Finite-Length Linear Schemes for Joint Source-Channel Coding over Gaussian Broadcast Channels with Feedback

Yonathan Murin, *Member, IEEE*, Yonatan Kaspi, *Member, IEEE*, Ron Dabora, *Senior Member, IEEE*, and Deniz Gündüz, *Senior Member, IEEE*

***Abstract*—We study linear encoding for a pair of correlated Gaussian sources transmitted over a two-user Gaussian broadcast channel in the presence of unit-delay noiseless feedback, abbreviated as the GBCF. Each pair of source samples is transmitted using a linear transmission scheme in a *finite* number of channel uses. We investigate three linear transmission schemes: A scheme based on the Ozarow-Leung (OL) code, a scheme based on the linear quadratic Gaussian (LQG) code of Ardestanizadeh et al., and a novel scheme derived in this work using a dynamic programming (DP) approach. For the OL and LQG schemes we present lower and upper bounds on the minimal number of channel uses needed to achieve a target mean-square error (MSE) pair. For the LQG scheme in the symmetric setting, we identify the optimal scaling of the sources, which results in a significant improvement of its finite horizon performance, and, in addition, characterize the (exact) minimal number of channel uses required to achieve a target MSE. Finally, for the symmetric setting, we show that for any fixed and finite number of channel uses, the DP scheme achieves an MSE lower than the MSE achieved by either the LQG or the OL schemes.**

I. Introduction

We study the transmission of a pair of correlated Gaussian sources over a two-user memoryless Gaussian broadcast channel (GBC) with correlated noise components at the receivers, in which the transmitter has access to noiseless causal feedback (FB) from both receivers. We abbreviate this channel as the GBCF. Motivated by practical broadcast scenarios with strict power, delay and complexity constraints, e.g., live multimedia broadcast [3], transmission of critical system parameters in a smart grid [4], or body-area sensor networks [5], we focus on uncoded linear transmission schemes, namely, schemes that do not encode over blocks of source symbol pairs.[1]

This work was supported in part by the Israel Science Foundation under grant 396/11, and by the European Research Council (ERC) through Starting Grant BEACON (agreement #677854). Parts of this work were presented at the IEEE Global Conference on Signal and Information Processing (GlobalSIP), December 2014, Atlanta, GA, [1], and at the IEEE Information Theory Workshop (ITW), April 2015, Jerusalem, Israel, [2].

Yonathan Murin is with the Department of Electrical Engineering, Stanford University, Stanford, CA. E-mail: moriny@stanford.edu.

Yonatan Kaspi was with the Information Theory and Applications (ITA) Center at the University of California, San Diego, CA. He is now with Goldman Sachs, New York, NY. E-mail: yonikaspi@gmail.com.

Ron Dabora is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel. E-mail: ron@ee.bgu.ac.il.

Deniz Gündüz is with the Department of Electrical and Electronic Engineering, Imperial College London, London, UK. E-mail: d.gunduz@imperial.ac.uk.

[1]This is motivated by the work [6] which showed that for a zero-delay source coding problem with memoryless sources, joint encoding of all the source symbols does not provide any gain.

Previous studies on GBCFs focused on the channel coding problem which assumes independent and uniformly distributed messages and characterized performance for the infinite horizon regime, i.e., the number of channel uses is unbounded. In the present work we study lossy joint source-channel coding (JSCC) for GBCFs focusing on the finite horizon regime: The sources are assumed to be correlated, and each source is to be reconstructed at its corresponding receiver within a target *non-zero* mean-square error (MSE) distortion. Our objective is to characterize the minimal number of channel uses required to achieve a target MSE pair.

We focus on *linear and memoryless transmission schemes* [7, Sec. III], i.e., the transmitted signal at any time index is restricted to be a linear combination of the encoder state at the same time index, while the encoder state is a linear combination of the state at the previous time index and the channel outputs from the previous transmission. In particular, we consider the following three transmission schemes: 1) A JSCC scheme based on the Ozarow-Leung (OL) channel coding scheme developed in [8], to which we refer as the *JSCC-OL scheme*; 2) A JSCC scheme based on the linear quadratic Gaussian (LQG) channel code derived in [9], to which we refer as the *JSCC-LQG scheme*; and 3) A novel JSCC transmission scheme, which is derived in this work, whose parameters are obtained using dynamic programming (DP) [10], to which we refer as the *JSCC-DP scheme*. While the JSCC-OL and the JSCC-DP schemes are time-varying and are designed based on signal processing arguments, the JSCC-LQG scheme is time-invariant and is based on a control theoretic approach. As we show in the sequel, these differences lead to different performances in the finite horizon regime.

A. Prior Work

While FB does not increase the capacity of memoryless point-to-point (PtP) channels [11], it was shown in [12] that for Gaussian PtP channels FB can reduce the complexity and delay required for achieving a target error probability. In fact, the scheme presented in [12] (referred herein as the SK scheme) achieves a *doubly* exponential decay in the probability of error with the number of transmitted symbols, whereas only a single exponential decay can be achieved without feedback. In the SK scheme, the receiver applies minimum MSE (MMSE) estimation to iteratively estimate the transmitted source (or message). Using the FB, the transmitter can track the estimation error at the receiver, and transmit it at the next

channel symbol. Thus, at each channel use, only the "missing information" is transmitted. The work [13] generalized this idea and presented the posterior matching principle for optimal transmission over memoryless PtP channels with FB: At each channel symbol the transmitter should send only information that is independent of the past transmitted symbols, and is relevant for the reconstruction of the transmitted message.

Differently from the situation for PtP channels, in multiuser channels FB may enlarge the capacity region. This was first demonstrated in [14] which showed that FB enlarges the capacity region of the memoryless multiple-access channel (MAC). Motivated by the optimality of the SK scheme for PtP Gaussian channels, the works [15] and [8] extended it to the two-user Gaussian MAC with FB (GMACF) and to the two-user GBCF, respectively. While for the GMACF this approach achieves the capacity region, for the GBCF this extension is generally suboptimal even though it achieves reliable communications at rate pairs which are outside the capacity region of the non-degraded GBC.[2] The OL scheme of [8] and the scheme of [15] were later extended to GBCFs and GMACFs with more than two users as well as to Gaussian interference channels with FB (GICFs) in [7]. Recently, in [17], we extended the OL scheme by using estimators with memory at the receivers instead of the memoryless estimators used in the original OL scheme of [8]. We note that the extended decoder does not always improve upon the memoryless decoder of [8], in fact, in some situations it may perform worse than the memoryless decoder of [8]. Finally, the work [18] used the scheme of [15] and the OL scheme of [8] to stabilize (in the mean square sense) two linear, discrete-time, scalar and time-invariant systems in closed-loop, via control over GMACFs and GBCFs, respectively. This approach was also used for stabilization over interference channels in [19].

An alternative approach to SK-type schemes is based on control theory. For Gaussian PtP channels [20] showed that solving an optimal LQG control problem leads to a capacity achieving FB transmission scheme, and presented control-oriented FB transmission schemes also for GBCFs, GMACFs and GICFs. In particular, for the two-user GBCF with *independent noise components* at the receivers, [20] presented a class of coding schemes which achieve rate pairs outside the achievable rate region of the OL scheme. Later, [9] used the LQG control framework to develop a FB coding scheme for the GBCF, referred herein as the *LQG scheme*, which does not require the noise components to be independent. It was also shown in [9] that when the noise components are independent and have the same variance, then the LQG scheme achieves rates higher than those achieved by the OL scheme and the scheme of [7] (for the case of more than two users). In fact, recently, [21] showed that for this scenario the LQG scheme achieves the maximal sum-rate among all possible linear-feedback schemes. GBCFs and GICFs were also studied in [22] which presented a (non-linear with memory) transmission scheme whose sum-rate approaches the corresponding full-cooperation bound,[3] as the signal-to-noise ratio (SNR) increases to infinity. Lastly, the recent work [23] showed that the capacity region of the GBCF with independent noises and only a common message cannot be achieved by linear feedback schemes such as the OL or LQG schemes.

While all the works on GBCFs reviewed above focus on the *achievable rates*, namely, bits per channel use that can be transmitted reliably as the number of channel uses goes to *infinity*, in the present work we study a JSCC problem. JSCC in multiuser networks with FB has been considered in several previous works. The work [24] presented sufficient conditions for lossy transmission of discrete memoryless (DM) correlated sources over a DM-MAC with FB which builds upon the hybrid coding scheme of [25]. Lossy transmission of correlated Gaussian sources over a two-user GMACF was studied in [26], in which sufficient conditions and necessary conditions for the achievability of an MSE pair were derived. In [26] it was also shown that for the symmetric setting, if the channel SNR is below a certain threshold, then an uncoded transmission scheme is optimal. While the works [24]–[26] focused on the scenario in which the source and channel bandwidths are matched, the work [27] considered scenarios in which these bandwidths are mismatched, and studied the transmission of correlated Gaussian sources over a two-user GMACF. For the symmetric setting, [27] presented upper and lower bounds on the energy-distortion tradeoff, i.e., the minimum transmission energy required to communicate a pair of sources over a noisy channel, such that the sources can be reconstructed within a specified target distortion. In [28] we study the energy-distortion tradeoff for the symmetric two-user GBCF. Here we remark that all the aforementioned works consider the infinite horizon regime.

