Asymptotic Capacity and Optimal Precoding in MIMO Multi-Hop Relay Networks

Nadia Fawaz, *Member, IEEE*, Keyvan Zarifi, *Member, IEEE*, Mérouane Debbah, *Senior Member, IEEE*, and David Gesbert, *Fellow, IEEE*

***Abstract*—A multihop relaying system is analyzed where data sent by a multi-antenna source is relayed by successive multi-antenna relays until it reaches a multi-antenna destination. Assuming correlated fading at each hop, each relay receives a faded version of the signal from the previous level, performs linear precoding and retransmits it to the next level. Using free probability theory and assuming that the noise power at relays— but not at destination—is negligible, the closed-form expression of the asymptotic instantaneous end-to-end mutual information is derived as the number of antennas at all levels grows large. The so-obtained deterministic expression is independent from the channel realizations while depending only on channel statistics. This expression is also shown to be equal to the asymptotic average end-to-end mutual information. The singular vectors of the optimal precoding matrices, maximizing the average mutual information with finite number of antennas at all levels, are also obtained. It turns out that these vectors are aligned to the eigenvectors of the channel correlation matrices. Thus, they can be determined using only the channel statistics. As the structure of the singular vectors of the optimal precoders is independent from the system size, it is also optimal in the asymptotic regime.**

***Index Terms*—Asymptotic capacity, correlated channel, free probability theory, multihop relay network, precoding.**

I. Introduction

RELAY communication systems have recently attracted much attention due to their potential to substantially improve the signal reception quality when the direct communication link between the source and the destination is not reliable. Due to its major practical importance as well as its significant technical challenge, deriving the capacity—or bounds on the capacity—of various relay communication schemes is growing to an entire field of research. Of interest is the derivation of capacity bounds for systems in which the source, the destination, and the relays are equipped with multiple antennas.

Several works have focused on the capacity of two-hop relay networks, such as [1]–[7]. Assuming fixed channel conditions, lower and upper bounds on the capacity of the two-hop multiple-input multiple output (MIMO) relay channel were derived in [1]. In the same paper, bounds on the ergodic capacity were also obtained when the communication links undergo i.i.d. Rayleigh fading. The capacity of a MIMO two-hop relay system was studied in [2] in the asymptotic case where the number of relay nodes grows large while the number of transmit and receive antennas remain constant. The scaling behavior of the capacity in two-hop amplify-and-forward (AF) networks was analyzed in [3]–[5] when the numbers of single-antenna sources, relays and destinations grow large. The achievable rates of a two-hop code-division multiple-access (CDMA) decode-and-forward (DF) relay system were derived in [8] when the numbers of transmit antennas and relays grow large. In [6], an ad hoc network with several source-destination pairs communicating through multiple AF-relays was studied and an upperbound on the asymptotic capacity in the low Signal-to-Noise Ratio (SNR) regime was obtained in the case where the numbers of source, relay and destination nodes grow large. The scaling behavior of the capacity of a two-hop MIMO relay channel was also studied in [7] for bi-directional transmissions. In [9] the optimal relay precoding matrix was derived for a two-hop relay system with perfect knowledge of the source-relay and relay-destination channel matrices at the relay.

Following the work in [10] on the asymptotic eigenvalue distribution of concatenated fading channels, several analyses were proposed for more general multihop relay networks, including [11]–[15]. In particular, considering multihop MIMO AF networks, the tradeoffs between rate, diversity, and network size were analyzed in [11], and the diversity-multiplexing tradeoff was derived in [12]. The asymptotic capacity of multihop MIMO AF relay systems was obtained in [13] when all channel links experience i.i.d. Rayleigh fading while the number of transmit and receive antennas, as well as the number of relays at each hop grow large with the same rate. Finally hierarchical multihop MIMO networks were studied in [15] and the scaling laws of capacity were derived when the network density increases.

Manuscript received December 27, 2008; revised July 23, 2010; accepted August 26, 2010. Date of current version March 16, 2011. This work was supported in part by the French Defense Body DGA, by BIONETS project (FP6-027748, http://www.bionets.eu) and in part by Alcatel-Lucent within the Alcatel-Lucent Chair on flexible radio at SUPELEC. The material in this paper was presented in part at the IEEE Information Theory Workshop (ITW 2008), Porto, Portugal, May 2008.

N. Fawaz is with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: nfawaz@mit.edu).

K. Zarifi is with the Institut National de la Recherche Scientifique-Énergie, Matériaux et Télécommunications (INRS-EMT), Université du Québec, Montreal, QC H5A 1K6, Canada, and also with Concordia University, Montréal, QC H3G 1M8, Canada (e-mail: keyvan.zarifi@emt.inrs.ca; zarifi@ieee.org).

M. Debbah is with the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, 91192, Gif-sur-Yvette, Cedex, France (e-mail: merouane.debbah@supelec.fr).

D. Gesbert is with the Mobile Communications Department, EURECOM, 06904 Sophia-Antipolis, France (e-mail: david.gesbert@eurecom.fr).

Communicated by A. Goldsmith, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2111830

In this paper, we study an N-hop MIMO relay communication system wherein data transmission from k_0 source antennas to k_N destination antennas is made possible through $N-1$ relay levels, each of which are equipped with $k_i,\ i=1,\ldots,N-1$ antennas. In this transmission chain with $N+1$ levels it is assumed that the direct communication link is only viable between two adjacent levels, due to large distances between nonadjacent levels for instance: each relay receives a faded version of the multidimensional signal transmitted from the previous level and, after linear precoding, retransmits it to the next level.

We consider the case where all communication links undergo Rayleigh flat fading and the fading channels at each hop (between two adjacent levels) may be correlated while the fading channels of any two different hops are independent. We assume that the channel at each hop is block-fading and that the channel coherence-time is long enough—with respect to codeword length—for the system to be in the nonergodic regime. As a consequence, the channel is a realization of a random matrix that is fixed during a coherence block. Therefore, the instantaneous end-to-end mutual information between the source and the destination can be viewed as an instance of a random variable.

Using tools from the free probability theory and assuming that the noise power at the relay levels, but not at the destination, is negligible, we derive a closed-form expression of the asymptotic instantaneous end-to-end mutual information between the source input and the destination output as the number of antennas at all levels grows large. This asymptotic expression is shown to be independent from the channel realizations and to only depend on the channel statistics. Therefore, as long as the statistical properties of the channel matrices at all hops do not change, the instantaneous mutual information asymptotically converges to the same deterministic expression for any arbitrary channel realization. This property has two major consequences. First, the mutual information in the asymptotic regime is not a random variable any more but a deterministic value representing an achievable rate. This means that when the channel is random but fixed during the transmission and the system size is large enough, the capacity in the sense of Shannon is not zero, on the contrary to the capacity of small size systems [16, Section 5.1]. Second, given the stationarity of channel statistical properties, the asymptotic instantaneous mutual information obtained in the nonergodic regime also serves as the asymptotic value of the average end-to-end mutual information between the source and the destination. Note that the latter is the same as the asymptotic ergodic end-to-end mutual information that would be obtained if the channel was an ergodic process.

We also obtain the singular vectors of the optimal precoding matrices that maximize the average mutual information of the system with a finite number of antennas at all levels. It is proven that the singular vectors of the optimal precoding matrices are also independent from the channel realizations and can be determined using only statistical knowledge of channel matrices at source and relays. We show that the so-obtained singular vectors are also optimal in the asymptotic regime of our concern. Finally, we apply the aforementioned results on the asymptotic mutual information and the structure of the optimal precoding matrices to several communications scenarios with different number of hops, and types of channel correlation.

The rest of the paper is organized as follows. Notations and the system model are presented in Section II. The end-to-end instantaneous mutual information in the asymptotic regime is derived in Section III, while the singular vectors of the optimal precoding matrices are obtained in Section IV. Theorems derived in Sections III and IV are applied to several MIMO communication scenarios in Section V. Numerical results are provided in Section VI and concluding remarks are drawn in Section VII.

II. System Model

Notation: $\mathbb{N}$ is the set of non-negative integers. Let $m < n \in \mathbb{N}$, the set of integers greater or equal to m and less or equal to n is denoted $\mathbb{N}_m^n \triangleq \{m, m+1, \ldots, n-1, n\}$. $\log$ denotes the logarithm in base 2 while $\ln$ is the logarithm in base e. $u(x)$ is the unit-step function defined by $u(x) = 0$ if $x < 0$; $u(x) = 1$ if $x \geq 0$. $K(m) \triangleq \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1-m\sin^2\theta}}$ is the complete elliptic integral of the first kind [17]. Matrices and vectors are represented by boldface upper and lower cases, respectively. $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$ stand for the transpose, the conjugate and the transpose conjugate of $\mathbf{A}$, respectively. The trace and the determinant of $\mathbf{A}$ are respectively denoted by $\mathrm{tr}(\mathbf{A})$ and $\det(\mathbf{A})$. $\lambda_{\mathbf{A}}(1), \ldots, \lambda_{\mathbf{A}}(n)$ represent the eigenvalues of an $n \times n$ matrix $\mathbf{A}$. The operator norm of $\mathbf{A}$ is defined by $\|\mathbf{A}\| \triangleq \sqrt{\max_i \lambda_{\mathbf{A}^H\mathbf{A}}(i)}$, while the Fröbenius norm of $\mathbf{A}$ is $\|\mathbf{A}\|_F \triangleq \sqrt{\mathrm{tr}(\mathbf{A}^H\mathbf{A})}$. The (i,j)th entry of matrix $\mathbf{A}_k$ is written $a_{ij}^{(k)}$. $\mathbf{I}_N$ is the identity matrix of size N. $\mathrm{E}[\cdot]$ is the statistical expectation operator, $\mathcal{H}(X)$ the entropy of a variable X, and $\mathcal{I}(X;Y)$ the mutual information between variables X and Y. $F_{\mathbf{\Omega}}^n(\cdot)$ is the empirical eigenvalue distribution of an $n \times n$ square matrix $\mathbf{\Omega}$ with real eigenvalues, while $F_{\mathbf{\Omega}}(\cdot)$ and $f_{\mathbf{\Omega}}(\cdot)$ are respectively its asymptotic eigenvalue distribution and its eigenvalue probability density function when its size n grows large. We denote the matrix product by $\bigotimes_{i=1}^N \mathbf{A}_i = \mathbf{A}_1\mathbf{A}_2\ldots\mathbf{A}_N$. Note that the matrix product is not commutative; therefore, the order of the index i in the product is important and in particular $(\bigotimes_{i=1}^N \mathbf{A}_i)^H = \bigotimes_{i=N}^1 \mathbf{A}_i^H$.

A. Multihop MIMO Relay Network

Consider Fig. 1 that shows a multihop relaying system with k_0 source antennas, k_N destination antennas and $N-1$ relaying levels. The $i-$th relaying level is equipped with k_i antennas. We assume that the noise power is negligible at all relays while at the destination the noise power is such that

$$\mathrm{E}[\mathbf{z}\mathbf{z}^H] = \sigma^2\mathbf{I} = \frac{1}{\eta}\mathbf{I} \tag{1}$$

where $\mathbf{z}$ is the circularly-symmetric zero-mean i.i.d. Gaussian noise vector at the destination. In effect, the simplifying noise-free relays assumption is made to have a white aggregate noise at the destination and, consequently, more tractable derivations. Note that several other authors have implicitly used a similar noise-free relay assumption by assuming that the noise at the destination of a MIMO multihop relay network is white. For instance, in [12] a multihop AF relay network is analyzed and it is proved that the resulting colored noise at the destination can

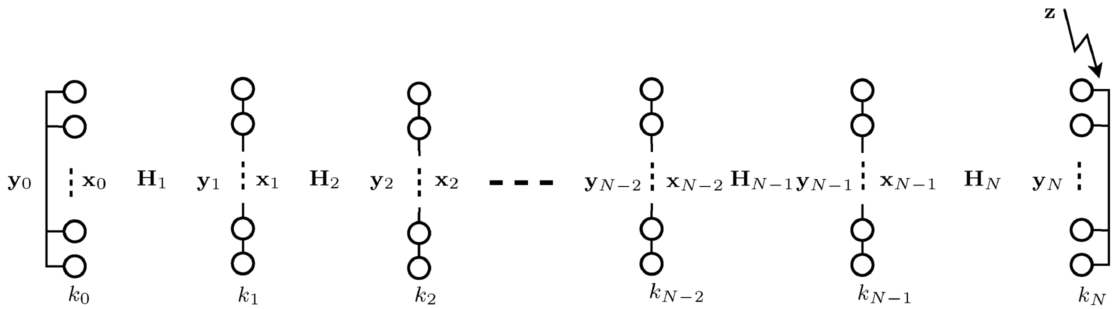


Fig. 1. Multilevel relaying system.

be well-approximated by white noise in the high SNR regime. In terms of practical relevance, the mutual information expression derived in the case of noise-free relays can be seen as an upper-bound for the case of noisy relays. When applied to a particular communication scenario, if the expressions obtained for perfect noise-free relays show that no gains in terms of rate result from relaying, then a more complex analysis with noisy relays will be irrelevant.

Throughout the paper, we assume that the correlated channel matrix at hop $i \in \{1, \ldots, N\}$ can be represented by the Kronecker model

$$\mathbf{H}_i \triangleq \mathbf{C}_{r,i}^{1/2} \boldsymbol{\Theta}_i \mathbf{C}_{t,i}^{1/2} \tag{2}$$

where $\mathbf{C}_{t,i}, \mathbf{C}_{r,i}$ are respectively the transmit and receive correlation matrices, $\boldsymbol{\Theta}_i$ are zero-mean i.i.d. Gaussian matrices independent over index i, with variance of the (k,l)th entry

$$\mathrm{E}[|\theta_{kl}^{(i)}|^2] = \frac{a_i}{k_{i-1}} \qquad \forall i \in \mathbb{N}_1^N \tag{3}$$

where $a_i = d_i^{-\beta}$ represents the pathloss attenuation with β and d_i denoting the pathloss exponent and the length of the ith hop, respectively. We also assume that channels matrices $\mathbf{H}_i, \ i = 1, \ldots, N$ remain constant during a coherence block of length L and vary independently from one channel coherence block to the next.

Note that no assumption is made on the structure of the channel correlation matrices. The particular case of i.i.d. Rayleigh fading channel can be obtained from the above Kronecker model when matrices $\mathbf{C}_{t,i}$ and $\mathbf{C}_{r,i}$ are set to identity. It should also be mentioned that by adapting the correlation matrices structure, the Kronecker model can be used to model relay-clustering. Given a total number of antennas k_i at level i, instead of considering that the relaying level consists of a single relay equipped with many antennas (k_i), we can consider that a relaying level contains n_i relays equipped with (k_i/n_i) antennas. Clustering has a direct impact on the structure of correlation matrices: when the k_i antennas at level i are distributed among several relays, correlation matrices become block-diagonal matrices, whose blocks represent the correlation between antennas at a relay, while antennas at different relays sufficiently separated in space are supposed uncorrelated. In the limit of a relaying level containing k_i relays equipped with a single antenna, we fall back to the case of uncorrelated fading with correlation matrices equal to identity.

Within one channel coherence block, the signal transmitted by the k_0 source antennas at time $l \in \{0, \ldots, L-1\}$ is given by the vector $\mathbf{x}_0(l) = \mathbf{P}_0 \mathbf{y}_0(l-1)$, where $\mathbf{P}_0$ is the source precoding matrix and $\mathbf{y}_0$ is a zero-mean random vector with

$$\mathrm{E}\{\mathbf{y}_0 \mathbf{y}_0^H\} = \mathbf{I}_{k_0} \tag{4}$$

which implies that

$$\mathrm{E}\{\mathbf{x}_0 \mathbf{x}_0^H\} = \mathbf{P}_0 \mathbf{P}_0^H. \tag{5}$$

Assuming that relays work in full-duplex mode, at time $l \in \{0, \ldots, L-1\}$ the relay at level i uses a precoding matrix $\mathbf{P}_i$ to linearly precode its received signal $\mathbf{y}_i(l-1) = \mathbf{H}_i \mathbf{x}_{i-1}(l-1)$ and form its transmitted signal

$$\mathbf{x}_i(l) = \mathbf{P}_i \mathbf{y}_i(l-1) \qquad \forall i \in \mathbb{N}_0^{N-1}. \tag{6}$$

The precoding matrices at source and relays $\mathbf{P}_i, \ i \in \mathbb{N}_0^{N-1}$ are subject to the per-node long-term average power constraints

$$\mathrm{tr}(\mathrm{E}[\mathbf{x}_i \mathbf{x}_i^H]) \le k_i \mathcal{P}_i \qquad \forall i \in \mathbb{N}_0^{N-1}. \tag{7}$$

The fact that $\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_{i-1}$, along with the variance $\mathrm{E}[|\theta_{kl}^{(i)}|^2] = \frac{a_i}{k_{i-1}}$ of $\mathbf{H}_i$ elements and with the power constraint $\mathrm{tr}(\mathrm{E}[\mathbf{x}_{i-1} \mathbf{x}_{i-1}^H]) \le k_{i-1} \mathcal{P}_{i-1}$ on $\mathbf{x}_{i-1}$, render the system of our concern equivalent to a system whose random channel elements $\theta_{kl}^{(i)}$ would be i.i.d. with variance a_i and whose power constraint on transmitted signal $\mathbf{x}_{i-1}$ would be finite and equal to $\mathcal{P}_{i-1}$. Having finite transmit power at each level, this equivalent system shows that adding antennas, i.e., increasing the system dimension, does not imply increasing the transmit power. Nonetheless, in order to use random matrix theory tools to derive the asymptotic instantaneous mutual information in Section III, the variance of random channel elements is required to be normalized by the size of the channel matrix. That is why the normalized model—channel variance (3) and power constraint (7)—was adopted.

It should also be noticed that choosing diagonal precoding matrices would reduce the above scheme to the simpler AF relaying strategy. Note that the proposed linear precoding relaying technique is adapted for high SNR regimes, but not for low SNR regimes. In the low SNR regime, known to be noise-limited, linear precoding performs poorly because power is wasted on forwarding noise, and other relaying strategies such as decode-and-forward are more appropriate [18], [19]. On the contrary in the high SNR regime, linear precoding techniques such as amplify-and-forward perform well [11], [20]. Finally, from a practical point of view, limited channel knowledge and simple linear precoding techniques at relays are particularly relevant for systems where relays have limited processing capabilities.

