Operating Half-Duplex Diamond Networks with Two Interfering Relays

Sarthak Jain, Martina Cardone, Soheil Mohajer
University of Minnesota, Minneapolis, MN 55455, USA, Email: {jain0122, mcardone, soheil}@umn.edu

***Abstract*—This paper considers a diamond network with two interfering relays, where the source communicates with the destination via a layer of 2 half-duplex relays that can communicate with each other. The main focus is on characterizing the 3 relay receive/transmit configuration states (out of the 4 possible ones) that suffice to achieve the approximate capacity of the network. Towards this end, the binary linear deterministic approximation of the Gaussian noise channel is analyzed, and explicit scheduling and relaying schemes are presented. These schemes quantify the amount of information that each relay is responsible for sending to the destination, as well as the fraction of time each relay should receive and transmit.**

I. Introduction

Characterizing the Shannon capacity for wireless relay networks is a long-standing open problem. When the n relays operate in half-duplex, i.e., at each point in time each relay can either receive or transmit but not both simultaneously, a *scheduling* question naturally arises: How should the n relays be scheduled for reception and transmission so that rates close to the Shannon capacity can be guaranteed?

In principle, in an n-relay half-duplex network, there are 2^n possible relay receive/transmit configuration states. However, in [1], it was surprisingly shown that out of these 2^n possible states, only $n+1$ states suffice to achieve the network *approximate capacity*, i.e., an additive gap approximation of the Shannon capacity, where the gap is only a function of n. This result leads to practically relevant questions, such as characterizing the set of $n+1$ *critical* states for each network.

In this work we provide an answer to this question for a diamond network with $n = 2$ interfering relays, where the source communicates with the destination by hopping information through one layer of two half-duplex relays that can communicate with each other. In particular, we analyze the binary linear deterministic approximation of the Gaussian noise channel, and we characterize a set of $n+1 = 3$ states that suffice to achieve the approximate capacity. Specifically, we show that among the two states where both relays are receiving or both relays are transmitting, at most one is active. This result, as well as the devised scheduling and relaying schemes can be translated to obtain similar results for the practically relevant Gaussian noise channel. Although simplistic, the considered network model captures three inherent aspects of wireless communication, namely: (i) its broadcast nature; (ii) signal superposition (at the relays and destination); (iii) and signal interference (at the relays). This last aspect of signal interference – which is not captured by the diamond network with two *non-interfering* relays studied in [2] – is particularly important, and occurs when two signals from the source and a transmitting relay interfere at a receiving relay. The design of relaying and scheduling schemes that properly deal with this interference is therefore of critical importance. We here show that for $n = 2$, the source can "neutralize" such interference at the receiving relay by appropriately precoding its transmitted information with the one sent by the transmitting relay. Although as the number of relays increases, i.e., $n > 2$ it is not clear how to properly deal with such interference, we believe that the derived results can be used as building blocks to study larger networks with an arbitrary number of relays.

Related Work. There is a large body of literature [3]–[7] that has shown that the cut-set bound provides a constant additive gap approximation for the Shannon capacity for Gaussian relay networks. Such approximate capacity for Gaussian half-duplex relay networks can be found by solving a linear program that involves 2^n cut constraints and 2^n variables representing the receive/transmit configuration states. It turns out that at most $n+1$ of these variables need to be non-zero [1], which are associated with the *critical* states. The critical state variables are known for the following networks: (i) diamond networks with $n = 2$ relays [2], where the source communicates with the destination via two *non-interfering* relays; (ii) line networks [8], where the source communicates with the destination through a path of n relays; and (iii) diamond networks with an arbitrary number n of relays under certain network conditions expressed in [9]. These results imply that the approximate capacity for these networks can be computed in polynomial time in n. An additional class of networks for which the approximate capacity can be computed in polynomial time is given by layered networks where the number of relays per layer is at most logarithmic in the number of layers [10].

Paper Organization. Section II introduces the notation, describes the Gaussian half-duplex 2-relay network with interference and summarizes known capacity results. Section III presents the main result of the paper, the proof of which is in Section IV. Specifically, Section III characterizes a set of (at most) 3 network states that suffice to characterize the approximate capacity of the binary-valued linear deterministic approximation of the Gaussian noise channel.

II. Notation, System Model and Known Results

Notation: We denote the set of integers $\{1, \dots, n\}$ by $[n]$; $0_{i\times j}$ is the all-zero matrix of dimension $i \times j$ and I_n is the identity

This research was supported in part by NSF under Award #1907785.

matrix of dimension n; for a matrix M, $|M|$ is the determinant of M, and $M_{\mathcal{A},\mathcal{B}}$ is the submatrix of M obtained by retaining all the rows indexed by set $\mathcal{A}$ and all the columns indexed by set $\mathcal{B}$; $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are the floor and ceiling operations, respectively, and $[a]^+ = \max\{a, 0\}$; $f \circ g(\cdot)$ is the function composition of f and g.

