The Approximate Capacity of the Gaussian N-Relay Diamond Network

Urs Niesen and Suhas N. Diggavi

Abstract

We consider the Gaussian "diamond" or parallel relay network, in which a source node transmits a message to a destination node with the help of N relays. Even for the symmetric setting, in which the channel gains to the relays are identical and the channel gains from the relays are identical, the capacity of this channel is unknown in general. The best known capacity approximation is up to an additive gap of order N bits and up to a multiplicative gap of order N^2, with both gaps independent of the channel gains.

In this paper, we approximate the capacity of the symmetric Gaussian N-relay diamond network up to an additive gap of 1.8 bits and up to a multiplicative gap of a factor 14. Both gaps are independent of the channel gains and, unlike the best previously known result, are also independent of the number of relays N in the network. Achievability is based on bursty amplify-and-forward, showing that this simple scheme is uniformly approximately optimal, both in the low-rate as well as in the high-rate regimes. The upper bound on capacity is based on a careful evaluation of the cut-set bound. We also present approximation results for the asymmetric Gaussian N-relay diamond network. In particular, we show that bursty amplify-and-forward combined with optimal relay selection achieves a rate within a factor $O(\log^4(N))$ of capacity with pre-constant in the order notation independent of the channel gains.

I. Introduction

Cooperation is a key feature of wireless communication. A simple canonical channel model capturing this feature is the "diamond" or parallel relay network introduced by Schein and Gallager [1], [2]. This network consists of a source node connected through a broadcast channel to N relays; the relays, in turn, are connected to the destination node through a multiple-access channel (see Fig. 1). The objective is

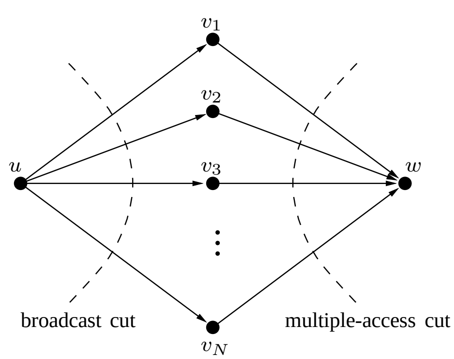

Fig. 1. The N-relay diamond network. The source node u transmits a message to the destination node w via the N relays $\{v_n\}_{n=1}^N$. The two cuts indicated in the figure are the broadcast cut (separating the source u from the relays $\{v_n\}$) and the multiple-access cut (separating the relays $\{v_n\}$ from the destination w).

to maximize the rate achievable between the source and the destination with the help of the N relays. Throughout this paper, we will be interested in the Gaussian version of this problem, in which both the broadcast and the multiple-access parts are subject to additive Gaussian noise. Moreover, for simplicity we will restrict attention in a significant part of the paper to the symmetric case, in which the channel

U. Niesen is with the Mathematics of Networks and Communications Research Department, Bell Labs, Alcatel-Lucent. Email: urs.niesen@alcatel-lucent.com. S. Diggavi is with the University of California, Los Angeles. Email: suhas@ee.ucla.edu.

gains within the multiple-access part and within the broadcast part of the network are identical (but are allowed to differ between the multiple-access and broadcast parts). However, we do show that some of the results for the symmetric setting can be extended to the asymmetric setting.

For the Gaussian 2-relay diamond network, the rates achievable with decode-and-forward and with amplify-and-forward at the relays were analyzed in [2]. It is shown there that these schemes achieve capacity in some regimes of signal-to-noise ratios (SNRs) of the broadcast and multiple-access parts of the diamond network. The asymptotic behavior of the N-relay Gaussian diamond network was investigated in [3]. In certain regimes of SNRs of the broadcast and multiple-access parts of the network, it is shown that amplify-and-forward is capacity achieving in the limit as $N \to \infty$. New achievable schemes for the Gaussian diamond network with bandwidth mismatch (i.e., the source and the relays have different bandwidth) were introduced in [4] and [5]. Perhaps surprisingly, these schemes lead to higher achievable rates than the ones obtained with amplify-and-forward and decode-and-forward even when the bandwidths at the source and the relays are identical. Half-duplex versions of the Gaussian diamond network, in which the relays cannot receive and transmit signals simultaneously, were considered in [6] and [7]. The capacity of a special class of 2-relay diamond networks is derived in [8]. For networks in this class, one relay receives the signal sent at the source without noise, and the destination node is connected to the relays by two orthogonal bit pipes of fixed rate. To the best of our knowledge, this is the only non-trivial example for which the capacity of the diamond network is known for all values of SNR. For the general Gaussian N-relay diamond network, the capacity is unknown.

Given the difficulty of determining the capacity of communication networks in general and of the diamond network in particular, it is natural to ask if it can at least be approximated. For high rates, such an approximation should be additive in nature, i.e., we would like to determine capacity up to an additive gap. For low rates, such an approximation should be multiplicative, i.e., we would like to determine capacity up to a multiplicative gap. If a communication strategy can be shown to have both small additive as well as multiplicative gaps, then this strategy is provably close to optimal both in the high rate as well as low rate regimes.

Additive approximations for channel capacity of communication networks were first derived in [9], where the capacity region of the two-user Gaussian interference channel is determined up to an additive gap of one bit. This was mainly enabled through a new outer bound for the interference channel. The approach of approximate capacity characterization was applied to general relay networks with single-source multicast in [10]. By introducing a new relaying strategy termed quantize-map-forward, capacity is derived up to an additive gap of $15n$ bits, where n is the number of nodes in the network. This additive gap was improved through the use of vector quantization at the relays [11], [12]. The sharpest known additive approximation gap is $1.26n$ bits for the complex Gaussian case (or $0.63n$ for the real case) [12]. Since the N-user diamond network is a special case of a relay network with a single source and destination and with $n = N + 2$ nodes, these results yield an additive approximation up to a gap of $0.63N + 1.26$ bits for this network (assuming real channel gains).

Multiplicative approximations were mostly analyzed for large wireless networks, for which the rate per source-destination pair is low. For a network with n nodes, the emphasis is on finding capacity approximations up to a small multiplicative factor in n. This approach was pioneered in [13]. Under a restricted model of communication, (essentially) the equal rate point of the capacity region of a wireless network with n randomly placed nodes was determined up to a constant multiplicative factor independent of n. Without the restrictive communication assumptions in [13], the problem becomes considerably harder. Approximations for the equal rate point under a Gaussian model were derived in [14] up to a multiplicative factor of $O(n^{\varepsilon})$ for any $\varepsilon > 0$. These approximation results were subsequently sharpened in [15], [16] to a factor $n^{O(1/\sqrt{\log(n)})}$. Under some conditions on the node placement, this factor can further be sharpened to $O(\log(n))$ [17]. Multiplicative approximations for arbitrary relay networks with single-source multicast (as opposed to wireless networks with multiple unicast, i.e., multiple separate source-destination pairs) were derived in [10]. For a network with maximum degree d, the capacity is approximated to within a factor of $2d(d+1)$. As pointed out earlier, the Gaussian N-relay diamond network is such a network with maximum degree $d = N$, and hence this result yields a multiplicative approximation up to a factor of $2N(N+1)$.

To summarize, the capacity region of the general Gaussian N-relay diamond network is not known. The best known additive approximation is up to a gap of $0.63N + 1.26$ bits, and the best known multiplicative approximation is up to a factor of $2N(N+1)$. In either case, the bounds degrade rather quickly as N increases. It is hence of interest to find approximation guarantees that behave better as a function of the number of relays N in the network. Ideally, we would like the approximation guarantees to be uniform in in the network size.

As a main result of this paper, we show that such a uniform approximation is indeed possible. More precisely, we find an additive approximation of the capacity of the symmetric Gaussian N-relay diamond network of gap at most 1.8 bits for any SNR and number of relays N. Moreover, we find a multiplicative approximation to the capacity up to at most a factor 14, again for any SNR and number of relays N. This is a significant improvement over the previously best known additive approximation of $0.63N + 1.26$ bits and multiplicative approximation of a factor $2N(N+1)$, especially for large values of N. In particular, as far as we know, this is the first such approximation result (both multiplicative as well as additive) that is independent of the number of network nodes for a nontrivial class of wireless networks.

We further show that bursty amplify-and-forward (first introduced in [2, p. 76]) with properly chosen duty cycle is close to capacity achieving for the diamond network simultaneously in the sense of multiplicative and additive approximation up to the aforementioned gaps. Hence, bursty amplify-and-forward with appropriately chosen duty cycle is a good communication scheme for the symmetric Gaussian N-relay diamond network both at low and at high SNRs, and independently of the number of relays N.

Some of these results can be extended to the asymmetric setting. For general (i.e., not necessarily symmetric) Gaussian N-relay diamond networks, we provide a factor $O(\log^4(N))$ multiplicative approximation of capacity, with pre-constant in the order notation independent of the channel gains. Achievability is based again on bursty amplify-and-forward, but this time a careful selection of relays is also necessary.

The main technical contribution of this paper is the upper bound on capacity. The standard way to obtain upper bounds on the capacity of the diamond network is to evaluate two particular cuts in the wireless network, namely the one separating the source from the relays (called the *broadcast cut* in the following) and the one separating the relays from the destination (called the *multiple-access cut* in the following) as depicted in Fig. 1. This approach is taken, for example, in [3]–[5]. In fact, for symmetric Gaussian N-relay diamond networks, whenever the capacity is known, it coincides with the minimum of these two cuts. We show in this paper that, in order to obtain uniform additive or multiplicative approximations for the capacity of this network, considering just these two cuts is not sufficient. Instead we need to *simultaneously* optimize over *all* possible 2^N cuts separating the source from the destination. Without this careful outer bound evaluation, we believe that the uniform (in network size) approximation would not have been possible.

The remainder of this paper is organized as follows. Section II formally introduces the problem statement. Section III presents the main results; the corresponding proofs are presented in Section IV. Section V contains concluding remarks.

II. Problem Statement

Consider the Gaussian N-relay diamond network as depicted in Fig. 2. The source node u transmits a message to the destination node w with the help of N parallel relays $\{v_1, \dots, v_N\}$. The channel inputs at time $t \in \mathbb{N}$ at nodes u and v_n are denoted by $X[t]$ and $X_n[t]$, respectively. The channel outputs at time

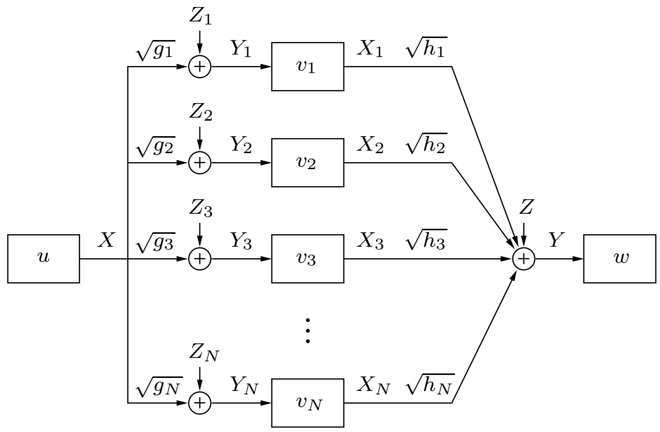

Fig. 2. The Gaussian N-relay diamond network.

