Gaussian Half-Duplex Relay Networks: Improved Constant Gap and Connections with the Assignment Problem

Martina Cardone, Daniela Tuninetti, Raymond Knopp and Umer Salim

***Abstract*—This paper considers a Gaussian relay network where a source transmits a message to a destination with the help of N half-duplex relays. The information theoretic cut-set upper bound to the capacity is shown to be achieved to within $1.96(N+2)$ bits by noisy network coding, thereby reducing the previously known gap. This gap is obtained as a special case of a more general constant gap result for Gaussian half-duplex multicast networks. It is then shown that the generalized Degrees-of-Freedom of this network is the solution of a linear program, where the coefficients of the linear inequality constraints are proved to be the solution of several linear programs referred as the assignment problem in graph theory, for which efficient numerical algorithms exist. The optimal schedule, that is, the optimal value of the 2^N possible transmit-receive configuration states for the relays, is investigated and known results for diamond networks are extended to general relay networks. It is shown, for the case of $N=2$ relays, that only $N+1=3$ out of the $2^N=4$ possible states have a strictly positive probability and suffice to characterize the capacity to within a constant gap. Extensive experimental results show that, for a general N-relay network with $N \leq 8$, the optimal schedule has at most $N+1$ states with a strictly positive probability. As an extension of a conjecture presented for diamond networks, it is conjectured that this result holds for any HD relay network and any number of relays. Finally, a network with $N=2$ relays is studied in detail to illustrate the channel conditions under which selecting the best relay is not optimal, and to highlight the nature of the rate gain due to multiple relays.**

***Index Terms*—Assignment problem, capacity to within a constant gap, generalized degrees-of-freedom, half-duplex, inner bound, outer bound, relay networks, weighted bipartite matching problem.**

Eurecom's research is partially supported by its industrial partners: BMW Group Research & Technology, IABG, Monaco Telecom, Orange, SAP, SFR, ST Microelectronics, Swisscom and Symantec. The research at Eurecom leading to these results has received funding from the EU Celtic+ Framework Program Project SHARING. The work of D. Tuninetti was partially funded by NSF under award number 0643954; the contents of this article are solely the responsibility of the author and do not necessarily represent the official views of the NSF. The research work carried out at Intel by U. Salim has received funding from the EU FP7 grant agreement iJOIN n° 317941.

M. Cardone would like to thank Paul de Kerret for insightful discussions.

M. Cardone and R. Knopp are with the Mobile Communications Department at Eurecom, Biot, 06410, France (e-mail: cardone@eurecom.fr; knopp@eurecom.fr). D. Tuninetti is with the Electrical and Computer Engineering Department of the University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: danielat@uic.edu). U. Salim is with the Algorithm Design group of Intel Mobile Communications, Sophia Antipolis, 06560, France (e-mail: umer.salim@intel.com).

The results in this paper were presented in part to the 2013 IEEE Information Theory Workshop (ITW) [1].

I. Introduction

Cooperation between nodes in a network has been proposed as a potential and promising technique to enhance the performance of wireless systems in terms of coverage, throughput, network generalized Degrees-of-Freedom (gDoF) and robustness / diversity. This last point is of great importance, especially in military and satellite communications, where redundancy and diversity play a significant role, by insuring a more reliable link between two networks (military communication) and two ground stations (satellite communication), with respect to the point-to-point communication.

The simplest form of collaboration can be modeled as a Relay Channel (RC) [2]. The RC is a multi-terminal network where a source conveys information to a destination with the help of one relay. The relay has no own data to send and its only purpose is to assist the source in the transmission. Motivated by the undeniable practical importance of the RC, in this paper we analyze a system where the communication between a source and a destination is assisted by multiple relays. In particular, we mainly focus on the enhancement in terms of gDoF due to the use of multiple Half-Duplex (HD) relays. A relay is said to work in HD mode if at any time / frequency instant it can not simultaneously transmit and receive. The HD modeling assumption is at present more practical than the Full-Duplex (FD) one. This is so because practical restrictions arise when a node can simultaneously transmit and receive, such as for example how well the self-interference can be canceled, making the implementation of FD relays challenging [3], [4].

A. Related work

The RC model was first introduced by van der Meulen [5] in 1971. Despite the significant research efforts, the capacity of the general memoryless RC is still unknown. In their seminal work [2], Cover and El Gamal proposed a general outer bound, now known as the *max-flow min-cut outer bound* or cut-set for short, and two achievable schemes: *decode-and-forward* (DF) and *compress-and-forward* (CF). The cut-set outer bound was shown to be tight for the degraded RC, the reversely degraded RC and the semi-deterministic RC [2], but it is not tight in general [6].

Although more study has been conducted for FD relays, there are some important references treating HD ones. In [7], the author studied the time-division duplexing RC. Both an outer bound, based on the cut-set argument, and an inner bound, based on *partial decode-and-forward* (PDF), were developed. In [7], the time instants at which the relay switches from listen to transmit and vice versa are assumed fixed, i.e., a priori known by all nodes; we refer to this mode of operation as *deterministic switch*. In [8], it was shown that higher rates can be achieved by considering a *random switch* at the relay. In this way the randomness that lies into the switch may be used to transmit (at most one bit per channel use of) further information to the destination. In [8], it was also shown how the memoryless FD framework incorporates the HD one as a special case, and as such there is no need to develop a separate theory for networks with HD nodes.

The pioneering work of [2] has been extended to networks with multiple relays. In [9], the authors proposed several inner and outer bounds for FD relay networks as a generalization of DF, CF and the cut-set bound; it was shown that DF achieves the ergodic capacity of a wireless Gaussian network with uniform phase fading if the phase information is locally available and the relays are close to the source node.

The exact characterization of the capacity region of a general memoryless network is challenging. Recently it has been advocated that progress can be made towards understanding the capacity by showing that achievable strategies are provably close to (easily computable) outer bounds [10]. As an example, in [11], the authors studied FD Gaussian relay networks with $N+2$ nodes (i.e., N relays, a source and a destination) and showed that the capacity can be achieved to within $\sum_{k=1}^{N+2} 5\min\{M_k, N_k\}$ bits with a network generalization of CF named *quantize-remap-and-forward* (QMF), where M_k and N_k are the number of transmit and receive antennas, respectively, of node k. Interestingly, the gap result remains valid for static and ergodic fading networks where the nodes operate either in FD mode or in HD mode; however [11] did not account for random switch in the outer bound. In [12], the authors demonstrated that the QMF scheme can be realized with nested lattice codes. Moreover they showed that for single antenna HD networks with N relays, by following the approach of [8], i.e., by also accounting for random switch in the outer bound, the gap is $8(N+2)$ bits [12]. Recently, for single antenna networks with N FD relays, the $5(N+2)$ bits gap of [11] was reduced to $2\times 0.63(N+2)$ bits (where the factor 2 accounts for complex-valued inputs) thanks to a novel ingenious generalization of CF named *noisy network coding* (NNC) [13].

The gap characterization of [13] is valid for a general multicast Gaussian network with FD nodes; the gap grows linearly with the number of nodes in the network, which could be a too coarse capacity characterization for networks with a large number of nodes. Smaller gaps can be obtained for more structured networks. For example, a *diamond network* [14] consists of a source, a destination and N relays where the source and the destination can not communicate directly and the relays can not communicate among themselves. In other words, a general Gaussian relay network with N relays is characterized by $(N+2)(N+1)$ generic channel link gains, while a diamond network has only $2N$ non-zero channel link gains. In [14] the case of $N=2$ relays was studied and an achievable region based on time sharing between DF and *amplify-and-forward* (AF) was proposed. In [15], the authors considered two specific configurations of a diamond network with a general number of relays (agents), where the relay-destination links are assumed to be lossless; in the first scenario the relays do not have decoding capabilities, while in the second scenario they do. Upper and lower bounds on the capacity were derived and evaluated for the Gaussian noise channel. Moreover the capacity of the deterministic channel when the relays can decode was characterized. The scenarios of [15] were further studied in [16] under the assumption of lossy relay-destination links and where each source-relay link and relay-destination link is a binary-symmetric channel. In [17], the authors analyzed the Gaussian diamond network with a direct link between the source and the destination and showed that 'uncoded forwarding' at the relays asymptotically achieves the cut-set upper bound when the number of relays goes to infinity. This strategy simply requires that each relay delays the input of one time unit and scales it to satisfy the power constraint.

The capacity of a general Gaussian FD diamond network is known to within $2\log(N+1)$ bits [18], [19]. If, in addition, the network is symmetric, that is, all source-relay links are equal and all relay-destination links are equal, the gap is less than 3.6 bits for any N [20].

HD diamond networks have been studied as well. In a HD diamond network with N relays, there are 2^N possible combinations of listening / transmitting states, since each relay, at a given time instant, can either transmit or receive. For the case of $N=2$ relays, [21] showed that out of $2^N=4$ possible states only $N+1=3$ states suffice to achieve the cut-set bound to within less than 4 bits. We refer to the states with strictly positive probability as *active states*. The achievable scheme of [21] is a clever extension of the two-hop DF strategy of [22]. In [21] a closed-form expression for the aforementioned active states, by assuming no power control and deterministic switch, was derived by solving the dual linear program (LP) associated with the LP derived from the cut-set bound. The work in [23] studied a general diamond network with $N=2$ relays and an 'antisymmetric' diamond network with $N=3$ relays and showed that a significant fraction of the capacity can be achieved by: (i) selecting a single relay, or (ii) selecting two relays and allowing them to work in a complementary fashion as in [21]. Inspired by [21], the authors of [23] also showed that, for a specific HD diamond network with $N=3$ relays, at most $N+1=4$ states out of the $2^N=8$ possible ones are active. The authors also numerically verified that for a general HD diamond network with $N\leq 7$ relays, at most $N+1$ states are active and conjectured that the same holds for any number of relays. In [24], this conjecture was proved for the case $N=3$ by using the *submodularity* of the cut-set bound.

Relay networks were also studied in [25], where an iterative algorithm was proposed to determine the optimal fraction of time each HD relay transmits/receives by using DF with deterministic switch.

B. Contributions

In this work we study a general Gaussian HD relay network, whose exact capacity is unknown, by following the approach proposed in [8]. Our main contributions can be summarized as follows:

1) We prove that NNC with deterministic switch achieves the cut-set bound (properly evaluated to account for random switch) to within $1.96(N+2)$ bits. This gap is smaller than the $8(N+2)$ bits gap available in the literature [12]. Our gap result for HD relay networks is obtained as a special case of a more general result for *multicast HD Gaussian networks*, which extends the 1.26 bits/node gap for the FD case [13] to a 1.96 bits/node gap for the HD case.
2) In order to determine the gDoF of the channel, one needs to find a tight high-SNR approximation for the different mutual information terms involved in the cut-set upper bound. As a result of independent interest, we show that such tight approximations can be found as the solution of Maximum Weighted Bipartite Matching (MWBM) problems, or assignment problems [26]. The MWBM problem is a special LP for which efficient polynomial-time algorithms, such as the Hungarian algorithm [27], exist. As an example, we show that this technique is useful to derive the gDoF of Gaussian broadcast networks with relays and to solve user scheduling problems.
3) We extend the results of [21] for a diamond network with $N=2$ relays to the general network. We show that, out of the $2^N = 4$ possible states, at most $N+1=3$ are sufficient to characterize the capacity to within a constant gap. Extensions beyond gDoF are also discussed. Similarly to [23], we verify through extensive numerical evaluations that, for a general relay network with $N \leq 8$ relays, at most $N+1$ states are sufficient to characterize the gDoF. Based on this evidence, we conjecture that the conjecture of [23] holds for any HD relay network.
4) We finally consider a general relay network with $N = 2$ relays. We highlight under which channel conditions a best-relay selection scheme is strictly suboptimal in terms of gDoF and we gain insights into the nature of the rate gain attainable in networks with multiple relays. For example, we show when the interaction between the relays, which is impossible in diamond networks, can strictly increase the gDoF.

C. Paper organization

The rest of the paper is organized as follows. Section II describes the channel model. Section III shows that the cut-set upper bound and the NNC lower bound for a general Gaussian HD multicast network are to within a constant gap from each another, from which the constant gap result for the relay network straightforwardly follows. Section IV proves the equivalence between the problem of finding the coefficients of the linear inequality constraints of the LP derived from the cut-set upper bound and the MWBM problem; it also shows that, for a 2-relay network, the number of active states in the cut-set upper bound is at most 3 for any SNR and any input distributions; it finally presents a conjecture regarding the maximum number of active states sufficient to characterize the cut-set upper bound for a general relay network and for any number of relays. Section V provides an example of a HD relay network with $N=2$ relays; it determines under which channel conditions the gDoF achieved with the best-relay selection strategy is strictly smaller than the gDoF attained by exploiting both relays; it provides insights into the synergies of multiple relays. Section VI concludes the paper. The proofs can be found in Appendix.

D. Notation

We use the notation convention of [28]. $[n_1 : n_2]$ is the set of integers from n_1 to $n_2 \geq n_1$. $[x]^+ := \max\{0, x\}$ for $x \in \mathbb{R}$. Y^j is a vector of length j with components $(Y_1, \ldots, Y_j)$. For an index set $\mathcal{A}$ we let $Y_{\mathcal{A}} = \{Y_j : j \in \mathcal{A}\}$. $\mathbf{0}_j$ denotes either the all-zero column vector of length j or the all-zero square matrix of dimension j, which one is usually clear from the context. $\mathbf{1}_j$ is a column vector of length j of all ones. $\mathbf{I}_j$ is the identity matrix of dimension j. $f_1(x) \doteq f_2(x)$ means that $\lim_{x\to\infty} f_1(x)/f_2(x) = 1$. $|A|$ indicates either the determinant of the matrix A or the cardinality of the set A, which one is usually clear from the context, while $\|a\|$ is the Euclidean length of the vector a. To indicate a sub matrix of the matrix $\mathbf{A}$ where only the rows indexed by the set $\mathcal{R}$ and the columns indexed by the set $\mathcal{C}$ are retained, we use the Matlab-inspired notation $\mathbf{A}_{\mathcal{R},\mathcal{C}}$. Moreover, for a square matrix $\mathbf{A}$, $\mathrm{diag}[\mathbf{A}]$ is a vector containing the diagonal elements of $\mathbf{A}$, while for a vector $\mathbf{a}$ we let $\mathrm{diag}[\mathbf{a}]$ be a diagonal matrix with the entries of $\mathbf{a}$ on the main diagonal, i.e., $\left[\mathrm{diag}[\mathbf{a}]\right]_{ij} = a_i\delta[i-j]$, where $\delta[n] = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0 \end{cases}$ is the Kronecker delta function. $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ indicates that X is a proper-complex Gaussian random variable with mean μ and variance σ^2.

