Full-Duplex Bidirectional MIMO: Achievable Rates Under Limited Dynamic Range

Brian P. Day, Adam R. Margetts, Daniel W. Bliss, and Philip Schniter

Abstract—**In this paper, we consider the problem of full-duplex bidirectional communication between a pair of modems, each with multiple transmit and receive antennas. The principal difficulty in implementing such a system is that, due to the close proximity of each modem's transmit antennas to its receive antennas, each modem's outgoing signal can exceed the dynamic range of its input circuitry, making it difficult—if not impossible—to recover the desired incoming signal. To address these challenges, we consider systems that use pilot-aided channel estimates to perform transmit beamforming, receive beamforming, and interference cancellation. Modeling transmitter/receiver dynamic-range limitations explicitly, we derive tight upper and lower bounds on the achievable sum-rate, and propose a transmission scheme based on maximization of the lower bound, which requires us to (numerically) solve a nonconvex optimization problem. In addition, we derive an analytic approximation to the achievable sum-rate, and show, numerically, that it is quite accurate. We then study the behavior of the sum-rate as a function of signal-to-noise ratio, interference-to-noise ratio, transmitter/receiver dynamic range, number of antennas, and training length, using optimized half-duplex signaling as a baseline.**

Index Terms—**Channel estimation, channel models, full-duplex, information theory, limited dynamic range, MIMO, wireless communication.**

I. INTRODUCTION

FULL-DUPLEX bidirectional communication between two multiple-input multiple-output (MIMO) wireless modems has the potential to nearly double the system spectral efficiency [1]. By full-duplex, we mean that the two modems perform simultaneous transmission and reception (STAR) at the same carrier frequency. The fundamental difficulty with STAR is that, due to the close proximity of a given modem's transmit antennas to its receive antennas, the modem's outgoing signal can overwhelm its receiver circuitry, making it impossible to recover the incoming signal. To avoid this problem, existing practical systems tend to communicate in half-duplex mode (e.g., time-division duplex or frequency-division duplex). In

this paper, we propose a realistic system model, including channel estimation errors and effects of limited dynamic range, and derive achievable-rate bounds for a proposed MIMO STAR protocol. It is shown that its spectral efficiency is uniformly better than optimized half-duplex and nearly double when operating within dynamic range constraints. Other advantages to full-duplex communication have been suggested in [2] from a medium-access-layer point of view. A related—but different—application of STAR is that of full-duplex relaying, where a relay node simultaneously receives from a source node and transmits to a destination node (see, e.g., [3] and [4]).

To facilitate STAR, existing techniques (e.g., [1]–[3] and [5]–[9]) propose to use various combinations of i) digital-domain suppression, ii) analog-domain suppression, and iii) propagation-domain suppression. Digital-domain suppression attempts to mitigate self-interference after analog-to-digital conversion (ADC) and will not be effective if the analog circuitry and/or ADCs are already overwhelmed by self-interference. Analog-domain suppression attempts to mitigate self-interference between the receiver's antenna(s) and ADC(s); this method was used early on (e.g., [10]) and continues to be the focus of many recent works (e.g., [2], [5], [9], [11], and [12]). Propagation-domain suppression attempts to mitigate self-interference in the wireless propagation environment, before it has a chance to overwhelm the analog and digital circuitry. This may be accomplished by passive isolation (e.g., via antenna separation and/or shielding), directional or polarized antennas (e.g., [7], [11], and [13]), careful antenna placement (e.g., [2]), and/or transmit beamforming (e.g., [1], [6], [8], and [14]).

In this work, we assume that each of the two modems uses $N_t \geq 1$ antennas for transmission and $N_r \geq 1$ different antennas for reception (i.e., MIMO[1] modems), and we assume a per-modem transmit power constraint. We then consider the problem of jointly optimizing the MIMO transmission and reception strategies in order to maximize the sum of the rates of reliable communication between the two modems (i.e., the sum-rate). Our work thus involves propagation-domain and digital-domain suppression; analog-domain suppression, while implicit in our system model, it is not explicitly discussed.

From our perspective, the primary challenges of MIMO-STAR are, in practice, due to the following:
1) high channel dynamic range (DR);
2) limited transmitter and receiver DR;
3) imperfect channel state information (CSI).
Channel DR refers to the ratio of the (nominal) interference-channel gain to the (nominal) desired-channel gain,

[1]We focus on the MIMO case, where $N_t > 1$ and $N_r > 1$, although our approach is general enough to handle the SIMO, MISO, and SISO cases, where $N_t = 1$ and/or $N_r = 1$.
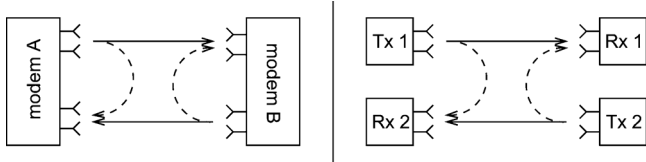
Fig. 1. Full-duplex bidirectional MIMO (left), and the 2×2 MIMO interference channel (right). Solid lines denote desired propagation and dashed lines denote interference.

which may be as high as 100 dB [3] due to the relative separation between intra- and inter-modem antenna pairs. Limited transmitter and receiver-DR is a natural consequence of non-ideal amplifiers, oscillators, ADCs, and digital-to-analog converters (DACs). Imperfect CSI can result for several reasons, including channel time-variation, additive noise, and DR limitations.

Moreover, we claim that it is the *combination* of high channel-DR, limited transmitter/receiver-DR, and imperfect CSI that is particularly challenging. For example, if a modem had infinite transmitter/receiver-DR and perfect CSI, then it could perfectly cancel its own self-interference, even under arbitrarily high channel-DR. In practice, though, because receiver-DR is limited and channel-DR is very high, self-interference can easily overwhelm the receiver circuitry, burying the desired signal component in the noise. To avoid this situation, self-interference must be *prevented* through, e.g., transmitter null-steering. But, with finite transmitter-DR and imperfect CSI, it is impossible to perfectly steer nulls. For all of these reasons, the MIMO-STAR problem is challenging.

Due to the practical importance of transmitter/receiver-DR and imperfect CSI, we model each artifact explicitly in this work. In particular, we model limited transmitter-DR by injecting, for each transmit antenna, an additive white Gaussian "transmitter noise" with variance κ times the energy of the intended transmit signal. Similarly, we model limited receiver-DR by injecting, for each receive antenna, an additive white Gaussian "receiver distortion" with variance β times the energy impinging on that receive antenna. Finally, we model CSI imperfections by assuming the use of pilot-aided least-squares (LS) channel estimation. To facilitate our sum-rate analysis, we assume that the channel is block-fading with a very long fading interval.

We now comment on the relationship between our problem and the two-user MIMO interference channel (ICh) problem (see, e.g., [15]). The MIMO-ICh problem concerns optimal simultaneous communication between two MIMO transmitter/receiver *pairs*, as illustrated by the right panel of Fig. 1. For the MIMO-ICh, the primary challenge is mitigation of other-user interference, and the typical goal is characterizing the maximum achievable rate pair under transmitter power constraints. At high signal-to-noise ratios, it suffices to use the (degrees-of-freedom optimal) *interference alignment* approach [16], wherein the user transmissions are linearly precoded so that the mutual interference is confined to a low-dimensional subspace. On the surface, our full-duplex bidirectional MIMO problem looks similar to the MIMO-ICh problem, except that the transmitter for one communicating pair is located in the same modem as the receiver for the other pair (see Fig. 1, left panel). The latter fact has profound impacts, however, because it implies that, in our

problem, the "other-user" codewords are perfectly known.[2] In fact, in our problem, "other-user" (i.e., self-) interference does not manifest directly, but indirectly through channel-estimation error and limited receiver-DR, both of which can become significant under very high channel-DR (e.g., 100 dB). In comparison, the MIMO-ICh has only rarely been studied under imperfect receiver CSI—see, e.g., the interference-alignment work [17]—and (to our knowledge) has never been studied under limited transmitter/receiver-DR.

The contributions of this paper are as follows. For the full-duplex bidirectional MIMO communication problem:

1) an explicit model for transmitter/receiver-DR limitations is proposed;
2) pilot-aided least-squares MIMO-channel estimation, under DR limitations, is analyzed;
3) the residual self-interference, resulting from DR limitations and channel-estimation error, is analyzed;
4) lower and upper bounds on the achievable sum-rate are derived as a function of the transmit covariance matrices, and a gradient-projection scheme to numerically compute the lower-bound-maximizing covariance matrices, subject to a power constraint, is proposed;
5) an analytic approximation of the maximum achievable sum-rate is proposed that enables one to straightforwardly predict when full-duplex operation can significantly outperform half-duplex operation;
6) explicit transmission and reception schemes are proposed based on sum-rate maximization principles;
7) the achievable sum-rate is numerically investigated as a function of signal-to-noise ratio, interference-to-noise ratio, transmitter/receiver dynamic range, number of antennas, and number of pilots.

Notation: We use $(\cdot)^{\mathsf{T}}$ to denote transpose, $(\cdot)^*$ conjugate, and $(\cdot)^{\mathsf{H}}$ conjugate transpose. For matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{M \times N}$, we use $\operatorname{tr}(\boldsymbol{A})$ to denote trace, $\det(\boldsymbol{A})$ to denote determinant, $\boldsymbol{A} \odot \boldsymbol{B}$ to denote elementwise (i.e., Hadamard) product, $\operatorname{sum}[\boldsymbol{A}] \in \mathbb{C}$ to denote the sum over all elements, $\operatorname{vec}(\boldsymbol{A}) \in \mathbb{C}^{MN}$ to denote vectorization, $\operatorname{diag}(\boldsymbol{A})$ to denote the diagonal matrix with the same diagonal elements as $\boldsymbol{A}$, $\operatorname{Diag}(\boldsymbol{a})$ to denote the diagonal matrix whose diagonal is constructed from the vector $\boldsymbol{a}$, and $[\boldsymbol{A}]_{m,n}$ to denote the element in the mth row and n^{th} column of $\boldsymbol{A}$. We denote expectation by $\operatorname{E}\{\cdot\}$, covariance by $\operatorname{Cov}\{\cdot\}$, statistical independence by $\perp\!\!\!\perp$, the circular complex Gaussian pdf with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{Q}$ by $\mathcal{CN}(\boldsymbol{m}, \boldsymbol{Q})$, and the Kronecker delta by δ_k. Finally, $\boldsymbol{I}$ denotes the identity matrix, $\mathbb{C}$ the complex field, and $\mathbb{Z}$ the integers.

II. System Model

Our bidirectional communication problem involves two modems ("A" and "B"), and thus two communicating transmitter-receiver pairs (as indexed by the variables $i, j \in \{1, 2\}$). We assume, without loss of generality, that modem A houses transmitter $j = 1$ and receiver $i = 2$, whereas modem B houses transmitter $j = 2$ and receiver $i = 1$. In the sequel, we use $t \in \mathbb{Z}^+$ to denote the channel-use index, $\boldsymbol{s}_j(t) \in \mathbb{C}^{N_{\mathrm{t}}}$ to denote the noisy signal radiated by the antenna array of transmitter j, and $\boldsymbol{u}_i(t) \in \mathbb{C}^{N_{\mathrm{r}}}$ to denote the undistorted signal collected

---

[2]Clearly, the MIMO-ICh problem becomes trivial when the other-user codewords are perfectly known.

by the antenna array of receiver i, where N_t is the number of transmit antennas and N_r is the number of receive antennas.

A. Propagation Channel

We assume that the signal radiated by transmitter j and collected by receiver i propagates through an additive white Gaussian noise (AWGN) corrupted Raleigh-fading MIMO channel $\boldsymbol{H}_{ij} \in \mathbb{C}^{N_r \times N_t}$. By "Rayleigh fading," we mean that $\text{vec}(\boldsymbol{H}_{ij}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_r N_t})$. The time-$t$ radiated signals $\{\boldsymbol{s}_j(t)\}_{j=1}^{2}$ are then related to each received signals $\boldsymbol{u}_i(t)$ via

$$\boldsymbol{u}_1(t) = \sqrt{\rho_1}\boldsymbol{H}_{11}\boldsymbol{s}_1(t) + \sqrt{\eta_1}\boldsymbol{H}_{12}\boldsymbol{s}_2(t) + \boldsymbol{n}_1(t) \quad (1)$$

$$\boldsymbol{u}_2(t) = \sqrt{\rho_2}\boldsymbol{H}_{22}\boldsymbol{s}_2(t) + \sqrt{\eta_2}\boldsymbol{H}_{21}\boldsymbol{s}_1(t) + \boldsymbol{n}_2(t). \quad (2)$$

In (1)–(2), $\boldsymbol{n}_i(t) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_r})$ denotes AWGN, $\rho_i > 0$ denotes the signal-to-noise ratio (SNR), and $\eta_i > 0$ denotes the interference-to-noise ratio (INR), all at receiver i. The size of η_i will depend on, e.g., antenna separation and analog-domain suppression.