$t \in \mathbb{N}$ at nodes w and v_n are denoted by $Y[t]$ and $Y_n[t]$. The channel inputs and outputs are related as

$$Y_n[t] \triangleq \sqrt{g_n} X[t] + Z_n[t],$$
$$Y[t] \triangleq \sum_{n=1}^{N} \sqrt{h_n} X_n[t] + Z[t],$$

where $(Z[t])_t, (Z_n[t])_{n,t}$ are independent and identically distributed Gaussian random variables with mean zero and variance one, independent of the channel inputs. The channel gains $(g_n)_{n=1}^N$ and $(h_n)_{n=1}^N$ are assumed to be real positive numbers, constant as a function of time, and known throughout the network.

A *T-length block code* for the diamond network is a collection of functions

$$f \colon \{1, \dots, M\} \to \mathbb{R}^T,$$
$$f_n \colon \mathbb{R}^T \to \mathbb{R}^T, \ \forall n \in \{1, \dots, N\},$$
$$\phi \colon \mathbb{R}^T \to \{1, \dots, M\}.$$

The encoding function f maps the message W, assumed to be uniformly distributed over the set $\{1, \dots, M\}$, to the channel inputs

$$(X[t])_{t=1}^T \triangleq f(W)$$

at the source node u. The function f_n maps the channel outputs $(Y_n[t])_{t=1}^T$ to the channel inputs

$$(X_n[t])_{t=1}^T \triangleq f_n\big((Y_n[t])_{t=1}^T\big)$$

at relay v_n.[1] The decoding function ϕ maps the channel outputs $(Y[t])_{t=1}^T$ at the destination node w into a reconstruction

$$\hat{W} \triangleq \phi\big((Y[t])_{t=1}^T\big).$$

We say the code satisfies a *unit average power constraint* if

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(X^2[t]) \leq 1,$$
$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(X_n^2[t]) \leq 1, \ \forall n \in \{1, \dots, N\}.$$

[1] Note that the functions f_n at the relays are not causal. This is to simplify notation; due to the layered nature of the network all results remain the same if causality is imposed.

The *rate* of the code is

$$\log(M)/T,$$

and its *average probability of error*

$$\mathbb{P}(\hat{W} \neq W).$$

A rate R is *achievable* if there exists a sequence of T-length block codes with unit average power constraint and rate at least R such that the average probability of error approaches zero as $T \to \infty$. The *capacity* $C\big(N, (g_n), (h_n)\big)$ of the diamond network is the supremum of all achievable rates.

A natural scheme for the diamond network is *amplify-and-forward*, in which each relay transmits a scaled version of the received signal. Formally,

$$X_n[t] = \alpha_n Y_n[t] = \alpha_n \sqrt{g_n} X[t] + \alpha_n Z_n[t],$$

where the constant α_n is chosen to satisfy the power constraint at the relay. Denote by $R_1\big(N, (g_n), (h_n)\big)$ the rate achieved by amplify-and-forward with optimal choice of $(\alpha_n)_{n=1}^N$. We point out that the optimization over $(\alpha_n)_{n=1}^N$ is necessary. While perhaps counterintuitive, it turns out that in the asymmetric setting the rate of amplify-and-forward is not always maximized when the relays use all available transmit power (see [2] for a discussion of this phenomenon in the two-relay case).

If the SNR at the relays is low (i.e., $g_n \ll 1$), it can be shown that simple amplify-and-forward is arbitrarily suboptimal. This is because the received signal power g_n at the relay v_n is much smaller than the noise power 1, and therefore the relay amplifies mostly noise. This effect can be mitigated by using *bursty amplify-and-forward* [2]. For a constant $\delta \in (0, 1]$, called the *duty cycle* in the following, we communicate for a fraction δ of time at average power $1/\delta$ using the amplify-and-forward scheme and stay silent for the remaining time. This satisfies the overall average unit power constraint. The resulting achievable rate is denoted by $R_\delta\big(N, (g_n), (h_n)\big)$. This notation is consistent, i.e., for $\delta = 1$ the simple and bursty amplify-and-forward schemes coincide and achieve both rate $R_1\big(N, (g_n), (h_n)\big)$.

A special case of the general diamond network described so far is the *symmetric* setting, in which $g_1 = g_2 = \ldots = g_N = g$ and $h_1 = h_2 \ldots = h_N = h$. With slight abuse of notation, we denote the capacity and rates achievable by bursty amplify-and-forward for the symmetric setting by $C(N, g, h)$ and $R_\delta(N, g, h)$.

Throughout this paper, we use bold font to denote vectors and matrices. $\log(\cdot)$ and $\ln(\cdot)$ denote the logarithms to base 2 and e, respectively. All capacities and rates are expressed in bits per channel use.

III. Main Results

The main results of this paper are additive and multiplicative capacity approximations for the Gaussian diamond relay network. We start with a discussion of symmetric networks in Section III-A. General asymmetric networks are treated in Section III-B.

A. Symmetric Diamond Networks

The first result lower bounds the rate achievable over a symmetric diamond network by using bursty amplify-and-forward with optimized duty cycle δ.

Theorem 1. *For every symmetric diamond network with $N \geq 2$ relays and channel gains $g, h > 0$, there exists a duty cycle $\delta^\star \in (0, 1]$ such that bursty amplify-and-forward achieves at least the rate*

$$R_{\delta^\star}(N, g, h) \geq \begin{cases} \frac{1}{2}\log\big(1 + \frac{1}{3}N\min\{g, Nh\}\big), & \text{if } \max\{g, Nh\} \geq 1 \\ \frac{1}{2}\ln(4/3)\log(1 + Ng), & \text{if } \max\{g, Nh\} < 1,\ g \leq h \\ \frac{1}{2}\log\big(1 + \frac{1}{3}N^2 gh\big), & \text{if } \max\{g, Nh\} < 1,\ g \in (h, N^2 h),\ N\sqrt{gh} \geq 1 \\ \frac{1}{2}\ln(4/3)\log(1 + N\sqrt{gh}), & \text{if } \max\{g, Nh\} < 1,\ g \in (h, N^2 h),\ N\sqrt{gh} < 1 \\ \frac{1}{2}\ln(4/3)\log(1 + N^2 h), & \text{if } \max\{g, Nh\} < 1,\ g \geq N^2 h. \end{cases}$$

The proof of Theorem 1 is presented in Section IV-A. Note that the optimal duty cycle $\delta^\star$ is allowed to depend on N, g, and h. In the high-rate regime, i.e., the first and third cases in Theorem 1, the duty cycle achieving the lower bound is $\delta^\star = 1$, and hence the bursty amplify-and-forward scheme reduces to simple amplify-and-forward. On the other hand, in the low-rate regime, i.e., the second, fourth, and fifth cases in Theorem 1, $\delta^\star < 1$, and (genuine) bursty amplify-and-forward is used.

Having established an achievable rate, the next theorem provides an upper bound on the capacity of the diamond network.

Theorem 2. *For every symmetric diamond network with $N \geq 2$ relays and channel gains $g, h > 0$, capacity is upper bounded by*

$$C(N,g,h) \leq \begin{cases} \frac{1}{2}\log\big(1+N\min\{g,Nh\}\big), & \text{if } \max\{g,Nh\} \geq 1 \\ \frac{1}{2}\log(1+Ng), & \text{if } \max\{g,Nh\} < 1,\ g \leq h \\ \frac{1}{2}\log\big(1+2N^2gh\big)+\frac{1}{2}, & \text{if } \max\{g,Nh\} < 1,\ g \in (h, N^2h),\ N\sqrt{gh} \geq 1 \\ \log(1+2N\sqrt{gh}), & \text{if } \max\{g,Nh\} < 1,\ g \in (h, N^2h),\ N\sqrt{gh} < 1 \\ \frac{1}{2}\log(1+N^2h), & \text{if } \max\{g,Nh\} < 1,\ g \geq N^2h. \end{cases}$$

The proof of Theorem 2 is presented in Section IV-B. As a corollary to Theorems 1 and 2, we obtain that bursty amplify-and-forward is close to optimal, in the sense that it achieves capacity both up to a constant additive gap as well as a constant multiplicative gap, where both constants are independent of the number of relays N and the channel gains g and h. This shows that optimized bursty amplify-and-forward is a good communication scheme for the symmetric diamond network both at low rates (due to the small multiplicative gap) as well as at high rates (due to the small additive gap).

Corollary 3. *For every symmetric diamond network with $N \geq 2$ relays and channel gains $g, h > 0$, there exists a duty cycle $\delta^\star \in (0,1]$ such that*

$$C(N,g,h) - R_{\delta^\star}(N,g,h) \leq 1 + \tfrac{1}{2}\log(3) \leq 1.8 \text{ bits},$$

and

$$\frac{C(N,g,h)}{R_{\delta^\star}(N,g,h)} \leq \frac{4}{\ln(4/3)} \leq 14.$$

The proof of Corollary 3 is presented in Section IV-C. We point out that choosing the duty cycle $\delta^\star$ as a function of N, g, and h, is not necessary to obtain the additive approximation result in Corollary 3. In fact, using only simple amplify-and-forward achieves the same additive approximation guarantee, i.e.,

$$C(N,g,h) - R_1(N,g,h) \leq 1.8 \text{ bits}$$

for all $N \geq 2$, $g, h > 0$. However, the same is not true if we are also interested in multiplicative approximation guarantees (at least in the low-rate regime). To achieve a constant additive approximation as well as constant multiplicative approximation, the duty cycle $\delta^\star$ is required to vary as a function of N, g, and h, and therefore bursty amplify-and-forward is required.

From Theorems 1 and 2, the capacity of the symmetric diamond network has three distinct regimes, depending on whether $g \leq h$, $h < g < N^2h$, or $g \geq N^2h$. In the first regime ($g \leq h$), the channel gain to the relays is weak compared to the channel gain to the destination, and the achievable rate is constrained by the broadcast part of the diamond network. The capacity in this regime is given approximately by

$$C(N,g,h) \approx \tfrac{1}{2}\log(1+Ng),$$

where the approximation is in the sense of Corollary 3, namely up to a multiplicative gap of factor 14 in the low-rate regime ($g \ll N^{-1}$) and up to an additive gap of 1.8 bits in the high-rate regime ($g \gg N^{-1}$). This is the capacity of a single-input multiple-output channel with unit power constraint, one transmit antenna, N receive antennas, and channel gain $\sqrt{g}$ between each of them. Thus, the broadcast cut in Fig. 1 in Section I is approximately tight in this regime.

In the third regime ($g \geq N^2 h$), the channel gain to the relays is strong compared to the channel gain to the destination, and the achievable rate is now constrained by the multiple-access part of the channel. The capacity in the third regime is given approximately by

$$C(N, g, h) \approx \tfrac{1}{2}\log(1 + N^2 h).$$

This is the capacity of a multiple-input single-output channel with unit per-antenna power constraint, N transmit antennas, one receive antenna, and channel gain $\sqrt{h}$ between each of them. Thus, the multiple-access cut in Fig. 1 is approximately tight in this regime. Observe that to achieve this rate the signals sent by the relays must be highly correlated and add up coherently at the destination.