Finally, we note that several works considered the GBC with *noisy* causal FB, see [29]–[32] and references therein. Particularly relevant to the context of the current work are the works [31] and [32] which studied channel coding for PtP Gaussian channels with noisy FB and for GBCs with noisy FB links, respectively. The transmission schemes studied in these works are based on the SK and OL schemes, respectively, while the noise in the feedback links was handled using a modulo-lattice precoding in both the direct and feedback links. Furthermore, [31] and [32] show that while adding noise to the feedback links results in performance degradation [31, Sec. V.D], many of the benefits of noiseless feedback can be carried over to the more practical setup of noisy feedback. This strengthens the motivation for the study presented in the following sections.

B. Main Contributions

In this work we study the transmission of a pair of correlated Gaussian sources over the two-user GBCF in the *finite horizon regime*, where each source is to be reconstructed at its corresponding receiver within a target MSE distortion. Our aim is to characterize the minimal number of channel uses required

[2]We note that feedback does not enlarge the capacity region of degraded GBCFs [16].

[3]In the full-cooperation bound each receiver knows the other receiver's channel output, see [22, Eqn. (1)].

to achieve the target MSE distortion pair. In the following we highlight our main contributions:

1) We adapt the OL scheme of [8] to the transmission of correlated Gaussian sources over GBCFs. This new JSCC scheme is referred to in the following as JSCC-OL. We first demonstrate that the initialization, which takes advantage of the correlation between the sources, is superior to the initialization suggested in [8]. Then, for the proposed JSCC-OL scheme we derive upper and lower bounds on the minimal number of channel uses needed to achieve a target pair of MSEs. We show that, *in contrast to the infinite horizon* regime, in the finite horizon regime there are many cases in which JSCC-OL outperforms JSCC-LQG. Lastly, we consider the symmetric setting, in which the sources have equal variances, the noise variances at the receivers are equal, and the target MSEs are equal. We show that for this setting in the low SNR regime, when the sources are independent, and the noises are independent, then JSCC-OL achieves approximately the same source-channel bandwidth ratio as the *best known* separation-based scheme which applies source and channel coding with *asymptotically large* blocklengths.[4] Since JSCC-OL applies linear encoding and decoding, this demonstrates the efficiency and attractiveness of the JSCC-OL scheme, even in the infinite horizon regime.
2) We adapt the LQG scheme of [9] to the transmission of correlated Gaussian sources over GBCFs in the finite horizon regime. This new JSCC scheme is referred to in the following as JSCC-LQG. As the original LQG scheme is optimized for the infinite horizon regime, it sometimes performs poorly when the horizon is finite. For this reason, our first contribution in the context of JSCC-LQG is the *derivation of a new decoder* based on the MMSE criterion, which outperforms the LQG decoder presented in [9] in the finite horizon regime, while achieving the same performance in the infinite horizon regime. For the general setting we derive *lower and upper bounds on the minimal number of channel uses needed to achieve a target pair of MSEs* with the JSCC-LQG scheme. For the symmetric setting, we show that, by properly scaling the transmitted sources, it is possible to achieve a the target MSE pair with significantly fewer channel transmissions than with the original initialization of [9]. We show that the proposed scaling technique[5]: 1) Optimally exploits the available transmission power, subject to a per-symbol average power constraint, 2) Minimizes the distance between the covariance matrices of the JSCC-LQG initial state and the covariance matrix of its steady-state (subject to the per-symbol average power constraint), and 3) Achieves the *same MSE exponents*[6] as in [9]. Thus, our proposed JSCC-LQG scheme is a *linear time-invariant* transmission scheme with *very good finite horizon performance and with the best known infinite horizon performance*. Finally, for the symmetric setting, we *explicitly* characterize the minimal number of channel uses required to achieve a target MSE pair via the roots of a second order polynomial, thus, providing a complete performance characterization of the JSCC-LQG in the finite horizon regime for the symmetric setting.
3) We present a new linear and memoryless transmission scheme based on DP, called JSCC-DP. For a finite number of channel uses, we show that the JSCC-DP scheme achieves lower MSE values than those achieved by the JSCC-LQG and the JSCC-OL schemes. Since finding the coefficients of this scheme becomes computationally infeasible as the number of channel uses becomes large, we also propose an approximate low-complexity version of the JSCC-DP scheme. Simulation results indicate that for moderate to high SNRs this approximate version has a negligible or no performance loss compared to the exact JSCC-DP scheme.

The rest of this paper is organized as follows: The problem formulation is introduced in Section II. We introduce the JSCC-OL and JSCC-LQG schemes in Sections III and IV, respectively. The JSCC-DP scheme is introduced in Section V, and a comparison of the three schemes along with numerical examples are presented in Section VI. Finally, concluding remarks are presented in Section VII.

II. Problem Definition

A. Notation

We use capital letters to denote random variables (RVs), e.g., X, and boldface letters to denote random column vectors, e.g., $\mathbf{X}$; the k'th element of a vector $\mathbf{X}$ is denoted by X_k, and we use $\mathbf{X}_k^j$ where $k<j$, to denote $(X_k, X_{k+1}, ..., X_{j-1}, X_j)$. We use sans-serif font to denote deterministic vectors and matrices: boldface letters denote vectors, e.g., $\mathsf{\mathbf{B}}$, while regular letters denote matrices, e.g., M. $[\mathsf{M}]_{m,n}$ denotes the entry at the m'th row and n'th column of a matrix M, and $\det(\mathsf{M})$ denotes the determinant of a square matrix M. I denotes the identity matrix. We use $\mathbb{E}\{\cdot\}$, $(\cdot)^T$, $\log(\cdot)$ and $\mathfrak{R}$ to denote expectation, transpose, natural basis logarithm, and the set of real numbers, respectively. We use $\mathcal{N}(\mu, \mathsf{Q})$ to denote the multivariate normal distribution with mean μ and covariance matrix Q. Finally, we define $[x]^+ \triangleq \max\{x, 0\}$, $\mathrm{sgn}(x)$ as the sign of x, where $\mathrm{sgn}(0) \triangleq 1$, and denote the ceiling function of x by $\lceil x \rceil$.

B. System Model

The two-user GBCF is depicted in Fig. 1. All the signals are real. The encoder observes a realization of a pair of correlated and jointly Gaussian sources, denoted by $\mathbf{S} = [S_1, S_2]^T$, and is required to send the source $S_i, i = 1, 2$, to the i'th receiver, denoted by Rx_i. Let $\mathbf{S} \sim \mathcal{N}(0, \mathsf{Q}_s)$, where the covariance matrix, Q_s, is given by:

$$\mathsf{Q}_s = \begin{bmatrix} \sigma_1^2 & \rho_s \sigma_1 \sigma_2 \\ \rho_s \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix},$$

[4]Note that for the considered setting, when the sources are independent, then separate source-channel coding is optimal.

[5]Note that finding this scaling factor requires a substantial technical effort as we must characterize the exact instantaneous transmission power of the LQG scheme. This is explained in detail in Subsection IV-E1.

[6]MSE exponents corresponds to the slope of decay of the logarithm of the MSE for sufficiently large number of channel uses.

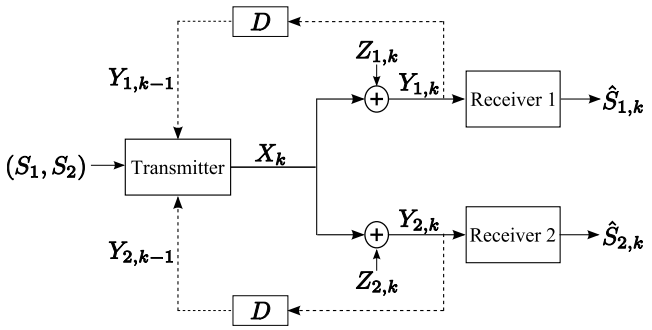


Fig. 1: Gaussian broadcast channel with correlated sources and feedback links. $\hat{S}_{1,k}$ and $\hat{S}_{2,k}$ are the reconstructions of S_1 and S_2, respectively, after the k'th channel use.

where

$$\sigma_i^2 = \mathbb{E}\left\{S_i^2\right\}, \quad \rho_s = \frac{\mathbb{E}\left\{S_1 S_2\right\}}{\sigma_1 \sigma_2}, \quad |\rho_s| < 1.$$

Each pair of source symbols is transmitted using K channel uses, indexed by $k = 1, 2, \ldots, K$. The channel outputs at the decoders are given by:

$$Y_{i,k} = X_k + Z_{i,k}, \quad i = 1, 2, \tag{1}$$

where X_k denotes the channel input at time k, the noise components $[Z_{1,k}, Z_{2,k}]^T \sim \mathcal{N}(0, \mathsf{Q}_z)$, are independent and identically distributed (i.i.d.) over $k = 1, 2, \ldots, K$, with covariance matrix Q_z given by:

$$\mathsf{Q}_z = \begin{bmatrix} \sigma_{z,1}^2 & \rho_z \sigma_{z,1} \sigma_{z,2} \\ \rho_z \sigma_{z,1} \sigma_{z,2} & \sigma_{z,2}^2 \end{bmatrix},$$

where

$$\rho_z = \frac{\mathbb{E}\left\{Z_1 Z_2\right\}}{\sigma_{z,1} \sigma_{z,2}}, \quad |\rho_z| < 1.$$