II. System model

A. General memoryless relay network

A memoryless relay network has one source (node 0), one destination (node $N+1$), and N relays indexed from 1 to N. It consists of $N+1$ input alphabets $(\mathcal{X}_1, \cdots, \mathcal{X}_N, \mathcal{X}_{N+1})$ (here $\mathcal{X}_i$ is the input alphabet of node i except for the source/node 0 where, for notation convenience, we use $\mathcal{X}_{N+1}$ rather than $\mathcal{X}_0$), $N+1$ output alphabets $(\mathcal{Y}_1, \cdots, \mathcal{Y}_N, \mathcal{Y}_{N+1})$ (here $\mathcal{Y}_i$ is the output alphabet of node i), and a transition probability $\mathbb{P}_{Y_{[1:N+1]}|X_{[1:N+1]}}$. The source has a message W uniformly distributed on $[1 : 2^{nR}]$ for the destination, where n denotes the codeword length and R the transmission rate in bits per channel use (logarithms are in base 2). At time i, $i \in [1 : n]$, the source maps its message W into a channel input symbol $X_{N+1,i}(W)$, and the k-th relay, $k \in [1 : N]$, maps its past channel observations into a channel input symbol $X_{k,i}\left(Y_k^{i-1}\right)$. The channel is assumed to be memoryless, that is, the following Markov chain holds for all $i \in [1 : n]$

$$(W, Y_{[1:N+1]}^{i-1}, X_{[1:N+1]}^{i-1}) \to X_{[1:N+1],i} \to Y_{[1:N+1],i}.$$

At time n, the destination outputs an estimate of the message based on all its channel observations as $\widehat{W}\left(Y_{N+1}^n\right)$. A rate R is said to be ϵ-achievable if there exists a sequence of codes indexed by the block length n such that $\mathbb{P}[\widehat{W} \neq W] \leq \epsilon$ for some $\epsilon \in [0,1]$. The capacity is the largest non-negative rate that is ϵ-achievable for any $\epsilon > 0$.

In this general memoryless framework, each relay can listen and transmit at the same time, i.e., it is a FD node. HD channels are a special case of the memoryless FD framework in the following sense [8]. With a slight abuse of notation compared to the previous paragraph, we let the channel input of the k-th relay, $k \in [1:N]$, be the pair (X_k, S_k), where $X_k \in \mathcal{X}_k$ as before and $S_k \in [0:1]$ is the *state* random variable that indicates whether the k-th relay is in receive-mode ($S_k = 0$) or in transmit-mode ($S_k = 1$). In the HD case the transition probability is specified as $\mathbb{P}_{Y_{[1:N+1]}|X_{[1:N+1]},S_{[1:N]}}$. In particular, when the k-th relay, $k \in [1:N]$, is listening ($S_k = 0$) the outputs are independent of X_k, while when the k-th relay is transmitting ($S_k = 1$) its output Y_k is independent of all other random variables. By adopting this convention, there is no need to develop a separate theory for HD channels, but the HD constraints can be incorporated into the FD framework. For example, with $N = 1$ relay the channel transition probability is specified by

$$\mathbb{P}_{Y_1,Y_2|X_1,X_2,S_1=0} = \mathbb{P}_{Y_1,Y_2|X_2,S_1=0}$$
$$\mathbb{P}_{Y_1,Y_2|X_1,X_2,S_1=1} = \mathbb{P}_{Y_2|X_1,X_2,S_1=1}\mathbb{P}_{Y_1|S_1=1}$$

where Y_1 is the relay output, Y_2 the destination output, X_1 the relay input, X_2 the source input, and S_1 the relay state. The generalization to an arbitrary number of relays N follows straightforwardly.

B. The Gaussian HD relay network

The single-antenna complex-valued power-constrained Gaussian HD relay network is described by the input/output relationship

$$\mathbf{Y} = \mathbf{H}_{\mathrm{eq}}\mathbf{X} + \mathbf{Z} \in \mathbb{C}^{N+1\times 1}, \tag{1a}$$
$$\mathbf{H}_{\mathrm{eq}} := \begin{bmatrix} \mathbf{I}_N - \mathrm{diag}[\mathbf{S}] & \mathbf{0}_N \\ \mathbf{0}_N^T & 1 \end{bmatrix} \mathbf{H} \begin{bmatrix} \mathrm{diag}[\mathbf{S}] & \mathbf{0}_N \\ \mathbf{0}_N^T & 1 \end{bmatrix} \tag{1b}$$

where

- $\mathbf{Y} := [Y_1, \ldots, Y_N, Y_{N+1}]^T \in \mathbb{C}^{N+1\times 1}$ is the vector of the received signals.
- $\mathbf{X} := [X_1, \ldots, X_N, X_{N+1}]^T \in \mathbb{C}^{N+1\times 1}$ is the vector of the transmitted signals (recall that, although the source is referred to as node 0, its input is indicated as X_{N+1} rather than X_0). Without loss of generality, we assume that the inputs are subject to the average power constraint $\mathbb{E}\left[|X_k|^2\right] \leq 1$, $k \in [1:N+1]$.
- $\mathbf{S} := [S_1, \ldots, S_N]^T \in [0:1]^N$ is the vector of the binary relay states, which takes into account if the k-th relay is receiving ($S_k = 0$) or transmitting ($S_k = 1$) for $k \in [1:N]$.
- $\mathbf{H} \in \mathbb{C}^{(N+1)\times(N+1)}$ is the constant channel matrix known by all terminals defined as

$$\mathbf{H} := \begin{bmatrix} \mathbf{H}_{\mathrm{r}\to\mathrm{r}} & \mathbf{H}_{\mathrm{s}\to\mathrm{r}} \\ \mathbf{H}_{\mathrm{r}\to\mathrm{d}} & \mathbf{H}_{\mathrm{s}\to\mathrm{d}} \end{bmatrix}. \tag{2}$$

  The entry in position (i,j) of the channel matrix in (2) represents the channel from node j to node i, $(i,j) \in [1:N+1]^2$. $\mathbf{H}$ is assumed to be static and therefore known to all nodes. In particular:
  - $\mathbf{H}_{\mathrm{r}\to\mathrm{r}} \in \mathbb{C}^{N\times N}$ defines the network connections among relays, i.e., $[\mathbf{H}_{\mathrm{r}\to\mathrm{r}}]_{ij}$, $(i,j) \in [1:N]^2$, is the channel gain from the j-th relay to the i-th relay. Notice that the entries on the main diagonal of $\mathbf{H}_{\mathrm{r}\to\mathrm{r}}$ do not matter for the channel capacity since the relays operate in HD mode;
  - $\mathbf{H}_{\mathrm{s}\to\mathrm{r}} \in \mathbb{C}^{N\times 1}$ is the column vector which contains the channel gains from the source to the relays, i.e., $[\mathbf{H}_{\mathrm{s}\to\mathrm{r}}]_{i,1}$, $i \in [1:N]$, is the channel gain from the source to the i-th relay;
  - $\mathbf{H}_{\mathrm{r}\to\mathrm{d}} \in \mathbb{C}^{1\times N}$ is the row vector which contains the channel gains from the relays to the destination, i.e., $[\mathbf{H}_{\mathrm{r}\to\mathrm{d}}]_{1,i}$, $i \in [1:N]$, is the channel gain from the i-th relay to the destination;
  - $\mathbf{H}_{\mathrm{s}\to\mathrm{d}} \in \mathbb{C}^{1\times 1}$ is the channel gain between the source and the destination (recall that by our notation the source input is indicated as X_{N+1} rather than X_0).
- $\mathbf{Z} := [Z_1, \ldots, Z_N, Z_{N+1}]^T \in \mathbb{C}^{N+1\times 1}$ is the jointly Gaussian noise vector. Without loss of generality, the noises are assumed to have zero mean and unit variance. Furthermore we assume, not without loss of generality [29], that the noises are independent, i.e., the covariance matrix of $\mathbf{Z}$ is the identity matrix.

The capacity of the Gaussian HD relay network in (1) is not known in general. In order to evaluate the ultimate performance of this system we make use of two metrics: the gDoF and the capacity to within a constant gap.

The capacity to within a constant gap is defined as:

Definition 1. *The capacity* C *of the Gaussian HD relay network in* (1) *is said to be known to within* GAP *bits if one can show an achievable rate* $R^{(\mathrm{in})}$ *and an outer bound* $R^{(\mathrm{out})}$ *such that*

$$R^{(\mathrm{in})} \leq \mathsf{C} \leq R^{(\mathrm{out})} \leq R^{(\mathrm{in})} + \mathsf{GAP}, \tag{3}$$

where GAP *is a constant that does not depend on the channel gains matrix* $\mathbf{H}$ *in* (1).

Knowing the capacity to within a constant gap implies the exact knowledge of the gDoF defined as:

Definition 2. *The gDoF of the Gaussian HD relay network in* (1) *is defined as*

$$\mathsf{d} := \lim_{\mathsf{SNR}\to+\infty} \frac{\mathsf{C}}{\log(1+\mathsf{SNR})}, \tag{4}$$

where C *is the capacity and* $\mathsf{SNR} \in \mathbb{R}^+$ *parameterizes the channel gains as* $|h_{ij}|^2 = \mathsf{SNR}^{\beta_{ij}}$*, for some fixed non-negative* β_{ij}, $(i,j) \in [1:N+1]^2$.

The gDoF in (4) is an exact characterization of the capacity at high-SNR, while the capacity to within a constant gap in (3) quantifies how far inner and outer bounds are in the worst SNR scenario.

III. Capacity to within a constant gap

This section is devoted to the capacity characterization of the Gaussian HD relay network in (1) to within a constant

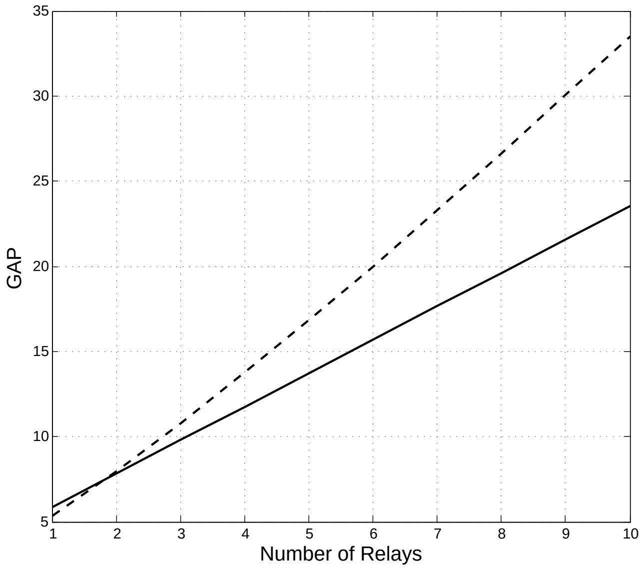

Fig. 1: Gap in (5) (solid curve) and gap in (6) (dashed curve) for the HD Gaussian relay network. The gap in (5) is smaller than that in (6) for any $N \geq 2$.

gap. We first adapt the cut-set upper bound [9] and the NNC lower bound [13] to the HD case by following the approach proposed in [8]. We then show that these bounds are at most a constant number of bits apart. Our result is:

Theorem 1. *The cut-set upper bound for the Gaussian HD relay network with N relays is achievable by NNC with deterministic switch to within*

$$\mathsf{GAP} \leq 1.96(N+2) \text{ bits}. \tag{5}$$

Proof: We first prove that for a *multicast* single-antenna complex-valued Gaussian network with HD power-constrained nodes the cut-set upper bound can be achieved to within 1.96 bits/node (while for the FD case the gap is 1.26 bits/node [13, Theorem 4]). The unicast Gaussian network with HD relays is a special multicast network with $K = N + 2$ nodes (i.e., one source, N relays, and one destination), which proves the claimed result in (5). The derivation of the gap result for the multicast network can be found in Appendix A. ■

Remark 1. The gap in Theorem 1 improves on the previously known gap result of $8(N + 2)$ bits [12]. □

Remark 2 (Single relay case)**.** The gap result in (5) for $N = 1$ gives $\mathsf{GAP} \leq 5.88$ bits, which is greater than the 1.61 bits gap we found in [30]. This is due to the fact that the bounding steps in the special case of $N = 1$ are tighter than those we used here for a general multicast network with K nodes.

Notice also that for a single relay, PDF is optimal to within 1 bit [30]. PDF has been extended to a general HD multi-relay network in [31]. However, to analytically evaluate this region and show that it achieves the cut-set upper bound to within a constant gap seems to be a challenging task, which is the main motivation for considering NNC here. □

Remark 3. In a preliminary version of this work [1], by using a bounding technique as in [32, pages 20-5, 20-7] we obtained

$$\mathsf{GAP} \cong \frac{N+2}{2} \log\left(4(N+2)\right) \text{ bits}. \tag{6}$$

As shown in Fig. 1 the gap in (5) is smaller that the one in (6) for $N \geq 2$ thanks to the tighter bound derived from [13, Lemma 1]. □

Remark 4 (Diamond networks)**.** A smaller gap than the one in (5) may be obtained for specific network topologies. For example, in [18] and [19] it was found that for a Gaussian FD diamond network with N relays the gap is of the order $\log(N)$, rather than linear in N [13]. Moreover, for a symmetric FD diamond network with N relays the gap does not depend on the number of relays and it is upper bounded by 3.6 bits [20]. The key difference between a general relay network and a diamond network is that for each subset $\mathcal{A}$ we have that $\text{rank}[\mathbf{H}_{\mathcal{A}}] \leq 2$; hence in (31b) we can use $\text{rank}[\mathbf{H}_{\mathcal{A}}] \leq \min\{|\mathcal{A}|, |\mathcal{A}^c|, 2\}$. With this and by numerically evaluating the resulting gap we obtain the result plotted in Fig. 2. From Fig. 2, we observe that the gap for the HD diamond network is in general smaller than the one computed for the general HD relay network; this is in line with what happens in FD. However, in FD for the diamond network the gap is logarithmic in N [18], [19], while the gap in Fig. 2 (solid curve) still grows linearly with N. This is due to the fact that the HD cut-set outer bound, as opposed to the FD one, contains the entropy of the state vector, which is upper bounded by the uniform distribution over the all possible states; this term contributes linearly in the number of nodes to the overall gap. Moreover, from Fig. 2 we observe that our gap (solid curve) is larger than the gap of order $N+3\log(N)$ from [24] (dashed curve). We believe that the reason is because our gap has been computed as a special case of a general HD multicast network while the one in [24] has been specifically derived for HD diamond relay networks. □

As we shall see in the next section, for a general HD network with N relays, only $N + 1$ states, out of the possible 2^N states, appear to be needed to characterize the cut-set upper bound. It is subject of current investigation on how to use this observation to develop bounds leading to a smaller gap.