As can be observed from Fig. 1, the signal received at the destination at time l is given by

$$\begin{aligned}\mathbf{y}_N(l) &= \mathbf{H}_N\mathbf{P}_{N-1}\mathbf{H}_{N-1}\mathbf{P}_{N-2}\dots\mathbf{H}_2\mathbf{P}_1\mathbf{H}_1\mathbf{P}_0\mathbf{y}_0(l-N)+\mathbf{z}\\ &= \mathbf{G}_N\mathbf{y}_0(l-N)+\mathbf{z}\end{aligned} \tag{8}$$

where the end-to-end equivalent channel is

$$\begin{aligned}\mathbf{G}_N &\triangleq \mathbf{H}_N\mathbf{P}_{N-1}\mathbf{H}_{N-1}\mathbf{P}_{N-2}\dots\mathbf{H}_2\mathbf{P}_1\mathbf{H}_1\mathbf{P}_0\\ &= \mathbf{C}_{r,N}^{1/2}\mathbf{\Theta}_N\mathbf{C}_{t,N}^{1/2}\mathbf{P}_{N-1}\mathbf{C}_{r,N-1}^{1/2}\mathbf{\Theta}_{N-1}\mathbf{C}_{t,N-1}^{1/2}\mathbf{P}_{N-2}\dots\\ &\quad\dots\mathbf{C}_{r,2}^{1/2}\mathbf{\Theta}_2\mathbf{C}_{t,2}^{1/2}\mathbf{P}_1\mathbf{C}_{r,1}^{1/2}\mathbf{\Theta}_1\mathbf{C}_{t,1}^{1/2}\mathbf{P}_0.\end{aligned} \tag{9}$$

Let us introduce the matrices

$$\begin{aligned}\mathbf{M}_0 &= \mathbf{C}_{t,1}^{1/2}\mathbf{P}_0\\ \mathbf{M}_i &= \mathbf{C}_{t,i+1}^{1/2}\mathbf{P}_i\mathbf{C}_{r,i}^{1/2} \qquad \forall i\in\mathbb{N}_1^{N-1}\\ \mathbf{M}_N &= \mathbf{C}_{r,N}^{1/2}.\end{aligned} \tag{10}$$

Then (9) can be rewritten as

$$\mathbf{G}_N = \mathbf{M}_N\mathbf{\Theta}_N\mathbf{M}_{N-1}\mathbf{\Theta}_{N-1}\dots\mathbf{M}_2\mathbf{\Theta}_2\mathbf{M}_1\mathbf{\Theta}_1\mathbf{M}_0. \tag{11}$$

For the sake of clarity, the dimensions of the matrices/vectors involved in our analysis are given below

$$\begin{array}{lll}\mathbf{x}_i : k_i\times 1 & \mathbf{y}_i : k_i\times 1 & \mathbf{P}_i : k_i\times k_i\\ \mathbf{H}_i : k_i\times k_i-1 & \mathbf{C}_{r,i} : k_i\times k_i & \mathbf{C}_{t,i} : k_i-1\times k_i-1\\ \mathbf{\Theta}_i : k_i\times k_i-1 & \mathbf{M}_i : k_i\times k_i. & \end{array}$$

In the sequel, we assume that the channel coherence time is large enough to consider the nonergodic case and consequently, time index l can be dropped. Finally, we define three channel-knowledge assumptions:

- Assumption $\mathbf{A_s}$, local statistical knowledge at source: the source has only statistical channel state information (CSI) of its forward channel $\mathbf{H}_1$, i.e., the source knows the transmit correlation matrix $\mathbf{C}_{t,1}$.
- Assumption $\mathbf{A_r}$, local statistical knowledge at relay: at the i^{th} relaying level, $i\in\mathbb{N}_1^{N-1}$, only statistical CSI of the backward channel $\mathbf{H}_i$ and forward channel $\mathbf{H}_{i+1}$ are available, i.e., relay i knows the receive correlation matrix $\mathbf{C}_{r,i}$ and the transmit correlation matrix $\mathbf{C}_{t,i+1}$.
- Assumption $\mathbf{A_d}$, end-to-end perfect knowledge at destination: the destination perfectly knows the end-to-end equivalent channel $\mathbf{G}_N$.

Throughout the paper, assumption $\mathbf{A_d}$ is always made. Assumption $\mathbf{A_d}$ is the single assumption on channel-knowledge necessary to derive the asymptotic mutual information in Section III, while the two extra assumptions $\mathbf{A_s}$ and $\mathbf{A_r}$ are also necessary in Section IV to obtain the singular vectors of the optimal precoding matrices.

B. Mutual Information

Consider the channel realization $\mathbf{G}_N$ in one channel coherence block. Under Assumption $\mathbf{A_d}$, the instantaneous end-to-end mutual information between channel input $\mathbf{y}_0$ and channel output $(\mathbf{y}_N, \mathbf{G}_N)$ in this channel coherence block is [16]

$$\begin{aligned}&\mathcal{I}(y_0; y_N|G_N=\mathbf{G}_N)\\ &\quad= \mathcal{H}(y_N|G_N=\mathbf{G}_N) - \underbrace{\mathcal{H}(y_N|y_0, G_N=\mathbf{G}_N)}_{\mathcal{H}(z)}\\ &\quad= \mathcal{H}(y_N|G_N=\mathbf{G}_N) - \mathcal{H}(z).\end{aligned} \tag{12}$$

The entropy of the noise vector is known to be $\mathcal{H}(z) = \log\det(\frac{\pi e}{\eta}\mathbf{I}_{k_N})$. Besides, $\mathbf{y}_0$ is zero-mean with variance $\mathrm{E}[\mathbf{y}_0\mathbf{y}_0^H] = \mathbf{I}_{k_0}$, thus given $\mathbf{G}_N$, the received signal $\mathbf{y}_N$ is zero-mean with variance $\mathbf{G}_N\mathbf{G}_N^H + \frac{1}{\eta}\mathbf{I}_{k_N}$. By [16, Lemma 2], we have the inequality $\mathcal{H}(y_N|G_N=\mathbf{G}_N) \le \log\det(\pi e\mathbf{G}_N\mathbf{G}_N^H + \frac{\pi e}{\eta}\mathbf{I}_{k_N})$, and the entropy is maximized when the latter inequality holds with equality. This occurs if $\mathbf{y}_N$ is circularly-symmetric complex Gaussian, which is the case when $\mathbf{y}_0$ is circularly-symmetric complex Gaussian. Therefore, throughout the rest of the paper we consider $\mathbf{y}_0$ to be a zero-mean circularly-symmetric complex Gaussian vector. As such, the instantaneous mutual information (12) can be rewritten as

$$\mathcal{I}(y_0; y_N|G_N=\mathbf{G}_N) = \log\det(\mathbf{I}_{k_N} + \eta\mathbf{G}_N\mathbf{G}_N^H). \tag{13}$$

Under Assumption $\mathbf{A_d}$, the average end-to-end mutual information between channel input $\mathbf{y}_0$ and channel output $(\mathbf{y}_N, \mathbf{G}_N)$ is

$$\begin{aligned}\mathcal{I}(y_0;(y_N,G_N)) &= \mathcal{I}(y_0;y_N|G_N) + \underbrace{\mathcal{I}(y_0;G_N)}_{0}\\ &= \mathcal{I}(y_0;y_N|G_N)\\ &= \mathrm{E}_{G_N}[\mathcal{I}(y_0;y_N|G_N=\mathbf{G}_N)]\\ &= \mathrm{E}_{G_N}[\log\det(\mathbf{I}_{k_N}+\eta\mathbf{G}_N\mathbf{G}_N^H)].\end{aligned} \tag{14}$$

To optimize the system, we are left with finding the precoders $\mathbf{P}_i$ that maximize the end-to-end mutual information (14) subject to power constraints (7). In other words, we need to find the maximum average end-to-end mutual information

$$\boldsymbol{C} \triangleq \max_{\{\mathbf{P}_i/\mathrm{tr}(\mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H])\le k_i\mathcal{P}_i\}_{i\in\mathbb{N}_0^{N-1}}} \mathrm{E}_{G_N}[\log\det(\mathbf{I}_{k_N}+\eta\,\mathbf{G}_N\mathbf{G}_N^H)]. \tag{15}$$

In Section IV, the problem of finding the singular vectors of the optimal precoders that maximize the average mutual information (15) is addressed under channel knowledge Assumptions $\mathbf{A_s}$, $\mathbf{A_r}$, and $\mathbf{A_d}$. Note that the nonergodic regime is considered; therefore, (14) represents only an average mutual information over channel realizations, and the solution to (15) does not necessarily represent the channel capacity in the Shannon sense—the supremum of achievable rates with arbitrary small probability of error—when the system size is small.

III. Asymptotic Mutual Information

In this section, we consider the instantaneous mutual information per source antenna between the source and the destination

$$\boldsymbol{I} \triangleq \frac{1}{k_0}\log\det(\mathbf{I}_{k_N}+\eta\mathbf{G}_N\mathbf{G}_N^H) \tag{16}$$

and derive its asymptotic value as the number of antennas $k_0, k_1, \ldots, k_N$ grow large. The following theorem holds.

Theorem 1: For the system described in Section II, assume that:

- channel knowledge assumption $\mathbf{A_d}$ holds;
- $k_0, k_1, \ldots, k_N \to \infty$ while $\frac{k_i}{k_N} \to \rho_i$ for all $i \in \mathbb{N}_0^N$;
- for all $i \in \mathbb{N}_0^N$, as $k_i \to \infty$, $\mathbf{M}_i^H\mathbf{M}_i$ has a limit eigenvalue distribution with a compact support.

Then the instantaneous mutual information per source antenna $\boldsymbol{I}$ converges almost surely to

$$\boldsymbol{I}_\infty = \frac{1}{\rho_0}\sum_{i=0}^{N}\rho_i \mathrm{E}\left[\log\left(1+\eta\frac{a_{i+1}}{\rho_i}h_i^N\Lambda_i\right)\right] - N\frac{\log e}{\rho_0}\eta\prod_{i=0}^{N}h_i \tag{17}$$

where $a_{N+1}=1$ by convention, $h_0, h_1, \ldots, h_N$ are the solutions of the system of $N+1$ equations

$$\prod_{j=0}^{N}h_j = \rho_i\mathrm{E}\left[\frac{h_i^N\Lambda_i}{\frac{\rho_i}{a_{i+1}}+\eta h_i^N\Lambda_i}\right] \qquad \forall i\in\mathbb{N}_0^N \tag{18}$$

and the expectation $\mathrm{E}[\cdot]$ in (17) and (18) is over Λ_i whose distribution is given by the asymptotic eigenvalue distribution $F_{\mathbf{M}_i^H\mathbf{M}_i}(\lambda)$ of $\mathbf{M}_i^H\mathbf{M}_i$.

The detailed proof of Theorem 1 is presented in Appendix B.

We would like to stress that (17) holds for any arbitrary set of precoding matrices $\mathbf{P}_i,\ i=0,\ldots,N-1$, if $\mathbf{M}_i^H\mathbf{M}_i$ has a compactly supported asymptotic eigenvalue distribution when the system dimensions grow large. We would like to point out that the power constraints on signals transmitted by the source or relays are not sufficient to guarantee the boundedness of the eigenvalues of $\mathbf{M}_i^H\mathbf{M}_i$. In fact, it is proved in Appendix C that these power constraints can be written as

$$\frac{1}{k_0}\mathrm{tr}(\mathbf{P}_0\mathbf{P}_0^H) \le \mathcal{P}_0,$$
$$\frac{a_i}{k_i}\mathrm{tr}(\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H)\prod_{k=0}^{i-1}\frac{a_k}{k_k}\mathrm{tr}(\mathbf{C}_{t,k+1}\mathbf{P}_k\mathbf{C}_{r,k}\mathbf{P}_k^H) \le \mathcal{P}_i, \quad \forall i\in\mathbb{N}_1^{N-1}. \tag{19}$$

In the asymptotic regime, $\lim_{k_i\to\infty}\frac{1}{k_i}\mathrm{tr}(\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H) = \mathrm{E}[\lambda_{\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H}]$ and $\lim_{k_k\to\infty}\frac{1}{k_k}\mathrm{tr}(\mathbf{C}_{t,k+1}\mathbf{P}_k\mathbf{C}_{r,k}\mathbf{P}_k^H) = \mathrm{E}[\Lambda_k]$. Therefore, the power constraints impose upper-bounds (19) on the product of the first-order moments of the eigenvalues of matrices $\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H$ and $\mathbf{M}_k^H\mathbf{M}_k$ in the asymptotic regime. Unfortunately, these upper-bounds do not prevent the eigenvalue distribution of $\mathbf{M}_i^H\mathbf{M}_i$ from having an unbounded support, and thus, the power constraints are a priori not sufficient to guarantee the compactness of the support of the asymptotic eigenvalue distribution of matrices $\mathbf{M}_i^H\mathbf{M}_i$. The assumption that matrices $\mathbf{M}_i^H\mathbf{M}_i$ have a compactly supported asymptotic eigenvalue distribution is a priori not an intrinsic property of the system model, but it was necessary to make that assumption in order to use Lemma 2 to prove Theorem 1.

Given a set of precoding matrices, it can be observed from (17) and (18) that the asymptotic expression is a deterministic value that depends only on channel statistics and not on a particular channel realization. In other words, for a given set of precoding matrices, as long as the statistical properties of the channel matrices do not change, the instantaneous mutual information always converges to the same deterministic achievable rate, regardless of the channel realization. From this observation, three results follow:

- **Result 1:** As the numbers of antennas at all levels grow large, the instantaneous mutual information is not a random variable anymore and the precoding matrices maximizing the asymptotic instantaneous mutual information can be found based only on knowledge of the channel statistics, without requiring any information regarding the instantaneous channel realizations.
- **Result 2:** When the channel is random but fixed during the transmission and the system size grows large enough, the Shannon capacity is not zero any more, on the contrary to the capacity of small-size nonergodic systems [16, Section 5.1].
- **Result 3:** The asymptotic instantaneous mutual information (17) obtained in the nonergodic regime also represents the asymptotic value of the average mutual information, whose expression is the same as the asymptotic ergodic end-to-end mutual information that would be obtained if the channel was an ergodic process.

It should also be mentioned that, according to the experimental results illustrated in Section VI, the system under consideration behaves like in the asymptotic regime even when it is equipped with a reasonable finite number of antennas at each level. Therefore, (17) can also be efficiently used to evaluate the instantaneous mutual information of a finite-size system.

IV. Optimal Transmission Strategy at Source and Relays

In previous section, the asymptotic instantaneous mutual information (17), (18) was derived considering arbitrary precoding matrices $\mathbf{P}_i, i\in\{0,\ldots,N-1\}$. In this section, we analyze the optimal linear precoding strategies $\mathbf{P}_i, i\in\{0,\ldots,N-1\}$ at source and relays that allow to maximize the average mutual information. We characterize the optimal transmit directions determined by the singular vectors of the precoding matrices at source and relays, for a system with finite $k_0, k_1, \ldots, k_N$. It turns out that those transmit direction are also the ones that maximize the asymptotic average mutual information. Moreover, from **Result 3** in Section III, it can be inferred that the singular vectors of the precoding matrices maximizing the asymptotic average mutual information are also optimal for the asymptotic instantaneous mutual information (17).

In future work, using the results on the optimal directions of transmission (singular vectors of $\mathbf{P}_i$) and the asymptotic mutual information (17)–(18), we intend to derive the optimal power allocation (singular values of $\mathbf{P}_i$) that maximize the asymptotic instantaneous/average mutual information (17) using only statistical knowledge of the channel at transmitting nodes.

The main result of this section is given by the following theorem:

Theorem 2: Consider the system described in Section II. For $i \in \{1,\dots,N\}$ let $\mathbf{C}_{t,i} = \mathbf{U}_{t,i}\mathbf{\Lambda}_{t,i}\mathbf{U}_{t,i}^H$ and $\mathbf{C}_{r,i} = \mathbf{U}_{r,i}\mathbf{\Lambda}_{r,i}\mathbf{U}_{r,i}^H$ be the eigenvalue decompositions of the correlation matrices $\mathbf{C}_{t,i}$ and $\mathbf{C}_{r,i}$, where $\mathbf{U}_{t,i}$ and $\mathbf{U}_{r,i}$ are unitary and $\mathbf{\Lambda}_{t,i}$ and $\mathbf{\Lambda}_{r,i}$ are diagonal, with their respective eigenvalues ordered in decreasing order. Then, under channel-knowledge assumptions $\mathbf{A_s}$, $\mathbf{A_r}$ and $\mathbf{A_d}$, the optimal linear precoding matrices that maximize the average mutual information under power constraints (7) can be written as

$$\begin{aligned}\mathbf{P}_0 &= \mathbf{U}_{t,1}\mathbf{\Lambda}_{P_0}\\ \mathbf{P}_i &= \mathbf{U}_{t,i+1}\mathbf{\Lambda}_{P_i}\mathbf{U}_{r,i}^H \qquad \forall i \in \mathbb{N}_1^{N-1}\end{aligned} \tag{20}$$

where $\mathbf{\Lambda}_{P_i}$ are diagonal matrices with non-negative real diagonal elements. Moreover, the singular vectors of the precoding matrices (20) are also the ones that maximize the asymptotic average mutual information. Since the asymptotic average mutual information has the same value as the asymptotic instantaneous mutual information, the singular vectors of the precoding matrices (20) are also optimal for the asymptotic instantaneous mutual information.

For the proof of Theorem 2, the reader is referred to Appendix C.

Theorem 2 indicates that to maximize the average mutual information:

- the source should align the eigenvectors of the transmit covariance matrix $\mathbf{Q} = \mathbf{P}_0\mathbf{P}_0^H$ to the eigenvectors of the transmit correlation matrix $\mathbf{C}_{t,1}$ of the first-hop channel $\mathbf{H}_1$. This alignment requires only local statistical channel knowledge $\mathbf{A_s}$. Note that similar results were previously obtained for both single-user [21] and multi-user [22] single-hop (without relays) MIMO system with covariance knowledge at the source.
- relay i should align the right singular vectors of its precoding matrix $\mathbf{P}_i$ to the eigenvectors of the receive correlation matrix $\mathbf{C}_{r,i}$, and the left singular vectors of $\mathbf{P}_i$ to the eigenvectors of the transmit correlation matrix $\mathbf{C}_{t,i+1}$. These alignments require only local statistical knowledge $\mathbf{A_r}$.

As power is non-negative, aligning the singular vectors of the precoders to the eigenvectors of channel correlation matrices allows to avoid wasting power on non-eigen directions. Moreover, it follows from Theorem 2 that the optimization of $\mathbf{P}_i$ can be divided into two decoupled problems: optimizing the transmit directions—singular vectors—on one hand, and optimizing the transmit powers—singular values—on the other hand.

