On the Fraction of Capacity One Relay can Achieve in Gaussian Half-Duplex Diamond Networks

Sarthak Jain, Soheil Mohajer, Martina Cardone
University of Minnesota, Minneapolis, MN 55404
Email: {jain0122, soheil, mcardone}@umn.edu

***Abstract*—This paper considers the Gaussian half-duplex diamond n-relay network, which consists of a broadcast hop between the source and n relays, and of a multiple access hop between the relays and the destination. The n relays do not communicate with each other and operate in half-duplex mode. The main focus of the paper is on answering the following question: What fraction of the *approximate capacity* of the entire network can be retained by only operating the highest-performing single relay? It is shown that a fraction $f = 1/(2+2\cos(\frac{2\pi}{n+2}))$ of the *approximate capacity* of the entire network can always be guaranteed. This fraction is also shown to be tight, that is, there exist Gaussian half-duplex diamond n-relay networks for which exactly an f fraction of the *approximate capacity* of the entire network can be achieved by using only the highest-performing relay.**

I. Introduction

Several practical challenges arise in wireless networks with relays. Relays must synchronize for reception and transmission, which might result in a highly-complex process when low-cost communication modules are needed. Operating all the relays might also bring to a severe power consumption, which cannot be sustained. With the goal of offering a suitable solution for these considerations, the authors of [1] introduced the *network simplification* problem, which seeks to provide fundamental guarantees on the fraction of the entire network capacity that can be retained when only a subset of the relays is operated.

In this paper, we investigate the network simplification problem in Gaussian half-duplex *diamond* n-relay networks, which consist of a broadcast hop between the source and the n relays, and of a multiple access hop between the n relays and the destination. There is no communication among the n relays. Moreover, the relays operate in half-duplex mode, i.e., at any point of time each relay can either receive or transmit, but not both simultaneously. We ask the following question: What fraction of the entire network *approximate capacity* can be retained by only operating the highest-performing relay? In particular, *approximate capacity* refers to a quantity that approximates the Shannon capacity within an additive gap which may depend on n, but not on the channel gains [2]–[6].

We show that, in any Gaussian half-duplex diamond n-relay network, there is always a single relay using which we can achieve at least $f = 1/(2 + 2\cos(\frac{2\pi}{n+2}))$ of the entire network approximate capacity. This fraction is independent of the values of the channel coefficients, and decreases as n increases. We also present networks for which this bound is *tight*, i.e., networks for which the approximate capacity of the highest-performing relay is exactly equal to an f fraction of the approximate capacity of the entire network.

The work of the authors was supported in part by the U.S. National Science Foundation under Grant CCF-1907785.

Related Work. The network simplification problem was pioneered in [1] in the context of Gaussian *full-duplex* diamond n-relay networks. The authors showed that there always exists a sub-network of k relays which has an approximate capacity of at least $k/(k + 1)$ fraction of the approximate capacity of the entire network. Recently, in [7] the authors analyzed the guarantee of selecting the highest-performing path in Gaussian full-duplex n-relay networks with arbitrary layered topology.

Very few results exist on the network simplification problem in half-duplex networks. In [8], the authors showed that any Gaussian half-duplex diamond n-relay network has a 2-relay sub-network that can achieve at least $1/2$ of the entire network approximate capacity. In [9], it was shown that operating $n-1$ relays always guarantees that an $(n-1)/n$ fraction of the entire network approximate capacity can be retained in any Gaussian half-duplex diamond n-relay network. For large networks (i.e., $n \gg 1$), the authors in [9] also showed that sub-networks with $k = 1$ and $k = 2$ relays can achieve $1/4$ and $1/2$ of the approximate capacity of the entire network, respectively. These guarantees are tight, and point out to a fundamental difference between half-duplex and full-duplex [1]: in the former case the fraction guarantee decreases as n increases, whereas in the latter case the fraction guarantee is independent of n.

In this work, we provide a complete answer to a question that was left open in [9]: What is the fundamental performance guarantee (in terms of ratio) when only the highest-performing $k = 1$ relay sub-network is operated, as a function of n?

Paper Organization. Section II describes the Gaussian half-duplex diamond n-relay network, and defines its approximate capacity. Section III formulates the problem and presents the main result of the paper. Section IV provides an overview of the proof of the main result.

II. Network Model

Notation. For integers $n_1, n_2 \in \mathbb{Z}$ with $n_2 \geq n_1$, we have $[n_1 : n_2] = \{\, x \mid x \in \mathbb{Z} \text{ and } n_1 \leq x \leq n_2 \}$. For a complex number $a \in \mathbb{C}$, $|a|$ denotes the magnitude of a. $\mathbb{E}[\cdot]$ denotes the expected value. Finally, $\lfloor x \rfloor$ is the floor of x.

