

Crystallization in Large Wireless Networks

Veniamin I. Morgenshtern and Helmut Bölcskei, *Senior Member, IEEE*

Abstract—We analyze fading interference relay networks where M single-antenna source–destination terminal pairs communicate concurrently and in the same frequency band through a set of K single-antenna relays using half-duplex two-hop relaying. Assuming that the relays have channel state information (CSI), it is shown that in the large- M limit, provided K grows fast enough as a function of M , the network “decouples” in the sense that the individual source–destination terminal pair capacities are strictly positive. The corresponding required rate of growth of K as a function of M is found to be sufficient to also make the individual source–destination fading links converge to nonfading links. We say that the network “crystallizes” as it breaks up into a set of effectively isolated “wires in the air.” A large-deviations analysis is performed to characterize the “crystallization” rate, i.e., the rate (as a function of M , K) at which the decoupled links converge to nonfading links. In the course of this analysis, we develop a new technique for characterizing the large-deviations behavior of certain sums of dependent random variables. For the case of no CSI at the relay level, assuming amplify-and-forward relaying, we compute the per source–destination terminal pair capacity for M , $K \rightarrow \infty$, with $K/M \rightarrow \beta$ fixed, using tools from large random matrix theory.

Index Terms—Amplify-and-forward, capacity scaling, crystallization, distributed orthogonalization, interference relay network, large-deviations theory, large random matrices, large wireless networks.

I. INTRODUCTION

THE capacity of the relay channel [3], [4] is still unknown in the general case. Recently, the problem has attracted significant attention, with progress being made on several aspects [5]. Sparked by [6], [7], analysis of the capacity¹ scaling behavior of large wireless (relay) networks has emerged as an interesting tool [8]–[14], [2], [1], which often allows to make stronger statements than a finite-number-of-nodes analysis. In parallel, the design of distributed space–time codes [15]–[17], the area of network coding [18], [19], and the understanding of the impact of relaying protocols and multiple-antenna terminals on network capacity [17], [20], [21] have seen remarkable activity.

Manuscript received August 24, 2006; revised January 16, 2007. This work was supported by Nokia Research Center, Helsinki, Finland and by the STREP project No. IST-027310 (MEMBRANE) within the Sixth Framework Programme of the European Commission. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005 and at the 43rd Annual Allerton Conference on Communications, Control and Computing, Monticello, IL, October 2005.

The authors are with the Communication Technology Laboratory, ETH Zurich, 8092 Zurich, Switzerland (e-mail: vmorgens@nari.ee.ethz.ch; boelcskei@nari.ee.ethz.ch).

Communicated by M. Gastpar, Guest Editor for the Special Issue on Relaying and Cooperation.

Digital Object Identifier 10.1109/TIT.2007.904789

¹Throughout the paper, when we talk about capacity, we mean the capacity induced by the considered protocols, not the capacity of the network itself.

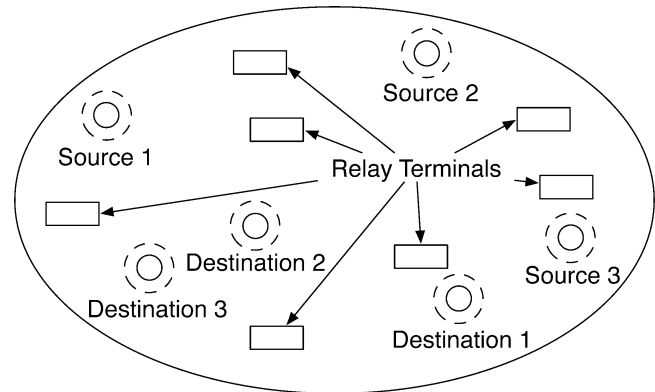


Fig. 1. Dense wireless interference relay network with dead zones around source and destination terminals. Each terminal employs one antenna.

This paper deals with interference fading relay networks where M single-antenna source–destination terminal pairs communicate concurrently and in the same frequency band through half-duplex two-hop relaying over a common set of K single-antenna relay terminals (see Fig. 1). Two setups are considered: i) the *coherent* case, where the relays have channel state information (CSI), perform matched filtering, and the destination terminals cannot cooperate, and ii) the *noncoherent* case, where the relays do not have CSI, perform amplify-and-forward (AF) relaying, and the destination terminals can cooperate. In the coherent case, the network operates in a completely distributed fashion, i.e., with no cooperation between any of the terminals whereas in the noncoherent case the destination terminals can cooperate and perform joint decoding.

A. Contributions and Relation to Previous Work

Our main contributions for the coherent case can be summarized as follows.

- We consider two different protocols, P1 introduced (for the finite- M case) in [1] and P2 introduced in [2]. P1 relies on the idea of relay partitioning (i.e., each relay is assigned to one source–destination terminal pair) and requires each relay terminal to know its assigned backward (source-to-relay) and forward (relay-to-destination) channel only. The relays perform matched filtering with respect to (w.r.t.) their assigned backward and forward channels. P2 does not use relay partitioning, requires each relay terminal to know all M backward and all M forward channels, and performs matched filtering w.r.t. all M backward and M forward links.

Previous work for the coherent case has established the power efficiency scaling of P2 for $M \rightarrow \infty$ with $K = M^2$ [2]; in [1] it was shown that for P1 with M fixed, in the $K \rightarrow \infty$ limit, network capacity scales as $C =$

$(M/2)\log(K) + O(1)$. The results in [1] and the corresponding proof techniques, however, rely heavily on M being fixed when $K \rightarrow \infty$. When $M, K \rightarrow \infty$, the amount of interference (at each destination terminal) grows with M . Establishing the corresponding network capacity scaling behavior, therefore, requires fundamentally new techniques, which are developed in this paper. In particular, we derive the network (ergodic) capacity scaling behavior for $M, K \rightarrow \infty$ for P1 and P2 by computing a lower and an upper bound on the per source–destination terminal pair capacity, and by showing that the bounds exhibit the same scaling (in M, K) behavior. The technique used to establish the lower bound is based on a result found in a completely different context in [22] and applied in [2] to derive the power efficiency scaling of P2. For our purposes, we need a slight generalization of the result in [22], which follows, in a straightforward fashion, from a result on nearest neighbor decoding reported in [23]. For the sake of completeness, we state, in Appendix E, the relevant inequality in the form needed in the context of this paper. The matching upper bound on the per source–destination terminal pair capacity poses significantly more technical challenges and is based on a large-deviations analysis of the individual link signal to interference plus noise ratio (SINR) random variables (RVs). In summary, we prove that in the large- M limit, provided the number of relay terminals K grows fast enough as a function of M , under both protocols P1 and P2 the network “decouples” in the sense that the individual source–destination terminal pair (ergodic) capacities are strictly positive. The corresponding minimum rates of growth are $K \propto M^3$ for P1 and $K \propto M^2$ for P2, with the per source–destination terminal pair capacity scaling (for $M, K \rightarrow \infty$) given by

$$C_{P1} = (1/2)\log(1 + \Theta(K/M^3))$$

and

$$C_{P2} = (1/2)\log(1 + \Theta(K/M^2))$$

respectively. The protocols P1 and P2 thus trade off CSI at the relays for the required (for the network to decouple) rate of growth of the number of relays. We hasten to add that an ergodic-capacity lower bound for P2 was previously established in [2]; this bound is restated (and reproved under slightly different assumptions) in this paper for the sake of completeness. It appears, however, that [2] does not establish the minimum rate of growth of the number of relays for the network to decouple.

- We analyze the network outage capacity behavior induced by P1 and P2 using a large-deviations approach. More specifically, we show that the growth rates $K \propto M^3$ in P1 and $K \propto M^2$ in P2 are sufficient to not only make the network decouple, but also to make the individual source–destination fading links converge to nonfading links. We say that the network “crystallizes” as it breaks up into a set of effectively isolated “wires in the air.” Each of the decoupled links experiences distributed spatial diversity (or relay diversity), with the corresponding diversity order going to infinity as $M \rightarrow \infty$. Consequently, in the large- M limit,

time diversity (achieved by coding over a sufficiently long time horizon) is not needed to achieve ergodic capacity. We obtain bounds on the outage capacity of the individual source–destination links, which allow to characterize the “crystallization” rate (more precisely a guaranteed “crystallization” rate as we do not know whether our bounds are tight), i.e., the rate (as a function of M, K) at which the decoupled links converge to nonfading links. In the course of this analysis, we develop a new technique for characterizing the large-deviations behavior of certain sums of dependent RVs. This technique builds on the well-known truncation approach and is reported in Appendix A.

- For P1 and P2, we establish the impact of cooperation at the relay level on network (ergodic) capacity scaling. More specifically, it is shown that, asymptotically in M and K , cooperation (realized by vector matched filtering) in groups of L relays leads to an L -fold reduction in the total number of relays needed to achieve a given per source–destination terminal pair capacity.

Previous work for the noncoherent (AF) case [1] demonstrated that for M fixed and $K \rightarrow \infty$, AF relaying turns the fading interference relay network into a fading point-to-point multiple-input multiple-output (MIMO) link, showing that the use of relays as active scatterers can recover spatial multiplexing gain in poor scattering environments. Our main contributions for the noncoherent (AF) case are as follows.

- Like in the coherent case, the proof techniques for the noncoherent (AF) case in [1] rely heavily on M being finite. Building on results reported in [24], we compute the $M, K \rightarrow \infty$ (with $K/M \rightarrow \beta$ fixed) per source–destination terminal pair capacity using tools from large-random-matrix theory [25], [26]. The limiting eigenvalue density function of the effective MIMO channel matrix between the source and destination terminals is characterized in terms of its Stieltjes transform as the unique solution of a fixed-point equation, which can be transformed into a fourth-order equation. Upon solving this fourth-order equation and applying the inverse Stieltjes transform, the remaining steps to computing the limiting eigenvalue density function, and based on that the asymptotic network capacity, need to be carried out numerically. We show that this can be accomplished in a straightforward fashion and provide a corresponding algorithm.
- We show that for $\beta \rightarrow \infty$, the fading AF relay network is turned into a fading point-to-point MIMO link (in a sense to be made precise in Section V), thus establishing the large- M, K analog of the result found previously for the finite- M and $K \rightarrow \infty$ case in [1].

B. Notation

The superscripts $T, H,$ and $*$ stand for transposition, conjugate transpose, and element-wise conjugation, respectively. $|\mathcal{X}|$ is the cardinality of the set \mathcal{X} . $\log(x)$ stands for the logarithm to the base 2, and $\ln(x)$ is the natural logarithm. $I[x] = 1$ if x is true and $I[x] = 0$ if x is false. $\delta[k] = 1$ for $k = 0$ and 0 otherwise. The unit step function $u(x) = 0$ for $x < 0$ and $u(x) = 1$ for $x \geq 0$. \mathbb{E} and Var denote the expectation and variance operator, respectively. $\lceil x \rceil$ stands for the smallest

integer greater than or equal to x . $\arg(x)$ stands for the argument of $x \in \mathbb{C}$. A circularly symmetric zero-mean complex Gaussian RV is an RV $Z = X + jY \sim \mathcal{CN}(0, \sigma^2)$, where X and Y are independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2/2)$. An exponentially distributed RV with parameter λ is a real-valued RV X with probability density function (pdf) given by $f_X(x) = \lambda \exp(-\lambda x)u(x)$. A Rayleigh-distributed RV with parameter α^2 is a real-valued RV X with pdf $f_X(x) = (x/\alpha^2) \exp(-x^2/(2\alpha^2))u(x)$. $\mathcal{U}(a, b)$ denotes the uniform distribution over the interval $[a, b]$. $\delta(x)$ is the Dirac delta distribution. The moment-generating function (MGF) of an RV X is defined as

$$M_X(s) \triangleq \int_{-\infty}^{\infty} e^{sx} f_X(x) dx.$$

$(x)^+ = x$ for $x > 0$ and 0 otherwise. For two functions $f(x)$ and $g(x)$, the notation $f(x) = O(g(x))$ means that $|f(x)/g(x)|$ remains bounded as $x \rightarrow \infty$. We write $g(x) = \Theta(f(x))$ to denote that $f(x) = O(g(x))$ and $g(x) = O(f(x))$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = o(g(x))$ means that $|f(x)/g(x)| \rightarrow 0$ as $x \rightarrow \infty$. Matrices and vectors (both deterministic and random) are denoted by upper and lower case, respectively, boldface letters. The element of a matrix \mathbf{X} in the n th row and m th column and the n th element of a vector \mathbf{x} are denoted as $[\mathbf{X}]_{n,m}$ and $[\mathbf{x}]_n$, respectively. $\lambda_i(\mathbf{X})$, $\lambda_{\min}(\mathbf{X})$, and $\lambda_{\max}(\mathbf{X})$ stand for the i th, the minimum, and the maximum eigenvalue of a matrix \mathbf{X} , respectively. $\mathbf{X} \circ \mathbf{Y}$ is the Hadamard (or element-wise) product of the matrices \mathbf{X} and \mathbf{Y} . $\|\mathbf{x}\|$ denotes the ℓ^2 -norm of the vector \mathbf{x} . $\Re z$ and $\Im z$ designate the real and imaginary part of $z \in \mathbb{C}$, respectively. $\mathbb{C}^+ \triangleq \{z \in \mathbb{C} | \Im z > 0\}$. For any $n, m \in \mathbb{N}$, $m \geq n$, $[n : m]$ denotes the natural numbers $\{n, n+1, \dots, m\}$.

C. Organization of the Paper

The rest of this paper is organized as follows. Section II describes the general channel model and the parts of the signal model² that pertain to both the coherent and the noncoherent case. Sections III and IV focus on the coherent case exclusively: Section III contains the large-deviations analysis of the individual link SINRs for P1 and P2. In Section IV, we present our ergodic-capacity scaling results, discuss the ‘‘crystallization’’ phenomenon, and study the impact of cooperation at the relay level. In Section V, we present our results on the asymptotic network capacity for the noncoherent (AF) case. We conclude in Section VI. The new technique to establish the large-deviations behavior of certain sums of dependent RVs is presented in Appendix A. Appendix B summarizes a set of (union) bounds used heavily throughout the paper. Appendices C and D contain the proofs of Theorems 1 and 6, respectively. The result from [23] needed for the proof of the ergodic capacity lower bounds for P1 and P2 is summarized in Appendix E. Appendix F contains some essentials from large-random-matrix theory needed in Section V. In Appendix G, we detail part of the solution of the fixed-point equation underlying the main result in Section V.

²The motivation for the channel model considered in this paper can be found in [1].

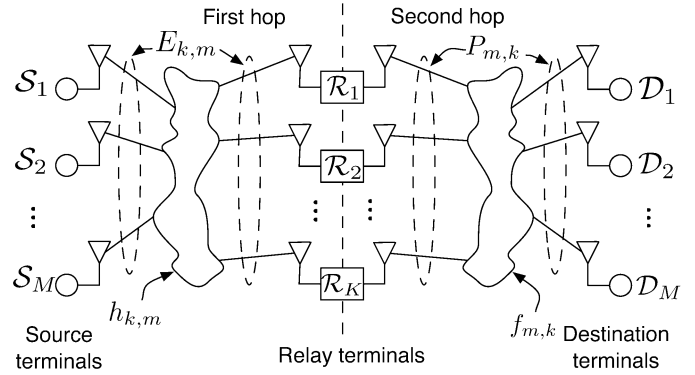


Fig. 2. Two-hop wireless relay network setup.

II. CHANNEL AND SIGNAL MODEL

In this section, we present the channel and signal model and additional basic assumptions. We restrict ourselves to the aspects that apply to both coherent and noncoherent networks and to both protocols considered in the coherent case. Relevant specifics for the coherent case will be provided in Sections III-A and B and for the noncoherent case in Section V.

A. General Assumptions

We consider an interference relay network (see Figs. 1 and 2) consisting of $K + 2M$ single-antenna terminals with M designated source–destination terminal pairs $\{\mathcal{S}_m, \mathcal{D}_m\}$ ($m \in [1 : M]$) and K relays \mathcal{R}_k ($k \in [1 : K]$). We assume a ‘‘dead-zone’’ of nonzero radius, free of relays, around each of the source and destination terminals, no direct link between the individual source–destination terminal pairs (e.g., due to large separation), and a domain of fixed area (i.e., dense network assumption). Transmission takes place in half-duplex fashion (the terminals cannot transmit and receive simultaneously) in two hops (also known as two-hop relaying) over two disjoint time slots. In the first time slot, the source terminals simultaneously broadcast their information to all the relay terminals (i.e., each relay terminal receives a superposition of all source signals). After processing the received signals, the relay terminals simultaneously broadcast the processed data to all the destination terminals during the second time slot. Our setup can be considered as an interference channel [27] with dedicated relays, hence the terminology *interference relay network*.

B. Channel and Signal Model

Throughout the paper, frequency-flat fading over the bandwidth of interest as well as perfectly synchronized transmission and reception between the terminals is assumed. For the finite- M and $K \rightarrow \infty$ case, it has been shown in [28] that the perfect-synchronization assumption can be relaxed, under quite general conditions on the synchronization errors, without impact on the capacity scaling laws. The input–output (I–O) relation for the link between the source terminals and the relay terminals during the first time slot is given by

$$\mathbf{r} = (\mathbf{E} \circ \mathbf{H}) \mathbf{s} + \mathbf{z} \quad (1)$$

where $\mathbf{r} = [r_1, r_2, \dots, r_K]^T$ with r_k denoting the signal received at the k th relay terminal, $\mathbf{E} \in \mathbb{R}^{K \times M}$ with $[\mathbf{E}]_{k,m} = \sqrt{E_{k,m}}$ where $E_{k,m}$ denotes the average energy received at \mathcal{R}_k through the $\mathcal{S}_m \rightarrow \mathcal{R}_k$ link³ (having accounted for path loss and shadowing in the $\mathcal{S}_m \rightarrow \mathcal{R}_k$ link), $\mathbf{H} \in \mathbb{C}^{K \times M}$ with $[\mathbf{H}]_{k,m} = h_{k,m}$ ($k \in [1:K]$, $m \in [1:M]$) where $h_{k,m} \sim \mathcal{CN}(0, 1)$ denotes the i.i.d. complex-valued channel gains corresponding to the $\mathcal{S}_m \rightarrow \mathcal{R}_k$ links, $\mathbf{s} = [s_1, s_2, \dots, s_M]^T$ where s_m is the zero-mean Gaussian signal transmitted by \mathcal{S}_m and the vector \mathbf{s} is i.i.d. temporally and spatially (across source terminals). Finally, $\mathbf{z} = [z_1, z_2, \dots, z_K]^T$ where $z_k \sim \mathcal{CN}(0, \sigma^2)$ is temporally and spatially (across relay terminals) white noise. The k th relay terminal processes its received signal r_k to produce the output signal t_k . The collection of output signals t_k , organized in the vector $\mathbf{t} = [t_1, t_2, \dots, t_K]^T$, is then broadcast to the destination terminals during the second time slot, while the source terminals are silent. The m th destination terminal receives the signal y_m with $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$ given by

$$\mathbf{y} = (\mathbf{P} \circ \mathbf{F}) \mathbf{t} + \mathbf{w} \quad (2)$$

where $\mathbf{P} \in \mathbb{R}^{M \times K}$ with $[\mathbf{P}]_{m,k} = \sqrt{P_{m,k}}$ and $P_{m,k}$ denotes the average energy received at \mathcal{D}_m through the $\mathcal{R}_k \rightarrow \mathcal{D}_m$ link (having accounted for path loss and shadowing in the $\mathcal{R}_k \rightarrow \mathcal{D}_m$ link). Furthermore, $\mathbf{F} \in \mathbb{C}^{M \times K}$ with $[\mathbf{F}]_{m,k} = f_{m,k}$ ($m \in [1:M]$, $k \in [1:K]$) where $f_{m,k} \sim \mathcal{CN}(0, 1)$ denotes the i.i.d. complex-valued channel gains corresponding to the $\mathcal{R}_k \rightarrow \mathcal{D}_m$ links, and $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ with $w_m \sim \mathcal{CN}(0, \sigma^2)$ being temporally and spatially (across destination terminals) white noise. Throughout the paper, we impose a per-source-terminal power constraint

$$\mathbb{E}[|s_m|^2] \leq 1/M, \quad m \in [1:M]$$

which results in the total transmit power trivially satisfying $\mathbb{E}[|\mathbf{s}|^2] \leq 1$. Furthermore, we impose a per-relay-terminal power constraint

$$\mathbb{E}[|t_k|^2] \leq P_{\text{rel}}/K, \quad k \in [1:K]$$

which results in the total power transmitted by the relay terminals satisfying $\mathbb{E}[|\mathbf{t}|^2] \leq P_{\text{rel}}$. As already mentioned above, path loss and shadowing are accounted for through the $E_{k,m}$ ($k \in [1:K]$, $m \in [1:M]$) (for the first hop), and the $P_{m,k}$ ($m \in [1:M]$, $k \in [1:K]$) (for the second hop). We assume that these parameters are deterministic, uniformly bounded from above (follows from the dead-zone assumption) and below (follows from considering a domain of fixed area) so that for all k, m

$$0 < \underline{E} \leq E_{k,m} \leq \bar{E} < \infty, \quad 0 < \underline{P} \leq P_{m,k} \leq \bar{P} < \infty. \quad (3)$$

Throughout the paper, we assume that the source terminals \mathcal{S}_m ($m \in [1:M]$) do not have CSI. The assumptions on CSI at the relays and the destination terminals depend on the setup (coherent or noncoherent case) and the protocol (in the coherent case) and will be made specific when needed.

A discussion of the motivation for the two scenarios analyzed in this paper can be found in [1].

³ $\mathcal{A} \rightarrow \mathcal{B}$ signifies communication from terminal \mathcal{A} to terminal \mathcal{B} .

III. THE COHERENT CASE

In this section, we describe the two protocols P1 and P2 and derive the corresponding SINR concentration results along with the resulting bounds on the individual source–destination link outage probability induced by P1 and P2. Note that the results in this section do not require ergodicity of \mathbf{H} and \mathbf{F} .

A. Protocol 1 (P1)

The basic setup was introduced in Section II. We shall next describe the specifics of P1. The K relay terminals are partitioned into M subsets \mathcal{M}_m ($m \in [1:M]$) with⁴ $|\mathcal{M}_m| = K/M$. The relays in \mathcal{M}_m are assumed to assist the m th source–destination terminal pair $\{\mathcal{S}_m, \mathcal{D}_m\}$. This assignment is succinctly described through the relay partitioning function $p: [1, K] \rightarrow [1, M]$ defined as

$$p(k) \triangleq m \Leftrightarrow \mathcal{R}_k \in \mathcal{M}_m.$$

We assume that the k th relay terminal has perfect knowledge of the phases $\arg(h_{k,p(k)})$ and $\arg(f_{p(k),k})$ of the single-input single-output (SISO) backward (from the perspective of the relay) channel $\mathcal{S}_{p(k)} \rightarrow \mathcal{R}_k$ and the corresponding forward channel $\mathcal{R}_k \rightarrow \mathcal{D}_{p(k)}$, respectively. We furthermore define

$$\tilde{h}_{k,p(k)} \triangleq \exp(j \arg(h_{k,p(k)}))$$

and

$$\tilde{f}_{p(k),k} \triangleq \exp(j \arg(f_{p(k),k})).$$

The signal r_k received at the k th relay terminal is first cophased w.r.t. the assigned backward channel followed by an energy normalization so that

$$u_k = d_{P1,k} \tilde{h}_{k,p(k)}^* r_k \quad (4)$$

where

$$d_{P1,k} \triangleq \sqrt{P_{\text{rel}}} \left[\frac{K}{M} \sum_{m=1}^M E_{k,m} + K\sigma^2 \right]^{-1/2} \quad (5)$$

ensures that the per-relay power constraint $\mathbb{E}[|u_k|^2] = P_{\text{rel}}/K$ is met. The relay terminal \mathcal{R}_k then computes the transmit signal t_k by cophasing w.r.t. its assigned forward channel, i.e.,

$$t_k = \tilde{f}_{p(k),k}^* u_k \quad (6)$$

which, obviously, satisfies $\mathbb{E}[|t_k|^2] \leq P_{\text{rel}}/K$ with equality and hence meets the total power constraint (across relays)

$$\mathbb{E}[|\mathbf{t}|^2] = \sum_{k=1}^K \mathbb{E}[|t_k|^2] = P_{\text{rel}}.$$

In summary, P1 ensures that the relays $\mathcal{R}_k \in \mathcal{M}_m$ forward the signal intended for \mathcal{D}_m , namely, the signal transmitted by \mathcal{S}_m , in a “doubly coherent” (w.r.t. backward and forward channels)

⁴For simplicity, we assume that K is an integer multiple of M . Moreover, in the remainder of the paper all results pertaining to P1 implicitly assume $K \geq M$.

fashion, whereas the signals transmitted by the source terminals $\mathcal{S}_{\hat{m}}$ with $\hat{m} \neq m$ are forwarded to \mathcal{D}_m in a “noncoherent” fashion (i.e., phase incoherence occurs either on the backward or the forward link or on both links). The idea underlying P1 has originally been introduced in [1] (for the finite- M case).

