Full-Duplex MIMO Relaying: Achievable Rates under Limited Dynamic Range

Brian P. Day,* Adam R. Margetts,† Daniel W. Bliss,† and Philip Schniter*
*Dept. of ECE, The Ohio State University, Columbus, OH 43210. Email: day.262@osu.edu, schniter@ece.osu.edu
†Advanced Sensor Techniques Group, MIT Lincoln Laboratory, Lexington, MA. Email: bliss@ll.mit.edu, margetts@ieee.org

***Abstract*—In this paper we consider the problem of full-duplex multiple-input multiple-output (MIMO) relaying between multi-antenna source and destination nodes. The principal difficulty in implementing such a system is that, due to the limited attenuation between the relay's transmit and receive antenna arrays, the relay's outgoing signal may overwhelm its limited-dynamic-range input circuitry, making it difficult—if not impossible—to recover the desired incoming signal. While explicitly modeling transmitter/receiver dynamic-range limitations and channel estimation error, we derive tight upper and lower bounds on the end-to-end achievable rate of decode-and-forward-based full-duplex MIMO relay systems, and propose a transmission scheme based on maximization of the lower bound. The maximization requires us to (numerically) solve a nonconvex optimization problem, for which we detail a novel approach based on bisection search and gradient projection. To gain insights into system design tradeoffs, we also derive an analytic approximation to the achievable rate and numerically demonstrate its accuracy.** [1]

I. Introduction

We consider the problem of communicating from source to destination nodes through a relay node. Traditional relay systems operate in a half-duplex, whereby the time-frequency signal-space used for the source-to-relay link is kept orthogonal to that used for the relay-to-destination link, such as with non-overlapping time periods or frequency bands. Half-duplex operation is used to avoid the high levels of relay self-interference that are faced with full-duplex operation, where the source and relay share a common time-frequency signal-space. For example, it is not unusual for the ratio between the relay's self-interference power and desired incoming signal power to exceed that of the relay's front-end hardware, making it impossible to recover the desired signal. The importance of *limited dynamic-range* (DR) cannot be overstressed; notice that, even if the self-interference signal was perfectly known, limited-DR renders perfect cancellation impossible.

Recently, multiple-input multiple-output (MIMO) relaying has been proposed as a means of increasing spectral efficiency (e.g., [1]). By MIMO relaying, we mean that the source, relay, and destination each use multiple antennas for both reception and transmission. MIMO relaying brings the possibility of full-duplex operation through *spatial* self-interference suppression (e.g., [2]–[6]). Still, the following fundamental questions about full-duplex MIMO relaying in the presence of self-interference remain: *1) What is the maximum achievable end-to-end throughput under a transmit power constraint? 2) How can the system be designed to achieve this throughput?*

In this paper, we aim to answer these two fundamental questions while paying special attention to the effects of both limited-DR and imperfect channel-state information (CSI). Limited-DR is a natural consequence of non-ideal amplifiers, oscillators, analog-to-digital converters (ADCs), and digital-to-analog converters (DACs). To model the effects of limited receiver-DR, we inject, at each receive antenna, an additive white Gaussian "receiver distortion" with variance β times the energy impinging on that receive antenna (where $\beta \ll 1$). Similarly, to model the effects of limited transmitter-DR, we inject, at each transmit antenna, an additive white Gaussian "transmitter noise" with variance κ times the energy of the intended transmit signal (where $\kappa \ll 1$). Imperfect CSI can result for several reasons, including channel time-variation, additive noise, and DR limitations. We focus on CSI imperfections that result from the use of pilot-aided least-squares (LS) channel estimation performed in the presence of limited-DR. Moreover, we consider regenerative relays that decode-and-forward (as in [2]–[6]), as opposed to simpler non-regenerative relays that only amplify-and-forward (also discussed in [4]).

The contributions of this paper (an abbreviated version of [7]) are as follows. For full-duplex MIMO relaying, an explicit model for transmitter/receiver-DR limitations is proposed; pilot-aided least-squares MIMO-channel estimation, under DR limitations, is analyzed; the residual self-interference, from DR limitations and channel-estimation error, is analyzed; lower and upper bounds on the achievable rate are derived; a transmission scheme is proposed based on maximizing the achievable-rate lower bound subject to a power constraint, requiring the solution of a nonconvex optimization problem, to which we apply bisection search and Gradient Projection; an analytic approximation of the maximum achievable rate is proposed; and, the achievable rate is numerically investigated as a function of signal-to-noise ratio, interference-to-noise ratio, and transmitter/receiver dynamic range.

II. System Model

We will use N_{s} and N_{r} to denote the number of transmit antennas at the source and relay, respectively, and M_{r} and M_{d} to denote the number of receive antennas at the relay and destination, respectively. Here and in the sequel, we use the

[1]This work was sponsored by the United States Air Force under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

subscripts s for source, r for relay, and d for destination, and we omit subscripts when referring to common quantities.

We assume that propagation between each transmitter-receiver pair can be characterized by a Raleigh-fading MIMO channel $\boldsymbol{H} \in \mathbb{C}^{M\times N}$ corrupted by additive white Gaussian noise (AWGN) $\boldsymbol{n}(t)$. By "Rayleigh fading," we mean that $\mathrm{vec}(\boldsymbol{H}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{MN})$, and by "AWGN," we mean that $\boldsymbol{n}(t) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$. The time-$t$ radiated signals $\boldsymbol{s}(t)$ are then related to the received signals $\boldsymbol{u}(t)$ via

$$\boldsymbol{u}_{\mathsf{r}}(t) \;=\; \sqrt{\rho_{\mathsf{r}}}\boldsymbol{H}_{\mathsf{sr}}\boldsymbol{s}_{\mathsf{s}}(t) + \sqrt{\eta_{\mathsf{r}}}\boldsymbol{H}_{\mathsf{rr}}\boldsymbol{s}_{\mathsf{r}}(t) + \boldsymbol{n}_{\mathsf{r}}(t) \qquad (1)$$

$$\boldsymbol{u}_{\mathsf{d}}(t) \;=\; \sqrt{\rho_{\mathsf{d}}}\boldsymbol{H}_{\mathsf{rd}}\boldsymbol{s}_{\mathsf{r}}(t) + \sqrt{\eta_{\mathsf{d}}}\boldsymbol{H}_{\mathsf{sd}}\boldsymbol{s}_{\mathsf{d}}(t) + \boldsymbol{n}_{\mathsf{d}}(t). \qquad (2)$$