We would like to draw the reader's attention to the fact that the proof of this theorem does not rely on the expression of the asymptotic mutual information given in (17). In fact, Theorem 2 is first proved in the nonasymptotic regime for an arbitrary set of $\{k_i\}_{i\in\mathbb{N}_0^N}$. As such, regardless of the system size, the singular vectors of the precoding matrices should always be aligned to the eigenvectors of the channel correlation matrices to maximize the average mutual information. In particular, the singular vectors of the precoding matrices that maximize the asymptotic average mutual information are also aligned to the eigenvectors of channel correlation matrices as in (20). Furthermore, from **Result 3** in Section III, we can conclude that the singular vectors given in (20) are also those that maximize the asymptotic instantaneous mutual information.

Finally, we would like to point out that a result similar to Theorem 2 was proved in [9] for a two-hop system with a single noisy relay, uncorrelated channels $\mathbf{H}_1$ and $\mathbf{H}_2$, and full channel knowledge at source, relay and destination: the left and right singular vectors of the optimal relay precoder were shown to be aligned to the eigenvectors of matrices $\mathbf{H}_1\mathbf{H}_1^H$ and $\mathbf{H}_2\mathbf{H}_2^H$, respectively. This result encourages us to believe that in the case of noisy relays, Theorem 2 may still hold for correlated channels, and statistical channel knowledge at source and relays.

V. Application to MIMO Communication Scenarios

In this section, Theorem 1 and Theorem 2 are applied to four different communication scenarios. In the first two scenarios, the special case of nonrelay assisted MIMO (N $=$ 1) without path-loss ($a_1 = 1$) is considered, and we show how (17) boils down to known results for the MIMO channel with or without correlation. In the third and fourth scenarios, a multihop MIMO system is considered and the asymptotic mutual information is developed in the uncorrelated and exponential correlation cases, respectively. The application of Theorem 1 and Theorem 2 to these scenarios will also serve as a base for simulations in Section VI, which validate the asymptotic expression in Theorem 1, and show the impact of relaying on the communication rate in presence or absence of correlation.

A. Single-Hop MIMO With Statistical CSI at Source

Consider a simple single-hop uncorrelated MIMO system with the same number of antennas at source and destination i.e., $\rho_0 = \rho_1 = 1$.

Assuming an i.i.d. Rayleigh fading channel i.e., $\mathbf{C}_{t,1} = \mathbf{C}_{r,1} = \mathbf{I}$ and equal power allocation at source antennas, the source precoder is $\mathbf{P}_0 = \sqrt{\mathcal{P}_0}\mathbf{I}$. Under these assumptions, the asymptotic mutual information (17) can easily be shown to be

$$\boldsymbol{I}_\infty = 2\log\left(\frac{1+\sqrt{1+4\eta\mathcal{P}_0}}{2}\right) - \frac{\log e}{4\eta\mathcal{P}_0}\left(\sqrt{1+4\eta\mathcal{P}_0}-1\right)^2. \tag{21}$$

It can be observed that the deterministic expression (21) depends only on the system characteristics and is independent from the channel realizations. Moreover, equal power allocation is known to be the capacity-achieving power allocation for a MIMO i.i.d. Rayleigh channel with statistical CSI at source [23, Section 3.3.2], [16]. As such, the asymptotic mutual information (21) also represents the asymptotic capacity of the system. We should also mention that (21) is similar to the expression of the asymptotic capacity per dimension previously derived in [23, Section 3.3.2] for the MIMO Rayleigh channel with equal number of transmit and receive antennas and statistical CSI at the transmitter.

In the more general case of correlated MIMO channel with separable correlation we have $\mathbf{H}_1 = \mathbf{C}_{r,1}^{1/2}\mathbf{\Theta}_1\mathbf{C}_{t,1}^{1/2}$. Let us denote the eigenvalue decomposition of $\mathbf{C}_{t,1} = \mathbf{U}_{t,1}\mathbf{\Lambda}_{t,1}\mathbf{U}_{t,1}^H$, where $\mathbf{\Lambda}_{t,1}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\mathbf{C}_{t,1}$ in the nonincreasing order and the unitary matrix $\mathbf{U}_{t,1}$ contains the corresponding eigenvectors. Defining the transmit covariance matrix $\mathbf{Q} \triangleq \mathrm{E}\left[\mathbf{x}_0\mathbf{x}_0^H\right] = \mathbf{P}_0\mathbf{P}_0^H$, it has been shown [21] that the capacity-achieving matrix $\mathbf{Q}^\star$ is given by

$$\mathbf{Q}^\star = \mathbf{U}_{t,1}\mathbf{\Lambda}_{\mathbf{Q}^\star}\mathbf{U}_{t,1}^H \tag{22}$$

where $\mathbf{\Lambda}_{\mathbf{Q}^\star}$ is a diagonal matrix containing the capacity-achieving power allocation. Under these assumptions, the asymptotic mutual information (17) becomes equivalent to the expression[1] obtained in [23, Theorem 3.7] for the capacity of the correlated MIMO channel with statistical CSI at transmitter.

B. Uncorrelated Multihop MIMO With Statistical CSI at Source and Relays

In this example, we consider an uncorrelated multihop MIMO system, i.e., all correlation matrices are equal to identity. Then by Theorem 2 the optimal precoding matrices should be diagonal. Assuming equal power allocation at source and relays, the precoding matrices are of the form $\mathbf{P}_i = \alpha_i\mathbf{I}_{k_i}$, where α_i is real positive and chosen to respect the power constraints.

Using the power constraint expression (19) in Appendix C, it can be shown by induction on i that the coefficients α_i in the uncorrelated case are given by

$$\begin{aligned}\alpha_0 &= \sqrt{\mathcal{P}_0}\\ \alpha_i &= \sqrt{\frac{\mathcal{P}_i}{a_i\mathcal{P}_{i-1}}} \qquad \forall i \in \mathbb{N}_1^{N-1}\\ \alpha_N &= 1.\end{aligned} \tag{23}$$

Then the asymptotic mutual information for the uncorrelated multihop MIMO system with equal power allocation is given by

$$\boldsymbol{I}_\infty = \sum_{i=0}^{N}\frac{\rho_i}{\rho_0}\log\left(1+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}\right) - N\frac{\log e}{\rho_0}\eta\prod_{i=0}^{N}h_i \tag{24}$$

where $h_0, h_1, \ldots, h_N$ are the solutions of the system of $N+1$ multivariate polynomial equations

$$\prod_{j=0}^{N}h_j = \frac{h_i^N\alpha_i^2 a_{i+1}}{1+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}} \qquad \forall i \in \mathbb{N}_0^N. \tag{25}$$

Note that the asymptotic mutual information is a deterministic value depending only on a few system characteristics: signal power $\mathcal{P}_i$, noise power $1/\eta$, pathloss a_i, number of hops N and ratio of the number of antennas ρ_i.

[1]The small differences between the expression derived from (17) and the capacity expression in [23, Theorem 3.7] are due to different normalization assumptions in [23]. In particular (17) is the mutual information per source antenna while the expression in [23] is the capacity per receive antenna.

C. Exponentially Correlated Multihop MIMO With Statistical CSI at Source and Relays

In this example, the asymptotic mutual information (17) is developed in the case of exponential correlation matrices and precoding matrices with singular vectors as in Theorem 2. Exponential correlation matrices are a common model of correlation in uniform linear antenna array (ULA) [24]–[26].

Exponential Correlation Model: We assume that Level i is equipped with a uniform linear array (ULA) of length L_i, characterized by its antenna spacing $l_i = L_i/k_i$ and its characteristic distances $\Delta_{t,i}$ and $\Delta_{r,i}$ proportional to transmit and receive spatial coherences, respectively. Then the receive and transmit correlation matrices at Level i can respectively be modeled by the following Hermitian Wiener-class[2] Toeplitz matrices [24]–[26]

$$\mathbf{C}_{r,i} = \begin{bmatrix} 1 & r_{r,i} & r_{r,i}^2 & \ldots & r_{r,i}^{k_i-1}\\ r_{r,i} & 1 & \ddots & \ddots & \vdots\\ r_{r,i}^2 & \ddots & \ddots & \ddots & r_{r,i}^2\\ \vdots & \ddots & \ddots & 1 & r_{r,i}\\ r_{r,i}^{k_i-1} & \ldots & r_{r,i}^2 & r_{r,i} & 1\end{bmatrix}_{k_i\times k_i}$$

and

$$\mathbf{C}_{t,i+1} = \begin{bmatrix} 1 & r_{t,i+1} & r_{t,i+1}^2 & \ldots & r_{t,i+1}^{k_i-1}\\ r_{t,i+1} & 1 & \ddots & \ddots & \vdots\\ r_{t,i+1}^2 & \ddots & \ddots & \ddots & r_{t,i+1}^2\\ \vdots & \ddots & \ddots & 1 & r_{t,i+1}\\ r_{t,i+1}^{k_i-1} & \ldots & r_{t,i+1}^2 & r_{t,i+1} & 1\end{bmatrix}_{k_i\times k_i} \tag{26}$$

where the antenna correlation at receive (resp. transmit) side $r_{r,i} = e^{-\frac{l_i}{\Delta_{r,i}}} \in [0,1)$ (resp. $r_{t,i+1} = e^{-\frac{l_i}{\Delta_{t,i}}} \in [0,1)$) is an exponential function of antenna spacing l_i and characteristic distance $\Delta_{r,i}$ (resp. $\Delta_{t,i}$) at relaying Level i.

Equal Power Allocation Over Optimal Precoding Directions: We further assume equal power allocation over the optimal directions, i.e., the singular values of $\mathbf{P}_i$ are chosen to be all equal: $\mathbf{\Lambda}_{P_i} = \alpha_i\mathbf{I}_{k_i}$, where α_i is real positive and chosen to respect the power constraint (7). Equal power allocation may not be the optimal power allocation scheme, but it is considered in this example for simplicity.

Using the power constraint expression for general correlation models (19) and considering precoding matrices $\mathbf{P}_i = \mathbf{U}_{r,i}^H(\alpha_i\mathbf{I}_{k_i})\mathbf{U}_{t,i+1}$ with singular vectors as in Theorem 2 and equal singular values α_i, we can show by induction on i

[2]A sequence of $n\times n$ Toeplitz Matrices $\mathbf{T}_n = [t_{k-j}]_{n\times n}$ is said to be in the Wiener class [27, Section 4.4] if the sequence $\{t_k\}$ of first-column and first-row elements is absolutely summable, i.e., $\lim_{n\to+\infty}\sum_{k=-n}^{n}|t_k| < +\infty$.If $|r_{r,i}| < 1$, then $\lim_{k_i\to+\infty}(\sum_{k=0}^{k_i-1}r_{r,i}^k + \sum_{k=-k_i-1}^{-1}r_{r,i}^{-k}) = \frac{1}{1-r_{r,i}} + \frac{1/r_{r,i}}{1-1/r_{r,i}} < \infty$, and consequently $\mathbf{C}_{r,i}$ is in the Wiener class. $\mathbf{C}_{t,i}$ is obviously also in the Wiener class if $|r_{t,i}| < 1$.

that the coefficients α_i respecting the power constraints for any correlation model are given by

$$\begin{aligned}\alpha_0 &= \sqrt{\mathcal{P}_0}\\ \alpha_i &= \sqrt{\frac{\mathcal{P}_i}{a_i\mathcal{P}_{i-1}}\frac{\operatorname{tr}(\mathbf{\Lambda}_{r,i-1})}{\operatorname{tr}(\mathbf{\Lambda}_{r,i})}\frac{k_i}{\operatorname{tr}(\mathbf{\Lambda}_{t,i}\mathbf{\Lambda}_{r,i-1})}}\;,\ \forall i\in\mathbb{N}_1^{N-1}\\ \alpha_N &= 1.\end{aligned}\tag{27}$$

We would like to point out that (27) is a general expression that holds not only for the exponential correlation model, but also for any correlation model as long as the singular vectors of the precoding matrices are chosen as in Theorem 2. Applying the exponential correlation model to (27) and making the dimensions of the system grow large, it can be shown that in the asymptotic regime, the α_i respecting the power constraint for the exponentially correlated system converge to the same value (23) as for the uncorrelated system.

Asymptotic Mutual Information: Under the assumptions of exponential channel correlation matrices, precoders with singular vectors as in Theorem 2, and equal power allocation over these precoding directions, we show in Appendix D that the asymptotic mutual information is given by (28) [see (28) and (29) at the bottom of the page], where $h_0, h_1, \dots, h_N$ are the solutions of the system of $N+1$ (29), and for all $i\in\mathbb{N}_0^N$

$$\begin{aligned}c_{r,i} &= \frac{1-r_{r,i}}{1+r_{r,i}}\\ c_{t,i+1} &= \frac{1-r_{t,i+1}}{1+r_{t,i+1}}\\ m_i &= 1-\frac{\left(\frac{c_{t,i+1}}{c_{r,i}}+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}\right)\left(\frac{c_{r,i}}{c_{t,i+1}}+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}\right)}{\left(\frac{1}{c_{r,i}c_{t,i+1}}+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}\right)\left(c_{r,i}c_{t,i+1}+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}\right)}\end{aligned}\tag{30}$$

(with the convention $r_{r,0} = r_{t,N+1} = 0$). Those expressions show that only a few relevant parameters affect the performance of this complex system: signal power $\mathcal{P}_i$, noise power $1/\eta$, pathloss a_i, number of hops N, ratio of the number of antennas ρ_i, and correlation ratios $c_{r,i}$ and $c_{t,i}$.

VI. Numerical Results

In this section, we present numerical results to validate Theorem 1 and to show that even with small k_i, for all $i\in\mathbb{N}_0^N$, the behavior of the system is close to its behavior in the asymptotic regime, making Theorem 1 a useful tool for optimization of finite-size systems as well as large networks.

A. Uncorrelated Multihop MIMO

The uncorrelated system described in Section V-B is first considered.

Fig. 2(a) plots the asymptotic mutual information from Theorem 1 as well as the instantaneous mutual information obtained for an arbitrary channel realization (shown as experimental curves in the figure) in the case of one, two or three hops. Experimental curves are drawn for systems with 10 antennas at source, destination and each relay, or 100 antennas at each level. When increasing the number of hops N, the distance between source and destination d is kept constant and $N-1$ relays are inserted between source and destination with equal spacing $d_i = d/N$ between each relaying level. In both examples, whose main purpose is not to optimize the system, but to validate the asymptotic formula in Theorem 1, matrices P_i are taken proportional to the identity matrix to simulate equal power allocation. The channel correlation matrices are also equal to the identity matrix to mimic the uncorrelated channel. Moreover, the pathloss exponent $\beta = 2$ is considered. We would like to point out that the experimental curves for different channel realizations produced similar results. As such, the experimental curve corresponding to a single channel realization is shown for the sake of clarity and conciseness.

Fig. 2(a) shows the perfect match between the instantaneous mutual information for an arbitrary channel realization with 100 antennas at each level and the asymptotic mutual information, validating Theorem 1 for large network dimensions. On the other hand, with 10 antennas at each level, it appears that the instantaneous mutual information of a system with a small number of antennas behaves very closely to the asymptotic regime, justifying the usefulness of the asymptotic formula even when evaluating the end-to-end mutual information of a system with small size.

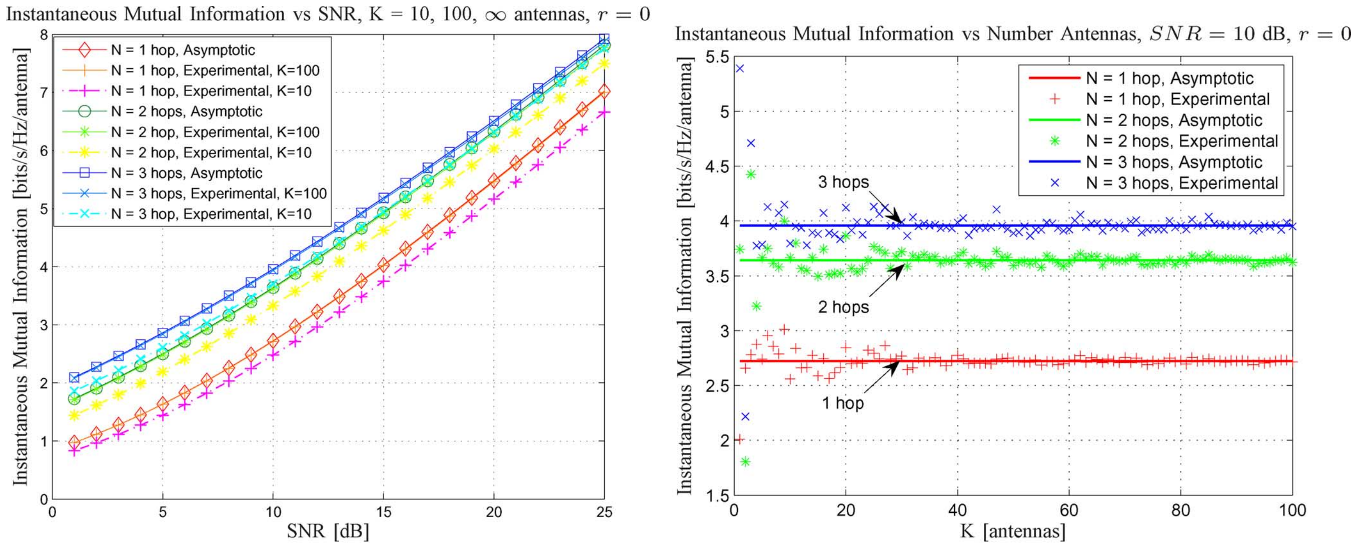
Finally, Fig. 2(b) plots the asymptotic mutual information for one, two, and three hops, as well as the value of the instantaneous mutual information for random channel realizations when the number of antennas at all levels increases. The concentration of the instantaneous mutual information values around the asymptotic limit when the system size increases shows the convergence of the instantaneous mutual information towards the asymptotic limit as the number of antennas grows large at all levels with the same rate.