We shall next derive the I–O relation for the SISO channels $\mathcal{S}_m \rightarrow \mathcal{D}_m$ ($m \in [1:M]$). The destination terminal \mathcal{D}_m receives doubly (backward and forward link) coherently combined contributions corresponding to the signal s_m , with interfering terms containing contributions from the signals $s_{\hat{m}}$ with $\hat{m} \neq m$ as well as noise, forwarded by the relays. Combining (1), (4), (6), and (2), it follows (after some straightforward algebra) that the signal received at \mathcal{D}_m ($m \in [1:M]$) is given by⁵

$$y_m = s_m \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^K a_k^{m,m}}_{\text{effective channel gain}} + \underbrace{\sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{K}} \sum_{k=1}^K a_k^{m,\hat{m}}}_{\text{interference}} + \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^K b_k^m z_k + w_m}_{\text{noise}} \quad (7)$$

where

$$a_k^{m,\hat{m}} \triangleq C_{\text{P1},k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \quad (8)$$

$$b_k^m \triangleq C_{\text{P1},k}^m \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* \quad (9)$$

with

$$C_{\text{P1},k}^{m,\hat{m}} = \sqrt{K} d_{\text{P1},k} \sqrt{P_{m,k} E_{k,\hat{m}}} \quad (10)$$

$$C_{\text{P1},k}^m = \sqrt{K} d_{\text{P1},k} \sqrt{P_{m,k}} \quad (11)$$

The normalization factor \sqrt{K} in (7), (10), and (11) is introduced for convenience of exposition. Using (3), it now follows that

$$\underline{C} \triangleq \sqrt{\frac{P \underline{E} P_{\text{rel}}}{\underline{E} + \sigma^2}} \leq C_{\text{P1},k}^{m,\hat{m}} \leq \sqrt{\frac{\overline{P} \overline{E} P_{\text{rel}}}{\overline{E} + \sigma^2}} \triangleq \overline{C} \quad (12)$$

$$\underline{c} \triangleq \sqrt{\frac{P P_{\text{rel}}}{\underline{E} + \sigma^2}} \leq C_{\text{P1},k}^m \leq \sqrt{\frac{\overline{P} P_{\text{rel}}}{\overline{E} + \sigma^2}} \triangleq \overline{c} \quad (13)$$

for all $k \in [1:K]$, $m \in [1:M]$, and $\hat{m} \in [1:M]$. In the following, it will be essential that the constants \underline{C} , \underline{c} , \overline{C} , and \overline{c} do not depend on M , K .

Since we assumed that the destination terminals \mathcal{D}_m ($m \in [1:M]$) cannot cooperate, the \mathcal{D}_m cannot perform joint decoding so that the network can be viewed as a collection of M SISO channels $\mathcal{S}_m \rightarrow \mathcal{D}_m$, i.e., as an interference channel

⁵The notation $\sum_{\hat{m} \neq m}$ stands for the summation over $\hat{m} \in [1:M]$ such that $\hat{m} \neq m$. If not specified, the upper limit of the summation is clear from the context.

with dedicated relays. We can see from (7) that each of these SISO channels consists of a fading effective channel, fading interference, caused by the source signals not intended for a given destination terminal, and finally a noise term incorporating thermal noise forwarded by the relays and thermal noise added at the destination terminals. In the remainder of this section, we make the conceptual assumption that each of the destination terminals \mathcal{D}_m has perfect knowledge of the fading and path loss and shadowing coefficients in the entire network, i.e., \mathcal{D}_m ($m \in [1:M]$) knows \mathbf{H} , \mathbf{F} , \mathbf{E} , and \mathbf{P} perfectly. An immediate consequence of this assumption is that \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the effective channel gain $(1/\sqrt{K}) \sum_{k=1}^K a_k^{m,m}$, the interference channel gains $(1/\sqrt{K}) \sum_{k=1}^K a_k^{m,\hat{m}}$ ($\hat{m} \neq m$), and the quantity $(1/\sqrt{K}) \sum_{k=1}^K b_k^m$. Conditioned on \mathbf{H} and \mathbf{F} , both the interference and the noise term in (7) are Gaussian, so that the mutual information for the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link is given by

$$I(y_m; s_m | \mathbf{H}, \mathbf{F}) = \frac{1}{2} \log \left(1 + \text{SINR}_m^{\text{P1}} \right) \quad (14)$$

where $\text{SINR}_m^{\text{P1}}$, defined in (15) at the bottom of the page, is the effective SINR in the SISO channel $\mathcal{S}_m \rightarrow \mathcal{D}_m$.

We conclude by noting that the large-deviations results in Section III-C rely heavily on the assumption that \mathcal{D}_m ($m \in [1:M]$) knows \mathbf{H} , \mathbf{F} , \mathbf{E} , and \mathbf{P} perfectly. The ergodic capacity-scaling results in Section IV will, however, be seen to require significantly less channel knowledge at the destination terminals.

B. Protocol 2 (P2)

The only difference between P1 and P2 is in the processing at the relays. Whereas in P1 the K relay terminals are partitioned into M clusters (of equal size) with each of these clusters assisting one particular source–destination terminal pair, in P2 each relay assists all source–destination terminal pairs so that relay partitioning is not needed. In turn, P2 requires that each relay knows the phases of all its M backward and M forward channels, i.e., \mathcal{R}_k needs knowledge of $\tilde{h}_{k,m}$ and $\tilde{f}_{m,k}$, respectively, for $m \in [1:M]$. Consequently, P2 requires significantly more CSI at the relays than P1. The relay processing stage in P2 computes

$$t_k = d_{\text{P2},k} \left(\sum_{m=1}^M \tilde{h}_{k,m}^* \tilde{f}_{m,k}^* \right) r_k \quad (16)$$

where

$$d_{\text{P2},k} \triangleq \sqrt{P_{\text{rel}}} \left[K \sum_{m=1}^M E_{k,m} + MK\sigma^2 \right]^{-1/2}$$

ensures that the power constraint $\mathbb{E}[|t_k|^2] = P_{\text{rel}}/K$ and hence

$$\mathbb{E}[|\mathbf{t}|^2] = \sum_{k=1}^K \mathbb{E}[|t_k|^2] = P_{\text{rel}}$$

is met.

$$\text{SINR}_m^{\text{P1}} \triangleq \frac{\left| \sum_{k=1}^K a_k^{m,m} \right|^2}{\left(\sum_{\hat{m} \neq m} \left| \sum_{k=1}^K a_k^{m,\hat{m}} \right|^2 + \sigma^2 M \sum_{k=1}^K |b_k^m|^2 + KM\sigma^2 \right)}. \quad (15)$$

$$\text{SINR}_m^{\text{P2}} \triangleq \frac{\left| \sum_{k=1}^K \sum_{\hat{m}=1}^M a_k^{m,m,\hat{m}} \right|^2}{\left(\sum_{\hat{m} \neq m} \left| \sum_{k=1}^K \sum_{\hat{m}=1}^M a_k^{m,\hat{m},\hat{m}} \right|^2 + \sigma^2 M \sum_{k=1}^K \left| \sum_{\hat{m}=1}^M b_k^{m,\hat{m}} \right|^2 + KM^2 \sigma^2 \right)}. \quad (21)$$

Again, we start by deriving the I–O relation for the SISO channels $\mathcal{S}_m \rightarrow \mathcal{D}_m$ ($m \in [1:M]$). Like in P1, the destination terminal \mathcal{D}_m receives doubly (backward and forward link) coherently combined contributions corresponding to the signal s_m , interfering terms containing contributions from the signals $s_{\hat{m}}$ with $\hat{m} \neq m$, as well as noise forwarded by the relays. Combining (1), (16), and (2), it follows that the signal received at \mathcal{D}_m ($m \in [1:M]$) is given by

$$\begin{aligned} y_m = & s_m \underbrace{\frac{1}{\sqrt{KM}} \sum_{k=1}^K \sum_{\hat{m}=1}^M a_k^{m,m,\hat{m}}}_{\text{effective channel gain}} \\ & + \underbrace{\sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{KM}} \sum_{k=1}^K \sum_{\hat{m}=1}^M a_k^{m,\hat{m},\hat{m}}}_{\text{interference}} \\ & + \underbrace{\frac{1}{\sqrt{KM}} \sum_{k=1}^K \sum_{\hat{m}=1}^M b_k^{m,\hat{m}} z_k + w_m}_{\text{noise}} \end{aligned} \quad (17)$$

where

$$\begin{aligned} a_k^{m,\hat{m},\hat{m}} & \triangleq C_{\text{P2},k}^{m,\hat{m}} \tilde{f}_{\hat{m},k}^* f_{m,k} \tilde{h}_{k,\hat{m}}^* h_{k,\hat{m}} \\ b_k^{m,\hat{m}} & \triangleq C_{\text{P2},k}^m \tilde{f}_{\hat{m},k}^* f_{m,k} \tilde{h}_{k,\hat{m}}^* \end{aligned}$$

with

$$C_{\text{P2},k}^{m,\hat{m}} \triangleq \sqrt{KM} d_{\text{P2},k} \sqrt{P_{m,k} E_{k,\hat{m}}} \quad (18)$$

$$C_{\text{P2},k}^m \triangleq \sqrt{KM} d_{\text{P2},k} \sqrt{P_{m,k}}. \quad (19)$$

Again, the normalization \sqrt{KM} in (17), (18), and (19) is introduced for convenience of exposition and

$$\underline{C} \leq C_{\text{P2},k}^{m,\hat{m}} \leq \bar{C}, \quad \underline{c} \leq C_{\text{P2},k}^m \leq \bar{c}$$

for all $k \in [1:K]$, $m \in [1:M]$, and $\hat{m} \in [1:M]$ with the constants \underline{C} , \underline{c} , \bar{C} , and \bar{c} not depending on M , K .

Recalling that we assume perfect knowledge of \mathbf{H} , \mathbf{F} , \mathbf{E} , and \mathbf{P} at each of the destination terminals, \mathcal{D}_m , the mutual information for the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link in P2 is given by

$$I(y_m; s_m | \mathbf{H}, \mathbf{F}) = \frac{1}{2} \log \left(1 + \text{SINR}_m^{\text{P2}} \right) \quad (20)$$

where $\text{SINR}_m^{\text{P2}}$, defined in (21) at the top of the page, is the effective SINR in the SISO channel $\mathcal{S}_m \rightarrow \mathcal{D}_m$.

C. Large-Deviations Analysis of SINR

Our goal in this section is to prove that $\text{SINR}_m^{\text{P1}}$ and $\text{SINR}_m^{\text{P2}}$ for $m \in [1:M]$ (and, thus, the corresponding mutual information quantities (14) and (20)) lie within “narrow intervals” around their mean values with⁶ “high probability” when M ,

⁶The precise meaning of “narrow intervals” and “high probability” is explained in the formulation of Theorems 1 and 2 in Section III-D.

$K \rightarrow \infty$. The technique we use to prove these *concentration results* is based on a large-deviations analysis and can be summarized as follows.

- i) Consider each sum in the numerator and denominator of (15) and (21) separately.
- ii) Represent the considered sum as a sum of independent RVs or as a sum of dependent complex-valued RVs with independent phases.
- iii) Find the mean value of the considered sum.
- iv) Employ a large-deviations analysis to prove that the considered sum lies within a narrow interval around its mean with high probability, i.e., establish a concentration result.
- v) Combine the concentration results for the separate sums using the union bounds summarized in Appendix B to obtain concentration results for $\text{SINR}_m^{\text{P1}}$ and $\text{SINR}_m^{\text{P2}}$.

1) *Chernoff Bounds*: Before embarking on a detailed discussion of the individual Steps i–v above, we note that a well-known technique to establish large-deviations results for sums of RVs (as required in Step iv above) is based on Chernoff bounds. This method, which yields the precise exponential behavior for the tails of the distributions under question, can, unfortunately, not be applied to all the sums in (15) and (21). To solve this problem, we develop a new technique, which allows to establish large-deviations results for certain sums of dependent complex-valued RVs with independent phases where the RVs occurring in the sum are such that their MGF does not need to be known. The new technique is based on the well-known idea of truncation of RVs and will, therefore, be called truncation technique. Even though truncation of RVs is a standard concept in probability theory, and in particular in large-deviations analysis, we could not find the specific approach developed in this paper in the literature. We therefore decided to present the truncation technique as a stand-alone concept and summarized the main results in Appendix A. Before proceeding, we note that even though the truncation technique has wider applicability than Chernoff bounds, it yields weaker exponents for the tails of the distributions under question.

Although the proofs of the main concentration results, Theorems 1 and 2 in Section III-D, are entirely based on the truncation technique, we still discuss the results of the application of Chernoff bounds (without giving all the details) in the following, restricting our attention to P1, to motivate the development of the truncation technique and to provide a reference for the quality (in terms of tightness of the bounds) of the results in Theorems 1 and 2. Moreover, the developments below introduce some of the key elements of the proofs of Theorems 1 and 2.

Following the approach outlined in Steps i–v above, we start by writing $\text{SINR}_m^{\text{P1}}$ as

$$\text{SINR}_m^{\text{P1}} = \frac{|S^{(1)} + S^{(2)}|^2}{S^{(3)} + \sigma^2 M S^{(4)} + KM \sigma^2} \quad (22)$$

and establishing bounds on the probability of large deviations of

$$S^{(1)} \triangleq \sum_{k:p(k)=m} C_{P1,k}^{m,m} |f_{m,k}| |h_{k,m}| \quad (23)$$

$$S^{(2)} \triangleq \sum_{k:p(k) \neq m} C_{P1,k}^{m,m} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,m} \quad (24)$$

$$S^{(3)} \triangleq \sum_{\hat{m} \neq m} \left| \sum_{k=1}^K C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \right|^2 \quad (25)$$

$$S^{(4)} \triangleq \sum_{k=1}^K (C_{P1,k}^m)^2 |f_{m,k}|^2. \quad (26)$$

We shall see in the following that the pdfs of the terms in $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$ have a structure that is simple enough for Chernoff bounds to be applicable. We start with the analysis of the simplest term, namely, $S^{(4)}$. To avoid unnecessary technical details and to simplify the exposition, we assume (only in the ensuing analysis of the large deviations behavior of $S^{(4)}$) that

$$C_{P1,k}^{m,\hat{m}} = C_{P1,k}^m = 1 \quad (27)$$

for all $m, \hat{m} \in [1 : M]$, $k \in [1 : K]$. Defining⁷ $X_k \triangleq |f_{m,k}|^2$, we have

$$S^{(4)} = \sum_{k=1}^K X_k$$

where the X_k are i.i.d. exponentially distributed with parameter $\lambda = 1$, i.e., $f_{X_k}(x) = \exp(-x) u(x)$ and hence $\mathbb{E}[X_k] = 1$. For convenience, we centralize X_k and define $Z_k \triangleq X_k - 1$. The MGF of Z_k is given by

$$M_{Z_k}(s) = \int_0^\infty e^{s(x-1)} e^{-x} dx = \frac{e^{-s}}{1-s}, \quad \Re s \leq 1. \quad (28)$$

Since the RVs Z_k are independent, we obtain, using the standard Chernoff bound (see, for example, [29, Sec. 5.4]), for $x > 0$

$$\begin{aligned} \mathbb{P} \left\{ \sum_{k=1}^K Z_k \geq x \right\} &\leq \min_{0 \leq s \leq 1} (M_{Z_k}(s))^K e^{-sx} \\ &= \min_{0 \leq s \leq 1} e^{-Ks - K \ln(1-s) - sx}. \end{aligned} \quad (29)$$

Because $(M_{Z_k}(s))^K \exp(-sx)$ is convex in s [29, Sec. 5.4], the minimum in (29) can easily be seen to be taken on for $s = x/(x+K)$, which gives

$$\mathbb{P} \left\{ \sum_{k=1}^K Z_k \geq x \right\} \leq e^{K \ln(x+K) - K \ln(K) - x}. \quad (30)$$

The corresponding relation for negative deviations ($x < 0$) is

$$\mathbb{P} \left\{ \sum_{k=1}^K Z_k \leq x \right\} \leq \begin{cases} e^{K \ln(x+K) - K \ln(K) - x}, & x > -K \\ 0, & x < -K. \end{cases} \quad (31)$$

⁷For notational convenience, we shall omit the index m in what follows.

Finally, setting $x = \sqrt{K}t$, we get the desired concentration result for the sum $S^{(4)}$ as

$$\mathbb{P} \left\{ S^{(4)} - K \geq \sqrt{K}t \right\} \leq e^{K \ln(1+t/\sqrt{K}) - \sqrt{K}t}, \quad t \geq 0 \quad (32)$$

$$\begin{aligned} \mathbb{P} \left\{ S^{(4)} - K \leq \sqrt{K}t \right\} \\ \leq \begin{cases} e^{K \ln(1+t/\sqrt{K}) - \sqrt{K}t}, & -\sqrt{K} < t \leq 0 \\ 0, & t \leq -\sqrt{K}. \end{cases} \end{aligned} \quad (33)$$

We now consider the case when K is large and $t = o(\sqrt{K})$ so that

$$\ln \left(1 + \frac{t}{\sqrt{K}} \right) = \frac{t}{\sqrt{K}} - \frac{t^2}{2K} + O \left(\left(\frac{t}{\sqrt{K}} \right)^3 \right). \quad (34)$$

If we omit higher (than second) order terms in (34), the bound in (32) and (33) can be compactly written as

$$\mathbb{P} \left\{ \left| S^{(4)} - K \right| \geq \sqrt{K}t \right\} \leq 2e^{-t^2/2}. \quad (35)$$

We can, therefore, conclude that the probability of large deviations of $S^{(4)}$ decays exponentially.

Similar concentration results, using Chernoff bounds, can be established for $S^{(1)}$ and $S^{(2)}$. The derivation is somewhat involved (as it requires establishing upper bounds on the MGF), does not provide insights into the problem and will, therefore, be omitted. Unfortunately, the simple technique used above to establish concentration results for $S^{(4)}$ (and applicable to $S^{(1)}$ and $S^{(2)}$) does not seem to be applicable to $S^{(3)}$. To see this, we start by noting that $S^{(3)}$ contains two classes of terms (in the sense of the properties of their pdf), i.e.,

$$S^{(3)} = S^{(31)} + S^{(32)} \quad (36)$$

with

$$S^{(31)} \triangleq \sum_{\hat{m} \neq m} \sum_{k=1}^K (C_{P1,k}^{m,\hat{m}})^2 |f_{m,k}|^2 |h_{k,\hat{m}}|^2 \quad (37)$$

$$\begin{aligned} S^{(32)} \triangleq &\sum_{\hat{m} \neq m} \sum_{k=1}^K \sum_{\hat{k} \neq k} C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \\ &\times C_{P1,\hat{k}}^{m,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}}. \end{aligned} \quad (38)$$

Now, there are two problems in applying the technique we have used so far to $S^{(3)}$: First, it seems very difficult to compute the MGFs for the individual terms in $S^{(31)}$ and $S^{(32)}$; second, the individual terms in $S^{(31)}$ and $S^{(32)}$ are not *jointly*⁸ independent across the summation indices. The first problem can probably be resolved using bounds on the exact MGFs (as can be done in the analysis of $S^{(1)}$ and $S^{(2)}$). The second problem, however, seems more fundamental. In particular, the individual terms in $S^{(31)}$ are independent across k but not across \hat{m} . In $S^{(32)}$, the individual terms are independent across k but not across \hat{k} and \hat{m} . Assuming that the problem of computing (or

⁸We write ‘‘jointly independent,’’ as opposed to ‘‘pairwise independent’’ here and in what follows to stress the fact that the joint pdf of the RVs under consideration can be factored into a product of the marginal pdfs. In several places throughout the paper we will deal with sets of RVs that turn out to be pairwise independent, but not jointly independent.

properly bounding) the MGFs is resolved, a natural way to overcome the second problem mentioned above would be to establish concentration results for the sums over k , i.e., for

$$\hat{S}_{\hat{m}}^{(31)} \triangleq \sum_{k=1}^K \left(C_{P1,k}^{m,\hat{m}} \right)^2 |f_{m,k}|^2 |h_{k,\hat{m}}|^2 \quad (39)$$

$$\begin{aligned} \hat{S}_{\hat{m},\hat{k}}^{(32)} &\triangleq \sum_{k=1}^K C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \\ &\quad \times C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}} \end{aligned} \quad (40)$$

and to employ the union bound for sums (Lemmas 2 and 4 in Appendix B) to obtain concentration results for $S^{(31)}$ and $S^{(32)}$. Unfortunately, this method, although applicable, yields results that are very loose in the sense of not reflecting the correct ‘‘order-of-magnitude behavior’’ of the typical deviations. To understand why this is the case, we perform an order-of-magnitude analysis as follows. For simplicity, we again assume that the condition (27) is satisfied. Note that for any $\hat{k}, k \in [1:K]$ such that $\hat{k} \neq k$ and any $\hat{m} \in [1:M]$ such that $\hat{m} \neq m$, we have

$$\mathbb{E} \left[\tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}} \right] = 0.$$

Chernoff bounding $\hat{S}_{\hat{m},\hat{k}}^{(32)}$ would, therefore, yield that

$$\mathbb{P} \left\{ \left| \hat{S}_{\hat{m},\hat{k}}^{(32)} \right| \geq \sqrt{K}t \right\}$$

decays exponentially⁹ in t . Then, applying the union bound for sums (Lemma 2) to $S^{(32)} = \sum_{\hat{m} \neq m} \sum_{\hat{k} \neq k} \hat{S}_{\hat{m},\hat{k}}^{(32)}$, we would conclude that

$$\mathbb{P} \left\{ \left| S^{(32)} \right| \geq (M-1)(K-1)\sqrt{K}t \right\} \quad (41)$$

decays exponentially in t . Even though the terms in $S^{(32)}$ are not completely independent across \hat{k} and \hat{m} , we will see in Section III-C2 that there is still enough independence between them for the truncation technique to reveal that

$$\mathbb{P} \left\{ \left| S^{(32)} \right| \geq \sqrt{(M-1)(K-1)K}t \right\} \quad (42)$$

decays exponentially in t , which is a much stronger concentration result than (41). The importance of the difference between (42) and (41) becomes clear if we consider $S^{(31)}$. Since $\hat{S}_{\hat{m}}^{(31)}$ is a sum over K independent terms, each of which satisfies $\mathbb{E}[|f_{m,k}|^2 |h_{k,\hat{m}}|^2] = 1$, Chernoff bounding would yield that

$$\mathbb{P} \left\{ \left| \hat{S}_{\hat{m}}^{(31)} - K \right| \geq \sqrt{K}t \right\}$$

decays exponentially in t . Applying the union bound to $S^{(31)} = \sum_{\hat{m} \neq m} \hat{S}_{\hat{m}}^{(31)}$, one can then show that

$$\mathbb{P} \left\{ \left| S^{(31)} - K(M-1) \right| \geq (M-1)\sqrt{K}t \right\} \quad (43)$$

decays exponentially in t . When M and K are large, we would now conclude from (41) and (43) that $S^{(3)} = S^{(31)} + S^{(32)}$ deviates around KM with a typical deviation of order $MK\sqrt{K}$. Since the typical deviations are larger (by a factor of \sqrt{K}) than the mean, the corresponding deviation result is useless. On the other hand, if we use the bound (42) combined with (43), again

assuming that M and K are large, we can conclude that $S^{(3)}$ deviates around KM with a typical deviation of order $\sqrt{MK} + M\sqrt{K}$, which is an order of magnitude smaller than the mean. As already mentioned, the truncation technique allows us to establish useful concentration results for sums with dependent terms such as that in (40).

2) *Application of the Truncation Technique:* In this subsection, we demonstrate how the desired concentration results for $S^{(31)}$ and $S^{(32)}$, defined in (37) and (38), respectively, can be obtained by application of the truncation technique. The following results will be used in the proof of Theorem 1 and will, therefore, be formulated for general $C_{P1,k}^{m,\hat{m}}$ and $C_{P1,k}^m$.

Analysis of $S^{(31)}$: Consider $\hat{S}_{\hat{m}}^{(31)}$. The variables $X_k \triangleq |f_{m,k}|^2$ and $Y_{k,\hat{m}} \triangleq |h_{k,\hat{m}}|^2$ are exponentially distributed with parameter $\lambda = 1$. Therefore, we have

$$\mathbb{P} \left\{ X_k \geq x \right\} = \mathbb{P} \left\{ Y_{k,\hat{m}} \geq x \right\} \leq e^{-x}, \quad x \geq 0, \text{ for all } k, \hat{m}.$$

Define $Z_{k,\hat{m}} \triangleq X_k Y_{k,\hat{m}}$. From the union bound for products it follows that

$$\mathbb{P} \left\{ Z_{k,\hat{m}} \geq x^2 \right\} = \mathbb{P} \left\{ X_k Y_{k,\hat{m}} \geq x^2 \right\} \leq 2e^{-x}$$

which yields

$$\mathbb{P} \left\{ Z_{k,\hat{m}} \geq x \right\} \leq 2e^{-\sqrt{x}}.$$

Next, using $\mathbb{E}[Z_{k,\hat{m}}] = 1$ and $\mathbb{E}[(Z_{k,\hat{m}})^2] = 4$ for all $k, \hat{m} \neq m$, and the independence of the RVs $Z_{k,\hat{m}}$ across $k \in [1:K]$, it follows from Corollary 2, taking into account (12), that for $K \geq 2$

$$\mathbb{P} \left\{ \left| \hat{S}_{\hat{m}}^{(31)} - \sum_{k=1}^K \left(C_{P1,k}^{m,\hat{m}} \right)^2 \right| \geq \sqrt{K}x \right\} \leq 6Ke^{-\Delta^{(31)}x^{2/5}}$$

where $\Delta^{(31)} \triangleq \min[1, (1/8)\bar{C}^{-4}]$. Applying the union bound for sums (see Lemma 2) and using (12), we finally obtain the desired¹⁰ concentration result for $S^{(31)}$ as

$$\begin{aligned} \mathbb{P} \left\{ S^{(31)} \geq (M-1)K\bar{C}^2 + (M-1)\sqrt{K}x \right\} \\ \leq 6(M-1)Ke^{-\Delta^{(31)}x^{2/5}} \end{aligned} \quad (44)$$

and

$$\begin{aligned} \mathbb{P} \left\{ S^{(31)} \leq (M-1)K\underline{C}^2 - (M-1)\sqrt{K}x \right\} \\ \leq 6(M-1)Ke^{-\Delta^{(31)}x^{2/5}}. \end{aligned} \quad (45)$$

Analysis of $S^{(32)}$: We start by rewriting (38) as

$$\begin{aligned} S^{(32)} &= \sqrt{K-1} \\ &\quad \times \sum_{\hat{m} \neq m} \sum_{k=1}^K C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} T_{\hat{m},k}^{(32)} \end{aligned} \quad (46)$$

where $T_{\hat{m},k}^{(32)}$ is defined as

$$T_{\hat{m},k}^{(32)} \triangleq \frac{1}{\sqrt{K-1}} \sum_{\hat{k} \neq k} C_{P1,\hat{k}}^{m,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}}.$$

¹⁰We note that we do not avoid using the union bound on $S^{(31)}$. It is important, however, that we do not use it when analyzing $S^{(32)}$.