B. Transmission Protocol

We assume that the signaling epoch $\mathcal{T}$ is partitioned into a training period $\mathcal{T}_{\text{train}}$ and a subsequent data communication period $\mathcal{T}_{\text{data}}$. The training period is further partitioned into two equal-length portions (i.e., $\mathcal{T}_{\text{train}}[1]$ and $\mathcal{T}_{\text{train}}[2]$) to avoid self-interference when learning the channel matrices. The data period is also partitioned into two equal-length portions (i.e., $\mathcal{T}_{\text{data}}[1]$ and $\mathcal{T}_{\text{data}}[2]$) over which the transmission parameters can be independently optimized. As we shall see in the sequel, such flexibility is critical when the INR η_i is large relative to the SNR ρ_i. Moreover, this partitioning of the data interval allows us to consider both half- and full-duplex transmission schemes as special cases of a more general transmission protocol. Within each of these four subperiods, we assume that the transmitted signals are zero-mean and wide-sense stationary. Finally, we note that our chosen transmission protocol is but one of many possible protocols,[3] and that the sum-rates derived in the sequel characterize only our chosen protocol.

C. Limited Transmitter Dynamic Range

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 $\boldsymbol{x}_j(t) \in \mathbb{C}^{N_t}$ denotes the jth transmitter's intended time-t transmit signal, and say $\boldsymbol{Q}_j \triangleq \text{Cov}\{\boldsymbol{x}_j(t)\}$ over the relevant time period (e.g., $t \in \mathcal{T}_{\text{data}}[1]$). We then write the time-t noisy radiated signal as

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

where $\boldsymbol{c}_j(t) \in \mathbb{C}^{N_t}$ denotes the transmitter noise. Typically, $\kappa \ll 1$. As shown by measurements of various hardware setups (e.g., [18] and [19]), the independent Gaussian noise model in (3) closely approximates the combined effects of additive power-amp noise, non-linearities in the DAC and power-amp,

---

[3]One might also consider partitioning the training and/or data intervals into two *non*-equal-length subperiods and then optimizing the relative lengths.
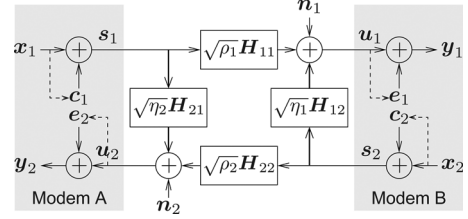

Fig. 2. A model of bidirectional MIMO communication under limited transmitter/receiver-DR. The dashed lines denote statistical dependence. In labeled quantities, the time index t has been suppressed for brevity.

and oscillator phase noise. Moreover, the dependence of the transmitter-noise variance on intended signal power in (3) follows directly from the definition of limited dynamic range.

D. Limited Receiver Dynamic Range

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 $\boldsymbol{u}_i(t) \in \mathbb{C}^{N_r}$ denotes the ith receiver's undistorted time-t received vector, and say $\boldsymbol{\Phi}_i \triangleq \text{Cov}\{\boldsymbol{u}_i(t)\}$ over the relevant time period (e.g., $t \in \mathcal{T}_{\text{data}}[1]$). We then write the distorted post-ADC received signal as

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

where $\boldsymbol{e}_i(t) \in \mathbb{C}^{N_r}$ is additive distortion. Typically, $\beta \ll 1$. From a theoretical perspective, automatic gain control (AGC) followed by dithered uniform quantization [20] yields quantization errors whose statistics closely match the model (4). More importantly, studies (e.g., [21]) have shown that the independent Gaussian distortion model (4) accurately captures the combined effects of additive AGC noise, non-linearities in the ADC and gain-control, and oscillator phase noise in practical hardware.

Fig. 2 summarizes our model. The dashed lines indicate that the distortion levels are proportional to mean energy level and not instantaneous value.

III. ANALYSIS OF ACHIEVABLE SUM-RATE

In this work, we consider systems that use the transmission protocol of Section II-B, pilot-aided channel estimation (as described in Section III-A), and partial self-interference cancellation (as described in Section III-B). For such systems, we derive upper and lower bounds on achievable sum-rate (in Section III-C), and we propose a novel signal design based on optimization of the sum-rate lower bound subject to a power constraint (in Section III-D). To better understand the behavior of the optimized sum-rate, we derive a simple approximation (in Section III-E) that is later shown to be surprisingly accurate.

A. Pilot-Aided Channel Estimation

In this section, we describe the pilot-aided channel estimation procedure that is used to learn the channel matrices $\{\boldsymbol{H}_{ij}\}$. In our protocol, the training interval consists of two subperiods, $\mathcal{T}_{\text{train}}[1]$ and $\mathcal{T}_{\text{train}}[2]$, each of duration TN_t channel uses (for some $T \in \mathbb{Z}^+$). For all $t \in \mathcal{T}_{\text{train}}[1]$, we assume that transmitter $j = 1$ transmits a known pilot signal and $j = 2$ remains silent, while, for all $t \in \mathcal{T}_{\text{train}}[2]$, $j = 2$ transmits and $j = 1$ remains

silent. As we shall see, it suffices to choose the pilot sequence $\boldsymbol{X}_j = [\boldsymbol{x}_j(1), \ldots, \boldsymbol{x}_j(TN_{\mathrm{t}})] \in \mathbb{C}^{N_{\mathrm{t}} \times TN_{\mathrm{t}}}$ arbitrarily so long as it satisfies $\frac{1}{2T}\boldsymbol{X}_j\boldsymbol{X}_j^{\mathsf{H}} = \boldsymbol{I}_{N_{\mathrm{t}}}$, where the scaling is chosen to satisfy a per-period power constraint of the form $\mathrm{tr}(\boldsymbol{Q}_j) = 2$, consistent with the data power constraints that will be described in the sequel.

Our limited transmitter/receiver-DR model implies that the (distorted) space-time pilot signal observed by receiver i is

$$\boldsymbol{Y}_i = \sqrt{\alpha_{ij}}\boldsymbol{H}_{ij}(\boldsymbol{X}_j + \boldsymbol{C}_j) + \boldsymbol{N}_i + \boldsymbol{E}_i \tag{5}$$

where, for notational convenience, we define

$$\alpha_{ij} \triangleq \begin{cases} \rho_i & \text{if } i = j \\ \eta_i & \text{if } i \neq j. \end{cases} \tag{6}$$

In (5), $\boldsymbol{C}_j, \boldsymbol{E}_i$ and $\boldsymbol{N}_i$ are $N_{\mathrm{t}} \times TN_{\mathrm{t}}$ matrices of transmitter noise, receiver distortion, and AWGN, respectively. At the conclusion of training, we assume that the ith receiver estimates the channels $\{\boldsymbol{H}_{ij}\}_{j=1}^2$ via least-squares (LS), yielding

$$\sqrt{\alpha_{ij}}\hat{\boldsymbol{H}}_{ij} \triangleq \frac{1}{2T}\boldsymbol{Y}_i\boldsymbol{X}_j^{\mathsf{H}} \tag{7}$$

and communicates them to the other modem.[4] In the sequel, it will be useful to decompose the channel estimate into the true channel plus some estimation error. In Appendix A, it is shown that such a decomposition takes the form of

$$\sqrt{\alpha_{ij}}\hat{\boldsymbol{H}}_{ij} = \sqrt{\alpha_{ij}}\boldsymbol{H}_{ij} + \boldsymbol{D}_{ij}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{ij} \tag{8}$$

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

$$\boldsymbol{D}_{ij} = \frac{1}{2T}\left[(1+\beta)\boldsymbol{I} + \alpha_{ij}\frac{2\kappa}{N_{\mathrm{t}}}\boldsymbol{H}_{ij}\boldsymbol{H}_{ij}^{\mathsf{H}} \right.$$
$$\left. + \alpha_{ij}\frac{2\beta}{N_{\mathrm{t}}}(1+\kappa)\mathrm{diag}(\boldsymbol{H}_{ij}\boldsymbol{H}_{ij}^{\mathsf{H}})\right] \tag{9}$$

characterizes the spatial covariance of the estimation error. Under $\beta \ll 1$ and $\kappa \ll 1$, the latter reduces to

$$\boldsymbol{D}_{ij} \approx \frac{1}{2T}\left[\boldsymbol{I} + \alpha_{ij}\frac{2\kappa}{N_{\mathrm{t}}}\boldsymbol{H}_{ij}\boldsymbol{H}_{ij}^{\mathsf{H}} + \alpha_{ij}\frac{2\beta}{N_{\mathrm{t}}}\mathrm{diag}(\boldsymbol{H}_{ij}\boldsymbol{H}_{ij}^{\mathsf{H}})\right]. \tag{10}$$

B. Partial Self-Interference Cancellation

As described earlier, the interference in our problem is a form of *self*-interference wherein the receiver $i \in \{1, 2\}$ knows the interfering codewords $\boldsymbol{x}_j$ from transmitter $j \neq i$. If this receiver knew perfectly the interfering channel $\boldsymbol{H}_{ij}$, and if there was no transmitter noise $\boldsymbol{c}_j$ nor receiver distortion $\boldsymbol{e}_i$, then the self-interference could be completely canceled. In general, however, this is not the case, and so the self-interference can only be partially canceled. Details are provided below.

Recall that the data communication period is partitioned into two subperiods, $\mathcal{T}_{\mathrm{data}}[1]$ and $\mathcal{T}_{\mathrm{data}}[2]$, and that—within each—the transmitted signals are wide-sense stationary. The (instantaneous, distorted) signal at receiver $i = 1$ and any time $t \in \mathcal{T}_{\mathrm{data}}[l]$ then takes the form

$$\boldsymbol{y}_1(t) = \sqrt{\rho_1}\boldsymbol{H}_{11}(\boldsymbol{x}_1(t) + \boldsymbol{c}_1(t))$$
$$+ \sqrt{\eta_1}\boldsymbol{H}_{12}(\boldsymbol{x}_2(t) + \boldsymbol{c}_2(t)) + \boldsymbol{n}_1(t) + \boldsymbol{e}_1(t) \tag{11}$$

---
[4]For our achievable-rate analysis, we assume a very long channel coherence interval, making negligible the relative overhead required for CSI exchange.

$$= (\sqrt{\rho_1}\hat{\boldsymbol{H}}_{11} - \boldsymbol{D}_{11}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{11})(\boldsymbol{x}_1(t) + \boldsymbol{c}_1(t))$$
$$+ (\sqrt{\eta_1}\hat{\boldsymbol{H}}_{12} - \boldsymbol{D}_{12}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{12})(\boldsymbol{x}_2(t) + \boldsymbol{c}_2(t))$$
$$+ \boldsymbol{n}_1(t) + \boldsymbol{e}_1(t). \tag{12}$$

Defining the aggregate noise term

$$\boldsymbol{v}_1(t) \triangleq \sqrt{\rho_1}\hat{\boldsymbol{H}}_{11}\boldsymbol{c}_1(t) - \boldsymbol{D}_{11}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{11}(\boldsymbol{x}_1(t) + \boldsymbol{c}_1(t))$$
$$+ \sqrt{\eta_1}\hat{\boldsymbol{H}}_{12}\boldsymbol{c}_2(t) - \boldsymbol{D}_{12}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{12}(\boldsymbol{x}_2(t) + \boldsymbol{c}_2(t))$$
$$+ \boldsymbol{n}_1(t) + \boldsymbol{e}_1(t), \tag{13}$$

we can write $\boldsymbol{y}_1(t) = \sqrt{\rho_1}\hat{\boldsymbol{H}}_{11}\boldsymbol{x}_1(t) + \sqrt{\eta_1}\hat{\boldsymbol{H}}_{12}\boldsymbol{x}_2(t) + \boldsymbol{v}_1(t)$, where the self-interference term $\sqrt{\eta_1}\hat{\boldsymbol{H}}_{12}\boldsymbol{x}_2(t)$ is known and thus can be canceled. The interference-canceled signal $\boldsymbol{z}_1(t) \triangleq \boldsymbol{y}_1(t) - \sqrt{\eta_1}\hat{\boldsymbol{H}}_{12}\boldsymbol{x}_2(t)$ can then be written as

$$\boldsymbol{z}_1(t) = \sqrt{\rho_1}\hat{\boldsymbol{H}}_{11}\boldsymbol{x}_1(t) + \boldsymbol{v}_1(t). \tag{14}$$

Equation (14) shows that, in effect, the information signal $\boldsymbol{x}_1(t)$ propagates through a known channel $\sqrt{\rho_1}\hat{\boldsymbol{H}}_{11}$ corrupted by an aggregate (possibly non-Gaussian) noise $\boldsymbol{v}_1(t)$.