In (1)-(2), $\rho_{\mathsf{r}} > 0$ and $\rho_{\mathsf{d}} > 0$ denote the signal-to-noise ratio (SNR) at the relay and destination, while $\eta_{\mathsf{r}} > 0$ and $\eta_{\mathsf{d}} > 0$ denote the interference-to-noise ratio (INR) at the relay and destination. (As described in the sequel, the destination treats the source-to-destination link as interference). The INR η_{r} will depend on the separation between, and orientation of, the relay's transmit and receive antenna arrays, whereas the INR η_{d} will depend on the separation between source and destination modems, so that typically $\eta_{\mathsf{d}} \ll \eta_{\mathsf{r}}$.

For full-duplex decode-and-forward relaying, we partition the time indices $t = 0, 1, 2, \ldots$ into a sequence of communication epochs $\{\mathcal{T}_i\}_{i=0}^{\infty}$ where, during epoch $\mathcal{T}_i \subset \mathbb{Z}^+$, the source communicates the i^{th} information packet to the relay, while simultaneously the relay communicates the $(i-1)^{th}$ information packet to the destination. Before the first data communication epoch, we assume the existence of a training epoch $\mathcal{T}_{\mathsf{train}}$ during which the modems estimate the channel state. From the estimated channel state, the data communication design parameters are optimized and the resulting parameters are used for every data communication epoch. Since the design and analysis will be identical for every data-communication epoch (as a consequence of channel time-invariance), we suppress the index i in the sequel and refer to an arbitrary data communication epoch as $\mathcal{T}_{\mathsf{data}}$.

The training epoch is partitioned into two equal-length periods (i.e., $\mathcal{T}_{\mathsf{train}}[1]$ and $\mathcal{T}_{\mathsf{train}}[2]$) to avoid self-interference when estimating the channel matrices. Each data epoch is also partitioned into two periods (i.e., $\mathcal{T}_{\mathsf{data}}[1]$ and $\mathcal{T}_{\mathsf{data}}[2]$) of normalized duration $\tau \in [0,1]$ and $1-\tau$, respectively, over which the transmission parameters can be independently optimized. As we shall see in the sequel, such flexibility is critical when the INR η_{r} is large relative to the SNR ρ_{r}. Moreover, this latter partitioning allows us to formulate both half- and full-duplex schemes as special cases of a more general transmission protocol. For use in the sequel, we find it convenient to define $\tau[1] \triangleq \tau$ and $\tau[2] \triangleq 1-\tau$. Within each of these periods, we assume that the transmitted signals are zero-mean and wide-sense stationary.

We model the effect of limited transmitter dynamic range (DR) by injecting, per transmit antenna, an independent zero-mean Gaussian "transmitter noise" whose variance is κ times the energy of the *intended* transmit signal at that antenna. In particular, say that $\boldsymbol{x}(t) \in \mathbb{C}^N$ denotes the transmitter's intended time-t transmit signal, and say $\boldsymbol{Q} \triangleq \mathrm{Cov}\{\boldsymbol{x}(t)\}$ over the relevant time period (e.g., $t \in \mathcal{T}_{\mathsf{data}}[1]$). We then write the time-t noisy radiated signal as

$$\boldsymbol{s}(t) = \boldsymbol{x}(t) + \boldsymbol{c}(t) \text{ s.t. } \begin{cases} \boldsymbol{c}(t) \sim \mathcal{CN}(\boldsymbol{0}, \kappa\,\mathrm{diag}(\boldsymbol{Q})) \\ \boldsymbol{c}(t) \perp\!\!\!\perp \boldsymbol{x}(t) \\ \boldsymbol{c}(t) \perp\!\!\!\perp \boldsymbol{c}(t')\big|_{t'\neq t} \end{cases}, \qquad (3)$$

where $\boldsymbol{c}(t) \in \mathbb{C}^N$ denotes transmitter noise and $\perp\!\!\!\perp$ statistical independence. Typically, $\kappa \ll 1$.

We model the effect of limited receiver-DR by injecting, per receive antenna, an independent zero-mean Gaussian "receiver distortion" whose variance is β times the energy collected by that antenna. In particular, say that $\boldsymbol{u}(t) \in \mathbb{C}^M$ denotes the receiver's undistorted time-t received vector, and say $\boldsymbol{\Phi} \triangleq \mathrm{Cov}\{\boldsymbol{u}(t)\}$ over the relevant time period (e.g., $t \in \mathcal{T}_{\mathsf{data}}[1]$). We then write the distorted post-ADC received signal as

$$\boldsymbol{y}(t) = \boldsymbol{u}(t) + \boldsymbol{e}(t) \text{ s.t. } \begin{cases} \boldsymbol{e}(t) \sim \mathcal{CN}(\boldsymbol{0}, \beta\,\mathrm{diag}(\boldsymbol{\Phi})) \\ \boldsymbol{e}(t) \perp\!\!\!\perp \boldsymbol{u}(t) \\ \boldsymbol{e}(t) \perp\!\!\!\perp \boldsymbol{e}(t')\big|_{t'\neq t} \end{cases}, \qquad (4)$$

where $\boldsymbol{e}(t) \in \mathbb{C}^M$ is additive distortion. Typically, $\beta \ll 1$. Justifications for these limited-DR models are given in [7]. Figure 1 summarizes our overall system model.

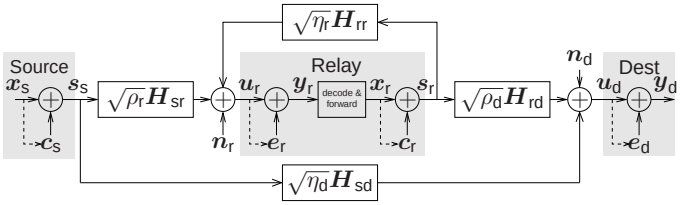

Fig. 1. Our model of full-duplex MIMO relaying under limited transmitter/receiver-DR. The dashed lines denote statistical dependence.

