{\mathbf{h}}[b_k] \neq 0 \text{ and } q_{k2} = 0, \forall k \in \mathrm{K}. \quad (56)$$

It follows from (53) that

$$I(\mathbf{x}, \mathbf{y}) = \mathbb{E}_{\mathbf{h}}\left[\log\left(1 + \rho y P_S + \rho^2 \sum_{k\in\mathrm{K}} \frac{q_{k1}\mu_k(\mathbf{h}) x_k z_k}{1 + \rho q_{k1} z_k + \rho \mu_k(\mathbf{h}) x_k}\right)\right]. \quad (57)$$

Observe that $\log(1 + x) \approx x \log e$ when $x \approx 0$. Therefore, when ρ is sufficiently small, we have

$$\begin{aligned} & I(\mathbf{x}, \mathbf{y}) \\ \approx\ & \mathbb{E}_{\mathbf{h}}\left[\left(\rho y P_S + \rho^2 \sum_{k\in\mathrm{K}} \frac{q_{k1}\mu_k(\mathbf{h}) x_k z_k}{1 + \rho q_{k1} z_k + \rho \mu_k(\mathbf{h}) x_k}\right) \log e\right] \\ \approx\ & \mathbb{E}_{\mathbf{h}}\left[\left(\rho y P_S + \rho^2 \sum_{k\in\mathrm{K}} q_{k1}\mu_k(\mathbf{h}) x_k z_k\right) \log e\right] \\ =\ & \mathbb{E}_{\mathbf{h}}\left[\left(\rho y P_S + \sum_{k\in\mathrm{K}} \Psi_k(q_{k1})\right) \log e\right] \end{aligned} \quad (58)$$

where

$$\Psi_k(q_{k1}) = \rho^2 q_{k1} \mu_k(\mathbf{h}) x_k z_k. \quad (59)$$

By comparing (55) and (58), it can be seen that the direct transmission is suboptimal. Therefore, for a capacity-achieving scheme, at least one relay must be active. We now consider an arbitrary pair of relays (R_i, R_j) in K such that R_j has a stronger cascaded S-R and R-D channels, i.e., $\mathbb{E}_{\mathbf{h}}[x_j z_j] \geq \mathbb{E}_{\mathbf{h}}[x_i z_i]$. We then have the following Lemma.

Lemma 1. *For the same amount of power q allocated to the broadcasting phases at the source associated with R_i and R_j in* K*, we have* $\mathbb{E}_{\mathbf{h}}[\Psi_j(q)] \geq \mathbb{E}_{\mathbf{h}}[\Psi_i(q)]$.

Proof: Let $P_{\text{off}}^k = \Pr[\mu_k(\mathbf{h}) = 0]$, $k = i, j$, be the probability that the relay R_k is silent at a given cooperative frame. The probability of the relay R_k being active in that cooperative frame is therefore $P_{\text{on}}^k = \Pr[\mu_k(\mathbf{h}) > 0] = 1 - P_{\text{off}}^k$. For simplicity, assume that the IAPC is the same for all relays. The average power $\bar{\mathbb{E}}_{\mathbf{h}}[\mu_k(\mathbf{h})]$ allocated to relay R_k can then be expressed as

$$\bar{\mathbb{E}}_{\mathbf{h}}[\mu_k(\mathbf{h})] = \begin{cases} \mathbb{E}_{\mathbf{h}}[\mu_k(\mathbf{h})] & \mathbb{E}_{\mathbf{h}}[\mu_k(\mathbf{h})] \leq P_k, \\ P_k & \mathbb{E}_{\mathbf{h}}[\mu_k(\mathbf{h})] > P_k, \end{cases} \quad (60)$$

where $\mathbb{E}_{\mathbf{h}}[\mu_k(\mathbf{h})]$ is the average power allocated to relay R_k under TAPC. Observe from (28) and (62) that under TAPC, the relay R_k remains silent at a given frame if and only if

$$\sqrt{D_k^2 - 4C_k\left(E_k + \frac{A_k}{\ln(2) N \lambda^*}\right)} \leq -D_k. \quad (61)$$