Denoting the $(\hat{\boldsymbol{H}}_{i1}, \hat{\boldsymbol{H}}_{i2})$-conditional covariance of $\boldsymbol{v}_i(t)$ during $t \in \mathcal{T}_{\mathrm{data}}[l]$ by $\hat{\boldsymbol{\Sigma}}_i[l] \triangleq \mathrm{Cov}\{\boldsymbol{v}_i(t)|\hat{\boldsymbol{H}}_{i1}, \hat{\boldsymbol{H}}_{i2}\}_{t \in \mathcal{T}_{\mathrm{data}}[l]}$, we show in Appendix B that

$$\hat{\boldsymbol{\Sigma}}_1[l] \approx \boldsymbol{I} + \kappa\rho_1\hat{\boldsymbol{H}}_{11}\mathrm{diag}(\boldsymbol{Q}_1[l])\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1[l])$$
$$+ \kappa\eta_1\hat{\boldsymbol{H}}_{12}\mathrm{diag}(\boldsymbol{Q}_2[l])\hat{\boldsymbol{H}}_{12}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2[l])$$
$$+ \beta\rho_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1[l]\hat{\boldsymbol{H}}_{11}^{\mathsf{H}})$$
$$+ \beta\eta_1\mathrm{diag}(\hat{\boldsymbol{H}}_{12}\boldsymbol{Q}_2[l]\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}) \tag{15}$$

where $\hat{\boldsymbol{D}}_{ij} \triangleq \mathrm{E}\{\boldsymbol{D}_{ij}|\hat{\boldsymbol{H}}_{ij}\}$ obeys

$$\hat{\boldsymbol{D}}_{ij} \approx \frac{1}{2T}\left[\boldsymbol{I} + \alpha_{ij}\frac{2\kappa}{N_{\mathrm{t}}}\hat{\boldsymbol{H}}_{ij}\hat{\boldsymbol{H}}_{ij}^{\mathsf{H}} + \alpha_{ij}\frac{2\beta}{N_{\mathrm{t}}}\mathrm{diag}(\hat{\boldsymbol{H}}_{ij}\hat{\boldsymbol{H}}_{ij}^{\mathsf{H}})\right] \tag{16}$$

and where the approximations in (15)–(16) follow from $\kappa \ll 1$ and $\beta \ll 1$. A similar analysis applies to $\hat{\boldsymbol{\Sigma}}_2[l]$. We note, for later use, that the channel estimation error terms $\hat{\boldsymbol{D}}_{ij}$ can be made arbitrarily small through appropriate choice of T.

C. Bounds on Achievable Sum-Rate

(14) succinctly characterizes the equivalent channel model, under limited transmitter/receiver-DR, pilot-aided LS MIMO-channel estimation, and partial self-interference cancellation, for transmitter/receiver pair $i = 1$ during data communication period $l \in \{1, 2\}$. (A similar model can be stated for pair $i = 2$.) Due to the channel estimation error components in (13), the aggregate noise $\boldsymbol{v}_1(t)$ is generally non-Gaussian, which complicates the analysis of the channel (14). It is known, however, that among all distributions on $\boldsymbol{v}_1(t)$ with fixed covariance $\hat{\boldsymbol{\Sigma}}_1[l]$, the Gaussian one is worst from a mutual-information perspective [22]. Thus, we can lower-bound the sum mutual information $I(\boldsymbol{Q})$, written as a function of the transmit covariance matrices $\boldsymbol{Q} \triangleq (\boldsymbol{Q}_1[1], \boldsymbol{Q}_1[2], \boldsymbol{Q}_2[1], \boldsymbol{Q}_2[2])$, as $I(\boldsymbol{Q}) \geq \underline{I}(\boldsymbol{Q})$, where [23]

$$\underline{I}(\boldsymbol{Q}) = \sum_{i=1}^2 \frac{1}{2}\sum_{l=1}^2 \log\det\left(\boldsymbol{I} + \rho_i\hat{\boldsymbol{H}}_{ii}\boldsymbol{Q}_i[l]\hat{\boldsymbol{H}}_{ii}^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}_i^{-1}[l]\right)$$
$$= \frac{1}{2}\sum_{i=1}^2\sum_{l=1}^2 \log\det\left(\rho_i\hat{\boldsymbol{H}}_{ii}\boldsymbol{Q}_i[l]\hat{\boldsymbol{H}}_{ii}^{\mathsf{H}} + \hat{\boldsymbol{\Sigma}}_i[l]\right)$$
$$- \log\det(\hat{\boldsymbol{\Sigma}}_i[l]). \tag{17}$$

Furthermore, standard communication theoretic arguments imply that it is possible to achieve a sum-rate equal to $\underline{I}(\boldsymbol{Q})$ in (17) by using independent Gaussian codebooks at each transmitter and maximum-likelihood detection at each receiver [23], as detailed in Section IV. Taking "log" in (17) to be base-2, the units of sum-rate are bits-per-channel-use (bpcu).

A straightforward upper bound $\bar{I}(\boldsymbol{Q})$ on achievable sum-rate then follows by ignoring the channel estimation error, in which case $\boldsymbol{v}_i(t)$ becomes Gaussian and the covariance matrix $\hat{\boldsymbol{\Sigma}}_i[l]$ takes the form of (15) with $\hat{\boldsymbol{D}}_{ij} = \boldsymbol{0}$. The resulting upper bound $\bar{I}(\boldsymbol{Q})$ then has exactly the form of (17), but with $\hat{\boldsymbol{\Sigma}}_i[l]$ computed using $\hat{\boldsymbol{D}}_{ij} = \boldsymbol{0}$. Moreover, the lower bound $\underline{I}(\boldsymbol{Q})$ converges to the upper bound $\bar{I}(\boldsymbol{Q})$ as the training length $T \to \infty$. As demonstrated in Fig. 5, the true mutual information $I(\boldsymbol{Q})$ is tightly sandwiched between the lower and upper bounds for practical training lengths T.

D. Transmit Covariance Optimization

We would now like to find the transmit covariance matrices $\boldsymbol{Q} \triangleq (\boldsymbol{Q}_1[1], \boldsymbol{Q}_1[2], \boldsymbol{Q}_2[1], \boldsymbol{Q}_2[2])$ that maximize the sum-rate lower bound $\underline{I}(\boldsymbol{Q})$ in (17) subject to the per-user power constraint (18b). This yields the optimization problem

$$\max_{\boldsymbol{Q}_1[1], \boldsymbol{Q}_1[2], \boldsymbol{Q}_2[1], \boldsymbol{Q}_2[2]} \underline{I}(\boldsymbol{Q}_1[1], \boldsymbol{Q}_1[2], \boldsymbol{Q}_2[1], \boldsymbol{Q}_2[2]) \quad (18a)$$

$$\text{s.t.} \ \frac{1}{2} \sum_{l=1}^{2} \text{tr}(\boldsymbol{Q}_i[l]) \leq 1, \ i = 1, 2 \quad (18b)$$

$$\boldsymbol{Q}_i[l] \geq 0, \quad \forall i, l \in \{1, 2\} \quad (18c)$$

where the inequality (18c) constrains each $\boldsymbol{Q}_i[l]$ to be positive semi-definite. We solve[5] this non-convex optimization problem via Gradient Projection (GP), taking inspiration from [24]. Our choice to partition the data period $\mathcal{T}_{\text{data}}$ into two subperiods (recall Section II-B) implies that half-duplex transmission is among the possible solutions to (18), thus implying that the optimizer of (18) can never be outperformed by half-duplex. When the data period is not subpartitioned, this desirable property does not hold, and performance suffers (as shown in Section V).

The GP algorithm [25] is defined as follows. For the generic problem of maximizing a function $f(\boldsymbol{a})$ over $\boldsymbol{a} \in \mathcal{A}$, the GP algorithm starts with an initialization $\boldsymbol{a}^{(0)}$ and iterates the following steps for $k = 0, 1, 2, 3, \ldots$

$$\tilde{\boldsymbol{a}}^{(k)} = \mathcal{P}_{\mathcal{A}}(\boldsymbol{a}^{(k)} + s^{(k)} \nabla f(\boldsymbol{a}^{(k)})) \quad (19)$$

$$\boldsymbol{a}^{(k+1)} = \boldsymbol{a}^{(k)} + \gamma^{(k)}(\tilde{\boldsymbol{a}}^{(k)} - \boldsymbol{a}^{(k)}) \quad (20)$$

where $\mathcal{P}_{\mathcal{A}}(\cdot)$ denotes projection onto the set $\mathcal{A}$ and $\nabla f(\cdot)$ denotes the gradient of $f(\cdot)$. The parameters $\gamma^{(k)} \in (0, 1]$ and $s^{(k)}$ act as stepsizes. In the sequel, we assume $s^{(k)} = 1 \ \forall k$.

In applying GP to the optimization problem (18), we first take gradient steps for $\boldsymbol{Q}_1[1]$ and $\boldsymbol{Q}_1[2]$, and then project onto the constraint set $\mathcal{A}$ defined by (18b)–(18c). Next, we take gradient steps for $\boldsymbol{Q}_2[1]$ and $\boldsymbol{Q}_2[2]$, and then project onto the constraint set. In summary, denoting the gradient by $\boldsymbol{G}_i[l] \triangleq \nabla_{\boldsymbol{Q}_i[l]} \underline{I}(\boldsymbol{Q})$, our GP algorithm iterates, for $i = 1$, the following steps to convergence:

$$\boldsymbol{P}_i^{(k)}[1] = \boldsymbol{Q}_i^{(k)}[1] + \boldsymbol{G}_i^{(k)}[1] \quad (21)$$

[5] In general, (18) is a non-convex optimization problem, and so finding the global maximum can be difficult. Although GP is not guaranteed to find the global maximum, our experience with different initializations suggests that, in our problem, GP is indeed finding the global maximum.

$$\boldsymbol{P}_i^{(k)}[2] = \boldsymbol{Q}_i^{(k)}[2] + \boldsymbol{G}_i^{(k)}[2] \quad (22)$$

$$(\tilde{\boldsymbol{Q}}_i^{(k)}[1], \tilde{\boldsymbol{Q}}_i^{(k)}[2]) = \mathcal{P}_{\mathcal{A}}(\boldsymbol{P}_i^{(k)}[1], \boldsymbol{P}_i^{(k)}[2]) \quad (23)$$

$$\boldsymbol{Q}_i^{(k+1)}[1] = \boldsymbol{Q}_i^{(k)}[1] + \gamma^{(k)}(\tilde{\boldsymbol{Q}}_i^{(k)}[1] - \boldsymbol{Q}_i^{(k)}[1]) \quad (24)$$

$$\boldsymbol{Q}_i^{(k+1)}[2] = \boldsymbol{Q}_i^{(k)}[2] + \gamma^{(k)}(\tilde{\boldsymbol{Q}}_i^{(k)}[2] - \boldsymbol{Q}_i^{(k)}[2]) \quad (25)$$

and then repeats the same for $i = 2$. An outer loop then repeats this pair of inner loops until the maximum change in $\{\boldsymbol{Q}_i^{(k)}[l]\}_{\forall i, l}$ (over k) is below a small positive threshold ϵ.

We now provide additional details on the GP steps. As for the gradient, Appendix C shows that, for $l = 1, 2$,

$$\boldsymbol{G}_1[l] = \frac{\rho_1}{\ln 2} \{\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}[\boldsymbol{R}_1^{-1}[l] + \beta \text{diag}(\boldsymbol{R}_1^{-1}[l] - \hat{\boldsymbol{\Sigma}}_1^{-1}[l])]\hat{\boldsymbol{H}}_{11}$$
$$+ \text{diag}[\kappa \hat{\boldsymbol{H}}_{11}^{\mathsf{H}}(\boldsymbol{R}_1^{-1}[l] - \hat{\boldsymbol{\Sigma}}_1^{-1}[l])\hat{\boldsymbol{H}}_{11}]\}$$
$$+ \frac{1}{\ln 2} \text{sum}[\hat{\boldsymbol{D}}_{11}^{*} \odot (\boldsymbol{R}_1^{-1}[l] - \hat{\boldsymbol{\Sigma}}_1^{-1}[l])]\boldsymbol{I}$$
$$+ \frac{\eta_1}{\ln 2} \{\text{diag}[\kappa \hat{\boldsymbol{H}}_{21}^{\mathsf{H}}(\boldsymbol{R}_2^{-1}[l] - \hat{\boldsymbol{\Sigma}}_2^{-1}[l])\hat{\boldsymbol{H}}_{21}]$$
$$+ \beta \hat{\boldsymbol{H}}_{11}^{\mathsf{H}} \text{diag}(\boldsymbol{R}_2^{-1}[l] - \hat{\boldsymbol{\Sigma}}_2^{-1}[l])\hat{\boldsymbol{H}}_{11}\}$$
$$+ \frac{1}{\ln 2} \text{sum}[\hat{\boldsymbol{D}}_{21}^{*} \odot (\boldsymbol{R}_2^{-1}[l] - \hat{\boldsymbol{\Sigma}}_2^{-1}[l])]\boldsymbol{I} \quad (26)$$

where $\boldsymbol{R}_i[l] \triangleq \hat{\boldsymbol{\Sigma}}_i[l] + \rho_i \hat{\boldsymbol{H}}_{ii} \boldsymbol{Q}_i[l] \hat{\boldsymbol{H}}_{ii}^{\mathsf{H}}$. A similar expression can be derived for $\boldsymbol{G}_2[l]$.