⁹We do not specify the exponent here.

The concentration result for $S^{(32)}$ (and other similar sums occurring in the proofs of Theorems 1 and 2) will be established by applying (one or multiple times) the following general steps.

- Establish a concentration result for $T_{\hat{m},k}^{(32)}$.
- Represent the terms on the right-hand side (RHS) of (46) in the form $C_{\text{PI},k}^{m,\hat{m}} Z_{\hat{m},k} \exp(j\hat{\phi}_{k,\hat{m}})$ where

$$Z_{\hat{m},k} \triangleq T_{\hat{m},k}^{(32)} |f_{m,k}| |h_{k,\hat{m}}|$$

and

$$\hat{\phi}_{k,\hat{m}} \triangleq \arg\left(\tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}}\right)$$

so that the sum $S^{(32)}$ can be written as

$$S^{(32)} \triangleq \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k=1}^K C_{\text{PI},k}^{m,\hat{m}} Z_{\hat{m},k} e^{j\hat{\phi}_{k,\hat{m}}}.$$

- Use the concentration result for $T_{\hat{m},k}^{(32)}$ together with the union bound for products (see Lemma 5) to establish bounds on the tail behavior of $Z_{\hat{m},k}$ and verify condition (109) in Theorem 10.
- If needed, split up the sum $S^{(32)}$ into several sums, so that the phases $\exp(j\hat{\phi}_{k,\hat{m}})$ are jointly independent in each of these sums and Theorem 10 can be applied (to each of these sums separately).
- Finally, apply Theorem 10 to each of the sums resulting in the previous step separately and use the union bound for sums to establish the desired concentration result for $S^{(32)}$.

Following this procedure, we start by deriving a concentration result for $T_{\hat{m},k}^{(32)}$. Since $T_{\hat{m},k}^{(32)}$ is of the same nature as $S^{(2)}$, we could, in principle, use Chernoff bounds. This would, however, lead to an exponent with a complicated dependence on t , which can be simplified only under certain assumptions on t , such as, e.g., $t = o(\sqrt{K})$ in (34). What we need is a simple universal bound for $\mathbb{P}\{|T_{\hat{m},k}^{(32)}| \geq x\}$, which is valid for all x and allows to verify condition (109) in Theorem 10 for $Z_{\hat{m},k}$. Such a bound can be obtained by applying the truncation technique to $T_{\hat{m},k}^{(32)}$ as follows. Define $X_{\hat{k}} \triangleq |f_{m,\hat{k}}|$, $Y_{\hat{k},\hat{m}} \triangleq |h_{\hat{k},\hat{m}}|$, and

$$\phi_{\hat{k},\hat{m}} \triangleq \arg\left(\tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}}\right)$$

so that

$$T_{\hat{m},k}^{(32)} = \frac{1}{\sqrt{K-1}} \sum_{\hat{k} \neq k} C_{\text{PI},\hat{k}}^{m,\hat{m}} X_{\hat{k}} Y_{\hat{k},\hat{m}} e^{j\phi_{\hat{k},\hat{m}}}.$$

The RVs $X_{\hat{k}}$ and $Y_{\hat{k},\hat{m}}$ (for all \hat{k}, \hat{m}) are Rayleigh distributed with parameter $\alpha^2 = 1/2$. Therefore, we have

$$\mathbb{P}\{X_{\hat{k}} \geq x\} = \mathbb{P}\{Y_{\hat{k},\hat{m}} \geq x\} \leq e^{-x^2}, \quad x \geq 0$$

and the union bound for products yields

$$\mathbb{P}\{X_{\hat{k}} Y_{\hat{k},\hat{m}} \geq x\} \leq 2e^{-x}, \quad x \geq 0 \quad (47)$$

which shows that condition (116) in Corollary 1 is satisfied.

Next, rewrite $\phi_{\hat{k},\hat{m}}$ as

$$\begin{aligned} \phi_{\hat{k},\hat{m}} &= \arg\left(\tilde{f}_{p(\hat{k}),\hat{k}}\right) \oplus \arg\left(f_{m,\hat{k}}^*\right) \\ &\quad \oplus \arg\left(\tilde{h}_{\hat{k},p(\hat{k})}\right) \oplus \arg\left(h_{\hat{k},\hat{m}}^*\right) \end{aligned} \quad (48)$$

where \oplus stands for addition modulo 2π . Because the f 's and the h 's in (48) are independent across $\hat{k} \in [1:K]$, it follows that the phases $\phi_{\hat{k},\hat{m}}$ are also independent across $\hat{k} \in [1:K]$, which is precisely what we need for the truncation technique to be applicable. Recalling that $m \neq \hat{m}$, and, therefore, either $p(\hat{k}) \neq m$ or $p(\hat{k}) \neq \hat{m}$, (48) implies that $\phi_{\hat{k},\hat{m}} \sim \mathcal{U}(-\pi, \pi)$ and hence $\mathbb{E}[\exp(j\phi_{\hat{k},\hat{m}})] = 0$ for all \hat{k}, \hat{m} . Since $\phi_{\hat{k},\hat{m}}$ is independent of $X_{\hat{k}}$ and $Y_{\hat{k},\hat{m}}$, we have $\mathbb{E}[\exp(j\phi_{\hat{k},\hat{m}}) X_{\hat{k}} Y_{\hat{k},\hat{m}}] = 0$ for all \hat{k}, \hat{m} and hence, $\mathbb{E}[T_{\hat{m},k}^{(32)}] = 0$ for all \hat{m}, k . Finally, applying Corollary 1 to $T_{\hat{m},k}^{(32)}$, taking into account (12), we get for $K \geq 2$ and $x \geq 0$ that

$$\mathbb{P}\{|T_{\hat{m},k}^{(32)}| \geq x\} \leq 8(K-1)e^{-\Delta^{(T)} x^{2/3}} \quad (49)$$

with $\Delta^{(T)} \triangleq 2^{-1/3} \min[1, (1/2)\bar{C}^{-2}]$.

We are now ready to establish the concentration result for $S^{(32)}$. First, rewrite $\phi_{k,\hat{m}}$ as

$$\begin{aligned} \phi_{k,\hat{m}} &\triangleq \arg\left(\tilde{f}_{p(k),k}^*\right) \oplus \arg\left(f_{m,k}\right) \\ &\quad \oplus \arg\left(\tilde{h}_{k,p(k)}^*\right) \oplus \arg\left(h_{k,\hat{m}}\right). \end{aligned} \quad (50)$$

Similar to $\phi_{\hat{k},\hat{m}}$ in (48), because $\hat{m} \neq m$ we conclude that $\phi_{k,\hat{m}} \sim \mathcal{U}(-\pi, \pi)$. Furthermore, because $\hat{k} \neq k$, the $\phi_{k,\hat{m}}$ are independent of $T_{\hat{m},k}^{(32)}$, and therefore also of $Z_{\hat{m},k}$ (for all k, \hat{m}). To apply Corollary 1 to $S^{(32)}$, the $\phi_{k,\hat{m}}$ are required to be jointly independent across $\hat{m} \in [1:M]$ for $\hat{m} \neq m$ and $k \in [1:K]$. It can be verified that this is not the case. There is, however, a simple way to resolve this problem by considering the two disjoint index sets

$$\begin{aligned} I_1 &\triangleq \left\{ (\hat{m}, k) \mid \hat{m} \in [1:M], \right. \\ &\quad \left. \hat{m} \neq m, k \in [1:K], p(k) \neq \hat{m} \right\} \\ I_2 &\triangleq \left\{ (\hat{m}, k) \mid \hat{m} \in [1:M], \right. \\ &\quad \left. \hat{m} \neq m, k \in [1:K], p(k) = \hat{m} \right\}. \end{aligned}$$

It follows by inspection that within each of the sets $\{\phi_{k,\hat{m}}\}_{(k,\hat{m}) \in I_1}$ and $\{\phi_{k,\hat{m}}\}_{(k,\hat{m}) \in I_2}$ the phases are jointly independent. Separating $S^{(32)}$ into two sums corresponding to the group of indices I_1 and I_2 , we get

$$S^{(32)} = S^{(321)} + S^{(322)} \quad (51)$$

with

$$\begin{aligned} S^{(321)} &\triangleq \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k:p(k) \neq \hat{m}} C_{\text{PI},k}^{m,\hat{m}} Z_{\hat{m},k} e^{j\phi_{k,\hat{m}}} \\ S^{(322)} &\triangleq \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k:p(k) = \hat{m}} C_{\text{PI},k}^{m,\hat{m}} Z_{\hat{m},k} e^{j\phi_{k,\hat{m}}}. \end{aligned}$$

$$L_{P1}(x) \triangleq \frac{\pi^2 \underline{C}^2 K}{16 \bar{C}_{SN}^2 M^3} \frac{\max\left[0, 1 - \frac{8}{\underline{C}\pi} \frac{M}{\sqrt{K}} x\right]^2}{\frac{\bar{C}^2}{\underline{C}_{SN}^2} + \frac{3}{\underline{C}_{SN}^2} \frac{x}{\sqrt{M}} + \frac{\sigma^2}{\underline{C}_{SN}^2} \left(\bar{c}^2 + \frac{x}{\sqrt{K}}\right) + \frac{\sigma^2}{\underline{C}_{SN}^2}} \quad (54)$$

$$U_{P1}(x) \triangleq \frac{\pi^2 \bar{C}^2 K}{16 \underline{C}_{SN}^2 M^3} \frac{\left(1 + \frac{8}{\underline{C}\pi} \frac{M}{\sqrt{K}} x\right)^2}{\max\left[0, \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{M-1}{M} - \frac{3}{\underline{C}_{SN}^2} \frac{x}{\sqrt{M}}\right] + \max\left[0, \frac{\sigma^2}{\underline{C}_{SN}^2} \left(\bar{c}^2 - \frac{x}{\sqrt{K}}\right)\right] + \frac{\sigma^2}{\underline{C}_{SN}^2}} \quad (55)$$

Applying the union bound for products first to $|f_{m,k}| |h_{k,\hat{m}}|$ as in (47), then to $Z_{\hat{m},k}$ using (49), and using the simple bound

$$2e^{-x} + 8(K-1)e^{-\Delta^{(T)}x^{1/3}} \leq 16(K-1)e^{-\Delta^{(T)}x^{1/3}}$$

which is valid for $x \geq 1$, we get

$$\mathbb{P}\{|Z_{\hat{m},k}| \geq x\} \leq 16(K-1)e^{-\Delta^{(T)}x^{1/3}}$$

for $K \geq 2$ and $x \geq 1$. Therefore, using

$$\mathbb{E}\left[Z_{\hat{m},k} \exp(j\hat{\phi}_{k,\hat{m}})\right] = 0, \quad \text{for all } k, \hat{m} \neq m$$

applying Corollary 1 to $S^{(321)}$ (which consists of $K(M-1)^2/M$ terms) and to $S^{(322)}$ (which consists of $K(M-1)/M$ terms) separately, taking into account (12), we obtain that for $K \geq 2$, $M > 2$, and $x \geq 1$

$$\begin{aligned} \mathbb{P}\left\{|S^{(321)}| \geq \sqrt{\frac{(K-1)K(M-1)^2}{M}} x\right\} \\ \leq 64 \frac{(K-1)K(M-1)^2}{M} e^{-\Delta^{(32)}x^{2/7}} \end{aligned} \quad (52)$$

and

$$\begin{aligned} \mathbb{P}\left\{|S^{(322)}| \geq \sqrt{\frac{(K-1)K(M-1)}{M}} x\right\} \\ \leq 64 \frac{(K-1)K(M-1)}{M} e^{-\Delta^{(32)}x^{2/7}} \end{aligned} \quad (53)$$

where $\Delta^{(32)} = 2^{-10/21} \min[1, (1/2)\bar{C}^{-2}]$. Combining (35) (and similar bounds for $S^{(1)}$ and $S^{(2)}$), (52), (53), (51), (44), (45), and (36), we can now state the final concentration result for SINR_m^{P1} by carrying out Step v in the summary presented in the first paragraph of Section III-C. Recall, however, that we used the classical Chernoff-bounding technique to establish the large-deviations behavior of $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$, whereas we employed the truncation technique to analyze the large-deviations behavior of $S^{(3)}$. Even though the Chernoff bounds are tighter than the bounds obtained through the truncation technique, the tightness of the final bounds for the tail behavior of SINR_m^{P1} and SINR_m^{P2} is determined by the weakest exponent in the bounds for the individual terms $S^{(1)}$, $S^{(2)}$, $S^{(3)}$, and $S^{(4)}$. Therefore, employing Chernoff bounds for $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$, and the truncation technique for $S^{(3)}$ will not lead to a significantly tighter final result, compared to the case where the truncation technique is used throughout. Motivated by this observation and for simplicity of exposition, we therefore decided to state the concentration results in Section III-D for SINR_m^{P1} and SINR_m^{P2} obtained by applying the truncation technique throughout.

D. Concentration Results for P1 and P2

In Section III-C, we outlined how the large-deviations behavior of the SINR (for P1 and P2) can be established based on the truncation technique and on union bounds. The resulting key statement, made precise in Theorems 1 and 2 below, is that the probability of the SINR falling outside a narrow interval around its mean is “exponentially small.” We proceed with the formal statement of the results.

Theorem 1: For any $K \geq 2$, $M \geq 2$, for any $x \geq 1$, the probability $P_{P1}(x)$ of the event

$$\text{SINR}_m^{P1} \notin [L_{P1}(x), U_{P1}(x)], \quad m \in [1 : M]$$

where $L_{P1}(x)$ and $U_{P1}(x)$ are as shown in (54) and (55) at the top of the page, respectively, with the constants \bar{C}_{SN} and \underline{C}_{SN} given by

$$\bar{C}_{SN} \triangleq \sqrt{\bar{C}^2 + \sigma^2 (\bar{c}^2 + 1)} \quad \underline{C}_{SN} \triangleq \sqrt{\underline{C}^2 + \sigma^2 (\underline{c}^2 + 1)}$$

satisfies the following inequality:

$$P_{P1}(x) \leq 302K^2 M e^{-\Delta_{P1}x^{2/7}} \quad (56)$$

with $\Delta_{P1} \triangleq \min\left[2^{-\frac{10}{21}}, 1/\left(2^{\frac{31}{21}}\bar{C}^2\right), 1/\left(8\bar{C}^4\right), 1/\left(4\bar{c}^4\right)\right]$.

Proof: See Appendix C. \square

Theorem 2: For any $K \geq 2$, $M \geq 2$, for any $x \geq 1$, the probability $P_{P2}(x)$ of the event

$$\text{SINR}_m^{P2} \notin [L_{P2}(x), U_{P2}(x)], \quad m \in [1 : M]$$

where $L_{P2}(x)$ and $U_{P2}(x)$ are given in (57) and (58) at the top of the following page, respectively, satisfies the following inequality:

$$P_{P2}(x) \leq 814K^2 M^3 e^{-\Delta_{P2}x^{2/9}} \quad (59)$$

with $\Delta_{P2} \triangleq \min\left[2^{-\frac{11}{9}}, 1/\left(2^{\frac{61}{36}}\bar{C}^2\right), 1/\left(8\bar{C}^4\right), 1/\left(4\bar{c}^4\right)\right]$.

Proof: The proof idea is the same as that underlying the proof of Theorem 1 with large parts of the proof itself being very similar to the proof of Theorem 1. For the sake of brevity the details of the proof are therefore omitted. \square

The concentration results in Theorems 1 and 2 form the basis for showing that, provided the rate of growth of K as a function of M is fast enough, the network “decouples” (see Theorems 3 and 4) and “crystallizes” (see Theorem 5). Moreover, as outlined in Theorem 5, the outage capacity behavior of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links can be inferred from (56) and (59).

$$L_{P2}(x) \triangleq \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{K}{M^2} \frac{\max\left[0, 1 - \frac{8}{\underline{C}\pi} \sqrt{\frac{M}{K}} x\right]^2}{\frac{\underline{C}^2}{\underline{C}_{SN}^2} + \frac{4}{\underline{C}_{SN}^2} \frac{x}{\min[\sqrt{M}, \sqrt{K}]} + \frac{\sigma^2}{\underline{C}_{SN}^2} \left(\underline{c}^2 + 2\frac{x}{\sqrt{K}}\right) + \frac{\sigma^2}{\underline{C}_{SN}^2}} \quad (57)$$

$$U_{P2}(x) \triangleq \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{K}{M^3} \frac{\left(1 + \frac{8}{\underline{C}\pi} \sqrt{\frac{M}{K}} x\right)^2}{\max\left[0, \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{M-1}{M} - \frac{4}{\underline{C}_{SN}^2} \frac{x}{\min[\sqrt{M}, \sqrt{K}]} \right] + \max\left[0, \frac{\sigma^2}{\underline{C}_{SN}^2} \left(\underline{c}^2 - 2\frac{x}{\sqrt{K}}\right)\right] + \frac{\sigma^2}{\underline{C}_{SN}^2}} \quad (58)$$

IV. ERGODIC CAPACITY AND COOPERATION AT THE RELAY LEVEL

The focus in the previous section was on establishing concentration results for the individual link SINRs for P1 and P2. Based on these results, in this section, we study the ergodic capacity realized by the two protocols and we establish the corresponding capacity scaling and outage capacity behavior.

A. Ergodic Capacity of P1 and P2

Throughout this section, we assume that all channels in the network are ergodic. The two main results are summarized as follows.

Theorem 3 (Ergodic Capacity of P1): Suppose that destination terminal \mathcal{D}_m ($m \in [1 : M]$) has perfect knowledge of the mean of the effective channel gain of the $S_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/(4\sqrt{K})) \sum_{k:p(k)=m} C_{P1,k}^{m,m}$. Then, for any $\epsilon, \delta > 0$ there exist $M_0, K_0 > 0$ such that for all $M \geq M_0, K \geq K_0$, the per source–destination terminal pair capacity achieved by P1 satisfies

$$\begin{aligned} \frac{1}{2} \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{K}{M^3} (1 - \epsilon) \right) &\leq C_{P1} \\ &\leq \frac{1}{2} \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{\max[K, M^{2+\delta}]}{M^3} (1 + \epsilon) \right). \end{aligned} \quad (60)$$

Theorem 4 (Ergodic Capacity of P2): Suppose that destination terminal \mathcal{D}_m ($m \in [1 : M]$) has perfect knowledge of the mean of the effective channel gain of the $S_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/(4\sqrt{KM})) \sum_{k=1}^K C_{P2,k}^{m,m}$. Then, for any $\epsilon, \delta > 0$ there exist $M_0, K_0 > 0$, such that for all $M \geq M_0, K \geq K_0$, the per source–destination terminal pair capacity achieved by P2 satisfies

$$\begin{aligned} \frac{1}{2} \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{K}{M^2} (1 - \epsilon) \right) &\leq C_{P2} \\ &\leq \frac{1}{2} \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{\max[K, M^{1+\delta}]}{M^2} (1 + \epsilon) \right). \end{aligned} \quad (61)$$

The proofs of Theorems 3 and 4 are very similar. Below we present the proof of Theorem 3 only. The proof of Theorem 4 is omitted.

Proof of Theorem 3: We start by establishing the lower bound in (60), the proof of which uses the result summarized in

Appendix E. To apply Lemma 7 in Appendix E, we start from (7) and define

$$\bar{F}_m \triangleq \frac{1}{\sqrt{K}} \sum_{k=1}^K \mathbb{E}[a_k^{m,m}]$$

$$\tilde{F}_m \triangleq \frac{1}{\sqrt{K}} \sum_{k=1}^K (a_k^{m,m} - \mathbb{E}[a_k^{m,m}])$$

$$W_m \triangleq \sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{K}} \sum_{k=1}^K a_k^{m,\hat{m}} + \frac{1}{\sqrt{K}} \sum_{k=1}^K b_k^m z_k + w_m.$$

With these definitions, we can now rewrite (7) as

$$y_m = \left(\bar{F}_m + \tilde{F}_m \right) s_m + W_m.$$

Straightforward, but tedious, manipulations yield

$$\begin{aligned} \bar{F}_m &= \frac{\pi}{4} \frac{1}{\sqrt{K}} \sum_{k:p(k)=m} C_{P1,k}^{m,m} \\ \text{Var}[\tilde{F}_m] &= \frac{1}{K} \left(\sum_{k=1}^K \left(C_{P1,k}^{m,m} \right)^2 - \frac{\pi^2}{16} \sum_{k:p(k)=m} \left(C_{P1,k}^{m,m} \right)^2 \right) \\ \text{Var}[W_m] &= \frac{1}{KM} \sum_{\hat{m} \neq m} \sum_{k=1}^K \left(C_{P1,k}^{m,\hat{m}} \right)^2 \\ &\quad + \frac{\sigma^2}{K} \sum_{k=1}^K \left(C_{P1,k}^{m,k} \right)^2 + \sigma^2. \end{aligned}$$

Next, we use (12) and (13) to lower-bound \bar{F}_m and upper-bound $\text{Var}[\tilde{F}_m]$ and $\text{Var}[W_m]$, substitute the resulting bounds into (150), and obtain¹¹

$$I(y_m; s_m) \geq \frac{1}{2} \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{(1/M)\underline{C}^2 + \underline{C}_{SN}^2} \frac{K}{M^3} \right). \quad (62)$$

Finally, fix $\epsilon > 0$ and set

$$M_0 = \frac{1 - \epsilon}{\epsilon} \frac{\underline{C}^2}{\underline{C}_{SN}^2}.$$

It then follows that for any $M \geq M_0$, the inequality

$$\frac{\underline{C}^2}{(1/M)\underline{C}^2 + \underline{C}_{SN}^2} \geq \frac{\underline{C}^2}{\underline{C}_{SN}^2} (1 - \epsilon)$$

¹¹We note that this bound is valid for arbitrary M and K and is, therefore, somewhat stronger than the asymptotic bound we are actually seeking.

is satisfied, which together with (62) completes the proof of the lower bound.

Proving the upper bound on C_{P1} in (60) turns out to be significantly more challenging. The method we use to this end is based on the concentration result for $\text{SINR}_m^{\text{P1}}$ in Theorem 1. We start by noting that the per-stream ergodic capacity can be upper-bounded by assuming that \mathcal{D}_m has perfect knowledge of \mathbf{H} and \mathbf{F} , i.e.,

$$\begin{aligned} C_m^{\text{P1}} &\leq \frac{1}{2} \mathbb{E}_{\mathbf{H}, \mathbf{F}} [I(y_m; s_m | \mathbf{H}, \mathbf{F})] \\ &= \frac{1}{2} \mathbb{E}_{\mathbf{H}, \mathbf{F}} \left[\log \left(1 + \text{SINR}_m^{\text{P1}} \right) \right] \\ &\leq \frac{1}{2} \log \left(1 + \mathbb{E}_{\mathbf{H}, \mathbf{F}} \left[\text{SINR}_m^{\text{P1}} \right] \right) \end{aligned}$$

where the last step follows from Jensen's inequality.