The Gaussian half-duplex diamond 2-relay network with interference, at each time instant, is described as

$$Y_i = (1 - S_i)(h_{is}X_s + S_j h_{ij} X_j + Z_i), \tag{1a}$$

$$Y_d = \sum_{i=1}^{2} S_i h_{di} X_i + Z_d, \tag{1b}$$

where in (1a) $i \in [2]$ and $j \in [2] \setminus \{i\}$. Here, (i) S_i is a binary variable that indicates the state of relay $i \in [2]$, where $S_i = 0$ means that relay i is in receive mode, and $S_i = 1$ indicates that relay i is in transmit mode; (ii) X_i is the channel input of node i that satisfies the unit average power constraint $\mathbb{E}[|X_i|^2] \leq 1$ for $i \in \{s, 1, 2\}$; (iii) h_{ij} with $i \in \{1, 2, d\}$ and $j \in \{s, 1, 2\}$ with $i \neq j$ is the *time-invariant*[1] complex channel gain from node j to node i; note that $h_{ds} = 0$; (iv) $Z_i \sim \mathcal{CN}(0, 1)$ is the complex additive white Gaussian noise at node $i \in \{1, 2, d\}$; and finally, (v) Y_i is the received signal by node $i \in \{1, 2, d\}$.

The Shannon capacity C^G of the network in (1) is not known, but it can be approximated by C satisfying $|\mathsf{C}^G - \mathsf{C}| \leq \kappa$, where κ is a constant, independent of the channel gains. The *approximate capacity* C can be obtained by solving the following optimization problem [1],

$$\mathsf{C} = \max_{\boldsymbol{\lambda}} \min_{\Omega \subseteq [2]} \left\{ \sum_{\mathcal{S} \subseteq [2]} \lambda_{\mathcal{S}} \cdot f(\mathcal{S}, \Omega) \right\}, \tag{2a}$$

$$f(\mathcal{S}, \Omega) = I(X_s, X_{\Omega \cap \mathcal{S}}; Y_d, Y_{\Omega^c \cap \mathcal{S}^c} | X_{\Omega^c \cap \mathcal{S}}, \mathcal{S}), \tag{2b}$$

where: (i) $\mathcal{S} \subseteq [2]$ corresponds to the state of the network in which relay i is in transmit mode if and only if $i \in \mathcal{S}$; (ii) $\lambda_{\mathcal{S}} \geq 0$ denotes the fraction of time that the network operates in state $\mathcal{S}$; note that $\sum_{\mathcal{S} \subseteq [2]} \lambda_{\mathcal{S}} = 1$; (iii) $\boldsymbol{\lambda}$ is the vector obtained by stacking together $\lambda_{\mathcal{S}}$ for all $\mathcal{S} \subseteq [2]$, and is referred to as a *schedule* of the network; (iv) $\Omega \subseteq [2]$ is used to denote a partition of the relays in the 'side of the source', i.e., $\{s\} \cup \Omega$ is a cut of the network; similarly, $\Omega^c = [2] \setminus \Omega$ denotes a partition of the relays in the 'side of the destination'.

In this work, we are interested in characterizing a set of 3 *critical* receive/transmit configuration states (out of the 4 possible ones) that suffice to characterize the approximate capacity in (2) under different network conditions. Towards this end, our approach consists of the following two steps:

Step 1. We characterize a set of *critical* states under different network conditions for the binary-valued linear deterministic approximation of the Gaussian noise channel in (1). This model, which was introduced in [3], captures – in a simple deterministic way – the interaction between interfering signals and neglects the noise.

[1]The channel coefficients are assumed to remain constant for the entire transmission duration, and hence are known to all the nodes in the network.

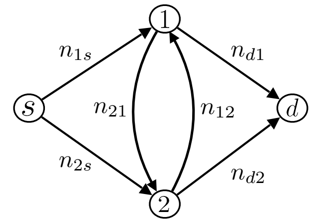


Fig. 1: System model.

Step 2. We translate the results obtained for the linear deterministic approximation channel into the original noisy Gaussian channel model in (1).

Because of the space limitations, we focus on Step 1, which is critical for the characterization of a set of *critical* states for the Gaussian noise network in (1). The binary-valued linear deterministic approximation of the Gaussian noise network in (1) has input-output relationship [3] (see also Fig. 1)

$$Y_i = (1 - S_i)(\mathbf{D}^{n-n_{is}} X_s + S_j \mathbf{D}^{n-n_{ij}} X_j), \tag{3a}$$

$$Y_d = \sum_{i=1}^{2} S_i \mathbf{D}^{n-n_{di}} X_i, \tag{3b}$$

$$\mathbf{D}^{n-m} = \left[\begin{array}{c|c} 0_{(n-m)\times m} & 0_{(n-m)\times(n-m)} \\ \hline I_m & 0_{m \times (n-m)} \end{array} \right], \tag{3c}$$