III. Analysis of Achievable Rate

A. Pilot-Aided Channel Estimation

We assume that, during the training epochs $\mathcal{T}_{\mathsf{train}}[1]$ and $\mathcal{T}_{\mathsf{train}}[2]$, least-squares estimates of the channels $\hat{\boldsymbol{H}}_{\mathsf{sr}}, \hat{\boldsymbol{H}}_{\mathsf{rr}}, \hat{\boldsymbol{H}}_{\mathsf{rd}}, \hat{\boldsymbol{H}}_{\mathsf{sd}}$ are obtained from TN-duration (for some $T \in \mathbb{Z}^+$) training sequences, according to the method described in [7]. There it was shown that the resulting estimates take the form

$$\sqrt{\alpha}\hat{\boldsymbol{H}} \;=\; \sqrt{\alpha}\boldsymbol{H} + \boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}}, \qquad (5)$$

where the entries of $\tilde{\boldsymbol{H}}$ are i.i.d $\mathcal{CN}(0,1)$, and where

$$\boldsymbol{D} \approx \tfrac{1}{2T}\big[\boldsymbol{I} + \alpha\tfrac{2\kappa}{N}\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}} + \alpha\tfrac{2\beta}{N}\,\mathrm{diag}(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})\big]. \qquad (6)$$

characterizes the spatial covariance of the estimation error under $\beta \ll 1$ and $\kappa \ll 1$. Above, $\alpha \in \{\rho_{\mathsf{r}}, \eta_{\mathsf{r}}, \rho_{\mathsf{d}}, \eta_{\mathsf{d}}\}$ for $\boldsymbol{H} \in \{\boldsymbol{H}_{\mathsf{sr}}, \boldsymbol{H}_{\mathsf{rr}}, \boldsymbol{H}_{\mathsf{rd}}, \boldsymbol{H}_{\mathsf{sd}}\}$, respectively.

B. Interference Cancellation and Equivalent Channel

Recalling that the data communication period is partitioned into two periods, $\mathcal{T}_{\mathsf{data}}[1]$ and $\mathcal{T}_{\mathsf{data}}[2]$, and that—within each—the transmitted signals are wide-sense stationary, the relay's (instantaneous, distorted) signal at any time $t \in \mathcal{T}_{\mathsf{data}}[l]$ is

$$\begin{aligned}\boldsymbol{y}_{\mathsf{r}}(t) \;=\; & (\sqrt{\rho_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{sr}} - \boldsymbol{D}_{\mathsf{sr}}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{\mathsf{sr}})(\boldsymbol{x}_{\mathsf{s}}(t)+\boldsymbol{c}_{\mathsf{s}}(t)) + \boldsymbol{n}_{\mathsf{r}}(t) \\ & + (\sqrt{\eta_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{rr}} - \boldsymbol{D}_{\mathsf{rr}}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{\mathsf{rr}})(\boldsymbol{x}_{\mathsf{r}}(t)+\boldsymbol{c}_{\mathsf{r}}(t)) + \boldsymbol{e}_{\mathsf{r}}(t), \quad (7)\end{aligned}$$

as implied by Fig. 1 and (5). Defining the aggregate noise term

$$\begin{aligned}\boldsymbol{v}_{\mathsf{r}}(t) \;\triangleq\; & \sqrt{\rho_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{sr}}\boldsymbol{c}_{\mathsf{s}}(t) - \boldsymbol{D}_{\mathsf{sr}}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{\mathsf{sr}}(\boldsymbol{x}_{\mathsf{s}}(t)+\boldsymbol{c}_{\mathsf{s}}(t)) + \boldsymbol{n}_{\mathsf{r}}(t) \\ & + \sqrt{\eta_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{rr}}\boldsymbol{c}_{\mathsf{r}}(t) - \boldsymbol{D}_{\mathsf{rr}}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{\mathsf{rr}}(\boldsymbol{x}_{\mathsf{r}}(t)+\boldsymbol{c}_{\mathsf{r}}(t)) + \boldsymbol{e}_{\mathsf{r}}(t), \quad (8)\end{aligned}$$

we can write $\boldsymbol{y}_{\mathsf{r}}(t) = \sqrt{\rho_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{sr}}\boldsymbol{x}_{\mathsf{s}}(t) + \sqrt{\eta_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{rr}}\boldsymbol{x}_{\mathsf{r}}(t) + \boldsymbol{v}_{\mathsf{r}}(t)$, where the self-interference term $\sqrt{\eta_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{rr}}\boldsymbol{x}_{\mathsf{r}}(t)$ is known and thus can be canceled. The interference-canceled signal $\boldsymbol{z}_{\mathsf{r}}(t) \triangleq \boldsymbol{y}_{\mathsf{r}}(t) - \sqrt{\eta_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{rr}}\boldsymbol{x}_{\mathsf{r}}(t)$ can then be written as

$$\boldsymbol{z}_{\mathsf{r}}(t) \;=\; \sqrt{\rho_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{sr}}\boldsymbol{x}_{\mathsf{s}}(t) + \boldsymbol{v}_{\mathsf{r}}(t). \quad (9)$$

Equation (9) shows that, in effect, the information signal $\boldsymbol{x}_{\mathsf{s}}(t)$ propagates through a known channel $\sqrt{\rho_{\mathsf{r}}}\hat{\boldsymbol{H}}_{\mathsf{sr}}$ corrupted by an aggregate (possibly non-Gaussian) noise $\boldsymbol{v}_{\mathsf{r}}(t)$, whose $(\hat{\boldsymbol{H}}_{\mathsf{sr}}, \hat{\boldsymbol{H}}_{\mathsf{rr}})$-conditional covariance we denote as $\hat{\boldsymbol{\Sigma}}_{\mathsf{r}}[l] \triangleq \mathrm{Cov}\{\boldsymbol{v}_{\mathsf{r}}(t) \,|\, \hat{\boldsymbol{H}}_{\mathsf{sr}}, \hat{\boldsymbol{H}}_{\mathsf{rr}}\}_{t\in\mathcal{T}_{\mathsf{data}}[l]}$. It can be shown [7] that