B. One-Sided Exponentially Correlated Multihop MIMO

Based on the model discussed in Section V-C, the one-sided exponentially correlated system is considered in this section. In

$$\boldsymbol{I}_\infty = \sum_{i=0}^{N}\frac{\rho_i}{\rho_0\pi^2}\int_{t=-\infty}^{+\infty}\int_{u=-\infty}^{+\infty}\log\left(1+c_{r,i}c_{t,i+1}\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}\frac{(1+t^2)}{(c_{r,i}^2+t^2)}\frac{(1+u^2)}{(c_{t,i+1}^2+u^2)}\right)\frac{dt}{1+t^2}\frac{du}{1+u^2} - N\frac{\log e}{\rho_0}\eta\prod_{i=0}^{N}h_i \tag{28}$$

$$\prod_{j=0}^{N}h_j = \frac{2}{\pi}\frac{h_i^N a_{i+1}\alpha_i^2}{\sqrt{c_{r,i}c_{t,i+1}+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}}\sqrt{\frac{1}{c_{r,i}c_{t,i+1}}+\frac{\eta h_i^N a_{i+1}\alpha_i^2}{\rho_i}}}K(m_i)\qquad \forall i\in\mathbb{N}_0^N \tag{29}$$



(a) (b)

Fig. 2. Uncorrelated case: asymptotic mutual information and instantaneous mutual information for single-hop MIMO, 2 hops, and 3 hops. (a) Mutual information versus SNR with $\mathrm{K} = 10, 100, \infty$ antennas. (b) Mutual information versus K_N, at $\mathrm{SNR} = 10$ dB.

the case of one-sided correlation, e.g., $r_{r,i} = 0$ and $r_{t,i} \geq 0$ for all $i \in \{0, \ldots, N\}$, the asymptotic mutual information (28), (29) is reduced to

$$\boldsymbol{I}_\infty = \sum_{i=0}^{N} \frac{\rho_i}{\rho_0 \pi} \int_{-\infty}^{+\infty} \log\left(1 + c_{t,i+1} \frac{\eta h_i^N a_{i+1} \alpha_i^2}{\rho_i} \frac{(1+u^2)}{(c_{t,i+1}^2 + u^2)}\right) \frac{du}{1+u^2} - N \frac{\log e}{\rho_0} \eta \prod_{i=0}^{N} h_i \tag{31}$$

where $h_0, h_1, \ldots, h_N$ are the solutions of the system of $N+1$ equations

$$\prod_{j=0}^{N} h_j = \frac{h_i^N a_{i+1} \alpha_i^2}{\sqrt{c_{t,i+1} + \frac{\eta h_i^N a_{i+1} \alpha_i^2}{\rho_i}} \sqrt{\frac{1}{c_{t,i+1}} + \frac{\eta h_i^N a_{i+1} \alpha_i^2}{\rho_i}}}. \tag{32}$$

One-sided correlation was considered to avoid the involved computation of the elliptic integral $K(m_i)$ in the system of (29), and therefore, to simplify simulations.

Fig. 3(a) plots the asymptotic mutual information for one, two or three hops, as well as the instantaneous mutual information obtained for an arbitrary channel realization (shown as experimental curves in the figure) for 10 and 100 antennas at each level. As in the uncorrelated case, the perfect match of the experimental and asymptotic curves in Fig. 3(a) with 100 antennas validates the asymptotic formula in Theorem 1 in the presence of correlation. Fig. 3(a) also shows that even for a small number of antennas, the system behaves closely to the asymptotic regime in the correlated case.

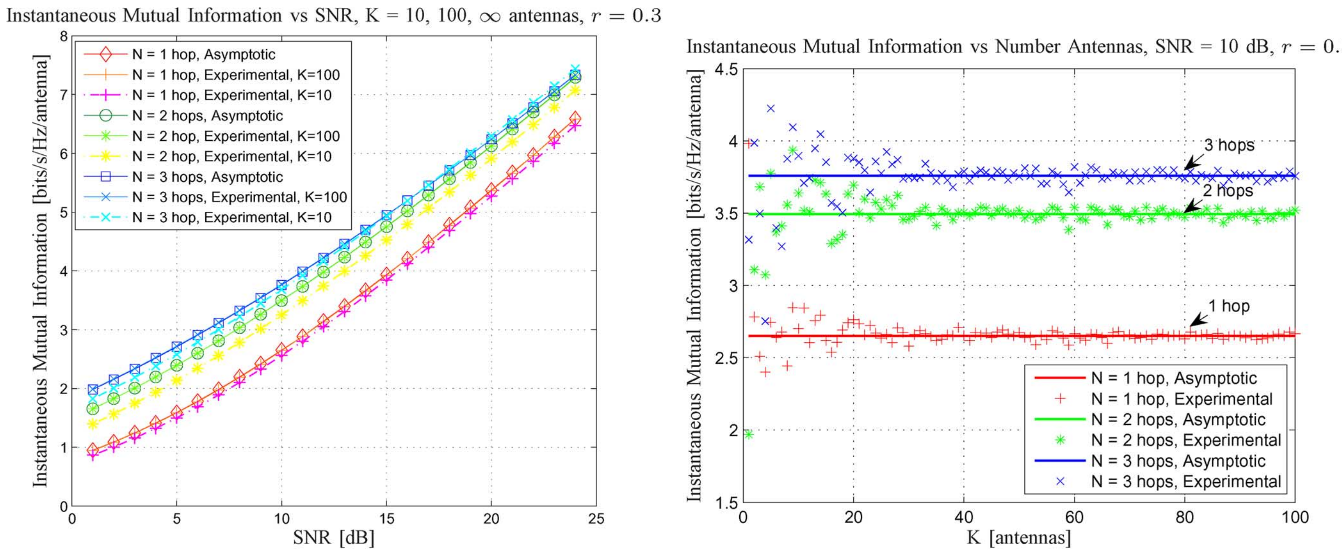
Finally, Fig. 3(b) plots the instantaneous mutual information for random channel realizations against the size of the system and shows its convergence towards the asymptotic mutual information when the number of antennas increases. We would like to mention that simulations for higher values of the correlation $r_{t,i}$ showed that convergence towards the asymptotic limit is slower when correlation increases.

VII. Conclusion and Research Perspectives

We studied a multihop MIMO relay network in the correlated fading environment, where relays at each level perform linear precoding on their received signal prior to retransmitting it to the next level. Using free probability theory, a closed-form expression of the instantaneous end-to-end mutual information was derived in the asymptotic regime where the number of antennas at all levels grows large. The asymptotic instantaneous end-to-end mutual information turns out to be a deterministic quantity that depends only on channel statistics and not on particular channel realizations. Moreover, it also serves as the asymptotic value of the average end-to-end mutual information. Simulation results verified that, even with a small number of antennas at each level, multihop systems behave closely to the asymptotic regime. This observation makes the derived asymptotic mutual information a powerful tool to optimize the instantaneous mutual information of finite-size systems with only statistical knowledge of the channel.

We also showed that for any system size the left and right singular vectors of the optimal precoding matrices that maximize the average mutual information are aligned, at each level, with the eigenvectors of the transmit and receive correlation matrices of the forward and backward channels, respectively. Thus, the singular vectors of the optimal precoding matrices can be determined with only local statistical channel knowledge at each level.

In the sequel, the analysis of the end-to-end mutual information in the asymptotic regime will first be extended to the case where noise impairs signal reception at each relaying level. Then, combining the expression of the asymptotic mutual information with the singular vectors of the optimal precoding matrices, future work will focus on optimizing the power allocation



(a) (b)

Fig. 3. One-sided exponential correlation case: asymptotic mutual information and instantaneous mutual information for $r = 0.3$, and single-hop MIMO, 2 hops, and 3 hops. (a) Mutual information versus SNR with $K = 10, 100, \infty$ antennas. (b) Mutual information versus K_N, at $\mathrm{SNR} = 10$ dB.

determined by the singular values of the precoding matrices. Finally future research directions also include the analysis of the relay-clustering effect, and the optimal size of clusters in correlated fading is expected to depend on the SNR regime.

Appendix A
Transforms and Lemmas

Transforms and lemmas used in the proofs of Theorems 1 and 2 are provided and proved in this appendix, while the proofs of Theorems 1 and 2 are detailed in Appendices B and C, respectively.

A) Transforms: Let $\mathbf{T}$ be a square matrix of size n with real eigenvalues $\lambda_{\mathbf{T}}(1), \ldots, \lambda_{\mathbf{T}}(n)$. The empirical eigenvalue distribution $F^n_{\mathbf{T}}$ of $\mathbf{T}$ is defined by

$$F^n_{\mathbf{T}}(x) \triangleq \frac{1}{n}\sum_{i=1}^{n} u(x - \lambda_{\mathbf{T}}(i)). \tag{33}$$

We define the following transformations [10]

$$Stieltjes\ transform\text{: } G_{\mathbf{T}}(s) \triangleq \int \frac{1}{\lambda - s} dF_{\mathbf{T}}(\lambda) \tag{34}$$

$$\Upsilon_{\mathbf{T}}(s) \triangleq \int \frac{s\lambda}{1 - s\lambda} dF_{\mathbf{T}}(\lambda) \tag{35}$$

$$S\text{-}transform\text{: } S_{\mathbf{T}}(z) \triangleq \frac{z+1}{z}\Upsilon^{-1}_{\mathbf{T}}(z) \tag{36}$$

where $\Upsilon^{-1}(\Upsilon(s)) = s$.

B) Lemmas: We present here the lemmas used in the proofs of Theorems 1 and 2. Lemmas 1, 3, 5, and 7 are proved in Appendix A-C, while Lemmas 2, 6, and 4 are taken from [27], [28], and [29], respectively.

Lemma 1: Consider an $n \times p$ matrix $\mathbf{A}$ and a $p \times n$ matrix $\mathbf{B}$, such that their product $\mathbf{AB}$ has non-negative real eigenvalues. Denote $\xi = \frac{p}{n}$. Then

$$S_{\mathbf{AB}}(z) = \frac{z+1}{z+\xi} S_{\mathbf{BA}}\left(\frac{z}{\xi}\right). \tag{37}$$

Note that Lemma 1 is a more general form of the results derived in [30, Eq. (1.2)], [10, Eq. (15)].

Lemma 2 ([28, Prop. 4.4.9 and 4.4.11]): For $n \in \mathbb{N}$, let $p(n) \in \mathbb{N}$ be such that $\frac{p(n)}{n} \to \xi$ as $n \to \infty$. Let:

- $\mathbf{\Theta}(n)$ be a $p(n) \times n$ complex Gaussian random matrix with i.i.d. elements with variance $\frac{1}{n}$.
- $\mathbf{A}(n)$ be an $n \times n$ constant matrix such that $\sup_n \|\mathbf{A}(n)\| < +\infty$ and $(\mathbf{A}(n), \mathbf{A}(n)^H)$ has the limit eigenvalue distribution μ.
- $\mathbf{B}(n)$ be a $p(n) \times p(n)$ Hermitian random matrix, independent from $\mathbf{\Theta}(n)$, with an empirical eigenvalue distribution converging almost surely to a compactly supported probability measure ν.

Then, as $n \to \infty$:

- the empirical eigenvalue distribution of $\mathbf{\Theta}(n)^H\mathbf{B}(n)\mathbf{\Theta}(n)$ converges almost surely to the compound free Poisson distribution $\pi_{\nu,\xi}$ [28];
- the family $(\{\mathbf{\Theta}(n)^H\mathbf{B}(n)\mathbf{\Theta}(n)\}, \{\mathbf{A}(n), \mathbf{A}(n)^H\})$ is asymptotically free almost everywhere.

Thus, the limiting eigenvalue distribution of $\mathbf{\Theta}(n)\mathbf{B}(n)\mathbf{\Theta}(n)^H\mathbf{A}(n)\mathbf{A}(n)^H$ is the free convolution $\pi_{\nu,\xi} \boxtimes \mu$ and its *S-transform* is

$$S_{\mathbf{\Theta B \Theta}^H\mathbf{AA}^H}(z) = S_{\mathbf{\Theta B \Theta}^H}(z) S_{\mathbf{AA}^H}(z). \tag{38}$$

Note that if the elements of $\mathbf{\Theta}(n)$ had variance $\frac{1}{p(n)}$ instead of $\frac{1}{n}$, $(\{\mathbf{\Theta}(n)^H\mathbf{B}(n)\mathbf{\Theta}(n)\}, \{\mathbf{A}(n), \mathbf{A}(n)^H\})$ would still be asymptotically free almost everywhere, and consequently, (38) would still hold.

Lemma 3: Consider an $n \times p$ matrix $\mathbf{A}$ with zero-mean i.i.d. entries with variance $\frac{a}{p}$. Assume that the dimensions go to infinity while $\frac{n}{p} \to \zeta$, then

$$S_{\mathbf{AA}^H}(z) = \frac{1}{a}\ \frac{1}{(1+\zeta z)}$$
$$S_{\mathbf{A}^H\mathbf{A}}(z) = \frac{1}{a}\ \frac{1}{(z+\zeta)}. \tag{39}$$

Lemma 4 ([29, Theorem H.1.h]): Let $\mathbf{A}$ and $\mathbf{B}$ be two positive semi-definite hermitian matrices of size $n \times n$. Let $\lambda_{\mathbf{A}}(i)$ and $\lambda_{\mathbf{B}}(i)$ be their decreasingly-ordered eigenvalues, respectively. Then the following inequality holds:

$$\sum_{i=1}^{n} \lambda_{\mathbf{A}}(i)\lambda_{\mathbf{B}}(n-i+1) \leq \operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^{n} \lambda_{\mathbf{AB}}(i) \leq \sum_{i=1}^{n} \lambda_{\mathbf{A}}(i)\lambda_{\mathbf{B}}(i). \tag{40}$$

Lemma 5: For $i \in \mathbb{N}_1^N$, let $\mathbf{A}_i$ be a $n_i \times n_{i-1}$ random matrix. Assume that:

- $\mathbf{A}_1, \ldots, \mathbf{A}_N$ are mutually independent,
- n_i goes to infinity while $\frac{n_i}{n_{i-1}} \to \zeta_i$,
- as n_i goes to infinity, the eigenvalue distribution of $\mathbf{A}_i\mathbf{A}_i^H$ converges almost surely in distribution to a compactly supported measure ν_i,
- as $n_1, \ldots, n_N$ go to infinity, the eigenvalue distribution of $(\bigotimes_{i=N}^1 \mathbf{A}_i)(\bigotimes_{i=N}^1 \mathbf{A}_i)^H$ converges almost surely in distribution to a measure μ_N.

Then μ_N is compactly supported.

Lemma 6 ([27, Theorem 9]): Let $\mathbf{T}_n$ be a sequence of Wiener-class Toeplitz matrices, characterized by the function $f(\lambda)$ with essential infimum m_f and essential supremum M_f. Let $\lambda_{\mathbf{T}_n}(1), \ldots, \lambda_{\mathbf{T}_n}(n)$ be the eigenvalues of $\mathbf{T}_n$ and s be any positive integer. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \lambda_{\mathbf{T}_n}^s(k) = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda)^s d\lambda. \tag{41}$$

Furthermore, if $f(\lambda)$ is real, or equivalently, the matrices $\mathbf{T}_n$ are all Hermitian, then for any function $g(\cdot)$ continuous on $[m_f, M_f]$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(\lambda_{\mathbf{T}_n}(k)) = \frac{1}{2\pi} \int_0^{2\pi} g(f(\lambda)) d\lambda. \tag{42}$$

Lemma 7: For $i \geq 1$, given a set of deterministic matrices $\{\mathbf{A}_k\}_{k\in\{0,\ldots,i\}}$ and a set of independent random matrices $\{\mathbf{\Theta}_k\}_{k\in\{1,\ldots,i\}}$, with i.i.d. zero-mean gaussian elements with variance σ_k^2

$$\operatorname{tr}\left(\mathrm{E}\left[\bigotimes_{k=i}^{1} \{\mathbf{A}_k \mathbf{\Theta}_k\} \mathbf{A}_0 \mathbf{A}_0^H \bigotimes_{k=1}^{i} \{\mathbf{\Theta}_k^H \mathbf{A}_k^H\} \right] \right) = \operatorname{tr}(\mathbf{A}_0\mathbf{A}_0^H) \prod_{k=1}^{i} \sigma_k^2 \operatorname{tr}(\mathbf{A}_k\mathbf{A}_k^H). \tag{43}$$

C) Proofs of Lemmas: The proofs of Lemmas 1, 3, 5, and 7 are given hereafter.

Proof of Lemma 1: Given two complex matrices $\mathbf{A}$ of size $m \times n$, and $\mathbf{B}$ of size $n \times m$, their products $\mathbf{AB}$ and $\mathbf{BA}$ have the same k nonzero eigenvalues $\lambda_{\mathbf{AB}}(1), \ldots, \lambda_{\mathbf{AB}}(k)$ with the same respective multiplicities $m_1, \ldots, m_k$. However, the multiplicities m_0 and m_0' of the 0-eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$, respectively, are related as follows:

$$m_0 + n = m_0' + m. \tag{44}$$

Assuming that $\mathbf{AB}$, and therefore, $\mathbf{BA}$, has real eigenvalues, we show hereafter that (37) holds.

The empirical eigenvalue distributions of $\mathbf{AB}$ and $\mathbf{BA}$ are defined by

$$F_{\mathbf{AB}}^m(\lambda) = \frac{m_0}{m} u(\lambda) + \frac{1}{m} \sum_{i=1}^{k} m_i u(\lambda - \lambda_{\mathbf{AB}}(i))$$

$$F_{\mathbf{BA}}^n(\lambda) = \frac{m_0'}{n} u(\lambda) + \frac{1}{n} \sum_{i=1}^{k} m_i u(\lambda - \lambda_{\mathbf{AB}}(i)). \tag{45}$$

Using (44), we get

$$F_{\mathbf{AB}}^m(\lambda) = \frac{n}{m} F_{\mathbf{BA}}^n(\lambda) + \left(1 - \frac{n}{m}\right) u(\lambda). \tag{46}$$

From (46), it is direct to show that

$$G_{\mathbf{AB}}(z) = \frac{n}{m} G_{\mathbf{BA}}(z) - \left(1 - \frac{n}{m}\right) \frac{1}{z}. \tag{47}$$

As $\Upsilon(s) = -1 - \frac{1}{s} G(\frac{1}{s})$, from (47), we obtain

$$\Upsilon_{\mathbf{AB}}(s) = \frac{n}{m} \Upsilon_{\mathbf{BA}}(s). \tag{48}$$

Finally, using $\{z = \Upsilon_{\mathbf{AB}}(s) = \frac{n}{m}\Upsilon_{\mathbf{BA}}(s)\} \Leftrightarrow \{\Upsilon_{\mathbf{AB}}^{-1}(z) = s = \Upsilon_{\mathbf{BA}}^{-1}\left(\frac{z}{n/m}\right)\}$ and the definition of the *S-transform* $S(z) \triangleq \frac{z+1}{z}\Upsilon^{-1}(z)$ yields the desired result

$$S_{\mathbf{AB}}(z) = \frac{z+1}{z + \frac{n}{m}} S_{\mathbf{BA}}\left(\frac{z}{n/m}\right). \tag{49}$$

This concludes the proof of Lemma 1. ■

Proof of Lemma 3: Consider an $n \times p$ matrix $\mathbf{A}$ with zero-mean i.i.d. entries with variance $\frac{a}{p}$. Let $\mathbf{X} = \frac{1}{\sqrt{a}}\mathbf{A}$ denote the normalized version of $\mathbf{A}$ with zero-mean i.i.d. entries of variance $\frac{1}{p}$ and define $\mathbf{Y} = a\mathbf{I}_n$ and $\mathbf{Z} = \mathbf{XX}^H\mathbf{Y} = \mathbf{AA}^H$. It is direct to show that $S_{\mathbf{Y}}(z) = \frac{1}{a}$. Using the latter result along with [10, Theorem 1], we obtain

$$S_{\mathbf{XX}^H}(z) = \frac{1}{(1+\zeta z)}$$

$$S_{\mathbf{AA}^H}(z) = S_{\mathbf{Z}}(z) = S_{\mathbf{XX}^H}(z) S_{\mathbf{Y}}(z) = \frac{1}{(1+\zeta z)} \frac{1}{a}. \tag{50}$$

Applying Lemma 1 to $S_{\mathbf{A}^H\mathbf{A}}(z)$ yields

$$S_{\mathbf{A}^H\mathbf{A}}(z) = \frac{z+1}{z+\zeta} S_{\mathbf{AA}^H}\left(\frac{z}{\zeta}\right) = \frac{1}{a} \frac{1}{(z+\zeta)}. \tag{51}$$

This completes the proof of Lemma 3. ■

Proof of Lemma 5: The proof of Lemma 5 is done by induction on N. For $N=1$, Lemma 5 obviously holds. Assuming that Lemma 5 holds for N, we now show that it also holds for $N+1$.