IV. Analysis of the optimal schedule

In general, for a N-relay network, 2^N states are possible. A capacity achieving scheme must optimize the probability with which each of these states occurs. In [21], it was proved that for a diamond network with $N = 2$ relays, out of the $2^N = 4$ possible states, at most $N + 1 = 3$ are active / have a non-zero probability. In [24], the authors extended the result of [21] to a diamond network with $N = 3$ relays. Based on numerical evidences, [23] conjectured that for a N-relay diamond network out of the 2^N possible states at most $N + 1$ states are active. Here we extend these results to a general Gaussian HD relay network as follows. The claim "out of 2^N possible states only $N + 1$ states are active as far as gDoF is concerned" is proved analytically for $N = 2$, shown to hold by numerical evaluations for $N \leq 8$ and conjectured to hold for any N. If the conjecture were true, it would show that HD relay networks have intrinsic properties regardless of

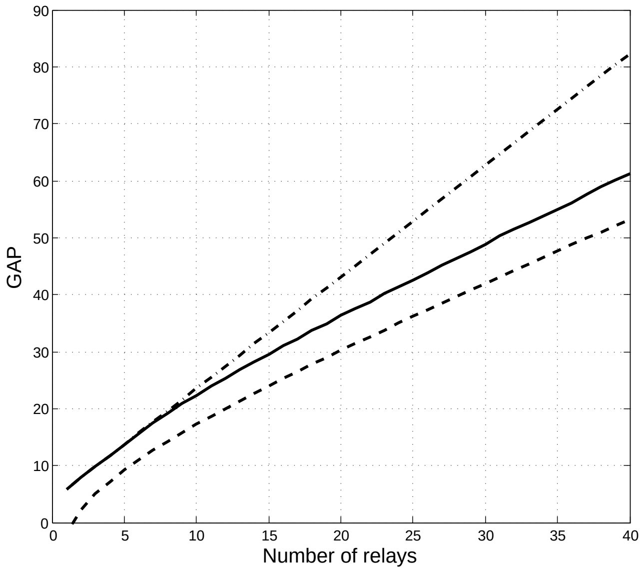

Fig. 2: Gap in (5) (dash-dotted curve), gap in (5) specialized to the HD diamond network (solid curve) and gap in [24] (dashed curve) for the HD diamond network.

their topology, i.e., known results for diamond networks are not a consequence of the simplified network topology.

In order to determine the gDoF we must find a tight high-SNR approximation for the different MIMO-type mutual information terms involved in the cut-set upper bound (see eq.(31a) in Appendix A). As a result of independent interest beyond the application to the Gaussian HD relay network studied in this paper, we first show that such an approximation can be found as the solution of a Maximum Weighted Bipartite Matching (MWBM) problem.

A. The maximum weighted bipartite matching (MWBM) problem and its relation to the gDoF

In graph theory, a weighted bipartite graph, or bigraph, is a graph whose vertices can be separated into two sets such that each edge in the graph has exactly one endpoint in each set. Moreover, a non-negative weight is associated with each edge in the bigraph [33]. For a general network we define the weight matrix $\mathbf{B}$ as follows: there is one set of n_1 nodes, where n_1 is the number of rows in $\mathbf{B}$, and another set of n_2 nodes, where n_2 is the number of columns in $\mathbf{B}$; the element $[\mathbf{B}]_{ij}$ is the weight of the edge between nodes i and j. A matching, or independent edge set, is a set of edges without common vertices [33]. The MWBM problem, or assignment problem, is defined as a matching where the sum of the edge weights in the matching has the maximal value [26]. The Hungarian algorithm is a polynomial time algorithm that efficiently solves the assignment problem [27]. Equipped with these definitions, we now show the following high-SNR approximation of the MIMO capacity:

Theorem 2. *Let $\mathbf{H} \in \mathbb{R}^{k\times n}$ be a full-rank matrix, where without loss of generality $k \leq n$. Let $\mathcal{S}_{n,k}$ be the set of all k-combinations of the integers in $[1:n]$ and $\mathcal{P}_{n,k}$ be the set of all k-permutations of the integers in $[1:n]$[1]. Then,*

$$\begin{aligned}|\mathbf{I}_k + \mathbf{H}\mathbf{H}^H| &= \sum_{\varsigma\in\mathcal{S}_{n,k}} \sum_{\pi\in\mathcal{P}_{n,k}} \prod_{i=1}^{k} \left|[\mathbf{H}_\varsigma]_{i,\pi(i)}\right|^2 + \quad T \\ &\doteq \mathsf{SNR}^{\mathsf{MWBM}(\mathbf{B})}, \end{aligned} \tag{7}$$

$$\mathsf{MWBM}(\mathbf{B}) := \max_{\varsigma\in\mathcal{S}_{n,k}} \max_{\pi\in\mathcal{P}_{n,k}} \sum_{i=1}^{k} [\mathbf{B}_\varsigma]_{i,\pi(i)}, \tag{8}$$

where $\mathbf{B}$ is the SNR-exponent matrix defined as $[\mathbf{B}]_{ij} = \beta_{ij} \geq 0 : |h_{ij}|^2 = \mathsf{SNR}^{\beta_{ij}}$, $\mathbf{H}_\varsigma$ and $\mathbf{B}_\varsigma$ are the square matrices obtained from $\mathbf{H}$ and $\mathbf{B}$, respectively, by retaining all rows and the columns indexed by ς, and T is the sum of terms that overall behave as $o\left(\mathsf{SNR}^{\mathsf{MWBM}(\mathbf{B})}\right)$.

Proof: The proof can be found in Appendix B. The expression in (8) is a possible way of writing the MWBM problem. ■

Theorem 2 establishes an interesting connection between the gDoF of a MIMO channel (with independent inputs) and graph theory. Notice that the high-SNR expression found in Theorem 2 holds for correlated inputs as well, as long as the average power constraint is a finite constant. More importantly, Theorem 2 allows to move from DoF, where all exponents β_{ij} have the same value, to gDoF, where different channel gains have different exponential behavior. DoF is essentially a characterization of the rank of the channel matrix; gDoF captures the potential advantage due to 'asymmetric' channel gains. gDoF, to the best of our knowledge, has been investigated so far only for Single-Input-Single-Output (SISO) networks with very few number of nodes; we believe that the reason is that in these cases one has only to consider equivalent Multiple-Input-Single-Output (MISO) and Single-Input-Multiple-Output (SIMO) channels, or to explicitly deal with determinants of matrices with small dimensions. Our result extends the gDoF analysis to any MIMO channel as we explain through some examples.

1) MISO and SIMO channels, i.e., the case $k = 1 \leq n$: In a MISO or SIMO channel, with channel vector $\mathbf{h} := [h_1, \ldots, h_n]$ such that $|h_k|^2 = \mathsf{SNR}^{\beta_k}, k \in [1:n]$, one trivially has

$$\begin{aligned}\log(1 + \|\mathbf{h}\|^2) &= \log\left(1 + \sum_{i=1}^{n} \mathsf{SNR}^{\beta_i}\right) \\ &\overset{\mathsf{SNR}\gg 1}{\doteq} \log\left(\mathsf{SNR}^{\max_{i\in[1:n]}\{\beta_i\}}\right).\end{aligned}$$

The corresponding MWBM problem has one set of vertices $\mathcal{A}_1$ consisting of $k = |\mathcal{A}_1| = 1$ node and the other set of vertices $\mathcal{A}_2$ consisting of $n = |\mathcal{A}_2| \geq 1$ nodes. The weights of the edges connecting the single vertex in $\mathcal{A}_1$ to the n vertices in $\mathcal{A}_2$ can be represented as the non-negative vector $\mathbf{B} = [\beta_1, \ldots, \beta_n]$. Clearly, the optimal $\mathsf{MWBM}(\mathbf{B}) =$

[1] The k-combinations and the k-permutations of the integers in $[1:n]$ are defined as sequences of a fixed length k of elements taken from a given set of size n such that no elements occurs more than once. Then, over this k-length sequence all the possible combinations $\mathcal{S}_{n,k}$ and all the possible permutations $\mathcal{P}_{n,k}$ are computed. With $\pi(i)$ we indicate the element in the i-th position of the permutation $\pi \in \mathcal{P}_{n,k}$.

$\max_{i\in[1:n]}\{\beta_i\}$ assigns the single vertex in $\mathcal{A}_1$ to the vertex in $\mathcal{A}_2$ that is connected to it through the edge with the maximum weight.

2) 2×2 MIMO channels, i.e., the case $k = n = 2$: As another example from the 2-user interference channel literature, consider the cut-set sum-rate upper bound

$$\mathbf{H} := \begin{bmatrix} h_{13} & h_{23} \\ h_{14} & h_{24} \end{bmatrix} = \begin{bmatrix} \sqrt{\mathsf{SNR}^{\beta_{13}}}\ \mathrm{e}^{\mathrm{j}\theta_{13}} & \sqrt{\mathsf{SNR}^{\beta_{23}}}\ \mathrm{e}^{\mathrm{j}\theta_{23}} \\ \sqrt{\mathsf{SNR}^{\beta_{14}}}\ \mathrm{e}^{\mathrm{j}\theta_{14}} & \sqrt{\mathsf{SNR}^{\beta_{24}}}\ \mathrm{e}^{\mathrm{j}\theta_{24}} \end{bmatrix}$$

$$\Longrightarrow \mathbf{B} := \begin{bmatrix} \beta_{13} & \beta_{23} \\ \beta_{14} & \beta_{24} \end{bmatrix},$$

$$\log\left(\mathbf{I}_2 + \mathbf{H}\mathbf{H}^H\right) \overset{\mathsf{SNR}\gg 1}{\doteq} \log\left(\mathsf{SNR}^{\max\{\beta_{13}+\beta_{24},\beta_{23}+\beta_{14}\}}\right).$$

The corresponding MWBM problem has one set of vertices $\mathcal{A}_1$ consisting of $k = |\mathcal{A}_1| = 2$ nodes (for future references let us refer to these vertices as nodes 1 and 2 – see first subscript in the channel gains) and the other set of vertices $\mathcal{A}_2$ consisting also of $n = |\mathcal{A}_2| = 2$ nodes (for future references let us refer to these vertices as nodes 3 and 4 – see second subscript in the channel gains). The weights of the edges connecting the vertices in $\mathcal{A}_1$ to the vertices in $\mathcal{A}_2$ can be represented as the non-negative weights $\beta_{ji},\ i \in [3:4],\ j \in [1:2]$. In this example, one possible matching assigns node 1 to node 3 and node 2 to node 4 (giving total weight $\beta_{13}+\beta_{24}$), while the other possible matching assigns node 2 to node 3 and node 1 to node 4 (giving total weight $\beta_{23}+\beta_{14}$); the best assignment is the one that gives the largest total weight.

Notice that the MWBM is a tight approximation of the 2×2 MIMO capacity only if the channel matrix is full rank, see [34, eq.(5) 1st line], but it is loose when the channel matrix is rank deficient, see [34, eq.(5) 2nd line, and compare with eq.(11)]. The reason is that the MWBM can not capture the impact of phases in MIMO situations. To exclude the case of a rank deficient channel matrix from our general setting for any value of k and n, we may proceed as in [35, page 2925]. Namely, we pose a reasonable distribution, such as for example the i.i.d. uniform distribution, on the phases $\theta_{ji},\ i\in[3:4],\ j\in[1:2]$, so that almost surely the channel matrix is full rank.

3) Relay-aided broadcast channels: In general, Theorem 2 allows to find the gDoF region for any Gaussian network whose capacity can be approximated to within a constant gap by linear combinations of $\log|\ldots|$ terms. For example, [36] studied a channel where one source communicates with K destinations with the help of L FD relays. The cut-set outer bound on the capacity region of such a network is shown to be achievable to within $O\left(N\log(N)\right)$ bits, where $N=K+L+1$ is the total number of nodes. This constant gap result implies the exact knowledge of the gDoF region. As an example of application of Theorem 2, we next show how to derive the sum gDoF of the relay-aided broadcast channel.

Consider a relay-aided broadcast channel with one source, K destinations, and $L = 1$ relay (the result can be straightforwardly extended to the case of multiple relays, of cooperation among destinations and with generalized feedback at the source). The source has input X_0, the relay has input X_1, the k-th destination has output

$$Y_k = \sqrt{\mathsf{SNR}^{\beta_{k,0}}}X_0 + \sqrt{\mathsf{SNR}^{\beta_{k,1}}}\ \mathrm{e}^{\mathrm{j}\theta_{k,1}}X_1 + Z_k,\ k \in [1:K],$$

and the relay has output

$$Y_{\mathrm{R}} = \sqrt{\mathsf{SNR}^{\beta_{\mathrm{R}}}}X_0 + Z_{\mathrm{R}},$$

where, since the channel is known to all nodes, each receiving node compensates for the phase of the link from the source. We assume that the phases $\{\theta_{k,1},\ k \in [1:K]\}$ are such that all involved channel (sub)matrices are full rank almost surely. Without loss of generality, we let $\beta_{1,0} = \max_{k\in[1:K]}\{\beta_{k,0}\}$, i.e., destination 1 has the strongest link from the source. We define the gDoF of destination k as $\mathsf{d}_k = \lim_{\mathsf{SNR}\to\infty}\frac{R_k}{\log(1+\mathsf{SNR})}$, $k \in [1:K]$. The capacity region of this relay-aided broadcast channel is to within a constant gap from the cut-set upper bound [36]. The cut-set outer bound yields for all $\mathcal{A} \subseteq [1:K],\ \mathcal{A} \neq \emptyset$,

$$\sum_{k\in\mathcal{A}} R_k \le I(X_0, X_1; Y_{\mathcal{A}}), \tag{9a}$$

$$\sum_{k\in\mathcal{A}} R_k \le I(X_0; Y_{\mathcal{A}}, Y_{\mathrm{R}}|X_1). \tag{9b}$$

The sum gDoF (and similarly for any other bounds) is the minimum of two terms: the first term from (9a) with $\mathcal{A} = [1:K]$ is

$$\sum_{k=1}^{K}\mathsf{d}_k \le \mathsf{MWBM}\left(\begin{bmatrix} \beta_{1,0} & \beta_{1,1} \\ \vdots & \vdots \\ \beta_{K,0} & \beta_{K,1} \end{bmatrix}\right) = \max_{j\in[2:K]}\{\beta_{1,0}+\beta_{j,1}, \beta_{j,0}+\beta_{1,1}\}, \tag{10a}$$

and the second term from (9b) with $\mathcal{A} = [1:K]$ is

$$\sum_{k=1}^{K}\mathsf{d}_k \le \mathsf{MWBM}\left(\begin{bmatrix} \beta_{1,0} & \ldots & \beta_{K,0} & \beta_{\mathrm{R}} \end{bmatrix}\right) = \max\{\beta_{1,0}, \beta_{\mathrm{R}}\}, \tag{10b}$$

from the assumption $\beta_{1,0} \ge \beta_{k,0},\ k \in [2:K]$. The closed-form expression for the gDoF in (10) sheds light into approximately optimal achievable schemes: if $\beta_{\mathrm{R}} \le \beta_{1,0} = \max_{k\in[1:K]}\{\beta_{k,0}\}$ the sum-gDoF is as for the broadcast channel without a relay (i.e., in practical wireless broadcast networks it might not be worth using a relay if the source-relay link is weaker than the strongest source-destination link), while if $\beta_{1,0} < \beta_{\mathrm{R}}$ it is sum-gDoF optimal to serve at most one extra destination in addition to destination 1 (see eq.(10a)). With L relays, it is sum-gDoF optimal to serve at most $L+1$ destinations; which subset of destinations to serve can be found by examining the 2^L MWBM-based bounds, in the spirit of (10).