To compute the projection $\mathcal{P}_{\mathcal{A}}(\boldsymbol{P}_i[1], \boldsymbol{P}_i[2])$, we first notice that, due to the Hermitian property of $\boldsymbol{P}_i[l]$, we can construct an eigenvalue decomposition $\boldsymbol{P}_i[l] = \boldsymbol{U}_i[l] \boldsymbol{\Lambda}_i[l] \boldsymbol{U}_i^{\mathsf{H}}[l]$ with unitary $\boldsymbol{U}_i[l]$ and real-valued $\boldsymbol{\Lambda}_i[l] = \text{Diag}(\lambda_{i,1}[l], \lambda_{i,2}[l], \ldots, \lambda_{i,N_t}[l])$. The projection of $(\boldsymbol{P}_i[1], \boldsymbol{P}_i[2])$ onto the constraints (18b)–(18c) then equals $\tilde{\boldsymbol{Q}}_i[l] = \boldsymbol{U}_i[l](\boldsymbol{\Lambda}_i[l] - \mu \boldsymbol{I})^+ \boldsymbol{U}_i^{\mathsf{H}}[l]$, where $(\boldsymbol{B})^+ = \max(\boldsymbol{B}, \boldsymbol{0})$ elementwise, and where μ is chosen such that $\frac{1}{2} \sum_{n=1}^{N_t} \sum_{l=1}^{2} \max(\lambda_{i,n}[l] - \mu, 0) = 1$. In essence, $\mathcal{P}_{\mathcal{A}}(\cdot)$ performs water-filling.

To adjust the stepsize $\gamma^{(k)}$, we use the Armijo stepsize rule [25], i.e., $\gamma^{(k)} = \nu^{m_k}$ where m_k is the smallest nonnegative integer that satisfies

$$\underline{I}(\boldsymbol{Q}^{(k+1)}) - \underline{I}(\boldsymbol{Q}^{(k)})$$
$$\geq \sigma \nu^{m_k} \sum_{i=1}^{2} \sum_{l=1}^{2} \text{tr}\left(\boldsymbol{G}_i^{(k)\mathsf{H}}[l](\tilde{\boldsymbol{Q}}_i^{(k)}[l] - \boldsymbol{Q}_i^{(k)}[l])\right) \quad (27)$$

for some constants σ, ν typically chosen so that $\sigma \in [10^{-5}, 10^{-1}]$ and $\nu \in [0.1, 0.5]$. Above, we used the shorthand $\boldsymbol{Q}^{(k)} \triangleq (\boldsymbol{Q}_1^{(k)}[1], \boldsymbol{Q}_1^{(k)}[2], \boldsymbol{Q}_2^{(k)}[1], \boldsymbol{Q}_2^{(k)}[2])$.

E. Sum-Rate Approximation

The complicated nature of the optimization problem (18) motivates us to approximate its solution, i.e., the transmit-covariance optimized sum-rate $I_* \triangleq \max_{\boldsymbol{Q} \in \mathbb{Q}} \underline{I}(\boldsymbol{Q})$, where $\mathbb{Q}$ represents the constraint set implied by (18b)–(18c). Here, we focus on the case of $T \to \infty$, where channel estimation error is driven to zero so that $\underline{I}(\boldsymbol{Q}) = I(\boldsymbol{Q}) = \bar{I}(\boldsymbol{Q})$, and we assume $\rho \triangleq \rho_1 = \rho_2$ and $\eta \triangleq \eta_1 = \eta_2$ for tractability.

Our approximation is built around the simplifying case that each $\boldsymbol{H}_{ij}$ is diagonal, although not necessarily square, with $N_{\min} \triangleq \min\{N_t, N_r\}$ identical diagonal entries equal to $\sqrt{\frac{N_t N_r}{N_{\min}}}$. (The latter value is chosen so that
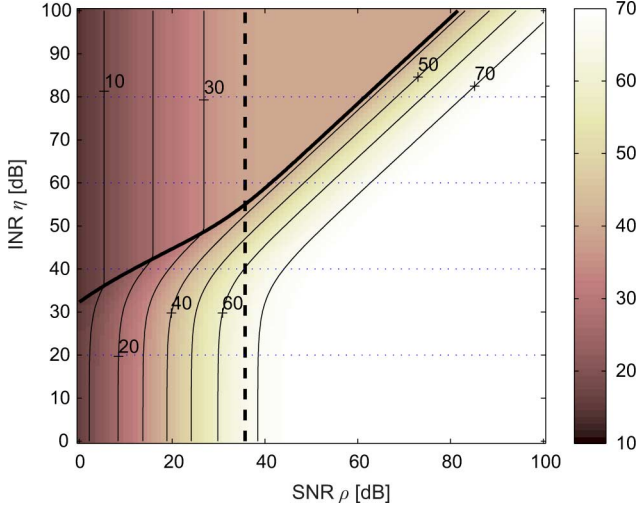
Fig. 3. Contour plot of the optimized-sum-rate approximation I_* versus SNR ρ and INR η, for $N_t = 3$, $N_r = 4$, and $\beta = \kappa = -40$ dB. The dark curve shows the boundary $\eta = \dfrac{\left(\sqrt{\xi^2 + \frac{2\rho\xi}{(\kappa+\beta)}} - (\xi - 2\rho) \right)}{2}$ between full- and half-duplex regimes, and the vertical dashed line shows the boundary $\rho = \frac{N_{\min}}{N_r(\kappa+\beta)}$ between SNR-limited and distortion-limited regimes.

$\mathrm{E}\{\mathrm{tr}(\boldsymbol{H}_{ij}\boldsymbol{H}_{ij}^\mathsf{H})\} = N_t N_r$, consistent with the assumptions of Section II-A.) In this case, the mutual information expression (17) becomes (for $j \neq i$)

$$I(\boldsymbol{Q}) \approx \frac{1}{2} \sum_{i,l} \log \det \left(\boldsymbol{I} + \rho \frac{N_t N_r}{N_{\min}} \boldsymbol{Q}_i[l] \left(\boldsymbol{I} + (\kappa + \beta) \frac{N_t N_r}{N_{\min}} \right. \right.$$
$$\left. \left. \times \left[\rho \, \mathrm{diag}(\boldsymbol{Q}_i[l]) + \eta \, \mathrm{diag}(\boldsymbol{Q}_j[l]) \right] \right)^{-1} \right). \quad (28)$$

When $\eta \ll \rho$, the η-dependent term in (28) can be ignored, after which it is straightforward to show that, under the constraints (18b)–(18c), the optimal covariances are the "full duplex" $\boldsymbol{Q}_{\mathrm{FD}} \triangleq (\frac{1}{N_t}\boldsymbol{I}, \frac{1}{N_t}\boldsymbol{I}, \frac{1}{N_t}\boldsymbol{I}, \frac{1}{N_t}\boldsymbol{I})$, for which (28) gives

$$I(\boldsymbol{Q}_{\mathrm{FD}}) \approx 2 N_{\min} \log \left(1 + \frac{\rho}{\frac{N_{\min}}{N_r} + (\kappa + \beta)(\rho + \eta)} \right). \quad (29)$$

When $\eta \gg \rho$, the η-dependent term in (28) dominates unless $\boldsymbol{Q}_j[l] = \boldsymbol{0}$. In this case, the optimal covariances are the "half duplex" ones $\boldsymbol{Q}_{\mathrm{HD}} \triangleq (\frac{2}{N_t}\boldsymbol{I}, \boldsymbol{0}, \boldsymbol{0}, \frac{2}{N_t}\boldsymbol{I})$, for which (28) gives

$$I(\boldsymbol{Q}_{\mathrm{HD}}) \approx N_{\min} \log \left(1 + \frac{\rho}{\frac{N_{\min}}{2N_r} + (\kappa + \beta)\rho} \right). \quad (30)$$

Finally, for any given pair (η, ρ), we approximate the optimized sum-rate as follows: $I_* \approx \max\{I(\boldsymbol{Q}_{\mathrm{FD}}), I(\boldsymbol{Q}_{\mathrm{HD}})\}$. From (29)–(30), it is straightforward to show that the boundary between full- and half-duplex occurs at

$$\eta = \frac{1}{2} \left(\sqrt{\xi^2 + \frac{2\rho\xi}{(\kappa+\beta)}} - (\xi - 2\rho) \right) \quad (31)$$

for $\xi \triangleq \frac{N_{\min}}{N_r(\kappa+\beta)} + 2\rho$.

We now make some additional observations about (29)–(30). First, suppose that $\eta \ll \rho$, in which case $\boldsymbol{Q}_{\mathrm{FD}}$ is appropriate. From (29), we see that $I(\boldsymbol{Q}_{\mathrm{FD}})$ will not significantly benefit from further increase in SNR ρ when $(\kappa + \beta)\rho > (\kappa + \beta)\eta + \frac{N_{\min}}{N_r}$,

i.e., when $\rho > \eta + \frac{N_{\min}}{N_r(\kappa+\beta)}$. Since $\eta \ll \rho$, this ρ-saturation occurs when $\rho > \frac{N_{\min}}{N_r(\kappa+\beta)}$. Next, suppose that $\eta \gg \rho$, in which case $\boldsymbol{Q}_{\mathrm{HD}}$ is appropriate. Here, (30) shows that $I(\boldsymbol{Q}_{\mathrm{HD}})$ will not significantly benefit from SNRs above $\rho = \frac{N_{\min}}{2N_r(\kappa+\beta)}$. Thus, in both the $\eta \ll \rho$ and $\eta \gg \rho$ cases, we can interpret $\rho \approx \frac{N_{\min}}{N_r(\kappa+\beta)}$ as the transition between SNR-limited and distortion-limited regimes. (See Fig. 3.)

Fig. 3 shows a contour plot of the proposed optimized-sum-rate approximation as a function of INR η and SNR ρ. We shall see in Section V that our approximation of the covariance-optimized sum-rate is surprisingly close, on average, to that found by solving (18) using gradient projection.

IV. TRANSMISSION AND RECEPTION SCHEMES

We now propose explicit transmission and reception schemes designed to maximize sum-rate. To start, we first recall the post-interference-cancellation model given by (14), where the aggregate noise $\{\boldsymbol{v}_i(t)\}_{t \in \mathcal{T}_{\mathrm{data}}[l]}$ has covariance $\hat{\boldsymbol{\Sigma}}_i[l]$ and the signal $\{\boldsymbol{x}_i(t)\}_{t \in \mathcal{T}_{\mathrm{data}}[l]}$ has covariance $\boldsymbol{Q}_i[l]$. Using a Cholesky decomposition of the aggregate-noise covariance matrix $\hat{\boldsymbol{\Sigma}}_i[l] = \boldsymbol{L}_i[l]\boldsymbol{L}_i^\mathsf{H}[l]$, and an eigen-decomposition of the transmit covariance matrix $\boldsymbol{Q}_i[l] = \boldsymbol{V}_i[l]\boldsymbol{S}_i[l]\boldsymbol{V}_i^\mathsf{H}[l]$, we can write the system model (14), for $t \in \mathcal{T}_{\mathrm{data}}[l]$, as

$$\underbrace{\boldsymbol{L}_i^{-1}[l]\boldsymbol{z}_i(t)}_{\triangleq \, \boldsymbol{r}_i(t)} = \underbrace{\sqrt{\rho_i}\boldsymbol{L}_i^{-1}[l]\hat{\boldsymbol{H}}_{ii}\boldsymbol{V}_i[l]\boldsymbol{S}_i^{\frac{1}{2}}[l]}_{\triangleq \, \boldsymbol{B}_i[l]} \boldsymbol{d}_i(t) + \underbrace{\boldsymbol{L}_i^{-1}[l]\boldsymbol{v}_i(t)}_{\triangleq \, \boldsymbol{w}_i(t)}$$
$$(32)$$

where the source $\{\boldsymbol{d}_i(t)\}_{t \in \mathcal{T}_{\mathrm{data}}[l]}$ and noise $\{\boldsymbol{w}_i(t)\}_{t \in \mathcal{T}_{\mathrm{data}}[l]}$ processes are both white (i.e., $\forall i, t$, $\mathrm{Cov}\{\boldsymbol{w}_i(t)\} = \boldsymbol{I}$ and $\mathrm{Cov}\{\boldsymbol{d}_i(t)\} = \boldsymbol{I}$). Denoting $N_i[l] \triangleq \mathrm{rank}(\boldsymbol{Q}_i[l])$, we have assumed (without loss of generality) that $\boldsymbol{S}_i[l] \in \mathbb{R}^{N_i[l] \times N_i[l]}$ is positive definite, and thus $\boldsymbol{B}_i[l] \in \mathbb{C}^{N_r \times N_i[l]}$.