Let $\mathbf{B} \triangleq [1, 1]^T$, $\mathbf{Y}_k \triangleq [Y_{1,k}, Y_{2,k}]^T$ and $\mathbf{Z}_k \triangleq [Z_{1,k}, Z_{2,k}]^T$. The signal model (1) can now be written in the following vector form:

$$\mathbf{Y}_k = \mathbf{B} X_k + \mathbf{Z}_k. \tag{2}$$

At time $k = 1, 2, \ldots, K$, Rx_i, $i = 1, 2$, uses its received channel outputs, $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k}$, to estimate S_i:

$$\hat{S}_{i,k} = g_{i,k}(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k}), \quad g_{i,k} : \mathfrak{R}^k \to \mathfrak{R}, \tag{3}$$

and the encoder maps the observed pair of sources and the received FB into a channel input via:

$$X_k = f_k(S_1, S_2, \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{k-1}), \quad f_k : \mathfrak{R}^{2k} \to \mathfrak{R}, \tag{4}$$

subject to a per-symbol average power constraint defined as:

$$\mathbb{E}\left\{X_k^2\right\} \le P, \quad \forall k = 1, 2, \ldots, K. \tag{5}$$

For a specific set of parameters $(\sigma_{z,1}^2, \sigma_{z,2}^2, \rho_z, \sigma_1^2, \sigma_2^2, \rho_s)$, we define a (D_1, D_2, K) code to be a collection of K encoding functions each satisfying (5), and two decoding functions such that:

$$\mathbb{E}\{(S_i - \hat{S}_{i,K})^2\} \le D_i, \quad 0 < D_i \le \sigma_i^2, \quad i = 1, 2. \tag{6}$$

For a given target MSE pair (D_1, D_2), our objective is to characterize the *minimal number* of channel uses K such that a (D_1, D_2, K) code exists. In the sequel, we let K_{SCHEME} denote the minimal number of channel uses required to achieve a pair of MSE distortion values, (D_1, D_2), by the scheme "SCHEME"$\in \{\text{OL}, \text{LQG}, \text{DP}\}$.

Remark 1. Note that in (5) we use a per-symbol average power constraint, similarly to [33, Eq. (22)] and [34, Sec. VII]. The per-symbol average power constraint is motivated by practical system implementation: Due to the finite dynamic range of power amplifiers [35, Ch. 9], the transmitter is not able to allocate power arbitrarily across time slots. This is particularly relevant for simple energy-limited sensor nodes which may benefit the most from the low-complexity linear encoding schemes proposed in this paper.

C. Linear and Memoryless JSCC for GBCFs

In this work we focus on the class of *linear and memoryless transmission schemes*, see, e.g., [7, Sec. III]: In this class of schemes, the transmitted signal at any time index, k, is restricted to be a linear function of the encoder state at time k, which, in turn, evolves as a linear combination of the encoder state at time $k-1$ and the channel outputs at time $k-1$. Letting $\mathbf{U}_k = [U_{1,k}, U_{2,k}]^T$ denote the encoder state vector at time k (each state is associated with one Tx-Rx link), the transmitted signal at time k, X_k, see (4), is generated as a *linear* (possibly time-varying) combination of the elements of $\mathbf{U}_k$: $X_k = \mathbf{T}_k^T \mathbf{U}_k$, where $\mathbf{T}_k = [t_{1,k}, t_{2,k}]^T$ are the combination weights. Furthermore, the encoder state vector is recursively obtained by:

$$U_{i,k} = \varphi_{i,k}(U_{i,k-1}, Y_{i,k-1}), \quad i = 1, 2, \quad k = 1, 2, \ldots, K, \tag{7}$$

where $\varphi_{i,k}(\cdot)$ is a *linear* mapping. In the following sections we state the three studied schemes, JSCC-OL, JSCC-LQG, and JSCC-DP, as instances of this class of transmission schemes. Using this general definition, we can highlight the fundamental differences between the schemes: In the JSCC-LQG scheme $\mathbf{T}_k$ and $\varphi_{i,k}$ do not depend on k and consequently the JSCC-LQG scheme is *time-invariant*. On the other hand, in the JSCC-OL and JSCC-DP schemes $\mathbf{T}_k$ and $\varphi_{i,k}$ are *time-varying*. In the JSCC-OL scheme $\varphi_{i,k}$ are based on linear and memoryless MMSE estimators of $U_{i,k}$ from $Y_{i,k-1}$, while in the JSCC-DP scheme $\varphi_{i,k}$ are computed recursively via DP. Clearly, while structure makes implementation and analysis easier, it may result in an inferior performance. Indeed, we show in Subsection VI-B that, even though the JSCC-LQG scheme achieves the best known MSE exponent, the JSCC-OL scheme can outperform the JSCC-LQG scheme in the finite horizon regime. This is because the time-varying nature of JSCC-OL allows for better exploiting the available power and the correlation between the sources. As the JSCC-DP scheme computes $\mathbf{T}_k$ and $\varphi_{i,k}$ recursively using the statistics of the signals, it can also adaptively change the transmission coefficients similarly to the JSCC-OL scheme. However, as these weights are obtained recursively, JSCC-DP achieves the smallest MSE at any a-priori specified number of channel uses.

Next, we recall some results and definitions from [8], and provide a finite horizon analysis of the JSCC-OL scheme.

$$\rho_k = \frac{(\rho_z\sigma_{z,1}\sigma_{z,2}\Xi + \xi_1^2\xi_2^2)\rho_{k-1} - \Psi_{k-1}^2\Xi \cdot g(1-\rho_{k-1}^2)\mathrm{sgn}(\rho_{k-1})}{\sqrt{\pi_1\pi_2}\sqrt{\sigma_{z,1}^2 + \Psi_{k-1}^2 g^2(1-\rho_{k-1}^2)}\sqrt{\sigma_{z,2}^2 + \Psi_{k-1}^2(1-\rho_{k-1}^2)}}. \tag{12}$$

III. JSCC via the OL Scheme

A. The OL Scheme for JSCC

In the JSCC-OL scheme, the transmitter generates the channel symbol to be transmitted at time k based on the previous channel outputs, available through the FB links. The transmitter first calculates the source estimates at the receivers, and obtains the estimation errors at each receiver. The transmitter then sends a linear combination of these estimation errors. Thus, at each time k, each receiver obtains its estimation error corrupted by a correlated noise term, consisting of the other receiver's estimation error and the additive channel noise. Each receiver then updates its estimate accordingly, thereby, decreasing the variance of its estimation error. The scheme is terminated after K_{OL} channel uses, where K_{OL} is chosen such that the target MSE for each source is achieved at the corresponding receiver.

***Setup*:** Let $\hat{S}_{i,k}$ be the estimate of S_i at Rx_i after the reception of the k'th channel output, $Y_{i,k}$. Letting $\epsilon_{i,k} \triangleq \hat{S}_{i,k} - S_i$ be the estimation error after k transmissions, and defining $\hat{\epsilon}_{i,k-1} \triangleq \hat{S}_{i,k-1} - \hat{S}_{i,k}$, we can write $\epsilon_{i,k} = \epsilon_{i,k-1} - \hat{\epsilon}_{i,k-1}$. Lastly, define $\alpha_{i,k} \triangleq \mathbb{E}\{\epsilon_{i,k}^2\}$ as the MSE at Rx_i after k transmissions, and $\rho_k \triangleq \frac{\mathbb{E}\{\epsilon_{1,k}\epsilon_{2,k}\}}{\sqrt{\alpha_{1,k}\alpha_{2,k}}}$ as the correlation coefficient between the estimation errors.

***Encoding*:** Set $\hat{S}_{i,0} = 0$, which yields $\epsilon_{i,0} = -S_i, \alpha_{i,0} = \mathbb{E}\left\{\epsilon_{i,0}^2\right\} = \sigma_i^2$, and $\rho_0 = \rho_s$. Next, for a given P, let $g > 0$ be a constant which controls the tradeoff between the information rate to Rx_1 and Rx_2, and define $\Psi_k \triangleq \sqrt{\frac{P}{1+g^2+2g|\rho_k|}}$. At the k'th iteration, $1 \le k \le K$, the transmitter sends

$$X_k = \Psi_{k-1} \cdot \left(\frac{\epsilon_{1,k-1}}{\sqrt{\alpha_{1,k-1}}} + \frac{\epsilon_{2,k-1}}{\sqrt{\alpha_{2,k-1}}} \cdot g \cdot \mathrm{sgn}(\rho_{k-1}) \right), \tag{8}$$

and the corresponding channel outputs are given in (1).