It follows from (56) that

$$A_k = -\rho^2 q x_k z_k (\rho q z_k + 1) \leq 0 \ \ k = i, j. \quad (62)$$

Hence, (61) is equivalent to

$$\rho^2 q x_k z_k \leq \lambda^* \ln(2), \quad (63)$$

when ρ is sufficiently small. Furthermore, since $\mathbb{E}_{\mathbf{h}}\left[\rho^2 q x_j z_j\right] > \mathbb{E}_{\mathbf{h}}\left[\rho^2 q x_i z_i\right]$, we have $P_{\text{on}}^j > P_{\text{on}}^i$. As a result, $\mathbb{E}_{\mathbf{h}}[\mu_j(\mathbf{h})] \geq \mathbb{E}_{\mathbf{h}}[\mu_i(\mathbf{h})]$. Combining this with (60), we have $\bar{\mathbb{E}}_{\mathbf{h}}[\mu_j(\mathbf{h})] > \bar{\mathbb{E}}_{\mathbf{h}}[\mu_i(\mathbf{h})]$ for the system under both TAPC and IAPC.

From the above analysis, it turns out that we always have

$$\mathbb{E}_{\mathbf{h}}\left[\rho^2 q \mu_j(\mathbf{h}) x_j z_j\right] > \mathbb{E}_{\mathbf{h}}\left[\rho^2 q \mu_i(\mathbf{h}) x_i z_i\right]. \quad (64)$$

Therefore, $\mathbb{E}_{\mathbf{h}}[\Psi_j(q)] \geq \mathbb{E}_{\mathbf{h}}[\Psi_i(q)]$. ■

Using Lemma 1, it is then easy to show that for a given total power allocated to the broadcasting phases at the source associated with a pair of relays R_i and R_j having $\mathbb{E}_{\mathbf{h}}[x_j z_j] \geq \mathbb{E}_{\mathbf{h}}[x_i z_i]$, it is more beneficial to allocate all the power to the broadcasting phase associated with R_j. The result can be straightforwardly extended to any number of relays. In particular, for a system having N relays, if R_o has the strongest cascaded S-R and R-D channels, the optimal power allocation at the source to achieve the capacity at low SNRs will be

$$\begin{cases} q_{o1} = P_S, & \\ q_{n1} = 0 & n = 1, ...N, n \neq o, \\ q_{n2} = 0 & n = 1, ...N. \end{cases} \quad (65)$$

This power allocation scheme is optimal at low SNR regimes and it is referred to as the selective power allocation scheme at the source, denoted as SPA-S. By using this scheme, one should allocate all the power to the relay having the strongest channel cascaded by the source-relay and relay-destination links and keep all other relays silent. As a result, the channel capacity can be simplified to $C = \frac{1}{2}\max_{\mu_o(\mathbf{h})} \int_{\mathbf{h}} I(\mathbf{x}, \mathbf{y}|\mathbf{h}) f(\mathbf{h})\, d\mathbf{h}$, and can be easily calculated. In particular, the optimal power adaptation solution at the relay R_o can be expressed as

$$\mu_o^*(\mathbf{h}) = \begin{cases} 0, & A_o \geq 0, \\ \left(\frac{-D_o + \sqrt{D_o^2 - 4C_o\left(E_o + \frac{A_o}{\ln(2) N \lambda_o^*}\right)}}{2C_o}\right)^+, & A_o < 0, \end{cases} \quad (66)$$

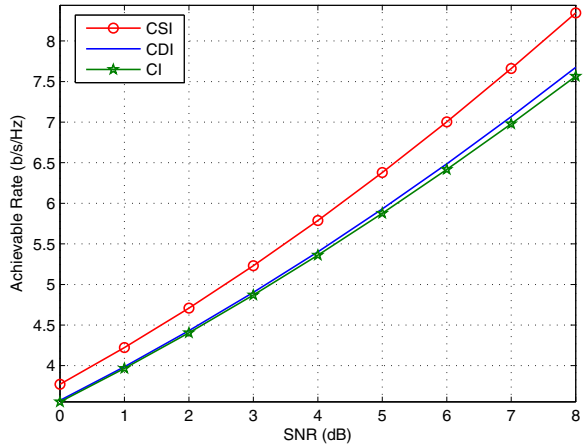

Fig. 2. The achievable rate of different NAF systems over the symmetric channel.