The Gaussian half-duplex diamond n-relay network $\mathcal{N}$ consists of two hops, as shown in Fig. 1: the broadcast

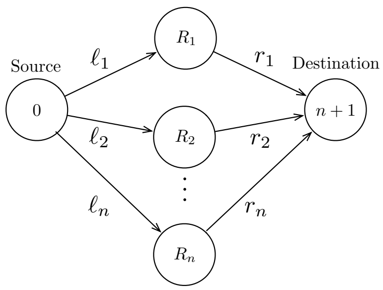


Fig. 1: Gaussian half-duplex diamond network with n relays.

hop between the source (node 0) and the set of n relays; and the multiple access hop between the n relays and the destination (node $n+1$). The n relays are non-interfering, and the source can communicate to the destination only by hopping information through them. Relays operate in half-duplex, and hence can either receive or transmit at any given time. The input/output relationship for $\mathcal{N}$ at time instance t is defined as

$$Y_{i,t} = (1 - S_{i,t})(h_{si}X_{0,t} + Z_{i,t}), \quad \forall i \in [1:n], \quad \text{(1a)}$$

$$Y_{n+1,t} = \sum_{i=1}^{n} S_{i,t}h_{id}X_{i,t} + Z_{n+1,t}, \quad \text{(1b)}$$

where: (i) $S_{i,t}$ is a binary variable indicating the state of relay R_i at time t; specifically, $S_{i,t} = 0$ and $S_{i,t} = 1$ mean that, at time t, R_i is in receive mode and transmit mode, respectively; (ii) $X_{i,t},\ \forall i \in [0:n]$ is the channel input of node i at time t that satisfies $\mathbb{E}[|X_{i,t}|^2] \leq 1$; (iii) h_{si} and h_{id} are the *time-invariant* complex channel gains from the source to R_i and from R_i to the destination, respectively; (iv) $Z_{i,t}\ , i \in [1:n{+}1]$ is the complex additive white Gaussian noise at node i at time t; noises are i.i.d. as $\mathcal{CN}(0,1)$; and (v) $Y_{i,t},\ \forall i \in [1:n+1]$ is the received signal by node i at time t.

The Shannon capacity $\mathsf{C}_n^G(\mathcal{N})$ for the network $\mathcal{N}$ in (1) is unknown in general. However, as shown in [3], [4], [6], it can be approximated to within a *constant additive gap* as

$$\left|\mathsf{C}_n^G(\mathcal{N}) - \mathsf{C}_n(\mathcal{N})\right| \leq \kappa_n,$$

where $\kappa_n = O(n)$ only depends on n and is independent of the channel coefficients and where $\mathsf{C}_n(\mathcal{N})$ is referred to as *approximate capacity* and is formally defined in Definition 1.

Definition 1. *The approximate capacity of the Gaussian half-duplex diamond n-relay network in (1) is defined as*

$$\begin{aligned} &\mathsf{C}_n(\mathcal{N}) = \max_{\boldsymbol{\lambda}}\ t \\ &\textit{s.t.}\ \ t \leq \sum_{\mathcal{S}\subseteq[1:n]} \lambda_{\mathcal{S}} \left(\max_{i\in\mathcal{S}^c\cap\Omega^c} \ell_i + \max_{i\in\mathcal{S}\cap\Omega} r_i \right),\ \forall \Omega \subseteq [1:n], \\ &\qquad \sum_{\mathcal{S}\subseteq[1:n]} \lambda_{\mathcal{S}} = 1,\ \lambda_{\mathcal{S}} \geq 0, \quad \forall \mathcal{S} \subseteq [1:n], \end{aligned} \quad \text{(2)}$$

where, $\forall i \in [1:n]$,

$$\ell_i = \log(1 + |h_{si}|^2), \ \ r_i = \log(1 + |h_{id}|^2).$$

In (2), we have that: (i) $\mathcal{S} \subseteq [1:n]$ is the network state in which relays $R_i, i \in \mathcal{S}$, are in transmitting mode, while the other relays are in receiving mode; (ii) $\lambda_{\mathcal{S}}$ is the time fraction that the network operates in state $\mathcal{S}$; (iii) $\boldsymbol{\lambda}$ is the vector obtained by stacking together $\lambda_{\mathcal{S}}, \forall \mathcal{S} \subseteq [1:n]$, and is referred to as a network *schedule*; (iv) $\Omega \subseteq [1:n]$ denotes a partition of the relays in the 'side of the source'; similarly, $\Omega^c = [1:n] \backslash \Omega$ is a partition of the relays in the 'side of the destination'; for a relay $R_i, i \in \Omega$, to contribute to the information flow we need $i \in \mathcal{S}$; similarly, for a relay $R_i, i \in \Omega^c$, to contribute to the information flow we need $i \in \mathcal{S}^c$.

III. Problem Statement and Main Result

In this section, we characterize fundamental guarantees on the approximate capacity of the *best* single relay sub-network, as a fraction of the approximate capacity of the entire network.