The most interesting regime is the second one ($h < g < N^2 h$). If $\max\{g, Nh\} \geq 1$, then the capacity is given approximately by

$$C(N, g, h) \approx \tfrac{1}{2}\log\big(1 + N\min\{g, Nh\}\big),$$

and again either the broadcast cut or the multiple-access cut are tight. If $\max\{g, Nh\} < 1$ the situation is more complicated. If $N\sqrt{gh} \geq 1$, then the capacity of the diamond network is approximately

$$C(N, g, h) \approx \tfrac{1}{2}\log(1 + N^2 gh),$$

and, if $N\sqrt{gh} < 1$,

$$C(N, g, h) \approx \tfrac{1}{2}\log\big(1 + N\sqrt{gh}\big).$$

In both cases, the capacity depends on the product of g and h, and not merely on the minimum of g and Nh. Hence, neither the broadcast cut nor the multiple-access cut are tight in this case. In fact, these bounds can be arbitrarily bad, both in terms of additive gap as well as multiplicative gap, as the next two examples illustrate.

For the additive gap, consider $g = N^{-5/8}$ and $h = N^{-9/8}$. Then $\max\{g, Nh\} = N^{-1/8} < 1$, $g = N^{1/2}h \in (h, N^2 h)$, and $N\sqrt{gh} = N^{1/8} \geq 1$, so that

$$\begin{aligned} C(N, g, h) &\approx \tfrac{1}{2}\log(1 + N^2 gh) \\ &= \tfrac{1}{2}\log(1 + N^{1/4}). \end{aligned}$$

On the other hand, the minimum of the broadcast and multiple-access cuts yields

$$\tfrac{1}{2}\log\big(1 + N\min\{g, Nh\}\big) = \tfrac{1}{2}\log(1 + N^{3/8}),$$

resulting in an additive gap of order $\Theta(\log(N))$ bits, which is unbounded as the number of relays $N \to \infty$.

For the multiplicative gap, consider $g = N^{-2}$ and $h = N^{-3}$. Then $\max\{g, Nh\} = N^{-2} < 1$, $g = Nh \in (h, N^2 h)$, and $N\sqrt{gh} = N^{-3/2} < 1$, so that

$$\begin{aligned} C(N, g, h) &\approx \tfrac{1}{2}\log\big(1 + N\sqrt{gh}\big) \\ &= \tfrac{1}{2}\log(1 + N^{-3/2}) \\ &\approx \tfrac{1}{2}\log(e)N^{-3/2}. \end{aligned}$$

On the other hand, the minimum of the broadcast and multiple-access cuts yields

$$\begin{aligned} \tfrac{1}{2}\log\big(1 + N\min\{g, Nh\}\big) &= \tfrac{1}{2}\log(1 + N^{-1}) \\ &\approx \tfrac{1}{2}\log(e)N^{-1}, \end{aligned}$$

resulting in a multiplicative gap of order $\Theta(\sqrt{N})$, which is again unbounded as the number of relays $N \to \infty$.

In the second regime, we thus need to take cuts other than the broadcast and multiple-access ones into account. The need for this can be understood as follows. Consider a general cut separating the source node u from the destination node w in the diamond network as shown in Fig. 3. Formally, let $S \subset \{1, \ldots, N\}$, and consider the cut from $u \cup \{v_n\}_{n \in S}$ to $w \cup \{v_n\}_{n \in S^c}$. Assume the signals $(X_n)_{n=1}^N$ sent from the relays to the destination are highly correlated. This results in the signal summing up coherently at the receiver, increasing the rate across the cut. At the same time, if the signals sent from the relays are highly correlated, then the signals $(X_n)_{n\in S^c}$ available at the relays on the other side of the cut can be used to estimate the signal received at the destination node. This decreases the rate across the cut. Thus, for general cuts, there is a tradeoff between the gain from coherent reception and the loss from prediction that come with increased signal correlation. This tradeoff is absent if we only consider the broadcast and multiple-access cuts. It is precisely this tradeoff that determines the behavior of the capacity of the diamond network in the second regime.

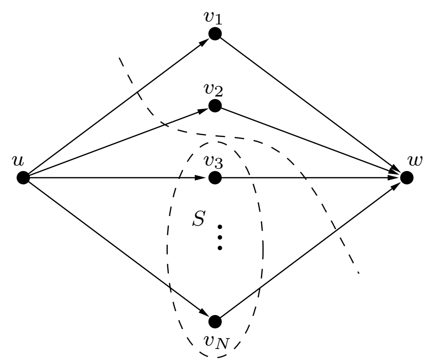

Fig. 3. A general cut in the diamond network. Here $S \subset \{1, \dots, N\}$, and the cut separates $u \cup \{v_n\}_{n\in S}$ from $w \cup \{v_n\}_{n\in S^c}$.

We point out that a (partial) decode-and-forward strategy is not sufficient to provide a uniform capacity approximation as in Corollary 3. Indeed, due to symmetry, *all* relays would be able to decode the source in any such strategy, which implies that decode-and-forward and partial decode-and-forward coincide in this case. The rate achievable with decode-and-forward is given by

$$\tfrac{1}{2}\log\big(1+\min\{g, N^2h\}\big).$$

Comparing this with Corollary 3, we see that (partial) decode-and-forward has an additive gap of at least $\Omega(\log(N))$ bits and a multiplicative gap of at least a factor $\Omega(N)$ to capacity. Similarly, compress-and-forward does not achieve constant (in the network size N) additive or multiplicative gaps to capacity, since it does not capture the gain from coherent signal addition at the destination. Finally, as was pointed out earlier, the traditional amplify-and-forward strategy does not yield a constant factor approximation of capacity. In fact, in can be shown that simple amplify-and-forward results in unbounded multiplicative gap even for $N = 2$. Therefore the bursty amplify-and-forward scheme introduced in [2] and advocated in this work has the nice property of being uniformly approximately optimal in both the additive sense and the multiplicative sense, as well as being a simple modification of the traditional amplify-and-forward scheme.

B. Asymmetric Diamond Networks

In the last section, we have considered *symmetric* diamond networks, in which the channel gain from the source u to the relay v_n is $\sqrt{g}$ and the channel gain from v_n to the destination w is $\sqrt{h}$ for all n. In this section, we show how some of the results can be extended to *asymmetric* diamond networks, in which the channel gains $(g_n)_{n=1}^N$ and $(h_n)_{n=1}^N$ are allowed to take arbitrary values.

For this asymmetric setting, it was shown in [10] that (bursty) amplify-and-forward does *not* achieve a constant (as a function of g_n and h_n) additive-gap approximation even when $N = 2$. However, we show here that bursty amplify-and-forward is approximately optimal in the sense of multiplicative approximation for any N, $(g_n)_{n=1}^N$, and $(h_n)_{n=1}^N$. More precisely, we show that the rate achieved by bursty amplify-and-forward combined with optimal relay selection is at most a factor $O(\log^4(N))$ from capacity uniformly in $(g_n)_{n=1}^N$ and $(h_n)_{n=1}^N$. While not constant in N, compared to the best previously known multiplicative approximation of a factor $2N(N+1)$, this is a significant improvement. Hence, at least in the low-rate regime, bursty amplify-and-forward is also a good communication scheme for asymmetric diamond networks.

Theorem 4. *There exists a universal constant $K < \infty$ such that for every diamond network with $N \geq 2$ relays and channel gains $(g_n)_{n=1}^N, (h_n)_{n=1}^N > 0$,*

$$\frac{C\big(N, (g_n), (h_n)\big)}{\sup_{\delta \in (0,1]} R_\delta\big(N, (g_n), (h_n)\big)} \leq K \log^4(N).$$

The proof of Theorem 4 is presented in Section IV-D. At a high level, achievability is proved as follows. Group the relays into classes such that all relays in the same class have approximately the same channel gains. Choose one such class, and set the constants $\alpha_n = 0$ for all relays not in this class (effectively disabling those relays). This relay-selection step reduces the original asymmetric network to a (almost) symmetric subnetwork. Theorem 1 can now be applied to this symmetric subnetwork to obtain a lower bound on the rate achievable with bursty amplify-and-forward. By maximizing over all possible classes, and hence all possible symmetric subnetworks, we get the largest rate achievable in this manner. The corresponding upper bound shows that this approach of relay selection combined with bursty amplify-and-forward is approximately optimal.

IV. Proofs

Sections IV-A, IV-B, and IV-C contain the proofs of Theorem 1 (achievability), Theorem 2 (upper bound), and Corollary 3 (approximation) for symmetric diamond networks. Section IV-D contains the proof of Theorem 4 for general asymmetric diamond networks.

A. Proof of Theorem 1 (Achievability for Symmetric Networks)

We start with a lemma computing the rate achievable by amplify-and-forward.

Lemma 5. *For every symmetric diamond network with $N \geq 2$ relays and channel gains $g, h > 0$, amplify-and-forward achieves*

$$R_1(N, g, h) = \tfrac{1}{2} \log\left(1 + \frac{N^2 gh}{1 + g + Nh}\right).$$

Proof: Recall that with amplify-and-forward relay v_n transmits

$$X_n[t] = \alpha Y_n[t] = \alpha \sqrt{g} X[t] + \alpha Z_n[t]$$

at time t, with constant $\alpha \geq 0$ chosen to satisfy the average unit power constraint. The received signal at the destination node w is

$$Y[t] = \alpha N \sqrt{gh} X[t] + \alpha \sqrt{h} \sum_{n=1}^{N} Z_n[t] + Z[t].$$

Observe that this describes a memoryless point-to-point channel with channel gain $\alpha N \sqrt{gh}$ and additive Gaussian noise of variance $1 + \alpha^2 N h$. $R_1(N, g, h)$ is the capacity of this point-to-point channel, optimized over the value of α.

For any value of $\alpha \geq 0$, the optimal distribution of the input X for this point-to-point channel is Gaussian with zero mean and variance one. The signal sent by the relays has power

$$\mathbb{E}(X_n^2) = \alpha^2 (1 + g),$$

and hence for

$$\alpha^2 \in \big[0, 1/(1+g)\big]$$

the average unit power constraints at the relays are satisfied. This implies that amplify-and-forward achieves a rate of

$$\begin{aligned} R_1(N,g,h) &= \max_{\alpha^2 \in [0,1/(1+g)]} \tfrac{1}{2}\log\Big(1 + \frac{\alpha^2 N^2 gh}{1+\alpha^2 Nh}\Big) \\ &= \tfrac{1}{2}\log\Big(1+\frac{N^2 gh}{1+g+Nh}\Big). \end{aligned}$$

■

The next lemma describes the rate achievable with bursty amplify-and-forward.

Lemma 6. *For every symmetric diamond network with* $N \geq 2$ *relays and channel gains* $g, h > 0$*, bursty amplify-and-forward with duty cycle* $\delta \in (0,1]$ *achieves*

$$R_\delta(N,g,h) = \tfrac{1}{2}\delta \log\Big(1 + \frac{N^2 gh/\delta^2}{1+g/\delta+Nh/\delta}\Big).$$

Proof: During the δ fraction of time we communicate, we are dealing with an equivalent channel with gains $\sqrt{g/\delta}$, $\sqrt{h/\delta}$ and with unit power constraints. The result now follows from Lemma 5 by taking into account that we only communicate a fraction δ of time. ■

Note that Lemmas 5 and 6 coincide for $\delta = 1$, as expected. We now proceed to the proof of Theorem 1. To simplify notation, set

$$R_\delta \triangleq R_\delta(N,g,h)$$

for $\delta \in (0,1]$.