where in (3a) $i \in [2]$ and $j \in [2] \setminus \{i\}$ and where n_{ij}'s with $i \in \{1, 2, d\}$, $j \in \{s, 1, 2\}$, and $i \neq j$ are such that $n_{ij} = \left\lceil \log |h_{ij}|^2 \right\rceil^+$. In (3), the vectors X_s, X_i, Y_d, Y_i with $i \in [2]$ are of length $n = \max\{n_{1s}, n_{2s}, n_{12}, n_{21}, n_{d1}, n_{d2}\}$, $\mathbf{D}$ is the so-called $n \times n$ shift matrix, and $S_i, i \in [2]$ is the i-th relay binary-value state random variable. The approximate capacity in (2) for the linear deterministic channel in (3) – denoted as $\mathsf{C}^{\rm LD}$ – can be found as the solution of the following optimization problem

$$\mathsf{C}^{\rm LD} = \max_{\boldsymbol{\lambda}} \ t \tag{4}$$

s.t.

$$t \leq \lambda_{\varnothing} \max\{n_{1s}, n_{2s}\} + \lambda_{\{2\}} n_{1s} + \lambda_{\{1\}} n_{2s} + \lambda_{\{1,2\}} 0 \triangleq g_1,$$
$$t \leq \lambda_{\varnothing} n_{1s} + \lambda_{\{2\}} p + \lambda_{\{1\}} 0 + \lambda_{\{1,2\}} n_{d2} \triangleq g_2,$$
$$t \leq \lambda_{\varnothing} n_{2s} + \lambda_{\{2\}} 0 + \lambda_{\{1\}} q + \lambda_{\{1,2\}} n_{d1} \triangleq g_3,$$
$$t \leq \lambda_{\varnothing} 0 + \lambda_{\{2\}} n_{d2} + \lambda_{\{1\}} n_{d1} + \lambda_{\{1,2\}} \max\{n_{d1}, n_{d2}\} \triangleq g_4,$$
$$g_5 \triangleq \left(\sum_{\mathcal{S} \subseteq [2]} \lambda_{\mathcal{S}} \right) - 1 \leq 0, \ \lambda_{\mathcal{S}} \geq 0, \quad \forall \mathcal{S} \subseteq [2],$$

where

$$p = \max\{n_{1s} + n_{d2}, n_{12}\}, \ q = \max\{n_{2s} + n_{d1}, n_{21}\}$$

are obtained by finding the rank of the incidence matrices corresponding to the linear deterministic model [3].

III. Set of Critical States

In this section, the main result of this paper is presented in Theorem 1. This result completely characterizes which (at most) $n + 1 = 3$ states achieve the approximate capacity $\mathsf{C}^{\rm LD}$ in (4) under different network conditions.

Theorem 1. *Let*

$$P = \begin{bmatrix} -n_{d1} & p - n_{d2} \\ n_{2s} & -p + n_{1s} \end{bmatrix}, \qquad Q = \begin{bmatrix} -n_{1s} & q - n_{2s} \\ n_{d2} & -q + n_{d1} \end{bmatrix}.$$

An optimal collection $\mathbb{S}$ of at most $n+1=3$ states, which suffice to achieve the approximate capacity C^{LD} in (4)*, is given as follows.*

(1) Case 1: $n_{1s} \le n_{2s}$ and $n_{d1} \le n_{d2}$

$$\begin{cases} \textit{(i)}\ \mathbb{S} = \{\varnothing, \{1\}, \{2\}\}, & \textit{if } |P| \ge 0, \\ \textit{(ii)}\ \mathbb{S} = \{\{1\}, \{2\}, \{1,2\}\}, & \textit{if } |Q| \ge 0, \\ \textit{(iii)}\ \mathbb{S} = \{\{1\}, \{2\}\}, & \textit{if } \max\{|P|, |Q|\} < 0. \end{cases}$$

(2) Case 2: $n_{1s} \le n_{2s}$ and $n_{d1} \ge n_{d2}$

$$\begin{cases} \textit{(i)}\ \mathbb{S} = \{\varnothing, \{1\}, \{2\}\}, & \textit{if } |P| \ge 0, \\ \textit{(ii)}\ \mathbb{S} = \{\{1\}, \{2\}, \{1,2\}\}, & \textit{if } |P| \le 0. \end{cases}$$

(3) Case 3: $n_{1s} \ge n_{2s}$ and $n_{d1} \le n_{d2}$

$$\begin{cases} \textit{(i)}\ \mathbb{S} = \{\varnothing, \{1\}, \{2\}\}, & \textit{if } -|Q| \ge 0, \\ \textit{(ii)}\ \mathbb{S} = \{\{1\}, \{2\}, \{1,2\}\}, & \textit{if } -|Q| \le 0. \end{cases}$$

(4) Case 4: $n_{1s} \ge n_{2s}$ and $n_{d1} \ge n_{d2}$

$$\begin{cases} \textit{(i)}\ \mathbb{S} = \{\varnothing, \{1\}, \{2\}\}, & \textit{if } -|Q| \ge 0, \\ \textit{(ii)}\ \mathbb{S} = \{\{1\}, \{2\}, \{1,2\}\}, & \textit{if } -|P| \ge 0, \\ \textit{(iii)}\ \mathbb{S} = \{\{1\}, \{2\}\}, & \textit{if } \max\{-|P|, -|Q|\} < 0. \end{cases}$$