This example shows that Theorem 2 is an efficient tool to characterize the gDoF region and it is not only restricted to networks with a small number of nodes. Moreover, this example shows that the result in Theorem 2 also represents an useful tool to solve user scheduling problems.

$$[\mathbf{A}]_{7,5} = \mathsf{MWBM}\left(\left[\begin{bmatrix}0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}\beta_{11}&\beta_{12}&\beta_{13}&\beta_{14}\\\beta_{21}&\beta_{22}&\beta_{23}&\beta_{24}\\\beta_{31}&\beta_{32}&\beta_{33}&\beta_{34}\\\beta_{41}&\beta_{42}&\beta_{43}&\beta_{44}\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{bmatrix}\right]_{\{3,4\},\{1,2,4\}}\right)$$
$$= \mathsf{MWBM}\left(\begin{bmatrix}\beta_{31}&0&\beta_{34}\\\beta_{41}&0&\beta_{44}\end{bmatrix}\right) = \max\{\beta_{31}+\beta_{44},\beta_{34}+\beta_{41}\} \quad (11)$$

B. The gDoF for a general N-relay network

With Theorem 2 we can express the gDoF d in (4) of the Gaussian HD relay network in (1) as a LP. In particular, let

$$\mathbf{f}^T := [\mathbf{0}_{2^N}^T, 1] \quad (12a)$$
$$\mathbf{x}^T := [\lambda_{\text{vect}}, \mathsf{d}] \quad (12b)$$
$$\lambda_{\text{vect}} := [\lambda_s] \in \mathbb{R}_+^{1\times 2^N} \text{ where } \lambda_s := \mathbb{P}[S_{[1:N]} = s] \in [0,1], \quad \forall s \in [0:1]^N : \sum_{s\in[0:1]^N} \lambda_s = 1, \quad (12c)$$

then

$$\mathsf{d} = \max\{\mathbf{f}^T\ \mathbf{x}\} \quad (12d)$$
$$\text{s.t.} \begin{bmatrix}-\mathbf{A} & \mathbf{1}_{2^N}\\ \mathbf{1}_{2^N}^T & 0\end{bmatrix}\mathbf{x} \le \mathbf{f}, \quad \mathbf{x} \ge 0, \quad (12e)$$

where the non-negative matrix $\mathbf{A} \in \mathbb{R}^{2^N\times 2^N}$ has entries (recall that, although the source is referred to as node 0, its input is indicated as X_{N+1} rather than X_0)

$$[\mathbf{A}]_{ij} := \lim_{\mathsf{SNR}\to+\infty} \frac{I(X_{\mathcal{A}_i\cup\{N+1\}}; Y_{\mathcal{A}_i^c\cup\{N+1\}}|X_{\mathcal{A}_i^c}, S_{[1:N]} = s_j)}{\log(1+\mathsf{SNR})}, \quad (12f)$$

where $\mathcal{A}_i$ and s_j are defined right after Theorem 3. In other words, each row of the matrix $\mathbf{A}$ refers to a possible cut in the network, while each column of $\mathbf{A}$ refers to a possible listening/transmitting configuration.

By a simple application of Theorem 2 we have that each entry of the matrix $\mathbf{A}$ can be evaluated by solving the corresponding MWBM problem. More formally

Theorem 3. *For the LP in eq.*(12)

$$[\mathbf{A}]_{ij} = \mathsf{MWBM}\left(\mathbf{B}_{\{N+1\}\cup(\mathcal{A}_i^c\cap\mathcal{A}_j^c),\ \{N+1\}\cup(\mathcal{A}_i\cap\mathcal{A}_j)}\right).$$

The notation in eq.(12f) and in Theorem 3 is as follows. $\mathbf{B}$ indicates the SNR-exponent matrix defined as $[\mathbf{B}]_{ij} = \beta_{ij} \ge 0 : |h_{ij}|^2 = \mathsf{SNR}^{\beta_{ij}}$, and the indices (i,j) have the following meaning. Index i refers to a "cut" in the network and index j to a "state of the relays". Both indices range in $[1:2^N]$ and must be seen as the decimal representation of a binary number with N bits. $\mathcal{A}_i$, $i \in [1:2^N]$, is the set of those relays who have a one in the corresponding binary representation of $i-1$ and s_j, $j \in [1:2^N]$, sets the state of a relay to the corresponding bit in the binary representation of $j-1$. Finally, we evaluate the MWBM of the bigraph with weight matrix

$$\left[\begin{bmatrix}\mathbf{I}_N - \text{diag}[s_j] & \mathbf{0}_N\\ \mathbf{0}_N^T & 1\end{bmatrix}\mathbf{B}\begin{bmatrix}\text{diag}[s_j] & \mathbf{0}_N\\ \mathbf{0}_N^T & 1\end{bmatrix}\right]_{\{N+1\}\cup\mathcal{A}_i^c,\{N+1\}\cup\mathcal{A}_i}$$
$$= \mathbf{B}_{\{N+1\}\cup(\mathcal{A}_i^c\cap\mathcal{A}_j^c),\ \{N+1\}\cup(\mathcal{A}_i\cap\mathcal{A}_j)},$$

where the equality follows from the following observation. Among the relays 'on the side of the destination' (indexed by $\mathcal{A}_i^c$) only those in receive mode matter (indexed by $\mathcal{A}_j^c$), therefore we can reduce the set of 'receiving nodes' from $\mathcal{A}_i^c$ to $\mathcal{A}_i^c \cap \mathcal{A}_j^c$. Similarly, among the relays 'on the side of the source' (indexed by $\mathcal{A}_i$) only those in transmit mode matter (indexed by $\mathcal{A}_j$), therefore we can reduce the set of 'transmitting nodes' from $\mathcal{A}_i$ to $\mathcal{A}_i \cap \mathcal{A}_j$. Notice that $\mathbf{B}_{\{N+1\}\cup(\mathcal{A}_i^c\cap\mathcal{A}_j^c),\ \{N+1\}\cup(\mathcal{A}_i\cap\mathcal{A}_j)}$ does not change if the roles of i and j are swapped, which implies that $[\mathbf{A}]_{ij} = [\mathbf{A}]_{ji}$, i.e., the matrix $\mathbf{A}$ is symmetric.

To better understand the notation, consider the following example.

Example 1 ($N = 3$, $i = 7$, and $j = 5$)**.** From $i-1 = 6 = 1\cdot 2^2 + 1\cdot 2^1 + 0\cdot 2^0$ we have $\mathcal{A}_7 = \{1,2\} = \{3\}^c$, meaning that relay 1 and relay 2 lie in the cut of the source and relay 3 lies in the cut of the destination. From $j-1 = 4 = 1\cdot 2^2 + 0\cdot 2^1 + 0\cdot 2^0$ we have $s_5 = [1,0,0]$, meaning that relay 1 is transmitting, and relays 2 and 3 are receiving (also $\mathcal{A}_5 = \{1\} = \{2,3\}^c$). With this we have

$$\{N+1\}\cup\mathcal{A}_i^c = \{4\}\cup\{3\} = \{3,4\},$$
$$\{N+1\}\cup\mathcal{A}_i = \{4\}\cup\{1,2\} = \{1,2,4\},$$
$$\{N+1\}\cup(\mathcal{A}_i^c\cap\mathcal{A}_j^c) = \{4\}\cup(\{3\}\cap\{2,3\}) = \{3,4\},$$
$$\{N+1\}\cup(\mathcal{A}_i\cap\mathcal{A}_j) = \{4\}\cup(\{1,2\}\cap\{1\}) = \{1,4\},$$

and therefore (11), at the top of the page, holds. Also

$$[\mathbf{A}]_{7,5} = \mathsf{MWBM}\left(\begin{bmatrix}\beta_{11}&\beta_{12}&\beta_{13}&\beta_{14}\\\beta_{21}&\beta_{22}&\beta_{23}&\beta_{24}\\\beta_{31}&\beta_{32}&\beta_{33}&\beta_{34}\\\beta_{41}&\beta_{42}&\beta_{43}&\beta_{44}\end{bmatrix}_{\{3,4\},\{1,4\}}\right)$$
$$= \mathsf{MWBM}\left(\begin{bmatrix}\beta_{31}&\beta_{34}\\\beta_{41}&\beta_{44}\end{bmatrix}\right)$$
$$= \max\{\beta_{31}+\beta_{44},\beta_{34}+\beta_{41}\}.\square$$

One interesting question is how many λ_s, i.e., λ_s is the probability the network is in state $s \in [0:1]^N$, are strictly positive [21], [23]. In [21], the authors analyzed the diamond network with $N = 2$ relays and showed that out of the $2^N = 4$ possible states only $N+1 = 3$ states are active. The proof considers the dual of the LP in (12). Here we extend the result of [21] to the fully-connected HD relay network with $N = 2$ relays; our proof identifies the channel conditions under which setting the probability of one of the states to zero is without loss of optimality. We have:

Theorem 4. *For a general HD relay network with $N = 2$ relays, there exists a schedule optimal to within a constant*

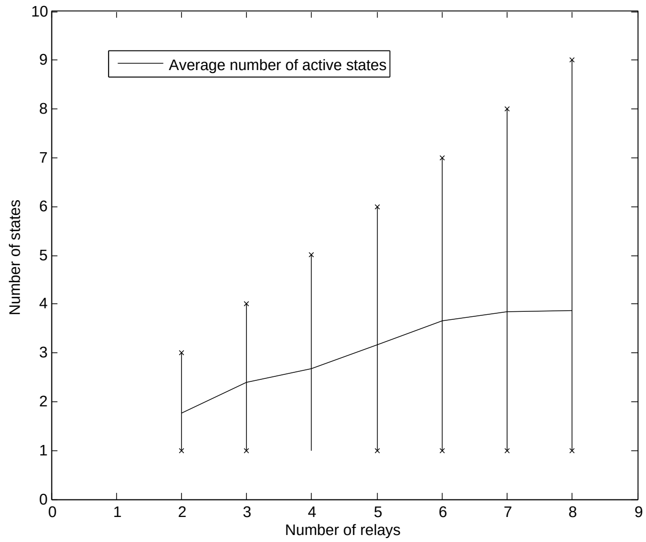

Fig. 3: Average, minimum and maximum number of active states to characterize the capacity of a HD relay network.

gap, and hence a schedule that maximizes d *in* (12d)*, with at most* $N+1=3$ *active states.*

Proof: The proof can be found in Appendix C. Notice that the statement holds for any memoryless HD relay network with $N=2$ relays. In particular, it holds for the Gaussian noise case, at any SNR and for any input distribution; it therefore holds for the gDoF as well. ■

We conjecture that for a general HD relay network with any number of relays Theorem 4 continues to hold, similarly to the conjecture presented in [23] for the diamond network.

Conjecture. *For a general HD relay network with N relays, there always exists a schedule optimal to within a constant gap, and hence a schedule that maximizes* d *in* (12d)*, with at most $N+1$ active states.*

The conjecture holds for the case of $N=2$ relays as proved in Theorem 4. For $N>2$, in our numerical simulations, we consider the gDoF. In particular, we proceeded through the following numerical evaluations: for each value of $N\le 8$, we randomly generated the SNR exponents of the channel gains, we computed the entries of $\mathbf{A}$ in (12) with the Hungarian algorithm, we solved the LP in (12) with the simplex method and we counted the number of constraints that equal the optimal gDoF (which is a known upper bound on the number of non zero entries of an optimal solution). The minimum and the maximum number of active states were found to be 1 and $N+1$, respectively, as shown in Fig. 3, which also shows the average number of active states computed by giving an equal weight to all the tried channels. If the reduction of the number of active states from exponential to linear as conjectured holds, it offers a simpler and more amenable way to design the network [23].

V. Fully-connected relay network with $N=2$ relays

To gain insights into how relays are best utilized, we consider a network with $N=2$ relays. The analysis presented here differs from the one in [23] in the following: (i) we study the fully-connected network, while in [23] only the diamond network is treated; (ii) we explicitly find under which channel conditions the gDoF performance is enhanced by exploiting both relays instead of using only the best one, and (iii) we provide insights into the nature of the rate gain in networks with multiple HD relays.

We consider the parameterization in (13) where, in order to increase the readability of the document, the SNR-exponents are indicated as

$$\left[\frac{\log(|h_{ij}|^2)}{\log(\mathsf{SNR})}\right]_{(i,j)\in[1:3]^2} = \begin{pmatrix} \star & \beta_1 & \alpha_{s1} \\ \beta_2 & \star & \alpha_{s2} \\ \alpha_{1d} & \alpha_{2d} & 1 \end{pmatrix}, \qquad (13)$$

where $\star$ denotes an entry that does not matter for channel capacity, α_{si} is the SNR-exponent on the link from the source to relay i, $i\in[1:2]$, α_{id} is the SNR-exponent on the link from relay i, $i\in[1:2]$, to the destination, β_i is the SNR-exponent on the link from relay j to relay i, $(i,j)\in[1:2]^2$ with $j\neq i$, and the direct link from the source to the destination (entry in position (3,3) in (13)) has SNR-exponent normalized to 1 without loss of generality. Notice that in order to consider a network without a direct link it suffices to consider all the other SNR-exponents to be larger than 1, or simply replace '1' with '0' in the discussion in the rest of the section.

We next derive the gDoF in both the FD and HD cases.