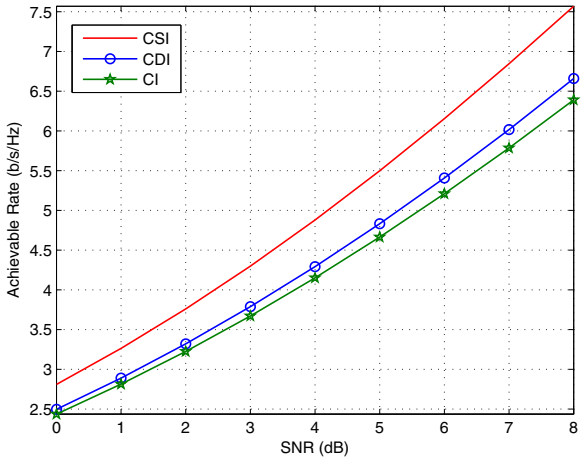

Fig. 3. The achievable rate of different NAF systems over the asymmetric channel.

while $\mu_n^*(\mathbf{h}) = 0, \quad n = 1, ...N, n \neq o,$ with

$$\begin{aligned} A_o &= -\rho^2 P_S x_o z_o (\rho P_S z_o + 1), \\ C_o &= \rho^2 x_o^2 (\rho P_S y + \rho P_S z_o + 1), \\ D_o &= \rho x_o (\rho P_S z_o + 1)(2\rho P_S y + \rho P_S z_o + 2), \\ E_o &= (\rho P_S y + 1)(\rho P_S z_o + 1)^2, \end{aligned} \tag{67}$$

and $\lambda_o^* > 0$ is the constant satisfying

$$\int_{\mathbf{h}} \mu_o^*(\mathbf{h}) f(\mathbf{h})\, d\mathbf{h} \leq \min(P_o, P_R). \tag{68}$$

Therefore, the system ergodic capacity at low SNRs under both TAPC and IAPC can be written as

$$C = \frac{1}{2} \int_{\mathbf{h}} \log \left[1 + \rho y P_S + \frac{\rho^2 P_S \mu_o^*(\mathbf{h})\, x_o z_o}{1 + \rho P_S z_o + \rho \mu_o^*(\mathbf{h})\, x_o} \right] f(\mathbf{h})\, d\mathbf{h}. \tag{69}$$

VI. Illustrative Results

In this section, numerical results are presented to illustrate the achievable rate and ergodic capacity obtained by our derived power allocation schemes using CSI. We will only focus on the systems having three relays, but the results can be obtained for any number of relays. Unless otherwise stated, the distance-dependent path-loss channel model [38] is adopted such that the channel variances depend on the path-loss exponent, i.e., $\phi_{SD} = d_{SD}^{-v}$, $\phi_{SR_n} = d_{SR_n}^{-v}$, $\phi_{RD_n} = d_{RD_n}^{-v}$ where $v = 3$ is chosen as the path-loss exponent. We consider both the symmetric channel in which we set $d_{SD} = d_{SR_n} = d_{RDn} = 1$, $n = 1, 2, 3$; and the asymmetric channel in which we set $d_{SD} = 1$, $d_{SR_1} = d_{RD_1} = 2$, $d_{SR_2} = d_{RD_2} = 0.5$, $d_{SR_3} = d_{RD_3} = 1$. In all cases, it is assumed that $P_S = P_R = 6$ and $P_1 = P_2 = P_3 = 0.4P_R$.

A. Achievable Rate: CSI vs. CDI/CI

In this subsection, we shall illustrate the gain in terms of achievable rate provided by having full CSI at the relays. The two conventional systems using CDI and CI will be used as benchmarks for comparison with our OPA-R scheme using CSI. Fig. 2 first plots the achievable rate of the CSI, CDI, and CI systems in the symmetric configuration. Note that all the considered systems use the equal power allocation at the source and have the same TAPC at the relays. Observe from Fig. 2 that the CSI system significantly outperforms the CDI and CI systems. The gains achieved by the CSI system over the CDI and CI systems are around 0.8dB and 1dB, respectively.