Note that $\mathsf{C}_n(\mathcal{N})$ in (2) is a function of the network $\mathcal{N}$ only through the link capacities $(\ell_i, r_i), i \in [1:n]$. Thus, with a slight notation abuse, we let $\mathcal{N} = \{(\ell_i, r_i), i \in [1:n]\}$. We also use $\mathcal{N}_i = \{(\ell_i, r_i)\}$ to denote a network with the source, relay R_i and destination. By solving (2) for the single $R_i, i \in [1:n]$, we obtain that the approximate capacity of $\mathcal{N}_i$ is

$$\mathsf{C}_1(\mathcal{N}_i) = \frac{\ell_i r_i}{\ell_i + r_i}.$$

We also define the *best single relay approximate capacity* of the network as the maximum approximate capacity among the single relay sub-networks, that is,

$$\mathsf{C}_1(\mathcal{N}) = \max_{i\in[1:n]} \mathsf{C}_1(\mathcal{N}_i).$$

Our goal is to find universal bounds on $\mathsf{C}_1(\mathcal{N})/\mathsf{C}_n(\mathcal{N})$. In particular, our main result is provided in the next theorem, whose sketch of the proof is provided in Section IV.

Theorem 1. *For any Gaussian half-duplex diamond n-relay network $\mathcal{N}$ with approximate capacity $\mathsf{C}_n(\mathcal{N})$, the best relay has an approximate capacity $\mathsf{C}_1(\mathcal{N})$ such that*

$$\frac{\mathsf{C}_1(\mathcal{N})}{\mathsf{C}_n(\mathcal{N})} \geq \frac{1}{2 + 2\cos\left(\frac{2\pi}{n+2}\right)}. \quad \text{(3)}$$

Moreover, the bound in (3) is tight, i.e., for any positive integer n, there exist Gaussian half-duplex diamond n-relay networks $\mathcal{N}$ for which $\mathsf{C}_1(\mathcal{N})$ satisfies the bound in (3) with equality.

Remark 1. *The bound in (3) for $n = 2$ and $n \to \infty$ reduces to*

$$\frac{\mathsf{C}_1(\mathcal{N})}{\mathsf{C}_n(\mathcal{N})} \geq \begin{cases} 1/2 & n = 2, \\ 1/4 & n \to \infty, \end{cases}$$

which subsumes the result of [9]. However, the bound in (3) provides a tight and non-asymptotic guarantee for all values of n, which was left as an open problem in [9].

Remark 2. *The bound in (3) has a pretty surprising behavior, which depends on the cosine of a function of the number of relays n, as also graphically shown in Fig. 2. This is fundamentally different from full-duplex [1], where the best relay has always a capacity that is at least $1/2$ of the entire network approximate capacity, independent of n.*

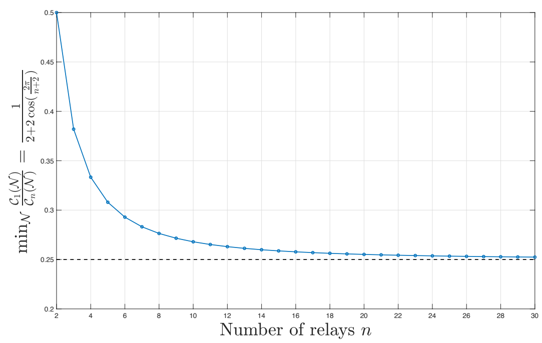


Fig. 2: Ratio $\mathsf{C}_1(\mathcal{N})/\mathsf{C}_n(\mathcal{N})$ in (3) as a function of n.

IV. Sketch of the Proof

In Section IV-A, we describe a few properties of the approximate capacity, which we then use in Section IV-B to present the main steps of the proof of (3). Finally, in Section IV-C we prove the tightness of (3), by providing some network realizations that satisfy the bound in (3) with equality.

A. Properties of the Approximate Capacity

The approximate capacity satisfies the following properties:

(P1) $\mathsf{C}_n(\mathcal{N})$ is a non-decreasing function of each point-to-point link capacity;

(P2) $\mathsf{C}_1(\mathcal{N})/\mathsf{C}_n(\mathcal{N})$ is invariant to scaling all the point-to-point link capacities by a constant factor;

(P3) $\mathsf{C}_1(\mathcal{N})/\mathsf{C}_n(\mathcal{N})$ is invariant to a relabelling of the relays.

Consider a network with the minimum $\mathsf{C}_1(\mathcal{N})/\mathsf{C}_n(\mathcal{N})$. For the relays with approximate capacity less than $\mathsf{C}_1(\mathcal{N})$, we can increase their channel coefficients until their approximate capacity reaches $\mathsf{C}_1(\mathcal{N})$. Then (P1) guarantees that $\mathsf{C}_n(\mathcal{N})$ does not decrease, and the ratio cannot increase. Then, we can normalize the channel gains so that all relays have unitary approximate capacity (by (P2)) and sort the relays w.r.t. their left-side links (by (P3)). This leads to the following lemma, which is formally proved in [10].