Now fix $\epsilon > 0$. To prove the upper bound in (60), it suffices to show that there exist $M_0, K_0 > 0$ such that for all $M \geq M_0$ and $K \geq K_0$

$$\mathbb{E}_{\mathbf{H}, \mathbf{F}} \left[\text{SINR}_m^{\text{P1}} \right] \leq A \frac{\max[K, M^{2+\delta}]}{M^3} (1 + \epsilon)$$

where we define

$$A \triangleq \frac{\pi^2 \bar{C}^2}{16 \underline{C}_{\text{SN}}^2}.$$

To simplify the exposition, we define

$$g(M, K) \triangleq \frac{1}{A} \text{SINR}_m^{\text{P1}}(M, K) \frac{M^3}{\max[K, M^{2+\delta}]}.$$

Note that we make the dependence of $\text{SINR}_m^{\text{P1}}$ on M and K explicit by using the notation $\text{SINR}_m^{\text{P1}}(M, K)$. In the remainder of the proof, we show that

$$\mathbb{E}_{\mathbf{H}, \mathbf{F}} [g(M, K)] \leq 1 + \epsilon \quad (63)$$

for M and K large enough. Let $f_g(x)$ denote the pdf of $g(M, K)$. Then, the expectation $\mathbb{E}_{\mathbf{H}, \mathbf{F}} [g(M, K)]$ can be written as

$$\begin{aligned} \mathbb{E}_{\mathbf{H}, \mathbf{F}} [g(M, K)] &= \int_0^\infty t f_g(t) dt \\ &= \int_0^{1+\epsilon_1} t f_g(t) dt + \int_{1+\epsilon_1}^\infty t f_g(t) dt \quad (64) \end{aligned}$$

where $\epsilon_1 > 0$ is chosen such that

$$1 + \epsilon_1 < 1 + \epsilon/3.$$

Consequently, we have

$$\begin{aligned} \int_0^{1+\epsilon_1} t f_g(t) dt &\leq (1 + \epsilon_1) \int_0^{1+\epsilon_1} f_g(t) dt \\ &\leq 1 + \epsilon_1 < 1 + \epsilon/3. \quad (65) \end{aligned}$$

For bounding the second integral on the RHS of (64), it is convenient to write the upper bound in Theorem 1 in the following form: there exist $\Delta > 0$, $\delta_1 > 0$, $\delta_2 > 0$, and $A_1, A_2, A_3 > 0$ such that for any $x \geq 1$ and $M, K \geq 2$

$$\mathbb{P} \left\{ g(M, K) \geq B(M, K, x) \right\} \leq A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta x^{2/7}} \quad (66)$$

with

$$B(M, K, x) \triangleq \frac{B^N(M, K, x)}{B^D(M, K, x)}$$

where $B^N(M, K, x)$ and $B^D(M, K, x)$ are given by

$$\begin{aligned} B^N(M, K, x) &\triangleq \frac{K}{\max[K, M^{2+\delta}]} \left(1 + A_1 \frac{M}{\sqrt{K}} x \right)^2 \quad (67) \\ B^D(M, K, x) &\triangleq \left(\underline{C}^2 \max \left[0, \frac{M-1}{M} - \frac{A_2 x}{\underline{C}^2 \sqrt{M}} \right] + \underline{c}^2 \sigma^2 \right. \\ &\quad \left. \times \max \left[0, 1 - \frac{x}{\underline{c}^2 \sqrt{K}} \right] + \sigma^2 \right) / \underline{C}_{\text{SN}}^2 \quad (68) \end{aligned}$$

The second integral on the RHS of (64) will be shown, for M and K large enough, to be upper-bounded by $2\epsilon/3$ by splitting it up and proving that

$$\int_{1+\epsilon_1}^{\lceil t_0 \rceil} t f_g(t) dt \leq \epsilon/3 \quad (69)$$

and

$$\int_{\lceil t_0 \rceil}^\infty t f_g(t) dt \leq \epsilon/3 \quad (70)$$

where the parameter $t_0 > 1 + \epsilon_1$, independent of M, K , will be chosen later. It will become clear later why we need to split up the second integral on the RHS of (64) according to (69) and (70). The integral in (69) can be bounded as follows:

$$\begin{aligned} \int_{1+\epsilon_1}^{\lceil t_0 \rceil} t f_g(t) dt &\leq \lceil t_0 \rceil \int_{1+\epsilon_1}^{\lceil t_0 \rceil} f_g(t) dt \\ &\leq \lceil t_0 \rceil \mathbb{P} \left\{ g(M, K) \geq 1 + \epsilon_1 \right\}. \end{aligned}$$

Set $x(M) = (\min[\sqrt{M}, M^\delta])^{1/3}$. With this choice of $x(M)$, it is not difficult to show that

$$\begin{aligned} \lim_{M, K \rightarrow \infty} A_1 \frac{M x(M)}{\sqrt{\max[K, M^{2+\delta}]}} &= 0 \\ \lim_{M, K \rightarrow \infty} A_2 \frac{x(M)}{\underline{C}^2 \sqrt{M}} &= 0 \\ \lim_{M, K \rightarrow \infty} \frac{x}{\underline{c}^2 \sqrt{K}} &= 0 \end{aligned}$$

which yields

$$\lim_{M, K \rightarrow \infty} B^N(M, K, x(M)) = \lim_{M, K \rightarrow \infty} \frac{K}{\max[K, M^{2+\delta}]} \leq 1. \quad (71)$$

Using $\underline{C}_{\text{SN}}^2 = \underline{C}^2 + \sigma^2 (\underline{c}^2 + 1)$, we can furthermore conclude that

$$\lim_{M, K \rightarrow \infty} B^D(M, K, x(M)) = 1$$

which, together with (71), implies that

$$\lim_{M, K \rightarrow \infty} B(M, K, x(M)) \leq 1.$$

We can, therefore, conclude that there exist $M_0^{(11)}, K_0^{(11)} > 0$ such that for any $M \geq M_0^{(11)}$ and $K \geq K_0^{(11)}$

$$B(M, K, x(M)) \leq 1 + \epsilon_1. \quad (72)$$

Trivially, we have

$$\lim_{M, K \rightarrow \infty} M^{\delta_1} K^{\delta_2} e^{-\Delta(x(M))^{2/7}} = 0$$

and, therefore, there exist $M_0^{(12)}, K_0^{(12)} > 0$ such that for any $M \geq M_0^{(12)}$ and $K \geq K_0^{(12)}$

$$A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta(x(M))^{2/7}} \leq \frac{\epsilon}{3 \lceil t_0 \rceil}. \quad (73)$$

Combining (72) and (73) and setting

$$M_0^{(1)} = \max[M_0^{(11)}, M_0^{(12)}], \quad K_0^{(1)} = \max[K_0^{(11)}, K_0^{(12)}]$$

we get that for any $M \geq M_0^{(1)}$ and $K \geq K_0^{(1)}$

$$\lceil t_0 \rceil \mathbb{P}\{g(M, K) \geq 1 + \epsilon_1\} \leq \epsilon/3 \quad (74)$$

which concludes the proof of (69).

To show (70), we note that

$$\int_{\lceil t_0 \rceil}^{\infty} t f_g(t) dt \leq \sum_{n=\lceil t_0 \rceil}^{\infty} (n+1) \mathbb{P}\{g(M, K) \geq n\} \triangleq S. \quad (75)$$

Expanding the square, upper-bounding x by x^2 in $B^N(M, K, x)$, and substituting the max terms in $B^D(M, K, x)$ by 0, we obtain the bound

$$\begin{aligned} B(M, K, x) &\leq \frac{K}{\max[K, M^{2+\delta}]} \frac{C_{\text{SN}}^2}{\sigma^2} \left(1 + \left(2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K} \right) x^2 \right) \\ &\triangleq B_1(M, K, x^2). \end{aligned} \quad (76)$$

Applying the change of variables $y = x^2$ in (76) and (66), we finally get

$$\begin{aligned} \mathbb{P}\{g(M, K) \geq B_1(M, K, \sqrt{y})\} &\leq \mathbb{P}\{g(M, K) \geq B(M, K, \sqrt{y})\} \\ &\leq A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta y^{1/7}}. \end{aligned} \quad (77)$$

Equating $B_1(M, K, y)$ with n and solving for y , we find that

$$\mathbb{P}\{g(M, K) \geq n\} \leq A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta(y_2(n, M, K))^{1/7}}$$

with

$$y_2(n, M, K) = \frac{\frac{\max[K, M^{2+\delta}]}{K} \left(\frac{\sigma^2}{C_{\text{SN}}^2} n - \frac{K}{\max[K, M^{2+\delta}]} \right)}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}}. \quad (78)$$

Now, S defined in (75) can be upper-bounded as

$$\begin{aligned} S &\leq 2 \sum_{n=\lceil t_0 \rceil}^{\infty} n \mathbb{P}\{g(M, K) \geq n\} \\ &\leq 2A_3 M^{\delta_1} K^{\delta_2} \sum_{n=\lceil t_0 \rceil}^{\infty} n e^{-\Delta(y_2(n, M, K))^{1/7}}. \end{aligned} \quad (79)$$

If n is such that $\sigma^2 n / C_{\text{SN}}^2 > 1$, then the expression in the parentheses in the numerator of (78) is strictly positive and it follows that

$$\lim_{M, K \rightarrow \infty} y_2(n, M, K) = \infty.$$

Therefore, if t_0 is chosen such that $\lceil t_0 \rceil > C_{\text{SN}}^2 / \sigma^2$, each term in the sum in (79) goes to zero exponentially fast in M, K . Note that the split-up in (69) and (70) was needed to enable us to choose t_0 large enough here. To simplify the exposition in the following, we set $t_0 = (2^7 + 1) C_{\text{SN}}^2 / \sigma^2$, so that

$$\left(\frac{\sigma^2}{C_{\text{SN}}^2} n - \frac{K}{\max[K, M^{2+\delta}]} \right)^{1/7} \geq 2$$

for $n \geq \lceil t_0 \rceil$. Next, we note that

$$\lim_{M, K \rightarrow \infty} \frac{\max[K, M^{2+\delta}]}{K} \frac{1}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}} = \infty$$

so that there exist $M_0^{(2)}, K_0^{(2)} > 0$ such that for any $M \geq M_0^{(2)}$ and $K \geq K_0^{(2)}$

$$\left(\frac{\max[K, M^{2+\delta}]}{K} \frac{1}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}} \right)^{1/7} \geq 2.$$

Now using that, trivially,

$$xy \geq x + y$$

for $x, y \geq 2$, we have for any $M \geq M_0^{(2)}, K \geq K_0^{(2)}$, and $n \geq \lceil t_0 \rceil$

$$\begin{aligned} (y_2(n, M, K))^{1/7} &\geq \left(\frac{\sigma^2}{C_{\text{SN}}^2} n - \frac{K}{\max[K, M^{2+\delta}]} \right)^{1/7} \\ &\quad + \left(\frac{\max[K, M^{2+\delta}]}{K} \frac{1}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}} \right)^{1/7} \end{aligned}$$

which yields

$$S \leq 2A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta \left(\frac{2A_1 M \sqrt{K} + A_1^2 M^2}{\max[K, M^{2+\delta}]} \right)^{-1/7}} \sum_{n=\lceil t_0 \rceil}^{\infty} h(n)$$

with

$$h(n) \triangleq n \exp \left(-\Delta \left(\frac{\sigma^2}{C_{\text{SN}}^2} n - 1 \right)^{1/7} \right).$$

Clearly, $h(n)$ decays fast enough for $\sum_{n=\lceil t_0 \rceil}^{\infty} h(n)$ to converge to a finite limit, in other words, there exists a constant $C < \infty$ (independent of M, K) such that

$$\sum_{n=\lceil t_0 \rceil}^{\infty} h(n) \leq C. \quad (80)$$

Moreover, it is easily seen that

$$\lim_{M, K \rightarrow \infty} M^{\delta_1} K^{\delta_2} e^{-\Delta \left(\frac{2A_1 M \sqrt{K} + A_1^2 M^2}{\max\{K, M^{2+\delta}\}} \right)^{-1/7}} = 0$$

which, together with (80), shows that S can be made arbitrarily small by choosing M and K large enough. More specifically, there exist $M_0^{(3)}, K_0^{(3)} > 0$ such that for any $M \geq M_0^{(3)}$ and $K \geq K_0^{(3)}$

$$S \leq \epsilon/3. \quad (81)$$

Taking

$$M_0 \triangleq \max\{M_0^{(1)}, M_0^{(2)}, M_0^{(3)}\}$$

$$K_0 \triangleq \max\{K_0^{(1)}, K_0^{(2)}, K_0^{(3)}\}$$

and combining (65), (74), and (81), we have shown (63), which completes the proof. \square

B. The ‘‘Crystallization’’ Phenomenon

As pointed out in the Introduction, the ‘‘crystallization’’ phenomenon occurs for $M, K \rightarrow \infty$, provided that K scales fast enough as a function of M , and manifests itself in two effects, namely, the decoupling of the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links and the convergence of each of the resulting SISO links to a non-fading link.

1) *Decoupling of the Network*: Theorems 3 and 4 show that in the $M, K \rightarrow \infty$ limit, the per-source destination terminal pair capacity scales as $C_{P1} = (1/2) \log(1 + \Theta(K/M^3))$ in P1 and $C_{P2} = (1/2) \log(1 + \Theta(K/M^2))$ in P2. We can, therefore, conclude that if $K \propto M^{3+\alpha}$ in P1 and $K \propto M^{2+\alpha}$ in P2 with $\alpha \geq 0$, apart from the factor $1/2$, which is due to the use of two time slots, P1 and P2 achieve full spatial multiplexing gain [30] (i.e., full sum-capacity pre-log) without any cooperation of the terminals in the network, not even the destination terminals. The corresponding distributed array gain (i.e., the factor inside the log) is given by M^α in both cases.

The fact that the per source–destination terminal pair capacity is strictly positive when K scales at least as fast as M^3 in P1 and at least as fast as M^2 in P2 shows that the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links in the network ‘‘decouple’’ in the sense that the SINR is strictly positive for each of the links. Note that this does not imply that the interference at the \mathcal{D}_m (created by $s_{\hat{m}}$ with $\hat{m} \neq m$) vanishes. Rather, if K scales fast enough, the signal power starts dominating the interference (plus noise) power. Since the corresponding upper and lower bounds in Theorems 3 and 4 exhibit the same scaling behavior, the $K \propto M^3$ and $K \propto M^2$, respectively, thresholds are fundamental in the sense of defining the critical scaling rate by delineating the regime where interference dominates over the signal and hence drives the per source–destination terminal pair capacity to zero from the regime where the signal dominates the interference and the per source–destination terminal pair capacity is strictly positive.

Further inspection of the upper and lower bounds in (60) and (61) reveals that, for fixed $\epsilon > 0$, unless all path loss and shadowing coefficients $E_{k,m}$ and $P_{m,k}$ ($k \in [1:K], m \in [1:M]$) are equal and hence $\bar{C}^2 = \underline{C}^2$ and $\bar{C}_{SN}^2 = \underline{C}_{SN}^2$, there is a gap (apart from that due to $\epsilon > 0$) between the bounds.

The order-of-magnitude reduction in the threshold for critical scaling in P2, when compared with P1, comes at the cost of each relay having to know all M backward and M forward channels. We can, therefore, conclude that P1 and P2 trade off the number of relay terminals for channel knowledge at the relays.

Finally, it is worthwhile to point out that in contrast to the finite- M results for P1 in [1], the destination terminals \mathcal{D}_m do not need knowledge of the fading coefficients $h_{k,m}$ and $f_{m,k}$. This can be seen by noting that the quantity $(\pi/(4\sqrt{K})) \sum_{k:p(k)=m} C_{P1,k}^{m,m}$, which has to be known at \mathcal{D}_m , depends on $E_{k,m}, P_{m,k}, K$, and M only. Moreover, the coefficient $(\pi/(4\sqrt{K})) \sum_{k:p(k)=m} C_{P1,k}^{m,m}$ can easily be acquired through training.

2) *Convergence to Nonfading Links and ‘‘Crystallization’’*: When the network decouples, it is interesting to ask how the decoupled SISO links behave (in terms of their fading statistics) when M and K grow large. The answer to this question follows from the concentration results in Theorems 1 and 2, which can be reformulated to establish upper bounds on the outage probability for the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links. For the sake of brevity, we focus on P1 in what follows. The goal is to arrive at a statement regarding

$$P_{\text{out},P1}(R) = \mathbb{P}\left\{\frac{1}{2} \log(1 + \text{SINR}_m^{P1}) \leq R\right\}$$

$$= \mathbb{P}\left\{\text{SINR}_m^{P1} \leq 2^{2R} - 1\right\}.$$

The corresponding result is summarized as follows.

Theorem 5 (Outage Probability for P1):

- 1) Assume that $K \geq 2, M \geq 2$, and $R \geq 0$ are such that

$$x(R) = \frac{1 - e_{P1}(M, K, R)}{\frac{16}{\underline{C}^2 \pi} \frac{M}{\sqrt{K}} + e_{P1}(M, K, R) \left(\frac{3}{\bar{C}_{SN}^2} \frac{1}{\sqrt{M}} + \frac{\sigma^2}{\bar{C}_{SN}^2} \frac{1}{\sqrt{K}} \right)} \geq 1 \quad (82)$$

where

$$e_{P1}(M, K, R) = \frac{16}{\pi^2} \frac{\bar{C}_{SN}^2}{\underline{C}^2} \frac{M^3}{K} (2^{2R} - 1).$$

Then, the individual link outage probability is upper-bounded as

$$P_{\text{out},P1}(R) \leq 151K^2 M e^{-\Delta_{P1} x(R)^{2/7}}. \quad (83)$$

- 2) Under the same conditions on K, M , and R as in 1), for any $\epsilon, \delta > 0, K \geq M^{3+\delta}$, and

$$R \leq \frac{1}{2} \log\left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{\bar{C}_{SN}^2} \frac{K}{M^3} (1 - \epsilon)\right), \quad (84)$$

we have

$$P_{\text{out},P1}(R) \leq \lim_{M, K \rightarrow \infty} 151K^2 M e^{-\Delta_{P1} x(R)^{2/7}} = 0.$$

Proof: We start with the proof of statement 1). Recall that Theorem 1 provides us with a parametric upper bound on

$\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L_{\text{P1}}(x)\}$ with $L_{\text{P1}}(x)$ defined in (54). Assuming that

$$x \leq \frac{C\pi\sqrt{K}}{16M} \tag{85}$$

and using $\overline{C}_{\text{SN}}^2 = \overline{C}^2 + \sigma^2(\overline{c}^2 + 1)$, we can lower-bound $L_{\text{P1}}(x)$ as

$$L_{\text{P1}}(x) \geq \frac{\pi^2}{16} \frac{\overline{C}^2}{\overline{C}_{\text{SN}}^2} \frac{K}{M^3} \frac{1 - \frac{16}{C\pi} \frac{M}{\sqrt{K}} x}{1 + \frac{3}{\overline{C}_{\text{SN}}^2} \frac{x}{\sqrt{M}} + \frac{\sigma^2}{\overline{C}_{\text{SN}}^2} \frac{x}{\sqrt{K}}} \triangleq L'_{\text{P1}}(x).$$

Solving

$$2^{2R} - 1 = L'_{\text{P1}}(x) \tag{86}$$

for $x(R)$ yields (82), which, by assumption, satisfies $x(R) \geq 1$. With

$$\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L'_{\text{P1}}(x)\} \leq \mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L_{\text{P1}}(x)\}$$

we can now apply¹² Theorem 1 to obtain

$$P_{\text{out,P1}}(R) \leq 151K^2Me^{-\Delta_{\text{P1}}x(R)^{2/7}}. \tag{87}$$

Finally, we note that $x(R)$ in (82) is trivially seen to satisfy (85). This concludes the proof of statement 1).

The proof of statement 2) is obtained by establishing a sufficient condition on $x(R)$, for any $R \geq 0$, to grow with increasing M (and by $K \geq M^{3+\delta}$ with increasing K). Using (82), it is easily verified that guaranteeing

$$0 \leq e_{\text{P1}}(M, K, R) \leq 1 - \epsilon$$

for some $0 < \epsilon < 1$ (independent of M, K) provides such a condition. The final result is now obtained by solving

$$e_{\text{P1}}(M, K, R) = \frac{16}{\pi^2} \frac{\overline{C}_{\text{SN}}^2}{\overline{C}^2} \frac{M^3}{K} (2^{2R} - 1) \leq 1 - \epsilon$$

for R . □

The implications of Theorem 5 are significant: For any transmission rate R less than the ergodic capacity (in the case $E_{k,m} = P_{m,k}$ for all k, m) or the ergodic capacity lower bound in Theorem 3 (in the case of general $E_{k,m}$ and $P_{m,k}$), the outage probability of each of the decoupled links goes to zero exponentially fast in the number of nodes in the network, provided K scales supercritically in M . We have thus shown that choosing the rate of growth of K fast enough for the network to decouple automatically guarantees that the decoupled SISO links converge to nonfading links. Equivalently, we can say that each of the decoupled links experiences a distributed spatial diversity (or, more precisely, relay diversity) order that goes to infinity as $M \rightarrow \infty$. Consequently, in the large- M limit time diversity (achieved by coding over a sufficiently long time horizon) is not needed to achieve ergodic capacity. We say that the network ‘‘crystallizes’’ as it breaks up into a set of effectively isolated ‘‘wires in the air.’’ From (83), we can furthermore infer the ‘‘crystallization’’ rate, i.e., the rate (as a function of M and K) at which the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links converge to nonfading links. We note, however, that the exponent $2/7$ (and

¹²Strictly speaking, one needs to use the upper bounds on $\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L_{\text{P1}}(x)\}$ derived in the last paragraph of Appendix C.

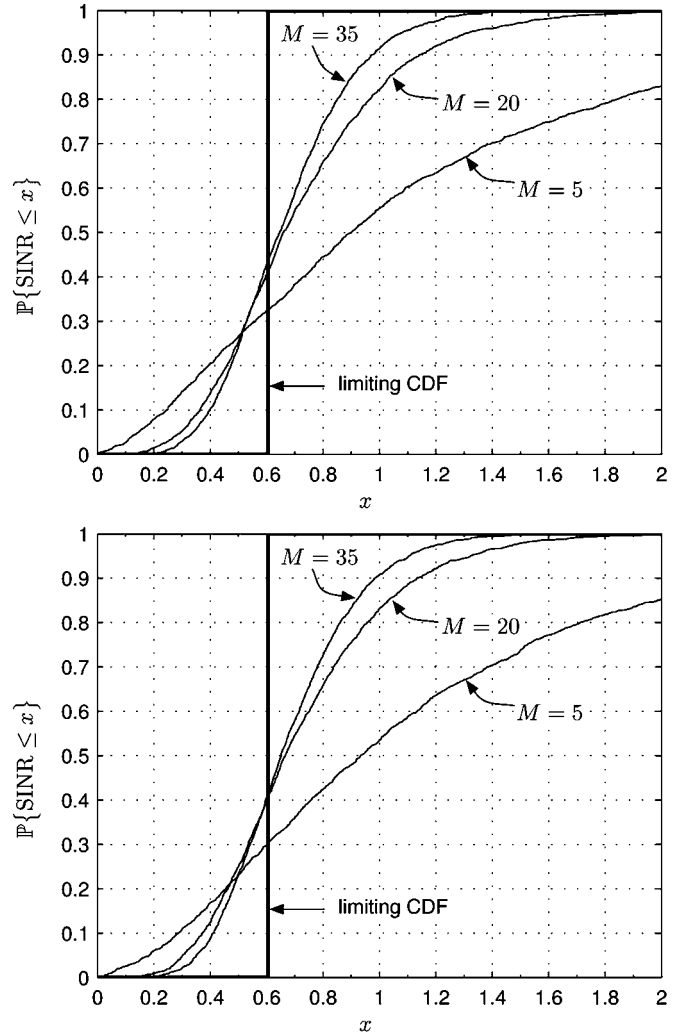


Fig. 3. Simulated (Monte Carlo) SINR CDFs for different values of M for (top) $K = M^3$ in P1 and (bottom) $K = M^2$ in P2.

$2/9$ for P2) is unlikely to be fundamental as it is probably a consequence of the application of the truncation technique. In this sense, we can only specify a guaranteed crystallization rate. We conclude by noting that the upper bound (87) (as well as the corresponding result for P2) tend to be rather loose. This is probably a consequence of the truncation technique and the use of union bounds to characterize the large-deviations behavior of the individual link SINR RVs.

3) *Numerical Results:* We shall finally provide numerical results quantifying the outage behavior of P1 and P2. For simplicity, we set $E_{k,m} = P_{m,k} = 1$ for all m, k , and $\sigma^2 = 0.01$ in both simulation examples. This choice for the path loss and shadowing parameters, although not representative of a real-world propagation scenario, isolates the dependence of our results on the network geometry. Moreover, it ensures that the distribution of the different SINR RVs for a given protocol is identical for all links so that it suffices to analyze the behavior of only one SINR RV for each of the two protocols. For $K = M^3$ in P1 and $K = M^2$ in P2, Fig. 3 shows the cumulative distribution functions (CDFs) (obtained through Monte Carlo simulation) of SINR^{P1} and SINR^{P2} , respectively, for different values

of M . We observe that, for increasing M , the CDFs approach a step function at the corresponding mean values, i.e., the SINR RVs, indeed, converge to a deterministic quantity, and, consequently, the underlying fading channel converges to a nonfading channel. The limiting mean values are given by the lower and upper bounds (which coincide in the case $E_{k,m} = P_{m,k} = 1$ for all m, k) in (60) and (61) for P1 and P2, respectively. We can furthermore see that for fixed M , the CDFs are very similar for P1 and P2 (recall, however, that $K = M^3$ in P1 and $K = M^2$ in P2), suggesting that the convergence behavior is similar for the two protocols. The difference in the theoretically predicted convergence exponents (2/7 for P1 and 2/9 for P2) therefore does not seem to be fundamental to the two protocols and may, indeed, be a consequence of our proof technique as already pointed out above.

C. Cooperation at the Relay Level

The analysis carried out so far was based on the assumption that the relays cannot cooperate. The purpose of this section is to investigate the impact of cooperation (in fact, a specific form of cooperation) at the relay level on the ergodic-capacity scaling behavior in the coherent case. Note that we continue to assume that the destination terminals cannot cooperate. Before proceeding, we would like to mention that concentration results and an outage analysis along the lines of the discussion in Sections III and IV-B are possible, but will be omitted for brevity of exposition.