The gain offered by CSI is slightly larger in the asymmetric configuration. In particular, Fig. 3 provides the achievable rate of the CSI, CDI, and CI systems in the asymmetric configuration. Note from Fig. 3 that the gains are as much as 1.1dB and 1.4dB over the CDI and CI systems, respectively. Again, this gain comes from the flexibility in sharing the TAPC among three relays, thanks to the knowledge of CSI.

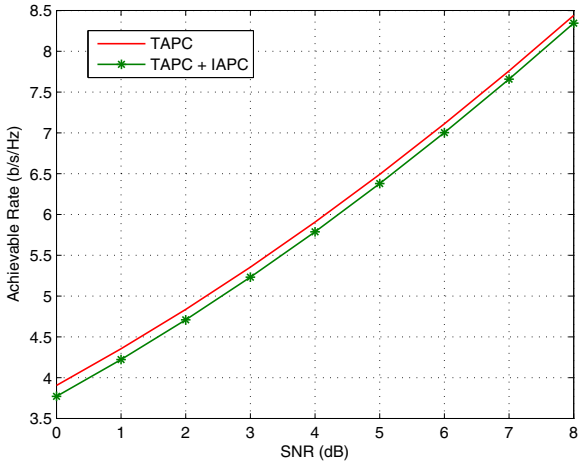

Fig. 4. The achievable rate of the NAF system with full CSI under both TAPC and IAPC and under TAPC only over the asymmetric channel.

Finally, Fig. 4 compares the achievable rate of the two NAF systems using OPA-R with full CSI and equal power allocation at the source under both TAPC and IAPC and under only TAPC over the asymmetric channel. As expected, there is a small penalty by imposing individual power constraint on each relay. The loss is of about 0.2dB at various SNR regions.

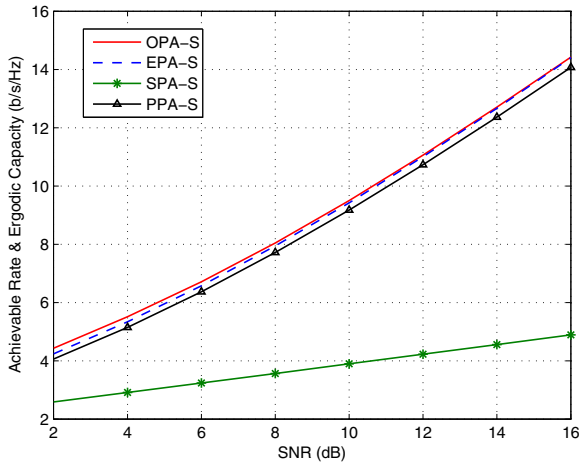

Fig. 5. The achievable rate and ergodic capacity of the NAF systems over the asymmetric channel in medium and high SNR regimes.

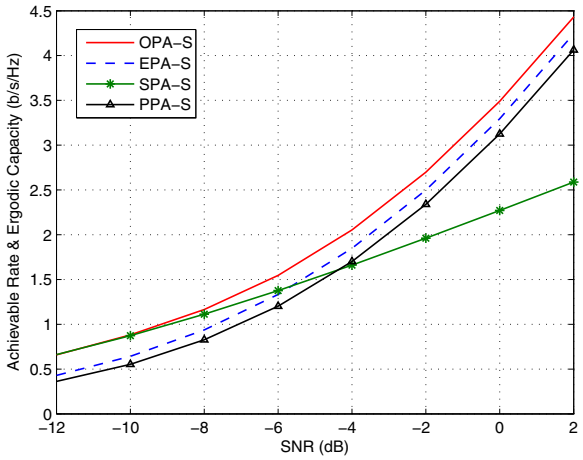

Fig. 6. The achievable rate and ergodic capacity of the NAF system over the asymmetric channel in low SNR regimes.