Remark 2. It follows from (8) that the average per-symbol transmission power of the JSCC-OL scheme is constant. Therefore, JSCC-OL inherently satisfies the average per-symbol power constraint in (5).

***Decoding*:** After the k'th channel use, the estimator that minimizes the instantaneous MSE, $\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\}$, is the conditional expectation [36, Eqn. (11.10)], i.e.,

$$\hat{S}_{i,k} = \mathbb{E}\left\{S_i | [Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k}]^T\right\}.$$

However, as successive channel outputs are not independent, the performance analysis of this estimator is highly complicated. For this reason, a simpler and suboptimal approach is considered in [8], in which Rx_i estimates $\epsilon_{i,k-1}$ based only on $Y_{i,k}$:

$$\hat{\epsilon}_{i,k-1} = \frac{\mathbb{E}\left\{\epsilon_{i,k-1}Y_{i,k}\right\}}{\mathbb{E}\left\{Y_{i,k}^2\right\}} Y_{i,k}. \tag{9}$$

Then, similarly to [37, Eq. (7)], the estimate of S_i is given by:

$$\hat{S}_{i,k} = -\sum_{m=1}^{k} \hat{\epsilon}_{i,m-1}. \tag{10}$$

Remark 3. In [15] it is shown that for the 2-user GMACF this approach is optimal in the MMSE sense. This follows as in the MAC setup both sources are estimated from the same channel output, thus, the estimation errors are orthogonal to the previous channel output. On the other hand, in the GBCF this approach is sub-optimal since $[Y_{1,1}, Y_{1,2}, \ldots, Y_{1,k-1}]^T$ is not necessarily orthogonal to $\epsilon_{2,k-1}$. In [17] we extended the estimator (9) to use $[Y_{i,k}, Y_{i,k-1}]^T$ instead of using only $Y_{i,k}$. This resulted in a transmission scheme which is linear but not memoryless.

Define $\pi_i \triangleq P + \sigma_{z,i}^2$, $\Xi \triangleq P + \sigma_{z,1}^2 + \sigma_{z,2}^2 - \rho_z\sigma_{z,1}\sigma_{z,2}$, and $\xi_i^2 \triangleq \sigma_{z,i}^2 - \rho_z\sigma_{z,1}\sigma_{z,2}$. In [8] the MSEs of the (memoryless) estimators in (9) are stated via the recursive expressions [8, Eqs. (5)–(6)]:

$$\alpha_{1,k} = \alpha_{1,k-1}\frac{\sigma_{z,1}^2 + \Psi_{k-1}^2 g^2(1-\rho_{k-1}^2)}{\pi_1}, \tag{11a}$$

$$\alpha_{2,k} = \alpha_{2,k-1}\frac{\sigma_{z,2}^2 + \Psi_{k-1}^2 (1-\rho_{k-1}^2)}{\pi_2}, \tag{11b}$$

and ρ_k is given by [8, Eqn. (7)] at the top of the page. The JSCC-OL scheme described above can be readily stated within the class of linear and memoryless schemes defined in Subsection II-C: The encoder state update for the JSCC-OL scheme can be expressed via (7) by setting $\mathbf{U}_k = [\epsilon_{1,k-1}, \epsilon_{2,k-1}]^T$, the transmitted signal X_k is obtained from the encoder states $\mathbf{U}_k$ via (8), and $U_{i,k+1} \equiv \epsilon_{i,k} = \epsilon_{i,k-1} - \hat{\epsilon}_{i,k-1}, \hat{\epsilon}_{i,k-1} = \frac{\mathbb{E}\{\epsilon_{i,k-1}Y_{i,k}\}}{\mathbb{E}\left\{Y_{i,k}^2\right\}} Y_{i,k}$. Observe that the linear estimation and transmission coefficients are time-varying.

B. Initialization of the JSCC-OL Scheme

In the above description of the JSCC-OL scheme the initialization is different than the original initialization in [8]. In this subsection we begin by motivating the initialization in Subsection III-A, i.e., $\epsilon_{i,0} = -S_i$ and $\rho_0 = \rho_s$, and then discuss alternative initialization approaches.

First, note that the instantaneous MSEs in (11) are monotonically decreasing functions of $|\rho_k|$. Thus, at least in the first transmission, there is a strong motivation to generate $\epsilon_{1,0}$ and $\epsilon_{2,0}$ as correlated as possible. Numerical simulations indicate that the benefits of highly correlated $\epsilon_{1,0}$ and $\epsilon_{2,0}$ carry beyond the first transmission. In particular, in [28, Appendix C] we analytically show that in the low SNR regime, $|\rho_k|$ slowly decreases towards zero, from its initial value, until it reaches a very small steady state value. Consequently, in the low SNR regime, initializing the scheme with correlated estimates yields substantial benefits over initializing with the steady-state

correlation. We conclude that, when the sources are correlated, initialization with $\epsilon_{i,0} = -S_i$ and $\rho_0 = \rho_s$, takes advantage of the correlation between the sources to rapidly decrease the MSE.

Next, we address the relevance of the initialization proposed in [8] to the problem studied in the current work. First, note that *the approach of [8] maximizes the achievable MSE exponents, and not the finite horizon performance*. In [8] it was shown that there exists a $\rho \in [0,1]$ such that a steady state is achieved in the sense that when $|\rho_{k-1}| = \rho$, then $\rho_k = -\rho_{k-1}$. This ρ^* can be obtained by setting $\rho_k = \rho$ and $\rho_{k-1} = -\rho$ in (12), finding the roots of the resulting sixth-order polynomial, and taking ρ^* as the largest root of this polynomial in $[0,1]$. In [8, pg. 669] it was also shown how to initialize the transmission to achieve a steady-state in (12), with $\rho_k = \rho^*, k \geq 2$. When an average per-symbol power constraint is applied, then the initialization suggested in [8] is suitable only if the steady state correlation coefficient satisfies $|\rho^*| \in \left[0, \frac{P}{\sqrt{(P+\sigma_{z,1}^2)(P+\sigma_{z,2}^2)}}\right)$. In the low SNR regime, $\frac{P}{\sqrt{(P+\sigma_{z,1}^2)(P+\sigma_{z,2}^2)}}$ is very small, and this initialization may result in a very slow decrease in the MSEs.

We now consider a general linear initialization which can achieve *any initial correlation value* ρ_0, such that $|\rho_0| \in [0,1)$. Let F be a deterministic matrix, $\mathbf{W} \sim \mathcal{N}(0, \mathsf{Q}_w)$ independent of $\mathbf{S}$, $\epsilon_0 = [\epsilon_{1,0}, \epsilon_{2,0}]^T$, and consider the following initialization:

$$\epsilon_0 = \mathsf{F}\mathbf{S} + \mathbf{W}. \tag{13}$$

Setting $\epsilon_{i,0} = -S_i$ is a special case of this general linear framework obtained with $\mathbf{W} = \mathbf{0}$ and $\mathsf{F} = -\mathsf{I}$. Since minimizing K_{OL} over all matrices F and over all covariance matrices Q_w is rather involved, in the following we aim at setting $\rho_{\epsilon_0} = \frac{\mathbb{E}\{\epsilon_{1,0}\epsilon_{2,0}\}}{\mathbb{E}\{\epsilon_{1,0}^2\}\mathbb{E}\{\epsilon_{2,0}^2\}}$ to be as large as possible. We next discuss two special instances of (13). To simplify the analytic treatment, we focus on the symmetric setting in which $\sigma_1^2 = \sigma_2^2 \triangleq \sigma_s^2, \sigma_{z,1}^2 = \sigma_{z,2}^2 \triangleq \sigma_z^2, D_1 = D_2 \triangleq D$, and we set $g = 1$ as no preference should be given to either of the sources. We begin by considering initialization using only noise addition, i.e., $\mathsf{F} = \mathsf{I}$.

Initialization via noise addition: Let $\mathsf{Q}_w = \sigma_w^2 \begin{bmatrix} 1 & \rho_w \\ \rho_w & 1 \end{bmatrix}$, and let ρ_0 be the desired correlation coefficient. We are interested in finding Q_w such that $\rho_{\epsilon_0} = \rho_0$. Since the transmitted signal in (8) is always scaled to satisfy the per-symbol average power constraint P, there are many pairs (ρ_w, σ_w^2) which result in $\rho_{\epsilon_0} = \rho_0$. To maximize the component of $\mathbf{S}$ in ϵ_0, we select the (ρ_w, σ_w^2) pair with the minimal σ_w^2. This pair is given by:

$$\sigma_w^2 = \frac{\sigma_s^2(\rho_s - \rho_0)}{\rho_0 + \text{sgn}(\rho_s - \rho_0)}, \quad \rho_w = \text{sgn}(\rho_0 - \rho_s).$$

By letting the JSCC-OL scheme transmit ϵ_0, the higher correlation coefficient facilitates a rapid decrease in the MSE at the receivers, at the cost of using some of the available power to transmit the noise vector $\mathbf{W}$. Thus, we have two contradicting effects: The increased ρ_0 *decreases* the MSE in estimating ϵ_0, but adding the noise $\mathbf{W}$ ϵ_0 increases the MSE in estimating $\mathbf{S}$. In Appendix A-A we derive a two-step MMSE estimator which first applies the OL scheme with $\epsilon_{i,0}$, and then applies MMSE estimation of S_i from the estimated $\epsilon_{i,0}$. The MSE of this estimator is lower bounded by the MSE of estimating S_i from $\epsilon_{i,0}$. Extensive numerical simulations indicate that the MSE achieved by this approach is higher than the MSE achieved by the initialization $\epsilon_{i,0} = -S_i$. We illustrate this point in the following example.