Remark 1. *The states $\{\{1\},\{2\}\}$ are active in all the cases. The proposed (optimal) relaying scheme only uses one of the links between the relays to transmit unique information. For instance, for all the cases where the condition only involves $|P|$ (respectively, $|Q|$), only the link from relay 2 to relay 1 (respectively, from relay 1 to relay 2) is used to transmit unique information. In these cases, we transmit p (respectively, q) bits of unique information through the cut $\Omega = \{2\}$ (respectively, $\Omega = \{1\}$) in state $\mathcal{S} = \{2\}$ (respectively, $\mathcal{S} = \{1\}$).*

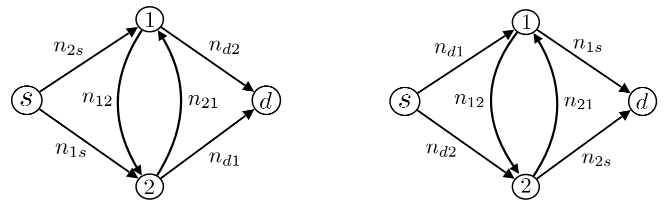
Remark 2. *If $p = n_{1s} + n_{d2}$ and $q = n_{2s} + n_{d1}$, then the existence of cross links n_{12} and n_{21} does not contribute to the approximate capacity, and the result in Theorem 1 reduces to that for a 2-relay diamond network in [2], namely if $n_{1s}n_{2s} \le n_{d1}n_{d2}$ then $\mathbb{S} = \{\varnothing, \{1\}, \{2\}\}$, otherwise $\mathbb{S} = \{\{1\}, \{2\}, \{1,2\}\}$. Moreover, even though communication schemes might be different, the approximate capacities of the two networks, i.e., with interfering relays (considered here) and with non-interfering relays (studied in [2]) are identical.*

In the rest of this section, we briefly present two results that will be used as building blocks for the proof of Theorem 1.

A. Karush-Kuhn-Tucker (KKT) Conditions

Define $\boldsymbol{\mu} = (\mu_1, \dots, \mu_5)$ and $\boldsymbol{\sigma} = (\sigma_\varnothing, \sigma_{\{1\}}, \sigma_{\{2\}}, \sigma_{\{1,2\}})$. The Lagrangian of the optimization problem in (4) is given by

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}, t) = -t + \sum_{i=1}^{4} \mu_i(t - g_i) + \mu_5(g_5) - \sum_{\mathcal{S} \subseteq [2]} \sigma_{\mathcal{S}} \lambda_{\mathcal{S}}. \tag{5}$$



(a) $\mathsf{f_L}(\mathcal{N})$. (b) $\mathsf{f_M}(\mathcal{N})$.

Fig. 2: Network tranformations.

The KKT conditions to verify the optimality of a given solution $(\boldsymbol{\lambda}^\star, t^\star)$ for the linear program in (4) are given as:

Primal Feasibility. We show the feasibility of a solution by presenting a transmission scheme and assessing its rate performance as well as deriving a network schedule for it.

Stationarity. We show the existence of a solution $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ such that $\frac{\partial}{\partial t}\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}^\star, t^\star) = 0$ and $\frac{\partial}{\partial \lambda_{\mathcal{S}}}\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}^\star, t^\star) = 0$ for every $\mathcal{S} \subseteq [2]$.

Complementary Slackness. We show that $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ that satisfy the stationarity condition are such that $\mu_i(t^\star - g_i^\star) = 0$ for every $i \in [4]$, $\mu_5 g_5^\star = 0$, $\sigma_{\mathcal{S}}\lambda_{\mathcal{S}}^\star = 0$, for every $\mathcal{S} \subseteq [2]$. Here, $g_i^\star$ is g_i in (4) evaluated at $\boldsymbol{\lambda}^\star$, for $i \in [5]$.

Dual Feasibility. We show that $\mu_i \ge 0$ for $i \in [5]$ and $\sigma_{\mathcal{S}} \ge 0$ for every $\mathcal{S} \subseteq [2]$.

B. Network Transformations

In the proof of Theorem 1, we will leverage two transformations of the network $\mathcal{N}$ in Fig. 1, as described below.