A. The full-duplex case

For the FD case, the cut-set bound is achievable to within $2\times 0.63\times 4 = 5.04$ bits with NNC [13]. As a consequence, it can be verified that the gDoF for the FD case is

$$\begin{aligned}\mathsf{d}^{(\mathrm{FD})}_{N=2} = \min\Big\{ &\max\left\{1,\alpha_{s1},\alpha_{s2}\right\}, \max\left\{\alpha_{s2}+\alpha_{1d},\beta_2+1\right\}, \\ &\max\left\{\alpha_{s1}+\alpha_{2d},\beta_1+1\right\}, \max\left\{1,\alpha_{1d},\alpha_{2d}\right\}\Big\}. \end{aligned} \qquad (14)$$

Notice that $\mathsf{d}^{(\mathrm{FD})}_{N=2}\ge 1$, i.e., the gDoF in (14) is no smaller than the gDoF that could be achieved without using the relays, that is, by communicating directly through the direct link to achieve gDoF $=1$. Notice also that the gDoF in (14) does not change if we exchange α_{s1} with α_{2d}, and α_{s2} with α_{1d}, i.e., if we swap the role of the source and destination. We aim to identify the channel conditions under which using both relays strictly improves the gDoF compared to the best-relay selection policy (which includes direct transmission from the source to the destination as a special case) that achieves

$$\begin{aligned}\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} &= \max\left\{1,\min\{\alpha_{s1},\alpha_{1d}\},\min\{\alpha_{s2},\alpha_{2d}\}\right\} \\ &\in [1,\mathsf{d}^{(\mathrm{FD})}_{N=2}]. \end{aligned} \qquad (15)$$

We distinguish the following cases:

Case 1): if

$$\text{either} \begin{cases} \alpha_{s1}\ge\alpha_{s2} \\ \alpha_{1d}\ge\alpha_{2d} \end{cases} \text{or} \begin{cases} \alpha_{s1}<\alpha_{s2} \\ \alpha_{1d}<\alpha_{2d} \end{cases}$$

then, since one of the relays is 'uniformly better' than the other, we immediately see that $\mathsf{d}^{(\mathrm{FD})}_{N=2}=\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}}$, so in this regime selecting the best relay for transmission is gDoF optimal.

$$
\begin{array}{ll}
\text{either} & \max\{1,\alpha_{s2},\alpha_{1d}\} < \min\{\alpha_{s1},\alpha_{2d}\} \le \max\left\{\alpha_{s2}+\alpha_{1d},\beta_2+1\right\} \\
\text{or} & \left\{ \max\left\{\alpha_{s2}+\alpha_{1d},\beta_2+1\right\} < \min\{\alpha_{s1},\alpha_{2d}\}\right\} \cap \mathcal{O}^c \\
\text{where} & \mathcal{O} := \{\beta_2=0,\alpha_{s2}+\alpha_{1d}\le 1\} \cup \{\alpha_{1d}=0,\beta_2+1\le\alpha_{s2}\} \cup \{\alpha_{s2}=0,\beta_2+1\le\alpha_{1d}\}, \\
\text{that is, for} & \max\{1,\alpha_{s2},\alpha_{1d}\} < \min\{\alpha_{s1},\alpha_{2d}\} \text{ except in region } \mathcal{O}
\end{array} \quad (16)
$$

Case 2): if not in Case 1, then we are in

$$\text{either} \begin{cases} \alpha_{s1} \ge \alpha_{s2} \\ \alpha_{1d} < \alpha_{2d} \end{cases} \text{or} \begin{cases} \alpha_{s1} < \alpha_{s2} \\ \alpha_{1d} \ge \alpha_{2d} \end{cases}.$$

Consider the case $\alpha_{s2} \le \alpha_{s1},\ \alpha_{1d} < \alpha_{2d}$ (the other one is obtained essentially by swapping the role of the relays). This corresponds to an 'asymmetric' situation where relay 1 has the best link from the source but relay 2 has the best link to the destination. In this case we would like to exploit the inter relay communication links (which are not present in a diamond network) to create a route source→relay1→relay2→destination in addition to the direct link source→destination. Indeed, in this case $\mathsf{d}^{(\mathrm{FD})}_{N=2}$ in (14) can be rewritten as

$$\mathsf{d}^{(\mathrm{FD})}_{N=2} = \min\Big\{ \max\left\{\alpha_{s2}+\alpha_{1d},\beta_2+1\right\}, \max\{1,\min\{\alpha_{s1},\alpha_{2d}\}\}\Big\}, \quad (17)$$

where the term $\max\{1,\min\{\alpha_{s1},\alpha_{2d}\}\}$ in (17) corresponds to the gDoF of a virtual single-relay channel such that the link from the source to the "virtual relay" has SNR-exponent α_{s1} and the link from the "virtual relay" to the destination has SNR-exponent α_{2d}. We aim to determine the subset of the channel parameters $\alpha_{s2} \le \alpha_{s1},\ \alpha_{1d} < \alpha_{2d}$ for which the gDoF in (17) is strictly larger than the 'best relay' gDoF in (15). The case $\alpha_{s2} \le \alpha_{s1},\ \alpha_{1d} < \alpha_{2d}$ subsumes the following possible orders of the channel gains:

case i	α_{1d}	α_{2d}	α_{s2}		α_{s1}		
case ii	α_{1d}		α_{s2}	α_{2d}	α_{s1}		
case iii	α_{1d}		α_{s2}		α_{s1}	α_{2d}	
case iv			α_{s2}	α_{1d}	α_{2d} α_{s1}		
case v			α_{s2}	α_{1d}	α_{s1}	α_{2d}	
case vi			α_{s2}		α_{s1}	α_{1d}	α_{2d}

We partition the set of channel parameters $\alpha_{s2} \le \alpha_{s1},\ \alpha_{1d} < \alpha_{2d}$ as follows:

- *Sub-case 2a)* (all but cases i and vi in the table): if
$$\max\{\alpha_{s2},\alpha_{1d}\} < \min\{\alpha_{s1},\alpha_{2d}\}, \quad (18)$$
then
$$\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} = \max\{1,\alpha_{s2},\alpha_{1d}\}, \quad (19)$$
which is strictly less than $\mathsf{d}^{(\mathrm{FD})}_{N=2}$ in (17) if (16), at the top of the page, holds.
- *Sub-case 2b)* (case i in the table above): if $\alpha_{1d} < \alpha_{2d} \le \alpha_{s2} \le \alpha_{s1}$, then the condition
$$\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} = \max\{1,\alpha_{2d}\} < \mathsf{d}^{(\mathrm{FD})}_{N=2} = \min\Big\{ \max\left\{\alpha_{s2}+\alpha_{1d},\beta_2+1\right\}, \max\{1,\alpha_{2d}\}\Big\}$$
is never verified, i.e., in this case $\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} = \mathsf{d}^{(\mathrm{FD})}_{N=2}$.
- *Sub-case 2c)* (case vi in the table above): if $\alpha_{s2} \le \alpha_{s1} \le \alpha_{1d} < \alpha_{2d}$, then
$$\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} = \max\{1,\alpha_{s1}\} < \mathsf{d}^{(\mathrm{FD})}_{N=2} = \min\Big\{ \max\left\{\alpha_{s2}+\alpha_{1d},\beta_2+1\right\}, \max\{1,\alpha_{s1}\}\Big\}$$
is never verified, i.e., in this case $\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} = \mathsf{d}^{(\mathrm{FD})}_{N=2}$.

To summarize, for a 2-relay network where the relays operate in FD, using both relays gives a strictly larger gDoF compared to only exploiting the best one if

$$\max\{1,\alpha_{s2},\alpha_{1d}\} < \min\{\alpha_{s1},\alpha_{2d}\} \text{ except for} \quad (20a)$$
$$\mathcal{O} := \{\beta_2=0,\alpha_{s2}+\alpha_{1d}\le 1\} \cup \{\alpha_{1d}=0,\beta_2+1\le\alpha_{s2}\} \cup \{\alpha_{s2}=0,\beta_2+1\le\alpha_{1d}\}. \quad (20b)$$

Recall that there is also a regime similar to (20) where the role of the relays is swapped.

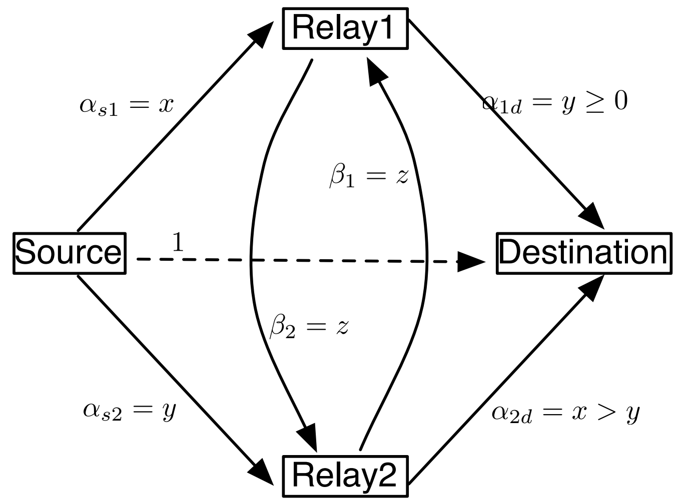

Fig. 4: Example of a two-relay network.

Fig. 4 gives an example of a network satisfying the conditions in (18), i.e., the assumption is $0 \le y < x$ without loss of generality. This is an 'asymmetric' network, i.e., one relay has the best link from the source and the other relay has the best link to the destination. By exploiting both relays, the system attains

$$\mathsf{d}^{(\mathrm{FD})}_{N=2} = \min\Big\{ \max\{1,x,y\}, \max\{2x,z+1\}, \max\{2y,z+1\}\Big\} = \min\Big\{ \max\{1,x\}, \max\{2y,z+1\}\Big\},$$

while, by using only the best relay, it achieves

$$\mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}} = \max\left\{1,\min\{x,y\}\right\} = \max\left\{1,y\right\}.$$

By (16), we have $\mathsf{d}^{(\mathrm{FD})}_{N=2} > \mathsf{d}^{(\mathrm{FD})}_{N=2,\text{best relay}}$ if

$$x > \max\left\{1, y\right\} \text{ except for } \left\{z = 0, y \leq \frac{1}{2}\right\}. \tag{21}$$

Note that a non-zero link between relay1 and relay2 allows to route the information through the path source→relay1→relay2→destination, which leads to an increase in terms of gDoF with respect to the best relay selection strategy.

B. The half-duplex case

With HD, the gDoF is given by (12), which with the notation in (13) becomes

$$\begin{aligned}\mathsf{d}^{(\mathrm{HD})}_{N=2} = \max\min\Big\{&\lambda_{00}D_1^{(0)}+\lambda_{01}D_1^{(1)}+\lambda_{10}D_1^{(2)}+\lambda_{11}D_1^{(3)},\\ &\lambda_{00}D_2^{(0)}+\lambda_{01}D_2^{(1)}+\lambda_{10}D_2^{(2)}+\lambda_{11}D_2^{(3)},\\ &\lambda_{00}D_3^{(0)}+\lambda_{01}D_3^{(1)}+\lambda_{10}D_3^{(2)}+\lambda_{11}D_3^{(3)},\\ &\lambda_{00}D_4^{(0)}+\lambda_{01}D_4^{(1)}+\lambda_{10}D_4^{(2)}+\lambda_{11}D_4^{(3)}\Big\}, \end{aligned} \tag{22}$$

where the maximization is over $\lambda_s, \ \forall s \in [0:1]^2$, with $\lambda_s = \mathbb{P}[S_{[1:2]} = s] \geq 0$, such that $\sum_{s\in[0:1]^2}\lambda_s = \lambda_{00} + \lambda_{01} + \lambda_{10} + \lambda_{11} = 1$, and

$$\begin{aligned} D_1^{(0)} &:= \max\left\{1, \alpha_{s1}, \alpha_{s2}\right\},\\ D_1^{(1)} = D_2^{(0)} &:= \max\left\{1, \alpha_{s1}\right\},\\ D_4^{(3)} &:= \max\left\{1, \alpha_{1d}, \alpha_{2d}\right\},\\ D_1^{(2)} = D_3^{(0)} &:= \max\left\{1, \alpha_{s2}\right\},\\ D_2^{(1)} &:= \max\left\{\alpha_{s1} + \alpha_{2d}, \beta_1 + 1\right\},\\ D_2^{(3)} = D_4^{(1)} &:= \max\left\{1, \alpha_{2d}\right\},\\ D_3^{(2)} &:= \max\left\{\alpha_{s2} + \alpha_{1d}, \beta_2 + 1\right\},\\ D_3^{(3)} = D_4^{(2)} &:= \max\left\{1, \alpha_{1d}\right\},\\ D_1^{(3)} = D_2^{(2)} = D_3^{(1)} = D_4^{(0)} &:= 1. \end{aligned}$$

For future reference, if only one relay helps the communication between the source and the destination then the achievable gDoF is [30]

$$\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}} = 1 + \max_{i\in[1:2]} \frac{[\alpha_{si}-1]^+ \ [\alpha_{id}-1]^+}{[\alpha_{si}-1]^+ + [\alpha_{id}-1]^+} \in [1, \mathsf{d}^{(\mathrm{HD})}_{N=2}]. \tag{23}$$

An analytical closed form solution for the optimal $\{\lambda_s\}$ in (22) is complex to find for general channel gain assignments. However, numerically it is a question of solving a LP, for which efficient numerical routines exist. By using Theorem 4, we can set either λ_{00} or λ_{11} to zero.