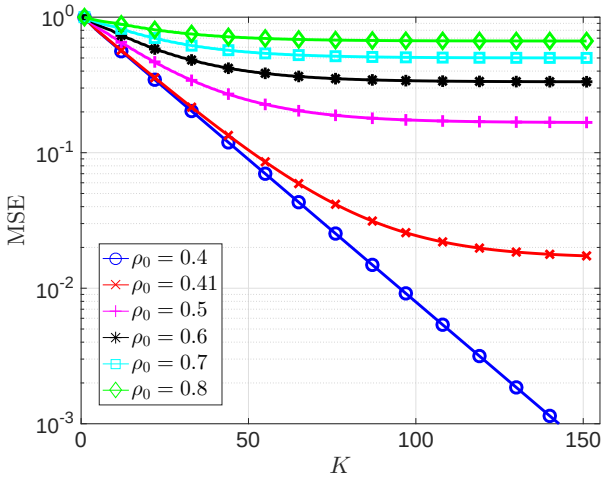


Fig. 2: MSE vs. number of channel uses for different values of ρ_0. $\rho_s = 0.4, \sigma_s^2 = 1, \rho_z = 0.3, \sigma_z^2 = 1$, and $P = 0.1$.

Example 1. Let $\rho_s = 0.4, \sigma_s^2 = 1, \rho_z = 0.3, \sigma_z^2 = 1$, and $P = 0.1$. Fig. 2 depicts the MSEs vs. K for this scenario. Note that $\rho_0 = 0.4$ corresponds to the initialization $\epsilon_{i,0} = -S_i$ as no noise need to be added. It can be observed that any value of ρ_0 other than $\rho_s = 0.4$ degrades the performance compared to the initialization $\epsilon_{i,0} = -S_i$. Furthermore, it can be observed that the MSE floor increases with ρ_0, as expected. This implies that for the two-step estimator, increasing $|\rho_0|$ increases the MSE due to the addition of the noise, rendering this initialization useless.

Remark 4. When the encoder and the decoders share a common source of randomness [38], then at least part of the added noise $\mathbf{W}$ can be eliminated at the receivers. In such a case, the above two-step estimator can achieve lower MSE than the MSE achieved with $\epsilon_{i,0} = -S_i$. Furthermore, in scenarios in which the noise $\mathbf{W}$ can be completely eliminated, the two-step estimator can achieve the MSE given in (11), where $|\rho_0|$ can be initialized to any value in the range $[0,1)$.

Remark 5. Note that if in the first step of the JSCC-OL scheme the receivers estimate S_i instead of $\epsilon_{i,0}$, then $\alpha_{i,1}$ and ρ_1 are no longer given by (11) and (12), respectively. On the other hand, as $\mathbf{S}$ is estimated directly from the channel outputs, the estimate does not suffer from an MSE floor. This observation motivates choosing (ρ_w, σ_w^2) to simultaneously minimize the MSE in estimating $\mathbf{S}$ after the first channel use, and maximize $|\rho_1|$, the correlation coefficient between the two estimation errors after the first channel use. Since this is a non-linear optimization problem which is difficult to solve analytically, we derived explicit expressions for the new

$\alpha_{i,1}$ and ρ_1, then we used a numerical search to find the ρ_0 which maximizes ρ_1, and chose the pair (ρ_w, σ_w^2) to minimize the new $\alpha_{i,1}$ while achieving the above maximizing ρ_0. An extensive numerical study indicates that for most combinations of scenario parameters, the maximal achievable $|\rho_1|$ is very close to $|\rho_s|$, thus, the gain from increasing the correlation ρ_0 beyond $|\rho_s|$ is minor. A non-negligible increase in ρ_1 is observed when $|\rho_s|$ is small and $P \gg \sigma_z^2$. Yet, even in these cases, the MSE obtained with this modified initialization was higher than the MSE achieved by initializing $\epsilon_{i,0} = -S_i$.

Next, we briefly discuss initialization via multiplication by F while $\mathbf{W} = \mathbf{0}$.

Initialization via (13) **with** $\mathbf{W} = \mathbf{0}$**:** Let $\mathsf{Q}_{\epsilon_0} = \sigma_{\epsilon_0}^2 \begin{bmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{bmatrix}, |\rho_0| < 1$, denote the covariance matrix of ϵ_0. The desired initialization is applied only through the product $\epsilon_0 = \mathsf{F}\mathbf{S}$. Since $|\rho_s| < 1$ and $|\rho_0| < 1$, the matrix F is unique and can be obtained by applying the Cholesky decomposition [39, Subsection 19.2.1.2] to both Q_s and Q_{ϵ_0}. Let the Cholesky decompositions of Q_s and Q_{ϵ_0} be $\mathsf{Q}_s = \mathsf{L}_s \mathsf{L}_s^T$, and $\mathsf{Q}_{\epsilon_0} = \mathsf{L}_{\epsilon_0} \mathsf{L}_{\epsilon_0}^T$, respectively. Using these decompositions we have $\mathsf{F} = \mathsf{L}_{\epsilon_0} \mathsf{L}_s^{-1}$. Note that from the definition of the Cholesky decomposition, the matrix $\mathsf{L}_{\epsilon_0} \mathsf{L}_s^{-1}$ is lower triangular. Therefore, $\epsilon_{1,0}$ is a scaled version of S_1, while $\epsilon_{2,0}$ is a linear combination of both S_1 and S_2. Similarly to the discussion for initialization via noise addition, one can either estimate $\mathbf{S}$ in two steps, first estimating ϵ_0 and then estimating $\mathbf{S}$ from ϵ_0, or estimate $\mathbf{S}$ directly from the channel outputs after the first step. The first approach, of the two steps estimator, results in an MSE floor at Rx_2 (if Rx_2 has access to S_1, then this MSE floor can be eliminated). The second approach, of direct estimation, typically achieves only a small increase in the correlation coefficient compared to $\epsilon_0 = -\mathbf{S}$, and no gains in the MSE for estimating $\mathbf{S}$ were observed.

In the next subsection we study the JSCC-OL scheme in the finite horizon regime with the initialization $\epsilon_{i,0} = -S_i$.

C. Finite Horizon Analysis of JSCC-OL

From (11) it follows that the MSEs at time instance k depend on ρ_{k-1}. However, as ρ_k is defined via the non-linear recursion (12), it follows that an explicit characterization of K_{OL} is highly complex. Thus, in the following theorem we present upper and lower bounds on K_{OL}.

Theorem 1. The JSCC-OL scheme with the decoder defined in (9) terminates within $K_{\mathrm{OL}}^{\mathrm{lb}} \leq K_{\mathrm{OL}} \leq K_{\mathrm{OL}}^{\mathrm{ub}}$ channel uses, where:

$$K_{\mathrm{OL}}^{\mathrm{ub}} = \left\lceil \frac{(1+g^2)}{P} \max\left\{ \pi_1 \log\left(\frac{\sigma_1^2}{D_1}\right), \frac{\pi_2}{g^2} \log\left(\frac{\sigma_2^2}{D_2}\right) \right\} \right\rceil, \quad (14a)$$

$$K_{\mathrm{OL}}^{\mathrm{lb}} = \left\lceil \max\left\{ \frac{\sigma_{z,1}^2}{P} \log\left(\frac{\sigma_1^2}{D_1}\right), \frac{\sigma_{z,2}^2}{P} \log\left(\frac{\sigma_2^2}{D_2}\right) \right\} \right\rceil. \quad (14b)$$

Proof. The upper and lower bounds in (14a) and (14b), respectively, are obtained via lower and upper bounding ρ_k in (11). A detailed proof is provided in Appendix A-B. □

D. JSCC-OL vs. Separate Source-Channel Coding

Next, we focus on the symmetric setting in which $\sigma_1^2 = \sigma_2^2 = \sigma_s^2, \sigma_{z,1}^2 = \sigma_{z,2}^2 = \sigma_z^2$ and $D_1 = D_2 = D$. As symmetry implies the same rate should be allocated for sending both sources, we set $g = 1$. In the following, we compare the source-channel bandwidth of the JSCC-OL scheme with that of separate source-channel coding (SSCC) for the GBCF when the sources and the noises are independent, i.e., $\rho_s = 0$ and $\rho_z = 0$. While the JSCC-OL scheme operates in the finite horizon regime, the SSCC scheme applies coding over blocks of source-pair samples, and uses asymptotically long channel codes. Clearly, by coding over multiple samples of source pairs one can obtain MSEs which are at least as low as the MSEs achieved by linear transmission schemes.