Relabelling relay nodes. This network transformation, denoted as $\mathsf{f_L}(\cdot)$, relabels relays 1 and 2, to obtain $\mathsf{f_L}(\mathcal{N})$ as shown in Fig. 2(a). With a slight abuse of notation, we apply the transformation to the individual network parameters, to obtain $\mathsf{f_L}(n_{1s}) = n_{2s}$, $\mathsf{f_L}(n_{2s}) = n_{1s}$, $\mathsf{f_L}(n_{12}) = n_{21}$, $\mathsf{f_L}(n_{21}) = n_{12}$, $\mathsf{f_L}(n_{d1}) = n_{d2}$, and $\mathsf{f_L}(n_{d2}) = n_{d1}$. This implies that $\mathsf{f_L}(p) = q$ and $\mathsf{f_L}(q) = p$. Moreover, the state $\mathcal{S} \subseteq [2]$ in the network $\mathcal{N}$ corresponds to the state $\mathsf{f_L}(\mathcal{S})$ in $\mathsf{f_L}(\mathcal{N})$, where $\mathsf{f_L}(\varnothing) = \varnothing$, $\mathsf{f_L}(\{1\}) = \{2\}$, $\mathsf{f_L}(\{2\}) = \{1\}$, and $\mathsf{f_L}(\{1,2\}) = \{1,2\}$.

Mirroring. This network transformation, denoted as $\mathsf{f_M}(\cdot)$, flips the direction of each link in the network $\mathcal{N}$ (thereby also switching the roles of the source s and the destination d). The resulting network $\mathsf{f_M}(\mathcal{N})$ is shown in Fig. 2(b), whose channel parameters are given by $\mathsf{f_M}(n_{1s}) = n_{d1}$, $\mathsf{f_M}(n_{2s}) = n_{d2}$, $\mathsf{f_M}(n_{12}) = n_{21}$, $\mathsf{f_M}(n_{21}) = n_{12}$, $\mathsf{f_M}(n_{d1}) = n_{1s}$, $\mathsf{f_M}(n_{d2}) = n_{2s}$. Consequently, we have $\mathsf{f_M}(p) = q$ and $\mathsf{f_M}(q) = p$. Moreover, the transformation $\mathsf{f_M}(\cdot)$ applied to the state $\mathcal{S} \subseteq [2]$ in $\mathcal{N}$ is given as $\mathsf{f_M}(\varnothing) = \{1,2\}$, $\mathsf{f_M}(\{1\}) = \{2\}$, $\mathsf{f_M}(\{2\}) = \{1\}$, and $\mathsf{f_M}(\{1,2\}) = \varnothing$.

It is not difficult to verify that the approximate capacity C^{LD} obtained by solving (4) is invariant to applying any combination of $\mathsf{f_L}(\cdot)$ and $\mathsf{f_M}(\cdot)$ to the network $\mathcal{N}$. Moreover, any transmission scheme for $\mathcal{N}$ can be translated to a scheme for a network obtained by these transformations.

IV. Proof of Theorem 1

In this section, we present the proof of Theorem 1, by showing the achievability and optimality for each separate case.

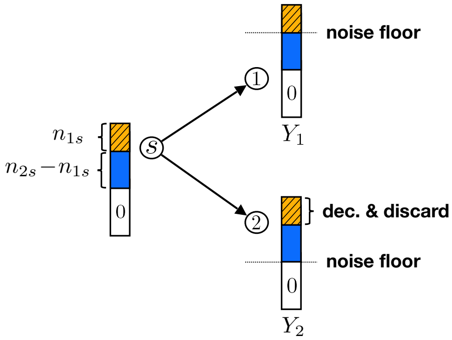


Fig. 3: Scheme for the state $\mathcal{S} = \varnothing$ in Cases 1(i) and 2(i).

■ **Case 1(i)** and **Case 2(i).** *Achievability.* We show the achievability part for these cases, by designing a transmission scheme and assessing its performance. We consider a block Markov coding scheme, where symbols decoded by the relays in block B will be transmitted in block $B+1$. For the proposed scheme, we characterize the number of *unique* bits that each relay is responsible to receive and transmit in each state in $\mathbb{S} = \{\varnothing, \{1\}, \{2\}\}$. This, together with flow preservation [9] constraints at each relay, leads to closed form expressions for $\{\lambda_{\mathcal{S}} : \mathcal{S} \in \mathbb{S}\}$ and the rate t. Next, we describe the operation of the network in each state $\mathcal{S} \in \mathbb{S}$.

$\underline{\mathcal{S} = \varnothing}$. The source s broadcasts a total of n_{2s} bits. Relay 1 receives the top n_{1s} bits. Relay 2 decodes all the n_{2s} bits, but discards the top n_{1s} bits and keeps the remaining $n_{2s} - n_{1s}$ bits for transmission in the future block (see Fig. 3).

$\underline{\mathcal{S} = \{2\}}$. In this state, the source transmits n_{1s} bits and Relay 2 broadcasts a total of $p - n_{1s}$ bits. The topmost n_{d2} bits of X_2 are intended for the destination d, and the remaining $p - n_{1s} - n_{d2}$ bits will be decoded by Relay 1, for retransmission in a future block. We can distinguish the following two sub-cases: (a) if $p = n_{12}$, then the bits from the source and Relay 2 reach Relay 1 at non-overlapping levels, and can be decoded by Relay 1, as shown in Fig. 4; and (b) if $p = n_{1s}+n_{d2}$, then the received sets of bits overlap at Relay 1 in $y = \min\{n_{1s}, n_{12}, n_{d2}, p-n_{12}\}$ levels. However, the source is aware of the content of these y bits sent by Relay 2, and can *neutralize* them at Relay 1 by performing a precoding before transmitting its n_{1s} bits [11]. This is illustrated in Fig. 5.