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\mathsf{r}}[l] \;\approx\; & \boldsymbol{I} + \kappa\rho_{\mathsf{r}}\hat{\boldsymbol{H}}_{\mathsf{sr}}\,\mathrm{diag}(\boldsymbol{Q}_{\mathsf{s}}[l])\hat{\boldsymbol{H}}_{\mathsf{sr}}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{\mathsf{sr}}\,\mathrm{tr}(\boldsymbol{Q}_{\mathsf{s}}[l]) \\ & + \kappa\eta_{\mathsf{r}}\hat{\boldsymbol{H}}_{\mathsf{rr}}\,\mathrm{diag}(\boldsymbol{Q}_{\mathsf{r}}[l])\hat{\boldsymbol{H}}_{\mathsf{rr}}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{\mathsf{rr}}\,\mathrm{tr}(\boldsymbol{Q}_{\mathsf{r}}[l]) \\ & + \beta\rho_{\mathsf{r}}\,\mathrm{diag}(\hat{\boldsymbol{H}}_{\mathsf{sr}}\boldsymbol{Q}_{\mathsf{s}}[l]\hat{\boldsymbol{H}}_{\mathsf{sr}}^{\mathsf{H}}) \\ & + \beta\eta_{\mathsf{r}}\,\mathrm{diag}(\hat{\boldsymbol{H}}_{\mathsf{rr}}\boldsymbol{Q}_{\mathsf{r}}[l]\hat{\boldsymbol{H}}_{\mathsf{rr}}^{\mathsf{H}}), \quad (10)\end{aligned}$$

where $\hat{\boldsymbol{D}}_{\mathsf{sr}} \triangleq \mathrm{E}\{\boldsymbol{D}_{\mathsf{sr}} \,|\, \hat{\boldsymbol{H}}_{\mathsf{sr}}\}$ and $\hat{\boldsymbol{D}}_{\mathsf{rr}} \triangleq \mathrm{E}\{\boldsymbol{D}_{\mathsf{rr}} \,|\, \hat{\boldsymbol{H}}_{\mathsf{rr}}\}$ obey

$$\hat{\boldsymbol{D}} \;\approx\; \tfrac{1}{2T}\big[\boldsymbol{I} + \alpha\tfrac{2\kappa}{N}\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}} + \alpha\tfrac{2\beta}{N}\,\mathrm{diag}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}})\big] \quad (11)$$

and where the approximations in (10)-(11) hold under $\kappa \ll 1$ and $\beta \ll 1$. An expression similar to (10) can be derived for $\hat{\boldsymbol{\Sigma}}_{\mathsf{d}}[l] \triangleq \mathrm{Cov}\{\boldsymbol{v}_{\mathsf{d}}(t) \,|\, \hat{\boldsymbol{H}}_{\mathsf{rd}}, \hat{\boldsymbol{H}}_{\mathsf{sd}}\}_{t\in\mathcal{T}_{\mathsf{data}}[l]}$.

C. Bounds on Achievable Rate

The end-to-end mutual information can be written, for a given time-sharing parameter τ, as [1]

$$I_\tau(\boldsymbol{\mathcal{Q}}) \;=\; \min\left\{\sum_{l=1}^{2}\tau[l]I_{\mathsf{sr}}(\boldsymbol{\mathcal{Q}}[l]), \sum_{l=1}^{2}\tau[l]I_{\mathsf{rd}}(\boldsymbol{\mathcal{Q}}[l])\right\}, \quad (12)$$

where $I_{\mathsf{sr}}(\boldsymbol{\mathcal{Q}}[l])$ and $I_{\mathsf{rd}}(\boldsymbol{\mathcal{Q}}[l])$ are the period-l mutual informations of the source-to-relay channel and relay-to-destination channel, respectively, and where $\boldsymbol{\mathcal{Q}}[l] \triangleq \big(\boldsymbol{Q}_{\mathsf{s}}[l], \boldsymbol{Q}_{\mathsf{r}}[l]\big)$ and $\boldsymbol{\mathcal{Q}} \triangleq \big(\boldsymbol{\mathcal{Q}}[1], \boldsymbol{\mathcal{Q}}[2]\big)$.

Mutual-information analysis is complicated by the fact that the aggregate noise terms $\boldsymbol{v}_{\mathsf{r}}$ and $\boldsymbol{v}_{\mathsf{d}}$ are, in general, non-Gaussian as a result of the channel-estimation-error components. However, it is known that, among all noise distributions of a given covariance, the Gaussian one is worst from a mutual-information perspective [8]. Thus, $I_{\mathsf{sr}}(\boldsymbol{\mathcal{Q}}[l])$ can be lower-bounded by [7]

$$\begin{aligned}\underline{I}_{\mathsf{sr}}(\boldsymbol{\mathcal{Q}}[l]) &= \log\det\big(\boldsymbol{I} + \rho_{\mathsf{r}}\hat{\boldsymbol{H}}_{\mathsf{sr}}\boldsymbol{Q}_{\mathsf{s}}[l]\hat{\boldsymbol{H}}_{\mathsf{sr}}^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}_{\mathsf{r}}^{-1}[l]\big) \quad (13)\\ &= \log\det\big(\rho_{\mathsf{r}}\hat{\boldsymbol{H}}_{\mathsf{sr}}\boldsymbol{Q}_{\mathsf{s}}[l]\hat{\boldsymbol{H}}_{\mathsf{sr}}^{\mathsf{H}} + \hat{\boldsymbol{\Sigma}}_{\mathsf{r}}[l]\big) - \log\det(\hat{\boldsymbol{\Sigma}}_{\mathsf{r}}[l])\end{aligned}$$

and $I_{\mathsf{rd}}(\boldsymbol{\mathcal{Q}}[l])$ similarly lower bounded by $\underline{I}_{\mathsf{rd}}(\boldsymbol{\mathcal{Q}}[l])$. The end-to-end τ-specific achievable-rate is then lower-bounded by

$$\underline{I}_\tau(\boldsymbol{\mathcal{Q}}) \;=\; \min\Bigg\{\underbrace{\sum_{l=1}^{2}\tau[l]\underline{I}_{\mathsf{sr}}(\boldsymbol{\mathcal{Q}}[l])}_{\triangleq\, \underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}})}, \underbrace{\sum_{l=1}^{2}\tau[l]\underline{I}_{\mathsf{rd}}(\boldsymbol{\mathcal{Q}}[l])}_{\triangleq\, \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})}\Bigg\}. \quad (14)$$