B. Effects of Power Allocation at the Souce

In this subsection, given the OPA-R scheme at the relays, we shall examine the impact of power allocation at the source. In particular, we shall consider the systems with OPA-R scheme at the relays incorporating with four power allocation schemes at the source as follows:

i) *OPA-S* - a capacity-achieving system with optimal power allocation at the source using our derived results from the MASSOL algorithm.
ii) *EPA-S* - a system with equal power allocation at the source in all the transmission phases, i.e., $q_{n1} = q_{n2} = 1$, $n = 1, 2, 3$. As demonstrated in Section V, this system is optimal at high SNR regimes.
iii) *SPA-S* - a system using only the relay having strongest cascaded source-relay and relay-destination channels. In our configuration, relay R_2 is used while relay R_1 and R_3 keep silent. The power allocation at the source is: $q_{11} = q_{12} = q_{22} = q_{31} = q_{32} = 0, q_{21} = 6$. This scheme was shown to be optimal at low SNR regimes in Section V.
iv) *PPA-S* - a system where the transmit power at the source is pre-selected as $q_{11} = 2, q_{12} = 1, q_{21} = q_{22} = 0.5, q_{31} = q_{32} = 1$.

For brevity of the presentation, we only present the results for the asymmetric channel under both TAPC and IAPC. Similar results can be obtained for other scenarios.

We first examine the four considered systems at medium and high SNR regimes. Specifically, the achievable rates of the four NAF systems are plotted in Fig. 5. It can be seen that the OPA-S system has the best performance since it is the capacity-achieving scheme. The OPA-S system provides significant gains over the SPA-S and PPA-S systems in a wide range of SNRs and slightly outperforms the EPA-S in medium SNR regions. At sufficiently high SNRs, the achievable rate of the EPA-S system approaches the channel capacity achieved by the OPA-S system, which confirms our analysis about the optimality of the EPA-S scheme at high SNR regimes.

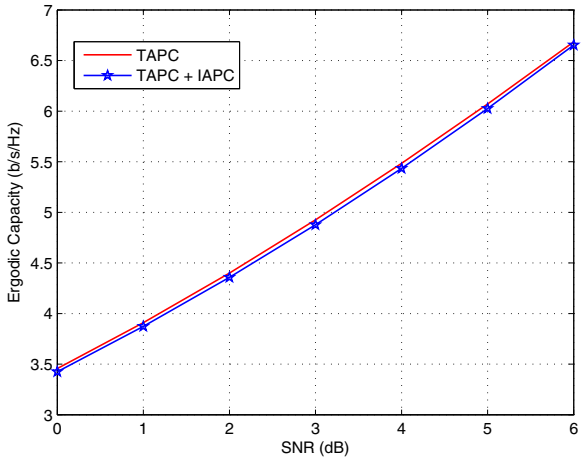

Fig. 7. The ergodic capacity of the NAF system under both TAPC and IAPC and under TAPC only over the asymmetric channel.

Fig. 6 compares the maximum achievable rates of the four considered systems at low SNR regions. Given that the OPA-S system is optimal, it can be seen that the OPA-S system achieves the best performance. Different from the high SNR scenario, the OPA-S system provides a gain ranging from 0.5dB to 2dB over the EPA-S system. When the SNR decreases, we can see that the SPA-S system achieves the same rate as that of the optimal OPA-S system, which is consistent with our analysis about the optimality of SPA-S at low SNRs in Section V.

C. Channel Capacity under both TAPC and IAPC and under TAPC only

Finally, in Fig. 7, we compare the capacity limits of two OPA-S schemes under both TAPC and IAPC and under TAPC only over the asymmetric channel. Similar to the achievable rate, we also observe that the rate decreases by imposing the individual power constraint on each relay. The loss is of about 0.1dB in a wide range of SNRs.

VII. Conclusion

This paper addressed the achievable rate and ergodic capacity of an NAF multi-relay network with CSI at the relays by examining the optimal power allocation schemes at the relays (OPA-R) and at the source (OPA-S) to maximize the achievable rate. At first, for a given power allocation scheme at the source, the OPA-R solutions were derived for the systems under both TAPC and IAPC. It was then shown that the proposed OPA-R schemes with CSI at the relays significantly outperform the conventional systems with CDI and CI. Furthermore, by jointly investigating the OPA-R and OPA-S schemes, we established the ergodic capacity for the considered systems. We also demonstrated that at high SNR regions, the equal power allocation scheme at the source (EPA-S) incorporated with the OPA-R scheme achieves the capacity. On the other hand, at low SNR regimes, it was shown that the capacity can be achieved by using only the relay with the strongest cascaded source-relay and relay-destination channels while the source spends all of its transmit power in the broadcasting phase associated with that relay.