Model (32) can be recognized as the classical "multiuser uplink with single transmit antennas at each user terminal and multiple receive antennas at the base station," as discussed in [23, p. 428-430]. In particular, (32) corresponds to $N_i[l]$ independently coded user terminals and a base station with N_r receive antennas. In this case, the sum-rate maximizing transmission/reception scheme employs independent Gaussian codebooks at each user terminal (or, in our case, for each element of $\boldsymbol{d}_i(t)$ during each subperiod l), and minimum-mean-squared error successive interference cancellation (MMSE-SIC) at the base-station (or, in our case, at the receiver during each subperiod l). Moreover, for each given pair (i, l), the sum-rate across codebooks must not exceed $\log_2 \det(\boldsymbol{I} + \rho_i \hat{\boldsymbol{H}}_{ii}\boldsymbol{Q}_i[l]\hat{\boldsymbol{H}}_{ii}\hat{\boldsymbol{\Sigma}}_i^{-1}[l])$ bpcu. The maximum allowable rate for each particular codebook depends on the interference level experienced during MMSE-SIC, which is dependent on the SIC cancellation order [23].

In summary: at the transmitter of each communicating pair i, the information bit-stream is partitioned into $N_i[1] + N_i[2]$ substreams, each of which is coded using an independent Gaussian codebook of unit variance and appropriate rate. Then, during each subperiod l, the coder output $\boldsymbol{d}_i(t)$ is linearly precoded to obtain the time-t transmit signal $\boldsymbol{x}_i(t) = \boldsymbol{V}_i[l]\boldsymbol{S}_i^{\frac{1}{2}}[l]\boldsymbol{d}_i(t)$ over $t \in \mathcal{T}_{\mathrm{data}}[l]$. At the receiver of each communicating pair i, over times $t \in \mathcal{T}_{\mathrm{data}}[l]$ for each subperiod l, the received signal $\boldsymbol{y}_i(t)$ first undergoes partial self-interference cancellation
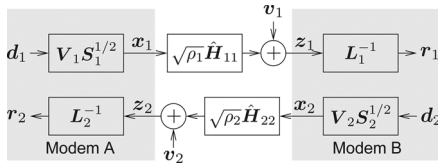
Fig. 4. Illustration of the proposed bidirectional MIMO transmission and reception schemes, based on the interference cancellation model (14) and its whitened version in (32). The time index (t) and subpartition index $[l]$ have been suppressed for brevity.

to obtain $z_i(t)$ (recall (14)), and then noise whitening to obtain $r_i(t) = L_i^{-1}[l]z_i(t)$. MMSE-SIC is then performed on $r_i(t)$ to decode each of the $N_i[l]$ substreams within $d_i(t)$, as described in [23, p. 361]. The overall scheme is illustrated in Fig. 4.

V. NUMERICAL RESULTS

In this section, we numerically investigate the behavior of the sum-rates achievable in full-duplex bidirectional MIMO communication under the proposed limited transmitter/receiver-DR and channel-estimation-error models. In particular, we study the average behavior of the GP-optimized sum-rate lower-bound $\max_Q \underline{I}(Q)$ as a function of SNR ρ_i, INR η_i, dynamic range κ^{-1} and β^{-1}, number of antennas N_t and N_r, and training length T. In doing so, our goal is to provide the system designer with an understanding of how sum-rate varies with each of these parameters. We also show the gains of full-duplex signaling over optimized half-duplex (OHD) signaling, of partial self-interference cancellation, and of the partitioning of the data period $\mathcal{T}_{data}$ into two distinct subperiods, $\mathcal{T}_{data}[1]$ and $\mathcal{T}_{data}[2]$. Finally, we show a close agreement between the GP-optimized sum-rate $\max_Q \underline{I}(Q)$ and the approximation proposed in Section III-E.

For the numerical results below, the propagation channel model from Section II-A and the limited transmitter/receiver-DR models from Section II-C and Section II-D were employed, pilot-aided channel estimation was implemented as in Section III-A, and the power constraint (18b) was applied, implying the channel-estimation-error covariance (9) and the aggregate-noise covariance (15). Throughout, we used training duration $T = 50$ (as justified below), Armijo parameters $\sigma = 0.01$ and $\nu = 0.2$, and GP stopping threshold $\epsilon = 0.01$. For brevity, we focused only on the case $\rho_1 = \rho_2 \triangleq \rho$ and $\eta_1 = \eta_2 \triangleq \eta$. All results were averaged over 1000 realizations, unless specified otherwise.

Below, we denote the full scheme proposed in Section III by "TCO-2-IC," which indicates the use of interference cancellation (IC) and transmit covariance optimization (TCO) performed individually over the 2 data subperiods (i.e., $\mathcal{T}_{data}[1]$ and $\mathcal{T}_{data}[2]$). To test the impact of IC and of two data subperiods, 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 subperiod (i.e., $Q_i[1] = Q_i[2]\ \forall i$), which we refer to as "TCO-1-IC." To optimize half-duplex, we used GP to maximize $\underline{I}(Q)$ under the power constraint (18b) and the additional half-duplex constraint $Q_1[2] = 0 = Q_2[1]$.

In Fig. 5, we investigate the role of channel-estimation training length T on the achievable sum-rate lower bound $\underline{I}(Q)$ of TCO-2-IC. There we see that the sum-rate increases rapidly in T for small T, but quickly saturates for larger values of T. This behavior can be understood from (15)–(16), which suggests that channel estimation error will have a negligible
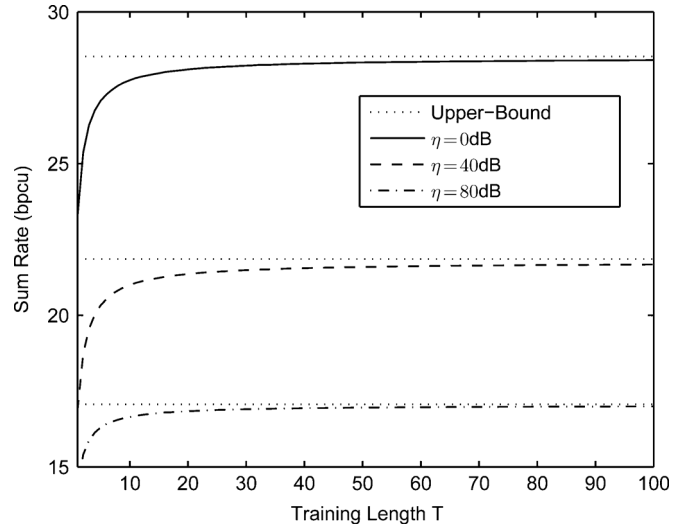

Fig. 5. Achievable sum-rate lower bound $\underline{I}(Q)$ for TCO-2-IC versus training interval T. Here, $N_t = 3$, $N_r = 4$, $\beta = \kappa = -40$ dB, $\rho = 15$ dB. Also shown is the corresponding upper bound $\bar{I}(Q)$ for each case.
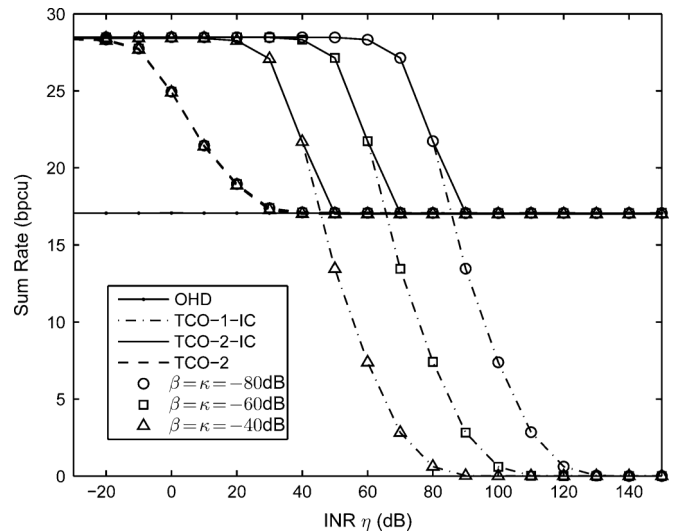

Fig. 6. Achievable sum-rate lower bound $\underline{I}(Q)$ for TCO-2-IC, TCO-2, TCO-1-IC, and OHD versus INR η. Here, $N_t = 3$, $N_r = 4$, $\rho = 15$ dB, and $T = 50$. OHD is plotted for $\beta = \kappa = -60$ dB, but was observed to give nearly identical results for all three values of $\beta = \kappa$.

effect on the noise covariance $\hat{\Sigma}_1[l]$ when $TN_t \gg 1$. Fig. 5 also shows the corresponding achievable sum-rate upper bound $\bar{I}(Q)$. These traces confirm that the nominal training length $T = 50$ ensures that $\underline{I}(Q) \approx \bar{I}(Q) \approx I(Q)$.

In Fig. 6, we examine sum-rate performance versus INR η for the TCO-2-IC, TCO-1-IC, TCO-2, and OHD schemes, using several different dynamic range parameters $\beta = \kappa$. For OHD, we see that sum-rate is invariant to INR η, as expected. For the proposed TCO-2-IC, we observe "full duplex" performance for low-to-mid values of η and a transition to OHD performance at high values of η, just as predicted by the approximation in Section III-E. In fact, the sum-rates in Fig. 6 are nearly identical to the approximate values in Fig. 3. To see the importance of two distinct data-communication periods, we study the TCO-1-IC trace, where we observe TCO-2-IC-like performance at low-to-mid values of η, but performance that drops
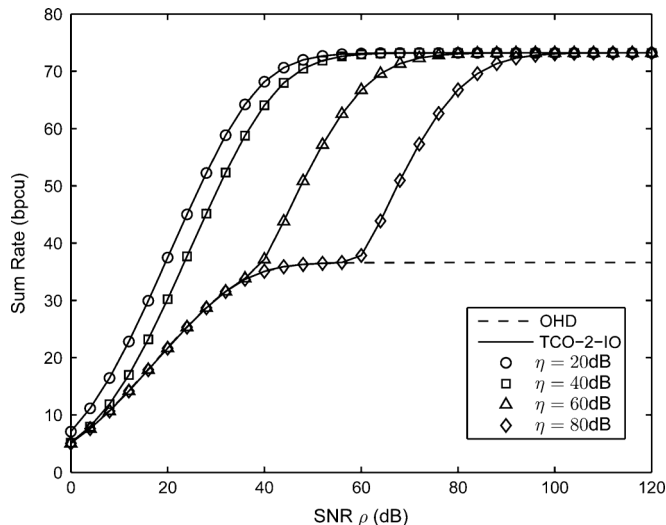
Fig. 7. Achievable sum-rate lower bound $\underline{I}(\boldsymbol{Q})$ for TCO-2-IC and OHD versus SNR ρ. Here, $N_t = 3$, $N_r = 4$, $\beta = \kappa = -40$ dB, and $T = 50$.

Fig. 8. Contour plot of the achievable sum-rate lower bound $\underline{I}(\boldsymbol{Q})$ for TCO-2-IC versus SNR ρ and INR η, for $N_t = 3$, $N_r = 4$, and $\beta = \kappa = -40$ dB, averaged over 200 realizations. The dark curve (i.e., full/half-duplex boundary) and dashed line (i.e., SNR/distortion-limited boundary) are identical to the ones in Fig. 3.

below OHD at high η. Essentially, TCO-1-IC forces full-duplex signaling at high INR η, 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 subperiods, similar to the interference-channel scheme [26]. Finally, from the TCO-2 trace, we conclude that partial interference cancellation is essential for all but extreme values of INR η.

In Fig. 7, we examine sum-rate of the proposed TCO-IC-2 and OHD versus SNR ρ, using the dynamic range parameters $\beta = \kappa = -40$ dB and various fixed values of INR η. All the behaviors in Fig. 7 are almost exactly as predicted by the sum-rate approximation described in Section III-E and illustrated in Fig. 3. In particular, we see OHD's sum-rate increase with SNR ρ up to the distortion-limited regime, i.e., $\rho \gtrsim \frac{N_{\min}}{N_r(\kappa+\beta)} \approx 36$ dB. For TCO-IC-2, we see sum-rate increase with ρ when $\rho \in [0, 36]$ (i.e., the SNR-limited regime), saturate when $\rho \in [36, \eta]$ (i.e., distortion-limited high-INR regime), increase again around $\rho \approx \eta$ (i.e., the transition to the low-INR regime), and then saturate when $\rho \gg \eta$ (i.e., the distortion-limited low-INR regime). In fact, the sum-rates in Fig. 7 are nearly identical to the approximations in Fig. 3.

In Fig. 8, we plot the GP-optimized sum-rate contours of the proposed TCO-IC-2 versus both SNR ρ and INR η, for comparison to the approximation in Fig. 3. Remarkably, the two plots look almost identical, confirming the accuracy of the approximation over a wide range of INRs η and SNRs ρ.