For the example in Fig. 4 the optimal schedule has $\lambda_{00} = \lambda_{11} = 0$ without loss of optimality, from Theorem 4. By letting $\lambda_{01} = \gamma \in [0, 1]$ and $\lambda_{10} = 1 - \gamma$ (recall $0 \leq y < x$ without

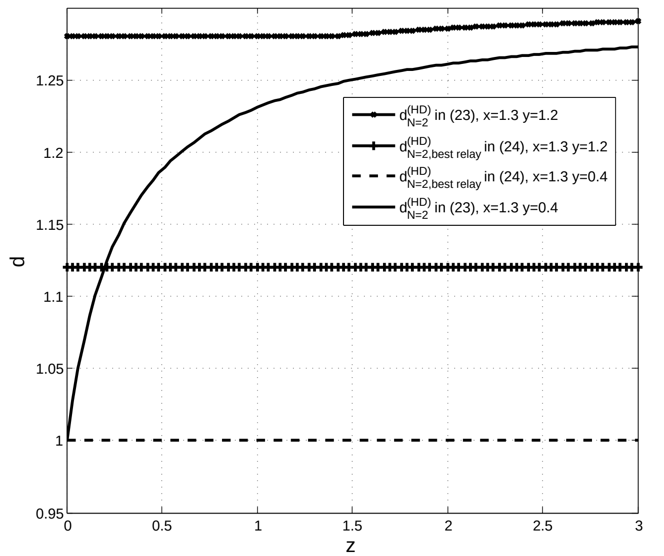

Fig. 5: $\mathsf{d}^{(\mathrm{HD})}_{N=2}$ in (24) and $\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}}$ in (25) for different values of $z \in [0, 3]$ in Fig. 4 and for $x = 1.3$, $y = 0.4, 1.2$.

loss of generality), the gDoF in (22) can be written as

$$\begin{aligned} \mathsf{d}^{(\mathrm{HD})}_{N=2} = \max_{\gamma\in[0,1]}\min\Big\{&\gamma\max\{1,x\}+(1-\gamma)\max\{1,y\}, &\text{(24a)}\\ &\gamma\max\{2x, z+1\} + (1-\gamma), &\text{(24b)}\\ &\gamma + (1-\gamma)\max\{2y, z+1\}\Big\} &\text{(24c)}\\ = 1+\min\Big\{&\frac{[x-1]^+\max\{2y-1,z\}}{[x-1]^+ + \max\{2y-1,z\} - [y-1]^+}, &\text{(24d)}\\ &\frac{\max\{2x-1,z\}\max\{2y-1,z\}}{\max\{2x-1,z\}+\max\{2y-1,z\}}\Big\}. &\text{(24e)} \end{aligned}$$

By using only the best relay as in (23), we would achieve

$$\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}} = 1 + \frac{[x-1]^+[y-1]^+}{[x-1]^+ + [y-1]^+}. \tag{25}$$

It can be easily seen that the best relay selection policy is strictly suboptimal if (21) is verified, as for the FD case. Considerations similar to those made for the FD case, can be made for the HD case as well. Fig. 5 shows, for different values of z, i.e., strength of the links between the two relays, the behaviors of $\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}}$ in (25) and $\mathsf{d}^{(\mathrm{HD})}_{N=2}$ in (24). Regarding the curves with $y = 0.4$, since we have $y < \frac{1}{2}$ and hence $\max\{2y-1, z\} = z, \forall z \geq 0$, $\mathsf{d}^{(\mathrm{HD})}_{N=2}$ in (24) is an increasing function of z. On the other hand, since $y < 1$, $\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}}$ in (25) is always equal to 1, i.e., direct transmission is gDoF optimal. We also notice that for $z = 0$, the two curves overlap since the condition in (21) holds. Regarding the curves with $y = 1.2$, we notice that $\mathsf{d}^{(\mathrm{HD})}_{N=2}$ in (24) is always strictly greater than $\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}}$ in (25), i.e., the channel conditions are such that the synergies between the two relays bring to an unbounded rate gain with respect to best relay selection. Moreover, $\mathsf{d}^{(\mathrm{HD})}_{N=2}$ in (24) starts to increase with z, when $\min\left\{\max\{2y-1, z\}, \max\{2x-1, z\}\right\} = \max\{2y-1, z\} = z$, i.e., $z = 1.4$ and best relay selection is always gDoF-wise greater than direct transmission, i.e., $\mathsf{d}^{(\mathrm{HD})}_{N=2,\text{best relay}} > 1$, since $\min\{x, y\} > 1$.

$$\mathcal{C}_{\text{multicast}} \supseteq \bigcup \Bigg\{ \sum_{i\in\mathcal{A}} R_i \leq I(X_{\mathcal{A}};\widehat{Y}_{\mathcal{A}^c}|X_{\mathcal{A}^c},S_{[1:K]},Q) - I(Y_{\mathcal{A}};\widehat{Y}_{\mathcal{A}}|\widehat{Y}_{\mathcal{A}^c},X_{[1:K]},S_{[1:K]},Q)$$
$$\text{such that } \mathcal{A}\subseteq[1:K],\ \mathcal{A}\neq\emptyset,\ \mathcal{A}^c\cap\mathcal{D}\neq\emptyset \Bigg\} \quad (26)$$

$$\mathcal{C}_{\text{multicast}} \supseteq \bigcup \Bigg\{ \sum_{i\in\mathcal{A}} R_i \ \leq \sum_{s\in[0:1]^K} \lambda_s \ \log\left|\mathbf{I}_{|\mathcal{A}^c|} + \frac{1}{1+\sigma^2}\ \mathbf{H}_{\mathcal{A},s}\mathbf{H}^H_{\mathcal{A},s}\right| - |\mathcal{A}|\log\left(1+\frac{1}{\sigma^2}\right)$$
$$\text{such that } \mathcal{A}\subseteq[1:K],\ \mathcal{A}\neq\emptyset,\ \mathcal{A}^c\cap\mathcal{D}\neq\emptyset \Bigg\} \quad (27)$$

VI. Conclusions

In this work we analyzed a network where a source communicates with a destination across a Gaussian channel and is assisted by N relays operating in half-duplex mode. We characterized the capacity to within a constant gap by using noisy network coding as achievable scheme. We also showed that this gap may be further reduced by considering more structured systems, such as the diamond network. We conjectured that the optimal schedule has at most $N+1$ active states, instead of the possible 2^N. This conjecture has been supported by the analytical proof in the special case of $N=2$ relays and in general by numerical evaluations. We finally analyzed a network with $N=2$ relays, and we showed under which channel conditions by exploiting both relays a strictly greater gDoF can be attained compared to a network where best-relay selection is used.

An interesting connection between the high-SNR approximation of the point-to-point MIMO capacity and the Maximum Weighted Bipartite Matching problem from graph theory has been discovered.

Appendix A
Gap result for Gaussian multicast HD networks and Proof of Theorem 1

Here we slightly change the notation compared to the channel model introduced in Section II-B. In particular, for notation convenience, we number the nodes from 1 to K, rather than from 0 to $N+1$. We aim to obtain a constant gap result for this class of networks that only depends on K, the total number of nodes, and for which the way the nodes are numbered is irrelevant.

A. Channel Model

A multicast network with K nodes is defined similarly to the multi-relay network introduced in Section II-B except that now each node $k\in[1:K]$, with channel input (X_k,S_k) and channel output Y_k, has an independent message of rate R_k to be decoded by the nodes indexed by $\mathcal{D}\subseteq[1:K]$. The channel input/output relationship of this HD Gaussian multicast network reads $\mathbf{Y}=(\mathbf{I}_K-\text{diag}[\mathbf{S}])\ \mathbf{H}\ \text{diag}[\mathbf{S}]\ \mathbf{X}+\mathbf{Z}$. We let $\mathcal{C}_{\text{multicast}}$ be the capacity region.

B. Inner Bound

The capacity of a HD Gaussian multicast network can be lower bounded by adapting the NNC scheme for the general memoryless network from [13] to the HD case by following the approach of [8]. In particular, NNC achieves the rate region (26) at the top of the page, where $\widehat{Y}_k$ represents a compressed version of Y_k, for $k\in[1:K]$, and where the union is over all input distributions that factorize as $\mathbb{P}_Q\prod_{k=1}^K \mathbb{P}_{X_k,S_k|Q}\mathbb{P}_{\widehat{Y}_k|Y_k,X_k,S_k,Q}$ and satisfy the power constraints. We consider jointly Gaussian inputs so as to get a rate region similar to [13, eq.(20)]. In all states $s\in[0:1]^K$, we consider i.i.d. $\mathcal{N}(0,1)$ inputs, time sharing random variable Q set to $Q=S_{[1:K]}$ (with this choice the nodes can coordinate), and compressed channel output $\widehat{Y}_k:=Y_k+\widehat{Z}_k$, $k\in[1:K]$, for $\widehat{Z}_k\sim\mathcal{N}(0,\sigma^2)$ independent of all other random variables and where the variance of $\widehat{Z}_k$ does not depend on the user index k. With this the NNC achievable region evaluates to (27) at the top of the page, where the union is over all $\lambda_s:=\mathbb{P}[S_{[1:K]}=s]\in[0,1],\ \forall s\in[0:1]^K:\ \sum_{s\in[0:1]^K}\lambda_s=1$ and over all $\sigma^2\in\mathbb{R}^+$, and where the matrix $\mathbf{H}_{\mathcal{A},s}\in\mathbb{C}^{|\mathcal{A}^c|\times|\mathcal{A}|}$ is defined as $\mathbf{H}_{\mathcal{A},s}:=\left[(\mathbf{I}_K-\text{diag}[s])\ \mathbf{H}\ \text{diag}[s]\right]_{\mathcal{A}^c,\mathcal{A}}$.

C. Outer Bound

The cut-set upper bound, adapted to the HD case by following [8], gives (28) at the top of the next page, where the union is over all joint input distributions $\mathbb{P}_{X^K,S^K}$ and satisfy the power constraints. Similarly to [13, eq.(19)], we upper bound each mutual information term as

$$\begin{aligned}
&I(X_{\mathcal{A}},S_{\mathcal{A}};Y_{\mathcal{A}^c}|X_{\mathcal{A}^c},S_{\mathcal{A}^c})\\
&=I(S_{\mathcal{A}};Y_{\mathcal{A}^c}|X_{\mathcal{A}^c},S_{\mathcal{A}^c})+I(X_{\mathcal{A}};Y_{\mathcal{A}^c}|X_{\mathcal{A}^c},S_{[1:K]})\\
&\leq H(S_{\mathcal{A}})+\sum_{s\in[0:1]^K}\lambda_s\ \log\left|\mathbf{I}_{|\mathcal{A}^c|}+\mathbf{H}_{\mathcal{A},s}\mathbf{K}_{\mathcal{A},s}\mathbf{H}^H_{\mathcal{A},s}\right| \quad (31a)\\
&\leq|\mathcal{A}|\log(2)+\sum_{s\in[0:1]^K}\lambda_s\ \log\left|\mathbf{I}_{|\mathcal{A}^c|}+\frac{1}{\gamma}\ \mathbf{H}_{\mathcal{A},s}\mathbf{H}^H_{\mathcal{A},s}\right|\\
&\quad+\sum_{s\in[0:1]^K}\lambda_s\ |\mathcal{A}|\ \frac{\log\left(\mathrm{e}\max\left\{1,\frac{\gamma}{\mathrm{e}}\ \frac{|\mathcal{A}|}{\text{rank}[\mathbf{H}_{\mathcal{A},s}]}\right\}\right)}{\max\left\{\frac{\mathrm{e}}{\gamma},\frac{|\mathcal{A}|}{\text{rank}[\mathbf{H}_{\mathcal{A},s}]}\right\}}, \quad (31b)
\end{aligned}$$

where: (i) $\mathbf{K}_{\mathcal{A},s}$ represents the covariance matrix of $X_{\mathcal{A}}$ conditioned on $S_{[1:K]}=s$; (ii) the inequality in (31a) follows

$$\mathcal{C}_{\text{multicast}} \subseteq \bigcup \left\{ \sum_{i\in\mathcal{A}} R_i \leq I(X_{\mathcal{A}}, S_{\mathcal{A}}; Y_{\mathcal{A}^c}|X_{\mathcal{A}^c}, S_{\mathcal{A}^c}) \text{ such that } \mathcal{A} \subseteq [1:K], \mathcal{A} \neq \emptyset, \mathcal{A}^c \cap \mathcal{D} \neq \emptyset \right\} \tag{28}$$

$$\begin{aligned} \mathcal{C}_{\text{multicast}} \subseteq \bigcup \Bigg\{ & \sum_{i\in\mathcal{A}} R_i \leq |\mathcal{A}|\log(2) + \sum_{s\in[0:1]^K} \lambda_s \ \log\left|\mathbf{I}_{|\mathcal{A}^c|} + \frac{1}{\gamma}\ \mathbf{H}_{\mathcal{A},s}\mathbf{H}_{\mathcal{A},s}^H\right| \\ & +|\mathcal{A}| \ \frac{\log\left(\mathrm{e}\max\left\{1, \frac{\gamma}{\mathrm{e}}\ \frac{|\mathcal{A}|}{\min\{|\mathcal{A}|,|\mathcal{A}^c|\}}\right\}\right)}{\max\left\{\frac{\mathrm{e}}{\gamma}, \frac{|\mathcal{A}|}{\min\{|\mathcal{A}|,|\mathcal{A}^c|\}}\right\}} \text{ such that } \mathcal{A} \subseteq [1:K],\ \mathcal{A} \neq \emptyset,\ \mathcal{A}^c \cap \mathcal{D} \neq \emptyset \Bigg\} \end{aligned} \tag{29}$$

$$\frac{\mathsf{GAP}}{K} \leq \min_{\gamma\geq \mathrm{e}-1} \max_{\mu\in[0,1]} \left\{\mu\ \log\left(\frac{2\gamma}{\gamma-1}\right) + \mu\ \min\left\{\frac{\gamma}{\mathrm{e}}, \frac{\min\{\mu,1-\mu\}}{\mu}\right\} \log\left(\max\left\{\mathrm{e}, \frac{\gamma\ \mu}{\min\{\mu,1-\mu\}}\right\}\right)\right\} \tag{30a}$$

$$\leq 1.96 \text{ bits/node} \tag{30b}$$

since conditioning reduces the entropy, since the entropy of a discrete random variable is non-negative, and by using the 'Gaussian maximizes entropy' principle; (iii) the inequality in (31b) follows since the entropy of a discrete random variable can be upper bounded as a function of the size of its support and from [13, Lemma 1] for all $\gamma \geq \mathrm{e}-1$. Finally, since the function $\frac{\log(\mathrm{e}\ \max\{1,x\})}{\max\{1,x\}}$ in (31b) is decreasing in x, the function in (31b) attains its maximum value when $\text{rank}[\mathbf{H}_{\mathcal{A},s}]$ is maximum, i.e., when $x = \frac{\gamma}{\mathrm{e}}\ \frac{|\mathcal{A}|}{\text{rank}[\mathbf{H}_{\mathcal{A},s}]} = \frac{\gamma}{\mathrm{e}}\ \frac{|\mathcal{A}|}{\min\{|\mathcal{A}|,|\mathcal{A}^c|\}}$, from which it thus follows that (29), at the top of the page, holds, where the union is over all $\lambda_s := \mathbb{P}[S_{[1:K]} = s] \in [0,1],\ \forall s \in [0:1]^K : \sum_{s\in[0:1]^K} \lambda_s = 1$ and where the parameter $\gamma \geq \mathrm{e}-1$ can be chosen to tighten the right hand side of (29).

D. Gap

We now proceed to bound the worst case gap (over $\mathcal{A}$) between the cut-set upper bound in (29) and the NNC lower bound in (27) (recall that the parameters γ and σ^2 can be chosen so as to tighten the bound). By choosing $\sigma^2 = \gamma - 1$ in (27) and by defining $\mu = \frac{|\mathcal{A}|}{K} \in [0,1]$, the gap is given by (30) at the top of the page, where the inequality in (30b) follows by numerical evaluations.

E. The HD multi-relay network

The unicast Gaussian network with HD relays is a special multicast network with $K = N+2$ nodes (i.e., one source, N relays, and one destination). With a single source, the inner bound in (27) allows to evaluate $R^{(\text{in})}$ in (3), while the upper bound in (29) allows to evaluate $R^{(\text{out})}$ in (3). From (30b) with $K = N+2$ nodes we deduce that the GAP in (3) for the HD multi-relay network is as in (5).