We consider the cases $\max\{g, Nh\} \geq 1$ and $\max\{g, Nh\} < 1$ separately. Assume first $\max\{g, Nh\} \geq 1$. Here we set $\delta = 1$, i.e., we use simple amplify-and-forward. By Lemma 5

$$\begin{aligned} R_1 &= \tfrac{1}{2}\log\Big(1+\frac{N^2 gh}{1+g+Nh}\Big) \\ &= \tfrac{1}{2}\log\Big(1+\frac{N\min\{g,Nh\}\max\{g,Nh\}}{1+\min\{g,Nh\}+\max\{g,Nh\}}\Big) \\ &\geq \tfrac{1}{2}\log\Big(1+\frac{N\min\{g,Nh\}\max\{g,Nh\}}{3\max\{g,Nh\}}\Big) \\ &= \tfrac{1}{2}\log\big(1+\tfrac{1}{3}N\min\{g,Nh\}\big), \end{aligned}$$

where we have used that $1 \leq \max\{g, Nh\}$ to obtain the inequality.

Assume in the following that $\max\{g, Nh\} < 1$. We consider the cases $g \leq h$, $g \in (h, N^2 h)$, and $g \geq N^2 h$ separately. Consider first $g \leq h$. Bursty amplify-and-forward with duty cycle $\delta = Ng \leq Nh \leq 1$ achieves by Lemma 6

$$\begin{aligned} R_\delta &= \tfrac{1}{2}Ng\log\Big(1+\frac{N^2 gh/(N^2g^2)}{1+g/(Ng)+Nh/(Ng)}\Big) \\ &= \tfrac{1}{2}Ng\log\Big(1+\frac{h}{g+g/N+h}\Big) \\ &\overset{(a)}{\geq} \tfrac{1}{2}Ng\log\Big(1+\frac{h}{h+h/N+h}\Big) \\ &\geq \tfrac{1}{2}Ng\log(4/3) \\ &\geq \tfrac{1}{2}\ln(4/3)\log(1+Ng), \end{aligned}$$

where in (a) we used $g \leq h$.

Consider then $g \in (h, N^2h)$. If $N\sqrt{gh} \geq 1$, then simple amplify-and-forward achieves by Lemma 5

$$\begin{aligned} R_1 &= \tfrac{1}{2}\log\Big(1 + \frac{N^2gh}{1+g+Nh}\Big) \\ &\geq \tfrac{1}{2}\log\big(1 + \tfrac{1}{3}N^2gh\big), \end{aligned}$$

where we have used that $1 + g + Nh \leq 3$, which follows from $\max\{g, Nh\} \leq 1$.

Still assuming $g \in (h, N^2h)$, if $N\sqrt{gh} < 1$,[2] then bursty amplify-and-forward with duty cycle $\delta = N\sqrt{gh} \leq 1$ achieves by Lemma 6

$$\begin{aligned} R_\delta &= \tfrac{1}{2}N\sqrt{gh}\log\Big(1 + \frac{N^2gh/(N^2gh)}{1 + g/(N\sqrt{gh}) + Nh/(N\sqrt{gh})}\Big) \\ &= \tfrac{1}{2}N\sqrt{gh}\log\Big(1 + \frac{1}{1 + \sqrt{g}/(N\sqrt{h}) + \sqrt{h}/\sqrt{g}}\Big) \\ &\overset{(b)}{\geq} \tfrac{1}{2}N\sqrt{gh}\log\Big(1 + \frac{1}{1 + \sqrt{N^2h}/(N\sqrt{h}) + \sqrt{h}/\sqrt{h}}\Big) \\ &= \tfrac{1}{2}N\sqrt{gh}\log(4/3) \\ &\geq \tfrac{1}{2}\ln(4/3)\log\big(1 + N\sqrt{gh}\big), \end{aligned}$$

where in (b) we have used that $g \leq N^2h$ and $g \geq h$.

Consider finally $g \geq N^2h$. Bursty amplify-and-forward with duty cycle $\delta = N^2h \leq g \leq 1$ achieves by Lemma 6

$$\begin{aligned} R_\delta &= \tfrac{1}{2}N^2h\log\Big(1 + \frac{N^2gh/(N^4h^2)}{1 + g/(N^2h) + Nh/(N^2h)}\Big) \\ &= \tfrac{1}{2}N^2h\log\Big(1 + \frac{g/(N^2h)}{1 + g/(N^2h) + 1/N}\Big) \\ &\overset{(c)}{\geq} \tfrac{1}{2}N^2h\log(4/3) \\ &\geq \tfrac{1}{2}\ln(4/3)\log(1 + N^2h), \end{aligned}$$

where in (c) we have used that $1/N \leq 1 \leq g/(N^2h)$. ■

B. Proof of Theorem 2 (Upper Bound for Symmetric Networks)

In this section, we derive an upper bound on the capacity of the Gaussian diamond network. The standard way to find such bounds is to start with the cut-set bound and then to simplify it further to obtain a closed-form expression. The derivation here starts with the cut-set bound as well, but differs in several key aspects from the standard approach, which we now highlight.

Let

$$[N] \triangleq \{1, 2, \dots, N\},$$

and for a subset $S \subset [N]$, define

$$S^c \triangleq [N] \setminus S.$$

By the cut-set bound [18, Theorem 14.10.1],

$$C(N, g, h) \leq \sup_{X, X_{[N]}} \min_{S \subset [N]} I\big(X, X_S; Y, Y_{S^c} \bigm| X_{S^c}\big),$$

[2]Note that $g \in (h, N^2h)$ and $N\sqrt{gh} < 1$ imply $\max\{g, Nh\} < 1$.

where the maximization is over random variables $X, X_{[N]}$ satisfying the power constraints $\mathbb{E}(X^2) \leq 1$, $\mathbb{E}(X_n^2) \leq 1$, and where $X_{\tilde{S}} \triangleq (X_n)_{n\in\tilde{S}}$ for any subset $\tilde{S} \subset [N]$ (see Fig. 3 in Section III-A). A short calculation (done in (6) below) reveals that

$$\sup_{X,X_{[N]}} \min_{S\subset[N]} I\big(X, X_S; Y, Y_{S^c} \bigm| X_{S^c}\big) \leq \sup_{X,X_{[N]}} \min_{S\subset[N]} \Big(I\big(X; Y_{S^c}\big) + I\big(X_S; Y \bigm| X_{S^c}\big)\Big). \tag{1}$$

In the right-hand side of (1), the first mutual information corresponds to the rate between the source nodes and the relays, and the second mutual information corresponds to the rate between the relays and the destination node.

One approach is to simplify this expression further through a sequence of two steps. The first step is to upper bound

$$\begin{aligned} I\big(X_S; Y \bigm| X_{S^c}\big) &= \mathcal{H}\big(Y \bigm| X_{S^c}\big) - \mathcal{H}(Z) \\ &\leq \mathcal{H}\Big(\sqrt{h}\textstyle\sum_{n\in S} X_n + Z\Big) - \mathcal{H}(Z) \\ &= I\Big(X_S; \sqrt{h}\textstyle\sum_{n\in S} X_n + Z\Big), \end{aligned}$$

where, in order to avoid confusion with the channel gain h, we denote the differential entropy by the non-standard symbol $\mathcal{H}$. This first step thus removes the conditioning on the signals X_{S^c} available at the destination side of the cut. The second step is to interchange the order of maximization and minimization. This yields

$$\begin{aligned} C(N, g, h) &\leq \min_{S\subset[N]} \sup_{X,X_{[N]}} \Big(I\big(X; Y_{S^c}\big) + I\Big(X_S; \sqrt{h}\textstyle\sum_{n\in S} X_n + Z\Big)\Big) \\ &= \min_{n\in\{0,\dots,N\}} \big(\tfrac{1}{2}\log(1 + (N-n)g) + \tfrac{1}{2}\log(1 + n^2 h)\big). \end{aligned} \tag{2}$$

This can be further upper bounded by considering only $n = 0$ or $n = N$, resulting in the minimum of the broadcast and multiple-access cut

$$C(N, g, h) \leq \min\big\{\tfrac{1}{2}\log(1 + Ng), \tfrac{1}{2}\log(1 + N^2 h)\big\}. \tag{3}$$

Neither of the upper bounds (2) and (3) are tight enough to obtain a constant gap approximation of the capacity (this can be seen from the two examples presented after Corollary 3).

In this paper, we also start the derivation of the upper bound from the cut-set bound (1), but we avoid taking the two simplifying steps mentioned in the last paragraph. Instead, we first show, using the symmetry in the problem, that the correlation between any two signals X_n and $X_{\tilde{n}}$ with $n \neq \tilde{n}$ can be assumed to be equal without loss of optimality. Using the resulting simple form of the covariance matrix allows us then to evaluate the term $I\big(X_S; Y | X_{S^c}\big)$ directly. This enables us to keep the conditioning on X_{S^c}, which yields a significantly tighter upper bound on capacity. The resulting upper bound is summarized in the following lemma.

Lemma 7. *For every symmetric diamond network with $N \geq 2$ relays and channel gains $g, h > 0$, capacity is upper bounded as*

$$\begin{aligned} &C(N, g, h) \\ &\quad\leq \sup_{\rho\in[0,1)} \min_{n\in\{0,\dots,N\}} \Big(\tfrac{1}{2}\log(1 + (N-n)g) + \tfrac{1}{2}\log\Big(1 + n\Big(1 + (n-1)\rho - \frac{n(N-n)\rho^2}{1 + (N-n-1)\rho}\Big)h\Big)\Big). \end{aligned}$$

The variable ρ appearing in the lemma can be interpreted as the correlation between the random variables $X_{[N]}$ as mentioned in the preceding discussion. Note that it is not clear a priori that this correlation ρ can be restricted to be nonnegative. This restriction is part of the assertion of the lemma. We also point out that it is important that $\rho = 1$ is excluded from the supremum in Lemma 7; the result is not true without this restriction.

It will be convenient in the following to work with a weaker version of Lemma 7. Note that, for $\rho \in [0,1)$,

$$\begin{aligned}1+(n-1)\rho-\frac{n(N-n)\rho^2}{1+(N-n-1)\rho} &\leq 1+n\rho-\frac{n(N-n)\rho^2}{1+(N-n)\rho} \\ &= \Big(\frac{N}{N-n}\Big)\Big(\frac{\frac{N-n}{N}+(N-n)\rho}{1+(N-n)\rho}\Big) \\ &\leq \frac{N}{N-n}.\end{aligned}$$

Hence

$$C(N,g,h) \leq \min_{n\in\{0,\dots,N\}} \bigg(\tfrac{1}{2}\log(1+(N-n)g)+\tfrac{1}{2}\log\Big(1+\frac{N^2}{N-n}h\Big)\bigg). \tag{4}$$

The upper bound (4) derived from Lemma 7 can be compared to the simpler bound (2). If $n = KN$ for some constant $K \in (0,1)$, then the factor multiplying the channel gain h in (2) is of order $\Theta(N^2)$. On the other hand, the same factor in (4) is of order $\Theta(N)$. Thus, the bound (4) can be considerably tighter than the simpler bound (2).

Proof of Lemma 7: By the cut-set bound [18, Theorem 14.10.1],

$$C \triangleq C(N,g,h) \leq \sup_{X,X_{[N]}} \min_{S\subset[N]} I\big(X,X_S;Y,Y_{S^c} \bigm| X_{S^c}\big), \tag{5}$$

where, as before, the maximization is over random variables $X, X_{[N]}$ satisfying the power constraints $\mathbb{E}(X^2) \leq 1$, $\mathbb{E}(X_n^2) \leq 1$. We evaluate (5) in two steps. First, we argue that the maximization over $X, X_{[N]}$ can be restricted to jointly Gaussian random variables such that each $\mathbb{E}(X_n^2) = 1$ and $\mathbb{E}(X_n X_{\tilde{n}}) = \rho$ for $n \neq \tilde{n}$ and some $\rho \in [-1/(N-1), 1]$. This simplifies the maximization to be over just the parameter ρ instead of N-dimensional distributions. Second, using the resulting simple form of the input distributions, we analytically evaluate the mutual information in (5) to obtain the stated bound.