Finally, it is noted that this paper considered an idealized i.i.d channel. It is certainly of practical interest to investigate a more realistic channel model such as the spatial channel model [39], [40]. Under this channel model, one needs to take into account spatial information such as the angle of departure, angle of arrival, and the spacing between the transmitter and receiver. Extending the results in this paper to such a realistic model is therefore a promising research direction.

Appendix A
Concavity of $\Lambda_n\left(\mathbf{Q}_n, \mu_n\left(\mathbf{h}\right)\right)$

From (30), it can be seen that we need to show $\log\det\left[\mathbf{I}_2 + \rho\mathbf{H}_n^\dagger\mathbf{H}_n\mathbf{Q}_n\right]$ is a concave function of $\mathbf{Q}_n$. Using $\det\left(\mathbf{I}+\mathbf{AB}\right) = \det\left(\mathbf{I}+\mathbf{BA}\right)$, we have:

$$\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}_n^\dagger\mathbf{H}_n\mathbf{Q}_n\right) = \log\det\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{Q}_n\mathbf{H}_n^\dagger\right). \tag{70}$$

Observe that $\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{Q}_n\mathbf{H}_n^\dagger\right)$ is Hermitian. Furthermore, given a positive semi-definite matrix $\mathbf{Q}_n$, for any non-zero vector $\mathbf{u}$ we have

$$\mathbf{u}^*\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{Q}_n\mathbf{H}_n^\dagger\right)\mathbf{u} = \left|\mathbf{u}\right|^2 + \rho\left(\mathbf{H}_n^\dagger\mathbf{u}\right)^*\mathbf{Q}_n\mathbf{H}_n^\dagger\mathbf{u} \geq 0. \tag{71}$$

Thus, $\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{Q}_n\mathbf{H}_n^\dagger\right)$ is Hermitian and positive semi-definite. Using Theorem 7.6.7 in [41], we can then conclude that $\log\det\left(.\right)$ is a concave function of $\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{Q}_n\mathbf{H}_n^\dagger\right)$. As a result, for any two given 2×2 positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$ and for all $\alpha\in(0,1)$ we obtain

$$\begin{aligned}&\log\det\left[\alpha\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{A}\mathbf{H}_n^\dagger\right) + \left(1-\alpha\right)\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{B}\mathbf{H}_n^\dagger\right)\right]\\ &\geq \alpha\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{A}\mathbf{H}_n^\dagger\right) + \left(1-\alpha\right)\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{B}\mathbf{H}_n^\dagger\right).\end{aligned} \tag{72}$$

Equivalently, we have

$$\begin{aligned}&\log\det\left[\mathbf{I}_2 + \rho\mathbf{H}_n\left(\alpha\mathbf{A} + \left(1-\alpha\right)\mathbf{B}\right)\mathbf{H}_n^\dagger\right]\\ &\geq \alpha\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{A}\mathbf{H}_n^\dagger\right) + \left(1-\alpha\right)\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}_n\mathbf{B}\mathbf{H}_n^\dagger\right).\end{aligned} \tag{73}$$

Thus, $\log\det\left[\mathbf{I}_2 + \rho\mathbf{H}_n^\dagger\mathbf{H}_n\mathbf{Q}_n\right]$ is the concave function of $\mathbf{Q}_n$ over the set of positive semi-definite matrices. As a consequence, $\Lambda_n\left(\mathbf{Q}_n, \mu_n\left(\mathbf{h}\right)\right)$ is the concave function of $\mathbf{Q}_n$.

References

[1] T. X. Tran, N. H. Tran, and H. R. Bahrami, "Optimal power sharing strategies in NAF multiple-relay networks with CSI," in *Proc. 2013 IEEE Int. Conf. Commun.*, pp. 4620–4624.

[2] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, pp. 74–80, Oct. 2004.

[3] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channel: performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1099–1109, Aug. 2004.

[4] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, Sept. 2005.

[6] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. 1427–1438, 2007.