The difference between the HD and the FD case is the factor 2 in (30a) for the HD case. Also notice that the HD gap of 1.96 bits/node is smaller than $(1+1.26)$ bits/node where 1.26 bits/node is the FD gap [13] and the extra 1 bit/node is due to random switch.

APPENDIX B
PROOF OF THEOREM 2

Let $\mathcal{S}_{n,k}$ be the set of all k-combinations of the integers in $[1:n]$ and $\mathcal{P}_{n,k}$ be the set of all k-permutations of the integers in $[1:n]$. Let $\sigma(\pi)$ be the sign / signature of the permutation π.

We start by demonstrating that the asymptotic behavior of $|\mathbf{I}_k + \mathbf{H}\mathbf{H}^H|$ is as that of $|\mathbf{H}\mathbf{H}^H|$, i.e., the identity matrix can be neglected. By using the determinant Leibniz formula [37], in fact we have,

$$\begin{aligned} |\mathbf{I}_k + \mathbf{H}\mathbf{H}^H| &= \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \prod_{i=1}^{k} \left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)} \\ &= \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \left\{ \left(\prod_{i=2}^{k} \left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)}\right) \left(\left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{1,\pi(1)}\right) \right\} \\ &= \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \left(\prod_{i=2}^{k} \left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)}\right) \delta[1-\pi(1)] \\ &+ \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \prod_{i=2}^{k} \left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)} \left[\mathbf{H}\mathbf{H}^H\right]_{1,\pi(1)}. \end{aligned}$$

Let

$$\mathsf{A}\,(\mathsf{SNR}) := \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \left(\prod_{i=2}^{k} \left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)}\right) \delta[1-\pi(1)]$$

$$\mathsf{B}\,(\mathsf{SNR}) := \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \prod_{i=2}^{k} \left[\mathbf{I}_k + \mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)} \left[\mathbf{H}\mathbf{H}^H\right]_{1,\pi(1)},$$

we have that $\mathsf{A}\,(\mathsf{SNR}) = o\,(\mathsf{B}\,(\mathsf{SNR}))$, because $\lim_{\mathsf{SNR}\to+\infty} \frac{\mathsf{A}(\mathsf{SNR})}{\mathsf{B}(\mathsf{SNR})} = 0$ where the SNR parameterizes the channel gains as $|h_{ij}|^2 = \mathsf{SNR}^{\beta_{ij}}$, for some non-negative β_{ij}. This is so because, as a function of SNR, $\mathsf{B}\,(\mathsf{SNR})$ grows faster than $\mathsf{A}\,(\mathsf{SNR})$ due to the term $\left[\mathbf{H}\mathbf{H}^H\right]_{1,\pi(1)}$. By induction it is possible to show that this

$$
\begin{aligned}
|\mathbf{H}\mathbf{H}^H| &\overset{(a)}{=} \sum_{\varsigma\in\mathcal{S}_{n,k}} |\mathbf{H}_\varsigma||\mathbf{H}_\varsigma^H| = \sum_{\varsigma\in\mathcal{S}_{n,k}} |\mathbf{H}_\varsigma|^2 \overset{(b)}{=} \sum_{\varsigma\in\mathcal{S}_{n,k}} \left| \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \prod_{i=1}^k [\mathbf{H}_\varsigma]_{i,\pi(i)} \right|^2 \\
&= \sum_{\varsigma\in\mathcal{S}_{n,k}} \left\{ \left(\sum_{\pi_1\in\mathcal{P}_{n,k}} \sigma(\pi_1) \prod_{i=1}^k [\mathbf{H}_\varsigma]_{i,\pi_1(i)} \right) \left(\sum_{\pi_2\in\mathcal{P}_{n,k}} \sigma(\pi_2) \prod_{j=1}^k [\mathbf{H}_\varsigma]_{j,\pi_2(j)} \right)^* \right\} \\
&= \sum_{\varsigma\in\mathcal{S}_{n,k}} \left\{ \left(\sum_{\pi\in\mathcal{P}_{n,k}} \prod_{i=1}^k \left| [\mathbf{H}_\varsigma]_{i,\pi(i)} \right|^2 \right) + \left(\sum_{\pi_1,\pi_2\in\mathcal{P}_{n,k},\pi_1\neq\pi_2} \sigma(\pi_1)\sigma(\pi_2) \prod_{i=1}^k \prod_{j=1}^k [\mathbf{H}_\varsigma]_{i,\pi_1(i)} \left([\mathbf{H}_\varsigma]_{j,\pi_2(j)} \right)^* \right) \right\} \\
&\overset{(c)}{\leq} \sum_{\varsigma\in\mathcal{S}_{n,k}} \left\{ \left(\sum_{\pi\in\mathcal{P}_{n,k}} \prod_{i=1}^k \left| [\mathbf{H}_\varsigma]_{i,\pi(i)} \right|^2 \right) + \left(\sum_{\pi_1,\pi_2\in\mathcal{P}_{n,k},\pi_1\neq\pi_2} \prod_{i=1}^k \prod_{j=1}^k \sqrt{ \left| [\mathbf{H}_\varsigma]_{i,\pi_1(i)} \right|^2 \left| [\mathbf{H}_\varsigma]_{j,\pi_2(j)} \right|^2 } \right) \right\} \\
&= \sum_{\varsigma\in\mathcal{S}_{n,k}} \left\{ \left(\sum_{\pi\in\mathcal{P}_{n,k}} \mathsf{SNR}^{\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi(i)}} \right) + \left(\sum_{\pi_1,\pi_2\in\mathcal{P}_{n,k},\pi_1\neq\pi_2} \mathsf{SNR}^{\frac{1}{2}\left(\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi_1(i)} + \sum_{j=1}^k [\mathbf{B}_\varsigma]_{j,\pi_2(j)} \right)} \right) \right\} \\
&\overset{(d)}{\doteq} \sum_{\varsigma\in\mathcal{S}_{n,k}} \left(\sum_{\pi\in\mathcal{P}_{n,k}} \mathsf{SNR}^{\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi(i)}} \right) \doteq \mathsf{SNR}^{\max_{\varsigma\in\mathcal{S}_{n,k}} \max_{\pi\in\mathcal{P}_{n,k}} \sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi(i)}}
\end{aligned} \tag{32}
$$

reasoning holds $\forall i \in [1:k]$ and hence

$$
|\mathbf{I}_k + \mathbf{H}\mathbf{H}^H| \doteq \sum_{\pi\in\mathcal{P}_{n,k}} \sigma(\pi) \prod_{i=1}^k \left[\mathbf{H}\mathbf{H}^H\right]_{i,\pi(i)} = |\mathbf{H}\mathbf{H}^H|.
$$

Therefore, we now focus on the study of $|\mathbf{H}\mathbf{H}^H|$. We have that (32), at the top of the page, holds, where the equalities / inequalities above are due to the following facts:

- equality (a): by applying the Cauchy-Binet formula [37] where $\mathbf{H}_\varsigma$ is the square matrix obtained from $\mathbf{H}$ by retaining all rows and those columns indexed by ς;
- equality (b): by applying the determinant Leibniz formula [37];
- inequality (c): by applying the Cauchy-Schwarz inequality [38];
- equality (d): when $\mathsf{SNR} \to \infty$, we have

$$
\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi(i)} \geq \frac{1}{2} \left(\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi_1(i)} + \sum_{j=1}^k [\mathbf{B}_\varsigma]_{j,\pi_2(j)} \right).
$$

Consider the following example. Let $|a|^2 = \mathsf{SNR}^{\beta_a}$, $|b|^2 = \mathsf{SNR}^{\beta_b}$, $|c|^2 = \mathsf{SNR}^{\beta_c}$, $|d|^2 = \mathsf{SNR}^{\beta_d}$

$$
|ab - cd|^2 \leq |a|^2|b|^2 + |c|^2|d|^2 + 2|a||b||c||d|.
$$

Now apply the gDoF formula, i.e.,

$$
\begin{aligned}
\mathsf{d} &:= \lim_{\mathsf{SNR}\to+\infty} \frac{\log\left(|a|^2|b|^2 + |c|^2|d|^2 + 2|a||b||c||d|\right)}{\log(1+\mathsf{SNR})} \\
&= \max\left\{ \beta_a + \beta_b,\ \beta_c + \beta_d,\ \frac{\beta_a + \beta_b + \beta_c + \beta_d}{2} \right\},
\end{aligned}
$$

but

$$
\begin{aligned}
\frac{\beta_a + \beta_b + \beta_c + \beta_d}{2} &\leq \frac{2\max\left\{\beta_a + \beta_b, \beta_c + \beta_d\right\}}{2} \\
&= \max\left\{\beta_a + \beta_b, \beta_c + \beta_d\right\}.
\end{aligned}
$$

Therefore, the term $\frac{\beta_a+\beta_b+\beta_c+\beta_d}{2}$ does not contribute in characterizing the gDoF. By direct induction, the above reasoning may be extended to a general number of terms leading to $\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi(i)} \geq \frac{1}{2}\left(\sum_{i=1}^k [\mathbf{B}_\varsigma]_{i,\pi_1(i)} + \sum_{j=1}^k [\mathbf{B}_\varsigma]_{j,\pi_2(j)} \right)$.

Appendix C
Proof of Theorem 4

In a HD relay network with $N = 2$, we have $2^N = 4$ possible states that may arise with probabilities $\lambda_s,\ \forall s \in [0:1]^2$, with $\lambda_s = \mathbb{P}[S_{[1:2]} = s] \geq 0$, such that $\sum_{s\in[0:1]^2} \lambda_s = \lambda_{00} + \lambda_{01} + \lambda_{10} + \lambda_{11} = 1$. Here we aim to demonstrate that a schedule with $\lambda_{00}\lambda_{11} = 0$ is optimal.

We start by solving a fairly general LP, which we will then specialize to the gDoF case as well as to the case of a general finite SNR.

A. General LP

Consider the LP in (33) at the top of the next page, where the different quantities $(a_u, b_u),\ u \in [1:2]$, and $D_v,\ v \in [1:4]$, are non-negative and will be defined later.

The proof is by contradiction. Assume that $\left[\lambda_{00}^{(\text{opt})}\ \ \lambda_{10}^{(\text{opt})}\ \ \lambda_{01}^{(\text{opt})}\ \ \lambda_{11}^{(\text{opt})}\right]$ is the optimal solution with $\lambda_{00}^{(\text{opt})} > 0$. This implies that for any $(\alpha, \beta, \gamma) \in [0,1]^3$ such that $\alpha + \beta + \gamma = 1$ we must have that (34), at the top of the next page, holds. Since $\lambda_{00}^{(\text{opt})} > 0$ by assumption, we can rewrite the above problem as in (35) at the top of the next page, for all $(\alpha, \beta, \gamma) \in [0,1]^3$ such that $\alpha + \beta + \gamma = 1$.

If we can find a triplet $(\alpha, \beta, \gamma) \in [0,1]^3 : \alpha + \beta + \gamma = 1$ for which (36), at the top of the next page, holds we reach a contradiction; hence, for this set of values we must have $\lambda_{00}^{(\text{opt})} = 0$.

$$\text{LP: } \max_{\lambda'\text{s}} \min \left\{ \begin{bmatrix} \max\{a_1,a_2\}+D_1 & a_2 & a_1 & 0 \\ a_2 & a_2+b_1+D_2 & 0 & b_1 \\ a_1 & 0 & a_1+b_2+D_3 & b_2 \\ 0 & b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} \lambda_{00} \\ \lambda_{10} \\ \lambda_{01} \\ \lambda_{11} \end{bmatrix} \right\} \tag{33}$$

$$\begin{aligned} &\min \left\{ \begin{bmatrix} \max\{a_1,a_2\}+D_1 & a_2 & a_1 & 0 \\ a_2 & a_2+b_1+D_2 & 0 & b_1 \\ a_1 & 0 & a_1+b_2+D_3 & b_2 \\ 0 & b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} \lambda_{00}^{(\text{opt})} \\ \lambda_{10}^{(\text{opt})} \\ \lambda_{01}^{(\text{opt})} \\ \lambda_{11}^{(\text{opt})} \end{bmatrix} \right\} \\ \geq &\min \left\{ \begin{bmatrix} \max\{a_1,a_2\}+D_1 & a_2 & a_1 & 0 \\ a_2 & a_2+b_1+D_2 & 0 & b_1 \\ a_1 & 0 & a_1+b_2+D_3 & b_2 \\ 0 & b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} 0 \\ \lambda_{10}^{(\text{opt})}+\lambda_{00}^{(\text{opt})}\alpha \\ \lambda_{01}^{(\text{opt})}+\lambda_{00}^{(\text{opt})}\beta \\ \lambda_{11}^{(\text{opt})}+\lambda_{00}^{(\text{opt})}\gamma \end{bmatrix} \right\} \end{aligned} \tag{34}$$

$$\begin{aligned} 0 = &\min \left\{ \begin{bmatrix} \max\{a_1,a_2\}+D_1 & a_2 & a_1 & 0 \\ a_2 & a_2+b_1+D_2 & 0 & b_1 \\ a_1 & 0 & a_1+b_2+D_3 & b_2 \\ 0 & b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\} \\ \geq &\min \left\{ \begin{bmatrix} \max\{a_1,a_2\}+D_1 & a_2 & a_1 & 0 \\ a_2 & a_2+b_1+D_2 & 0 & b_1 \\ a_1 & 0 & a_1+b_2+D_3 & b_2 \\ 0 & b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha \\ \beta \\ \gamma \end{bmatrix} \right\} \\ = &\min \left\{ \begin{bmatrix} a_2 & a_1 & 0 \\ a_2+b_1+D_2 & 0 & b_1 \\ 0 & a_1+b_2+D_3 & b_2 \\ b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \right\} \end{aligned} \tag{35}$$

$$\min \left\{ \begin{bmatrix} a_2 & a_1 & 0 \\ a_2+b_1+D_2 & 0 & b_1 \\ 0 & a_1+b_2+D_3 & b_2 \\ b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \right\} > 0 \tag{36}$$

Assume $b_1 b_2 \leq a_1 a_2$ and define

$$\begin{aligned} \alpha &= \frac{b_2}{a_2+b_2}, \ \beta = \frac{b_1}{a_1+b_1}, \\ \gamma &= 1-\alpha-\beta = \frac{a_1a_2-b_1b_2}{(a_1+b_1)(a_2+b_2)} \end{aligned}$$

which is a valid assignment since all coefficients are non-negative and sum to one. With this we have that (37), at the top of the next page, holds. Hence, for $b_1b_2 \leq a_1a_2$ and $(a_1,a_2,b_1,b_2) \neq (0,0,0,0)$ we must have $\lambda_{00}^{(\text{opt})} = 0$. A similar reasoning shows that if $b_1b_2 \geq a_1a_2$ and $(a_1,a_2,b_1,b_2) \neq (0,0,0,0)$ we must have $\lambda_{11}^{(\text{opt})} = 0$. It is easy to show that if $\min\{a_1,a_2\} = 0$ then $\lambda_{00}^{(\text{opt})} = 0$, without loss of optimality. Similarly if $\min\{b_1,b_2\} = 0$ then $\lambda_{11}^{(\text{opt})} = 0$, without loss of optimality. This is because, under these conditions, one of the constraints in (33) becomes redundant and therefore, by contradiction, it is easy to show that either $\lambda_{00}^{(\text{opt})} = 0$ or $\lambda_{11}^{(\text{opt})} = 0$ is optimal.