Lemma 1. *Let $\mathcal{N}^\star$ be the collection of half-duplex diamond n-relay networks with minimum $\mathsf{C}_1(\cdot)/\mathsf{C}_n(\cdot)$. Then, there exists a network $\mathcal{N} \in \mathcal{N}^\star$ that satisfies*

$$1 \le \ell_1 \le \ell_2 \le \ldots \le \ell_{n-1} \le \ell_n \le \infty, \tag{4a}$$
$$\infty \le r_n \le r_{n-1} \le \ldots \le r_2 \le r_1 \le 1, \tag{4b}$$
$$\mathsf{C}_1(\mathcal{N}_i) = \ell_i r_i/(\ell_i + r_i) = 1, \quad \forall i \in [1:n]. \tag{4c}$$

We also need the following lemma, that is proved in [10].

Lemma 2. *Let $\mathcal{A}$ be any set, and $\{f_i(\cdot), i \in [1:t]\}$ be any set of functions. Then, the two optimization problems given by*

$$\begin{aligned} &\max_{\mathbf{x}\in\mathcal{A}} \ y \\ &\text{s.t.} \quad y \le f_i(\mathbf{x}), \quad i \in [1:t], \end{aligned} \tag{5}$$

and

$$\begin{aligned} &\min_{\boldsymbol{\mu}} \max_{\mathbf{x}\in\mathcal{A}} \ \sum\nolimits_{i=1}^{t} \mu_i f_i(\mathbf{x}) \\ &\qquad \text{s.t.} \quad \mu_i \ge 0, \ i \in [1:t], \quad \sum\nolimits_{i=1}^{t} \mu_i = 1, \end{aligned} \tag{6}$$

have identical solutions.

B. Overview of the Proof for the Fraction Guarantee in (3)

The result in Lemma 1 implies that, for any positive integer n, there always exists a network $\mathcal{N}$ for which $\mathsf{C}_1(\mathcal{N})/\mathsf{C}_n(\mathcal{N})$ is minimum and $\mathsf{C}_1(\mathcal{N}_i) = 1, \forall i \in [1:n]$; hence, also $\mathsf{C}_1(\mathcal{N}) = 1$. Thus, proving (3) reduces to proving that, for any Gaussian half-duplex diamond n-relay network $\mathcal{N}$ with unitary single relay approximate capacities, we always have $\mathsf{C}_n(\mathcal{N}) \le \sigma_n + 2$, where $\sigma_n = 2\cos(\frac{2\pi}{n+2})$, or equivalently,

$$\max_{\mathcal{N}:\mathsf{C}_1(\mathcal{N}_i)=1,\forall i\in[1:n]} \mathsf{C}_n(\mathcal{N}) \le \sigma_n + 2. \tag{7}$$

Step 1: Equivalent formulation of (7). We define $z_i \triangleq \ell_i - 1, i \in [1:n]$, which together with $\mathsf{C}_1(\mathcal{N}_i) = 1$ implies $r_i = \frac{1}{z_i} + 1$. Thus, the class of networks of interest can be parameterized by $\mathbf{z} = [z_1, z_2, \ldots, z_n]$, where $0 \le z_1 \le z_2 \le \ldots \le z_n \le \infty$ because of the condition in (4a). By using the definition in (2), our optimization problem in (7) can be written as

$$\begin{aligned} \mathrm{OPT}_0 \ &= \max_{\mathbf{z}} \max_{\boldsymbol{\lambda}} \Gamma \\ &\text{s.t. } \Gamma \le \sum_{\mathcal{S}\subseteq[1:n]} \lambda_{\mathcal{S}} \left(\max_{i\in\mathcal{S}^c\cap\Omega^c} \ell_i + \max_{i\in\mathcal{S}\cap\Omega} r_i \right), \ \forall \Omega \subseteq [1:n], \\ &\quad \sum_{\mathcal{S}\subseteq[1:n]} \lambda_{\mathcal{S}} = 1, \ \lambda_{\mathcal{S}} \ge 0, \ \forall \mathcal{S} \subseteq [1:n], \\ &\quad \ell_i = 1 + z_i, \quad r_i = 1 + \frac{1}{z_i}, \quad i \in [1:n], \\ &\quad 0 \le z_1 \le z_2 \le \cdots \le z_n \le \infty. \end{aligned} \tag{8}$$