Consider a coding scheme which requires (on average) K channel uses to send m samples of source pairs in order to achieve a target MSE D. We define the *source-channel bandwidth ratio* of this scheme as $\kappa \triangleq K/m$.

As the JSCC-OL scheme applies uncoded transmission, then $m = 1$, and its source-channel bandwidth ratio is given by $\kappa_{\mathrm{OL}} = K_{\mathrm{OL}}$. It was shown in [40, Thm. 2] that for the symmetric setting with $\rho_s = 0$, SSCC is optimal. Let κ_{sep} denote the source-channel bandwidth ratio of the optimal SSCC, which applies the optimal source compression followed by a capacity achieving channel code. Since $\rho_s = 0$, the optimal source code compresses each of the Gaussian sources separately via an optimal rate-distortion code [41, Thm. 13.3.2], resulting in two independent messages. As the optimal channel code for the GBCF is not known, in the following we consider upper and lower bounds on κ_{sep}.

A lower bound on κ_{sep} is obtained by using the upper bound on the symmetric achievable rate for the GBCF, stated in [8, pg. 671], i.e., by letting one of the receivers have access to both channel outputs. Applying a simple manipulation to the results in [8, pg. 671] we obtain that if R is a symmetric achievable rate for the GBCF, and $\rho_z = 0$, then:

$$\begin{aligned} R &< \frac{1}{2} \log\left(\sqrt{\frac{9}{4} + \frac{2P}{\sigma_z^2}} - \frac{1}{2} \right) \\ &\overset{(a)}{=} \frac{1}{2} \log\left(\sqrt{\frac{9}{4} + 2\mathrm{SNR}} - \frac{1}{2} \right), \end{aligned} \quad (15)$$

where in (a) we set $\mathrm{SNR} \triangleq \frac{P}{\sigma_z^2}$. For completeness, this analysis is provided in Appendix A-C. The lower bound on κ_{sep} is thus given by:

$$\kappa_{\mathrm{sep}} \geq \frac{\log\left(\frac{\sigma_s^2}{D}\right)}{\log\left(\sqrt{\frac{9}{4} + 2\mathrm{SNR}} - \frac{1}{2}\right)} \triangleq \kappa_{\mathrm{sep}}^{\mathrm{lb}}.$$

An upper bound on κ_{sep} is obtained by using the LQG channel code of [9], which is the best known channel code for the GBCF:

$$\kappa_{\mathrm{sep}} \leq \frac{\log\left(\frac{\sigma_s^2}{D}\right)}{2\log|a_1|} \triangleq \kappa_{\mathrm{sep}}^{\mathrm{ub}},$$

where a_1 is defined in [9, Eq. (14)]. A detailed description of the LQG scheme is provided in the following Section IV-A. Recall that for $\rho_z = 0$ the LQG code of [9] is the optimal

linear channel coding scheme in the sense of maximal sum-rate [21], which motivates focusing on $\rho_z = 0$. In the following proposition we upper bound the terms $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}}$ and $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{lb}}$:

Proposition 1. In the symmetric setting with $\rho_s = \rho_z = 0$, $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}}$ and $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{lb}}$ are upper bounded by:

$$K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}} \leq \left\lceil 2 \log \left(\frac{\sigma_s^2}{D} \right) \right\rceil, \tag{16a}$$

$$K_{\text{OL}} - \kappa_{\text{sep}}^{\text{lb}} \leq \left\lceil \left(2 + \frac{2}{\text{SNR}} - \frac{1}{\sqrt{2\text{SNR}}} \right) \log \left(\frac{\sigma_s^2}{D} \right) \right\rceil. \tag{16b}$$

Proof. The proof of (16a) is detailed in Appendix A-D. The proof of (16b) is detailed in Appendix A-E. □

We note that the right-hand side (RHS) of (16a) is independent of the SNR, while both K_{OL} and $\kappa_{\text{sep}}^{\text{ub}}$ increase when SNR decreases. Therefore, (16a) implies that for low enough SNR, $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}} \ll K_{\text{OL}}$, which implies that the gap becomes negligible compared to K_{OL} and $\kappa_{\text{sep}}^{\text{ub}}$ for sufficiently low SNR. For instance, letting $\sigma_s^2 = 1$ and $D = 10^{-2}$, we obtain $\left\lceil 2 \log \frac{\sigma_s^2}{D} \right\rceil = 10$. For $P = 0.001$ and $\sigma_z^2 = 1$, we have $\kappa_{\text{sep}}^{\text{ub}} = 9213$ and therefore $K_{\text{OL}} \leq 9223$; thus, $(K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}})/K_{\text{OL}} \approx 10^{-3}$. An explicit calculation of K_{OL} via (11)–(12) results in $K_{\text{OL}} = 9213$, thus, for this setting $K_{\text{OL}} = \kappa_{\text{sep}}^{\text{ub}}$, indeed demonstrating that the gap in (16a) is negligible compared to K_{OL} and $\kappa_{\text{sep}}^{\text{ub}}$. It should be noted that $\kappa_{\text{sep}}^{\text{ub}}$ is achieved by applying source and channel coding with an *asymptotically large blocklength*. In particular, in contrast to the JSCC-OL scheme, coding takes place over multiple samples of source pairs, and for $\kappa_{\text{sep}}^{\text{ub}}$ to be approached *infinitely many channel uses* are required, which results in a large delay and a high complexity. On the other hand, the JSCC-OL scheme uses a *finite* number of channel symbols for the transmission of a *single* source pair. In spite of this fundamental difference, Prop. 1 shows that in the low SNR regime the performance loss of the JSCC-OL scheme compared to the SSCC scheme is negligible.

We also note that in the low SNR regime the RHS of (16b) is approximately given by $\left\lceil \log \left(\frac{\sigma_s^2}{D} \right) \frac{2}{\text{SNR}} \right\rceil$. While the RHS of (16b) only constitutes an upper bound, simulation results indicate that indeed, in the low SNR regime, $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{lb}}$ increases proportionally to $\frac{1}{\text{SNR}}$. Based on the simulation results, we conjecture that this negative result is due to the fact that the upper bound in (15) based on [8, pg. 671] is not tight.

Next, we discuss a control theoretic approach for the problem of transmitting correlated Gaussian sources over the GBCF.

IV. JSCC Based on the LQG Scheme

It was observed in [20] and [42] that there is a natural duality between the problem of FB stabilization and communications over PtP Gaussian channels with FB. Based on this duality, results and tools from control theory were used to design channel codes for PtP Gaussian channels with FB. This duality was also exploited to construct channel codes for multiuser Gaussian channels with FB: In [20] a duality between communications over the GBCF with unit-variance independent noise components, and a FB stabilization problem was established; yet, [20] did not present an explicit FB communications scheme. In [9], a scheme which belongs to the class of schemes analyzed in [20] was presented. This scheme, referred to as the LQG scheme, also supports the communications for GBCFs with correlated noise components.

The LQG scheme achieves the best known information *rates* over the GBCF. Furthermore, [9, Lemma 1] characterizes a linear relationship between the achievable rates and the achievable MSE exponents, which correspond to the slope of decay of the logarithm of the MSE for sufficiently large number of channel uses. Therefore, for sufficiently large number of channel uses, higher achievable rates correspond to higher rate of decay of the MSE. This implies that for large enough number of channel uses (or low enough MSEs), the LQG scheme is preferable compared to schemes which achieve lower rates. This, together with the time-invariant property of the LQG scheme, motivates studying the JSCC-LQG scheme in the finite horizon regime.

We emphasize that the original LQG scheme in [9] is designed to optimize the *infinite* horizon channel coding rate, while in this work we focus on *JSCC* in the *finite* horizon regime. These two differences give rise to two challenging research problems: The first problem is the characterization of the minimal number of transmissions required by the JSCC-LQG scheme to achieve a target MSE pair given a specific initialization, and the second is the optimization of the initial transmission of the JSCC-LQG scheme. In this section we address these two research problems. The main difficulty in designing a time-varying finite horizon JSCC-LQG scheme for GBCFs, follows since constraining the controller as in (5) results in minimization problems which do not have explicit solutions. Furthermore, typically, an LQG scheme is designed for a specified LTI system [10, Ch. 4.1], and it is not clear how to specify an LTI system to achieve a target finite horizon performance. Therefore, our approach in this section is to adapt the LQG scheme of [9] to the transmission of Gaussian sources over GBCFs and analyze the finite horizon behavior of the adapted schemes. We investigate how to initialize the new JSCC-LQG scheme in order to minimize the number of channel symbols required to achieve a target MSE pair. For the symmetric setting, we introduce a new initialization by scaling the transmitted signals, which results in a significantly better finite horizon performance *without* degrading the infinite horizon performance. In particular, we show that the new proposed scaling technique: 1) Optimally exploits the available transmission power subject to the per-symbol average power constraint in (5), 2) Minimizes the distance between the covariance matrix of the JSCC-LQG initial state and the covariance matrix of its steady-state (subject to the per-symbol average power constraint in (5)), and 3) Achieves the *same MSE exponents* as in [9]. Thus, the JSCC-LQG scheme derived in this work is time-invariant, achieves the best known infinite horizon performance, and achieves a very good performance in the finite horizon regime.