We first recall that the eigenvalues of Gramian matrices $\mathbf{A}\mathbf{A}^H$ are non-negative. Thus, the support of μ_{N+1} is lower-bounded by 0, and we are left with showing that it is also upper-bounded.

Denoting $\mathbf{B}_N = (\bigotimes_{i=N}^1 \mathbf{A}_i)(\bigotimes_{i=N}^1 \mathbf{A}_i)^H$, we can write

$$\mathbf{B}_{N+1} = \mathbf{A}_{N+1}\mathbf{B}_N\mathbf{A}_{N+1}^H. \tag{52}$$

For a matrix $\mathbf{A}$, let $\lambda_{\mathbf{A},\max}$ denote its largest eigenvalue. The largest eigenvalue of $\mathbf{B}_{N+1}$ is given by

$$\begin{aligned}
&\lambda_{\mathbf{B}_{N+1},\max}\\
&= \max_{\mathbf{x}} \frac{\mathbf{x}^H\,\mathbf{B}_{N+1}\,\mathbf{x}}{\mathbf{x}^H\mathbf{x}}\\
&= \max_{\mathbf{x}} \frac{\mathbf{x}^H\,\mathbf{A}_{N+1}\mathbf{B}_N\mathbf{A}_{N+1}^H\,\mathbf{x}}{\mathbf{x}^H\mathbf{x}}\\
&= \max_{\mathbf{x}} \frac{\mathrm{tr}(\mathbf{B}_N\,\mathbf{A}_{N+1}^H\mathbf{x}\mathbf{x}^H\mathbf{A}_{N+1})}{\mathbf{x}^H\mathbf{x}}\\
&\le \max_{\mathbf{x}} \frac{\sum_{k=1}^{n_N}\lambda_{\mathbf{B}_N}(k)\,\lambda_{\mathbf{A}_{N+1}^H\mathbf{x}\mathbf{x}^H\mathbf{A}_{N+1}}(k)}{\mathbf{x}^H\mathbf{x}}, \text{ by Lemma 4}\\
&\le \max_{\mathbf{x}} \;\lambda_{\mathbf{B}_N,\max}\frac{\sum_{k=1}^{n_N}\lambda_{\mathbf{A}_{N+1}^H\mathbf{x}\mathbf{x}^H\mathbf{A}_{N+1}}(k)}{\mathbf{x}^H\mathbf{x}}\\
&= \lambda_{\mathbf{B}_N,\max}\;\max_{\mathbf{x}} \frac{\mathrm{tr}(\mathbf{A}_{N+1}^H\mathbf{x}\mathbf{x}^H\mathbf{A}_{N+1})}{\mathbf{x}^H\mathbf{x}}\\
&= \lambda_{\mathbf{B}_N,\max}\;\max_{\mathbf{x}} \frac{\mathbf{x}^H\mathbf{A}_{N+1}\mathbf{A}_{N+1}^H\mathbf{x}}{\mathbf{x}^H\mathbf{x}}\\
&= \lambda_{\mathbf{B}_N,\max}\;\lambda_{\mathbf{A}_{N+1}\mathbf{A}_{N+1}^H,\max}.
\end{aligned} \tag{53}$$

To simplify notations, we rename the random variables as follows:

$$X = \lambda_{\mathbf{B}_{N+1},\max} \quad Y = \lambda_{\mathbf{B}_N,\max} \quad Z = \lambda_{\mathbf{A}_{N+1}\mathbf{A}_{N+1}^H,\max}. \tag{54}$$

Then (53) can be rewritten

$$X \le YZ. \tag{55}$$

Let $a \ge 0$, by (55) we have

$$F_X(a) = \Pr\{X < a\} \ge \Pr\{YZ < a\} = F_{YZ}(a) \tag{56}$$

which still holds for the asymptotic distributions as $n_1,\ldots,n_{N+1} \to \infty$, while $\frac{n_i}{n_{i-1}} \to \zeta_i$. Denoting the plane region $\mathcal{D}_a = \{x,y \ge 0 / xy < a\}$, we can write

$$\begin{aligned}
&F_{YZ}(a)\\
&= \int\!\!\int_{y,z\in\mathcal{D}_a} f_{Y,Z}(y,z)dydz\\
&= \int\!\!\int_{y,z\in\mathcal{D}_a} f_Y(y)f_Z(z)dydz, \text{ by independence of } Y \text{ and } Z\\
&= \int_{y=0}^{+\infty}\left(\int_{z=0}^{a/y} f_Z(z)dz\right) f_Y(y)dy\\
&= \int_{y=0}^{+\infty} F_Z\left(\frac{a}{y}\right) f_Y(y)dy.
\end{aligned} \tag{57}$$

By assumption, the distributions of $\mathbf{A}_{N+1}\mathbf{A}_{N+1}^H$ and $\mathbf{B}_N$ converge almost surely to compactly supported measures. Thus, their largest eigenvalues are asymptotically upper-bounded and the support of the asymptotic distributions of Y and Z are upper-bounded, i.e.,

$$\begin{aligned}
&\exists c_y \ge 0 \text{ such that } \forall y \ge c_y,\ F_Y(y) = 1 \quad (f_Y(y) = 0)\\
&\exists c_z \ge 0 \text{ such that } \forall z \ge c_z,\ F_Z(z) = 1 \quad (f_Z(z) = 0).
\end{aligned} \tag{58}$$

Let $a \ge c_y\,c_z$, then for all $0 < y \le c_y$, we have $\frac{a}{y} \ge \frac{a}{c_y} \ge c_z$ and $F_Z\left(\frac{a}{y}\right) = 1$, as the dimensions go to infinity with constant rates. Therefore, in the asymptotic regime, we have

$$\begin{aligned}
F_{YZ}(a) &= \int_{y=0}^{c_y} F_Z\left(\frac{a}{y}\right) f_Y(y)dy\\
&= \int_{y=0}^{c_y} 1 f_Y(y)dy = F_Y(c_y) = 1.
\end{aligned} \tag{59}$$

Combining (56) and (59), we get $F_X(a) = 1$ for $a > c_y\,c_z$. Thus, there exists a constant c_x such that $0 \le c_x \le c_y\,c_z$ and $\forall x \ge c_x$, $F_X(x) = 1$, which means that the support of the asymptotic distribution of X is upper-bounded. As a consequence, the support of the asymptotic eigenvalue distribution of $\mathbf{B}_{N+1}$ is also upper-bounded. Therefore, the support of μ_{N+1} is upper-bounded, which concludes the proof. ■

Proof of Lemma 7: The proof of Lemma 7 is done by induction. We first prove that Lemma 7 holds for $i=1$. To that purpose, we define the matrix $\mathbf{B} = \mathbf{A}_1\mathbf{\Theta}_1\mathbf{A}_0\mathbf{A}_0^H\mathbf{\Theta}_1^H\mathbf{A}_1^H$. Then

$$\mathrm{tr}(\mathrm{E}[\mathbf{A}_1\mathbf{\Theta}_1\mathbf{A}_0\mathbf{A}_0^H\mathbf{\Theta}_1^H\mathbf{A}_1^H]) = \mathrm{tr}(\mathrm{E}[\mathbf{B}]) = \sum_{j=1}^{k_1}\mathrm{E}[b_{jj}]. \tag{60}$$

The expectation of the j^{th} diagonal element b_{jj} of $\mathbf{B}$ is

$$\begin{aligned}
\mathrm{E}[b_{jj}] &= \sum_{k,l,m,n,p}\mathrm{E}[a_{jk}^{(1)}\theta_{kl}^{(1)}a_{lm}^{(0)}a_{nm}^{(0)*}\theta_{pn}^{(1)*}a_{jp}^{(1)*}]\\
&= \sum_{k,l,m}|a_{jk}^{(1)}|^2|a_{lm}^{(0)}|^2\underbrace{\mathrm{E}[|\theta_{kl}^{(1)}|^2]}_{\sigma_1^2}\\
&= \sigma_1^2\sum_k|a_{jk}^{(1)}|^2\sum_{l,m}|a_{lm}^{(0)}|^2.
\end{aligned} \tag{61}$$

where the second equality is due to the fact that $\mathrm{E}[\theta_{kl}^{(1)}\theta_{pn}^{(1)*}] = \sigma_1^2\delta_{k,p}\delta_{l,n}$. It follows from (60) and (61) that

$$\begin{aligned}
\mathrm{tr}(\mathrm{E}[\mathbf{B}]) &= \sigma_1^2\sum_{j,k}|a_{jk}^{(1)}|^2\sum_{l,m}|a_{lm}^{(0)}|^2\\
&= \sigma_1^2\mathrm{tr}(\mathbf{A}_1\mathbf{A}_1^H)\mathrm{tr}(\mathbf{A}_0\mathbf{A}_0^H)
\end{aligned} \tag{62}$$

which shows that Lemma 7 holds for $i=1$.

Now, assuming that Lemma 7 holds for $i-1$, we show it also holds for i. We define the matrix $\mathbf{B}_i = \bigotimes_{k=i}^1\{\mathbf{A}_k\mathbf{\Theta}_k\}\mathbf{A}_0\mathbf{A}_0^H\bigotimes_{k=1}^i\{\mathbf{\Theta}_k^H\mathbf{A}_k^H\}$. Then

$$\mathrm{tr}(\mathrm{E}[\mathbf{B}_i]) = \sum_{j=1}^{k_1}\mathrm{E}[b_{jj}^{(i)}] = \mathrm{tr}(\mathrm{E}[\mathbf{A}_i\mathbf{\Theta}_i\mathbf{B}_{i-1}\mathbf{\Theta}_i^H\mathbf{A}_i^H]). \tag{63}$$

The expectation of the j^{th} diagonal element $b_{jj}^{(i)}$ of $\mathbf{B}_i$ is

$$\begin{aligned}\mathrm{E}[b_{jj}^{(i)}] &= \sum_{k,l,m,n} \mathrm{E}[a_{jk}^{(i)}\theta_{kl}^{(i)}b_{lm}^{(i-1)}\theta_{nm}^{(i)*}a_{jn}^{(i)*}] \\ &= \sum_{k,l}|a_{jk}^{(i)}|^2\mathrm{E}[b_{ll}^{(i-1)}]\underbrace{\mathrm{E}[|\theta_{kl}^{(i)}|^2]}_{\sigma_i^2} \\ &= \sigma_i^2\sum_k |a_{jk}^{(i)}|^2\sum_l \mathrm{E}[b_{ll}^{(i-1)}] \end{aligned} \tag{64}$$

where the second equality is due to the independence of $\boldsymbol{\Theta}_i$ and $\mathbf{B}_{i-1}$ and to the fact that $\mathrm{E}[\theta_{kn}^{(i)}\theta_{lm}^{(i)*}] = \sigma_i^2\delta_{k,p}\delta_{l,n}$. Thus, (63) becomes

$$\begin{aligned}\mathrm{tr}(\mathrm{E}[\mathbf{B}_i]) &= \sigma_i^2\sum_{j,k}|a_{jk}^{(i)}|^2\sum_l \mathrm{E}[b_{ll}^{(i-1)}] \\ &= \sigma_i^2\mathrm{tr}(\mathbf{A}_i\mathbf{A}_i^H)\mathrm{tr}(\mathrm{E}[\mathbf{B}_{i-1}]) \\ &= \sigma_i^2\mathrm{tr}(\mathbf{A}_i\mathbf{A}_i^H)\mathrm{tr}(\mathbf{A}_0\mathbf{A}_0^H)\prod_{k=1}^{i-1}\sigma_k^2\mathrm{tr}(\mathbf{A}_k\mathbf{A}_k^H) \\ &= \mathrm{tr}(\mathbf{A}_0\mathbf{A}_0^H)\prod_{k=1}^{i}\sigma_k^2\mathrm{tr}(\mathbf{A}_k\mathbf{A}_k^H)\end{aligned} \tag{65}$$

which shows that if Lemma 7 holds for $i-1$, then it holds for i.

Therefore, Lemma 7 holds for any $i \geq 1$, which concludes the proof. ∎

Appendix B
Proof of Theorem 1

In this appendix, we first list the main steps of the proof of Theorem 1 and then present the detailed proof of each step. Note that the proof of Theorem 1 uses tools from the free probability theory introduced in Appendix A. The proof of Theorem 1 consists of the following four steps.

1) Obtain $S_{\mathbf{G}_N\mathbf{G}_N^H}(z)$.
2) Use $S_{\mathbf{G}_N\mathbf{G}_N^H}(z)$ to find $\Upsilon_{\mathbf{G}_N\mathbf{G}_N^H}(z)$.
3) Use $\Upsilon_{\mathbf{G}_N\mathbf{G}_N^H}(z)$ to obtain $d\boldsymbol{I}/d\eta$.
4) Integrate $d\boldsymbol{I}/d\eta$ to obtain $\boldsymbol{I}$ itself.

- **First Step: obtain** $S_{\mathbf{G}_N\mathbf{G}_N^H}(z)$

Theorem 3: As $k_i, \forall i \in \mathbb{N}_0^N$, go to infinity with the same rate, the S-transform of $\mathbf{G}_N\mathbf{G}_N^H$ is given by

$$S_{\mathbf{G}_N\mathbf{G}_N^H}(z) = S_{\mathbf{M}_N^H\mathbf{M}_N}(z)\prod_{i=1}^{N}\frac{\rho_{i-1}}{a_i}\frac{1}{(z+\rho_{i-1})}S_{\mathbf{M}_{i-1}^H\mathbf{M}_{i-1}}\left(\frac{z}{\rho_{i-1}}\right). \tag{66}$$

Proof: The proof is done by induction using Lemmas 1, 3, 2. First, we prove (66) for $N=1$. Note that

$$\mathbf{G}_1\mathbf{G}_1^H = \mathbf{M}_1\boldsymbol{\Theta}_1\mathbf{M}_0\mathbf{M}_0^H\boldsymbol{\Theta}_1^H\mathbf{M}_1^H \tag{67}$$

therefore

$$\begin{aligned}&S_{\mathbf{G}_1\mathbf{G}_1^H}(z) \\ &= S_{\boldsymbol{\Theta}_1\mathbf{M}_0\mathbf{M}_0^H\boldsymbol{\Theta}_1^H\mathbf{M}_1^H\mathbf{M}_1}(z), \text{ by Lemma 1} \\ &= S_{\boldsymbol{\Theta}_1\mathbf{M}_0\mathbf{M}_0^H\boldsymbol{\Theta}_1^H}(z)S_{\mathbf{M}_1^H\mathbf{M}_1}(z), \text{ by Lemma 2} \\ &= \frac{z+1}{z+\frac{k_0}{k_1}}S_{\mathbf{M}_0\mathbf{M}_0^H\boldsymbol{\Theta}_1^H\boldsymbol{\Theta}_1}\left(\frac{z}{\frac{k_0}{k_1}}\right)S_{\mathbf{M}_1^H\mathbf{M}_1}(z), \text{ by Lemma 1} \\ &= \frac{z+1}{z+\frac{k_0}{k_1}}S_{\mathbf{M}_0\mathbf{M}_0^H}\left(\frac{z}{\frac{k_0}{k_1}}\right)S_{\boldsymbol{\Theta}_1^H\boldsymbol{\Theta}_1}\left(\frac{z}{\frac{k_0}{k_1}}\right)S_{\mathbf{M}_1^H\mathbf{M}_1}(z), \text{ by Lemma 2} \\ &= \frac{z+1}{z+\frac{k_0}{k_1}}S_{\mathbf{M}_0\mathbf{M}_0^H}\left(\frac{z}{\frac{k_0}{k_1}}\right)\frac{1}{a_1}\frac{1}{\frac{z}{\frac{k_0}{k_1}}+\frac{k_1}{k_0}}S_{\mathbf{M}_1^H\mathbf{M}_1}(z), \text{ by Lemma 3} \\ &= S_{\mathbf{M}_1^H\mathbf{M}_1}(z)\frac{\rho_0}{a_1}\frac{1}{z+\rho_0}S_{\mathbf{M}_0^H\mathbf{M}_0}\left(\frac{z}{\rho_0}\right), \text{ by Lemma 1.}\end{aligned} \tag{68}$$

Now, we need to prove that if (66) holds for $N=q$, it also holds for $N=q+1$. Note that

$$\begin{aligned}\mathbf{G}_{q+1}\mathbf{G}_{q+1}^H = \mathbf{M}_{q+1}\boldsymbol{\Theta}_{q+1}\mathbf{M}_q\boldsymbol{\Theta}_q\ldots\mathbf{M}_1\boldsymbol{\Theta}_1\mathbf{M}_0 \\ \times\mathbf{M}_0^H\boldsymbol{\Theta}_1^H\mathbf{M}_1^H\ldots\boldsymbol{\Theta}_q^H\mathbf{M}_q^H\boldsymbol{\Theta}_{q+1}^H\mathbf{M}_{q+1}^H.\end{aligned} \tag{69}$$