Step 2: Reducing the number of constraints. The optimization problem in (8) has one constraint for each possible partition of the relays $\Omega \subseteq [1:n]$. We now focus on a small class of such partitions parameterized as $\Omega_t, \forall t \in [0:n]$, where

$$\Omega_t = [t+1:n], \qquad \text{and} \qquad \Omega_t^c = [1:t]. \tag{9}$$

With this, the right-hand-side of the cut constraint corresponding to Ω_t in (8) can be simplified as

$$\begin{aligned} &\sum_{\mathcal{S}\subseteq[1:n]} \lambda_{\mathcal{S}} \left(\max_{i\in\mathcal{S}^c\cap\Omega_t^c} \ell_i + \max_{i\in\mathcal{S}\cap\Omega_t} r_i \right) \\ &\overset{\text{(a)}}{\le} (1-\alpha_t)\ell_{t-1} + \alpha_t \ell_t + (1-\alpha_{t+1}) r_{t+1} + \alpha_{t+1} r_{t+2} \\ &\overset{\text{(b)}}{=} \bar{\alpha}_t z_{t-1} + \alpha_t z_t + \bar{\alpha}_{t+1} \frac{1}{z_{t+1}} + \alpha_{t+1} \frac{1}{z_{t+2}} + 2 \\ &\triangleq g_t(\mathbf{z}, \boldsymbol{\alpha}), \end{aligned} \tag{10}$$

where $\alpha_t = \sum_{\mathcal{S}:t\notin\mathcal{S}} \lambda_{\mathcal{S}} = 1 - \bar{\alpha}_t$; (a) follows from (4a) and (4b), and in (b) we substituted $\ell_t = 1 + z_t$ and $r_t = 1 + 1/z_t$ for $t \in [1:n]$. We define $z_i = -1$ for $i \notin [1:n]$.

By ignoring all the cut constraints except those in $\{\Omega_t : t \in [0:n]\}$, we obtain

$$\begin{aligned} \mathrm{OPT}_1 = &\max_{\mathbf{z},\boldsymbol{\alpha}} \Gamma \\ &\text{s.t.} \quad \Gamma \le g_t(\mathbf{z}, \boldsymbol{\alpha}), \quad \forall t \in [0:n], \\ &\qquad \alpha_i \in [0,1], \quad \forall i \in [0:n+1], \\ &\qquad 0 \le z_1 \le z_2 \le \ldots \le z_n, \\ &\qquad z_{-1} = z_0 = z_{n+1} = z_{n+2} = -1. \end{aligned} \tag{11}$$

It is clear that $\mathrm{OPT}_0 \leq \mathrm{OPT}_1$ and by means of Lemma 2, we have that $\mathrm{OPT}_1 = \mathrm{OPT}_2$, where

$$\begin{aligned} \mathrm{OPT}_2 = \min_{\boldsymbol{\mu}} \max_{\boldsymbol{z},\boldsymbol{\alpha}} \ & h(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\alpha}) \\ \text{s.t.} \ & \mu_t \geq 0, \quad \forall t \in [0:n], \\ & \textstyle\sum_{t=0}^{n} \mu_t = 1, \\ & \alpha_i \in [0,1], \quad \forall i \in [0:n+1], \\ & 0 \leq z_1 \leq z_2 \leq \ldots \leq z_n, \\ & z_{-1} = z_0 = z_{n+1} = z_{n+2} = -1, \end{aligned} \tag{12a}$$

with

$$h(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\alpha}) = \textstyle\sum_{t=0}^{n} \mu_t g_t(\boldsymbol{z}, \boldsymbol{\alpha}). \tag{12b}$$

Step 3: Grouping optimum $z_t^\star$'s. By taking the derivative of $h(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\alpha})$ in (12b) with respect to each variable z_t, it is easy to see that $h(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\alpha})$ is a convex function in z_t for any fixed coefficient vectors $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$. Hence, at the optimum point $(\boldsymbol{\mu}^\star, \boldsymbol{z}^\star, \boldsymbol{\alpha}^\star)$ for (12), each z_t should take one of its extreme values. Since $z_{t-1} \leq z_t \leq z_{t+1}$, this implies that for the optimum vector $\boldsymbol{z}^\star = [z_1^\star, z_2^\star, \cdots, z_n^\star]$ we have $z_t^\star \in \{z_{t-1}^\star, z_{t+1}^\star\}$ for $t \in [2:n-1]$, $z_1^\star \in \{0, z_2^\star\}$, and $z_n^\star \in \{z_{n-1}^\star, \infty\}$. Therefore, $(z_1^\star, z_2^\star, \cdots, z_n^\star)$ can be grouped into:

$$\begin{aligned} & z_1^\star = \cdots = z_{t_1}^\star = \beta_1, \quad z_{t_1+1}^\star = \cdots = z_{t_2}^\star = \beta_2, \\ & \cdots \quad z_{t_{m-1}+1}^\star = \cdots = z_{t_m}^\star = \beta_m, \end{aligned} \tag{13}$$

where $0 \leq \beta_1 < \beta_2 < \cdots < \beta_{m-1} < \beta_m \leq \infty$. Note that $t_j - t_{j-1}$ (with $t_0 = 0$) is the number of z_i's whose optimum value equals β_j. It is not difficult to see that [10]: (i) $t_1 \geq 1$ if $\beta_1 = 0$, (ii) $t_1 \geq 2$ if $\beta_1 > 0$, (iii) $t_i - t_{i-1} \geq 2$ for $i \in [2:m-1]$ that implies $1 \leq m \leq \lfloor \frac{n+2}{2} \rfloor$, (iv) $t_m - t_{m-1} \geq 1$ if $\beta_m = \infty$ and (v) $t_m - t_{m-1} \geq 2$ if $\beta_m < \infty$.