Moreover, the rate $\underline{I}_\tau(\boldsymbol{\mathcal{Q}})$ bits-per-channel-use (bpcu) can be achieved via independent Gaussian codebooks at the transmitters and maximum-likelihood detection at the receivers [8].

A straightforward achievable-rate upper bound $\overline{I}_\tau(\boldsymbol{\mathcal{Q}})$ results from the case of perfect CSI (i.e., $\hat{\boldsymbol{D}} = \boldsymbol{0}$), where $\boldsymbol{v}_{\mathsf{r}}(t)$ and $\boldsymbol{v}_{\mathsf{d}}(t)$ are Gaussian.

IV. Transmit Covariance Optimization

We would now like to find the transmit covariance matrices $\boldsymbol{\mathcal{Q}}$ that maximize the achievable-rate lower bound $\underline{I}_\tau(\boldsymbol{\mathcal{Q}})$ in (14) subject to the per-link power constraint $\boldsymbol{\mathcal{Q}} \in \mathbb{Q}_\tau$, where

$$\begin{aligned}\mathbb{Q}_\tau \triangleq \Bigg\{\boldsymbol{\mathcal{Q}} \text{ s.t. } & \sum_{l=1}^{2}\tau[l]\,\mathrm{tr}\big(\boldsymbol{Q}_{\mathsf{s}}[l]\big) \le 1, \sum_{l=1}^{2}\tau[l]\,\mathrm{tr}\big(\boldsymbol{Q}_{\mathsf{r}}[l]\big) \le 1, \\ & \boldsymbol{Q}_{\mathsf{s}}[l] = \boldsymbol{Q}_{\mathsf{s}}^{\mathsf{H}}[l] \ge 0,\, \boldsymbol{Q}_{\mathsf{r}}[l] = \boldsymbol{Q}_{\mathsf{r}}^{\mathsf{H}}[l] \ge 0\Bigg\}, \quad (15)\end{aligned}$$

and subsequently optimize the time-sharing parameter τ. We now denote the optimal (i.e., maximin) rate, for a given τ, by

$$\underline{I}_{*,\tau} \triangleq \max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_\tau} \min\big\{\underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}}), \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})\big\}, \quad (16)$$

and we use $\mathbb{Q}_{*,\tau}$ to denote the corresponding set of maximin designs $\boldsymbol{\mathcal{Q}}$ (which are, in general, not unique). Then, with $\tau_* \triangleq \arg\max_{\tau\in[0,1]} \underline{I}_{*,\tau}$, the optimal rate is $\underline{I}_* \triangleq \underline{I}_{*,\tau_*}$, and the corresponding set of maximin designs is $\mathbb{Q}_* \triangleq \mathbb{Q}_{*,\tau_*}$.

It is important to realize that, among the maximin designs $\mathbb{Q}_{*,\tau}$, there exists at least one "link-equalizing" design, i.e., $\exists\boldsymbol{\mathcal{Q}} \in \mathbb{Q}_{*,\tau}$ s.t. $\underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}}) = \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})$. To see why this is the case, notice that, given any maximin design $\boldsymbol{\mathcal{Q}}$ such that $\underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}}) > \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})$, a simple scaling of $\boldsymbol{Q}_{\mathsf{s}}[l]$ can yield $\underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}}) = \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})$, and thus an equalizing design. A similar argument can be made when $\underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}}) > \underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}})$.

Referring to the set of *all* link-equalizing designs (maximin or otherwise), for a given τ, as

$$\mathbb{Q}_{=,\tau} \triangleq \big\{\boldsymbol{\mathcal{Q}} \in \mathbb{Q}_\tau \;\text{ s.t. }\; \underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}}) = \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})\big\}, \quad (17)$$

the maximin equalizing design can be found by solving either $\arg\max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_{=,\tau}} \underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}})$ or $\arg\max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_{=,\tau}} \underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}})$, where the equivalence is due to the equalizing property. More generally, the maximin equalizing design can be found by solving

$$\arg\max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_{=,\tau}} \underline{I}_\tau(\boldsymbol{\mathcal{Q}}, \zeta) \quad (18)$$

with *any* fixed $\zeta \in [0,1]$ and the ζ-weighted sum-rate

$$\underline{I}_\tau(\boldsymbol{\mathcal{Q}},\zeta) \triangleq \zeta \underline{I}_{\mathsf{sr},\tau}(\boldsymbol{\mathcal{Q}}) + (1-\zeta)\underline{I}_{\mathsf{rd},\tau}(\boldsymbol{\mathcal{Q}}). \tag{19}$$

To find the maximin equalizing design, we propose relaxing the constraint on $\boldsymbol{\mathcal{Q}}$ from $\mathbb{Q}_{=,\tau}$ to $\mathbb{Q}_\tau$, yielding the ζ-weighted-sum-rate optimization problem

$$\boldsymbol{\mathcal{Q}}_{*,\tau}(\zeta) = \arg\max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_\tau} \underline{I}_\tau(\boldsymbol{\mathcal{Q}},\zeta). \tag{20}$$

At each bisection step, we use Gradient Projection (GP) to solve[2] the τ-specific, ζ-weighted-sum-rate optimization problem (20). See [7] for details.