We start by simplifying the mutual information in (5) for a fixed cut $S \subset [N]$. We have

$$\begin{aligned}I\big(X,X_S;&Y,Y_{S^c} \bigm| X_{S^c}\big) \\ &= \mathcal{H}\big(Y,Y_{S^c} \bigm| X_{S^c}\big) - \mathcal{H}\big(Y,Y_{S^c} \bigm| X,X_{[N]}\big) \\ &= \mathcal{H}\big(Y_{S^c} \bigm| X_{S^c}\big) + \mathcal{H}\big(Y \bigm| Y_{S^c},X_{S^c}\big) - \mathcal{H}\big(Y_{S^c} \bigm| X,X_{[N]}\big) - \mathcal{H}\big(Y \bigm| Y_{S^c},X,X_{[N]}\big) \\ &\leq \mathcal{H}\big(Y_{S^c}\big) + \mathcal{H}\big(Y \bigm| X_{S^c}\big) - \mathcal{H}\big(Y_{S^c} \bigm| X\big) - \mathcal{H}\big(Y \bigm| X_{[N]}\big) \\ &= I\big(X;Y_{S^c}\big) + I\big(X_S;Y \bigm| X_{S^c}\big),\end{aligned} \tag{6}$$

where we have used that

$$\mathcal{H}\big(Y_{S^c} \bigm| X,X_{[N]}\big) = \mathcal{H}\big(Z_{S^c}\big) = \mathcal{H}\big(Y_{S^c} \bigm| X\big),$$

and that

$$\mathcal{H}\big(Y \bigm| Y_{S^c},X,X_{[N]}\big) = \mathcal{H}(Z) = \mathcal{H}\big(Y \bigm| X_{[N]}\big).$$

Combining (5) and (6) yields

$$C \leq \sup_{X,X_{[N]}} \min_{S\subset[N]} \Big(I\big(X;Y_{S^c}\big) + I\big(X_S;Y \bigm| X_{S^c}\big)\Big). \tag{7}$$

For the first term in (7),

$$I\big(X;Y_{S^c}\big) \leq \tfrac{1}{2}\log(1+|S^c|g), \tag{8}$$

since the channel from X to Y_{S^c} is a Gaussian single-input multiple-output channel with channel gains $\sqrt{g}$. For the second term in (7),

$$\begin{aligned} I\big(X_S; Y \bigm| X_{S^c}\big) &= \mathcal{H}\big(Y \bigm| X_{S^c}\big) - \mathcal{H}\big(Y \bigm| X_{[N]}\big) \\ &= \mathcal{H}\Big(\sqrt{h}\textstyle\sum_{n\in S}\big(X_n - \beta_n(X_{S^c})\big) + Z \Bigm| X_{S^c}\Big) - \mathcal{H}(Z) \\ &\leq \mathcal{H}\Big(\sqrt{h}\textstyle\sum_{n\in S}\big(X_n - \beta_n(X_{S^c})\big) + Z\Big) - \mathcal{H}(Z), \end{aligned} \tag{9}$$

for any choice of functions $\beta_n(X_{S^c})$ for $n \in S$. In particular, let $\beta_n(X_{S^c})$ be the minimum mean-square error estimator for X_n based on X_{S^c}.

Let $X_{[N]}$ have covariance matrix $\boldsymbol{Q}$. Then, by [19, Theorem 1.2.11], $(X_n - \beta_n(X_{S^c}))_{n\in S}$ has covariance matrix

$$\boldsymbol{Q}_{S|S^c} \triangleq \boldsymbol{Q}_{S,S} - \boldsymbol{Q}_{S,S^c}\boldsymbol{Q}^{-}_{S^c,S^c}\boldsymbol{Q}_{S^c,S}, \tag{10}$$

where, for any subsets $S_1, S_2 \subset [N]$, $\boldsymbol{Q}_{S_1,S_2}$ is the submatrix of $\boldsymbol{Q}$ induced by the rows S_1 and columns S_2, and where $\boldsymbol{Q}^{-}_{S^c,S^c}$ is the *Moore-Penrose generalized inverse* of the matrix $\boldsymbol{Q}_{S^c,S^c}$. The matrix $\boldsymbol{Q}_{S|S^c}$ is called the *generalized Schur complement* of $\boldsymbol{Q}_{S^c,S^c}$ in $\boldsymbol{Q}$. Note that if $\boldsymbol{Q}_{S^c,S^c}$ is invertible, then $\boldsymbol{Q}^{-}_{S^c,S^c} = \boldsymbol{Q}^{-1}_{S^c,S^c}$ and the generalized Schur complement reduces to the standard Schur complement.

Before proceeding, we need to introduce some notation. Denote by $\boldsymbol{I}_a$ the $a \times a$ identity matrix, and by $\mathbf{1}_{a,b}$ the $a \times b$ matrix of ones. To simplify notation, we will write $\mathbf{1}$ for the column vector $\mathbf{1}_{a,1}$, whenever the dimension is clear from the context. With these definitions,

$$\mathcal{H}\Big(\sqrt{h}\textstyle\sum_{n\in S}\big(X_n - \beta_n(X_{S^c})\big) + Z\Big) - \mathcal{H}(Z) \leq \frac{1}{2}\log\big(1 + h\mathbf{1}^T\boldsymbol{Q}_{S|S^c}\mathbf{1}\big). \tag{11}$$

Substituting (8), (9), and (11) into (7) yields

$$C \leq \sup_{\substack{\boldsymbol{Q}\geq 0: \\ q_{n,n}\leq 1 \forall n\in[N]}} \min_{S\subset[N]} \Big(\tfrac{1}{2}\log(1+|S^c|g) + \tfrac{1}{2}\log\big(1 + h\mathbf{1}^T\boldsymbol{Q}_{S|S^c}\mathbf{1}\big)\Big),$$

where $\boldsymbol{Q} \geq 0$ denotes that $\boldsymbol{Q}$ is a positive semi-definite matrix. We have thus simplified the maximization over input distributions to a maximization over covariance matrices. The next step is to show that the covariance matrix $\boldsymbol{Q}$ can be restricted without loss of optimality to have the form

$$\rho\mathbf{1}_{N,N} + (1-\rho)\boldsymbol{I}_N,$$

and hence the maximization over covariance matrices can be further simplified to a maximization over just the scalar correlation parameter ρ.[3]

For convenience of notation, define

$$\psi_S(\boldsymbol{Q}) \triangleq \tfrac{1}{2}\log\big(1 + h\mathbf{1}^T\boldsymbol{Q}_{S|S^c}\mathbf{1}\big)$$

and

$$\psi(\boldsymbol{Q}) \triangleq \min_{S\subset[N]} \Big(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q})\Big),$$

so that

$$C \leq \sup_{\substack{\boldsymbol{Q}\geq 0: \\ q_{n,n}\leq 1 \forall n\in[N]}} \psi(\boldsymbol{Q}). \tag{12}$$

[3]Upon completion of this work, we realized that a somewhat similar argument as in this step was used in [20, Section III] for the Gaussian multiple-access channel with feedback.

Consider a covariance matrix $\boldsymbol{Q} \geq 0$, and let $\boldsymbol{P}$ be any permutation matrix on $[N]$. Note that $\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P} \geq 0$. Moreover, by symmetry,[4]

$$\begin{aligned}\psi(\boldsymbol{Q}) &= \min_{S\subset[N]} \left(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q})\right)\\ &= \min_{S\subset[N]} \left(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S\big(\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}\big)\right)\\ &= \psi\big(\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}\big),\end{aligned}$$

and thus $\psi(\cdot)$ is invariant under permutation.

Now, the generalized Schur complement is matrix-concave over the set of positive semi-definite matrices [21, Theorem 3.1] (see also [22, p. 469] for the corresponding result for positive definite matrices). More precisely, if $\boldsymbol{Q} = \lambda_1\boldsymbol{Q}^1 + \lambda_2\boldsymbol{Q}^2$ with $\lambda_1 \in [0,1]$, $\lambda_2 = 1-\lambda_1$, then

$$\boldsymbol{Q}_{S|S^c} \geq \lambda_1\boldsymbol{Q}^1_{S|S^c} + \lambda_2\boldsymbol{Q}^2_{S|S^c},$$

i.e.,

$$\boldsymbol{Q}_{S|S^c} - \big(\lambda_1\boldsymbol{Q}^1_{S|S^c} + \lambda_2\boldsymbol{Q}^2_{S|S^c}\big)$$

is a positive semi-definite matrix. Therefore,

$$\mathbf{1}^T\big(\boldsymbol{Q}_{S|S^c} - \big(\lambda_1\boldsymbol{Q}^1_{S|S^c} + \lambda_2\boldsymbol{Q}^2_{S|S^c}\big)\big)\mathbf{1} \geq 0,$$

implying that

$$\begin{aligned}\tfrac{1}{2}\log\big(1+h\mathbf{1}^T\boldsymbol{Q}_{S|S^c}\mathbf{1}\big) &\geq \tfrac{1}{2}\log\big(1+\lambda_1 h\mathbf{1}^T\boldsymbol{Q}^1_{S|S^c}\mathbf{1} + \lambda_2 h\mathbf{1}^T\boldsymbol{Q}^2_{S|S^c}\mathbf{1}\big)\\ &\geq \lambda_1\tfrac{1}{2}\log\big(1+h\mathbf{1}^T\boldsymbol{Q}^1_{S|S^c}\mathbf{1}\big) + \lambda_2\tfrac{1}{2}\log\big(1+h\mathbf{1}^T\boldsymbol{Q}^2_{S|S^c}\mathbf{1}\big).\end{aligned}$$

Thus $\psi_S(\boldsymbol{Q})$ is concave in $\boldsymbol{Q}$. Finally,

$$\begin{aligned}&\min_{S\subset[N]} \left(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q})\right)\\ &\quad\geq \min_{S\subset[N]} \Big(\lambda_1\big(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q}^1)\big) + \lambda_2\big(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q}^2)\big)\Big)\\ &\quad\geq \lambda_1 \min_{S\subset[N]} \big(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q}^1)\big) + \lambda_2 \min_{S\subset[N]} \big(\tfrac{1}{2}\log(1+|S^c|g) + \psi_S(\boldsymbol{Q}^2)\big),\end{aligned}$$

and hence $\psi(\boldsymbol{Q})$ is also concave in $\boldsymbol{Q}$.

Fix $\varepsilon > 0$, and assume that $\boldsymbol{Q}^\star$ achieves ε-optimality, i.e., $\boldsymbol{Q}^\star \geq 0$, $q^\star_{n,n} \leq 1$ for all $n \in [N]$, and

$$\psi(\boldsymbol{Q}^\star) \geq \sup_{\substack{\boldsymbol{Q}\geq 0:\\ q_{n,n}\leq 1 \forall n\in[N]}} \psi(\boldsymbol{Q}) - \varepsilon.$$

Set

$$\boldsymbol{Q} = \frac{1}{N!}\sum_{\boldsymbol{P}} \boldsymbol{P}^T\boldsymbol{Q}^\star\boldsymbol{P},$$

where the sum is over all $N!$ permutation matrices on $[N]$.