$\underline{\mathcal{S} = \{1\}}$. The source sends n_{2s} bits intended for Relay 2 and Relay 1 broadcasts n_{d1} bits intended for the destination. There is no flow of information from Relay 1 to Relay 2, and any potential interference caused by Relay 1 at Relay 2 can be neutralized by a precoding at the source.

Note that the number of bits sent by each relay cannot exceed the number of received ones. At an optimal point, however, we have preservation of information, i.e., the number of received and decoded bits (excluding the discarded ones) at each relay is *equal* to the number of bits transmitted by that relay. This and the fact that $\lambda_\varnothing + \lambda_{\{1\}} + \lambda_{\{2\}} = 1$ lead to

$$\underbrace{\begin{bmatrix} n_{1s} & -n_{d1} & (p-n_{d2}) \\ (n_{2s}-n_{1s}) & n_{2s} & -(p-n_{1s}) \\ 1 & 1 & 1 \end{bmatrix}}_{H} \begin{bmatrix} \lambda_\varnothing \\ \lambda_{\{1\}} \\ \lambda_{\{2\}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (6)$$

Therefore, we have

$$\lambda_\varnothing = \frac{|H_{\{1,2\},\{2,3\}}|}{|H|} = \frac{|P|}{|H|}, \qquad \lambda_{\{1\}} = \frac{-|H_{\{1,2\},\{1,3\}}|}{|H|}, \qquad \lambda_{\{2\}} = \frac{|H_{\{1,2\}\{1,2\}}|}{|H|}. \quad (7)$$

It is not difficult to verify that the conditions of Cases 1(i) and 2(i) imply that $\lambda_\varnothing$, $\lambda_{\{1\}}$ and $\lambda_{\{2\}}$ in (7) are non-negative. The total rate t achieved by this scheme is hence given by

$$t = \lambda_{\{1\}} n_{d1} + \lambda_{\{2\}} n_{d2} = \lambda_\varnothing n_{2s} + \lambda_{\{2\}} n_{1s} + \lambda_{\{1\}} n_{2s}. \quad (8)$$

Note that among these bits, $(\lambda_\varnothing + \lambda_{\{2\}}) n_{1s}$ bits go through the path $s \to 1 \to d$, $\lambda_{\{2\}} n_{d2}$ bits traverse the path $s \to 2 \to d$, and $\lambda_{\{2\}}[n_{12} - n_{1s} - n_{d2}]^+$ bits are delivered to the destination via the path $s \to 2 \to 1 \to d$.

Optimality/Converse. In order to prove the optimality of the solution given by (7)-(8), we show that the KKT conditions are satisfied. Setting $\mu_3 = \sigma_\varnothing = \sigma_{\{1\}} = \sigma_{\{2\}} = 0$, and equating the partial derivatives of the Lagrangian in (5) to zero, we obtain a system of five linear equations in five variables, namely μ_1, μ_2, μ_4, μ_5, and $\sigma_{\{1,2\}}$. It is not difficult to verify that the solution of this linear system satisfies the stationarity, complementary slackness and dual feasibility constraints.

■ **Case 1(ii).** Applying the mirroring transformation on the network $\mathcal{N}$, we get $\mathsf{f}_\mathsf{M}(\mathcal{N})$, for which we have

$$|\mathsf{f}_\mathsf{M}(P)| = \begin{vmatrix} -\mathsf{f}_\mathsf{M}(n_{d1}) & \mathsf{f}_\mathsf{M}(p) - \mathsf{f}_\mathsf{M}(n_{d2}) \\ \mathsf{f}_\mathsf{M}(n_{2s}) & -\mathsf{f}_\mathsf{M}(p) + \mathsf{f}_\mathsf{M}(n_{1s}) \end{vmatrix} = |Q| \geq 0.$$

Therefore, $\mathsf{f}_\mathsf{M}(\mathcal{N})$ satisfies the conditions of Case 1(i), where the set of critical states is given by $\mathbb{S} = \{\varnothing, \{1\}, \{2\}\}$. Hence, as described in Section III, the set of critical states of $\mathcal{N}$ is given by $\mathsf{f}_\mathsf{M}^{-1}(\mathbb{S}) = \{\{1,2\}, \{2\}, \{1\}\}$. The approximate capacity and the relaying scheme can be immediately obtained by translating those of Case 1(i).

■ **Case 1(iii).** *Achievability.* We start by describing the network operation for each state $\mathcal{S} \in \mathbb{S} = \{\{1\}, \{2\}\}$.