V. Achievable-Rate Approximation

The complicated nature of the optimization problem (16) motivates us to approximate its solution, i.e., the covariance-optimized achievable rate $\underline{I}_* = \max_{\tau\in[0,1]}\max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_\tau}\underline{I}_\tau(\boldsymbol{\mathcal{Q}})$. In doing so, we focus on the case of $T\to\infty$, where channel estimation error is driven to zero so that $\underline{I}_\tau(\boldsymbol{\mathcal{Q}}) = I_\tau(\boldsymbol{\mathcal{Q}}) = \overline{I}_\tau(\boldsymbol{\mathcal{Q}})$. In addition, for tractability, we restrict ourselves to the case $N_{\mathsf{s}} = N_{\mathsf{r}} = N$ and $M_{\mathsf{r}} = M_{\mathsf{d}} = M$ (i.e., N transmit antennas and M receive antennas at each node), the case $\eta_{\mathsf{d}} = 0$ (i.e., no direct source-to-destination link), and the case $\tau = \frac{1}{2}$ (i.e., equal time-sharing).

Our approximation is built around the simplifying case that the channel matrices $\{\boldsymbol{H}_{\mathsf{sr}}, \boldsymbol{H}_{\mathsf{rr}}, \boldsymbol{H}_{\mathsf{rd}}\}$ are each diagonal, although not necessarily square, and have $R \triangleq \min\{M,N\}$ identical diagonal entries equal to $\sqrt{MN/R}$. (The latter value is chosen so that $\mathrm{E}\{\mathrm{tr}(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})\} = MN$ as assumed in Section II.) In this case, the mutual information (14) becomes

$$\begin{aligned} I_\tau(\boldsymbol{\mathcal{Q}}) \approx \frac{1}{2}\min\Bigg\{ &\sum_{l=1}^{2}\log\det\Big(\boldsymbol{I} + \rho_{\mathsf{r}}\tfrac{NM}{R}\boldsymbol{Q}_{\mathsf{s}}[l]\big(\boldsymbol{I} + (\kappa+\beta) \\ &\times \tfrac{NM}{R}\big[\rho_{\mathsf{r}}\,\mathrm{diag}(\boldsymbol{Q}_{\mathsf{s}}[l]) + \eta_{\mathsf{r}}\,\mathrm{diag}(\boldsymbol{Q}_{\mathsf{r}}[l])\big]\big)^{-1}\Big), \\ &\sum_{l=1}^{2}\log\det\Big(\boldsymbol{I} + \rho_{\mathsf{d}}\tfrac{NM}{R}\boldsymbol{Q}_{\mathsf{r}}[l]\big(\boldsymbol{I} + (\kappa+\beta) \\ &\times \tfrac{NM}{R}\rho_{\mathsf{d}}\,\mathrm{diag}(\boldsymbol{Q}_{\mathsf{r}}[l])\big)^{-1}\Big)\Bigg\}. \end{aligned} \tag{21}$$

When $\eta_{\mathsf{r}} \ll \rho_{\mathsf{r}}$, the η_{r}-dependent terms in (21) can be ignored, after which it is straightforward to show that, under the constraint (15), the optimal covariances are the "full duplex" $\boldsymbol{\mathcal{Q}}_{\mathsf{FD}} \triangleq (\frac{1}{N}\boldsymbol{I}, \frac{1}{N}\boldsymbol{I}, \frac{1}{N}\boldsymbol{I}, \frac{1}{N}\boldsymbol{I})$, for which (21) gives

$$\begin{aligned} I(\boldsymbol{\mathcal{Q}}_{\mathsf{FD}}) &\approx R\log\Big(1+\min\Big\{\tfrac{\rho_{\mathsf{r}}}{\frac{R}{M}+(\kappa+\beta)(\rho_{\mathsf{r}}+\eta_{\mathsf{r}})}, \tfrac{\rho_{\mathsf{d}}}{\frac{R}{M}+(\kappa+\beta)\rho_{\mathsf{d}}}\Big\}\Big) \\ &= \begin{cases} R\log\Big(1+\frac{\rho_{\mathsf{d}}}{\frac{R}{M}+(\kappa+\beta)\rho_{\mathsf{d}}}\Big) & \text{if } \frac{\rho_{\mathsf{r}}}{\rho_{\mathsf{d}}} \geq 1+\frac{(\kappa+\beta)\eta_{\mathsf{r}}M}{R} \\ R\log\Big(1+\frac{\rho_{\mathsf{r}}}{\frac{R}{M}+(\kappa+\beta)(\rho_{\mathsf{r}}+\eta_{\mathsf{r}})}\Big) & \text{else.} \end{cases} \end{aligned} \tag{22}$$

When $\eta_{\mathsf{r}} \gg \rho_{\mathsf{r}}$, the η_{r}-dependent term in (21) dominates unless $\boldsymbol{Q}_{\mathsf{r}}[l] = \boldsymbol{0}$. In this case, the optimal covariances are the "half duplex" ones $\boldsymbol{\mathcal{Q}}_{\mathsf{HD}} \triangleq (\frac{2}{N}\boldsymbol{I}, \boldsymbol{0}, \boldsymbol{0}, \frac{2}{N}\boldsymbol{I})$, for which (21) gives

$$I(\boldsymbol{\mathcal{Q}}_{\mathsf{HD}}) \approx \begin{cases} \frac{R}{2}\log\Big(1+\frac{\rho_{\mathsf{d}}}{\frac{R}{2M}+(\kappa+\beta)\rho_{\mathsf{d}}}\Big) & \text{if } \frac{\rho_{\mathsf{r}}}{\rho_{\mathsf{d}}} \geq 1 \\ \frac{R}{2}\log\Big(1+\frac{\rho_{\mathsf{r}}}{\frac{R}{2M}+(\kappa+\beta)\rho_{\mathsf{r}}}\Big) & \text{else.} \end{cases} \tag{23}$$

Finally, given any triple $(\rho_{\mathsf{r}}, \eta_{\mathsf{r}}, \rho_{\mathsf{d}})$, we approximate the achievable rate as follows: $I_* \approx \max\{I(\boldsymbol{\mathcal{Q}}_{\mathsf{FD}}), I(\boldsymbol{\mathcal{Q}}_{\mathsf{HD}})\}$.