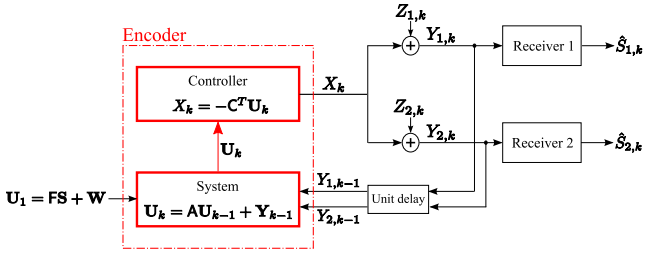


Fig. 3: Control system modeling of transmission over the GBCF. The states of the system are denoted by $\mathbf{U}_k = [U_{1,k}, U_{1,k}]^T$. The controller generates a scalar signal X_k, and the noisy channel outputs $Y_{1,k}$ and $Y_{2,k}$, represented as the vector $\mathbf{Y}_k = [Y_{1,k}, Y_{1,k}]^T$, are fed back to the system with a unit delay. $(\hat{S}_{1,k}, \hat{S}_{2,k})$ are the reconstructions of (S_1, S_2) after the k'th channel use.

A. The LQG scheme for JSCC

In this section we investigate the application of the LQG scheme of [9] to the transmission of a pair of Gaussian sources over the GBCF. The JSCC-LQG scheme is derived by mapping the FB control problem into a linear code for the GBCF. The asymptotic performance of this scheme is determined by the eigenvalues of the open-loop matrix of a linear system with unit memory representing the encoder. These eigenvalues are determined by the minimal power required to stabilize the system using the FB. In the finite horizon regime, the JSCC-LQG scheme is terminated after K channel uses when the target MSE pair is met.

Consider a two-dimensional unstable dynamical system, depicted in Fig. 3, which is stabilized by a controller observing the entire system state vector, $\mathbf{U}_k = [U_{1,k}, U_{2,k}]^T$. At time k the controller outputs a scalar signal X_k, which is received after being corrupted by additive Gaussian noises. Recall that $\mathbf{Y}_k$ is the noisy received control signal at the output of the channel at time k, stated in (2), and let $\mathsf{A} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$ with $a_i \in \Re, a_1 \neq a_2, |a_i| > 1$.[7] Recall (13) in which F is a deterministic matrix, and $\mathbf{W} \sim \mathcal{N}(0, \mathsf{Q}_w)$ independent of $\mathbf{S}$. The source encoding is incorporated into the control problem by generating the initial as be a function of the source samples:

$$\mathbf{U}_1 = \mathsf{F}\mathbf{S} + \mathbf{W}. \tag{17}$$

The system state vector at time k, $\mathbf{U}_k$, recursively evolves via:

$$\mathbf{U}_k = \mathsf{A}\mathbf{U}_{k-1} + \mathbf{Y}_{k-1}, \quad k = 2, 3, \ldots, K. \tag{18}$$

***Encoding*:** In the communications problem that correspond to the control problem, the encoder is the combination of the system stated in (18) and the controller, see Fig. 3. At each time index, the encoder recursively computes $\mathbf{U}_k$, and transmits X_k obtained from $\mathbf{U}_k$ using the linear controller presented in [9, Lemma 4]: $X_k = -\mathbf{C}^T\mathbf{U}_k$, where $\mathbf{C} = [c_1, c_2]^T$. The vector $\mathbf{C}$ is given by $\mathbf{C} = (\mathbf{B}^T\mathsf{G}\mathbf{B} + 1)^{-1}\mathsf{A}\mathsf{G}^T\mathbf{B}$, where $\mathbf{B}$ is defined above (2), and G is the unique positive-definite solution of the discrete algebraic Riccati equation (DARE) [9, Eq. (22)]:

$$\mathsf{G} = \mathsf{A}^T\mathsf{G}\mathsf{A} - \mathsf{A}^T\mathsf{G}\mathbf{B}(\mathbf{B}^T\mathsf{G}\mathbf{B} + 1)^{-1}\mathbf{B}^T\mathsf{G}\mathsf{A}, \tag{19}$$

[7]Note that if $a_1 = a_2$ then the pair $(\mathsf{A}, \mathbf{B})$ is not controllable, see [10, Def. 4.1.1 and Prop. 4.4.1].

such that the magnitudes of both eigenvalues of the matrix $\mathsf{A} - \mathbf{B}\mathbf{C}^T$ are smaller than 1. It follows from [9, Lemma 4] that, as $k \to \infty$, the covariance matrix of $\mathbf{U}_k$, $\mathsf{Q}_{u,k}$, converges to the solution of the discrete algebraic Lyapunov equation [9, Eq. (23)]:

$$\mathsf{Q}_u = (\mathsf{A} - \mathbf{B}\mathbf{C}^T)\mathsf{Q}_u(\mathsf{A} - \mathbf{B}\mathbf{C}^T)^T + \mathsf{Q}_z, \tag{20}$$

where the solution of (20) is restricted to be a positive semidefinite matrix. Finally, in [9, Lemma 4] the matrix A is obtained from the minimum asymptotic average power via: $P(\mathsf{A}, \mathsf{Q}_z) = \mathbf{C}^T\mathsf{Q}_u\mathbf{C} = \mathrm{trace}(\mathsf{G}\mathsf{Q}_z)$, see [9, Eq. (24)].

The JSCC-LQG scheme described above can be readily stated within the class of linear and memoryless schemes defined in Subsection II-C: The encoder state update for the JSCC-LQG scheme stated in (18) exactly follows the relationship in (7), while the transmitted signal is given by $X_k = \mathbf{T}_k^T\mathbf{U}_k$ with $\mathbf{T}_k^T = -\mathbf{C}^T$.

***Decoding*:** The work [9] considered $\mathbf{U}_1$ generated with $\mathsf{F} = \mathsf{I}$ and $\mathbf{W} = \mathbf{0}$ in (17), and used the so-called "zero trajectory" (ZT) detector. This detector recursively estimates $U_{i,k}$ via [9, Eq. (18)]:

$$\hat{U}_{i,1} = 0, \quad \hat{U}_{i,k} = a_i\hat{U}_{i,k-1} + Y_{i,k-1}, \tag{21}$$

for $k = 2, 3, \ldots, K+1$. Then, it estimates S_i from $\hat{U}_{i,k+1}$ via [9, Subsec. IV.A]:

$$\hat{S}_{i,k} = -a_i^{-k}\hat{U}_{i,k+1}, \tag{22}$$

which results in the MSE [9, proof of Lemma 3]:

$$\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\} = a_i^{-2k}\mathbb{E}\{U_{i,k+1}^2\}. \tag{23}$$

Remark 6. Note that in contrast to the JSCC-OL scheme, in the JSCC-LQG scheme the encoder and the decoders are decoupled. More precisely, in the JSCC-OL scheme the transmitted signal at time k is a linear combination of the estimation errors at time $k-1$ On the other hand, in the JSCC-LQG scheme the transmitted signal at time k, X_k, depends only on $\mathbf{U}_k$, and is not a function of $\hat{U}_{i,k}$ or $\hat{S}_{i,k}$.

B. Initialization of the JSCC-LQG Scheme

Similarly to Subsection III-B, optimizing over all deterministic matrices F and over all correlation matrices Q_w in (17) is computationally very intensive. An alternative approach is to select F and $\mathbf{W}$, such that $\mathsf{Q}_{u,1}$, the covariance matrix of $\mathbf{U}_1$, will be equal to Q_u, which is the solution of (20). The motivation for this approach is two-fold:

1) The LQG has the best known infinite horizon performance, i.e, MSE exponent, thus, it is preferable that the system will achieve this MSE exponent for *every* k.
2) When $\mathsf{Q}_{u,1} = \mathsf{Q}_u$ then $P_k = P, \forall k$. Therefore, this initialization leads to optimal utilization of the available transmission power.

To rigorously analyze the impact of JSCC-LQG initialization via (17) on the finite horizon performance, we define a distance between two correlation matrices, e.g., $\mathsf{Q}_{u,1}$ and Q_u, by $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1}) = ||\mathsf{Q}_u - \mathsf{Q}_{u,1}||_{\mathrm{F}}$, where $||\mathsf{Q}||_{\mathrm{F}} =$

$\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2}([\mathsf{Q}]_{i,j})^2}$ is the Frobenius matrix norm, see, e.g., [7, Sec. IV.A].

Note that the initialization objectives in the finite horizon regime for the JSCC-LQG and for the JSCC-OL schemes are fundamentally different: The JSCC-OL initialization aims at increasing $|\rho_0|$, while the JSCC-LQG initialization aims at minimizing $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$ via properly selecting F and $\mathbf{W}$. In fact, in some cases, e.g., when $\rho^* < |\rho_s|$, increasing $|\rho_0|$ increases the distance between the initial and steady states.