Next, we use (13) and the above relations to upper bound $g_t(\boldsymbol{z}^\star, \boldsymbol{\alpha})$ in (10) for all $t \in \{t_1, t_2, \ldots, t_{m-1}\}$. We have

$$\begin{aligned} & g_{t_i}(\boldsymbol{z}^\star, \boldsymbol{\alpha}) \\ & = \bar{\alpha}_{t_i} z_{t_i-1}^\star + \alpha_{t_i} z_{t_i}^\star + \bar{\alpha}_{t_i+1} \frac{1}{z_{t_i+1}^\star} + \alpha_{t_i+1} \frac{1}{z_{t_i+2}^\star} + 2 \\ & \leq 2 + \beta_i + \frac{1}{\beta_{i+1}} \triangleq G_i(\boldsymbol{\beta}). \end{aligned} \tag{14}$$

Similarly, for $t \in \{0, n\}$, we obtain

$$g_0(\boldsymbol{z}^\star, \boldsymbol{\alpha}) \leq 1 + \frac{1}{\beta_1} \triangleq G_0(\boldsymbol{\beta}), \tag{15}$$

$$\text{and} \quad g_n(\boldsymbol{z}^\star, \boldsymbol{\alpha}) \leq 1 + \beta_m \triangleq G_m(\boldsymbol{\beta}). \tag{16}$$

Next, we use (14)-(16) to further upper bound the objective function $h(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\alpha})$ of OPT_2 – defined in (12b) – as follows:

$$\begin{aligned} h(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\alpha}) &= \sum_{i=0}^{n} \mu_i g_i(\boldsymbol{z}^\star, \boldsymbol{\alpha}) \\ &\leq \sum_{i=0}^{m} \mu_{t_i} G_i(\boldsymbol{\beta}) + \sum_{i \notin \{t_0, \ldots, t_m\}} \mu_i g_i(\boldsymbol{z}^\star, \boldsymbol{\alpha}). \end{aligned} \tag{17}$$

Step 4: Further reducing the number of constraints. The optimization problem in (12) involves a minimization over $\boldsymbol{\mu}$. Thus, setting more restrictions on the variable $\boldsymbol{\mu}$ can only increase the optimum cost. We set $\mu_t = 0$ for $t \notin \{t_0 = 0, t_1, t_2, \ldots, t_m = n\}$, and $\mu_{t_i} = \tilde{\mu}_i$ for $i = \{0, 1, \ldots, m\}$. Here $\tilde{\mu}_i$'s are arbitrary non-negative variables that sum up to 1. Incorporating this and the bound in (17) into (12) gives us:

$$\begin{aligned} \mathrm{OPT}_3 = \min_{\tilde{\boldsymbol{\mu}}} \max_{m \in [1:\lfloor \frac{n+2}{2} \rfloor]} \max_{\boldsymbol{\beta}} \ & \sum_{t=0}^{m} \tilde{\mu}_t G_t(\boldsymbol{\beta}) \\ \text{s.t.} \ & \tilde{\mu}_t \geq 0, \quad \forall t \in [0:m], \\ & \sum_{t=0}^{m} \tilde{\mu}_t = 1, \\ & 0 \leq \beta_1 < \beta_2 < \cdots < \beta_m \leq \infty. \end{aligned} \tag{18}$$

Note that $\mathrm{OPT}_2 \leq \mathrm{OPT}_3$ since: (i) the objective function in (18) is an upper bound for that of (12), and (ii) the feasible set for $\boldsymbol{\mu}$ in (12) is a super set of that of $\tilde{\boldsymbol{\mu}}$ in (18). Applying Lemma 2 on the optimization problem in (18), we get:

$$\begin{aligned} \mathrm{OPT}_4 = \max_{m \in [1:\lfloor \frac{n+2}{2} \rfloor]} \ & \max_{\boldsymbol{\beta}} \ \Phi \\ \text{s.t.} \ & \Phi \leq G_i(\boldsymbol{\beta}), \quad \forall i \in [0:m], \\ & 0 \leq \beta_1 < \beta_2 < \cdots < \beta_m \leq \infty, \end{aligned} \tag{19}$$

where $G_i(\boldsymbol{\beta})$'s are defined in (14)-(16). Note that Lemma 2 implies that $\mathrm{OPT}_3 = \mathrm{OPT}_4$.

Step 5: Solving the inner optimization problem in (19). We fix m in the optimization problem in (19) and further analyze the inner optimization problem. This yields

$$\begin{aligned} \mathrm{OPT}_5(m) = \max_{\boldsymbol{\beta}} \ & \Phi \\ \text{s.t.} \ & \Phi \leq G_i(\boldsymbol{\beta}), \quad \forall i \in [0:m], \\ & 0 \leq \beta_1 < \beta_2 < \cdots < \beta_m \leq \infty, \end{aligned} \tag{20}$$

for every fixed $m \in \left[1 : \lfloor \frac{n+2}{2} \rfloor\right]$. The following lemma, whose proof can be found in [10], highlights some important properties of the optimum solution of the optimization problem in (20).