Finally, in Fig. 9, we explore the sum-rate of TCO-2-IC and OHD versus the number of antennas, N_t and N_r, for fixed values of SNR $\rho = 15$ dB, INR $\eta = 60$ dB, and transmitter/receiver-DR $\beta = \kappa = -60$ dB. We recall, from Fig. 6, that these parameters correspond to the interesting region between half-duplex and full-duplex. In Fig. 9, we see that sum-rate increases with both N_r and N_t, as expected. More interesting is the sum-rate behavior when the total number of antennas is fixed, e.g., at $N_t + N_r = 7$, as illustrated by the triangles in Fig. 9. The fact that the configuration $(N_t, N_r) = (3, 4)$ outperforms $(N_t, N_r) = (4, 3)$ is predicted by the approximations (29)–(30): given fixed $N_{\min}$ (here, $N_{\min} = 3$), one should strive to maximize N_r.
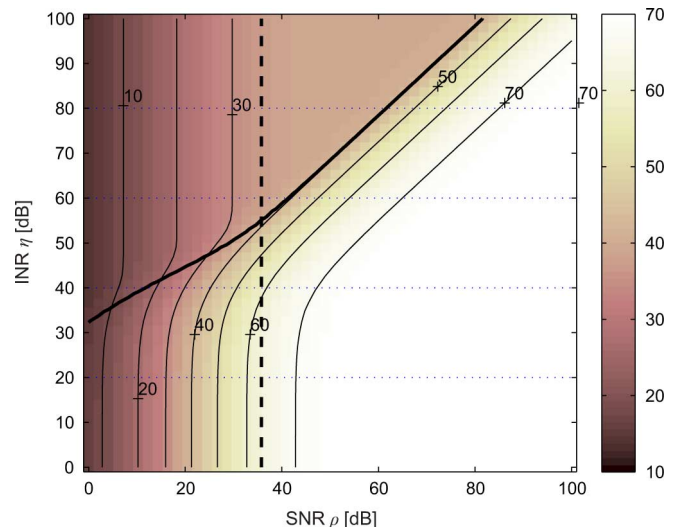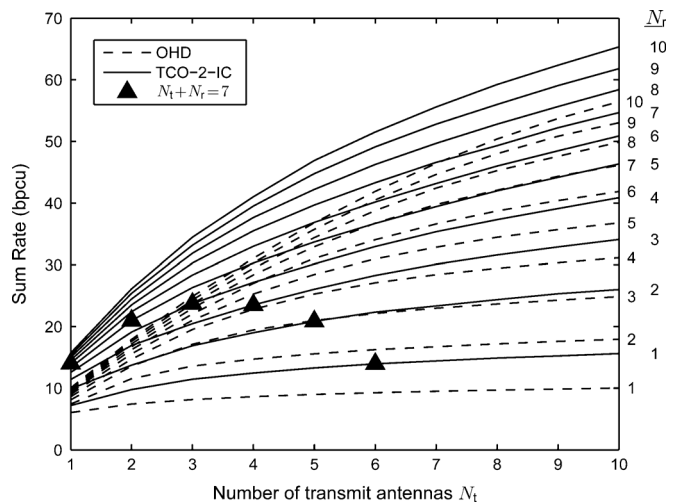

Fig. 9. Achievable sum-rate lower bound $\underline{I}(\boldsymbol{Q})$ for TCO-2-IC and OHD versus number of antennas N_t with various N_r. Here, $\rho = 15$ dB, $\eta = 60$ dB, $\beta = \kappa = -60$ dB, and $T = 50$.

VI. CONCLUSION

We considered the problem of full-duplex bidirectional communication between a pair of MIMO modems in the presence of noise, limited transmitter/receiver-dynamic range, imperfect CSI, and very high levels of self-interference. Using explicit Gaussian models for dynamic-range limitation and pilot-aided channel estimation error, we derived upper and lower bounds on the achievable sum-rate that tighten as the number of pilots increases. Furthermore, we proposed a transmission scheme based on maximizing the sum-rate lower bound, which involves the numerical solution of a nonconvex optimization problem, to which we applied gradient projection. In addition, we derived an analytic approximation to the achievable sum-rate that agrees closely with the results of the numerical optimization and allows one to straightforwardly predict when full-duplex operation can

significantly outperform half-duplex operation. Finally, we numerically studied the behavior of achievable sum-rate as a function of signal-to-noise ratio, interference-to-noise ratio, transmitter/receiver dynamic range, number of antennas, and number of pilots.

In future work, we plan to investigate the effect of practical coding/decoding schemes, and to extend our approach to time-varying channels and to systems with more than two modems.

APPENDIX A
CHANNEL ESTIMATION DETAILS

In this appendix, we derive certain details of Section III-A. While doing so, we suppress the i and j subscripts for brevity. Under limited transmitter-DR, the undistorted received space-time signal is

$$U = \sqrt{\alpha} H (X + C) + N \tag{33}$$

where the spatial correlation[6] of the non-distorted pilot signal X equals $\frac{2}{N_t} I$ and hence the spatial correlation of the transmitter distortion C equals $\frac{2\kappa}{N_t} I$. Conditioned on H, the spatial correlation of U is then

$$\Phi = \frac{2\alpha(1+\kappa)}{N_t} HH^H + I, \tag{34}$$

and hence the H-conditional spatial correlation of the receiver distortion E equals

$$\beta \mathrm{diag}(\Phi) = \beta \left(\frac{2\alpha(1+\kappa)}{N_t} \mathrm{diag}(HH^H) + I \right). \tag{35}$$

Given (5), the distorted received signal Y can be written as

$$Y = \sqrt{\alpha} H X + W \tag{36}$$

where

$$W \triangleq \sqrt{\alpha} H C + N + E \tag{37}$$

is aggregate complex Gaussian noise that is temporally white with H-conditional spatial correlation

$$\frac{2\alpha\kappa}{N_t} HH^H + I + \beta \left(\frac{2\alpha(1+\kappa)}{N_t} \mathrm{diag}(HH^H) + I \right). \tag{38}$$

Due to the fact that $\frac{1}{2T} X X^H = I$, the channel estimate (7) takes the form

$$\sqrt{\alpha} \hat{H} = \frac{1}{2T} Y X^H = \sqrt{\alpha} H + \frac{1}{2T} W X^H \tag{39}$$

where $\frac{1}{2T} W X^H$ is Gaussian channel estimation error. We now analyze the H-conditional correlations among the elements of the channel estimation error matrix. We begin by noticing

$$\mathrm{E}\left\{ \left[\frac{1}{2T} W X^H \right]_{m,p} \left[\frac{1}{2T} W X^H \right]_{n,q}^* \Big| H \right\}$$

$$= \left(\frac{1}{2T} \right)^2 \mathrm{E}\left\{ \sum_k [W]_{m,k} [X]_{p,k}^* \sum_l [X]_{q,l} [W]_{n,l}^* \Big| H \right\} \tag{40}$$

$$= \left(\frac{1}{2T} \right)^2 \sum_{k,l} [X]_{p,k}^* [X]_{q,l} \mathrm{E}\left\{ [W]_{m,k} [W]_{n,l}^* \Big| H \right\}. \tag{41}$$

[6]The spatial correlation of $X = [x(1), \ldots, x(TN_t)]$ is defined as $\mathrm{E}\{x(t)x(t)^H\} = \mathrm{E}\{\frac{1}{TN_t} \sum_{t=1}^{TN_t} x(t)x(t)^H\} = \mathrm{E}\{\frac{1}{TN_t} XX^H\}$.

To find $\mathrm{E}\{ [W]_{m,k} [W]_{n,l}^* | H \}$, we recall that

$$\mathrm{E}\{ [N]_{m,k} [N]_{n,l}^* | H \} = \delta_{m-n} \delta_{k-l} \tag{42}$$

$$\mathrm{E}\{ [C]_{q,k} [C]_{p,l}^* | H \} = \frac{2\kappa}{N_t} \delta_{q-p} \delta_{k-l} \tag{43}$$

$$\mathrm{E}\{ [E]_{m,k} [E]_{n,l}^* | H \} = \beta [\Phi]_{m,m} \delta_{m-n} \delta_{k-l} \tag{44}$$

implying that

$$\mathrm{E}\{ [W]_{m,k} [W]_{n,l}^* | H \}$$

$$= \alpha \sum_{q,p} [H]_{m,q} [H]_{n,p}^* \mathrm{E}\{ [C]_{q,k} [C]_{p,l}^* | H \} \tag{45}$$

$$+ \mathrm{E}\{ [N]_{m,k} [N]_{n,l}^* | H \} + \mathrm{E}\{ [E]_{m,k} [E]_{n,l}^* | H \}$$

$$= \delta_{k-l} \left(\alpha \frac{2\kappa}{N_t} \sum_p [H]_{m,p} [H]_{n,p}^* + \delta_{m-n} + \beta [\Phi]_{m,m} \delta_{m-n} \right) \tag{46}$$

which implies that

$$\mathrm{E}\left\{ \left[\frac{1}{2T} W X^H \right]_{m,p} \left[\frac{1}{2T} W X^H \right]_{n,q}^* \Big| H \right\}$$

$$= \left(\frac{1}{2T} \right)^2 \sum_k [X]_{p,k}^* [X]_{q,k}$$

$$\times \left(\alpha \frac{2\kappa}{N_t} \sum_p [H]_{m,p} [H]_{n,p}^* + \delta_{m-n} + \beta [\Phi]_{m,m} \delta_{m-n} \right) \tag{47}$$

$$= \frac{1}{2T} \delta_{p-q} \left(\alpha \frac{2\kappa}{N_t} \sum_p [H]_{m,p} [H]_{n,p}^* + \delta_{m-n} + \beta [\Phi]_{m,m} \delta_{m-n} \right) \tag{48}$$

where the latter expression follows from the fact that $\sum_k [X]_{p,k}^* [X]_{q,k} = 2T \delta_{p-q}$, as implied by $\frac{1}{2T} X X^H = I$. Equation (48) implies that the channel estimation error is temporally white with H-conditional spatial correlation

$$D \triangleq \frac{1}{2T} \left(\alpha \frac{2\kappa}{N_t} HH^H + I + \beta \mathrm{diag}(\Phi) \right) \tag{49}$$

$$= \frac{1}{2T} \left(\alpha \frac{2\kappa}{N_t} HH^H + I + \beta \left(\alpha \frac{2(1+\kappa)}{N_t} \mathrm{diag}(HH^H) + I \right) \right). \tag{50}$$

Our final claim is that the channel estimation error $\frac{1}{2T} W X^H$ is statistically equivalent to $D^{\frac{1}{2}} \tilde{H}$, with $\tilde{H} \in \mathbb{C}^{N_r \times N_t}$ constructed from i.i.d $\mathcal{CN}(0,1)$ entries. This can be seen from the following:

$$\mathrm{E}\left\{ [D^{\frac{1}{2}} \tilde{H}]_{m,p} [D^{\frac{1}{2}} \tilde{H}]_{n,q}^* \right\}$$

$$= \mathrm{E}\left\{ \sum_k [D^{\frac{1}{2}}]_{m,k} [\tilde{H}]_{k,p} \sum_l [D^{\frac{1}{2}}]_{n,l}^* [\tilde{H}]_{l,q}^* \right\} \tag{51}$$

$$= \sum_{k,l} [D^{\frac{1}{2}}]_{m,k} [D^{\frac{1}{2}}]_{n,l}^* \mathrm{E}\left\{ [\tilde{H}]_{k,p} [\tilde{H}]_{l,q}^* \right\} \tag{52}$$

$$= \delta_{p-q} \sum_k [D^{\frac{1}{2}}]_{m,k} [D^{\frac{1}{2}}]_{n,k}^* = \delta_{p-q} [D]_{m,n} \tag{53}$$

where we used the fact that $\mathrm{E}\{ [\tilde{H}]_{k,p} [\tilde{H}]_{l,q}^* \} = \delta_{k-l} \delta_{p-q}$.

APPENDIX B
INTERFERENCE CANCELLATION DETAILS

In this appendix, we characterize the channel-estimate-conditioned covariance of the aggregate interference $\boldsymbol{v}_1$, whose expression was given in (13). The i and j subscripts and subperiod index $[l]$ are suppressed when possible, for brevity.