$\underline{\mathcal{S} = \{i\}, i \in [2]}$. The source s broadcasts n_{js} bits that are decoded by Relay $j \in [2] \setminus \{i\}$. In the same time, Relay i broadcasts a total of $n_{di} + x_i$ bits, among which the topmost n_{di} bits are intended for d (and will be discarded by Relay j) and x_i bits should be decoded at Relay j. Note that Relay j can decode n_{js} bits from s and x_i bits from Relay i only if

$$0 \leq x_i \leq n_{ji} - n_{di} - n_{js}, \qquad i \neq j. \quad (9)$$

Similar to Case 1(i), a potential interference at Relay j can be neutralized by precoding at the source.

Using the information preservation principle as we did for Case 1(i) and the fact that $\lambda_{\{1\}} + \lambda_{\{2\}} = 1$, we can write

$$\begin{bmatrix} -(n_{d1}+x_1) & n_{1s}+x_2 \\ n_{2s}+x_1 & -(n_{d2}+x_2) \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_{\{1\}} \\ \lambda_{\{2\}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (10)$$

This implies that

$$\lambda_{\{1\}} = \frac{n_{d2} - n_{1s}}{(n_{2s} - n_{1s}) + (n_{d2} - n_{d1})} \geq 0, \qquad \lambda_{\{2\}} = \frac{n_{2s} - n_{d1}}{(n_{2s} - n_{1s}) + (n_{d2} - n_{d1})} \geq 0, \quad (11)$$

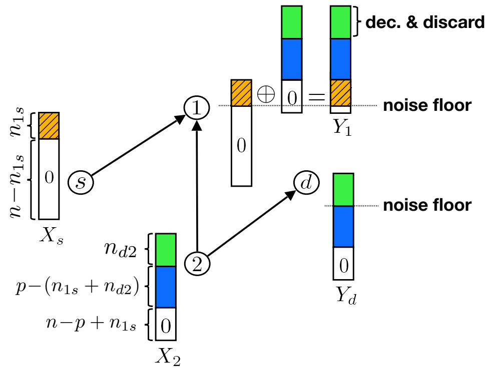


Fig. 4: Scheme for the state $\mathcal{S} = \{2\}$ in Cases 1(i) and 2(i): no precoding at source is needed.

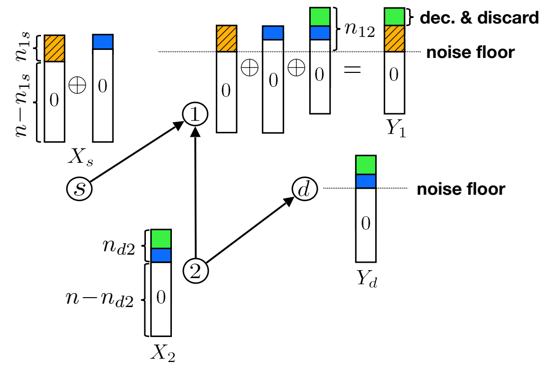


Fig. 5: Scheme for the state $\mathcal{S} = \{2\}$ in Cases 1(i) and 2(i): precoding at source is needed.

where the inequalities hold since $|P| < 0$ and $|Q| < 0$ in this case. The total achievable rate t is then given by

$$t = \lambda_{\{1\}} n_{d1} + \lambda_{\{2\}} n_{d2} = \lambda_{\{1\}} n_{2s} + \lambda_{\{2\}} n_{1s}. \qquad (12)$$

To conclude the proof of primal feasibility, we need to show that there exist some x_1 and x_2 satisfying (9).

Using the conditions of Case 1(iii), it is easy to verify that

$$x_1 = \frac{[n_{1s}n_{2s} - n_{d1}n_{d2}]^+}{n_{d2} - n_{1s}} \quad x_2 = \frac{[n_{d1}n_{d2} - n_{1s}n_{2s}]^+}{n_{2s} - n_{d1}} \qquad (13)$$

are non-negative and satisfy (9). Moreover, it can be readily seen from (13) that we always have either $x_1 = 0$ or $x_2 = 0$.

Finally, we note that since x_i is the number of bits sent from Relay i to Relay j, it should be an integer, which is not necessarily the case in (13). However, since $n_{ji} - n_{di} - n_{js}$ is integer, if x_i satisfies (9), so does $\lceil x_i \rceil$. Therefore, in order to send x_i bits from Relay i to Relay j, we time share between sending $\lceil x_i \rceil$ bits and $\lfloor x_i \rfloor$ bits.

Optimality/Converse. In order to show the optimality of the solution given by (11)-(12), we prove that it satisfies the KKT conditions. Setting $\mu_2 = \mu_3 = \sigma_{\{1\}} = \sigma_{\{2\}} = 0$ and equating the partial derivatives of the Lagrangian in (5) to zero, we obtain a system of five linear equations in five variables, namely μ_1, μ_4, μ_5, $\sigma_{\varnothing}$, $\sigma_{\{1,2\}}$. It is not difficult to show that the solution to this system satisfies the stationarity, complementary slackness and dual feasibility constraints.