Cooperation at the relay level will be accounted for by grouping the K single-antenna relay terminals into Q groups

$$\mathcal{G}_q \triangleq \left\{ \mathcal{R}_{(q-1)L+1}, \mathcal{R}_{(q-1)L+2}, \dots, \mathcal{R}_{qL} \right\}, \quad q \in [1:Q]$$

with L relays in each group¹³ and by assuming that the relays in each group can fully cooperate, but cooperation across groups is not possible. In order to simplify the exposition, in the remainder of this section, we think of a group \mathcal{G}_q ($q \in [1:Q]$) as a single relay element with L antenna elements and use the term “vector-relay (v-relay)” terminal to address the L -antenna relays $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_Q$. For $q \in [1:Q]$ and $m \in [1:M]$, the following notation will be used:

$$\begin{aligned} \mathbf{r}_q &\triangleq [r_{(q-1)L+1}, r_{(q-1)L+2}, \dots, r_{qL}]^T \\ \mathbf{t}_q &\triangleq [t_{(q-1)L+1}, t_{(q-1)L+2}, \dots, t_{qL}]^T \\ \mathbf{z}_q &\triangleq [z_{(q-1)L+1}, z_{(q-1)L+2}, \dots, z_{qL}]^T \\ \mathbf{h}_{q,m} &\triangleq [h_{(q-1)L+1,m}, h_{(q-1)L+2,m}, \dots, h_{qL,m}]^T \\ \mathbf{f}_{m,q} &\triangleq [f_{m,(q-1)L+1}, f_{m,(q-1)L+2}, \dots, f_{m,qL}]^T \end{aligned}$$

where \mathbf{r}_q and \mathbf{t}_q are the (L -dimensional) vector-valued signals received and transmitted by the q th v-relay, respectively, \mathbf{z}_q is additive noise at the q th v-relay, $\mathbf{h}_{q,m}$ contains the channel gains for the $\mathcal{S}_m \rightarrow \mathcal{G}_q$ link, and $\mathbf{f}_{m,q}$ contains the channel gains for the $\mathcal{G}_q \rightarrow \mathcal{D}_m$ link. Additionally, for simplicity, we assume that relays belonging to a given group q are located close to each other so that

$$\hat{E}_{q,m} \triangleq E_{(q-1)L+1,m} = E_{(q-1)L+2,m} = \dots = E_{qL,m}$$

¹³For simplicity, we assume that Q divides K so that $K = QL$.

$\hat{P}_{m,q} \triangleq P_{m,(q-1)L+1} = P_{m,(q-1)L+2} = \dots = P_{m,qL}$ for $q \in [1:Q]$ and $m \in [1:M]$. With this notation, the I-O relations (1) and (2) for the $\mathcal{S}_m \rightarrow \mathcal{G}_q$ links and the $\mathcal{G}_q \rightarrow \mathcal{D}_m$ links can be written as

$$\mathbf{r}_q = \sum_{m=1}^M \hat{E}_{q,m} \mathbf{h}_{q,m} s_m + \mathbf{z}_q, \quad q \in [1:Q]$$

and

$$y_m = \sum_{q=1}^Q \hat{P}_{m,q} \mathbf{f}_{m,q}^T \mathbf{t}_q + w_m, \quad m \in [1:M]$$

respectively. Next, we describe the generalization of the protocols P1 and P2 to the case of v-relays making the aspect of cooperation at the relay level explicit.

1) *P1 for the Cooperative Case:* Like in the case of single-antenna relays (described in Section III-A), we partition the Q v-relay terminals into M subsets \mathcal{M}_m ($m \in [1:M]$) with¹⁴ $|\mathcal{M}_m| = Q/M$. The v-relays (each of which has L antenna elements) in \mathcal{M}_m are assumed to assist the m th source-destination terminal pair $\{\mathcal{S}_m, \mathcal{D}_m\}$, and the relay partitioning function $p: [1, Q] \rightarrow [1, M]$ is defined as

$$p(q) \triangleq m \Leftrightarrow \mathcal{G}_q \in \mathcal{M}_m.$$

We assume that the q th v-relay terminal has perfect knowledge of the phases of the single-input multiple-output backward channel $\mathcal{S}_{p(q)} \rightarrow \mathcal{G}_q$ and the phases of the corresponding multiple-input single-output forward channel $\mathcal{G}_q \rightarrow \mathcal{D}_{p(q)}$. This implies that perfect knowledge of the vectors

$$\tilde{\mathbf{h}}_{q,p(q)} \triangleq \left[e^{j \arg([\mathbf{h}_{q,p(q)}]_1)}, e^{j \arg([\mathbf{h}_{q,p(q)}]_2)}, \dots, e^{j \arg([\mathbf{h}_{q,p(q)}]_L)} \right]^T$$

and

$$\tilde{\mathbf{f}}_{p(q),q} \triangleq \left[e^{j \arg([\mathbf{f}_{p(q),q}]_1)}, e^{j \arg([\mathbf{f}_{p(q),q}]_2)}, \dots, e^{j \arg([\mathbf{f}_{p(q),q}]_L)} \right]^T$$

is available at \mathcal{G}_q . The signal \mathbf{r}_q received at the q th v-relay terminal is phase-matched-filtered first w.r.t. the assigned backward channel $\mathcal{S}_{p(q)} \rightarrow \mathcal{G}_q$ and then w.r.t. the assigned forward channel $\mathcal{G}_q \rightarrow \mathcal{D}_{p(q)}$ followed by a normalization so that

$$\mathbf{t}_q = d_{P1,q} \tilde{\mathbf{f}}_{p(q),q}^* \left(\tilde{\mathbf{h}}_{q,p(q)}^H \mathbf{r}_q \right) \quad (88)$$

where¹⁵ the choice

$$d_{P1,q} \triangleq \frac{1}{L} \sqrt{P_{\text{rel}}} \times \left[\frac{Q}{M} \sum_{m=1}^M \hat{E}_{q,m} + \frac{\pi(L-1)Q}{4M} \hat{E}_{q,p(q)} + Q\sigma^2 \right]^{-1/2}$$

ensures that the per-v-relay power constraint

$$\mathbb{E} \left[\|\mathbf{t}_q\|^2 \right] = P_{\text{rel}}/Q, \quad q \in [1:Q]$$

¹⁴For simplicity, we assume that M divides Q .

¹⁵The quantity $d_{P1,q}$, used in this section is (for $L > 1$) different from $d_{P1,k}$ defined in (5). We use the same symbol for notational simplicity and employ the index q (instead of k) consistently, in order to resolve potential ambiguities. The same comment applies to other variables redefined in this section.

and consequently the total (across v-relays) power constraint

$$\sum_{q=1}^Q \mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}$$

is met. As in the single-antenna relay (i.e., noncooperative) case, P1 ensures that the relays $\mathcal{G}_q \in \mathcal{M}_m$ forward the signal intended for \mathcal{D}_m in a “doubly coherent” (w.r.t. the assigned backward and forward channel) fashion whereas the signals transmitted by the source terminals $\mathcal{S}_{\hat{m}}$ with $\hat{m} \neq m$ are forwarded to \mathcal{D}_m in a “noncoherent” fashion (i.e., phase incoherence occurs either on the backward or the forward link or on both links). From (88), we can see that cooperation in groups of L single-antenna relays is realized by phase combining on the backward and forward links of each v-relay. More sophisticated forms of cooperation such as equalization on the backward link and precoding on the forward link are certainly possible, but are beyond the scope of this paper.

2) *P2 for the Cooperative Case:* Like in the case of single-antenna relays (i.e., the noncooperative case), P2 requires that each relay, in fact here v-relay, knows the phases of all its M vector-valued backward and forward channels, i.e., \mathcal{G}_q needs knowledge of $\tilde{\mathbf{h}}_{q,m}$ and $\tilde{\mathbf{f}}_{m,q}$, respectively, for $m \in [1:M]$. The relay processing stage in P2 computes

$$\mathbf{t}_q = d_{\text{P2},q} \left(\sum_{m=1}^M \tilde{\mathbf{f}}_{m,q}^* \tilde{\mathbf{h}}_{q,m}^H \right) \mathbf{r}_q$$

where

$$d_{\text{P2},q} \triangleq \frac{1}{L} \sqrt{P_{\text{rel}}} \times \left[Q \sum_{m=1}^M \hat{E}_{q,m} + \frac{\pi(L-1)Q}{4M} \sum_{m=1}^M \hat{E}_{q,m} + MQ\sigma^2 \right]^{-1/2}$$

ensures that the per-v-relay power constraint

$$\mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}/Q, \quad q \in [1:Q]$$

and, consequently, the total (across relays) power constraint

$$\sum_{q=1}^Q \mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}$$

is met.

3) *Ergodic-Capacity Results:* We are now ready to establish the impact of cooperation at the relay level on the ergodic capacity scaling laws for P1 and P2. Our results are summarized in Theorems 6 and 7 as follows.

Theorem 6 (Ergodic Capacity of P1 With Cooperation): Suppose that destination terminal \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the mean of the effective channel gain of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/4)L^2 \sum_{q:p(q)=m} d_{\text{P1},q} \hat{P}_{m,q} \hat{E}_{q,m}$. Then, for any $\epsilon, \delta > 0$, there exist $M_0, Q_0 > 0$ such that for

all $M \geq M_0$ and $Q \geq Q_0$, the per source–destination terminal pair capacity achieved by P1 satisfies¹⁶

$$\frac{1}{2} \log \left(1 + \frac{\pi^2 QL^2}{16 M^3} \frac{\bar{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon) \right) \leq C_{\text{P1}} \leq \frac{1}{2} \log \left(1 + \frac{\pi^2 \max[Q, M^{2+\delta}] L^2}{16 M^3} \frac{\bar{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon) \right). \quad (89)$$

Theorem 7 (Ergodic Capacity of P2 With Cooperation): Suppose that destination terminal \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the mean of the effective channel gain of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/4)L^2 \sum_{q=1}^Q d_{\text{P2},q} \hat{P}_{m,q} \hat{E}_{q,m}$. Then, for any $\epsilon, \delta > 0$, there exist M_0, Q_0 such that for all $M \geq M_0, Q \geq Q_0$ the per source–destination terminal pair capacity achieved by P2 satisfies

$$\frac{1}{2} \log \left(1 + \frac{\pi^2 QL^2}{16 M^2} \frac{\bar{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon) \right) \leq C_{\text{P2}} \leq \frac{1}{2} \log \left(1 + \frac{\pi^2 \max[Q, M^{1+\delta}] L^2}{16 M^2} \frac{\bar{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon) \right). \quad (90)$$

Proof of Theorems 6 and 7: The upper bounds in (89) and (90) are again established based on a concentration result for the individual link SINRs and the lower bounds build on the technique summarized in Appendix E. The proofs of Theorems 6 and 7 are almost identical to the proofs of Theorems 3 and 4, respectively, and do not require new techniques. There is, however, one important aspect in which Theorems 6 and 7 differ from Theorems 3 and 4, namely, the appearance of the factor L^2 in (89) and (90). To demonstrate where this factor comes from, we provide the proof of the ergodic capacity lower bound for P1 in Appendix D. The proofs of the remaining statements will be omitted for brevity of exposition. \square

Discussion of Results: Just like in the noncooperative (i.e., single-antenna relay) case, we can conclude that asymptotically in M if $K \propto M^{3+\alpha}$ in P1 and $K \propto M^{2+\alpha}$ in P2 with $\alpha > 0$, the network decouples.

The effect of cooperation (through phase matched-filtering) at the relay level manifests itself in the presence of the factor L^2 inside the log in the bounds for C_{P1} and C_{P2} stated in Theorems 6 and 7, respectively. We can summarize the results of Theorems 6 and 7 as¹⁷

$$C_{\text{P1}} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{QL^2}{M^3} \right) \right) \\ C_{\text{P2}} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{QL^2}{M^2} \right) \right).$$

We can, therefore, conclude that the per-stream array gain A is given by $A_{\text{P1}} = QL^2/M^3$ for P1 and $A_{\text{P2}} = QL^2/M^2$

¹⁶Note that the quantities $\bar{C}_{\text{SN}}, \underline{C}, \bar{C}$, and $\underline{C}_{\text{SN}}$ used in this section have been defined in Section III.

¹⁷Note that we use the $\Theta(\cdot)$ notation only to hide the dependence on $\underline{E}, \bar{E}, \underline{P}$, and \bar{P} . Strictly speaking, as L is finite it should also be hidden under the $\Theta(\cdot)$ notation. However, our goal is to exhibit the impact of cooperation at the relay level on C_{P1} and C_{P2} , which is the reason for making the dependence on L explicit.

for P2. On a conceptual level, the array gain can be decomposed into a contribution due to distributed array gain A_d and a contribution due to cooperation at the relay level (realized by phase matching on backward and forward links) A_c , i.e., $A = A_d A_c$ with $A_{d,P1} = QL/M^3$, $A_{d,P2} = QL/M^2$, and $A_{c,P1} = A_{c,P2} = L$. To illustrate the impact of cooperation at the relay level, we compare a network with K noncooperating single-antenna relays to a network with a total of $K = QL$ single-antenna relays cooperating in groups of L single-antenna relays. In the case where there is no cooperation at the relay level, we have

$$C_{P1}^{(nc)} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{K}{M^3} \right) \right)$$

whereas if the relays cooperate in groups of L single-antenna relays, we get

$$C_{P1}^{(c)} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{KL}{M^3} \right) \right).$$

Cooperation at the relay level (realized by phase matched-filtering) in groups of L single-antenna relays therefore yields an L -fold increase in the effective per-stream SINR due to additional array gain given by $A_c = L$. Equivalently, the total number of single-antenna relays needed to achieve a given per source–destination terminal pair capacity is reduced by a factor of L through cooperation in groups of L single-antenna relay elements. The conclusions for P2 are identical.

As already pointed out earlier, the network decouples into effectively isolated source–destination pair links for any finite $L > 1$. Even though a concentration analysis along the lines of Theorems 1 and 2 was not performed (for the sake of brevity), it can be shown that for finite $L > 1$, the individual links converge to nonfading links as $M, Q \rightarrow \infty$, provided that Q scales supercritically as a function of M .

Numerical Example: We conclude this section with a numerical example that demonstrates the impact of cooperation at the relay level, where we use the same parameters as in the simulation examples at the end of Section IV-B. Fig. 4 shows the SINR CDF for P1 with $L = 4$ and $QL = M^3$ (the case $L = 1$ shown in Fig. 3 is included for reference). We observe that, as pointed out above, for increasing M , we, indeed, get convergence of the fading link to a nonfading link. Moreover, we can also see that increasing L for fixed M results in higher per source–destination terminal pair capacity, but at the same time slows down convergence (w.r.t. M and hence also Q) of the link SINRs to their deterministic limits.

V. NONCOHERENT (AF) RELAY NETWORKS

So far, we have considered coherent relay networks, where each relay terminal knows its assigned backward and forward channels (P1) or all backward and forward channels (P2) perfectly. In the following, we relax this assumption and study networks with no CSI at the relay terminals, i.e., noncoherent relay networks. In particular, we investigate a simple AF architecture where the relay terminals, in the second time slot, forward

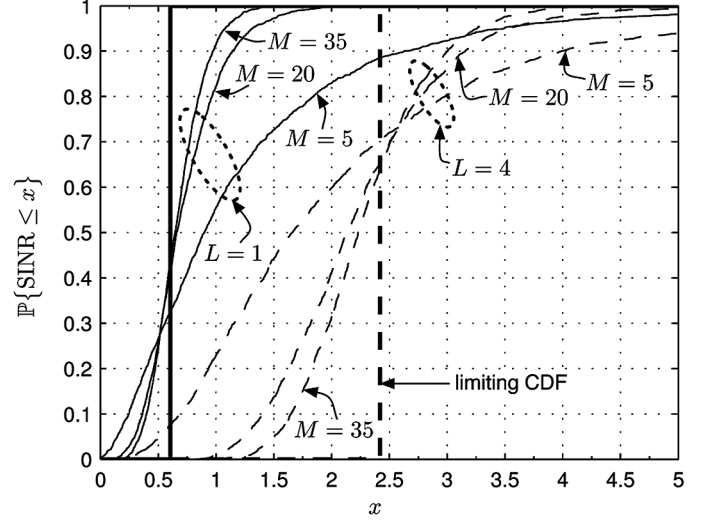


Fig. 4. Simulated (Monte Carlo) SINR CDFs for different values of M for $QL = M^3$ in P1 with $L = 1$ and $L = 4$.

(without additional processing) a scaled version of the signal received in the first time slot. As already mentioned in Section II, the source terminals do not have CSI. The destination terminals cooperate and perform joint decoding. The assumptions on CSI at the destination terminals will be made specific in Section V-B.

A. The AF Protocol

Throughout this section, we assume that $E_{k,m} = P_{m,k} = 1$ for all $m \in [1 : M]$, $k \in [1 : K]$. This assumption is conceptual as the technique used to derive the main result in this section does not seem to be applicable for general $E_{k,m}$ and $P_{m,k}$. On the other hand, the results in this section do not require \mathbf{H} and \mathbf{F} to have Gaussian entries. Upon reception of r_k , the k th relay terminal simply scales the received signal to obtain $t_k = (d/\sqrt{K})r_k$. Choosing $d = \sqrt{P_{\text{rel}}/(1 + \sigma^2)}$ ensures that the per-relay power constraint $\mathbb{E}[|t_k|^2] \leq P_{\text{rel}}/K$ and, hence, the total power constraint $\mathbb{E}[|\mathbf{t}|^2] \leq P_{\text{rel}}$ is met.

With these assumptions, inserting (1) into (2), we get the following I–O relation:

$$\mathbf{y} = \frac{d}{\sqrt{K}} \mathbf{F} \mathbf{H} \mathbf{s} + \frac{d}{\sqrt{K}} \mathbf{F} \mathbf{z} + \mathbf{w}. \quad (91)$$

In the remainder of this section, we assume that the *jointly decoding* destination terminals have access to the realizations of \mathbf{H} and \mathbf{F} . In fact, as shown in the subsequent analysis, knowledge of $\mathbf{F} \mathbf{H}$ and \mathbf{F} is sufficient.

B. Capacity of the AF Protocol

Based on the I–O relation (91), we shall next study the behavior of $I(\mathbf{y}; \mathbf{s} | \mathbf{F} \mathbf{H}, \mathbf{F})$ when $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$. We start by noting that

$$I(\mathbf{y}; \mathbf{s} | \mathbf{F} \mathbf{H}, \mathbf{F}) = \log \det \left(\mathbf{I} + \frac{d^2}{\sigma^2 M K} \mathbf{H}^H \mathbf{F}^H \left(\frac{d^2}{K} \mathbf{F} \mathbf{F}^H + \mathbf{I} \right)^{-1} \mathbf{F} \mathbf{H} \right).$$

Since the destination terminals perform joint decoding, the ergodic capacity per source–destination terminal pair is given by

$$C_{\text{AF}} = \frac{1}{2} \mathbb{E} \left[\frac{1}{M} \sum_{k=1}^K \log \left(1 + \frac{1}{\sigma^2} \lambda_k \left(\frac{1}{M} \mathbf{H} \mathbf{H}^H \mathbf{T} \right) \right) \right] \quad (92)$$

where

$$\mathbf{T} \triangleq \frac{d^2}{K} \mathbf{F}^H \left(\mathbf{I} + \frac{d^2}{K} \mathbf{F} \mathbf{F}^H \right)^{-1} \mathbf{F}$$

and the factor 1/2 in (92) results from the fact that data is transmitted over two time slots.

C. Asymptotic Capacity Behavior

To compute C_{AF} in the $M, K \rightarrow \infty$ limit with $K/M \rightarrow \beta$, we start by analyzing the corresponding asymptotic behavior of $\lambda_k((1/M)\mathbf{H}\mathbf{H}^H\mathbf{T})$. To this end, we define the empirical spectral distribution (ESD) of a matrix (random or deterministic).

Definition 1: Let $\mathbf{X} \in \mathbb{C}^{N \times N}$ be a Hermitian matrix. The ESD of \mathbf{X} is defined as

$$F_{\mathbf{X}}^N(x) \triangleq \frac{1}{N} \sum_{n=1}^N I[\lambda_n(\mathbf{X}) \leq x].$$

For random \mathbf{X} , the quantity $F_{\mathbf{X}}^N(x)$ is random as well, i.e., it is an RV for each x . In the following, our goal is to prove the convergence (in the sense defined below), when $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$, and $\beta \in (0, \infty)$, of $F_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}^K(x)$ to a deterministic limit and to find the corresponding limiting eigenvalue distribution.

Definition 2: We say that the ESD $F_{\mathbf{X}}^N(x)$ of a random Hermitian matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$ converges almost surely (a.s.) to a deterministic limiting function $F_{\mathbf{X}}(x)$, when $N \rightarrow \infty$, if for any $\epsilon > 0$ there exists an $N_0 > 0$ such that for all $N \geq N_0$ a.s.

$$\sup_{x \in \mathbb{R}} |F_{\mathbf{X}}^N(x) - F_{\mathbf{X}}(x)| \leq \epsilon.$$

To prove the convergence of $F_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}^K(x)$ to a deterministic limiting function, we start by analyzing $F_{\mathbf{T}}^K(x)$.

Lemma 1: For $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$, the ESD $F_{\mathbf{T}}^K(x)$ converges a.s. to a nonrandom limiting distribution $F_{\mathbf{T}}(x)$ with corresponding density given by¹⁸

$$\begin{aligned} f_{\mathbf{T}}(x) &= \frac{\sqrt{(1+\gamma_1)(1+\gamma_2)}}{2\pi d^2 x(1-x)^2} \\ &\times \sqrt{\left(\frac{\gamma_2}{1+\gamma_2} - x \right)^+ \left(x - \frac{\gamma_1}{1+\gamma_1} \right)^+} \\ &+ \left[1 - \frac{1}{\beta} \right]^+ \delta(x) \end{aligned} \quad (93)$$

where $\gamma_1 \triangleq d^2(1 - 1/\sqrt{\beta})^2$ and $\gamma_2 \triangleq d^2(1 + 1/\sqrt{\beta})^2$.

¹⁸Note that (93) implies that $f_{\mathbf{T}}(x)$ is compactly supported in the interval $[\gamma_1/(1+\gamma_1), \gamma_2/(1+\gamma_2)]$.