Recalling that $\hat{\boldsymbol{D}} \triangleq \mathrm{E}\{\boldsymbol{D}|\hat{\boldsymbol{H}}\}$, we first establish that

$$\mathrm{Cov}\{\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}}\boldsymbol{x}|\hat{\boldsymbol{H}}\} = \hat{\boldsymbol{D}}\mathrm{tr}(\mathrm{Cov}(\boldsymbol{x})) \qquad (54)$$

which will be useful in the sequel. To show (54), we examine the $(m,n)^{\mathrm{th}}$ element of the covariance matrix:

$$\left[\mathrm{Cov}\left\{\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}}\boldsymbol{x}\Big|\hat{\boldsymbol{H}}\right\}\right]_{m,n}$$

$$= \mathrm{E}\left\{\left[\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}}\boldsymbol{x}\right]_m [\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}}\boldsymbol{x}]_n^* \Big|\hat{\boldsymbol{H}}\right\} \qquad (55)$$

$$= \mathrm{E}\left\{\sum_{p,r}\left[\boldsymbol{D}^{\frac{1}{2}}\right]_{m,p}\left[\tilde{\boldsymbol{H}}\right]_{p,r}[\boldsymbol{x}]_r \sum_{q,t}\left[\boldsymbol{D}^{\frac{1}{2}}\right]_{n,q}^*\left[\tilde{\boldsymbol{H}}\right]_{q,t}^*[\boldsymbol{x}]_t^* \Big|\hat{\boldsymbol{H}}\right\} \qquad (56)$$

$$= \sum_{p,r,q,t}\left\{\mathrm{E}\left\{\left[\boldsymbol{D}^{\frac{1}{2}}\right]_{m,p}\left[\boldsymbol{D}^{\frac{1}{2}}\right]_{n,q}^*|\hat{\boldsymbol{H}}\right\}\right.$$

$$\left.\times \underbrace{\mathrm{E}\{[\tilde{\boldsymbol{H}}]_{p,r}[\tilde{\boldsymbol{H}}]_{q,t}^*\}}_{\delta_{p-q}\delta_{r-t}}\mathrm{E}\{[\boldsymbol{x}]_r[\boldsymbol{x}]_t^*\}\right\} \qquad (57)$$

$$= \mathrm{E}\left\{\sum_p[\boldsymbol{D}^{\frac{1}{2}}]_{m,p}[\boldsymbol{D}^{\frac{1}{2}}]_{n,p}^*\Big|\hat{\boldsymbol{H}}\right\}\mathrm{E}\left\{\sum_r[\boldsymbol{x}]_r[\boldsymbol{x}]_r^*\right\} \qquad (58)$$

$$= [\hat{\boldsymbol{D}}]_{m,n}\mathrm{tr}(\mathrm{Cov}\{\boldsymbol{x}\}). \qquad (59)$$

Rewriting the previous equality in matrix form, we get (54). As a corollary, we note that

$$\mathrm{E}\{(\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}})\mathrm{Cov}\{\boldsymbol{x}\}(\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{H}})^{\mathsf{H}}|\hat{\boldsymbol{H}}\} = \hat{\boldsymbol{D}}\mathrm{tr}(\mathrm{Cov}\{\boldsymbol{x}\}) \qquad (60)$$

which will also be useful in the sequel.

Next we characterize the $(\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12})$-conditional covariance of the receiver distortion $\boldsymbol{e}_1$. Recalling that $\mathrm{Cov}\{\boldsymbol{e}_1\} = \beta\mathrm{diag}(\boldsymbol{\Phi}_1)$ where $\boldsymbol{\Phi}_1 = \mathrm{Cov}\{\boldsymbol{u}_1\}$, we have $\mathrm{Cov}\{\boldsymbol{e}_1|\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12}\} = \beta\mathrm{diag}(\hat{\boldsymbol{\Phi}}_1)$ where $\hat{\boldsymbol{\Phi}}_1 \triangleq \mathrm{Cov}\{\boldsymbol{u}_1|\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12}\}$. Then, given that $\boldsymbol{u}_1 = \boldsymbol{y}_1 - \boldsymbol{e}_1$ with $\boldsymbol{y}_1$ from (12), and using the fact that $\mathrm{Cov}(\boldsymbol{x}_i + \boldsymbol{c}_i) = \boldsymbol{Q}_i + \kappa\mathrm{diag}(\boldsymbol{Q}_i)$, we get

$$\hat{\boldsymbol{\Phi}}_1 = \rho_1\hat{\boldsymbol{H}}_{11}\big(\boldsymbol{Q}_1 + \kappa\mathrm{diag}(\boldsymbol{Q}_1)\big)\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}$$
$$+ \mathrm{E}\{(\boldsymbol{D}_{11}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{11})(\boldsymbol{Q}_1 + \kappa\mathrm{diag}(\boldsymbol{Q}_1))(\boldsymbol{D}_{11}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{11})^{\mathsf{H}}|\hat{\boldsymbol{H}}_{11}\}$$
$$+ \eta_1\hat{\boldsymbol{H}}_{12}\big(\boldsymbol{Q}_2 + \kappa\mathrm{diag}(\boldsymbol{Q}_2)\big)\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}$$
$$+ \mathrm{E}\{(\boldsymbol{D}_{12}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{12})(\boldsymbol{Q}_2 + \kappa\mathrm{diag}(\boldsymbol{Q}_2))(\boldsymbol{D}_{12}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{12})^{\mathsf{H}}|\hat{\boldsymbol{H}}_{12}\} + \boldsymbol{I} \qquad (61)$$

$$= \rho_1\hat{\boldsymbol{H}}_{11}\big(\boldsymbol{Q}_1 + \kappa\mathrm{diag}(\boldsymbol{Q}_1)\big)\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1 + \kappa\mathrm{diag}(\boldsymbol{Q}_1))$$
$$+ \eta_1\hat{\boldsymbol{H}}_{12}\big(\boldsymbol{Q}_2 + \kappa\mathrm{diag}(\boldsymbol{Q}_2)\big)\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}$$
$$+ \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2 + \kappa\mathrm{diag}(\boldsymbol{Q}_2)) + \boldsymbol{I}. \qquad (62)$$

Then,

$$\hat{\boldsymbol{\Phi}}_1 = \rho_1\hat{\boldsymbol{H}}_{11}\big(\boldsymbol{Q}_1 + \kappa\mathrm{diag}(\boldsymbol{Q}_1)\big)\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + (1+\kappa)\hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1)$$
$$+ \eta_1\hat{\boldsymbol{H}}_{12}\big(\boldsymbol{Q}_2 + \kappa\mathrm{diag}(\boldsymbol{Q}_2)\big)\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}$$
$$+ (1+\kappa)\hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2) + \boldsymbol{I} \qquad (63)$$

$$\approx \rho_1\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1)$$
$$+ \eta_1\hat{\boldsymbol{H}}_{12}\boldsymbol{Q}_2\hat{\boldsymbol{H}}_{12}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2) + \boldsymbol{I} \qquad (64)$$

where, for the approximation, we assumed $\kappa \ll 1$. Thus,

$$\mathrm{Cov}\{\boldsymbol{e}_1|\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12}\}$$
$$\approx \beta\big(\rho_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}) + \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1)$$
$$+ \eta_1\mathrm{diag}(\hat{\boldsymbol{H}}_{12}\boldsymbol{Q}_2\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}) + \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2) + \boldsymbol{I}\big). \qquad (65)$$

Finally we are ready to characterize $\hat{\boldsymbol{\Sigma}}_1$, the $(\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12})$-conditional covariance of $\boldsymbol{v}_1$. From (13),

$$\hat{\boldsymbol{\Sigma}}_1 = \kappa\rho_1\mathrm{E}\{\boldsymbol{H}_{11}\mathrm{diag}(\boldsymbol{Q}_1)\boldsymbol{H}_{11}^{\mathsf{H}}|\hat{\boldsymbol{H}}_{11}\} + \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1)$$
$$+ \kappa\eta_1\mathrm{E}\{\boldsymbol{H}_{12}\mathrm{diag}(\boldsymbol{Q}_2)\boldsymbol{H}_{12}^{\mathsf{H}}|\hat{\boldsymbol{H}}_{12}\}$$
$$+ \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2) + \boldsymbol{I} + \mathrm{Cov}\{\boldsymbol{e}_1|\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12}\} \qquad (66)$$

$$= \kappa\rho_1\hat{\boldsymbol{H}}_{11}\mathrm{diag}(\boldsymbol{Q}_1)\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \boldsymbol{I} + \mathrm{Cov}\{\boldsymbol{e}_1|\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12}\}$$
$$+ \kappa\mathrm{E}\{(\boldsymbol{D}_{11}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{11})\mathrm{diag}(\boldsymbol{Q}_1)(\boldsymbol{D}_{11}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{11})^{\mathsf{H}}|\hat{\boldsymbol{H}}_{11}\}$$
$$+ \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1) + \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2) \qquad (67)$$
$$+ \kappa\eta_1\hat{\boldsymbol{H}}_{12}\mathrm{diag}(\boldsymbol{Q}_2)\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}$$
$$+ \kappa\mathrm{E}\{(\boldsymbol{D}_{12}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{12})\mathrm{diag}(\boldsymbol{Q}_2)(\boldsymbol{D}_{12}^{\frac{1}{2}}\tilde{\boldsymbol{H}}_{12})^{\mathsf{H}}|\hat{\boldsymbol{H}}_{12}\}$$

$$= \kappa\rho_1\hat{\boldsymbol{H}}_{11}\mathrm{diag}(\boldsymbol{Q}_1)\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + (1+\kappa)\hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1)$$
$$+ \kappa\eta_1\hat{\boldsymbol{H}}_{12}\mathrm{diag}(\boldsymbol{Q}_2)\hat{\boldsymbol{H}}_{12}^{\mathsf{H}} + (1+\kappa)\hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2)$$
$$+ \boldsymbol{I} + \mathrm{Cov}\{\boldsymbol{e}_1|\hat{\boldsymbol{H}}_{11},\hat{\boldsymbol{H}}_{12}\} \qquad (68)$$

$$\approx \boldsymbol{I} + \kappa\rho_1\hat{\boldsymbol{H}}_{11}\mathrm{diag}(\boldsymbol{Q}_1)\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{11}\mathrm{tr}(\boldsymbol{Q}_1)$$
$$+ \kappa\eta_1\hat{\boldsymbol{H}}_{12}\mathrm{diag}(\boldsymbol{Q}_2)\hat{\boldsymbol{H}}_{12}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{12}\mathrm{tr}(\boldsymbol{Q}_2)$$
$$+ \beta\rho_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}) + \beta\eta_1\mathrm{diag}(\hat{\boldsymbol{H}}_{12}\boldsymbol{Q}_2\hat{\boldsymbol{H}}_{12}^{\mathsf{H}}) \qquad (69)$$

where, for the approximation, we assumed $\kappa \ll 1$ and $\beta \ll 1$, and we leveraged (65).

APPENDIX C
GRADIENT DETAILS

In this appendix, we derive an expression for the gradient $\nabla_{\boldsymbol{Q}_1[l]}\underline{I}(\boldsymbol{Q}_1[1],\boldsymbol{Q}_1[2],\boldsymbol{Q}_2[1],\boldsymbol{Q}[2])$ for $l \in \{1,2\}$ by first deriving an expression for the derivative $\frac{\partial\underline{I}}{\partial\boldsymbol{Q}_i[l]}$ and then using the fact that $\nabla_{\boldsymbol{Q}_i[l]}\underline{I} = 2\left(\frac{\partial\underline{I}}{\partial\boldsymbol{Q}_i[l]}\right)^*$.

To do this, we first consider the related problem of computing the derivative $\frac{\partial\det(\boldsymbol{Y})}{\partial\boldsymbol{X}}$, where

$$\boldsymbol{Y} \triangleq \boldsymbol{C}\mathrm{diag}(\boldsymbol{X})\boldsymbol{D} + \mathrm{diag}(\boldsymbol{E}\boldsymbol{X}\boldsymbol{F}) + \boldsymbol{G}\mathrm{tr}(\boldsymbol{X}) + \boldsymbol{Z} \qquad (70)$$

and where (70) can be written elementwise as

$$[\boldsymbol{Y}]_{i,j} = \sum_{m,n}[\boldsymbol{C}]_{i,m}[\boldsymbol{X}]_{m,n}[\boldsymbol{D}]_{n,j}\delta_{m-n} + [\boldsymbol{Z}]_{i,j}$$
$$+ \sum_{p,q}[\boldsymbol{E}]_{i,p}[\boldsymbol{X}]_{p,q}[\boldsymbol{F}]_{q,j}\delta_{i-j} + [\boldsymbol{G}]_{i,j}\sum_t[\boldsymbol{X}]_{t,t}. \qquad (71)$$

$$\frac{\partial}{\partial \boldsymbol{Q}_1[l]} \underline{I}(\boldsymbol{Q}_1[1], \boldsymbol{Q}_1[2], \boldsymbol{Q}_2[1], \boldsymbol{Q}_2[2])$$

$$= \frac{\partial}{\partial \boldsymbol{Q}_1[l]} \frac{1}{2} \left(\log\det(\boldsymbol{R}_1[l]) - \log\det(\hat{\boldsymbol{\Sigma}}_1[l]) + \log\det(\boldsymbol{R}_2[l]) - \log\det(\hat{\boldsymbol{\Sigma}}_2[l]) \right) \tag{76}$$

$$= \frac{\partial}{\partial \boldsymbol{Q}_1[l]} \frac{1}{2} \Big\{ \log\det\left(\rho_1\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1[l]\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \kappa\rho_1\hat{\boldsymbol{H}}_{11}\mathrm{diag}(\boldsymbol{Q}_1[l])\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \beta\rho_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1[l]\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}) + \hat{\boldsymbol{D}}_{11}\mathrm{tr}\boldsymbol{Q}_1[l] + \boldsymbol{Z}_1[l] \right)$$

$$- \log\det\left(\kappa\rho_1\hat{\boldsymbol{H}}_{11}\mathrm{diag}(\boldsymbol{Q}_1[l])\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} + \beta\rho_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1[l]\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}) + \hat{\boldsymbol{D}}_{11}\mathrm{tr}\boldsymbol{Q}_1[l] + \boldsymbol{Z}_2[l] \right)$$

$$+ \log\det\left(\beta\eta_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1[l]\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}) + \kappa\eta_1\hat{\boldsymbol{H}}_{21}\mathrm{diag}(\boldsymbol{Q}_1[l])\hat{\boldsymbol{H}}_{21}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{21}\mathrm{tr}\boldsymbol{Q}_1[l] + \boldsymbol{Z}_3[l] \right)$$