[7] J. Luo, R. S. Blum, L. J. Cimini, L. J. Greenstein, and A. M. Haimovich, "Decode-and-forward cooperative diversity with power allocation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 793–799, 2007.

[8] O. Somekh, O. Simeone, H. V. Poor, and S. Shamai, "Cellular systems with full-duplex compress-and-forward relaying and cooperative base stations," in *Proc. 2008 IEEE Int. Symp. Inform. Theory*, pp. 2086–2090.

[9] S. Simoens, O. Muñoz, and J. Vidal, "Achievable rates of compress-and-forward cooperative relaying on Gaussian vector channels," in *Proc. 2007 IEEE Int. Conf. Commun.*, pp. 4225–4231.

[10] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4152–4172, Dec. 2005.

[11] H. Bölcskei, R.Nabar, O. Oyman, and A. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 1433–1444, June 2006.

[12] J. Z. B. Wang and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 29–43, Jan. 2005.

[13] S. Shrestha and K. Chang, "Analysis of outage capacity performance for cooperative DF and AF relaying in dissimilar Rayleigh fading channels," *Proc. 2008 IEEE Int. Symp. Inform. Theory*, pp. 494–498.

[14] Y. Ding, J. Zhang, and K. M. Wong, "Ergodic channel capacities for the amplify-and-forward half-duplex cooperative systems," *IEEE Trans. Inf. Theory*, vol. 55, pp. 713–730, Feb. 2009.

[15] K. Loa, C. C. Wu, S. T. Sheu, Y. Yuan, M. Chion, D. Huo, and L. Xu, "IMT-Advanced relay standards," *IEEE Commun. Mag.*, vol. 25, pp. 40–48, Aug. 2010.

[16] M. Hasna and M.-S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, pp. 216–218, May 2003.

[17] M. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1126–1131, Nov. 2003.

[18] M. Hasna and M.-S. Alouini, "Optimal power allocation for relayed transmissions over Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1999–2004, Nov. 2004.

[19] I. Hammerstrom and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2798–2802, Aug. 2007.

[20] G. Farhadi and N. Beaulieu, "Power-optimized amplify-and-forward multi-hop relaying systems," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4634–4643, Sept. 2009.

[21] L. Sanguinetti, A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, pp. 1331–1346, Sept. 2012.

[22] Y. Ding, J. K. Zhang, and K. M. Wong, "The amplify-and-forward half-duplex cooperative system: pairwise error probability and precoder design," *IEEE Trans. Signal Process.*, vol. 55, pp. 605–617, Feb. 2007.

[23] W. Zhang and K. B. Letaief, "Bandwidth efficient cooperative diversity for wireless networks," in *Proc. 2007 IEEE Global Telecommun. Conf.*, pp. 2942–2946.

[24] L. J. Rodriguez, A. Helmy, N. H. Tran, and T. Le-Ngoc, "Optimal power adaption for NAF relaying with channel side information," in *Proc. 2012 IEEE Int. Conf. Commun.*, pp. 4537–4541.

[25] L. J. Rodriguez, N. H. Tran, A. Helmy, and T. Le-Ngoc, "Optimal power adaption for cooperative AF relaying with channel side information," *IEEE Trans. Veh. Technol.*, pp. 3164–3174, Sept. 2013.