Lemma 3. *For every integer m, there exists some solution $(\boldsymbol{\beta}^\star, \Phi^\star)$ for the optimization problem in (20) that satisfies*

$$G_i(\boldsymbol{\beta}^\star) = \Phi^\star, \qquad \forall i \in [1:m-1].$$

Moreover, if $\beta_1^\star > 0$, we have $G_0(\boldsymbol{\beta}^\star) = \Phi^\star$, and similarly, if $\beta_m^\star < \infty$, then $G_m(\boldsymbol{\beta}^\star) = \Phi^\star$.

We now analyze the structure of $\mathrm{OPT}_5(m)$: for a given m, we find the optimal $\boldsymbol{\beta}^\star$ that satisfies Lemma 3. In particular, we should consider four different cases, depending on whether $\beta_1^\star = 0$ or $\beta_1^\star > 0$ and $\beta_m^\star = \infty$ or $\beta_m^\star < \infty$. For illustration, we here focus on the case $\beta_1^\star > 0$ and $\beta_m^\star < \infty$; the detailed analysis of the other three cases can be found in [10]. For the case $\beta_1^\star > 0$ and $\beta_m^\star < \infty$, we define

$$b_0 = 1, \qquad b_i = \frac{1}{\prod_{k=1}^{i} \beta_k^\star}, \ \forall i \in [1:m], \tag{21}$$

$$\Longrightarrow \beta_i^\star = \frac{b_{i-1}}{b_i}, \qquad \forall i \in [1:m]. \tag{22}$$

Using this change of variables and the fact that $G_i(\boldsymbol{\beta}^\star) = \mathrm{OPT}_5(m), i \in [1:m-1]$ (see Lemma 3), we get that

$$G_i(\boldsymbol{\beta}^\star) = 2 + \beta_i^\star + \frac{1}{\beta_{i+1}^\star} = 2 + \frac{b_{i-1}}{b_i} + \frac{b_{i+1}}{b_i},$$

for every i in $[1:m-1]$. Then, for given n and m, we define:

$$\sigma_{n,m} \triangleq \mathrm{OPT}_5(m) - 2 = \frac{b_{i-1}}{b_i} + \frac{b_{i+1}}{b_i}, \ \forall i \in [1:m-1], \quad (23)$$

which implies

$$b_{i+1} - \sigma_{n,m} b_i + b_{i-1} = 0, \quad \forall i \in [1:m-1]. \quad (24)$$

The above expression is an order 2 linear homogeneous recurrence equation, whose solution is given by [11]:

$$b_i = uU^i + vV^i, \qquad i \in [0:m], \quad (25)$$

where U and V are the roots of the characteristic equation of (24), that is, $X^2 - \sigma_{n,m} X + 1 = 0$, and u and v are constants, that can be found from the initial conditions.

Since we assume $\beta_1^\star > 0$ and $\beta_m^\star < \infty$, from Lemma 3 we have

$$\frac{1}{\beta_1^\star} = \beta_m^\star = 1 + \sigma_{n,m},$$

where $\beta_m^\star = b_{m-1}/b_m$ from (22). From (21) we also have that $b_0 = 1$ and $b_1 = 1/\beta_1^\star$ from which we can find the values of u and v in (25). Using all these results, we can find the optimum value of $\sigma_{n,m}$ in (23), as stated in the following proposition, whose complete proof is provided in [10, Appendix A].

Proposition 1. *The optimal value $\sigma_{n,m}$ in (23) is given by*

$$\sigma_{n,m} = 2\cos\left(\frac{2\pi}{2m+2}\right). \quad (26)$$

Step 6: Optimizing $\sigma_{n,m}$ in (26) over m. From (19), we have

$$\mathrm{OPT}_4 = \max_{m \in \left[1:\lfloor \frac{n+2}{2} \rfloor\right]} \mathrm{OPT}_5(m) = 2 + \max_{m \in \left[1:\lfloor \frac{n+2}{2} \rfloor\right]} \sigma_{n,m}, \quad (27)$$

where the second equality follows from (23) where $\sigma_{n,m}$ is given in (26). The next proposition provides the optimum m, and hence the optimum solution for the problem in (27).