Proof: We start with the singular value decomposition

$$\frac{d}{\sqrt{K}} \mathbf{F} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}$$

where the columns of $\mathbf{U} \in \mathbb{C}^{M,M}$ are the eigenvectors of the matrix $(d^2/K)\mathbf{F}\mathbf{F}^H$, the columns of $\mathbf{V}^H \in \mathbb{C}^{K,K}$ are the eigenvectors of $(d^2/K)\mathbf{F}^H\mathbf{F}$, and the matrix $\mathbf{\Sigma} \in \mathbb{R}^{M,K}$ contains $R = \min(M, K)$ nonzero entries $\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{RR}$, which are the positive square roots of the nonzero eigenvalues of the matrix $(d^2/K)\mathbf{F}\mathbf{F}^H$. Defining $\mathbf{\Lambda} \triangleq \mathbf{\Sigma} \mathbf{\Sigma}^H \in \mathbb{R}^{M,M}$, we have

$$\mathbf{T} = \mathbf{V}^H \mathbf{\Sigma}^H (\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{\Sigma} \mathbf{V}.$$

By inspection, it follows that

$$F_{\Sigma^H(\mathbf{I}+\mathbf{\Lambda})^{-1}\Sigma}^K(x) = \frac{M}{K} F_{\mathbf{\Lambda}}^M \left(\frac{x}{1-x} \right) + \left(1 - \frac{M}{K} \right) u(x). \quad (94)$$

As $F_{\mathbf{\Lambda}}^M(x) = F_{(d^2/K)\mathbf{F}\mathbf{F}^H}^M(x)$, by the Marčenko–Pastur law (see Theorem 12 in Appendix F), we conclude that $F_{\mathbf{\Lambda}}^M(x)$ converges a.s. to a limiting nonrandom distribution $F_{\mathbf{\Lambda}}(x)$ with corresponding density

$$f_{\mathbf{\Lambda}}(x) = \frac{\beta}{2\pi x d^2} \sqrt{(\gamma_2 - x)^+ (x - \gamma_1)^+} + [1 - \beta]^+ \delta(x). \quad (95)$$

From (94) we can, therefore, conclude that $F_{\Sigma^H(\mathbf{I}+\mathbf{\Lambda})^{-1}\Sigma}^K(x)$ converges a.s. to a nonrandom limit given by

$$F_{\Sigma^H(\mathbf{I}+\mathbf{\Lambda})^{-1}\Sigma}(x) = \frac{1}{\beta} F_{\mathbf{\Lambda}} \left(\frac{x}{1-x} \right) + \left(1 - \frac{1}{\beta} \right) u(x). \quad (96)$$

Taking the derivative w.r.t. x on both sides of (96), the density corresponding to $F_{\Sigma^H(\mathbf{I}+\mathbf{\Lambda})^{-1}\Sigma}(x)$ is obtained as

$$\begin{aligned} f_{\Sigma^H(\mathbf{I}+\mathbf{\Lambda})^{-1}\Sigma}(x) \\ = \frac{1}{\beta} f_{\mathbf{\Lambda}} \left(\frac{x}{1-x} \right) \frac{1}{(1-x)^2} + \left(1 - \frac{1}{\beta} \right) \delta(x). \end{aligned} \quad (97)$$

We obtain the final result in (93) now by noting that $f_{\mathbf{T}}(x) = f_{\Sigma^H(\mathbf{I}+\mathbf{\Lambda})^{-1}\Sigma}(x)$ because of the unitarity of \mathbf{V} and by inserting (95) into (97) and carrying out straightforward algebraic manipulations. \square

Based on Lemma 1, we can now apply Theorem 11 (Appendix F) to conclude that $F_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}^K(x)$ converges a.s. to a deterministic function $F_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ as $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$. The corresponding limiting density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ is obtained through the application of the Stieltjes inversion formula (151) to the solution of the fixed-point equation

$$G(z) = \underbrace{\int_{-\infty}^{\infty} \frac{f_{\mathbf{T}}(x) dx}{x(1-\beta-\beta z G(z)) - z}}_I, \quad z \in \mathbb{C}^+ \quad (98)$$

in the set

$$\{G(z) \in \mathbb{C} \mid -(1-\beta)/z + \beta G(z) \in \mathbb{C}^+\}, \quad z \in \mathbb{C}^+ \quad (99)$$

where we used the symbol $G(z)$ to denote the Stieltjes transform $G_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(z)$. In the following, for brevity, we write

G instead of $G(z)$. To solve (98), we first compute the integral I on the RHS of (98). We substitute $f_{\mathbf{T}}(x)$ from (93) into (98) and define

$$\eta_1 \triangleq \frac{\gamma_1}{1 + \gamma_1}, \quad \eta_2 \triangleq \frac{\gamma_2}{1 + \gamma_2}, \quad \rho \triangleq \frac{\sqrt{(1 + \gamma_1)(1 + \gamma_2)}}{2\pi d^2}$$

to obtain

$$I = -\frac{1}{z} \left[1 - \frac{1}{\beta} \right]^+ + \frac{1}{z} \underbrace{\int_{\eta_1}^{\eta_2} \frac{\rho \sqrt{(\eta_2 - x)(x - \eta_1)} dx}{x(1-x)^2 \left(x \left(\frac{1-\beta}{z} - \beta G \right) - 1 \right)}_{\hat{I}}. \quad (100)$$

The integral \hat{I} is computed in Appendix G. Employing the notation introduced in Appendix G, we can finally write the fixed-point equation (98) as

$$Gz = - \left[1 - \frac{1}{\beta} \right]^+ + \chi \left(A_1 \hat{I}_1 + A_2 \hat{I}_2 + A_3 \hat{I}_3 + A_4 \hat{I}_4 \right). \quad (101)$$

It is tedious, but straightforward, to show that for any $\beta > 0$

$$- \left[1 - \frac{1}{\beta} \right]^+ + \chi A_1 \hat{I}_1 = -\frac{\beta - 1}{2\beta}$$

so that (101) can be written as

$$Gz + \frac{\beta - 1}{2\beta} - \chi A_2 \hat{I}_2 - \chi A_3 \hat{I}_3 = \chi A_4 \hat{I}_4. \quad (102)$$

Next, multiplying (102) by $2d^2\beta(G\beta z + z + \beta - 1)^2$, squaring both sides, and introducing the auxiliary variable

$$\hat{G} \triangleq -\frac{1-\beta}{z} + \beta G$$

we obtain after straightforward, but tedious, manipulations that \hat{G} must satisfy the following quartic equation:

$$\hat{G}^4 + a_3 \hat{G}^3 + a_2 \hat{G}^2 + a_1 \hat{G} + a_0 = 0 \quad (103)$$

with the coefficients

$$a_3 = \frac{1}{z}(2z - \beta + 1) \quad a_2 = \frac{1}{z} \left(z - \beta + 3 - \frac{\beta}{d^2} \right)$$

$$a_1 = \frac{1}{z^2} \left(2z - \beta + 1 - \frac{\beta}{d^2} \right) \quad a_0 = \frac{1}{z^2}.$$

The quartic equation (103) can be solved analytically. The resulting expressions are, however, very lengthy, do not lead to interesting insights, and will therefore be omitted. It is important to note, however, that (103) has two pairs of complex conjugate roots. The solutions of (103) will henceforth be denoted as \hat{G}_1 , \hat{G}_1^* , \hat{G}_2 , and \hat{G}_2^* . We recall that our goal is to find the unique solution G of the fixed point equation (98) such that $\hat{G} = -(1 - \beta)/z + \beta G \in \mathbb{C}^+$ for all $z \in \mathbb{C}^+$. Therefore, in each point $z \in \mathbb{C}^+$, we can immediately eliminate the two solutions (out of the four) that have a negative imaginary part. In practice, this can be done conveniently by constructing the functions

$$\hat{G}'_1 \triangleq \Re \hat{G}_1 + j \Im \hat{G}_1 \quad \text{and} \quad \hat{G}'_2 \triangleq \Re \hat{G}_2 + j \Im \hat{G}_2$$

which can be computed analytically, satisfy (103), and are in \mathbb{C}^+ for any $z \in \mathbb{C}^+$. Next, note that (102) has a unique solution in the set (99), which is also the unique solution of (98). We can obtain this solution $G(z)$, $z \in \mathbb{C}^+$, by substituting

$$G_1 = (1/\beta)(\hat{G}'_1 - (\beta - 1)/z) \quad \text{and} \quad G_2 = (1/\beta)(\hat{G}'_2 - (\beta - 1)/z)$$

into (102) and checking which of the two satisfies the equation. Unfortunately, it seems that this verification cannot be formalized in the sense of identifying the unique solution of (102) in analytic form. The primary reason for this is that to check *algebraically* if G_1 and G_2 satisfy (102), we have to perform a noninvertible transformation (squaring) of (102), which doubles the number of solutions of this equation, and results in G_1 and G_2 both satisfying the resulting formula. The second reason is that depending on the values of the parameters $\beta > 0$, $d > 0$, the correct solution is either G_1 or G_2 , and the dependence between G_1 , G_2 , β , and d has a complicated structure. Starting from the analytical expressions for G_1 and G_2 , we can identify, however, for any fixed $\beta > 0$, $d > 0$, the density function $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x) = (1/\pi) \lim_{y \rightarrow 0^+} \Im[G(x + jy)]$ corresponding to the unique solution of (102) [and hence of (98)] numerically. This is accomplished as follows. We know that, for given x , $\lim_{y \rightarrow 0^+} \Im[G(x + jy)]$ is either equal to

$$L_1(x) \triangleq \lim_{y \rightarrow 0^+} \Im[G_1(x + jy)]$$

or

$$L_2(x) \triangleq \lim_{y \rightarrow 0^+} \Im[G_2(x + jy)].$$

Even though the functions $L_1(x)$ and $L_2(x)$ can be computed analytically (with the resulting expressions being very lengthy and involved), it seems that for any fixed $x > 0$, the correct choice between the values $L_1(x)$ and $L_2(x)$ can only be made numerically. The following algorithm constitutes one possibility to solve this problem.

Algorithm—Choice of the Limit

Input: $x > 0$

- 1) Choose a small enough $y > 0$
- 2) Substitute $G_1(x + jy)$ and $G_2(x + jy)$ into (102)
- 3) If $G_1(x + jy)$ satisfies (102), then

return $L_1(x)$

otherwise

return $L_2(x)$

This algorithm includes a heuristic element. The following comments are therefore in order.

- In Step 1 of the algorithm, the choice of y cannot be formalized in the sense of giving an indication of how small it has to be as a function of β and d . On the one hand, y has to be strictly greater than zero, because (102) in general holds in \mathbb{C}^+ only and does not need to hold either for $G_1(x + j0)$ or for $G_2(x + j0)$. On the other hand, y should

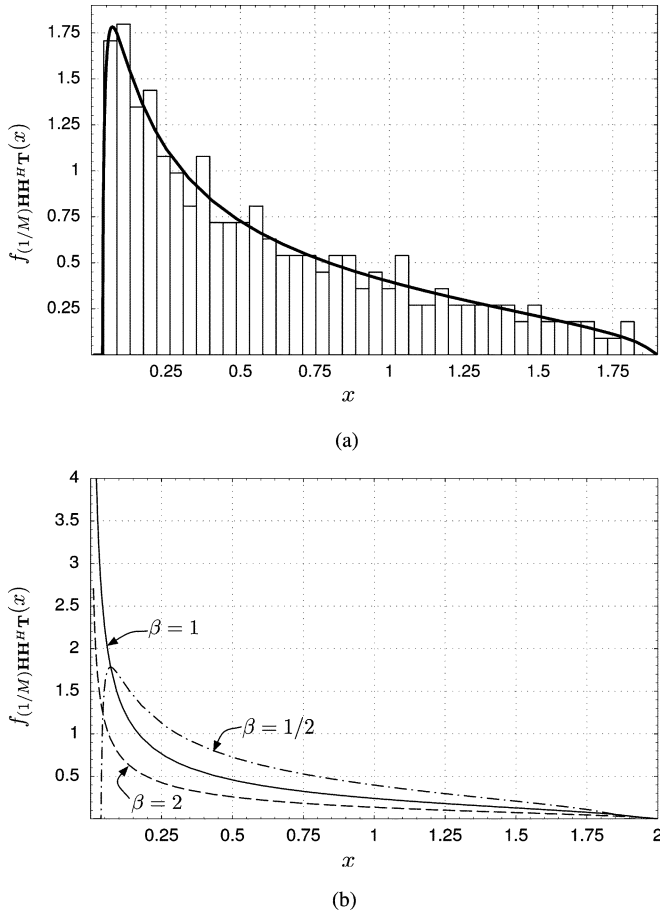


Fig. 5. Limiting density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ (a) for $\beta = 1/2$ and $d = 1$ along with its histogram (Monte Carlo) and (b) for different values of $\beta = 2, 1, 1/2$, and $d = 1$.

be small enough for $G_1(x + jy)$ to be close to $L_1(x)$ and $G_2(x + jy)$ to be close to $L_2(x)$. The correctness of the output of the algorithm is justified by the fact that $G(z)$ is analytic in \mathbb{C}^+ (see Definition 3 in Appendix F).

- In Step 3, the check whether $G_1(x + jy)$ satisfies (102) is performed numerically. Therefore, rounding errors will arise. It turns out, however, that in practice, unless $|L_1(x) - L_2(x)|$ is very small (in this case it does not matter which of the two values we choose), the solution of (102) yields a clear indication of whether $G_1(x + jy)$ or $G_2(x + jy)$ is the correct choice.
- To compute the density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ using the proposed algorithm, we need to run Steps 1–3 for every x . It will be proved later that $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ is always compactly supported and bounds for its support will be given in analytic form (as a function of β and d). Since the algorithm consists of very basic arithmetic operations only, it is very fast and can easily be run on a dense grid inside the support region of $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$.

As an example, for $d = 1$ and $\beta = 1/2$, Fig. 5(a) shows the density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ obtained by the algorithm formulated above along with the histogram of the same density obtained through Monte Carlo simulation. We can see that the two curves match very closely and that our method allows to obtain a much more refined picture of the limiting density. Fig. 5(b) shows the

density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ for $\beta = 2, 1, 1/2$ obtained through our algorithm. We can see that the density function is always compactly supported.

The final step in computing the asymptotic capacity of the AF relay network is to take the limit $K, M \rightarrow \infty$ with $K/M \rightarrow \beta$ in (92) and to evaluate the resulting integral

$$C_{\text{AF}}^\beta \triangleq \frac{\beta}{2} \int_0^\infty \log\left(1 + \frac{x}{\sigma^2}\right) f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x) dx \quad (104)$$

numerically. The evaluation of (104) is drastically simplified if we consider that $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ is compactly supported. The corresponding interval boundaries (or, more specifically, bounds thereon) can be computed analytically as a function of β and d . We start by noting that the second part of Theorem 12 in Appendix F implies that a.s.

$$\lim_{M \rightarrow \infty} \lambda_{\max}\left((1/M)\mathbf{H}\mathbf{H}^H\right) = (1 + \sqrt{\beta})^2.$$

From (97) and Theorem 12, it follows that a.s.

$$\lambda_{\max}(\mathbf{T}) = d^2(1 + \sqrt{\beta})^2 / (\beta + d^2(1 + \sqrt{\beta})^2).$$

For any realization of \mathbf{H} and \mathbf{T} and any M, K , by the submultiplicativity of the spectral norm, we have

$$\lambda_{\max}\left((1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}\right) \leq \lambda_{\max}\left((1/M)\mathbf{H}\mathbf{H}^H\right) \lambda_{\max}(\mathbf{T})$$

which implies that for $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ a.s.

$$\lambda_{\max}\left((1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}\right) \leq \frac{d^2(1 + \sqrt{\beta})^4}{\beta + d^2(1 + \sqrt{\beta})^2} \triangleq x_{\max}.$$

We can thus conclude that $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ is compactly supported on the interval¹⁹ $[0, x_{\max}]$. Consequently, the integral in (104) becomes

$$C_{\text{AF}}^\beta = \frac{\beta}{2} \int_0^{x_{\max}} \log\left(1 + \frac{x}{\sigma^2}\right) f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x) dx$$

which we can compute numerically, using any standard method for numerical integration and employing the algorithm described above to evaluate $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ at the required grid points. Using this procedure, we computed C_{AF}^β as a function of β for $d = 1$ with the result depicted in Fig. 6. We can see that for $\beta < 1$ (i.e., $K < M$), C_{AF}^β increases very quickly with β , which is because the corresponding effective MIMO channel matrix builds up rank and hence spatial multiplexing gain. For $\beta > 1$ (i.e., $K > M$), when the effective MIMO channel matrix is already full rank with high probability, the curve flattens out and for $\beta \rightarrow \infty$, the capacity C_{AF}^β seems to converge to a finite value. In Section V-D, we prove that C_{AF}^β indeed converges to a finite limit as $\beta \rightarrow \infty$. This result has an interesting interpretation as it allows to relate the AF relay network to a point-to-point MIMO channel.

D. Convergence to Point-to-Point MIMO Channel

In [1], it was shown that for finite M , as $K \rightarrow \infty$, the two-hop AF relay network capacity converges to half the capacity of a point-to-point MIMO link; the factor 1/2 penalty comes from the fact that communication takes place over two time slots.

¹⁹The actual supporting interval of $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ may, in fact, be smaller.

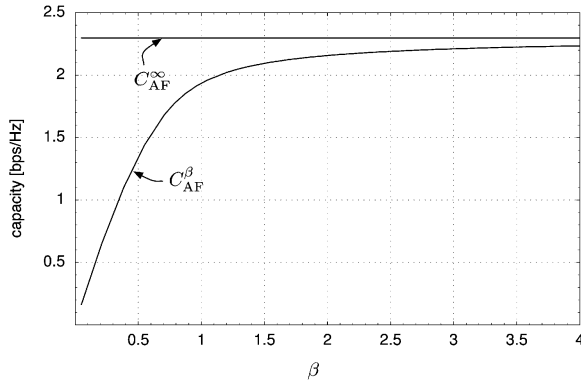


Fig. 6. Capacity C_{AF}^β as a function of β for $d = 1$ and $\sigma^2 = 0.01$.

In the following, we demonstrate that the result in [1] can be generalized to the $M, K \rightarrow \infty$ case. More specifically, we show that for $\beta \rightarrow \infty$, the asymptotic ($M, K \rightarrow \infty$) capacity of the two-hop AF relay network is equal to half the asymptotic ($M \rightarrow \infty$) capacity of a point-to-point MIMO channel with M transmit and M receive antennas. We start by dividing (103) by β and taking the limit²⁰ $\beta \rightarrow \infty$, which yields the quadratic equation

$$z\hat{G}^2 + z\left(1 + \frac{1}{d^2}\right)\hat{G} + \left(1 + \frac{1}{d^2}\right) = 0. \quad (105)$$

The two solutions of (105) are given by

$$\hat{G}_{1,2}(z) = \frac{-z\left(1 + \frac{1}{d^2}\right) \pm \sqrt{z^2\left(1 + \frac{1}{d^2}\right)^2 - 4z\left(1 + \frac{1}{d^2}\right)}}{2z}. \quad (106)$$

Applying the Stieltjes inversion formula (151) to (106) and choosing the solution that yields a positive density function, we obtain

$$\begin{aligned} & \beta f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x) \\ &= \frac{1}{\pi} \lim_{y \rightarrow 0^+} \Im[\beta G(x + jy)] \\ &= \frac{1}{\pi} \lim_{y \rightarrow 0^+} \Im[\hat{G}(x + jy)] \\ &= \frac{1}{2\pi x} \sqrt{\left[4x\left(1 + \frac{1}{d^2}\right) - x^2\left(1 + \frac{1}{d^2}\right)^2\right]^+}. \end{aligned} \quad (107)$$

Inserting (107) into (104) and changing the integration variable according to $u \triangleq x\left(1 + \frac{1}{d^2}\right)$, we find that $C_{AF}^\beta \xrightarrow{\beta \rightarrow \infty} C_{AF}^\infty$, where

$$C_{AF}^\infty \triangleq \frac{1}{4\pi} \int_0^4 \sqrt{\frac{4}{u} - 1} \log\left(1 + \frac{d^2}{(d^2 + 1)\sigma^2} u\right) du. \quad (108)$$

Comparing (108) with [31, eq. (13)], it follows that for $\beta \rightarrow \infty$ the asymptotic $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ per source–destination terminal pair capacity in the two-hop AF relay network is equal to half the asymptotic ($M \rightarrow \infty$) per-antenna capacity in

²⁰It is important that first we take the limit $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ and afterwards let $\beta \rightarrow \infty$.

a point-to-point MIMO link with M transmit and M receive antennas, provided the signal-to-noise ratio (SNR) in the relay case is defined as $\text{SNR} \triangleq d^2 / ((d^2 + 1)\sigma^2)$. For M and K large, it is easy to verify that this choice corresponds to the SNR at each destination terminal in the AF relay network. In this sense, we can conclude that for $\beta \rightarrow \infty$ the AF relay network “converges” to a point-to-point MIMO link with the same received SNR.

VI. CONCLUSION

The minimum rate of growth of the number of relays K , as a function of the number of source–destination terminal pairs M , for coherent fading interference relay networks to decouple was shown to be $K \propto M^3$ under protocol P1 and $K \propto M^2$ under protocol P2. P1 requires relay partitioning and the knowledge of one backward and one forward fading coefficient at each relay, whereas P2 does not need relay partitioning, but requires that each relay knows all its M backward and M forward fading coefficients. The protocols P1 and P2 are thus found to trade off CSI at the relays for the required (for the network to decouple) rate of growth of K as a function of M .

We found that cooperation at the relay level in groups of L relays, both for P1 and P2, results in an L -fold reduction of the total number of relays needed to achieve a given per source–destination terminal pair capacity. An interesting open question in this context is whether more sophisticated signal processing at the relays (such as equalization for the backward link and precoding for the forward link) could lead to improved capacity scaling behavior.

It was furthermore shown that the critical growth rates $K \propto M^3$ in P1 and $K \propto M^2$ in P2 are sufficient to not only make the network decouple, but to also make the individual source–destination fading links converge to nonfading links. We say that the network “crystallizes” as it breaks up into a set of effectively isolated “wires in the air.” More pictorially, the decoupled links experience increasing distributed spatial (or more specifically relay) diversity. Consequently, in the large- M limit time diversity (achieved by coding over a sufficiently long time horizon) is not needed to achieve ergodic capacity. We furthermore characterized the “crystallization” rate (more precisely a guaranteed “crystallization” rate as we do not know whether our bounds are tight), i.e., the rate (as a function of M, K) at which the decoupled links converge to nonfading links. In the course of our analysis, we developed a new technique for characterizing the large-deviations behavior of certain sums of dependent random variables.

For noncoherent fading interference relay networks with AF relaying and joint decoding at the cooperating destination terminals, we computed the asymptotic (in M and K with $K/M \rightarrow \beta$ fixed) network capacity using tools from large random-matrix theory. To the best of our knowledge, this is the first application of large random-matrix theory to characterize the capacity behavior of large fading networks. An elegant extension of this approach to the case of multiple layers of relays was recently reported in [32]. We furthermore demonstrated that for $\beta \rightarrow \infty$, the relay network converges to a point-to-point MIMO link. This generalizes the finite- M result in [1] and shows that the use of relays as active scatterers can recover spatial multiplexing gain in poor scattering environments, even if the number of transmit

and receive antennas grows large. More importantly, our result shows that linear increase in the number of relays as a function of transmit–receive antennas is sufficient for this to happen.

The large-deviations analysis, along with the notion of decoupling of the network, as carried out in this paper, could serve as a general tool to assess the impact of protocols, processing at the relays, propagation conditions, routing, and scheduling on network outage and ergodic capacity performance. More specifically, an interesting question is under which conditions “crystallization” can happen in a general network and, if it occurs, what the corresponding “crystallization” rate would be. It has to be noted, however, that, in view of the technical difficulties posed by the basic case analyzed in this paper, it is unclear whether this framework can yield substantial analytical insights into the above-mentioned questions.

Finally, we note that if we interpret our results in terms of per-node throughput, we find that P1 achieves $O(1/n^{2/3})$ whereas P2 realizes $O(1/\sqrt{n})$. The scaling law for P2 is exactly the same as the behavior established by Gupta and Kumar in [6] and the per-node throughput goes to zero. On the other hand, it is interesting to observe that we can get an $O(1/\sqrt{n})$ throughput without imposing any assumptions on the path-loss behavior. General conclusions on the impact of fading on the network-capacity scaling law cannot be drawn as we are considering a specific setup and specific protocols. It was recently shown [33], however, that under optimistic assumptions on CSI in the network $O(1)$ throughput can be achieved using hierarchical cooperation.

APPENDIX A

TRUNCATION OF RVs AND LARGE DEVIATIONS

We start by recalling the famous Hoeffding inequality along with an important variation that will be central for our developments.

Theorem 8 (Hoeffding [34]): Let X_1, X_2, \dots, X_N be independent real-valued RVs and $A_n \leq X_n \leq B_n$ for $n \in [1:N]$. Let $S_N = \sum_{n=1}^N X_n$. Then

$$\mathbb{P}\{S_N - \mathbb{E}[S_N] \geq Nx\} \leq \exp\left(-\frac{2N^2x^2}{\sum_{n=1}^N (B_n - A_n)^2}\right).$$

Theorem 9 (Maurer [35]): Let X_1, X_2, \dots, X_N be independent real-valued RVs with $X_n \geq 0$ and $\mathbb{E}[X_n^2] < \infty$ for $n \in [1:N]$. Let $S_N = \sum_{n=1}^N X_n$. Then

$$\mathbb{P}\{S_N - \mathbb{E}[S_N] \leq -Nx\} \leq \exp\left(-\frac{N^2x^2}{2\sum_{n=1}^N \mathbb{E}[X_n^2]}\right).$$

The following theorem builds on the Hoeffding inequality (Theorem 8) and constitutes the core of the truncation technique.

Theorem 10: Assume the following on a common probability space.

- The real-valued RVs X_1, X_2, \dots, X_N (possibly dependent) have marginal distribution functions $F_{X_n}(x)$, $n \in [1:N]$. The tails of these distributions are exponentially decaying uniformly in n , i.e., there exist $B > 0$,

$\alpha > 0, \beta > 0$, and $x_0 > 0$ such that for $x \geq x_0 > 0$ and $n \in [1:N]$

$$\mathbb{P}\{|X_n| \geq x\} = 1 - F_{X_n}(x) + F_{X_n}(-x) \leq Be^{-\alpha x^\beta}. \quad (109)$$

- The real-valued RVs $\phi_1, \phi_2, \dots, \phi_N$ are jointly independent and satisfy

$$-1 \leq \phi_n \leq 1, \quad \mathbb{E}[\phi_n] = 0, \quad n \in [1:N].$$

- The real-valued deterministic nonnegative coefficients A_1, A_2, \dots, A_N are uniformly bounded from above, i.e., there exists a constant A independent of n such that

$$0 \leq A_n \leq A, \quad n \in [1:N].$$

- The set of RVs $\{X_n\}_{n=1}^N$ is independent of the set $\{\phi_n\}_{n=1}^N$.