Note that $\boldsymbol{Q}$ is positive semi-definite and satisfies $q_{n,n} \leq 1$ for all $n \in [N]$. Moreover, using the concavity and invariance under permutation of $\psi(\cdot)$, we obtain

$$\begin{aligned}\psi(\boldsymbol{Q}) &\geq \frac{1}{N!}\sum_{\boldsymbol{P}} \psi\big(\boldsymbol{P}^T\boldsymbol{Q}^\star\boldsymbol{P}\big)\\ &= \psi(\boldsymbol{Q}^\star),\end{aligned}$$

[4]Note that the minimization over $S \subset [N]$ is crucial for this fact to hold. Indeed, $\psi_S(\boldsymbol{Q}) \neq \psi_S\big(\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}\big)$ in general.

and hence $\boldsymbol{Q}$ is also an ε-optimal covariance matrix. Note that this $\boldsymbol{Q}$ has the form

$$\rho \mathbf{1}_{N,N} + \kappa \boldsymbol{I}_N,$$

for $\kappa \leq 1-\rho]$, and thus we can restrict the maximization of $\psi(\boldsymbol{Q})$ to matrices of this form. Since the generalized Schur complement is monotonically increasing over the set of positive semi-definite matrices [21, Theorem 3.1], we can further restrict the value of κ to be $1-\rho$. Denote the resulting matrix by $\boldsymbol{Q}^{\rho}$, i.e.,

$$\boldsymbol{Q}^{\rho} \triangleq \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}.$$

Note that $\boldsymbol{Q}^{\rho}$ is positive semi-definite only if $\rho \in [-1/(N-1), 1]$ (since otherwise the eigenvalue corresponding to the eigenvector $\mathbf{1}$ is negative).

The upper bound on capacity in (12) can thus be simplified to

$$\begin{aligned} C &\leq \sup_{\rho \in [-1/(N-1),1]} \min_{S \subset [N]} \left(\tfrac{1}{2}\log(1+|S^c|g) + \tfrac{1}{2}\log\left(1+h\mathbf{1}^T \boldsymbol{Q}^{\rho}_{S|S^c}\mathbf{1}\right)\right) \\ &= \sup_{\rho \in [-1/(N-1),1]} \min_{n \in \{0,\dots,N\}} \left(\tfrac{1}{2}\log(1+(N-n)g) + \tfrac{1}{2}\log\left(1+h\mathbf{1}^T \boldsymbol{Q}^{\rho}_{[n]|[n]^c}\mathbf{1}\right)\right), \end{aligned} \tag{13}$$

where $[0]$ is understood as the empty set and $[0]^c \triangleq [N]$. Observe that the minimization in (13) is over integers $n \in \{0,\dots,N\}$ as opposed to subsets $S \subset [N]$ due to the symmetry in $\boldsymbol{Q}^{\rho}$. Note furthermore that instead of maximizing over arbitrary input distributions, we only have to maximize over the single real number ρ.

We now compute the expression in parentheses in (13) analytically. To this end, we need to compute $\boldsymbol{Q}^{\rho}_{[n]|[n]^c}$, which, by (10), involves the computation of the generalized inverse $(\boldsymbol{Q}^{\rho}_{[n]^c,[n]^c})^{-}$. We will first consider the case when $\boldsymbol{Q}^{\rho}_{[n]^c,[n]^c}$ is invertible, and then consider the remaining cases in which $\boldsymbol{Q}^{\rho}_{[n]^c,[n]^c}$ is not invertible. If $n \in \{1,\dots,N-1\}$ and $\rho \in [-1/(N-1),1)$, then $\boldsymbol{Q}^{\rho}_{[n]^c,[n]^c}$ is invertible, and after some algebra, we obtain

$$\mathbf{1}^T \boldsymbol{Q}^{\rho}_{[n]|[n]^c}\mathbf{1} = n\Big(1+(n-1)\rho - \frac{n(N-n)\rho^2}{1+(N-n-1)\rho}\Big). \tag{14}$$

We now consider the remaining cases, in which $\boldsymbol{Q}^{\rho}_{[n]^c,[n]^c}$ is not invertible. If $\rho = 1$ and $n \in \{1,\dots,N-1\}$, then

$$\mathbf{1}^T \boldsymbol{Q}^{1}_{[n]|[n]^c}\mathbf{1} = 0. \tag{15}$$

If $n = 0$, then

$$\mathbf{1}^T \boldsymbol{Q}^{\rho}_{[0]|[0]^c}\mathbf{1} = 0, \tag{16}$$

and if $n = N$, then

$$\mathbf{1}^T \boldsymbol{Q}^{\rho}_{[N]|[N]^c}\mathbf{1} = \mathbf{1}^T \boldsymbol{Q}^{\rho}\mathbf{1} = N(1+(N-1)\rho), \tag{17}$$

both for any $\rho \in [-1/(N-1), 1]$.

Denote by

$$\eta(\rho,n) \triangleq n\Big(1+(n-1)\rho - \frac{n(N-n)\rho^2}{1+(N-n-1)\rho}\Big)$$

the right-hand side of (14). Note that $\eta(\rho,n)$ is well defined for all $\rho \in [-1/(N-1),1]$, $n \in \{0,\dots,N\}$ except for $\eta(1,N)$ and $\eta(-1/(N-1),0)$ (for which the expression involves dividing zero by zero).

Moreover, from (14)–(17) we see that whenever $\eta(\rho, n)$ is well defined, it is equal to $\mathbf{1}^T \boldsymbol{Q}^{\rho}_{[n]|[n]^c}\mathbf{1}$. For the two cases in which $\eta(\rho, n)$ is not well defined, we have from (15)–(17) that for any $n \in [N]$,

$$\lim_{\rho \uparrow 1} \eta(\rho, n) = \mathbf{1}^T \boldsymbol{Q}^{1}_{[n]|[n]^c}\mathbf{1}, \tag{18a}$$

$$\lim_{\rho \downarrow -1/(N-1)} \eta(\rho, n) = \mathbf{1}^T \boldsymbol{Q}^{-1/(N-1)}_{[n]|[n]^c}\mathbf{1}, \tag{18b}$$

and in particular this holds for $n = 0$ and $n = N$. Thus we can write $\mathbf{1}^T \boldsymbol{Q}^{\rho}_{[n]|[n]^c}\mathbf{1}$ compactly as a function of $\eta(\rho, n)$ for any $n \in [N]$ and do not need to consider the boundary cases $n = 0$, $n = N$, and $\rho = -1/(N-1)$, $\rho = 1$ separately. Substituting (14) and (18) into (13), we obtain

$$C \leq \sup_{\rho \in (-1/(N-1),1)} \min_{n \in \{0,\dots,N\}} \Bigg(\tfrac{1}{2}\log(1+(N-n)g) + \tfrac{1}{2}\log\bigg(1+n\Big(1+(n-1)\rho - \frac{n(N-n)\rho^2}{1+(N-n-1)\rho}\Big)h\bigg)\Bigg). \tag{19}$$

Observe that the supremum in (19) is only over $\rho \in (-1/(N-1), 1)$ as opposed to $\rho \in [-1/(N-1), 1]$ as in (13).

We finally argue that the supremum can be restricted to be over values $\rho \in [0, 1)$. Consider the derivative with respect to ρ of the multiplier of the h term in (19),

$$\frac{d}{d\rho} n\Big(1 + (n-1)\rho - \frac{n(N-n)\rho^2}{1+(N-n-1)\rho}\Big) = n\Big((n-1) - n(N-n)\rho \frac{2+(N-n-1)\rho}{\big(1+(N-n-1)\rho\big)^2}\Big).$$

If $\rho \in (-1/(N-1), 0)$, then this derivative is non-negative, and thus the multiplier of h in (19) is non-decreasing in that range of ρ. Since this is true simultaneously for all $n \in \{0, \dots, N\}$, we can restrict the supremum to be over the range $\rho \in [0, 1)$. This proves the lemma. ■

We now proceed to the proof of Theorem 2. As before, we denote the capacity of the diamond network by

$$C \triangleq C(N, g, h).$$

We again consider the cases $\max\{g, Nh\} \geq 1$ and $\max\{g, Nh\} < 1$ separately. Assume first that $\max\{g, Nh\} \geq 1$. Capacity is upper bounded by the minimum of the simple broadcast and multiple-access cuts

$$\begin{aligned} C &\leq \min\left\{\tfrac{1}{2}\log(1+Ng), \tfrac{1}{2}\log(1+N^2h)\right\} \\ &= \tfrac{1}{2}\log\big(1+N\min\{g, Nh\}\big). \end{aligned} \tag{20}$$

Observe that (20) is valid regardless of the value of $\max\{g, Nh\}$.

Assume in the following that $\max\{g, Nh\} < 1$. As before, we treat the cases $g \leq h$, $g \in (h, N^2h)$, and $g \geq N^2h$ separately. Consider first $g \leq h$. Using the upper bound in (20), we obtain

$$C \leq \tfrac{1}{2}\log\big(1+Ng\big).$$

Consider then $g \in (h, N^2h)$. If $N\sqrt{gh} \geq 1$, then the simplified form (4) of Lemma 7 with $N - n = \lceil N^2h \rceil \in \{0, \dots, N\}$ (since $Nh \leq 1$) yields

$$\begin{aligned} C &\leq \tfrac{1}{2}\log(1+(N-n)g) + \tfrac{1}{2}\log\Big(1+\frac{N^2}{N-n}h\Big) \\ &= \tfrac{1}{2}\log(1+\lceil N^2h\rceil g) + \tfrac{1}{2}\log\left(1+\frac{N^2}{\lceil N^2h\rceil}h\right) \\ &\leq \tfrac{1}{2}\log(1+g+N^2gh) + \tfrac{1}{2} \\ &\leq \tfrac{1}{2}\log(1+2N^2gh) + \tfrac{1}{2}, \end{aligned}$$

where we have used that $g \leq N^2 gh$ since

$$N^2 h \geq \sqrt{g}\sqrt{N^2 h} = N\sqrt{gh} \geq 1.$$

Still assuming $g \in (h, N^2 h)$, if $N\sqrt{gh} < 1$, then the simplified form (4) of Lemma 7 with $N - n = \lceil N\sqrt{h/g}\,\rceil \in \{0, \dots, N\}$ (since $g \geq h$ and hence $\sqrt{h/g} \leq 1$) shows that

$$\begin{aligned}
C &\leq \tfrac{1}{2}\log(1 + (N-n)g) + \tfrac{1}{2}\log\Big(1 + \frac{N^2}{N-n}h\Big) \\
&= \tfrac{1}{2}\log\big(1 + \lceil N\sqrt{h/g}\,\rceil g\big) + \tfrac{1}{2}\log\bigg(1 + \frac{N^2}{\lceil N\sqrt{h/g}\,\rceil}h\bigg) \\
&\leq \tfrac{1}{2}\log\big(1 + g + N\sqrt{gh}\big) + \tfrac{1}{2}\log\big(1 + N\sqrt{gh}\big) \\
&\leq \log\big(1 + 2N\sqrt{gh}\big),
\end{aligned}$$

where we have used that

$$g \leq \sqrt{g}\sqrt{N^2 h} = N\sqrt{gh}.$$

Finally, consider $g \geq N^2 h$. The upper bound (20) yields

$$C \leq \tfrac{1}{2}\log\big(1 + N^2 h\big),$$

concluding the proof. ■

C. Proof of Corollary 3 (Capacity Approximation for Symmetric Networks)

The corollary follows directly from Theorems 1 and 2 using the inequalities

$$\log(1 + ax) \begin{cases} \geq a\log(1+x), & \text{for } a \in [0,1], x \geq 0 \\ \leq a\log(1+x), & \text{for } a \geq 1, x \geq 0, \end{cases}$$

and

$$\log(1 + ax) \begin{cases} \geq \log(a) + \log(1+x), & \text{for } a \in [0,1], x \geq 0 \\ \leq \log(a) + \log(1+x), & \text{for } a \geq 1, x \geq 0. \end{cases}$$

■

D. Proof of Theorem 4 (Capacity Approximation for Asymmetric Networks)

The idea of the proof is as follows. Group the relays into classes such that all relays in the same class have approximately the same channel gains. We argue that the number of classes needed is on the order $\Theta(\log^2(N))$. Choose one such class, and set the constants $\alpha_n = 0$ for all relays not in this class. This effectively reduces the network to a (almost) symmetric one, which we have already analyzed in the earlier parts of this paper. By maximizing over which class to choose, we get the largest rate achievable in this manner. This yields a lower bound on $R_\delta\big(N, (g_n), (h_n)\big)$. We then argue that this approach is close to optimal, by showing that capacity $C\big(N, (g_n), (h_n)\big)$ is upper bounded by $\Theta(\log^4(N))$ times the maximum of the capacities of these classes.