From (22)-(23), using $\theta \triangleq \frac{R}{M(\kappa+\beta)}$, it is straightforward to show that the approximated system operates as follows.

1) Say $\frac{\rho_{\mathsf{r}}}{\rho_{\mathsf{d}}} \leq 1$. Then full-duplex is used iff
$$\eta_{\mathsf{r}} \leq \tfrac{1}{2}\sqrt{(\theta+2\rho_{\mathsf{r}})^2 + \tfrac{2\rho_{\mathsf{r}}}{\kappa+\beta}(\theta+2\rho_{\mathsf{r}})} - \tfrac{1}{2}\theta. \tag{24}$$
For either half- or full-duplex, I_* is invariant to ρ_{d}, i.e., the source-to-relay link is the limiting one.
2) Say $1 \leq \frac{\rho_{\mathsf{r}}}{\rho_{\mathsf{d}}} \leq 1 + \frac{(\kappa+\beta)\eta_{\mathsf{r}}M}{R}$. Full-duplex is used iff
$$\eta_{\mathsf{r}} \leq \tfrac{\rho_{\mathsf{r}}}{2\rho_{\mathsf{d}}}\sqrt{(\theta+2\rho_{\mathsf{d}})^2 + \tfrac{2\rho_{\mathsf{d}}}{\kappa+\beta}(\theta+2\rho_{\mathsf{d}})} - \theta\big(1-\tfrac{\rho_{\mathsf{r}}}{2\rho_{\mathsf{d}}}\big). \tag{25}$$
3) Say $1 + \frac{(\kappa+\beta)\eta_{\mathsf{r}}M}{R} \leq \frac{\rho_{\mathsf{r}}}{\rho_{\mathsf{d}}}$, or equivalently $\eta_{\mathsf{r}} \leq \eta_{\mathsf{crit}} \triangleq \big(\frac{\rho_{\mathsf{r}}}{\rho_{\mathsf{d}}}-1\big)\frac{R}{M(\kappa+\beta)}$. Then full-duplex is always used, and I_* is invariant to ρ_{r} and η_{r}, i.e., the rate is limited by the relay-to-destination link.

Figure 2 shows a contour plot of the proposed achievable-rate approximation as a function of INR η_{r} and SNR ρ_{r}, for the case that $\rho_{\mathsf{r}}/\rho_{\mathsf{d}} = 2$. We shall see in Section VI that our approximation of the covariance-optimized achievable-rate is reasonably close to that found by solving (16) using bisection/GP.

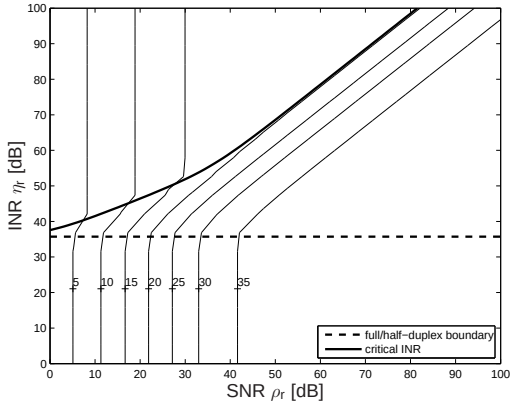

Fig. 2. Contour plot of the approximated achievable rate I_* versus relay SNR ρ_{r} and INR η_{r}, for $N=3$, $M=4$, $\beta=\kappa=-40$dB, and $\rho_{\mathsf{r}}/\rho_{\mathsf{d}}=2$. The horizontal dashed line shows the INR η_{crit}, and the dark curve shows the boundary between full- and half-duplex regimes described in (25).

VI. Numerical Results and Conclusions

We now study the average behavior of the bisection/GP-optimized rate $\underline{I}_* = \max_\tau\max_{\boldsymbol{\mathcal{Q}}\in\mathbb{Q}_\tau}\underline{I}_\tau(\boldsymbol{\mathcal{Q}})$ as a function of SNRs ρ_{r} and ρ_{d}; INRs η_{r} and η_{d}; and dynamic range parameters κ and β. We also investigate the role of interference cancellation, the role of two distinct data periods, the role of τ-optimization, and the relation to optimized half-duplex (OHD) signaling. In doing so, we find close agreement with the achievable-rate approximation proposed in Section V and

[2]Because (16) is generally non-convex, finding the global maximum can be difficult. Although GP is guaranteed only to find a local, and not global, maximum, our experience with different initializations suggests that GP is indeed finding the global maximum in our problem.

illustrated in Fig. 2. All results below used $N \triangleq N_{\mathsf{s}} = N_{\mathsf{r}}$ transmit antennas, $M \triangleq M_{\mathsf{r}} = M_{\mathsf{d}}$ receive antennas, the SNR ratio $\rho_{\mathsf{r}}/\rho_{\mathsf{d}} = 2$, the destination INR $\eta_{\mathsf{d}} = 1$, training duration $T = 50$, optimization of time-share $\tau \in \{0.1, 0.2, \ldots, 0.9\}$, and were averaged over 100 realizations.

Below, we denote the full scheme proposed in Section IV by "TCO-2-IC," which indicates the use of interference cancellation (IC) and transmit covariance optimization (TCO) performed individually over the 2 data periods (i.e., $\mathcal{T}_{\mathsf{data}}[1]$ and $\mathcal{T}_{\mathsf{data}}[2]$). To test the impact of IC and of two data periods, we also implemented the proposed scheme but without IC, which we refer to as "TCO-2," as well as the proposed scheme with only one data period (i.e., $\boldsymbol{Q}_i[1] = \boldsymbol{Q}_i[2]\ \forall i$), which we refer to as "TCO-1-IC." To optimize half-duplex, we used GP to maximize the sum-rate $\underline{I}_{\tau}(\boldsymbol{\mathcal{Q}}, \frac{1}{2})$ under the power constraint (15) and the half-duplex constraint $\boldsymbol{Q}_1[2] = \boldsymbol{0} = \boldsymbol{Q}_2[1]$; τ-optimization was performed as described above.