Let $S_N = \sum_{n=1}^N A_n X_n \phi_n$. Then, for all $N > 0$ and $x > 0$ such that $x \geq x_0^{(2+\beta)/2}$

$$\begin{aligned} \mathbb{P}\{|S_N| \geq \sqrt{N}x\} \\ \leq 2 \max[2, NB] \exp\left(-\min\left[\frac{1}{2A^2}, \alpha\right] x^{\frac{2\beta}{2+\beta}}\right). \end{aligned} \quad (110)$$

Proof: The proof is based on the idea of truncation of the RVs X_n . We start by fixing N and choosing t such that $(Nt^2)^\gamma \geq x_0$. The truncation parameter $0 < \gamma < 1$ will be chosen later. Next, we truncate the RVs X_n , $n \in [1:N]$, according to

$$\hat{X}_n \triangleq X_n I\{|X_n| \leq (Nt^2)^\gamma\}.$$

Define $\hat{S}_N \triangleq \sum_{n=1}^N A_n \hat{X}_n \phi_n$. Note that the independence of $\{X_n\}_{n=1}^N$ and $\{\phi_n\}_{n=1}^N$ and the condition $\mathbb{E}[\phi_n] = 0$ ($n \in [1:N]$) implies that $\mathbb{E}[S_N] = \mathbb{E}[\hat{S}_N] = 0$. Let I_n denote the event that X_n is equal to its truncated version, i.e., $I_n \triangleq \{X_n = \hat{X}_n\}$ and, \bar{I}_n the event that $X_n \neq \hat{X}_n$, i.e., $\bar{I}_n \triangleq \{X_n \neq \hat{X}_n\}$. With these definitions, distinguishing the events where either all X_n are equal to their truncated version, i.e., $\bigcap_{n=1}^N I_n$ and where at least one of the X_n is not equal to its truncated version, i.e., $\bigcup_{n=1}^N \bar{I}_n$, we get

$$\begin{aligned} \mathbb{P}\{|S_N| \geq Nt\} \\ = \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcap_{n=1}^N I_n\right\} \mathbb{P}\left\{\bigcap_{n=1}^N I_n\right\} \\ + \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcup_{n=1}^N \bar{I}_n\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \bar{I}_n\right\} \\ = \mathbb{P}\left\{|\hat{S}_N| \geq Nt\right\} \mathbb{P}\left\{\bigcap_{n=1}^N I_n\right\} \\ + \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcup_{n=1}^N \bar{I}_n\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \bar{I}_n\right\} \\ \leq \mathbb{P}\left\{|\hat{S}_N| \geq Nt\right\} + \sum_{n=1}^N \mathbb{P}\{\bar{I}_n\} \end{aligned} \quad (111)$$

where the last step follows by using the trivial bounds

$$\mathbb{P}\left\{\bigcap_{n=1}^N I_n\right\} \leq 1, \quad \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcup_{n=1}^N \bar{I}_n\right\} \leq 1$$

and applying the union bound to $\mathbb{P}\{\bigcup_{n=1}^N \bar{I}_n\}$. Since $-1 \leq \phi_n \leq 1$, $n \in [1:N]$, we obtain the following bounds for the individual terms in \hat{S}_N :

$$-A_n(Nt^2)^\gamma \leq A_n \hat{X}_n \phi_n \leq A_n(Nt^2)^\gamma, \quad n \in [1:N].$$

Moreover, owing to the independence of the ϕ_n , conditioned on the set $\mathcal{X} \triangleq \{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_N\}$, the RVs $A_n \hat{X}_n \phi_n$ are independent. Therefore, using Bayes' rule and the Hoeffding inequality (Theorem 8), noting that $\mathbb{E}[\hat{S}_N | \mathcal{X}] = 0$, we can conclude that

$$\begin{aligned} \mathbb{P}\left\{|\hat{S}_N| \geq Nt\right\} &= \mathbb{E}_{\mathcal{X}} \left[\mathbb{P}\left\{|\hat{S}_N - \mathbb{E}[\hat{S}_N | \mathcal{X}]| \geq Nt \mid \mathcal{X}\right\} \right] \\ &\leq 2 \exp\left(-\frac{N^2 t^2}{2 \sum_{n=1}^N A_n^2 (Nt^2)^{2\gamma}}\right) \\ &\leq 2 \exp\left(-\frac{(Nt^2)^{1-2\gamma}}{2A^2}\right). \end{aligned} \quad (112)$$

Next, using (109), and assuming [this will be justified in (114)] that $(Nt^2)^\gamma \geq x_0$, we have

$$\begin{aligned} \mathbb{P}\{\bar{I}_n\} &= \mathbb{P}\{X_n \neq \hat{X}_n\} \\ &= \mathbb{P}\{|X_n| \geq (Nt^2)^\gamma\} \leq B e^{-\alpha(Nt^2)^\gamma}. \end{aligned} \quad (113)$$

To get the fastest possible exponential decay in (111), we need to choose the free parameter γ such that it maximizes $\min[1 - 2\gamma, \gamma\beta]$, which is the solution that makes the exponents of t in (112) and (113) equal and is given by $\gamma = 1/(2+\beta)$. Finally, setting $t = x/\sqrt{N}$ results in

$$(Nt^2)^\gamma = x^{2\gamma} = x^{2/(2+\beta)} \geq x_0 \quad (114)$$

as required. Combining (111), (112) and (113), we finally obtain

$$\begin{aligned} \mathbb{P}\left\{|S_N| \geq \sqrt{N}x\right\} \\ \leq 2 \exp\left(-\frac{1}{2A^2} x^{\frac{2\beta}{2+\beta}}\right) + NB \exp\left(-\alpha x^{\frac{2\beta}{2+\beta}}\right). \end{aligned} \quad (115)$$

The final result (110) is a trivial upper bound to (115). \square

The following corollary is the generalization of Theorem 10 to the complex-valued case and is used repeatedly in the proofs of Theorems 1 and 2.

Corollary 1: Assume the following on a common probability space.

- The absolute values of the complex-valued (possibly dependent) RVs X_1, X_2, \dots, X_N have marginal distribution functions $F_{X_n}(x)$, $n \in [1:N]$. The tails of these distributions are exponentially decaying uniformly in n , i.e., there exist $B > 0$, $\alpha > 0$, $\beta > 0$, and $x_0 > 0$ such that for $x \geq x_0 > 0$ and $n \in [1:N]$

$$\mathbb{P}\{|X_n| \geq x\} = 1 - F_{X_n}(x) \leq B e^{-\alpha x^\beta}. \quad (116)$$

- The real-valued RVs $\phi_1, \phi_2, \dots, \phi_N$ are jointly independent and satisfy $\phi_n \sim \mathcal{U}(-\pi, \pi)$ and hence $\mathbb{E}[e^{j\phi_n}] = 0$ for all $n \in [1:N]$.

- The real-valued deterministic nonnegative coefficients A_1, A_2, \dots, A_N are uniformly bounded from above, i.e., there exists a constant A independent of n such that

$$0 \leq A_n \leq A, \quad n \in [1:N].$$

- The set of RVs $\{X_n\}_{n=1}^N$ is independent of the set $\{\phi_n\}_{n=1}^N$.
Let $S_N = \sum_{n=1}^N A_n X_n e^{j\phi_n}$. Then, for all $N > 0$ and $x > 0$ such that $x \geq x_0^{(2+\beta)/2}$

$$\begin{aligned} \mathbb{P}\left\{|S_N| \geq \sqrt{N}x\right\} \\ \leq 4 \max[2, NB] \exp\left(-\min\left[\frac{1}{2A^2}, \alpha\right] 2^{-\frac{\beta}{\beta+2}} x^{\frac{2\beta}{\beta+2}}\right). \end{aligned}$$

Proof: Apply Theorem 10 to $\Re S_N$ and $\Im S_N$ separately and combine the two bounds using the Pythagorean union bound (Lemma 3). \square

The following corollary is a modification of Theorem 10 for the case of independent nonnegative RVs and is used repeatedly in the proofs of Theorems 1 and 2.

Corollary 2: Assume the following on a common probability space.

- The real-valued nonnegative RVs X_1, X_2, \dots, X_N are jointly independent and have marginal distribution functions $F_{X_n}(x)$, $n \in [1:N]$. The right tails of these distributions are exponentially decaying uniformly in n , i.e., there exist $B > 0$, $\alpha > 0$, $\beta > 0$, and $x_0 > 0$ such that for all $x \geq x_0 > 0$ and $n \in [1:N]$

$$\mathbb{P}\{X_n \geq x\} = 1 - F_{X_n}(x) \leq B e^{-\alpha x^\beta}. \quad (117)$$

- The expectations $\mathbb{E}[X_n^2]$ are uniformly bounded from above, i.e., there exists a constant C independent of n such that

$$\mathbb{E}[X_n^2] \leq C, \quad n \in [1:N]. \quad (118)$$

- The real-valued deterministic nonnegative coefficients A_1, A_2, \dots, A_N are uniformly bounded from above, i.e., there exists a constant A independent of n such that

$$0 \leq A_n \leq A, \quad n \in [1:N]. \quad (119)$$

Let $S_N = \sum_{n=1}^N A_n X_n$. Then, for all $N > 0$ and $x > 0$ such that $x \geq x_0^{(2+\beta)/2}$

$$\begin{aligned} \mathbb{P}\left\{|S_N - \mathbb{E}[S_N]| \geq \sqrt{N}x\right\} \\ \leq 3 \max[1, NB] \exp\left(-\min\left[\frac{2}{A^2}, \alpha, \frac{1}{2A^2 C}\right] x^{\frac{2\beta}{\beta+2}}\right). \end{aligned} \quad (120)$$

Proof: The proof idea of this corollary is similar to that used in Theorem 10. However, there are several technical details, which do not occur in the proof of Theorem 10. We have,

therefore, decided to present the full version of the proof of Corollary 2.

Unlike in the proof of Theorem 10, here we have $\mathbb{E}[S_N] \neq 0$. To obtain an upper bound on $\mathbb{P}\{|S_N - \mathbb{E}[S_N]| \geq \sqrt{N}x\}$, we establish an upper bound on $\mathbb{P}\{S_N \geq \mathbb{E}[S_N] + \sqrt{N}x\}$ and on $\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\}$ and use the union bound to combine the results.

We start by deriving an upper bound on $\mathbb{P}\{S_N \geq \mathbb{E}[S_N] + \sqrt{N}x\}$. Following the same steps as in the proof of Theorem 10, we define the truncation parameter $0 < \gamma < 1$, which will be chosen later. Fix N and choose t such that $(Nt^2)^\gamma \geq x_0$. We truncate the RVs X_n ($n \in [1:N]$) according to

$$\hat{X}_n \triangleq X_n I[X_n \leq (Nt^2)^\gamma]$$

and define $\hat{S}_N \triangleq \sum_{n=1}^N A_n \hat{X}_n$. It is easily seen that $\mathbb{E}[S_N] \geq \mathbb{E}[\hat{S}_N]$ and therefore

$$\mathbb{P}\{S_N \geq \mathbb{E}[S_N] + Nt\} \leq \mathbb{P}\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt\}. \quad (121)$$

Let I_n denote the event that X_n is equal to its truncated version, i.e., $I_n \triangleq \{X_n = \hat{X}_n\}$, and \bar{I}_n the event that $X_n \neq \hat{X}_n$, i.e., $\bar{I}_n \triangleq \{X_n \neq \hat{X}_n\}$. With these definitions, distinguishing the events where either all X_n are equal to their truncated version, i.e., $\bigcap_{n=1}^N I_n$ and where at least one of the X_n is not equal to its truncated version, i.e., $\bigcup_{n=1}^N \bar{I}_n$, we get

$$\begin{aligned} & \mathbb{P}\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt\} \\ &= \mathbb{P}\left\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt \mid \bigcap_{n=1}^N I_n\right\} \mathbb{P}\left\{\bigcap_{n=1}^N I_n\right\} \\ & \quad + \mathbb{P}\left\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt \mid \bigcup_{n=1}^N \bar{I}_n\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \bar{I}_n\right\} \\ &= \mathbb{P}\{\hat{S}_N \geq \mathbb{E}[\hat{S}_N] + Nt\} \mathbb{P}\left\{\bigcap_{n=1}^N I_n\right\} \\ & \quad + \mathbb{P}\left\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt \mid \bigcup_{n=1}^N \bar{I}_n\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \bar{I}_n\right\} \\ & \leq \mathbb{P}\{\hat{S}_N \geq \mathbb{E}[\hat{S}_N] + Nt\} + \sum_{n=1}^N \mathbb{P}\{\bar{I}_n\} \end{aligned} \quad (122)$$

where the last step is obtained by using the trivial bounds

$$\mathbb{P}\left\{\bigcap_{n=1}^N I_n\right\} \leq 1, \quad \mathbb{P}\left\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt \mid \bigcup_{n=1}^N \bar{I}_n\right\} \leq 1$$

and applying the union bound to $\mathbb{P}\{\bigcup_{n=1}^N \bar{I}_n\}$. The individual terms in \hat{S}_N are bounded according to

$$0 \leq A_n \hat{X}_n \leq A_n (Nt^2)^\gamma, \quad n \in [1:N].$$

Using Bayes' rule and the Hoeffding inequality (Theorem 8), we can conclude that

$$\begin{aligned} & \mathbb{P}\{\hat{S}_N \geq \mathbb{E}[\hat{S}_N] + Nt\} \\ & \leq \exp\left(-\frac{2N^2 t^2}{\sum_{n=1}^N A_n^2 (Nt^2)^{2\gamma}}\right) \\ & \leq \exp\left(-\frac{2(Nt^2)^{1-2\gamma}}{A^2}\right). \end{aligned} \quad (123)$$

Next, using (117), and assuming [this will be justified in (125)] that $(Nt^2)^\gamma \geq x_0$, we have

$$\begin{aligned} \mathbb{P}\{\bar{I}_n\} &= \mathbb{P}\{X_n \neq \hat{X}_n\} \\ &= \mathbb{P}\{X_n \geq (Nt^2)^\gamma\} \leq B e^{-\alpha(Nt^2)^\gamma}. \end{aligned} \quad (124)$$

To get the fastest possible exponential decay in (122), we need to choose the free parameter γ such that it maximizes $\min[1 - 2\gamma, \gamma\beta]$, which is the solution that makes the exponents of t in (123) and (124) equal and is given by $\gamma = 1/(2 + \beta)$. Finally, setting $t = x/\sqrt{N}$ results in

$$(Nt^2)^\gamma = x^{2\gamma} = x^{2/(2+\beta)} \geq x_0 \quad (125)$$

as required. Combining (121)–(124), we obtain

$$\begin{aligned} & \mathbb{P}\{S_N \geq \mathbb{E}[S_N] + \sqrt{N}x\} \\ & \leq \exp\left(-\frac{2}{A^2} x^{\frac{2\beta}{2+\beta}}\right) + NB \exp\left(-\alpha x^{\frac{2\beta}{2+\beta}}\right). \end{aligned} \quad (126)$$

It remains to establish an upper bound on $\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\}$. From Theorem 9, it follows that

$$\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\} \leq \exp\left(-\frac{Nx^2}{2\sum_{n=1}^N \mathbb{E}[A_n^2 X_n^2]}\right)$$

which, using (118) and (119), can be further upper-bounded as

$$\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\} \leq \exp\left(-\frac{x^2}{2A^2 C}\right). \quad (127)$$

Combining (126) and (127) and using the union bound, we obtain

$$\begin{aligned} \mathbb{P}\{|S_N - \mathbb{E}[S_N]| \geq \sqrt{N}x\} & \leq \exp\left(-\frac{2}{A^2} x^{\frac{2\beta}{2+\beta}}\right) \\ & \quad + NB \exp\left(-\alpha x^{\frac{2\beta}{2+\beta}}\right) + \exp\left(-\frac{x^2}{2A^2 C}\right). \end{aligned} \quad (128)$$

The final result (120) is a trivial upper bound to (128). \square

APPENDIX B UNION BOUNDS

In this appendix, as a reference, we present several variations of union bounds for probability that we use frequently throughout the paper.

Lemma 2 (Union Bound for Sums): Assume the complex-valued RVs X_1, X_2, \dots, X_N are such that

$$\mathbb{P}\{|X_n| \geq C_n\} \leq P_n, \quad n \in [1:N]$$

where C_1, C_2, \dots, C_N and P_1, P_2, \dots, P_N are fixed positive constants. Then

$$\mathbb{P}\left\{\left|\sum_{n=1}^N X_n\right| \geq \sum_{n=1}^N C_n\right\} \leq \sum_{n=1}^N P_n.$$

Proof: Let A_n denote the event that $|X_n| \geq C_n$, $n \in [1:N]$. Let B denote the event that $|\sum_{n=1}^N X_n| \geq \sum_{n=1}^N C_n$. By inspection, it follows that $B \Rightarrow \bigcup_{n=1}^N A_n$, which implies $\mathbb{P}\{B\} \leq \sum_{n=1}^N \mathbb{P}\{A_n\}$. \square

The proofs of the remaining union bounds follow exactly the same pattern as the proof of Lemma 2 and will hence be omitted.

Lemma 3 (Pythagorean Union Bound): Assume the complex-valued RV X is such that

$$\mathbb{P}\{|\Re X| \geq C_R\} \leq P_R \quad \text{and} \quad \mathbb{P}\{|\Im X| \geq C_I\} \leq P_I$$

where C_R, C_I, P_R , and P_I are fixed positive constants. Then

$$\mathbb{P}\{|X| \geq \sqrt{C_R^2 + C_I^2}\} \leq P_R + P_I.$$

Lemma 4 (Union Bound for Mixed Sums): Assume that the complex-valued RVs X_1, X_2, \dots, X_N are such that

$$\mathbb{P}\{|X_n| \geq C_n\} \leq P_n, \quad n \in [1:N]$$

where C_1, C_2, \dots, C_N and P_1, P_2, \dots, P_N are fixed positive constants; then, the following statements hold.

1) If the real-valued RVs $X'_1, X'_2, \dots, X'_{N'}$ are such that

$$\mathbb{P}\{X'_n \leq C'_n\} \leq P'_n, \quad n \in [1:N']$$

where $C'_1, C'_2, \dots, C'_{N'}$ and $P'_1, P'_2, \dots, P'_{N'}$ are fixed positive constants, then

$$\mathbb{P}\left\{\left|\sum_{n=1}^N X_n + \sum_{n=1}^{N'} X'_n\right| \leq \max\left[0, \sum_{n=1}^{N'} C'_n - \sum_{n=1}^N C_n\right]\right\} \\ \leq \sum_{n=1}^N P_n + \sum_{n=1}^{N'} P'_n.$$

2) If the real-valued RVs $X'_1, X'_2, \dots, X'_{N'}$ are such that

$$\mathbb{P}\{X'_n \geq C'_n\} \leq P'_n, \quad n \in [1:N']$$

then

$$\mathbb{P}\left\{\left|\sum_{n=1}^N X_n + \sum_{n=1}^{N'} X'_n\right| \geq \sum_{n=1}^{N'} C'_n + \sum_{n=1}^N C_n\right\} \\ \leq \sum_{n=1}^N P_n + \sum_{n=1}^{N'} P'_n.$$

Lemma 5 (Union Bound for Products): Assume the complex-valued RVs X_1, X_2, \dots, X_N are such that

$$\mathbb{P}\{|X_n| \geq C_n\} \leq P_n, \quad n \in [1:N]$$

where C_1, C_2, \dots, C_N and P_1, P_2, \dots, P_N are fixed positive constants. Then

$$\mathbb{P}\left\{\left|\prod_{n=1}^N X_n\right| \geq \prod_{n=1}^N C_n\right\} \leq \sum_{n=1}^N P_n.$$

Lemma 6 (Union Bound for Fractions): If for real-valued positive RVs X_1 and X_2 and positive constants C_1, C_2 and P_1, P_2

$$\mathbb{P}\{X_1 \geq C_1\} \leq P_1 \quad \text{and} \quad \mathbb{P}\{X_2 \leq C_2\} \leq P_2$$

then

$$\mathbb{P}\{X_1/X_2 \geq C_1/C_2\} \leq P_1 + P_2.$$

If, in turn

$$\mathbb{P}\{X_1 \leq C_1\} \leq P_1 \quad \text{and} \quad \mathbb{P}\{X_2 \geq C_2\} \leq P_2$$

then

$$\mathbb{P}\{X_1/X_2 \leq C_1/C_2\} \leq P_1 + P_2.$$

APPENDIX C PROOF OF THEOREM 1

We start by recalling that we want to establish a concentration result for $\text{SINR}_m^{\text{P1}}$, given by (22), using the truncation technique throughout. As already mentioned, this entails establishing the large-deviations behavior of $S^{(1)}, S^{(2)}, S^{(3)}$, and $S^{(4)}$. For $S^{(3)}$, this has already been done in Section III-C2. It remains to establish the corresponding (based on the truncation technique) concentration results for $S^{(1)}, S^{(2)}$, and $S^{(4)}$ defined by (23), (24), and (26), respectively.

A. Analysis of $S^{(1)}$

The sum $S^{(1)}$ can be written as

$$S^{(1)} = \sum_{k:p(k)=m} C_{\text{P1},k}^{m,m} Z_k^{(1)} \quad (129)$$

with

$$Z_k^{(1)} \triangleq |f_{m,k}| |h_{k,m}|.$$

For any $k \in [1:K]$ such that $p(k) = m$, we have

$$\mathbb{E}[Z_k^{(1)}] = \pi/4 \quad \text{and} \quad \mathbb{E}\left[\left(Z_k^{(1)}\right)^2\right] = 1.$$

Application of the union bound for products yields

$$\mathbb{P}\{Z_k^{(1)} \geq x\} \leq 2e^{-x}, \quad x \geq 0.$$

Noting that the sum $S^{(1)}$ contains K/M terms, which are jointly independent, taking into account (12), and using Corollary 2, we get for $x \geq 0$ and $K/M \geq 1$

$$\mathbb{P}\left\{\left|S^{(1)} - \frac{\pi}{4} \sum_{k:p(k)=m} C_{\text{P1},k}^{m,m}\right| \geq \sqrt{\frac{K}{M}} x\right\} \leq 6 \frac{K}{M} e^{-\Delta^{(1)} x^{2/3}}$$

with $\Delta^{(1)} = \min\left[1, 1/(2\bar{C}^2)\right]$. Finally, using (12), it follows that

$$\mathbb{P}\left\{S^{(1)} \geq \frac{\pi\bar{C}K}{4M} + \sqrt{\frac{K}{M}}x\right\} \leq 6\frac{K}{M}e^{-\Delta^{(1)}x^{2/3}} \quad (130)$$

and

$$\mathbb{P}\left\{S^{(1)} \leq \frac{\pi\bar{C}K}{4M} - \sqrt{\frac{K}{M}}x\right\} \leq 6\frac{K}{M}e^{-\Delta^{(1)}x^{2/3}} \quad (131)$$

for any $x \geq 0$ and $K/M \geq 1$.

B. Analysis of $S^{(2)}$

The sum $S^{(2)}$ can be written as

$$S^{(2)} = \sum_{k:p(k) \neq m} C_{P1,k}^{m,m} Z_k^{(2)}$$

with

$$Z_k^{(2)} \triangleq \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,m}.$$

For any $k \in [1:K]$ such that $p(k) \neq m$, we have $\mathbb{E}[Z_k^{(2)}] = 0$. Application of the union bound for products yields

$$\mathbb{P}\left\{|Z_k^{(2)}| \geq x\right\} \leq 2e^{-x}, \quad x \geq 0.$$

Noting that the sum $S^{(2)}$ contains $K(M-1)/M$ terms, which are jointly independent, taking into account (12), and using Corollary 1, we get for $x \geq 0$ and $K(M-1)/M \geq 1$

$$\mathbb{P}\left\{|S^{(2)}| \geq \sqrt{\frac{K(M-1)}{M}}x\right\} \leq 8\frac{K(M-1)}{M}e^{-\Delta^{(2)}x^{2/3}} \quad (132)$$

with $\Delta^{(2)} = 2^{-\frac{1}{3}} \min\left[1, 1/(2\bar{C}^2)\right]$.

C. Analysis of $S^{(4)}$

The sum $S^{(4)}$ can be written as

$$S^{(4)} = \sum_{k=1}^K (C_{P1,k}^m)^2 Z_k^{(4)}$$

with

$$Z_k^{(4)} = |f_{m,k}|^2.$$

Since $Z_k^{(4)}$ is exponentially distributed with parameter $\lambda = 1$, we have

$$\mathbb{P}\left\{Z_k^{(4)} \geq x\right\} \leq e^{-x}, \quad k \in [1:K], \quad x \geq 0.$$

Noting that the sum $S^{(4)}$ contains K jointly independent terms, taking into account (13), and using

$$\mathbb{E}\left[Z_k^{(4)}\right] = 1 \quad \text{and} \quad \mathbb{E}\left[\left(Z_k^{(4)}\right)^2\right] = 2, \quad k \in [1:K]$$

we get for $x \geq 0$ and $K \geq 1$

$$\mathbb{P}\left\{\left|S^{(4)} - \sum_{k=1}^K (C_{P1,k}^m)^2\right| \geq \sqrt{K}x\right\} \leq 3Ke^{-\Delta^{(4)}x^{2/3}}$$

with $\Delta^{(4)} = \min\left[1, 1/(4\bar{C}^4)\right]$. Therefore, using (13), it follows that

$$\mathbb{P}\left\{S^{(4)} \geq K\bar{C}^2 + \sqrt{K}x\right\} \leq 3Ke^{-\Delta^{(4)}x^{2/3}} \quad (133)$$

and

$$\mathbb{P}\left\{S^{(4)} \leq K\bar{C}^2 - \sqrt{K}x\right\} \leq 3Ke^{-\Delta^{(4)}x^{2/3}}. \quad (134)$$

We are now ready to carry out the final Step v of the program outlined in the first paragraph of Section III-C. The concentration result for SINR_{P1}^m is expressed in terms of upper bounds on $\mathbb{P}\{\text{SINR}_m^{P1} \geq \hat{U}_{P1}\}$ and $\mathbb{P}\{\text{SINR}_m^{P1} \leq \hat{L}_{P1}\}$, where the exact form of \hat{U}_{P1} and \hat{L}_{P1} is specified below.

To establish an upper bound on $\mathbb{P}\{\text{SINR}_m^{P1} \geq \hat{U}_{P1}\}$, we proceed as follows.