[26] T. X. Tran, N. H. Tran, H. R. Bahrami, H. Dinh, and S. Sastry, "On achievable rate and ergodic capacity of OAF multiple-relay networks with CSI," in *Proc. 2013 IEEE Veh. Technol. Conf. — Spring*, pp. 1–6.
[27] A. Murugan, K. Azarian, and H. E. Gamal, "Cooperative lattice coding and decoding," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 268–279, Feb. 2007.
[28] S. Yang and J. -C. Belfiore, "Optimal space-time codes for the MIMO amplify-and-forward cooperative channel," *IEEE Trans. Inf. Theory*, vol. 53, pp. 647–663, Feb. 2007.
[29] K. Tammer, "The application of parametric optimization and imbedding for the foundation and realization of a generalized primal decomposition approach," *Parametric Optimization and Related Topics, Mathematical Research*, vol. 35, pp. 376–386, Akademie-Verlag, Berlin, 1987.
[30] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3114–3123, Aug. 2007.
[31] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
[32] H. Weldegiorgis, *Penalty and Barrier Function Methods for Constrained Optimization*. LAP LAMBERT Academic Publishing, 2011.
[33] P. A. Jensen and J. F. Bard, *Operations Research Models and Methods*. Wiley, 2002.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge. University Press, 2004.
[35] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Maxtrix Anal. Applic.*, vol. 19, pp. 499–533, 1998.
[36] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, 2013.
[37] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.
[38] N. Ahmed, M. Khojastepour, and B. Aazhang, "Outage minimization and optimal power control for the fading relay channel," in *Proc. 2004 IEEE Information Theory Workshop*, pp. 458–462.
[39] 3rd Generation Partnership Project (3GPP), "Spatial channel model for multipe input multiple output (MIMO) simulations (3GPP TR 25.996 version 6.1.0 release 6)," ETSI, Tech. Rep., 2003.
[40] J. Salo, G. D. Galdo, J. Salmi, P. Kyosti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996)." Available: http://www.tkk.fi/Units/Radio/scm/, Jan. 2005.
[41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

Tuyen X. Tran received the B.Eng. degree (Honors Program) in Electronics & Telecommunications from Hanoi University of Technology, Vietnam in 2011, and the M.Sc. degree in Electrical & Computer Engineering from the University of Akron, OH, USA in 2013. His research interest lies in the areas of information theory, cooperative wireless communications and MIMO networks. He received the Dean's Fellowship Award in recognition of outstanding graduate work and research achievement from the University of Akron in 2012-2013.

Nghi H. Tran received the B.Eng. degree from Hanoi University of Technology, Vietnam in 2002, the M.Sc. degree (with Graduate Thesis Award) and the Ph.D. degree from the University of Saskatchewan, Canada in 2004 and 2008, respectively, all in Electrical & Computer Engineering. From May 2008 to July 2010, he was at McGill University as a Postdoctoral Scholar under the prestigious Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowship. From August 2010 to July 2011, Dr. Tran was at McGill University as a Research Associate. He also worked as a Consultant in the satellite industry. Since August 2011, Dr. Tran has been an Assistant Professor in the Department of Electrical and Computer Engineering, University of Akron, OH, USA. Dr. Tran's research interests span the areas of signal processing and communication and information theories for wireless systems and networks. Dr. Tran has been serving as a TPC member for a number of flagship IEEE conferences. He is a TPC co-Chair of the Workshop on Trusted Communications with Physical Layer Security for IEEE GLOBECOM 2014.

Hamid Reza Bahrami is currently an Assistant Professor at the Department of Electrical and Computer Engineering at The University of Akron, OH, USA. His main area of research includes wireless communication, information theory, and applications of signal processing in communication. He received his Ph.D. degree in electrical engineering from McGill University, Montreal, Canada in 2008. From 2007 to 2009 and prior to joining the University of Akron, he was a scientist at Wavesat Inc. Montreal, Canada. Dr. Bahrami has B.Sc. and M.Sc. degrees both in Electrical Engineering from Sharif University of Technology and University of Tehran, Iran, respectively. He has served as the Editor for Transactions on Emerging Telecommunications Technologies, and as the Technical Program Committee member of numerous IEEE conferences including IEEE GLOBECOM and International Conference on Communications (ICC). He is currently a member of the IEEE, the IEEE Communications Society, and the IEEE Vehicular Technology Society.

Dr. Shivakumar Sastry received his Ph.D. degree in Computer Engineering and Science from Case Western Reserve University. He earned an M.Sc.(Engg) degree from Indian Institute of Science and an MS Computer Science degree from University of Central Florida. He has extensive industry experience in software engineering and automation systems and is an expert in graph algorithms and modeling. His current interests are in sensor-actuator systems for automation, provable adaptation, diagnostics and prognostics, and predictable monitoring. He is currently an Associate Professor at University of Akron. Here, he leads the Complex Engineered Systems Lab. Prior to joining Akron, he was Senior Research Scientist and Advisory Software Developer with Rockwell Automation. Dr. Sastry is senior member of IEEE.