Therefore

$$\begin{aligned}S_{\mathbf{G}_{q+1}\mathbf{G}_{q+1}^H}(z) &= S_{\mathbf{M}_{q+1}\ldots\mathbf{M}_{q+1}^H}(z) \\ &= S_{\boldsymbol{\Theta}_{q+1}\mathbf{M}_q\ldots\mathbf{M}_q^H\boldsymbol{\Theta}_{q+1}^H\mathbf{M}_{q+1}^H\mathbf{M}_{q+1}}(z),\end{aligned} \tag{70}$$

by Lemma 1. The empirical eigenvalue distribution of Wishart matrices $\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^H$ converges almost surely to the Marčenko-Pastur law whose support is compact. Moreover, by assumption, the empirical eigenvalue distribution of $\mathbf{M}_i^H\mathbf{M}_i, i=0,\ldots,N+1$ converges to an asymptotic distribution with a compact support. Thus, by Lemma 5, the asymptotic eigenvalue distribution of $\mathbf{M}_q\boldsymbol{\Theta}_q\ldots\boldsymbol{\Theta}_q^H\mathbf{M}_q^H$ has a compact support. Therefore, Lemma 2 can be applied to (70) to show that

$$\begin{aligned}&S_{\mathbf{G}_{q+1}\mathbf{G}_{q+1}^H}(z) \\ &= S_{\boldsymbol{\Theta}_{q+1}\ldots\boldsymbol{\Theta}_{q+1}^H}(z)S_{\mathbf{M}_{q+1}^H\mathbf{M}_{q+1}}(z), \text{ by Lemma 2} \\ &= \frac{z+1}{z+\frac{k_q}{k_{q+1}}}S_{\mathbf{M}_q\ldots\mathbf{M}_q^H\boldsymbol{\Theta}_{q+1}^H\boldsymbol{\Theta}_{q+1}}\left(\frac{z}{\frac{k_q}{k_{q+1}}}\right) \\ &\quad\times S_{\mathbf{M}_{q+1}^H\mathbf{M}_{q+1}}(z), \text{ by Lemma 1} \\ &= \frac{z+1}{z+\frac{k_q}{k_{q+1}}}S_{\mathbf{M}_q\ldots\mathbf{M}_q^H}\left(\frac{z}{\frac{k_q}{k_{q+1}}}\right)S_{\boldsymbol{\Theta}_{q+1}^H\boldsymbol{\Theta}_{q+1}}\left(\frac{z}{\frac{k_q}{k_{q+1}}}\right) \\ &\quad\times S_{\mathbf{M}_{q+1}^H\mathbf{M}_{q+1}}(z), \text{ by Lemma 2} \\ &= \frac{z+1}{z+\frac{k_q}{k_{q+1}}}\left(S_{\mathbf{M}_q^H\mathbf{M}_q}\left(\frac{z}{\frac{k_q}{k_{q+1}}}\right)\right.\end{aligned}$$

$$\times \prod_{i=1}^{q} \frac{\frac{k_{i-1}}{k_q}}{a_i} \frac{1}{\frac{z}{\frac{k_q}{k_{q+1}}} + \frac{k_{i-1}}{k_q}} S_{\mathbf{M}_{i-1}^H \mathbf{M}_{i-1}} \left(\frac{\left(\frac{z}{\frac{k_q}{k_{q+1}}} \right)}{\frac{k_{i-1}}{k_q}} \right)$$

$$\times \frac{1}{a_{q+1}} \frac{1}{\frac{k_{q+1}}{k_q} + \frac{z}{\frac{k_q}{k_{q+1}}}} S_{\mathbf{M}_{q+1}^H \mathbf{M}_{q+1}}(z) \text{ , by Lemma 3}$$

$$= \frac{z+1}{z + \frac{k_q}{k_{q+1}}} S_{\mathbf{M}_{q+1}^H \mathbf{M}_{q+1}}(z) \frac{\frac{k_q}{k_{q+1}}}{a_{q+1}} \frac{1}{z+1} S_{\mathbf{M}_q^H \mathbf{M}_q} \left(\frac{z}{\frac{k_q}{k_{q+1}}} \right)$$

$$\times \prod_{i=1}^{q} \frac{\frac{k_{i-1}}{k_{q+1}}}{a_i} \frac{1}{z + \frac{k_{i-1}}{k_{q+1}}} S_{\mathbf{M}_{i-1}^H \mathbf{M}_{i-1}} \left(\frac{z}{\frac{k_{i-1}}{k_{q+1}}} \right)$$

$$= S_{\mathbf{M}_{q+1}^H \mathbf{M}_{q+1}}(z) \prod_{i=1}^{q+1} \frac{\frac{k_{i-1}}{k_{q+1}}}{a_i} \frac{1}{z + \frac{k_{i-1}}{k_{q+1}}} S_{\mathbf{M}_{i-1}^H \mathbf{M}_{i-1}} \left(\frac{z}{\frac{k_{i-1}}{k_{q+1}}} \right)$$

$$= S_{\mathbf{M}_{q+1}^H \mathbf{M}_{q+1}}(z) \prod_{i=1}^{q+1} \frac{\rho_{i-1}}{a_i} \frac{1}{(z+\rho_{i-1})} S_{\mathbf{M}_{i-1}^H \mathbf{M}_{i-1}} \left(\frac{z}{\rho_{i-1}} \right). \tag{71}$$

The proof is complete. ■

- **Second Step: use $S_{\mathbf{G}_N \mathbf{G}_N^H}(z)$ to find $\Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}(z)$**

Theorem 4: Let us define $a_{N+1} = 1$. We have

$$s\Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}^N(s) = \prod_{i=0}^{N} \frac{\rho_i}{a_{i+1}} \; \Upsilon_{M_i^H M_i}^{-1} \left(\frac{\Upsilon_{\mathbf{G}_N \mathbf{G}_N^H(s)}}{\rho_i} \right). \tag{72}$$

Proof: From (66) it follows that

$$\frac{z}{z+1} S_{\mathbf{G}_N \mathbf{G}_N^H}(z) = \frac{z}{z+1} S_{\mathbf{M}_N^H \mathbf{M}_N}(z) \times \prod_{i=1}^{N} \frac{\rho_{i-1}}{a_i} \frac{1}{z+\rho_{i-1}} \frac{\frac{z}{\rho_{i-1}}+1}{\frac{z}{\rho_{i-1}}} \frac{\frac{z}{\rho_{i-1}}}{\frac{z}{\rho_{i-1}}+1} S_{\mathbf{M}_{i-1}^H \mathbf{M}_{i-1}} \left(\frac{z}{\rho_{i-1}} \right). \tag{73}$$

Using (36) in (73), we obtain

$$\Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}^{-1}(z) = \frac{1}{z^N} \Upsilon_{M_N^H M_N}^{-1}(z) \prod_{i=1}^{N} \frac{\rho_{i-1}}{a_i} \; \Upsilon_{M_{i-1}^H M_{i-1}}^{-1} \left(\frac{z}{\rho_{i-1}} \right) \tag{74}$$

or equivalently

$$\Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}^{-1}(z) = \frac{1}{z^N} \prod_{i=0}^{N} \frac{\rho_i}{a_{i+1}} \; \Upsilon_{M_i^H M_i}^{-1} \left(\frac{z}{\rho_i} \right). \tag{75}$$

Substituting $z = \Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}(s)$ in (75), (72) follows. This completes the proof. ■

- **Third Step: use $\Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}(z)$ to obtain $d\boldsymbol{I}/d\eta$**

Theorem 5: In the asymptotic regime, as $k_0, k_1, \ldots, k_N$ go to infinity while $\frac{k_i}{k_N} \to \rho_i, i = 0, \ldots, N$, the derivative of the instantaneous mutual information is given by

$$\frac{d\boldsymbol{I}_\infty}{d\eta} = \frac{1}{\rho_0 \ln 2} \prod_{i=0}^{N} h_i \tag{76}$$

where $h_0, h_1, \ldots, h_N$ are the solutions to the following set of $N+1$ equations

$$\prod_{j=0}^{N} h_j = \rho_i \mathrm{E} \left[\frac{h_i^N \Lambda_i}{\frac{\rho_i}{a_i+1} + \eta h_i^N \Lambda_i} \right] \qquad \forall i \in \mathbb{N}_0^N. \tag{77}$$

The expectation in (77) is over Λ_i whose probability distribution function is given by $F_{\mathbf{M}_i^H \mathbf{M}_i}(\lambda)$ (convention: $a_{N+1} = 1$).

Proof: First, we note that

$$\begin{aligned} \boldsymbol{I} &= \frac{1}{k_0} \log\det(\mathbf{I} + \eta \mathbf{G}_N \mathbf{G}_N^H) \\ &= \frac{1}{k_0} \sum_{i=1}^{k_N} \log(1 + \eta \lambda_{\mathbf{G}_N \mathbf{G}_N^H}(i)) \\ &= \frac{k_N}{k_0} \frac{1}{k_N} \sum_{i=1}^{k_N} \log(1 + \eta \lambda_{\mathbf{G}_N \mathbf{G}_N^H}(i)) \\ &= \frac{k_N}{k_0} \int \log(1+\eta\lambda) dF_{\mathbf{G}_N \mathbf{G}_N^H}^{k_N}(\lambda) \\ &\xrightarrow{a.s.} \frac{1}{\rho_0} \int \log(1+\eta\lambda) dF_{\mathbf{G}_N \mathbf{G}_N^H}(\lambda) \\ &= \frac{1}{\rho_0 \ln 2} \int \ln(1+\eta\lambda) dF_{\mathbf{G}_N \mathbf{G}_N^H}(\lambda) \end{aligned} \tag{78}$$

where $F_{\mathbf{G}_N \mathbf{G}_N^H}^{k_N}(\lambda)$ is the (nonasymptotic) empirical eigenvalue distribution of $\mathbf{G}_N \mathbf{G}_N^H$, that converges almost-surely to the asymptotic empirical eigenvalue distribution $F_{\mathbf{G}_N \mathbf{G}_N^H}$, whose support is compact. Indeed, the empirical eigenvalue distribution of Wishart matrices $\mathbf{\Theta}_i \mathbf{\Theta}_i^H$ converges almost surely to the Marčenko-Pastur law whose support is compact, and by assumption, for $i \in \mathbb{N}_0^{N+1}$ the empirical eigenvalue distribution of $\mathbf{M}_i^H \mathbf{M}_i$ converges to an asymptotic distribution with a compact support. Therefore, according to Lemma 5, the asymptotic eigenvalue distribution of $\mathbf{G}_N \mathbf{G}_N^H$ has a compact support. The log function is continuous, thus bounded on the compact support of the asymptotic eigenvalue distribution of $\mathbf{G}_N \mathbf{G}_N^H$. This enables the application of the bounded convergence theorem to obtain the almost-sure convergence in (78).

It follows from (78) that

$$\begin{aligned} \frac{d\boldsymbol{I}_\infty}{d\eta} &= \frac{1}{\rho_0 \ln 2} \int \frac{\lambda}{1+\eta\lambda} dF_{\mathbf{G}_N \mathbf{G}_N^H}(\lambda) \\ &= \frac{1}{-\rho_0 \eta \ln 2} \int \frac{-\eta\lambda}{1-(-\eta)\lambda} dF_{\mathbf{G}_N \mathbf{G}_N^H}(\lambda) \\ &= \frac{1}{-\rho_0 \eta \ln 2} \Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}(-\eta). \end{aligned} \tag{79}$$

Let us denote

$$t = \Upsilon_{\mathbf{G}_N \mathbf{G}_N^H}(-\eta) \tag{80}$$

$$g_i = \Upsilon^{-1}_{\mathbf{M}_i^H \mathbf{M}_i}\left(\frac{t}{\rho_i}\right) \qquad \forall i \in \mathbb{N}_0^N \tag{81}$$

and, for the sake of simplicity, let $\alpha = \rho_0 \ln 2$. From (79), we have

$$t = -\eta\alpha \frac{d\boldsymbol{I}_\infty}{d\eta}. \tag{82}$$

Substituting $s = -\eta$ in (72) and using (80) and (81), it follows that

$$-\eta t^N = \prod_{i=0}^{N} \frac{\rho_i}{a_{i+1}} g_i. \tag{83}$$

Finally, from (81) and the very definition of Υ in (35), we obtain

$$t = \rho_i \int \frac{g_i \lambda}{1 - g_i \lambda} dF_{\mathbf{M}_i^H \mathbf{M}_i}(\lambda) \qquad \forall i \in \mathbb{N}_0^N. \tag{84}$$

Substituting (82) in (83) and (84) yields

$$(-\eta)^{N+1}\left(\alpha \frac{d\boldsymbol{I}}{d\eta}\right)^N = \prod_{i=0}^{N} \frac{\rho_i}{a_{i+1}} g_i \tag{85}$$

and

$$-\eta\left(\alpha \frac{d\boldsymbol{I}_\infty}{d\eta}\right) = \rho_i \int \frac{g_i \lambda}{1 - g_i \lambda} dF_{\mathbf{M}_i^H \mathbf{M}_i}(\lambda) \qquad \forall i \in \mathbb{N}_0^N. \tag{86}$$

Letting

$$h_i = \left(\frac{\rho_i}{a_{i+1}}\right)^{\frac{1}{N}} \left(\frac{g_i}{-\eta}\right)^{\frac{1}{N}} \tag{87}$$

it follows from (85) that

$$\alpha \frac{d\boldsymbol{I}_\infty}{d\eta} = \prod_{i=0}^{N} h_i. \tag{88}$$

Using (87) and (88) in (86), we obtain

$$-\eta \prod_{j=0}^{N} h_j = \rho_i \int \frac{-\eta h_i^N \frac{a_{i+1}}{\rho_i}\lambda}{1-(-\eta)h_i^N \frac{a_{i+1}}{\rho_i}\lambda} dF_{\mathbf{M}_i^H \mathbf{M}_i}(\lambda), \quad \forall i \in \mathbb{N}_0^N \tag{89}$$

or, equivalently

$$\begin{aligned}\prod_{j=0}^{N} h_j &= \rho_i \int \frac{h_i^N \lambda}{\frac{\rho_i}{a_{i+1}} + \eta h_i^N \lambda} dF_{\mathbf{M}_i^H \mathbf{M}_i}(\lambda) \\ &= \rho_i \mathrm{E}\left[\frac{h_i^N \Lambda_i}{\frac{\rho_i}{a_{i+1}} + \eta h_i^N \Lambda_i}\right], \quad \forall i \in \mathbb{N}_0^N. \end{aligned} \tag{90}$$

This, along with (88), complete the proof. ∎

- **Fourth Step: integrate $d\boldsymbol{I}/d\eta$ to obtain $\boldsymbol{I}$ itself**

The last step of the proof of Theorem 1 is accomplished by computing the derivative of $\boldsymbol{I}_\infty$ in (17) with respect to η and showing that the derivative matches (76). This shows that (17) is one primitive function of $\frac{d\boldsymbol{I}_\infty}{d\eta}$. Since primitive functions of $\frac{d\boldsymbol{I}_\infty}{d\eta}$ differ by a constant, the constant was chosen such that the mutual information (17) is zero when SNR η goes to zero: $\lim_{\eta \to 0} \boldsymbol{I}_\infty(\eta) = 0$.

We now proceed with computing the derivative of $\boldsymbol{I}_\infty$. If (17) holds, then we have (recall $\alpha = \rho_0 \ln 2$)

$$\alpha \boldsymbol{I}_\infty = \sum_{i=0}^{N} \rho_i \mathrm{E}\left[\ln\left(1 + \frac{\eta a_{i+1}}{\rho_i} h_i^N \Lambda_i\right)\right] - N\eta \prod_{i=0}^{N} h_i. \tag{91}$$

From (91), we have

$$\begin{aligned} &\alpha \frac{d\boldsymbol{I}_\infty}{d\eta} \\ &= \sum_{i=0}^{N} \rho_i \mathrm{E}\left[\frac{\Lambda_i\left(h_i^N + N\eta h_i^{N-1} h_i'\right)}{\frac{\rho_i}{a_{i+1}}(1 + \frac{\eta a_{i+1}}{\rho_i} h_i^N \Lambda_i)}\right] \\ &\quad - N\prod_{i=0}^{N} h_i - N\eta\left(\sum_{i=0}^{N} h_i' \prod_{\substack{j=0 \\ j\neq i}}^{N} h_j\right) \\ &= \sum_{i=0}^{N} \rho_i \mathrm{E}\left[\frac{\Lambda_i h_i^N}{\frac{\rho_i}{a_{i+1}} + \eta h_i^N \Lambda_i}\right] + N\eta \sum_{i=0}^{N} \frac{h_i'}{h_i} \rho_i \mathrm{E}\left[\frac{\Lambda_i h_i^N}{\frac{\rho_i}{a_{i+1}} + \eta h_i^N \Lambda_i}\right] \\ &\quad - N\prod_{i=0}^{N} h_i - N\eta\left(\sum_{i=0}^{N} \frac{h_i'}{h_i} \prod_{j=0}^{N} h_j\right) \\ &= \sum_{i=0}^{N} \prod_{j=0}^{N} h_j + N\eta\left(\sum_{i=0}^{N} \frac{h_i'}{h_i} \prod_{j=0}^{N} h_j\right) \\ &\quad - N\prod_{i=0}^{N} h_i - N\eta\left(\sum_{i=0}^{N} \frac{h_i'}{h_i} \prod_{j=0}^{N} h_j\right) \\ &= (N+1)\prod_{j=0}^{N} h_j - N\prod_{j=0}^{N} h_j = \prod_{j=0}^{N} h_j \end{aligned} \tag{92}$$

where $h_i' \triangleq \frac{dh_i}{d\eta}$ and the third line is due to (18). Equation (76) immediately follows from (92). This completes the proof. ∎

Appendix C
Proof of Theorem 2

In this appendix, we provide the proof of Theorem 2. The proof of this theorem is based on [29, Theorem H.1.h] that is reiterated in Lemma 4. Note that, [29, Theorem H.1.h] has been used before to characterize the source precoder maximizing the average mutual information of single-user [21] and multi-user [22] single-hop MIMO systems with covariance knowledge at source, and to obtain the relay precoder maximizing the instantaneous mutual information of a two-hop MIMO system with full CSI at the relay [9]. We extend the results of [9], [21], [22] to suit the MIMO multihop relaying system of our concern.

The proof consists of three following steps.

- **Step 1:** Use the singular value decomposition (SVD) $\mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^H = \mathbf{\Lambda}_{t,i+1}^{1/2} \mathbf{U}_{t,i+1}^H \mathbf{P}_i \mathbf{U}_{r,i} \mathbf{\Lambda}_{r,i}^{1/2}$ and show that unitary matrices $\mathbf{U}_i$ and $\mathbf{V}_i$ impact the maximization of the average mutual information through the power constraints only, while diagonal matrices $\mathbf{D}_i$ affect both the mutual information expression and the power constraints.

- **Step 2:** Represent the power constraint expression as a function of $\mathbf{D}_i, \mathbf{U}_i, \mathbf{V}_i$ and channel correlation matrices only.
- **Step 3:** Show that the directions minimizing the trace in the power constraint are those given in Theorem 2, regardless of the singular values contained in $\mathbf{D}_i$.