- 1) Apply Part 2 of Lemma 4 to (130) and (132) to establish a stochastic upper bound²¹ for $|S^{(1)} + S^{(2)}|$.
- 2) Apply Part 1 of Lemma 4 to (45), (52), and (53) to establish a stochastic lower bound²² for $|S^{(3)}|$.
- 3) Apply Part 1 of Lemma 4 to the result from Step 2 and (134) to establish a stochastic lower bound for $S^{(3)} + \sigma^2MS^{(4)} + KM\sigma^2$.
- 4) Apply the union bound for fractions (Lemma 6) to the stochastic upper bound from Step 1 and to the stochastic lower bound from Step 3 to establish the final result:

$$\mathbb{P}\left\{\text{SINR}_m^{P1} \geq \hat{U}_{P1}\right\} \leq P_{P1}^U \quad (135)$$

with

$$\hat{U}_{P1} \triangleq \frac{\pi^2}{16} \frac{\bar{C}^2}{\underline{C}_{SN}^2} \frac{K}{M^3} \frac{\hat{U}_{P1}^N}{\hat{U}_{P1}^D} \quad (136)$$

and P_{P1}^U , \hat{U}_{P1}^N , and \hat{U}_{P1}^D defined as

$$\begin{aligned} P_{P1}^U &\triangleq 6\frac{K}{M}e^{-\Delta^{(1)}x_1^{2/3}} \\ &+ 8\frac{K(M-1)}{M}e^{-\Delta^{(2)}x_2^{2/3}} \\ &+ 6(M-1)Ke^{-\Delta^{(31)}x_{31}^{2/5}} \\ &+ 64\frac{(K-1)K(M-1)^2}{M}e^{-\Delta^{(32)}x_{321}^{2/7}} \\ &+ 64\frac{(K-1)K(M-1)}{M}e^{-\Delta^{(32)}x_{322}^{2/7}} \\ &+ 3Ke^{-\Delta^{(4)}x_4^{2/3}} \end{aligned} \quad (137)$$

²¹For an RV X , a ‘‘stochastic upper bound’’ in this context means a bound of the form $\mathbb{P}\{X \geq A\} \leq P$.

²²For an RV X , a ‘‘stochastic lower bound’’ in this context means a bound of the form $\mathbb{P}\{X \leq A\} \leq P$.

$$\hat{U}_{P1}^N \triangleq \left(1 + \frac{4}{\underline{C}_\pi} \sqrt{\frac{M}{K}} x_1 + \frac{4}{\underline{C}_\pi} \sqrt{\frac{M(M-1)}{K}} x_2 \right)^2 \quad (138)$$

$$\hat{U}_{P1}^D \triangleq \max \left[0, \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{M-1}{M} - \frac{1}{\underline{C}_{SN}^2} \frac{M-1}{M\sqrt{K}} x_{31} \right. \\ \left. - \frac{1}{\underline{C}_{SN}^2} \sqrt{\frac{(K-1)(M-1)^2}{KM^3}} x_{321} \right. \\ \left. - \frac{1}{\underline{C}_{SN}^2} \sqrt{\frac{(K-1)(M-1)}{KM^3}} x_{322} \right] \\ + \frac{\sigma^2}{\underline{C}_{SN}^2} \max \left[0, \underline{c}^2 - \frac{1}{\sqrt{K}} x_4 \right] + \frac{\sigma^2}{\underline{C}_{SN}^2}. \quad (139)$$

An upper bound on $\mathbb{P}\{\text{SINR}_m^{P1} \leq \hat{L}_{P1}\}$ can be obtained as follows.

- 1) Apply Part 1 of Lemma 4 to (131) and (132) to establish a stochastic lower bound for $|S^{(1)} + S^{(2)}|$.
- 2) Apply Part 2 of Lemma 4 to (44), (52), and (53) to establish a stochastic upper bound for $|S^{(3)}|$.
- 3) Apply Part 2 of Lemma 4 to the result from Step 2 and to (133) to establish a stochastic upper bound for $S^{(3)} + \sigma^2 MS^{(4)} + KM\sigma^2$.
- 4) Apply the union bound for fractions to the stochastic lower bound from Step 1 and to the stochastic upper bound from Step 3 to establish the final result:

$$\mathbb{P}\{\text{SINR}_m^{P1} \leq \hat{L}_{P1}\} \leq P_{P1}^U \quad (140)$$

with

$$\hat{L}_{P1} \triangleq \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{K}{M^3} \frac{\hat{L}_{P1}^N}{\hat{L}_{P1}^D} \quad (141)$$

and \hat{L}_{P1}^N and \hat{L}_{P1}^D defined as

$$\hat{L}_{P1}^N \triangleq \max \left[0, 1 - \frac{4}{\underline{C}_\pi} \sqrt{\frac{M}{K}} x_1 \right. \\ \left. - \frac{4}{\underline{C}_\pi} \sqrt{\frac{M(M-1)}{K}} x_2 \right]^2 \quad (142)$$

$$\hat{L}_{P1}^D \triangleq \frac{\underline{C}^2}{\underline{C}_{SN}^2} \frac{M-1}{M} + \frac{1}{\underline{C}_{SN}^2} \frac{M-1}{M\sqrt{K}} x_{31} \\ + \frac{1}{\underline{C}_{SN}^2} \sqrt{\frac{(K-1)(M-1)^2}{KM^3}} x_{321} \\ + \frac{1}{\underline{C}_{SN}^2} \sqrt{\frac{(K-1)(M-1)}{KM^3}} x_{322} \\ + \frac{\sigma^2}{\underline{C}_{SN}^2} \left(\underline{c}^2 + \frac{1}{\sqrt{K}} x_4 \right) + \frac{\sigma^2}{\underline{C}_{SN}^2}. \quad (143)$$

The result presented in Theorem 1 is a simpler and slightly weaker form of the bounds (135) and (140). To obtain this simplification we proceed as follows. Set

$$x_1 = x_2 = x_{31} = x_4 = x_{321} = x_{322} = x$$

in (137), (138), (139), (142), and (143). Note that in this case $L_{P1}(x) \leq \hat{L}_{P1}(x)$ and $U_{P1}(x) \geq \hat{U}_{P1}(x)$ and therefore

$$\mathbb{P}\{\text{SINR}_m^{P1} \geq U_{P1}\} \leq \mathbb{P}\{\text{SINR}_m^{P1} \geq \hat{U}_{P1}\} \leq P_{P1}^U \quad (144)$$

$$\mathbb{P}\{\text{SINR}_m^{P1} \leq L_{P1}\} \leq \mathbb{P}\{\text{SINR}_m^{P1} \leq \hat{L}_{P1}\} \leq P_{P1}^U. \quad (145)$$

Finally, combine the bounds (144) and (145) according to

$$\mathbb{P}\left\{ \left(\text{SINR}_m^{P1} \geq U_{P1} \right) \cup \left(\text{SINR}_m^{P1} \leq L_{P1} \right) \right\} \\ \leq \mathbb{P}\{\text{SINR}_m^{P1} \geq U_{P1}\} + \mathbb{P}\{\text{SINR}_m^{P1} \leq L_{P1}\} \leq 2P_{P1}^U$$

and note that $2P_{P1}^U$ is upper-bounded by the RHS of (56). \square

APPENDIX D

PROOF OF LOWER BOUND IN THEOREM 6

As already mentioned in the main body of the paper, the proof of the lower bound in (89) is based on the technique summarized in Appendix E. After straightforward algebra, it follows that the I-O relation of the SISO channel between the terminals \mathcal{S}_m and \mathcal{D}_m ($m \in [1:M]$) is given by

$$y_m = \left(\bar{F}_m + \tilde{F}_m \right) s_m + W_m$$

where

$$\bar{F}_m \triangleq \frac{1}{\sqrt{Q}} \sum_{q=1}^Q \mathbb{E}[a_q^{m,m}]$$

$$\tilde{F}_m \triangleq \frac{1}{\sqrt{Q}} \sum_{q=1}^Q (a_q^{m,m} - \mathbb{E}[a_q^{m,m}])$$

$$W_m \triangleq \sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{Q}} \sum_{q=1}^Q a_q^{m,\hat{m}} + \frac{1}{\sqrt{Q}} \\ \times \sum_{q=1}^Q b_q^m \tilde{\mathbf{h}}_{q,p(q)}^H \mathbf{z}_q + w_m$$

and

$$a_q^{m,\hat{m}} \triangleq C_{P1,q}^{m,\hat{m}} \left(\tilde{\mathbf{f}}_{p(q),q}^H \mathbf{f}_{m,q} \right) \left(\tilde{\mathbf{h}}_{q,p(q)}^H \mathbf{h}_{q,\hat{m}} \right)$$

$$b_q^m \triangleq C_{P1,q}^m \left(\tilde{\mathbf{f}}_{p(q),q}^H \mathbf{f}_{m,q} \right)$$

$$C_{P1,q}^{m,\hat{m}} \triangleq \sqrt{Q} d_{P1,q} \hat{P}_{m,q} \hat{E}_{q,\hat{m}}$$

$$C_{P1,q}^m \triangleq \sqrt{Q} d_{P1,q} \hat{P}_{m,q}$$

It is not difficult, but tedious, to verify that

$$\bar{F}_m = \frac{\pi}{4} \frac{L^2}{\sqrt{Q}} \sum_{q:p(q)=m} C_{P1,q}^{m,m} \quad (146)$$

$$\text{Var}[\tilde{F}_m] = \frac{L^2}{Q} \sum_{q:p(q) \neq m} \left(C_{P1,q}^{m,m} \right)^2 \\ + \frac{(L + (\pi/4)(L-1)L)^2 - (\pi^2/16)L^4}{Q} \\ \times \sum_{q:p(q)=m} \left(C_{P1,q}^{m,m} \right)^2 \quad (147)$$

$$\begin{aligned}
\text{Var}[W_m] &= \frac{L^2 + (\pi/4)L^2(L-1)}{QM} \sum_{\hat{m} \neq m} \sum_{q:p(q)=m} \left(C_{P1,q}^{m,\hat{m}} \right)^2 \\
&+ \frac{L^2 + (\pi/4)L^2(L-1)}{QM} \sum_{\hat{m} \neq m} \sum_{q:p(q)=\hat{m}} \left(C_{P1,q}^{m,\hat{m}} \right)^2 \\
&+ \frac{L^2}{QM} \sum_{\hat{m} \neq m} \sum_{\substack{q:p(q) \neq m \\ p(q) \neq \hat{m}}} \left(C_{P1,q}^{m,\hat{m}} \right)^2 \\
&+ \frac{L^2 + (\pi/4)L^2(L-1)}{Q} \sigma^2 \sum_{q:p(q)=m} \left(C_{P1,q}^m \right)^2 \\
&+ \frac{L^2}{Q} \sigma^2 \sum_{q:p(q) \neq m} \left(C_{P1,q}^m \right)^2 + \sigma^2. \tag{148}
\end{aligned}$$

Using (3), we lower-bound \bar{F}_m and upper-bound $\text{Var}[\tilde{F}_m]$ and $\text{Var}[W_m]$, substitute the resulting bounds into (150), and obtain

$$I(y_m; s_m) \geq \frac{1}{2} \log \left(1 + \frac{\pi^2 Q}{16 M^3} \underline{f}(M, L) \right) \tag{149}$$

where

$$\begin{aligned}
\underline{f}(M, L) &= \frac{P \underline{E} P_{\text{rel}} L^2}{\left(\bar{E} + \frac{\pi(L-1)}{4M} \bar{E} + \sigma^2 \right) \left(\epsilon(M, L) + \bar{C}^2 + \sigma^2 \bar{c}^2 + \sigma^2 \right)}
\end{aligned}$$

with

$$\begin{aligned}
\epsilon(M, L) &= \frac{\bar{C}^2}{M} + \frac{(1 + (\pi/4)(L-1))^2 \bar{C}^2}{M^2} \\
&+ \frac{(1 + (\pi/4)(L-1)) (2\bar{C}^2 + \sigma^2 \bar{c}^2)}{M}.
\end{aligned}$$

Finally, since L is finite, it follows by inspection that $\lim_{M \rightarrow \infty} \epsilon(M, L) = 0$ and, therefore

$$\lim_{M \rightarrow \infty} \underline{f}(M, L) = \frac{L^2 \bar{C}^2}{\bar{C}_{\text{SN}}^2}$$

which, together with (149), concludes the proof. \square

APPENDIX E

LOWER BOUND ON CHANNEL CAPACITY WITH IMPERFECT CHANNEL KNOWLEDGE

The following lemma is obtained by recognizing that the expression in [23, eq. (66)] is trivially a lower bound to $I(X; Y)$ in (150) below. For completeness, we present the result in the form needed in this paper. For the proof of the (general) statement the interested reader is referred to [23].

Lemma 7: Consider a SISO channel with I–O relation

$$Y = FX + W$$

where $X \sim \mathcal{CN}(0, \sigma_X^2)$, W is zero-mean noise²³ with variance σ_W^2 , F is the random channel gain with variance σ_F^2 , and Y is the output of the channel. Assume that F can be decomposed as

$$F = \bar{F} + \tilde{F}$$

where $\bar{F} = \mathbb{E}[F]$ is known at the receiver and \tilde{F} with $\mathbb{E}[\tilde{F}] = 0$ is not known at the receiver. Assume that X is statistically independent²⁴ of both F and W . Then, the mutual information $I(X; Y)$ can be lower-bounded as follows:

$$I(X; Y) \geq \log \left(1 + \frac{\bar{F}^2 \sigma_X^2}{\sigma_F^2 \sigma_X^2 + \sigma_W^2} \right). \tag{150}$$

APPENDIX F

SOME ESSENTIALS FROM LARGE RANDOM-MATRIX THEORY

In this appendix, we briefly summarize the basic definitions and results from large random-matrix theory used in this paper. An excellent tutorial on this subject is [25].

Definition 3 (Stieltjes Transform): Let $F(x)$ be a distribution function with density $f(x)$. The analytic function

$$G_F(z) \triangleq \int \frac{f(x)}{x-z} dx, \quad z \in \mathbb{C}^+$$

is called the Stieltjes transform of $F(x)$.

Lemma 8 (Inversion Formula): Let $G_F(z)$ be the Stieltjes transform of a distribution function $F(x)$. The corresponding density function can be obtained as

$$f(x) = \frac{1}{\pi} \lim_{y \rightarrow 0^+} \Im [G_F(x + jy)]. \tag{151}$$

Theorem 11 (Silverstein [24]): Define the following quantities on a common probability space.

- The random matrix $\mathbf{X} \in \mathbb{C}^{N \times N'}$ has i.i.d. zero-mean entries with variance one.
- The random matrix $\mathbf{Y} \in \mathbb{C}^{N \times N}$ is Hermitian nonnegative definite with $F_{\mathbf{Y}}^N(x)$, for $N \rightarrow \infty$, converging on $[0, \infty)$ a.s. to a nonrandom distribution function $F_{\mathbf{Y}}(x)$ with corresponding density $f_{\mathbf{Y}}(x)$.

Assume that the matrices \mathbf{X} and \mathbf{Y} are statistically independent. Then, for $N, N' \rightarrow \infty$ with $N/N' \rightarrow \beta$,

$$F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}^N(x) \xrightarrow{\text{a.s.}} F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}(x)$$

with its Stieltjes transform $G_{F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}}(z)$ satisfying

$$\begin{aligned}
G_{F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}}(z) &= \int_{-\infty}^{\infty} \frac{f_{\mathbf{Y}}(x) dx}{x(1 - \beta - \beta z G_{F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}}(z)) - z}, \quad z \in \mathbb{C}^+.
\end{aligned}$$

The solution of this fixed-point equation is unique in the set

$$\left\{ G_{F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}}(z) \in \mathbb{C} \mid -\frac{1-\beta}{z} + \beta G_{F_{(1/N')\mathbf{X}\mathbf{X}^H\mathbf{Y}}}(z) \in \mathbb{C}^+ \right\}.$$

We shall furthermore use the Marčenko–Pastur law as stated in [36].

²⁴In [22, Sec. III], it is assumed that X , F , and W are statistically independent. The condition required here is weaker: F and W need not be statistically independent.

²³In contrast to [22, Sec. III], the noise is not necessarily Gaussian.

Theorem 12 (Marčenko–Pastur [37]): Assume that the matrix $\mathbf{X} \in \mathbb{C}^{N \times N'}$ has i.i.d. zero-mean entries with variance d^2 . Then, for $N, N' \rightarrow \infty$ with $N'/N \rightarrow \beta$, the ESD of $(1/N')\mathbf{X}\mathbf{X}^H$ converges a.s. to a limiting distribution function with density

$$f_{(1/N')\mathbf{X}\mathbf{X}^H}(x) = \frac{\beta}{2\pi x d^2} \sqrt{(\gamma_2 - x)^+ (x - \gamma_1)^+} + [1 - \beta]^+ \delta(x)$$

where $\gamma_1 = d^2(1 - 1/\sqrt{\beta})^2$ and $\gamma_2 = d^2(1 + 1/\sqrt{\beta})^2$.

Under the same assumptions as in the first statement, if, in addition, the entries of \mathbf{X} have finite fourth moments, then a.s.

$$\begin{aligned} \lim_{N' \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N'} \mathbf{X}\mathbf{X}^H \right) &= \gamma_1 \\ \lim_{N' \rightarrow \infty} \lambda_{\max} \left(\frac{1}{N'} \mathbf{X}\mathbf{X}^H \right) &= \gamma_2. \end{aligned}$$

APPENDIX G

COMPUTATION OF THE INTEGRAL \hat{I} IN (100)

In the following, we detail the computation of the integral

$$\hat{I} \triangleq \rho \int_{\eta_1}^{\eta_2} \frac{\sqrt{(\eta_2 - x)(x - \eta_1)} dx}{x(1-x)^2 \left(x \left(\frac{1-\beta}{z} - \beta G \right) - 1 \right)}$$

on the RHS of (100). With the change of variables

$$t = \sqrt{\frac{x - \eta_1}{\eta_2 - x}}$$

and the notation

$$\begin{aligned} \mu_1 &\triangleq 1 - \eta_1 & \nu_1 &\triangleq \eta_1 \left(\frac{1-\beta}{z} - \beta G \right) - 1 \\ \mu_2 &\triangleq 1 - \eta_2 & \nu_2 &\triangleq \eta_2 \left(\frac{1-\beta}{z} - \beta G \right) - 1 \end{aligned}$$

the integral \hat{I} can be written as

$$\hat{I} = 2(\eta_2 - \eta_1)^2 \rho \int_0^\infty \frac{t^2(t^2 + 1) dt}{(\eta_2 t^2 + \eta_1)(\mu_2 t^2 + \mu_1)^2 (\nu_2 t^2 + \nu_1)}.$$

To simplify further, we introduce the notation

$$\kappa_1 \triangleq -\frac{\eta_1}{\eta_2}, \quad \kappa_2 \triangleq -\frac{\mu_1}{\mu_2}, \quad \kappa_3 \triangleq -\frac{\nu_1}{\nu_2}, \quad \chi \triangleq \frac{2(\eta_2 - \eta_1)^2 \rho}{\eta_2 \mu_2^2 \nu_2}$$

so that

$$\hat{I} = \chi \int_0^\infty \frac{t^2(t^2 + 1) dt}{(t^2 - \kappa_1)(t^2 - \kappa_2)^2(t^2 - \kappa_3)}. \quad (152)$$

Upon partial fraction expansion of the integrand in (152), we obtain

$$\hat{I} = \chi(A_1 \hat{I}_1 + A_2 \hat{I}_2 + A_3 \hat{I}_3 + A_4 \hat{I}_4)$$

where

$$\begin{aligned} \hat{I}_1 &\triangleq \int_0^\infty \frac{dt}{t^2 - \kappa_1} & \hat{I}_2 &\triangleq \int_0^\infty \frac{dt}{(t^2 - \kappa_2)^2} \\ \hat{I}_3 &\triangleq \int_0^\infty \frac{dt}{t^2 - \kappa_2} & \hat{I}_4 &\triangleq \int_0^\infty \frac{dt}{t^2 - \kappa_3} \end{aligned} \quad (153)$$

with

$$A_1 = \frac{\kappa_1(\kappa_1 + 1)}{(\kappa_1 - \kappa_2)^2(\kappa_1 - \kappa_3)} \quad (154)$$

$$A_2 = \frac{\kappa_2(\kappa_2 + 1)}{(\kappa_2 - \kappa_1)(\kappa_2 - \kappa_3)} \quad (155)$$

$$A_3 = \frac{-\kappa_2^2 - \kappa_1 \kappa_2^2 + \kappa_1 \kappa_3 + 2\kappa_1 \kappa_2 \kappa_3 - \kappa_2^2 \kappa_3}{(\kappa_2 - \kappa_1)^2(\kappa_2 - \kappa_3)^2} \quad (156)$$

$$A_4 = \frac{\kappa_3(\kappa_3 + 1)}{(\kappa_3 - \kappa_1)(\kappa_3 - \kappa_2)^2}. \quad (157)$$

The integrals in (153) can be evaluated resulting in

$$\hat{I}_1 = \frac{1}{\sqrt{-\kappa_1}} \arctan \frac{t}{\sqrt{-\kappa_1}} \Big|_0^\infty = \frac{\pi}{2\sqrt{-\kappa_1}} \quad (158)$$

$$\begin{aligned} \hat{I}_2 &= -\frac{t}{2\kappa_2(t^2 - \kappa_2)} \Big|_0^\infty - \frac{1}{2\kappa_2 \sqrt{-\kappa_2}} \arctan \frac{t}{\sqrt{-\kappa_2}} \Big|_0^\infty \\ &= -\frac{\pi}{4\kappa_2 \sqrt{-\kappa_2}} \end{aligned} \quad (159)$$

$$\hat{I}_3 = \frac{1}{\sqrt{-\kappa_2}} \arctan \frac{t}{\sqrt{-\kappa_2}} \Big|_0^\infty = \frac{\pi}{2\sqrt{-\kappa_2}} \quad (160)$$

$$\hat{I}_4 = \frac{1}{\sqrt{-\kappa_3}} \arctan \frac{t}{\sqrt{-\kappa_3}} \Big|_0^\infty = \frac{\pi}{2\sqrt{-\kappa_3}}. \quad (161)$$

The quantity κ_3 is complex-valued, and the arctan and square root in (160) are understood as the principal values of these functions in \mathbb{C} as defined in [38].

Finally, by inspection, combining (158)–(161) with (154)–(157) and resubstituting the values of the parameters $\kappa_1, \kappa_2, \kappa_3, \chi, \rho, \mu_1, \mu_2, \eta_1, \eta_2, \nu_1, \nu_2, \gamma_1$, and γ_2 , after straightforward but tedious simplifications, we find

$$\begin{aligned} \chi A_1 \hat{I}_1 &= \frac{(\sqrt{\beta} + 1) |\sqrt{\beta} - 1|}{2\beta} \\ \chi A_2 \hat{I}_2 &= -\frac{z}{\sqrt{\beta}(G\beta z + z + \beta - 1)} \\ \chi A_3 \hat{I}_3 &= -\frac{z d^2 (\sqrt{\beta} - 1)^2 (G\beta z + z + \beta - 1)}{2d^2 \beta (G\beta z + z + \beta - 1)^2} \\ &\quad + \frac{z\beta(G\beta z + \beta - 1)}{2d^2 \beta (G\beta z + z + \beta - 1)^2} \\ \chi A_4 \hat{I}_4 &= -\frac{(G\beta z + \beta - 1)}{2d^2 \beta (G\beta z + z + \beta - 1)^2} \\ &\quad \times \sqrt{\frac{d^2(G\beta z + z + \beta - 1)(\sqrt{\beta} - 1)^2 + z\beta}{d^2(G\beta z + z + \beta - 1)(\sqrt{\beta} + 1)^2 + z\beta}} \\ &\quad \times \left(d^2(G\beta z + z + \beta - 1)(\sqrt{\beta} + 1)^2 + z\beta \right). \end{aligned}$$

ACKNOWLEDGMENT

The authors are indebted to Prof. O. Zeitouni for suggesting the application of the truncation technique to establish the large-deviations behavior of the sums of dependent random variables occurring in the proofs of Theorems 1 and 2. We are furthermore grateful to Prof. Zeitouni for pointing out an error in an earlier version of Theorem 10 and for suggesting the correction.

Helpful discussions with Prof. Zeitouni on noncoherent (AF) relay networks are acknowledged as well. We would furthermore like to thank A. Dana for pointing out that P2 as introduced in [2] leads to decoupling of the network.

REFERENCES

- [1] H. Bölcskei, R. U. Nabar, Ö. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.
- [2] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and ad-hoc wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2890–2914, Jul. 2006.
- [3] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, no. 1, pp. 120–154, 1971.
- [4] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
- [5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [6] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2002.
- [7] M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.
- [8] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.
- [9] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Oct. 2002.
- [10] O. Lévêque and Í. E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 858–865, Mar. 2005.
- [11] L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.
- [12] A. Jovičić, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2555–2565, Nov. 2004.
- [13] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.
- [14] B. Wang, J. Zhang, and L. Zheng, "Achievable rates and scaling laws of power-constrained wireless sensory relay networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4084–4104, Sep. 2006.
- [15] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [16] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [17] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.
- [18] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [19] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
- [20] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.
- [21] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.
- [22] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [23] A. Lapidth and S. Shamai (Shitz), "Fading channels: How perfect need side information be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [24] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, pp. 331–339, Nov. 1995.
- [25] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Commun. and Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [26] R. R. Müller, "Applications of large random matrices in communications engineering," in *Proc. Int. Conf. Advances in the Internet, Processing, Systems, and Interdisciplinary Research (IPSI)*, Sveti Stefan, Montenegro, Oct. 2003.
- [27] A. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 60–71, Jan. 1978.
- [28] R. U. Nabar and H. Bölcskei, "Capacity scaling laws in asynchronous relay networks," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2004, pp. 502–511.
- [29] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [30] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [31] Í. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
- [32] S. Yeh and O. Lévêque, "Asymptotic capacity of multi-level amplify-and-forward relay networks," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1436–1440.
- [33] 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.
- [34] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, pp. 13–30, Mar. 1963.
- [35] A. Maurer, "A bound on the deviation probability for sums of non-negative random variables," *J. Inequalities Pure Appl. Math.*, vol. 4, no. 1, 2003.
- [36] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices," *Statistica Sinica*, vol. 9, pp. 611–677, 1999.
- [37] V. A. Marčenko and L. A. Pastur, "Distribution of some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.
- [38] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1964.