$$- \log\det\left(\beta\eta_1\mathrm{diag}(\hat{\boldsymbol{H}}_{11}\boldsymbol{Q}_1[l]\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}) + \kappa\eta_1\hat{\boldsymbol{H}}_{21}\mathrm{diag}(\boldsymbol{Q}_1[l])\hat{\boldsymbol{H}}_{21}^{\mathsf{H}} + \hat{\boldsymbol{D}}_{21}\mathrm{tr}\boldsymbol{Q}_1[l] + \boldsymbol{Z}_4[l] \right) \Big\} \tag{77}$$

$$= \frac{\rho_1}{\ln 2} \Big\{ \left(\hat{\boldsymbol{H}}_{11}^{\mathsf{H}} \left[\boldsymbol{R}_1^{-1}[l] + \beta\mathrm{diag}(\boldsymbol{R}_1^{-1}[l] - \hat{\boldsymbol{\Sigma}}_1^{-1}[l]) \right] \hat{\boldsymbol{H}}_{11} \right)^{\mathsf{T}} + \kappa\mathrm{diag}\left(\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}(\boldsymbol{R}_1^{-1}[l] - \hat{\boldsymbol{\Sigma}}_1^{-1}[l])\hat{\boldsymbol{H}}_{11} \right) \Big\}$$

$$+ \frac{1}{\ln 2}\mathrm{sum}\left(\hat{\boldsymbol{D}}_{11} \odot (\boldsymbol{R}_1^{-1}[l] - \hat{\boldsymbol{\Sigma}}_1^{-1}[l])^{\mathsf{T}} \right)\boldsymbol{I} + \frac{\eta_1}{\ln 2} \Big\{ \kappa\mathrm{diag}\left(\hat{\boldsymbol{H}}_{21}^{\mathsf{H}}(\boldsymbol{R}_2^{-1}[l] - \hat{\boldsymbol{\Sigma}}_2^{-1}[l])\hat{\boldsymbol{H}}_{21} \right)$$

$$+ \beta\left(\hat{\boldsymbol{H}}_{11}^{\mathsf{H}}\mathrm{diag}(\boldsymbol{R}_2^{-1}[l] - \hat{\boldsymbol{\Sigma}}_2^{-1}[l])\hat{\boldsymbol{H}}_{11} \right)^{\mathsf{T}} \Big\} + \frac{1}{\ln 2}\mathrm{sum}\left(\hat{\boldsymbol{D}}_{21} \odot (\boldsymbol{R}_2^{-1}[l] - \hat{\boldsymbol{\Sigma}}_2^{-1}[l])^{\mathsf{T}} \right)\boldsymbol{I} \tag{78}$$

Notice that, for $\boldsymbol{V}_{r,s}$ defined as a zero-valued matrix except for a unity element at row r and column s, we have

$$\frac{\partial \det(\boldsymbol{Y})}{\partial \boldsymbol{X}} = \sum_{r,s} \boldsymbol{V}_{r,s} \frac{\partial \det(\boldsymbol{Y})}{\partial [\boldsymbol{X}]_{r,s}}$$

$$= \sum_{r,s} \boldsymbol{V}_{r,s} \sum_{i,j} \frac{\partial \det(\boldsymbol{Y})}{\partial [\boldsymbol{Y}]_{i,j}} \frac{\partial [\boldsymbol{Y}]_{i,j}}{\partial [\boldsymbol{X}]_{r,s}}. \tag{72}$$

Then, using (72), we get

$$\frac{\partial \det(\boldsymbol{Y})}{\partial \boldsymbol{X}} = \sum_{r,s} \boldsymbol{V}_{r,s} \sum_{i,j} \frac{\partial \det(\boldsymbol{Y})}{\partial [\boldsymbol{Y}]_{i,j}} \Big[[\boldsymbol{C}]_{i,r}[\boldsymbol{D}]_{s,j}\delta_{r-s}$$

$$+ [\boldsymbol{E}]_{i,r}[\boldsymbol{F}]_{s,j}\delta_{i-j} + [\boldsymbol{G}]_{i,j}\delta_{r-s} \Big] \tag{73}$$

$$= \mathrm{diag}\left(\boldsymbol{D} \left[\frac{\partial \det \boldsymbol{Y}}{\partial \boldsymbol{Y}} \right]^{\mathsf{T}} \boldsymbol{C} \right)$$

$$+ \left[\boldsymbol{F}\mathrm{diag}\left(\frac{\partial \det \boldsymbol{Y}}{\partial \boldsymbol{Y}} \right)^{\mathsf{T}} \boldsymbol{E} \right]^{\mathsf{T}}$$

$$+ \mathrm{sum}\left[\boldsymbol{G} \odot \left[\frac{\partial \det \boldsymbol{Y}}{\partial \boldsymbol{Y}} \right] \right]\boldsymbol{I} \tag{74}$$

$$= \det(\boldsymbol{Y})\Big\{ \mathrm{diag}\left(\boldsymbol{D}\boldsymbol{Y}^{-1}\boldsymbol{C} \right) + \left[\boldsymbol{F}\mathrm{diag}\left(\boldsymbol{Y}^{-1} \right)\boldsymbol{E} \right]^{\mathsf{T}}$$

$$+ \mathrm{sum}\left(\boldsymbol{G} \odot (\boldsymbol{Y}^{-1})^{\mathsf{T}} \right)\boldsymbol{I} \Big\} \tag{75}$$

where, for the last step, we used the fact that $\frac{\partial \det(\boldsymbol{Y})}{\partial \boldsymbol{Y}} = \det(\boldsymbol{Y})(\boldsymbol{Y}^{-1})^{\mathsf{T}}$.

Applying (75) to (17), we can obtain an expression for $\frac{\partial I}{\partial \boldsymbol{Q}_i[l]}$. To do so, we think of $\boldsymbol{Z}$ in (70) as representing the terms in $\underline{I}$ that have zero derivative with respect to $\boldsymbol{Q}_i[l]$. Using the shorthand notation $\boldsymbol{R}_i[l] \triangleq \rho_i\hat{\boldsymbol{H}}_{ii}\boldsymbol{Q}_i[l]\hat{\boldsymbol{H}}_{ii}^{\mathsf{H}} + \hat{\boldsymbol{\Sigma}}_i[l]$, and recalling the expression for $\hat{\boldsymbol{\Sigma}}_i[l]$ in (15), the result for $i = 1$ and $l \in \{1, 2\}$ is shown in (76)–(78) at the top of the page.

Finally, using $\boldsymbol{G}_1[l] = 2\left(\frac{\partial I}{\partial \boldsymbol{Q}_1[l]} \right)^*$ and leveraging the fact that $\boldsymbol{R}_i[l]$ and $\hat{\boldsymbol{\Sigma}}_i[l]$ are Hermitian matrices, we get the expres-sion for $\boldsymbol{G}_1[l]$ in (26). A similar expression can be derived for $\boldsymbol{G}_2[l]$.

REFERENCES

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

[2] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. ACM Int. Conf. Mobile Comput. Netw.*, Chicago, IL, Sep. 2010, pp. 1–12.

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

[4] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, 2012 [Online]. Available: http://arxiv.org/abs/1111.2618, to be published

[5] S. Chen, M. Beach, and J. McGeehan, "Division-free duplex for wire-less applications," *Electron. Lett.*, vol. 34, no. 2, pp. 147–148, 1998.

[6] J. Sangiamwong, T. Asai, J. Hagiwara, Y. Okumura, and T. Ohya, "Joint multi-filter design for full-duplex MU-MIMO relaying," pre-sented at the IEEE Veh. Technol. Conf., Barcelona, Spain, Apr. 2009.

[7] S. Sohaib and D. K. C. So, "Asynchronous polarized coopera-tive MIMO communication," in *Proc. IEEE Veh. Technol. Conf.*, Barcelona, Spain, Apr. 2009, pp. 1–5.

[8] T. Riihonen, S. Werner, and R. Wichman, "Spatial loop interference suppression in full-duplex MIMO relays," in *Proc. Asilomar Conf. Sig-nals, Syst., Comput.*, Pacific Grove, CA, Nov. 2009, pp. 1508–1512.

[9] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2010, pp. 1558–1562.

[10] A. Raghavan, E. Gebara, E. M. Tentzeris, and J. Laskar, "Analysis and design of an interference canceller for collocated radios," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 11, pp. 3498–3508, 2005.

[11] B. Radunovic, D. Gunawardena, P. Key, A. P. N. Singh, V. Balan, and G. Dejean, "Rethinking indoor wireless: Low power, low fre-quency, full duplex," Microsoft Research, MSR-TR-2009-148, Jul. 24, 2009 [Online]. Available: http://research.microsoft.com/pubs/104950/TR.pdf

[12] M. Jain, J. I. Choi, T. M. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. ACM Int. Conf. Mobile Comput. Netw.*, Las Vegas, NV, Sep. 2011, pp. 301–312.

[13] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless Systems," 2011 [Online]. Available: http://arxiv.org/abs/1107.1276

[14] D. Senaratne and C. Tellambura, "Beamforming for space division duplexing," in *Proc. IEEE Int. Conf. Commun.*, Kyoto, Japan, Jun. 2011, pp. 1–5.

[15] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Capacity regions and sum-rate capacities of vector Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 5030–5044, Oct. 2010.

[16] S. A. Jafar and M. J. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2637–2642, Jul. 2007.

[17] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," in *Proc. IEEE Int. Conf. Commun.*, Dresden, Germany, Jun. 2009, pp. 1–5.

[18] G. Santella and F. Mazzenga, "A hybrid analytical-simulation procedure for performance evaluation in M-QAM-OFDM schemes in presence of nonlinear distortions," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 142–151, Feb. 1998.

[19] H. Suzuki, T. V. A. Tran, I. B. Collings, G. Daniels, and M. Hedley, "Transmitter noise effect on the performance of a MIMO-OFDM hardware implementation achieving improved coverage," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 867–876, Aug. 2008.

[20] R. M. Gray and T. G. Stockham, Jr, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, pp. 805–812, May 1993.

[21] W. Namgoong, "Modeling and analysis of nonlinearities and mismatches in AC-coupled direct-conversion receiver," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 163–173, Jan. 2005.

[22] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.

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

[24] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, pp. 2839–2848, Nov. 2003.

[25] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

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

Brian P. Day received the B.S. in electrical and computer engineering from The Ohio State University, Columbus, in 2010.

Since 2010, he has been working toward the Ph.D. degree in electrical and computer engineering at The Ohio State University. His primary research interests are full-duplex communication, signal processing, and optimization.

Adam R. Margetts received the dual B.S. degree in electrical engineering and mathematics from Utah State University, Logan, in 2000 and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 2002 and 2005, respectively.

He has been with the Massachusetts Institute of Technology's Lincoln Laboratory, Lexington, MA, since 2005. He holds two patents in the area of signal processing for communications. His current research interests include distributed transmit beamforming, cooperative communications, full-duplex relay systems, space-time coding, and wireless networking.

Daniel W. Bliss received the B.S.E.E. in electrical engineering from Arizona State University, Phoenix, in 1989, and the M.S. degree in physics and the Ph.D. degree from the University of California at San Diego in 1995 and 1997, respectively.

From 1989 to 1991, he worked at General Dynamics, where he designed avionics for the Atlas-Centaur launch vehicle and performed research and development of fault-tolerant avionics. As a member of the superconducting magnet group at General Dynamics from 1991 to 1993, he performed magnetic field calculations and optimization for high-energy particle-accelerator superconducting magnets. His doctoral work (1993–1997) was in the area of high-energy particle physics, searching for bound states of gluons, studying the two-photon production of Hadronic final states, and investigating innovative techniques for lattice-gauge-theory calculations. Since 1997, he has been employed by the Massachusetts Institute of Technology's Lincoln Laboratory, Lexington, where he focuses on adaptive signal processing, parameter estimation bounds, and information theoretic performance bounds for multisensor systems. He is currently a Senior Member of the Technical Staff at MIT Lincoln Laboratory in the Advanced Sensor Techniques Group. His current research topics include multiple-input multiple-output (MIMO) wireless communications, MIMO radar, cognitive radios, radio network performance bounds, geolocation techniques, channel phenomenology, and signal processing and machine learning for anticipatory medical monitoring.

Philip Schniter received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 2000.

From 1993 to 1996, he was a Systems Engineer with Tektronix Inc., Beaverton, OR. After receiving the Ph.D. degree, in 2000, he joined the Department of Electrical and Computer Engineering at The Ohio State University, Columbus, where he is currently an Associate Professor and a Member of the Information Processing Systems (IPS) Lab. In 2008–2009, he was a Visiting Professor at Eurecom, Sophia Antipolis, France, and Supélec, Gif-sur-Yvette, France. His areas of interest include statistical signal processing, wireless communications and networks, and machine learning.

Dr. Schniter received the National Science Foundation CAREER Award, in 2003.