■ **Case 2(ii).** Applying the transformation $\mathsf{f_L} \circ \mathsf{f_M}(\cdot)$ to the network $\mathcal{N}$, we get $\mathsf{f_L} \circ \mathsf{f_M}(\mathcal{N})$, which still lies in Case 2. Moreover, we have that $|\mathsf{f_L} \circ \mathsf{f_M}(P)|$ can be evaluated as

$$\begin{vmatrix} -\mathsf{f_L} \circ \mathsf{f_M}(n_{d1}) & \mathsf{f_L} \circ \mathsf{f_M}(p) - \mathsf{f_L} \circ \mathsf{f_M}(n_{d2}) \\ \mathsf{f_L} \circ \mathsf{f_M}(n_{2s}) & -\mathsf{f_L} \circ \mathsf{f_M}(p) + \mathsf{f_L} \circ \mathsf{f_M}(n_{1s}) \end{vmatrix} = -|P| \geq 0.$$

Hence, $\mathsf{f_L} \circ \mathsf{f_M}(\mathcal{N})$ satisfies the conditions of Case 2(i) with critical states $\mathbb{S} = \{\varnothing, \{1\}, \{2\}\}$. Therefore, the set of critical states of $\mathcal{N}$ is given by $\mathsf{f_M}^{-1} \circ \mathsf{f_L}^{-1}(\mathbb{S}) = \{\{1,2\}, \{1\}, \{2\}\}$. Finally, by applying the mirroring followed by the relabelling transformations, the approximate capacity and the relaying scheme for Case 2(i) can be translated to Case 2(ii).

■ **Case 3.** This case can be converted to Case 2 using the transformation $\mathsf{f_L}(\cdot)$. Note that we have $|\mathsf{f_L}(P)| = -|Q|$, and hence the set of critical states is $\mathsf{f_L}^{-1}(\{\varnothing, \{1\}, \{2\}\}) = \{\varnothing, \{2\}, \{1\}\}$ when $-|Q| \geq 0$, and it is $\mathsf{f_L}^{-1}(\{\{1\}, \{2\}, \{1,2\}\}) = \{\{2\}, \{1\}, \{1,2\}\}$ when $-|Q| \leq 0$. The approximate capacity and the relaying scheme can be found by applying the described relabelling transformation on those used for Case 2.

■ **Case 4.** Similarly, a network in Case 4 can be converted to one in Case 1 using the relabelling transformation $\mathsf{f_L}(\cdot)$. Noting that $|\mathsf{f_L}(P)| = -|Q|$ and $|\mathsf{f_L}(Q)| = -|P|$, the set of critical states can be found as $\mathsf{f_L}^{-1}(\{\varnothing, \{1\}, \{2\}\}) = \{\varnothing, \{2\}, \{1\}\}$ for $-|Q| \geq 0$, and as $\mathsf{f_L}^{-1}(\{\{1\}, \{2\}, \{1,2\}\}) = \{\{2\}, \{1\}, \{1,2\}\}$ when $-|P| \geq 0$. Finally, for $-|P| < 0$ and $-|Q| < 0$ the critical states are given by $\mathsf{f_L}^{-1}(\{\{1\}, \{2\}\}) = \{\{2\}, \{1\}\}$. Applying $\mathsf{f_L}(\cdot)$ on the approximate capacity and the relaying scheme for Case 1, we can find those for Case 4.

References

[1] M. Cardone, D. Tuninetti, and R. Knopp, “On the optimality of simple schedules for networks with multiple half-duplex relays,” *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4120–4134, July 2016.

[2] H. Bagheri, A. S. Motahari, and A. K. Khandani, “On the capacity of the half-duplex diamond channel,” in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 649–653.

[3] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[4] A. Özgür and S. N. Diggavi, “Approximately achieving Gaussian relay network capacity with lattice-based QMF codes,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8275–8294, December 2013.

[5] S. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, “Noisy network coding,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132 –3152, 2011.

[6] S. H. Lim, K. T. Kim, and Y. H. Kim, “Distributed decode-forward for multicast,” in *IEEE International Symposium on Information Theory (ISIT)*, June 2014, pp. 636–640.

[7] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim, “Gaussian half-duplex relay networks: improved constant gap and connections with the assignment problem,” *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3559 – 3575, June 2014.

[8] Y. H. Ezzeldin, M. Cardone, C. Fragouli, and D. Tuninetti, “Efficiently finding simple schedules in Gaussian half-duplex relay line networks,” in *IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 471–475.

[9] S. Jain, M. Elyasi, M. Cardone, and S. Mohajer, “On simple scheduling in half-duplex relay diamond networks,” in *IEEE International Symposium on Information Theory (ISIT)*, July 2019.

[10] R. H. Etkin, F. Parvaresh, I. Shomorony, and A. S. Avestimehr, “Computing half-duplex schedules in Gaussian relay networks via min-cut approximations,” *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7204–7220, November 2014.

[11] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. Tse, “Transmission techniques for relay-interference networks,” in *46th Annual Allerton Conf. on Commun., Control, and Comp.* IEEE, 2008, pp. 467–474.