Before detailing each step, we recall that the maximum average mutual information is given by

$$\mathcal{C} \triangleq \max_{\{\mathbf{P}_i/\mathrm{tr}(\mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H]) \le k_i\mathcal{P}_i\}_{i\in\mathbb{N}_0^{N-1}}} \mathrm{E}\left[\log\det(\mathbf{I}_{k_N}+\eta\,\mathbf{G}_N\mathbf{G}_N^H)\right] \tag{93}$$

and we define the conventions $a_0 = 1$, and $\mathbf{C}_{r,0} = \mathbf{I}_{k_0}$. Note that the latter implies that $\mathbf{U}_{r,0} = \mathbf{I}_{k_0}$ and $\mathbf{\Lambda}_{r,0} = \mathbf{I}_{k_0}$.

- **Step 1: clarify how the average mutual information depends on the transmit directions and the transmit powers**

For $i \in \{1,\ldots,N\}$ we define

$$\mathbf{\Theta}'_i = \mathbf{U}_{r,i}^H\mathbf{\Theta}_i\mathbf{U}_{t,i}. \tag{94}$$

Since $\mathbf{\Theta}_i$ is zero-mean i.i.d. complex Gaussian, thus bi-unitarily invariant, and $\mathbf{U}_{r,i}$ and $\mathbf{U}_{t,i}$ are unitary matrices, $\mathbf{\Theta}'_i$ has the same distribution as $\mathbf{\Theta}_i$.

For $i \in \{0,\ldots,N-1\}$, we consider the following SVD

$$\mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^H = \mathbf{\Lambda}_{t,i+1}^{1/2}\mathbf{U}_{t,i+1}^H\mathbf{P}_i\mathbf{U}_{r,i}\mathbf{\Lambda}_{r,i}^{1/2} \tag{95}$$

where $\mathbf{U}_i, \mathbf{V}_i$ are unitary matrices, $\mathbf{D}_i$ is a real diagonal matrix with non-negative diagonal elements in the nonincreasing order of amplitude.

We now rewrite the average mutual information as a function of matrices $\mathbf{U}_i$, $\mathbf{V}_i$ and $\mathbf{D}_i$, in order to take the maximization in (15) over $\mathbf{U}_i$, $\mathbf{V}_i$ and $\mathbf{D}_i$ instead of $\mathbf{P}_i$. Using (94) and (95) the average mutual information $\mathcal{I}$ can be expressed in terms of matrices $\mathbf{\Theta}'_i$, $\mathbf{U}_i$, $\mathbf{V}_i$ and $\mathbf{D}_i$ as

$$\begin{aligned}\mathcal{I} &\triangleq \mathrm{E}\left[\log\det(\mathbf{I}_{k_N}+\eta\,\mathbf{G}_N\mathbf{G}_N^H)\right]\\ &= \mathrm{E}\left[\log\det(\mathbf{I}_{k_N}\right.\\ &\quad + \eta\mathbf{U}_{r,N}\mathbf{\Lambda}_{r,N}^{1/2}\mathbf{\Theta}'_N\mathbf{U}_{N-1}\mathbf{D}_{N-1}\mathbf{V}_{N-1}^H\mathbf{\Theta}'_{N-1}\ldots\\ &\quad \ldots\mathbf{U}_1 D_1\mathbf{V}_1^H\mathbf{\Theta}'_1\mathbf{U}_0\mathbf{D}_0\mathbf{V}_0^H\mathbf{V}_0\mathbf{D}_0^H\mathbf{U}_0^H\mathbf{\Theta}_1'^H\mathbf{V}_1 D_1^H\mathbf{U}_1^H\ldots\\ &\quad \left.\ldots\mathbf{\Theta}_{N-1}'^H\mathbf{V}_{N-1}\mathbf{D}_{N-1}^H\mathbf{U}_{N-1}^H\mathbf{\Theta}_N'^H\mathbf{\Lambda}_{r,N}^{1/2}\mathbf{U}_{r,N}^H)\right]\end{aligned} \tag{96}$$

$\mathbf{\Theta}'_i$ being zero-mean i.i.d. complex Gaussian, multiplying it by unitary matrices does not change its distribution. Therefore, $\mathbf{\Theta}''_i = \mathbf{V}_i^H\mathbf{\Theta}'_i\mathbf{U}_{i-1}$ has the same distribution as $\mathbf{\Theta}'_i$ and the average mutual information can be rewritten

$$\begin{aligned}&\mathcal{I}\\ &= \mathrm{E}\left[\log\det(\mathbf{I}_{k_N}+\eta\,\mathbf{\Lambda}_{r,N}^{1/2}\mathbf{\Theta}''_N\mathbf{D}_{N-1}\mathbf{\Theta}''_{N-1}\ldots\mathbf{D}_1\mathbf{\Theta}''_1\mathbf{D}_0\right.\\ &\quad\left.\times\mathbf{D}_0^H\mathbf{\Theta}_1''^H\mathbf{D}_1^H\ldots\mathbf{\Theta}_{N-1}''^H\mathbf{D}_{N-1}^H\mathbf{\Theta}_N''^H\mathbf{\Lambda}_{r,N}^{1/2})\right]\\ &= \mathrm{E}\left[\log\det(\mathbf{I}_{k_N}+\eta\mathbf{\Lambda}_{r,N}^{1/2}\bigotimes_{i=N}^{1}\{\mathbf{\Theta}''_i\mathbf{D}_{i-1}\}\bigotimes_{i=1}^{N}\{\mathbf{D}_{i-1}^H\mathbf{\Theta}_i''^H\}\mathbf{\Lambda}_{r,N}^{1/2})\right].\end{aligned} \tag{97}$$

Therefore, the maximum average mutual information can then be represented as

$$\begin{aligned}\mathcal{C} =& \max_{\{\mathbf{D}_i,\mathbf{U}_i,\mathbf{V}_i/\mathrm{tr}(\mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H])\le k_i\mathcal{P}_i\}_{i\in\mathbb{N}_0^{N-1}}}\\ &\mathrm{E}\left[\log\det(\mathbf{I}_{k_N}+\eta\mathbf{\Lambda}_{r,N}^{1/2}\bigotimes_{i=N}^{1}\{\mathbf{\Theta}''_i\mathbf{D}_{i-1}\}\bigotimes_{i=1}^{N}\{\mathbf{D}_{i-1}^H\mathbf{\Theta}_i''^H\}\mathbf{\Lambda}_{r,N}^{1/2})\right].\end{aligned} \tag{98}$$

Expression (97) shows that the average mutual information $\mathcal{I}$ does not depend on the matrices $\mathbf{U}_i$ and $\mathbf{V}_i$, which determine the transmit directions at source and relays, but only depends on the singular values contained in matrices $\mathbf{D}_i$. Nevertheless, as shown by (98), the maximum average mutual information $\mathcal{C}$ depends on the matrices $\mathbf{U}_i, \mathbf{V}_i$—and thus on the transmit directions—through the power constraints.

- **Step 2: give the expression of the power constraints in function of $\mathbf{D}_i, \mathbf{U}_i, \mathbf{V}_i$ and channel correlation matrices**

We show hereunder that the average power of transmitted signal $\mathbf{x}_i$ at ith relaying level is given by

$$\mathrm{tr}(\mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H]) = a_i\mathrm{tr}(\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H)\prod_{k=0}^{i-1}\frac{a_k}{k_k}\mathrm{tr}(\mathbf{C}_{t,k+1}\mathbf{P}_k\mathbf{C}_{r,k}\mathbf{P}_k^H). \tag{99}$$

Proof: The average power of transmitted signal $\mathbf{x}_i$ can be written as

$$\mathrm{tr}(\mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H]) = \mathrm{tr}(\mathrm{E}[\bigotimes_{k=i}^{1}\{\mathbf{A}_k\mathbf{\Theta}_k\}\mathbf{A}_0\mathbf{A}_0^H\bigotimes_{k=1}^{i}\{\mathbf{\Theta}_k^H\mathbf{A}_k^H\}])$$

with

$$\begin{aligned}\mathbf{A}_i &= \mathbf{P}_i\mathbf{C}_{r,i}^{1/2}\\ \mathbf{A}_k &= \mathbf{M}_k = \mathbf{C}_{t,k+1}^{1/2}\mathbf{P}_k\mathbf{C}_{r,k}^{1/2} \qquad \forall k\in\mathbb{N}_0^{i-1}\\ \sigma_k^2 &= \frac{a_k}{k_{k-1}}.\end{aligned} \tag{100}$$

Applying Lemma 7 to $\mathrm{tr}(\mathrm{E}\{\mathbf{x}_i\mathbf{x}_i^H\})$ yields

$$\begin{aligned}&\mathrm{tr}(\mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H])\\ &= \mathrm{tr}(\mathbf{C}_{t,1}\mathbf{P}_0\mathbf{C}_{r,0}\mathbf{P}_0^H)\prod_{k=1}^{i-1}\frac{a_k}{k_{k-1}}\mathrm{tr}(\mathbf{C}_{t,k+1}\mathbf{P}_k\mathbf{C}_{r,k}\mathbf{P}_k^H)\\ &\quad\times\frac{a_i}{k_{i-1}}\mathrm{tr}(\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H)\\ &= a_i\mathrm{tr}(\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H)\prod_{k=0}^{i-1}\frac{a_k}{k_k}\mathrm{tr}(\mathbf{C}_{t,k+1}\mathbf{P}_k\mathbf{C}_{r,k}\mathbf{P}_k^H)\end{aligned} \tag{101}$$

which concludes the proof. ■

Using (99) in the power constraints (7), those constraints can be rewritten as a product of trace-factors as in (19). In order to express (19) in function of matrices $\mathbf{U}_i$, $\mathbf{V}_i$ and $\mathbf{D}_i$, we first rewrite (95) as

$$\mathbf{P}_i = \mathbf{U}_{t,i+1}\mathbf{\Lambda}_{t,i+1}^{-1/2}\mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^H\mathbf{\Lambda}_{r,i}^{-1/2}\mathbf{U}_{r,i}^H \tag{102}$$

and use (102) in (19) to obtain

$$\begin{aligned}&\operatorname{tr}(\mathbf{P}_i\mathbf{C}_{r,i}\mathbf{P}_i^H)\\&=\operatorname{tr}(\mathbf{U}_{t,i+1}\mathbf{\Lambda}_{t,i+1}^{-1/2}\mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^H\mathbf{\Lambda}_{r,i}^{-1/2}\mathbf{U}_{r,i}^H\quad\mathbf{U}_{r,i}\mathbf{\Lambda}_{r,i}\mathbf{U}_{r,i}^H\\&\quad\times\mathbf{U}_{r,i}\mathbf{\Lambda}_{r,i}^{-1/2}\mathbf{V}_i\mathbf{D}_i^H\mathbf{U}_i^H\mathbf{\Lambda}_{t,i+1}^{-1/2}\mathbf{U}_{t,i+1}^H)\\&=\operatorname{tr}(\mathbf{\Lambda}_{t,i+1}^{-1}\mathbf{U}_i\mathbf{D}_i^2\mathbf{U}_i^H)\end{aligned}\tag{103}$$

and

$$\operatorname{tr}(\mathbf{C}_{t,k+1}\mathbf{P}_k\mathbf{C}_{r,k}\mathbf{P}_k^H)=\operatorname{tr}(\mathbf{D}_k\mathbf{D}_k^H)=\operatorname{tr}(\mathbf{D}_k^2)\tag{104}$$

where $\mathbf{D}_i^2=\mathbf{D}_i\mathbf{D}_i^H$ is a real diagonal matrix with non-negative diagonal elements in nonincreasing order. This leads to the following expression of the power constraints in function of $\mathbf{U}_i,\mathbf{D}_i$

$$\begin{aligned}\operatorname{tr}(\mathbf{\Lambda}_{t,1}^{-1}\mathbf{U}_0\mathbf{D}_0^2\mathbf{U}_0^H)&\le k_0\mathcal{P}_0\\a_i\operatorname{tr}(\mathbf{\Lambda}_{t,i+1}^{-1}\mathbf{U}_i\mathbf{D}_i^2\mathbf{U}_i^H)&\le\frac{k_i\mathcal{P}_i}{\prod_{k=0}^{i-1}\frac{a_k}{k_k}\operatorname{tr}(\mathbf{D}_k^2)},\ \forall i\in\mathbb{N}_2^{N-1}.\end{aligned}\tag{105}$$

It was shown in Step 1 that matrices $\mathbf{V}_i$ do not have an impact on the expression of the average mutual information $\mathcal{I}$ (97), and surprisingly (105) now shows that matrices $\mathbf{V}_i$ do not have an impact on the power constraints either. In fact, as can be observed from (105), the power constraints depend only on matrices $\mathbf{U}_i$ and $\mathbf{D}_i$. It should also be noticed that matrix $\mathbf{U}_i$ has an impact on the power constraint of the ith relay only.

- **Step 3: give the optimal transmit directions**

To determine the optimal directions of transmission at source, we apply Lemma 4 to the source power constraint (105) $\operatorname{tr}(\mathbf{\Lambda}_{t,1}^{-1}\mathbf{U}_0\mathbf{D}_0^2\mathbf{U}_0^H)\le k_0\mathcal{P}_0$, and conclude that for all choices of diagonal elements of $\mathbf{D}_0^2$, the matrix $\mathbf{U}_0$ that minimizes the trace $\operatorname{tr}(\mathbf{\Lambda}_{t,1}^{-1}\mathbf{U}_0\mathbf{D}_0^2\mathbf{U}_0^H)$ is $\mathbf{U}_0=I_{k_0}$. Therefore, the source precoder becomes

$$\begin{aligned}\mathbf{P}_0&=\mathbf{U}_{t,1}\mathbf{\Lambda}_{t,1}^{-1/2}\mathbf{D}_0\mathbf{V}_0^H\mathbf{\Lambda}_{r,0}^{-1/2}\mathbf{U}_{r,0}^H=\mathbf{U}_{t,1}\mathbf{\Lambda}_{t,1}^{-1/2}\mathbf{D}_0\mathbf{V}_0^H\\&=\mathbf{U}_{t,1}\mathbf{\Lambda}_{P_0}\mathbf{V}_0^H.\end{aligned}\tag{106}$$

This recalls the known result (22) in the single-hop MIMO case, where the optimal precoding covariance matrix at source was shown [21], [22] to be

$$\mathbf{Q}^\star\triangleq\mathrm{E}[\mathbf{x}_0\mathbf{x}_0^H]=\mathbf{P}_0\mathbf{P}_0^H=\mathbf{U}_{t,1}\mathbf{\Lambda}_{\mathbf{Q}^\star}\mathbf{U}_{t,1}^H.\tag{107}$$

Similarly, to determine the optimal direction of transmission at ith relaying level, we apply Lemma 4 to the ith power constraint: for all choices of diagonal elements of $\mathbf{D}_i^2$, the matrix $\mathbf{U}_i$ that minimizes the trace $\operatorname{tr}(\mathbf{\Lambda}_{t,i+1}^{-1}\mathbf{U}_i\mathbf{D}_i^2\mathbf{U}_i^H)$ is $\mathbf{U}_i=I_{k_i}$. This leads to the precoding matrix at level i

$$\mathbf{P}_i=\mathbf{U}_{t,i+1}\mathbf{\Lambda}_{t,i+1}^{-1/2}\mathbf{D}_i\mathbf{V}_i^H\mathbf{\Lambda}_{r,i}^{-1/2}\mathbf{U}_{r,i}^H.\tag{108}$$

Now since matrices $\mathbf{V}_i, i\in\{0,\dots,N-1\}$ have an impact neither on the expression of the average mutual information nor on the power constraints, they can be chosen to be equal to identity: $\mathbf{V}_i=\mathbf{I}, i\in\{0,\dots,N-1\}$. This leads to the (nonunique but simple) optimal precoding matrices

$$\begin{aligned}\mathbf{P}_0&=\mathbf{U}_{t,1}\mathbf{\Lambda}_{P_0}\\\mathbf{P}_i&=\mathbf{U}_{t,i+1}\mathbf{\Lambda}_{P_i}\mathbf{U}_{r,i}^H\end{aligned}\tag{109}$$

with the diagonal matrices $\mathbf{\Lambda}_{P_i}=\mathbf{\Lambda}_{t,i+1}^{-1/2}\mathbf{D}_i\mathbf{\Lambda}_{r,i}^{-1/2}$ containing the singular values of $\mathbf{P}_i$.

This completes the proof of Theorem 2. ■

Appendix D
Proof of the Asymptotic Mutual Information With Exponential Correlations

In this appendix, we provide the proof of the asymptotic mutual information (28), obtained under the assumptions of exponential channel correlations, precoding matrices with singular vectors as in Theorem 2, and optimal power allocation over these directions.