Recall the notation

$$[N] \triangleq \{1, \dots, N\}$$

and, for $S \subset [N]$,

$$S^c \triangleq [N] \setminus S.$$

Furthermore, in this section, we will use

$$\begin{aligned}
g_S &\triangleq (g_n)_{n \in S}, \\
h_S &\triangleq (h_n)_{n \in S}
\end{aligned}$$

for $S \subset [N]$, and

$$ah_S \triangleq (ah_n)_{n\in S}$$

for scalar $a \in \mathbb{R}$.

We want to partition $[N]$ into subsets such that for n and $\tilde{n}$ in the same subset the relays v_n and $v_{\tilde{n}}$ have approximately the same channel gains. Moreover, we want the number of required subsets to be small. This is not directly possible if the channel gains are very different. For example, consider $g_n = h_n = 2^n$; note, however, that in this case most of the relays are very weak compared to the strongest one, and could hence be disregarded without too much loss in rate. We formalize this idea by allowing some "overload" subsets (in the language of quantization theory) in the partition of $[N]$, which correspond to relays that may have very different channel gains, but that are all too weak to have much impact on achievable rates.

Define

$$g^\star \triangleq \max_{n\in[N]} \min\{g_n, N^2 h_n\},$$
$$h^\star \triangleq \max_{n\in[N]} \min\{h_n, g_n\}.$$

The quantities $g^\star$ and $h^\star$ are essentially the largest channel gains, accounting for situations in which one of the channel gains g_n, h_n clearly dominates the other one. If we let n be such that $h^\star = \min\{h_n, g_n\}$, then

$$g^\star \geq \min\{g_n, N^2 h_n\} \geq \min\{g_n, h_n\} = h^\star. \tag{21}$$

Similarly, if n is such that $g^\star = \min\{g_n, N^2 h_n\}$, then

$$h^\star \geq \min\{h_n, g_n\} \geq N^{-2} \min\{N^2 h_n, g_n\} = N^{-2} g^\star. \tag{22}$$

Thus, $g^\star$ and $h^\star$ can not be too different.

We are now ready to introduce the partition of $[N]$ mentioned above. We start with the "overload" subsets. Define the sets

$$T^1 \triangleq \big\{n \in [N] : g_n \leq N^{-3} g^\star\big\},$$
$$T^2 \triangleq \big\{n \in [N] \setminus T^1 : h_n \leq N^{-3} h^\star\big\},$$

i.e., T^1 and T^2 correspond to those relays that have channel gains that are very weak compared to the strongest one in the network. Set

$$L \triangleq \lfloor 3 \log(N) \rfloor.$$

For $\ell \in \{0, \ldots, L\}$, define

$$T^1_\ell \triangleq \big\{n \in [N] \setminus (T^1 \cup T^2) : g_n \in (2^{-\ell-1} g^\star, 2^{-\ell} g^\star], h_n \geq g_n, \big\},$$
$$T^2_\ell \triangleq \big\{n \in [N] \setminus (T^1 \cup T^2 \cup_{\tilde{\ell}} T^1_{\tilde{\ell}}) : g_n \geq N^2 h_n, h_n \in (2^{-\ell-1} h^\star, 2^{-\ell} h^\star]\big\},$$

i.e., $\{T^1_\ell\}$ and $\{T^2_\ell\}$ quantize those channel gains for which one of g_n, h_n dominates the other one. Finally, define for $k, \ell \in \{0, \ldots, L\}$,

$$S_{k,\ell} \triangleq \big\{n \in [N] \setminus \big(T^1 \cup T^2 \cup_{\tilde{\ell}} (T^1_{\tilde{\ell}} \cup T^2_{\tilde{\ell}})\big) : g_n \in (2^{-k-1} g^\star, 2^{-k} g^\star], h_n \in (2^{-\ell-1} h^\star, 2^{-\ell} h^\star]\big\}.$$

The subsets $\{S_{k,\ell}\}$ quantize the remaining channel gains. The number of sets T^1, T^2, $\{T^1_\ell\}$, $\{T^2_\ell\}, \{S_{k,\ell}\}$ is equal to

$$\tilde{L} \triangleq (L+1)^2 + 2(L+1) + 2 = \Theta(\log^2(N)).$$

We argue that T^1, T^2, $\{T^1_\ell\}$, $\{T^2_\ell\}, \{S_{k,\ell}\}$ partition $[N]$. The sets are clearly disjoint, so we only need to show that their union covers $[N]$. If either $g_n \leq N^{-3} g^\star$ or $h_n \leq N^{-3} h^\star$ then $n \in T^1 \cup T^2$. Assume in the following discussion that $g_n > N^{-3}g^\star$ and $h_n > N^{-3}h^\star$. If $g_n \leq g^\star$ and $h_n \leq h^\star$, then n is an element of $\{T_\ell^1\}$, $\{T_\ell^2\}$, or $\{S_{k,\ell}\}$. If $g_n > g^\star$, then

$$h_n \leq N^2 h_n \leq g^\star \leq g_n,$$

so that $g_n \geq N^2 h_n$ and $h_n = \min\{h_n, g_n\} \leq h^\star$. This implies that $n \in \cup_\ell T_\ell^2$. If $h_n > h^\star$, then

$$g_n \leq h^\star \leq h_n,$$

so that $h_n \geq g_n$ and $g_n = \min\{g_n, N^2 h_n\} \leq g^\star$. This implies that $n \in \cup_\ell T_\ell^1$. Together, this proves that we have properly partitioned $[N]$.

We are now ready for the proof of the upper bound on capacity. We argue that the capacity of the diamond network with N relays cannot be much larger than the sum of the capacities of the $\tilde{L}$ subchannels induced by the partition of $[N]$ defined above. Formally, we argue that

$$C\big(N, g_{[N]}, h_{[N]}\big) \leq C\big(|T^1|, g_{T^1}, 2\tilde{L}h_{T^1}\big) + C\big(|T^2|, g_{T^2}, 2h_{T^2}\big) + \sum_{i=1}^{2}\sum_{\ell=0}^{L} C\big(|T_\ell^i|, g_{T_\ell^i}, 2\tilde{L}h_{T_\ell^i}\big) + \sum_{k,\ell=0}^{L} C\big(|S_{k,\ell}|, g_{S_{k,\ell}}, 2\tilde{L}h_{S_{k,\ell}}\big). \tag{23}$$

To see this, note that the right-hand side is the capacity of $\tilde{L}$ parallel diamond networks each with unit input power constraint. Moreover, increasing each channel gain $\sqrt{h_n}$ by a factor of $\sqrt{2\tilde{L}}$ (or $\sqrt{2}$ in the case of T^2) is equivalent to reducing the power of the additive noise at the destination node of the parallel networks by a factor $1/(2\tilde{L})$ (or $1/2$ for T^2). We can now use these parallel networks to simulate the original N-relay diamond network by forcing the input (at the source node u) to all the parallel networks to be identical, and by summing up the outputs (at the destination node w) of the parallel networks. This proves (23).

Next, we argue that the capacities of the asymmetric subnetworks in (23) can be upper bounded by the capacities of symmetric diamond networks. Consider the subset $S_{k,\ell}$. Since capacity is increasing in the channel gains,

$$C\big(|S_{k,\ell}|, g_{S_{k,\ell}}, 2\tilde{L}h_{S_{k,\ell}}\big) \leq C\big(|S_{k,\ell}|, 2^{-k}g^\star, \tilde{L}2^{1-\ell}h^\star\big). \tag{24}$$

Observe that the right-hand side is the capacity of a *symmetric* diamond network. Consider then T_ℓ^i. By the same argument

$$C\big(|T_\ell^1|, g_{T_\ell^1}, 2\tilde{L}h_{T_\ell^1}\big) \leq C\big(|T_\ell^1|, 2^{-\ell}g^\star, \infty\big), \tag{25}$$

and

$$C\big(|T_\ell^2|, g_{T_\ell^2}, 2\tilde{L}h_{T_\ell^2}\big) \leq C\big(|T_\ell^2|, \infty, \tilde{L}2^{1-\ell}h^\star\big). \tag{26}$$

It remains to consider T^1 and T^2. For the set T^1, we have

$$C\big(|T^1|, g_{T^1}, 2\tilde{L}h_{T^1}\big) \leq C\big(N, N^{-3}g^\star, \infty\big).$$

From Theorem 2,

$$C\big(N, N^{-3}g^\star, \infty\big) \leq \tfrac{1}{2}\log(1 + N^{-2}g^\star).$$

By the definition of $g^\star$, there exists at least one n such that $g_n \geq g^\star$ and $h_n \geq N^{-2}g^\star$. Using just this one relay v_n, a rate of at least

$$\tfrac{1}{2}\log(1 + N^{-2}g^\star)$$

is achievable.[5] For this n, we have

$$g_n \geq g^\star > N^{-3}g^\star,$$

[5]This rate is achievable, for example, with decode-and-forward. Note that we use decode-and-forward here only as a proof technique to obtain the upper bound on capacity. Achievability is based exclusively on (bursty) amplify-and-forward.

and hence $n \notin T^1$. Moreover, using (21),

$$h_n \geq N^{-2} g^\star \geq N^{-2} h^\star > N^{-3} h^\star,$$

and hence $n \notin T^2$. This n is therefore an element of one of the subsets $\{T^1_\ell\}$, $\{T^2_\ell\}$, $\{S_{k,\ell}\}$, and we obtain from (24)–(26),

$$\begin{aligned} C\big(\lvert T^1\rvert, g_{T^1}, \tilde{L} h_{T^1}\big) \leq \max\Big\{ & \max_{\ell\in\{0,\ldots,L\}} C\big(\lvert T^1_\ell\rvert, 2^{-\ell} g^\star, \infty\big), \max_{\ell\in\{0,\ldots,L\}} C\big(\lvert T^2_\ell\rvert, \infty, \tilde{L} 2^{1-\ell} h^\star\big), \\ & \max_{k,\ell\in\{0,\ldots,L\}} C\big(\lvert S_{k,\ell}\rvert, 2^{-k} g^\star, \tilde{L} 2^{1-\ell} h^\star\big)\Big\}. \end{aligned} \tag{27}$$