B. The cut-set upper bound cast in the form of the general LP for a finite SNR

From our previous discussion, we restrict attention to the case $\min\{a_1,a_2,b_1,b_2\} \neq 0$. After straightforward manipulations the cut-set bound, for $N = 2$, can be further upper bounded as (38) at the top of the next page, where the term $\max_{S_1,S_2}\{I(X_\text{s};Y_\text{d}|X_1,X_2,S_1,S_2)\} \leq \log(1+|h_{33}|^2)$ and where

$$\begin{aligned} a'_2 &:= I(X_3;Y_2|Y_3,X_1,X_2,S_1=0,S_2=0), \\ c'_2 &:= I(X_3;Y_2|Y_3,X_1,X_2,S_1=1,S_2=0), \\ a'_1 &:= I(X_3;Y_1|Y_3,X_1,X_2,S_1=0,S_2=0), \\ d'_1 &:= I(X_3;Y_1|Y_3,X_1,X_2,S_1=0,S_2=1), \\ a_2 &:= \max\{a'_2,c'_2\}, \\ a_1 &:= \max\{a'_1,d'_1\}, \\ b'_1 &:= I(X_1;Y_3|X_2,S_1=1,S_2=1), \\ c'_1 &:= I(X_1;Y_3|X_2,S_1=1,S_2=2), \\ b'_2 &:= I(X_2;Y_3|X_1,S_1=1,S_2=1), \\ d'_2 &:= I(X_2;Y_3|X_1,S_1=0,S_2=1) \end{aligned}$$

$$
\begin{aligned}
&\max_{(\alpha,\beta,\gamma)\in[0,1]^3:\alpha+\beta+\gamma=1} \min\left\{\begin{bmatrix} a_2 & a_1 & 0 \\ a_2+b_1+D_2 & 0 & b_1 \\ 0 & a_1+b_2+D_3 & b_2 \\ b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix}\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix}\right\} \\
&\geq \min\left\{\begin{bmatrix} a_2 & a_1 & 0 \\ a_2+b_1+D_2 & 0 & b_1 \\ 0 & a_1+b_2+D_3 & b_2 \\ b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix}\begin{bmatrix}\frac{b_2}{a_2+b_2}\\ \frac{b_1}{a_1+b_1}\\ \frac{a_1a_2-b_1b_2}{(a_1+b_1)(a_2+b_2)}\end{bmatrix}\right\} \\
&= \min\left\{\begin{bmatrix} a_2 & a_1 & 0 \\ b_1 & b_2 & \max\{b_1,b_2\}+D_4 \end{bmatrix}\begin{bmatrix}\frac{b_2}{a_2+b_2}\\ \frac{b_1}{a_1+b_1}\\ \frac{a_1a_2-b_1b_2}{(a_1+b_1)(a_2+b_2)}\end{bmatrix}\right\} \\
&> 0 \text{ if } (a_1,a_2,b_1,b_2)\neq(0,0,0,0)
\end{aligned} \quad (37)
$$

$$
R^{(\text{cut-set},N=2)} \leq 2\log(2) + \max_{S_1,S_2}\{I(X_\text{s};Y_\text{d}|X_1,X_2,S_1,S_2)\} + \text{eq.}(33) \quad (38)
$$

and

$$
\begin{aligned}
b_1 &:= \max\{b_1',c_1'\}, \\
b_2 &:= \max\{b_2',d_2'\}, \\
D_1 &:= I(X_3;Y_1,Y_2|Y_3,X_1,X_2,S_1=0,S_2=0)+ \\
&\quad -\max\{I(X_3;Y_1|Y_3,X_1,X_2,S_1=0,S_2=0), \\
&\qquad I(X_3;Y_2|Y_3,X_1,X_2,S_1=0,S_2=0)\}, \\
D_2 &:= I(X_1;Y_2|Y_3,X_2,S_1=1,S_2=0), \\
D_3 &:= I(X_2;Y_1|Y_3,X_1,S_1=0,S_2=1), \\
D_4 &:= I(X_1,X_2;Y_3|S_1=1,S_2=1)+ \\
&\quad -\max\{I(X_1;Y_3|X_2,S_1=1,S_2=1), \\
&\qquad I(X_2;Y_3|X_1,S_1=1,S_2=1)\}.
\end{aligned}
$$

C. The cut-set upper bound cast in the form of the general LP for gDoF

If one is interested in the gDoF for the Gaussian noise case, it suffices to consider

$$
\mathsf{d} \leq 1 + \text{eq.}(33),
$$

which is the high-SNR approximation of (38) where the direct link from the source to the destination has SNR-exponent normalized to 1, i.e., $|h_{33}|^2 = \mathsf{SNR}^1$, without loss of generality. In this case the different quantities in (33) can be simply found by evaluating the different mutual information terms above and by using the definition in (4). We obtain

$$
\begin{aligned}
&a_2' = c_2' = a_2 := [\alpha_{s2}-1]^+, \\
&a_1' = d_1' = a_1 := [\alpha_{s1}-1]^+, \\
&b_1' = c_1' = b_1 := [\alpha_{1d}-1]^+, \\
&b_2' = d_2' = b_2 := [\alpha_{2d}-1]^+, \\
&D_1 := 0, \\
&D_2 := \max\{\alpha_{1d}+\alpha_{s2}-1,\beta_2\} - [\alpha_{1d}-1]^+ - [\alpha_{s2}-1]^+, \\
&D_3 := \max\{\alpha_{2d}+\alpha_{s1}-1,\beta_1\} - [\alpha_{2d}-1]^+ - [\alpha_{s1}-1]^+, \\
&D_4 := 0,
\end{aligned}
$$

where α_{si} is the SNR-exponent on the link from the source to relay i, $i \in [1:2]$, α_{id} is the SNR-exponent on the link from relay i, $i \in [1:2]$, to the destination and β_i is the SNR-exponent on the link from relay j to relay i, $(i,j) \in [1:2]^2$ with $j \neq i$.

REFERENCES

[1] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim, "Gaussian half-duplex relay networks: Improved gap and a connection with the assignment problem," in *IEEE Information Theory Workshop (ITW) 2013*, September 2013, pp. 1–5.

[2] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572 – 584, September 1979.

[3] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers, 2010*, 2010, pp. 1558–1562.

[4] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in *Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2011*, 2011, pp. 2002–2006.

[5] E. C. van der Meulen, "Three-terminal communication channel," *Advances in Applied Probability*, vol. 3, pp. 120–154, 1971.

[6] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 921 –930, March 2009.

[7] A. Host-Madsen, "On the capacity of wireless relaying," in *IEEE 56th Vehicular Technology Conference (VTC) 2002*, 2002, vol. 3, pp. 1333 – 1337 vol.3.

[8] G. Kramer, "Models and theory for relay channels with receive constraints," in *42nd Annual Allerton Conference on Communication, Control, and Computing*, September 2004, pp. 1312–1321.

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

[10] A.S. Avestimehr, *Wireless network information flow: a deterministic approach*, Ph.D. thesis, EECS Department, University of California, Berkeley, October 2008.

[11] 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.

[12] 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.

[13] S.H. 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.

[14] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *IEEE International Symposium on Information Theory (ISIT) 2000*, June 2000, p. 22.

[15] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3008–3023, July 2008.

[16] L. Ong, S.J. Johnson, and C.M. Kellett, "On the capacity of the binary-symmetric parallel-relay network," *arxiv:1207.3574*, 2012.

[17] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 765–779, March 2005.

[18] B. Chern and A. Özgür, "Achieving the capacity of the n-relay Gaussian diamond network within log(n) bits," in *IEEE Information Theory Workshop (ITW) 2012, (also arXiv:1207.5660)*, September 2012.

[19] A. Sengupta, I-Hsiang Wang, and C. Fragouli, "Optimizing quantize-map-and-forward relaying for Gaussian diamond networks," in *IEEE Information Theory Workshop (ITW) 2012*, 2012, pp. 381–385.

[20] U. Niesen and S. Diggavi, "Approximate capacity of the Gaussian n-relay diamond network," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 845–859, February 2012.

[21] H. Bagheri, A.S. Motahari, and A.K. Khandani, "On the capacity of the half-duplex diamond channel," in *IEEE International Symposium on Information Theory (ISIT), 2010*, June 2010, pp. 649 –653.

[22] F. Xue and S. Sandhu, "Cooperation in a half-duplex Gaussian diamond relay channel," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3806 –3814, October 2007.

[23] S. Brahma, A. Özgür, and C. Fragouli, "Simple schedules for half-duplex networks," in *IEEE International Symposium on Information Theory (ISIT), 2012*, July 2012, pp. 1112 –1116.

[24] S. Brahma, C. Fragouli, and A. Özgür, "On the structure of approximately optimal schedules for half-duplex diamond networks," in *51st Annual Allerton Conference on Communication, Control, and Computing*, September 2013.

[25] L. Ong, M. Motani, and S. J. Johnson, "On capacity and optimal scheduling for the half-duplex multiple-relay channel," *IEEE Transactions on Information Theory*, vol. 58, no. 9, pp. 5770 –5784, September 2012.

[26] R.E. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*, SIAM, 2009.

[27] H.W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistic Quart.*, vol. 2, pp. 83–97, 1955.

[28] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, Cambridge U.K., 2011.

[29] L. Zhang, J. Jiang, A.J. Goldsmith, and S. Cui, "Study of Gaussian relay channels with correlated noises," *IEEE Transactions on Communications*, vol. 59, no. 3, pp. 863 –876, March 2011.

[30] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim, "The capacity to within a constant gap of the Gaussian half-duplex relay channel," in *IEEE International Symposium on Information Theory (ISIT) 2013*, July 2013, pp. 749–753.

[31] P. Rost and G. Fettweis, "Achievable rates in Gaussian half-duplex multiple relay networks," in *IEEE Global Telecommunications Conference (GLOBECOM), 2008*, December 2008, pp. 1–6.

[32] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," *arxiv:1001.3404*, 2010.

[33] L. Lovasz and M.D. Plummer, *Matching Theory*, American Mathematical Society, 2009.

[34] V.M. Prabhakaran and P. Viswanath, "Interference channels with source cooperation," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 156 –186, January 2011.

[35] I.H. Wang and D.N.C. Tse, "Interference mitigation through limited transmitter cooperation," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2941–2965, May 2011.

[36] S. Kannan, A. Raja, and P. Viswanath, "Approximately optimal wireless broadcasting," *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7154–7167, December 2012.

[37] G. Broida and S.G. Williamson, *A comprehensive Introduction to Linear Algebra*, Addison-Wesley, 1989.

[38] J.M. Steele, *The Cauchy-Schwarz Master Class: an introduction to the art of mathematical inequalities*, Cambridge University Press, 2004.

Martina Cardone received the B.S. (Telecommunications Engineering) and M.S. (Telecommunications Engineering) degrees summa cum laude from the Politecnico di Torino, Italy, in 2009 and 2011, respectively and the M.S. degree in Telecommunications Engineering from Telecom ParisTech, Paris, France, in 2011, as part of a double degree program. Since October 2011, she is a Ph.D. student at EURECOM, France. Her current research focuses on multi-user information theory and its applications to wireless channels with cooperation among the nodes.

Daniela Tuninetti received her M.S. in Telecommunication Engineering from Politecnico di Torino (Italy) in 1998, and her Ph.D. in Electrical Engineering from ENST/Telecom ParisTech (with work done at EURECOM in Sophia Antipolis, France) in 2002. From 2002 to 2004 she was a postdoctoral research associate at the School of Communication and Computer Science at the EPFL/Swiss Federal Institute of Technology in Lausanne. Since January 2005, she is with the Department of Electrical and Computer Engineering at the University of Illinois at Chicago, Chicago, IL USA, where she currently is an Associate Professor. Dr. Tuninetti was the editor-in-chief of the IEEE Information Theory Society Newsletter from 2006 to 2008, and an associate editor for the IEEE Communication Letters from 2006 to 2009. She currently serves as an editor for the IEEE Transactions on Wireless Communications. She regularly serves on the Technical Program Committee of IEEE workshops and conferences, and she was the Communication Theory symposium co-chair of the 2010 IEEE International Conference on Communications (ICC 2010). Dr. Tuninetti received the best student paper award at the European Wireless Conference in 2002, and was the recipient of an NSF CAREER award in 2007. Her research interests are in the ultimate performance limits of wireless interference networks, with special emphasis on cognition and user cooperation.

Raymond Knopp is professor in the Mobile Communications Department at EURECOM. He received the B.Eng. (Honours) and the M.Eng. degrees in Electrical Engineering from McGill University, Montreal, Canada, in 1992 and 1993, respectively. From 1993-1997 he was a research assistant in the Mobile Communications Department at EURECOM working towards the PhD degree in Communication Systems from the Swiss Federal Institute of Technology (EPFL), Lausanne. From 1997-2000 he was a research associate in the Mobile Communications Laboratory (LCM) of the Communication Systems Department of EPFL. His current research and teaching interests are in the area of digital communications, software radio architectures, and implementation aspects of signal processing systems and real-time wireless networking protocols. He has a proven track record in managing both fundamental and experimental research projects at an international level and is also technical coordinator of the OpenAirInterface.org open-source wireless radio platform initiative which aims to bridge the gap between cutting-edge theoretical advances in wireless communications and practical designs.

Umer Salim received the Ph.D. and M.S. degrees, both in electrical engineering with specialization in communication theory and signal processing from EURECOM, France, and Supelec, France, respectively. He has several years of research experience in digital communications and signal processing and has published several papers in well-known conferences and journals. His mains areas of interest include signal processing techniques for multi-cell multi-user MIMO systems, novel and practical CSI feedback design techniques and analysis, information theoretic analysis of cognitive radio, and multi-user information theory in general. He has been serving as the reviewer for IEEE Trans. on Information Theory, IEEE Trans. on Wireless Communications, IEEE Trans. on Signal Processing and numerous well-known conferences.
Dr. Salim co-authored a paper which received the best paper award at the European Wireless Conference 2011.
Dr. Salim is currently working at Intel in the department of Systems engineering where the main focus is on the design of advanced receivers for future wireless standards. At Intel, he has designed sophisticated interference cancellation algorithms which are in use in modern high-end smart-phones and tablets.