In Fig. 3, we examine achievable-rate performance versus INR η_{r} for the TCO-2-IC, TCO-1-IC, TCO-2, and OHD schemes, using different dynamic range parameters $\beta = \kappa$. For OHD, we see that rate is invariant to INR η_{r}, as expected. For the proposed TCO-2-IC, we observe "full duplex" performance for low-to-mid values of η_{r} and a transition to OHD performance at high values of η_{r}, just as predicted by the approximation in Section V. In fact, the rates in Fig. 3 are very close to the approximated values in Fig. 2. To see the importance of two distinct data-communication periods, we examine the TCO-1-IC trace, where we observe TCO-2-IC-like performance at low-to-midrange values of η_{r}, but performance that drops below OHD at high η_{r}. Essentially, TCO-1-IC forces full-duplex signaling at high INR η_{r}, where half-duplex signaling is optimal, while TCO-2-IC facilitates the possibility of half-duplex signaling through the use of two distinct data-communication periods, similar to the MIMO-interference-channel scheme in [9]. The effect of τ-optimization can be seen by comparing the two OHD traces, one which uses the fixed value $\tau = 0.5$ and the other which uses the optimized value $\tau = \tau_*$. The separation between these traces shows that τ-optimization gives a small but noticeable rate gain. Finally, by examining the TCO-2 trace, we conclude that partial interference cancellation is very important for all but extremely low or high values of INR η_{r}.

In Fig. 4, we examine the rate of the proposed TCO-IC-2 and OHD versus SNR ρ_{r}, using the dynamic range parameters $\beta = \kappa = -40$dB, $\eta_{\mathsf{d}} = 1$, and two fixed values of INR η_{r}. All the behaviors in Fig. 4 are predicted by the rate approximation described in Section V and illustrated in Fig. 2. In particular, at the low INR of $\eta_{\mathsf{r}} = 20$dB, TCO-IC-2 operates in the full-duplex regime for all values of SNR ρ_{r}. Meanwhile, at the high INR of $\eta_{\mathsf{r}} = 60$dB, TCO-IC-2 operates in half-duplex at low values of SNR ρ_{r}, but switches to full-duplex after ρ_{r} exceeds a threshold.

In the full paper [7], we also examine rate performance versus training length T and for various combinations of transmit and receive antennas (M, N).

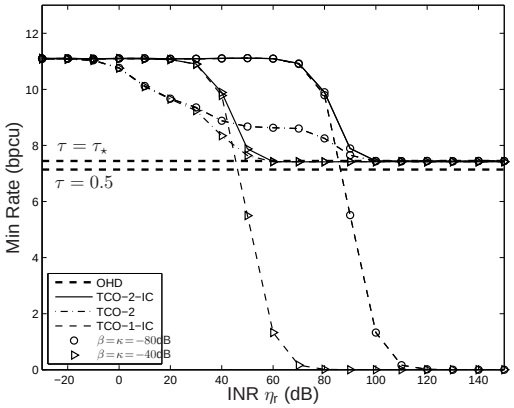

Fig. 3. Achievable-rate lower bound $\underline{I}_*$ for TCO-2-IC, TCO-2, TCO-1-IC, and OHD versus INR η_{r}. Here, $N = 3$, $M = 4$, $\rho_{\mathsf{r}} = 15$dB, $\rho_{\mathsf{r}}/\rho_{\mathsf{d}} = 2$, $\eta_{\mathsf{d}} = 0$dB, and $T = 50$. OHD is plotted for $\beta = \kappa = -40$dB, but was observed to give nearly identical rate for $\beta = \kappa = -80$dB. Both fixed-time-share ($\tau = 0.5$) and optimized-time-share ($\tau = \tau_*$) versions of OHD are shown.

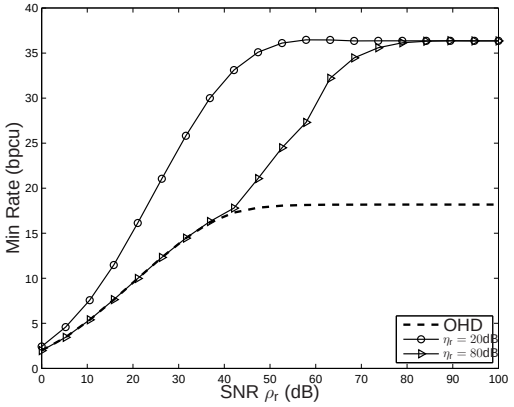

Fig. 4. Achievable-rate lower bound $\underline{I}_*$ for TCO-2-IC and OHD versus SNR ρ_{r}. Here, $\rho_{\mathsf{r}}/\rho_{\mathsf{d}} = 2$, $\eta_{\mathsf{d}} = 0$dB, $N = 3$, $M = 4$, $\beta = \kappa = -40$dB, and $T = 50$. OHD in this figure is optimized over τ.

REFERENCES

[1] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 29–43, Jan. 2005.

[2] D. W. Bliss, P. A. Parker, and A. R. Margetts, "Simultaneous transmission and reception for improved wireless network performance," in *Proc. IEEE Workshop Statist. Signal Process.*, (Madison, WI), pp. 478–482, Aug. 2007.

[3] P. Larsson and M. Prytz, "MIMO on-frequency repeater with self-interference cancellation and mitigation," in *Proc. IEEE Veh. Tech. Conf.*, (Barcelona, Spain), Apr. 2009.

[4] Y. Hua, "An overview of beamforming and power allocation for MIMO relays," in *Proc. IEEE Military Commun. Conf.*, (San Jose, CA), pp. 375–380, Nov. 2010.

[5] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex mimo relays," *IEEE Trans. Signal Process.*, vol. 59, pp. 5983–5993, Dec. 2011.

[6] T. Riihonen, S. Werner, and R. Wichman, "Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation," in *Proc. Asilomar Conf. Signals Syst. Comput.*, (Pacific Grove, CA), Nov. 2011.

[7] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *arXiv:1111.2618*, Nov. 2011.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge University Press, 2005.

[9] Y. Rong and Y. Hua, "Optimal power schedule for distributed MIMO links," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2896–2900, Aug. 2008.