A) Optimal Precoding Directions: For $i\in\mathbb{N}_1^N$, the eigenvalue decompositions of channel correlation matrices $\mathbf{C}_{t,i}$ and $\mathbf{C}_{r,i}$ can be written as

$$\begin{aligned}\mathbf{C}_{t,i}&=\mathbf{U}_{t,i}\mathbf{\Lambda}_{t,i}\mathbf{U}_{t,i}^H\\\mathbf{C}_{r,i}&=\mathbf{U}_{r,i}\mathbf{\Lambda}_{r,i}\mathbf{U}_{r,i}^H\end{aligned}\tag{110}$$

where $\mathbf{U}_{t,i}$ and $\mathbf{U}_{r,i}$ are unitary, and $\mathbf{\Lambda}_{t,i}$ and $\mathbf{\Lambda}_{r,i}$ are diagonal with their respective eigenvalues ordered in decreasing order. Following Theorem 2, we consider precoding matrices of the form $\mathbf{P}_i=\mathbf{U}_{t,i+1}\mathbf{\Lambda}_{P_i}\mathbf{U}_{r,i}^H$, i.e., the singular vectors of $\mathbf{P}_i$ are optimally aligned to the eigenvectors of channel correlation matrices. Consequently, we can rewrite matrices $\mathbf{M}_i^H\mathbf{M}_i$ (10) as

$$\begin{aligned}\mathbf{M}_0^H\mathbf{M}_0&=\mathbf{U}_{t,1}^H\mathbf{\Lambda}_{P_0}^2\mathbf{\Lambda}_{t,1}\mathbf{U}_{t,1}\\\mathbf{M}_i^H\mathbf{M}_i&=\mathbf{U}_{r,i}^H\mathbf{\Lambda}_{r,i}\mathbf{\Lambda}_{P_i}^2\mathbf{\Lambda}_{t,i+1}\mathbf{U}_{r,i}\qquad\forall i\in\mathbb{N}_1^{N-1}\\\mathbf{M}_N^H\mathbf{M}_N&=\mathbf{U}_{r,N}^H\mathbf{\Lambda}_{r,N}\mathbf{U}_{r,N}.\end{aligned}\tag{111}$$

Thus, the eigenvalues of matrices $\mathbf{M}_i^H\mathbf{M}_i$ are contained in the following diagonal matrices

$$\begin{aligned}\mathbf{\Lambda}_0&=\mathbf{\Lambda}_{P_0}^2\mathbf{\Lambda}_{t,1}\\\mathbf{\Lambda}_i&=\mathbf{\Lambda}_{r,i}\mathbf{\Lambda}_{P_i}^2\mathbf{\Lambda}_{t,i+1}\qquad\forall i\in\mathbb{N}_1^{N-1}\\\mathbf{\Lambda}_N&=\mathbf{\Lambda}_{r,N}.\end{aligned}\tag{112}$$

The asymptotic mutual information, given by (17) and (18), involves expectations of functions of Λ_i whose distribution is given by the asymptotic eigenvalue distribution $F_{\mathbf{M}_i^H\mathbf{M}_i}(\lambda)$ of $\mathbf{M}_i^H\mathbf{M}_i$. Equation (112) shows that a function $g_1(\Lambda_i)$ can be written as a function $g_2(\Lambda_{P_i}^2,\Lambda_{r,i},\Lambda_{t,i+1})$, where the variables $\Lambda_{P_i}^2,\Lambda_{r,i}$, and $\Lambda_{t,i+1}$ are characterized by the asymptotic eigenvalue distributions $F_{\mathbf{P}_i^H\mathbf{P}_i}(\lambda)$, $F_{\mathbf{C}_{r,i}}(\lambda)$, and $F_{\mathbf{C}_{t,i+1}}(\lambda)$ of matrices $\mathbf{P}_i^H\mathbf{P}_i$, $\mathbf{C}_{r,i}$ and $\mathbf{C}_{t,i+1}$, respectively. Therefore, expectations in (17) and (18) can be computed using the asymptotic joint distribution of $(\Lambda_{P_i}^2,\Lambda_{r,i},\Lambda_{t,i+1})$ instead of the distribution $F_{\mathbf{M}_i^H\mathbf{M}_i}(\lambda)$. To simplify notations, we rename the variables as follows:

$$X=\Lambda_{P_i}^2\qquad Y=\Lambda_{r,i}\qquad Z=\Lambda_{t,i+1}.\tag{113}$$

$$\boldsymbol{I}_\infty = \sum_{i=0}^{N} \frac{\rho_i}{\rho_0 (2\pi)^2} \int_{\lambda=0}^{2\pi} \int_{\nu=0}^{2\pi} \log\left(1 + h_i^N \frac{\eta a_{i+1} \alpha_i^2 (1-r_{r,i}^2)(1-r_{t,i+1}^2)}{\rho_i |1-r_{r,i}e^{j\lambda}|^2 |1-r_{t,i+1}e^{j\nu}|^2}\right) d\lambda\, d\nu - N \frac{\log e}{\rho_0} \eta \prod_{i=0}^{N} h_i \tag{120}$$

$$\prod_{j=0}^{N} h_j = \frac{\rho_i}{(2\pi)^2} \int_{\lambda=0}^{2\pi} \int_{\nu=0}^{2\pi} \frac{h_i^N a_{i+1} \alpha_i^2 (1-r_{r,i}^2)(1-r_{t,i+1}^2)}{\rho_i |1-r_{r,i}e^{j\lambda}|^2 |1-r_{t,i+1}e^{j\nu}|^2 + \eta h_i^N a_{i+1} \alpha_i^2 (1-r_{r,i}^2)(1-r_{t,i+1}^2)}\, d\lambda\, d\nu\ ,\ \forall i \in \mathbb{N}_0^N \tag{121}$$

Then, the expectation of a function $g_1(\Lambda_i)$ can be written

$$\begin{aligned}
&\mathrm{E}[g_1(\Lambda_i)] \\
&= \mathrm{E}[g_2(X,Y,Z)] \\
&= \int_z \int_y \int_x g_2(x,y,z) f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz \\
&= \int_z \int_y \int_x g_2(x,y,z) f_{X|Y,Z}(x|y,z) f_{Y|Z}(y|z) f_Z(z) dx\, dy\, dz.
\end{aligned} \tag{114}$$

B) Exponential Correlation Model: So far, general correlation matrices were considered. We now introduce the exponential correlation model (26) and further develop (114) for the distributions $f_{Y|Z}(y|z)$ and $f_Z(z)$ resulting from that particular correlation model.

As k_i grows large, the sequence of Toeplitz matrices $\mathbf{C}_{r,i}$ of size $k_i \times k_i$, defined in (26), is fully characterized by the continuous real function $f_{r,i}$, defined for $\lambda \in [0, 2\pi)$ by [27, Section 4.1]

$$\begin{aligned}
f_{r,i}(\lambda) &= \lim_{k_i \to +\infty} \left(\sum_{k=0}^{k_i-1} r_{r,i}^k e^{jk\lambda} + \sum_{k=-(k_i-1)}^{-1} r_{r,i}^{-k} e^{jk\lambda} \right) \\
&= \frac{1}{1-r_{r,i}e^{j\lambda}} + \frac{r_{r,i}e^{-j\lambda}}{1-r_{r,i}e^{-j\lambda}} \\
&= \frac{1-r_{r,i}^2}{|1-r_{r,i}e^{j\lambda}|^2}.
\end{aligned} \tag{115}$$

We also denote the essential infimum and supremum of $f_{r,i}$ by $m_{f_{r,i}}$ and $M_{f_{r,i}}$, respectively [27, Section 4.1]. In a similar way, we can define the continuous real function $f_{t,i+1}$ characterizing the sequence of Toeplitz matrices $\mathbf{C}_{t,i+1}$ by replacing $r_{r,i}$ in (115) by $r_{t,i+1}$, and we denote by $m_{f_{t,i+1}}$ and $M_{f_{t,i+1}}$ its essential infimum and supremum, respectively.

By the Szegö Theorem [27, Theorem 9], recalled in Lemma 6, for any real function $g(\cdot)$ (resp. $h(\cdot)$) continuous on $[m_{f_{r,i}}, M_{f_{r,i}}]$ (resp. $[m_{f_{t,i+1}}, M_{f_{t,i+1}}]$), we have

$$\begin{aligned}
\int_y g(y) f_Y(y)\, dy &\triangleq \lim_{k_i \to +\infty} \frac{1}{k_i} \sum_{k=1}^{k_i} g\left(\lambda_{\mathbf{C}_{r,i}}(k)\right) \\
&= \frac{1}{2\pi} \int_0^{2\pi} g\left(f_{r,i}(\lambda)\right)\, d\lambda \\
\int_z h(z) f_Z(z)\, dz &\triangleq \lim_{k_i \to +\infty} \frac{1}{k_i} \sum_{k=1}^{k_i} h\left(\lambda_{\mathbf{C}_{t,i+1}}(k)\right) \\
&= \frac{1}{2\pi} \int_0^{2\pi} h\left(f_{t,i+1}(\nu)\right)\, d\nu.
\end{aligned} \tag{116}$$

Assuming that variables $Y = \Lambda_{r,i}$ and $Z = \Lambda_{t,i+1}$ are independent, and applying the Szegö Theorem to (114), we can write

$$\begin{aligned}
&\mathrm{E}[g_1(\Lambda_i)] \\
&= \int_z \int_y \underbrace{\left(\int_x g_2(x,y,z) f_{X|Y,Z}(x|y,z) dx \right)}_{g_3(y,z)} f_Y(y) f_Z(z) dy\, dz \\
&= \int_z \left(\int_y g_3(y,z) f_Y(y)\, dy \right) f_Z(z)\, dz \\
&= \int_z \left(\frac{1}{2\pi} \int_{\lambda=0}^{2\pi} g_3\left(f_{r,i}(\lambda), z\right)\, d\lambda \right) f_Z(z)\, dz\ ,\ \text{by (116)} \\
&= \frac{1}{2\pi} \int_{\lambda=0}^{2\pi} \left(\int_z g_3\left(f_{r,i}(\lambda), z\right)\ f_Z(z)\, dz \right)\, d\lambda \\
&= \frac{1}{(2\pi)^2} \int_{\lambda=0}^{2\pi} \int_{\nu=0}^{2\pi} g_3\left(f_{r,i}(\lambda), f_{t,i+1}(\nu)\right)\, d\lambda\, d\nu\ ,\ \text{by (116)}.
\end{aligned} \tag{117}$$

C) Equal Power Allocation Over Optimal Precoding Directions: We now evaluate (117) in the case of equal power allocation over the optimal directions given in (27). From (27) it can be seen that $X = \Lambda_{P_i}^2 = \alpha_i^2$ is independent from Y and Z, thus $f_{X|Y,Z}(x|y,z) = f_X(x) = \delta(x - \alpha_i^2)$. Consequently

$$g_3(y,z) = \int_x g_2(x,y,z) \delta(x - \alpha_i^2)\, dx = g_2(\alpha_i^2, y, z) \tag{118}$$

and (117) becomes

$$\begin{aligned}
&\mathrm{E}[g_1(\Lambda_i)] = \\
&\frac{1}{(2\pi)^2} \int_{\lambda=0}^{2\pi} \int_{\nu=0}^{2\pi} g_2\left(\alpha_i^2, \frac{1-r_{r,i}^2}{|1-r_{r,i}e^{j\lambda}|^2}, \frac{1-r_{t,i+1}^2}{|1-r_{t,i+1}e^{j\nu}|^2}\right) d\lambda\, d\nu.
\end{aligned} \tag{119}$$

D) Asymptotic Mutual Information: Using (119) in (17) with $g_2(x,y,z) = \log\left(1 + \eta \frac{a_{i+1}}{\rho_i} h_i^N xyz\right)$ gives the expression of the asymptotic mutual information (120) [see (120) and (121) at the top of the page], where $h_0, h_1, \ldots, h_N$ are the solutions of the system of $N+1$ (121), obtained by using (119) in (18) with $g_2(x,y,z) = \frac{h_i^N \Lambda_i xyz}{\frac{\rho_i}{a_{i+1}} + \eta h_i^N xyz}$ (with the convention $r_{r,0} = r_{t,N+1} = 0$). Applying the changes of variables

$$\begin{aligned}
t &= \tan\left(\frac{\lambda}{2}\right),\ \text{thus}\ \cos(\lambda) = \frac{1-t^2}{1+t^2}\ \text{ and }\ d\lambda = \frac{2du}{1+t^2} \\
u &= \tan\left(\frac{\nu}{2}\right),\ \text{thus}\ \cos(\nu) = \frac{1-u^2}{1+u^2}\ \text{ and }\ d\nu = \frac{2du}{1+u^2}
\end{aligned} \tag{122}$$

and performing some algebraic manipulations that are skipped for the sake of conciseness, (120) and (121) can be rewritten as in (28). This concludes the proof.

ACKNOWLEDGMENT

The authors would like to thank B.H. Khan for his valuable help.

REFERENCES

[1] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 29–43, Jan. 2005.

[2] H. Bölcskei, R. Nabar, O. Oyman, and A. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[3] V. Morgenshtern and H. Bölcskei, "Capacity of large amplify-and-forward relay networks," presented at the IEEE Communication Theory Workshop, May 2006.

[4] V. Morgenshtern and H. Bölcskei, "Large random matrix analysis of relay networks," presented at the Allerton Conf., 2006.

[5] V. Morgenshtern and H. Bölcskei, "Crystallization in large wireless networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3319–3349, Oct. 2007.

[6] H. Li, Z. Han, and H. Poor, "Asymptotic analysis of large cooperative relay networks using random matrix theory," *EURASIP J. Advances in Signal Processing*, Feb. 2008.

[7] R. Vaze and R. W. J. Heath, "Capacity scaling for MIMO two-way relaying," *IEEE Trans. Inf. Theory*, submitted for publication.

[8] L. Cottatellucci, T. Chan, and N. Fawaz, "Large system design and analysis of protocols for decode-forward relay networks," presented at the ICST WiOpt/PhysComNet, Berlin, Germany, Apr. 2008.

[9] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398–1407, Apr. 2007.

[10] R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2086–2091, Jul. 2002.

[11] S. Borade, L. Zheng, and R. Gallager, "Amplify-and-forward in wireless relay networks: Rate, diversity, and network size," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3302–3318, Oct. 2007.

[12] S. Yang and J.-C. Belfiore, "Diversity of MIMO multihop relay channels," *IEEE Trans. Inf. Theory*, submitted for publication.

[13] S. Yeh and O. Leveque, "Asymptotic capacity of multi-level amplify-and-forward relay networks," presented at the IEEE ISIT, Jun. 2007.

[14] N. Fawaz, K. Zarifi, M. Debbah, and D. Gesbert, "Asymptotic capacity and optimal precoding strategy of multi-level precode & forward in correlated channels," presented at the IEEE Information Theory Workshop, Porto, Portugal, May 2008.

[15] A.Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[16] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, 1999.

[17] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.

[18] I. Maric and R. Yates, "Bandwidth and power allocation for cooperative strategies in Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1880–1889, Apr. 2010.

[19] N. Fawaz and M. Médard, "On the non-coherent wideband multipath fading relay channel," presented at the IEEE Int. Symp. Information Theory, Austin, TX, Jun. 2010.

[20] I. Marić, A. Goldsmith, and M. Médard, "Analog network coding in the high-SNR regime," presented at the IEEE WiNC, Jun. 2010.

[21] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[22] A. Soysal and S. Ulukus, "Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 1402–1412, Sep. 2007.

[23] A. Tulino and S. Verdu, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*. Boston, MA: NOW, 2004, vol. 1.

[24] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, pp. 1350–1359, Sep. 2001.

[25] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1350–1359, Jul. 2004.

[26] C. Oestges, B. Clerckx, M. Guillaud, and M. Debbah, "Dual-polarized wireless communications: From propagation models to system performance evaluation," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 4019–4031, Oct. 2008.

[27] R. Gray, "Toeplitz and circulant matrices: A review," in *Foundations and Trends in Communications and Information Theory*. Boston, MA: NOW, 2006, vol. 2.

[28] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. Providence, RI: American Mathematical Society, 2000.

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

[30] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, Nov. 1995.

Nadia Fawaz (S'07–A'08–M'10) received the Diplôme d'ingénieur (M.Sc.) in 2005 and the Ph.D. degree in 2008, both in electrical engineering, from the École Nationale Supérieure des Télécommunications de Paris (ENST) and EURECOM, France.

She is currently a postdoctoral researcher in the Research Laboratory of Electronics (RLE) at the Massachusetts Institute of Technology (MIT), Cambridge. Prior to that, she worked as an engineering intern at Bouygues Telecom, France, in 2004, as a research intern in the Laboratoire d'Electronique et Technologie de l'Information (LETI) at the Commissariatà l'Energie Atomique (CEA), France, in 2005, as a research intern at Samsung Advanced Institute of Technology (SAIT), South Korea, in 2006, and as a visiting graduate student at MIT in 2007. Her current research interests include information theory and random matrix theory applied to wireless communications, in particular, relay networks.

Dr. Fawaz received the graduate and postgraduate fellowships from the Délégation Généraleà l'Armement (DGA), France, was awarded the HITACHI prize for the best Master thesis in 2005, and the DGA prize—jointly awarded by the French Ministries of Defense and of Education and Research—for the best Ph.D. thesis in 2010.

Keyvan Zarifi (S'04–M'08) received the Ph.D. degree (with the highest honors) in electrical and computer engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2007.

From January 2002 until March 2005, he was with the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. From April 2005 until September 2007, he was with the Darmstadt University of Technology. From September 2002 until March 2003, he was a visiting scholar at the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. Since September 2007, he has been jointly with Institut National de la Recherche Scientifique-Énergie, Matériaux, et Télécommunications (INRS-EMT), Université du Québec, and Concordia University, Montreal, QC, Canada, as a postdoctoral Fellow. His research interests include statistical signal processing, wireless sensor networks, MIMO and cooperative communications, and blind estimation and detection techniques.

Dr. Zarifi received the Postdoctoral Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) in 2008.

Mérouane Debbah (S'01–A'03–M'04–SM'08) was born in Madrid, Spain. He entered the École Normale Supérieure de Cachan (France) in 1996 where he received the M.Sc and Ph.D. degrees in 1999 and 2002, respectively.

From 1999 to 2002, he worked for Motorola Labs on wireless local area networks and prospective fourth generation systems. From 2002 until 2003, he was appointed Senior Researcher at the Vienna Research Center for Telecommunications (FTW) (Vienna, Austria) working on MIMO wireless channel modeling issues. From 2003 until 2007, he joined the Mobile Communications department of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. He is presently a Professor at SUPELEC (Gif-sur-Yvette, France), holder of the Alcatel-Lucent Chair on Flexible Radio. His research interests are in information theory, signal processing and wireless communications.

Dr. Debbah is the recipient of the "Mario Boella" prize award in 2005, the 2007 General Symposium IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award, as well as the Valuetools 2007, Valuetools 2008, and CrownCom2009 best student paper awards. He is a WWRF fellow.

David Gesbert (S'96–M'99–SM'06–F'11) received the Ph.D. degree from the École Nationale Supérieure des Télécommunications, France, in 1997.

He is currently a Professor in the Mobile Communications Department, EURECOM, France, where he heads the Communications Theory Group. From 1997 to 1999, he was with the Information Systems Laboratory, Stanford University, Stanford, CA. In 1999, he was a founding engineer of Iospan Wireless, Inc., San Jose, CA, a startup company pioneering MIMO-OFDM (now Intel). Between 2001 and 2003, he was with the Department of Informatics, University of Oslo, Norway, as an adjunct professor. He has published about 170 papers and several patents all in the area of signal processing, communications, and wireless networks. He coauthored the book *Space Time Wireless Communications: From Parameter Estimation to MIMO Systems* (Cambridge, 2006).

Dr. Gesbert was a co-editor of several special issues on wireless networks and communications theory for *JSAC* (2003, 2007, and 2009), the *EURASIP Journal on Applied Signal Processing* (2004 and 2007), and *Wireless Communications Magazine* (2006). He served on the IEEE Signal Processing for Communications Technical Committee (2003–2008). He's an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the *EURASIP Journal on Wireless Communications and Networking*. He authored or coauthored papers winning the 2004 IEEE Best Tutorial Paper Award (Communications Society) for a 2003 JSAC paper on MIMO systems, the 2005 Best Paper (Young Author) Award for Signal Processing Society journals, and the Best Paper Award for the 2004 ACM MSWiM workshop.