Similarly,

$$\begin{aligned} C\big(\lvert T^2\rvert, g_{T^2}, 2h_{T^2}\big) & \leq C\big(N, \infty, 2N^{-3} h^\star\big) \\ & \leq \tfrac{1}{2}\log(1 + 2N^{-1} h^\star) \\ & \leq \tfrac{1}{2}\log(1 + h^\star). \end{aligned}$$

By the definition of $h^\star$ there exists at least one n such that $h_n \geq h^\star$ and $g_n \geq h^\star$. Using this relay v_n alone, we achieve at least a rate of

$$\tfrac{1}{2}\log(1 + h^\star).$$

For this n,

$$h_n \geq h^\star > N^{-3} h^\star,$$

and hence $n \notin T^2$. Moreover, using (22),

$$g_n \geq h^\star \geq N^{-2} g^\star > N^{-3} g^\star,$$

and hence $n \notin T^1$. This n is therefore an element of one of the subsets $\{T^1_\ell\}$, $\{T^2_\ell\}$, $\{S_{k,\ell}\}$, and we obtain again from (24)–(26),

$$\begin{aligned} C\big(\lvert T^2\rvert, g_{T^2}, 2h_{T^2}\big) \leq \max\Big\{ & \max_{\ell\in\{0,\ldots,L\}} C\big(\lvert T^1_\ell\rvert, 2^{-\ell} g^\star, \infty\big), \max_{\ell\in\{0,\ldots,L\}} C\big(\lvert T^2_\ell\rvert, \infty, \tilde{L} 2^{1-\ell} h^\star\big), \\ & \max_{k,\ell\in\{0,\ldots,L\}} C\big(\lvert S_{k,\ell}\rvert, 2^{-k} g^\star, \tilde{L} 2^{1-\ell} h^\star\big)\Big\}. \end{aligned} \tag{28}$$

Substituting (24)–(28) into (23), we obtain

$$\begin{aligned} C\big(N, g_{[N]}, h_{[N]}\big) \leq \tilde{L} \max\Big\{ & \max_{\ell\in\{0,\ldots,L\}} C\big(\lvert T^1_\ell\rvert, 2^{-\ell} g^\star, \infty\big), \max_{\ell\in\{0,\ldots,L\}} C\big(\lvert T^2_\ell\rvert, \infty, \tilde{L} 2^{1-\ell} h^\star\big), \\ & \max_{k,\ell\in\{0,\ldots,L\}} C\big(\lvert S_{k,\ell}\rvert, 2^{-k} g^\star, \tilde{L} 2^{1-\ell} h^\star\big)\Big\}. \end{aligned} \tag{29}$$

This concludes the proof of the upper bound on capacity.

We continue with the proof of achievability. Fix $k, \ell \in \{0, \ldots, L\}$, and recall that α_n is the constant determining the amplification at relay v_n. Assume we set $\alpha_n = 0$ for all $n \notin S_{k,\ell}$. This results in a network in which all but the relays in $S_{k,\ell}$ are removed. Thus

$$R_\delta\big(N, g_{[N]}, h_{[N]}\big) \geq R_\delta\big(\lvert S_{k,\ell}\rvert, g_{S_{k,\ell}}, h_{S_{k,\ell}}\big). \tag{30}$$

Moreover, since R_δ is increasing in the channel gains,

$$R_\delta\big(\lvert S_{k,\ell}\rvert, g_{S_{k,\ell}}, h_{S_{k,\ell}}\big) \geq R_\delta\big(\lvert S_{k,\ell}\rvert, 2^{-k-1} g^\star, 2^{-\ell-1} h^\star\big). \tag{31}$$

With this, we have lower bounded the rate achievable for the asymmetric diamond network by the one of a symmetric diamond network (with fewer relays and smaller channel gains). We can thus apply the results from Section III-A to obtain

$$\sup_{\delta\in(0,1]} R_\delta\big(|S_{k,\ell}|, 2^{-k-1}g^\star, 2^{-\ell-1}h^\star\big) \geq \frac{1}{112\tilde{L}} C\big(|S_{k,\ell}|, 2^{-k}g^\star, \tilde{L}2^{1-\ell}h^\star\big), \tag{32}$$

where the factor $1/(112\tilde{L}) = 1/(8\tilde{L}\times 14)$ is composed of a factor $8\tilde{L}$ to offset the increase of the channel gains to the relay by two and the increase of the channel gains from the relays by $4\tilde{L}$ (see Theorem 1) and of a factor 14 to go from rate achievable with bursty amplify-and-forward to capacity (see Theorem 2 and Corollary 3).

Combining (30), (31), and (32) yields

$$\sup_{\delta\in(0,1]} R_\delta\big(N, g_{[N]}, h_{[N]}\big) \geq \frac{1}{112\tilde{L}} C\big(|S_{k,\ell}|, 2^{-k}g^\star, \tilde{L}2^{1-\ell}h^\star\big). \tag{33}$$

A similar argument, setting $\alpha_n = 0$ for n outside T_ℓ^i, shows that

$$\begin{aligned} \sup_{\delta\in(0,1]} R_\delta\big(N, g_{[N]}, h_{[N]}\big) &\geq \sup_{\delta\in(0,1]} R_\delta\big(|T_\ell^1|, g_{T_\ell^1}, h_{T_\ell^1}\big), \\ &\geq \sup_{\delta\in(0,1]} R_\delta\big(|T_\ell^1|, 2^{-\ell-1}g^\star, 2^{-\ell-1}g^\star\big), \\ &\geq \frac{1}{28} C\big(|T_\ell^1|, 2^{-\ell}g^\star, \infty\big), \end{aligned} \tag{34}$$

and

$$\begin{aligned} \sup_{\delta\in(0,1]} R_\delta\big(N, g_{[N]}, h_{[N]}\big) &\geq \sup_{\delta\in(0,1]} R_\delta\big(|T_\ell^2|, g_{T_\ell^2}, h_{T_\ell^2}\big), \\ &\geq \sup_{\delta\in(0,1]} R_\delta\big(|T_\ell^2|, N^2 2^{-\ell-1}h^\star, 2^{-\ell-1}h^\star\big), \\ &\geq \frac{1}{56\tilde{L}} C\big(|T_\ell^1|, \infty, \tilde{L}2^{1-\ell}h^\star\big), \end{aligned} \tag{35}$$

for all $\ell \in \{0, \dots, L\}$.

We can optimize over the lower bounds in (33), (34), and (35) to obtain

$$\begin{aligned} \sup_{\delta\in(0,1]} R_\delta\big(N, g_{[N]}, h_{[N]}\big) \geq \frac{1}{112\tilde{L}} \max\Bigg\{ &\max_{\ell\in\{0,\dots,L\}} C\big(|T_\ell^1|, 2^{-\ell}g^\star, \infty\big), \max_{\ell\in\{0,\dots,L\}} C\big(|T_\ell^2|, \infty, \tilde{L}2^{1-\ell}h^\star\big), \\ &\max_{k,\ell\in\{0,\dots,L\}} C\big(|S_{k,\ell}|, 2^{-k}g^\star, \tilde{L}2^{1-\ell}h^\star\big)\Bigg\}. \end{aligned}$$

Comparing this with the upper bound (29) shows that

$$C(N, g_{[N]}, h_{[N]}) \leq 112\tilde{L}^2 \sup_{\delta\in(0,1]} R_\delta(N, g_{[N]}, h_{[N]}).$$

Using that

$$\tilde{L} \leq (3\log(N) + 1)^2 + 6\log(N) + 4$$

shows that there exists a universal constant $K < \infty$ (and, in particular, independent of $g_{[N]}$, $h_{[N]}$, and N) such that $112\tilde{L}^2 \leq K\log^4(N)$ for $N \geq 2$. This concludes the proof of the theorem. ∎

V. Conclusion

We presented an approximation of the capacity of the symmetric Gaussian N-relay diamond network. The capacity was characterized up to a 1.8 bit additive gap and a factor 14 multiplicative gap uniformly for all channel gains and number of relays. The inner bound in this approximate characterization relies on bursty amplify-and-forward, showing that this scheme is good simultaneously at low and high rates, uniformly in the channel gains and in the number of relays N. The upper bound resulted from a careful evaluation of the cut-set bound. We argued that all 2^N possible cuts in the diamond network need to be evaluated simultaneously, and that the standard approach of only considering the minimum of the broadcast and multiple-access cuts is insufficient to derive uniform capacity approximations. We extended this approach to asymmetric diamond networks, for which we showed that bursty amplify-and-forward achieves capacity up to a multiplicative gap of a factor $O(\log^4(N))$ with pre-constant in the order notation independent of the channel gains.

The results in this paper show that, at least for symmetric diamond networks, it is possible to derive capacity approximations that are independent of the network size. Deriving such uniform capacity approximations for general networks remains an open problem.

References

[1] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE ISIT*, p. 22, June 2000.

[2] B. Schein, *Distributed Coordination in Network Information Theory*. PhD thesis, Massachusetts Institute of Technology, 2001.

[3] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 765–779, Mar. 2005.

[4] Y. Kochman, A. Khina, U. Erez, and R. Zamir, "Rematch and forward for parallel relay networks," in *Proc. IEEE ISIT*, pp. 767–771, July 2008.

[5] S. S. C. Rezaei, S. O. Gharan, and A. K. Khandani, "A new achievable rate for the Gaussian parallel relay channel," in *Proc. IEEE ISIT*, pp. 194–198, June 2009.

[6] F. Xue and S. Sandhu, "Cooperation in a half-duplex Gaussian diamond relay channel," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3806–3814, Oct. 2007.

[7] H. Bagheri, A. S. Motahari, and A. K. Khandani, "On the capacity of the half-duplex diamond channel," *arXiv:0911.1426 [cs.IT]*, Nov. 2009. submitted to IEEE Transactions on Information Theory.

[8] W. Kang and S. Ulukus, "Capacity of a class of diamond channels," *IEEE Trans. Inf. Theory*, vol. 57, pp. 4955–4960, Aug. 2011.

[9] R. Etkin and D. N. C. Tse, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5534–5562, Dec. 2008.

[10] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, Apr. 2011.

[11] A. Özgür and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," in *Proc. IEEE ISIT*, pp. 669–673, June 2010. See also arXiv:1005.1284 [cs.IT].

[12] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, pp. 3132–3152, May 2011.

[13] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[14] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3549–3572, Oct. 2007.

[15] J. Ghaderi, L.-L. Xie, and X. Shen, "Hierarchical cooperation in ad hoc networks: Optimal clustering and achievable throughput," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3425–3436, Aug. 2009.

[16] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *IEEE Trans. Inf. Theory*, vol. 56, pp. 3959–3982, Sept. 2009.

[17] U. Niesen, "Interference alignment in dense wireless networks," *IEEE Trans. Inf. Theory*, vol. 57, pp. 2889–2901, May 2011.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[19] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[20] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, pp. 711–716, Sept. 1987.

[21] C.-K. Li and R. Mathias, "Extremal characterizations of the Schur complement and resulting inequalities," *SIAM Review*, vol. 42, pp. 233–246, June 2000.

[22] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.