Proposition 2. *The optimal solution for the optimization problem in (27) is given by*

$$\mathrm{OPT}_4 = 2 + 2\cos\left(\frac{2\pi}{n+2}\right).$$

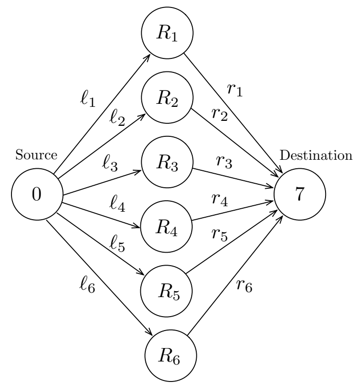
Proof. We focus on the case $\beta_1^\star > 0$ and $\beta_m^\star < \infty$, and relegate the other three cases to [10]. For this case, we have $t_1 \geq 2$ and $t_i - t_{i-1} \geq 2$ for $i \in [2:m]$. Thus, since $t_m = n$, we get

$$n = t_m = \sum_{i=2}^{m} (t_i - t_{i-1}) + t_1 \geq 2(m-1) + 2 = 2m,$$

which implies $m \leq \frac{n}{2}$, and hence

$$\mathrm{OPT}_4 = 2 + \max_{m \leq \frac{n}{2}} 2\cos\left(\frac{2\pi}{2m+2}\right) = 2 + 2\cos\left(\frac{2\pi}{n+2}\right).$$

$\square$



(a)

i	ℓ_i	r_i
1	$\sqrt{2}$	$2+\sqrt{2}$
2	$\sqrt{2}$	$2+\sqrt{2}$
3	2	2
4	2	2
5	$2+\sqrt{2}$	$\sqrt{2}$
6	$2+\sqrt{2}$	$\sqrt{2}$

(b)

i	ℓ_i	r_i
1	1	$L \to \infty$
2	1.8019	2.2470
3	1.8019	2.2470
4	3.2470	1.4450
5	3.2470	1.4450

(c)

Fig. 3: Gaussian half-duplex diamond networks for which the bound in (3) is tight. The table in (b) shows the link capacities for the network defined in (28) with $n = 6$, and the table in (c) indicates the link capacities of the network given in (29) with $n = 5$.

Step 7: Collecting all the results together. We have proved that for any Gaussian half-duplex diamond n-relay network $\mathcal{N}$ with unitary single relay approximate capacities we always have

$$\mathsf{C}_n(\mathcal{N}) = \mathrm{OPT}_0 \leq \mathrm{OPT}_1 = \mathrm{OPT}_2 \leq \mathrm{OPT}_3$$
$$= \mathrm{OPT}_4 = 2 + 2\cos\left(\frac{2\pi}{n+2}\right),$$

which concludes the proof of the ratio guarantee in Theorem 1.

C. Tightness of the Bound in (3)

To conclude the proof of Theorem 1, we need to show that the bound in (3) is tight. Towards this end, we next present two network realizations for which the ratio in (3) is indeed satisfied with equality. The detailed analysis of these networks can be found in [10]. In what follows we define $\theta = \frac{2\pi}{n+2}$.

Case 1: Let $n = 2k$ be an even integer. A Gaussian half-duplex diamond n-relay network $\mathcal{N}$ for which the ratio in (3) is tight is given by (see Fig. 3(b) for $n = 6$)

$$\begin{aligned} \ell_{2i} = \ell_{2i-1} &= \frac{2\sin(\theta)\sin(i\theta)}{\cos(i\theta) - \cos((i+1)\theta)}, \ i \in [1:k], \\ r_{2i} = r_{2i-1} &= \frac{2\sin(\theta)\sin(i\theta)}{\cos((i-1)\theta) - \cos(i\theta)}, \ i \in [1:k]. \end{aligned} \quad (28)$$

Case 2: Let $n = 2k+1$ be an odd number. A Gaussian half-duplex diamond n-relay network $\mathcal{N}$ for which the ratio in (3) is tight is given by (see Fig. 3(c) for $n = 5$)

$$\begin{aligned} \ell_1 &= 1, \qquad r_1 = L \to \infty, \\ \ell_{2i} = \ell_{2i+1} &= \frac{\sin(i\theta) + \sin((i+1)\theta)}{\sin((i+1)\theta)}, \ i \in [1:k], \\ r_{2i} = r_{2i+1} &= \frac{\sin(i\theta) + \sin((i+1)\theta)}{\sin(i\theta)}, \ i \in [1:k]. \end{aligned} \quad (29)$$

References

[1] C. Nazaroglu, A. Özgür, and C. Fragouli, “Wireless network simplification: The Gaussian N-relay diamond network,” *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6329–6341, October 2014.

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

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

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

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

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

[7] Y. H. Ezzeldin, A. Sengupta, and C. Fragouli, “Wireless network simplification: Beyond diamond networks,” in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2594–2598.

[8] S. Brahma and C. Fragouli, “A simple relaying strategy for diamond networks,” in *Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT)*, June 2014, pp. 1922–1926.

[9] Y. H. Ezzeldin, M. Cardone, C. Fragouli, and D. Tuninetti, “Network simplification in half-duplex: Building on submodularity,” *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6801–6818, October 2019.

[10] S. Jain, S. Mohajer, and M. Cardone, “Best relay selection in Gaussian half-duplex diamond networks,” *arXiv:2001.02851*, January 2020.

[11] H. S. Wilf, *generatingfunctionology*. AK Peters/CRC Press, 2005.