In the next two subsections we consider the special case of initialization with $\mathsf{F} = \mathsf{I}$ and $\mathbf{W} = \mathbf{0}$: In Subsection IV-C we present a new MMSE decoder which achieves MSE pairs lower than the MSE pairs achieved by the ZT decoder presented in [9], and in Subsection IV-D we analyze the finite horizon performance of the JSCC-LQG scheme with this initialization in the general setting. The JSCC-LQG with the general initialization $\mathbf{U}_1 = \mathsf{F}\mathbf{S} + \mathbf{W}$ is studied in Subsection IV-E, in which we focus on the symmetric setting.

C. An Improved JSCC-LQG Decoder

In this subsection we consider initialization with $\mathsf{F} = \mathsf{I}$ and $\mathbf{W} = \mathbf{0}$. Note that the decoding rule (22) is not necessarily optimal in the instantaneous MMSE sense. Let $\mathsf{M} \triangleq \mathsf{A} - \mathbf{B}\mathbf{C}^T$ denote the closed-loop matrix, and let $\mathsf{Q}_{u,k} \triangleq \mathbb{E}\left\{\mathbf{U}_k\mathbf{U}_k^T\right\}$ denote the state covariance matrix at time k, with $\mathsf{Q}_{u,1} = \mathsf{Q}_s$. In (B.1) we show that the closed-loop dynamics of the system is given by:

$$\mathbf{U}_k = (\mathsf{A} - \mathbf{B}\mathbf{C}^T)\mathbf{U}_{k-1} + \mathbf{Z}_{k-1},$$

while in (B.7) we show that $\mathsf{Q}_{u,k}$ is given by:

$$\mathbb{E}\left\{\mathbf{U}_k\mathbf{U}_k^T\right\} = \mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1} + \sum_{l=0}^{k-2}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l.$$

The MMSE estimator of S_i, based on the observation $\hat{U}_{i,k+1}$ in (21) is stated in the following theorem:

Theorem 2. The MMSE estimator of $S_i, i = 1, 2$, at time k, based on the observation $\hat{U}_{i,k+1}$ computed via (21), is:

$$\hat{S}_{i,k} = \frac{[\mathsf{M}^k\mathsf{Q}_s]_{i,i} - \sigma_i^2 a_i^k}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2 a_i^{2k}}\hat{U}_{i,k+1}. \quad (24)$$

Furthermore, the MSE of $\hat{S}_{i,k}$ is given by:

$$\mathbb{E}\left\{(S_i - \hat{S}_{i,k})^2\right\} = \frac{\sigma_i^2[\mathsf{Q}_{u,k+1}]_{i,i} - \left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2 a_i^{2k}}, \quad (25)$$

and as $k \to \infty$ the MSE expression in (25) coincides with the MSE of the decoder in (22).

Proof. The proof is provided in Appendix B-A. □

Remark 7. Since the estimator in (24) is the optimal estimator of S_i based on the observation $\hat{U}_{i,k+1}$, it clearly achieves an MSE value smaller than or equal to that achieved by the estimator in (22). In particular, (24) outperforms (22) in the finite horizon regime, i.e., for large MSEs, see Fig. 6.

D. Finite Horizon Analysis of JSCC-LQG

Next, we study the JSCC-LQG scheme with the decoder (24) in the finite horizon regime. We begin with the average instantaneous transmission power which we denote by P_k. In contrast to the JSCC-OL scheme in which $P_k = P, \forall k$, in the JSCC-LQG scheme P_k varies with k. While the LQG theory implies that $P_k \to P$ asymptotically as $k \to \infty$, it does not constrain P_k for any finite k, hence P_k may be larger than P, thus, violating the per-symbol average power constraint in (5). This implies that for specific P, σ_1^2 and σ_2^2, there are pairs of sources which cannot be transmitted using the JSCC-LQG scheme with the initialization $\mathbf{U}_1 = \mathbf{S}$, i.e., setting $\mathsf{F} = \mathsf{I}$ and $\mathbf{W} = \mathbf{0}$ in (17). In the following subsection, we present a sufficient condition under which the JSCC-LQG scheme, initialized with $\mathbf{U}_1 = \mathbf{S}$, satisfies (5). For the symmetric setting we use the same approach to find a necessary and sufficient condition, see Subsection IV-E.

1) A Sufficient Condition for Satisfying the Average Per-Symbol Power Constraint: Let $[\lambda_1, \lambda_2]^T$ denote the eigenvalues of the closed-loop matrix M, and let $\mathsf{V} = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$ be a matrix whose columns are the corresponding eigenvectors of M. Recall that $\mathbf{C} = [c_1, c_2]^T$ and for $\varsigma_1, \varsigma_2 \geq 0$ and $\rho \in (-1, 1)$, define:

$$\omega_1(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{c_1(\varsigma_1 v_1 v_4 - \rho\varsigma_2 v_1 v_2) + c_2(\varsigma_1 v_3 v_4 - \rho\varsigma_2 v_2 v_3)}{\det(\mathsf{V})} \quad (26a)$$

$$\omega_2(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{c_1(\rho\varsigma_2 v_1 v_2 - \varsigma_1 v_2 v_3) + c_2(\rho\varsigma_2 v_1 v_4 - \varsigma_1 v_3 v_4)}{\det(\mathsf{V})} \quad (26b)$$

$$\omega_3(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{-\varsigma_2\sqrt{1-\rho^2}(c_1 v_1 v_2 + c_2 v_2 v_3)}{\det(\mathsf{V})} \quad (26c)$$

$$\omega_4(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{\varsigma_2\sqrt{1-\rho^2}(c_1 v_1 v_2 + c_2 v_1 v_4)}{\det(\mathsf{V})}. \quad (26d)$$

Further define:

$$\alpha_1(\varsigma_1, \varsigma_2, \rho) \triangleq \omega_1^2(\varsigma_1, \varsigma_2, \rho) + \omega_3^2(\varsigma_1, \varsigma_2, \rho) \quad (27a)$$

$$\alpha_2(\varsigma_1, \varsigma_2, \rho) \triangleq \omega_2^2(\varsigma_1, \varsigma_2, \rho) + \omega_4^2(\varsigma_1, \varsigma_2, \rho) \quad (27b)$$

$$\alpha_3(\varsigma_1, \varsigma_2, \rho) \triangleq 2\omega_1(\varsigma_1, \varsigma_2, \rho)\omega_2(\varsigma_1, \varsigma_2, \rho) + 2\omega_3(\varsigma_1, \varsigma_2, \rho)\omega_4(\varsigma_1, \varsigma_2, \rho), \quad (27c)$$

and finally define:

$$\eta_1(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{\alpha_1(\varsigma_1, \varsigma_2, \rho)}{1 - \lambda_1^2} \quad (28a)$$

$$\eta_2(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{\alpha_2(\varsigma_1, \varsigma_2, \rho)}{1 - \lambda_2^2} \quad (28b)$$

$$\eta_3(\varsigma_1, \varsigma_2, \rho) \triangleq \frac{\alpha_3(\varsigma_1, \varsigma_2, \rho)}{1 - \lambda_1\lambda_2}. \quad (28c)$$

The following proposition characterizes source pairs for which the per-symbol average power constraint in (5) is satisfied when the JSCC-LQG scheme is used:

$$K_{\text{LQG}}^{\text{ub}} = \left\lceil \max\left\{ \frac{\left[\log\left(\frac{\vartheta_1}{D_1}\right)\right]^+}{2\log|a_1|}, \frac{\left[\log\left(\frac{\vartheta_2}{D_2}\right)\right]^+}{2\log|a_2|} \right\} \right\rceil \tag{30a}$$

$$K_{\text{LQG}}^{\text{lb}} = \left\lceil \max\left\{ \frac{\left[\log\left(\frac{\left[\sigma_1^2\sigma_{z,1}^2-\beta_1^2-D_1\sigma_{z,1}^2\right]^+}{(2\beta_1+\sigma_1^2)D_1}\right)\right]^+}{2\log|a_1|}, \frac{\left[\log\left(\frac{\left[\sigma_2^2\sigma_{z,2}^2-\beta_2^2-D_2\sigma_{z,2}^2\right]^+}{(2\beta_2+\sigma_2^2)D_2}\right)\right]^+}{2\log|a_2|} \right\} \right\rceil. \tag{30b}$$

Proposition 2. If the following condition holds for every $k = 1, 2, 3, \ldots$:

$$\begin{aligned}\lambda_1^{2(k-1)}\big(\alpha_1(\sigma_1,\sigma_2,\rho_s) - \eta_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\big) \\ + \lambda_2^{2(k-1)}\big(\alpha_2(\sigma_1,\sigma_2,\rho_s) - \eta_2(\sigma_{z,1},\sigma_{z,2},\rho_z)\big) \\ + (\lambda_1\lambda_2)^{k-1}\big(\alpha_3(\sigma_1,\sigma_2,\rho_s) - \eta_3(\sigma_{z,1},\sigma_{z,2},\rho_z)\big) \le 0,\end{aligned} \tag{29}$$

then the JSCC-LQG scheme satisfies the per-symbol average power constraint in (5).

Proof. The proof is provided in Appendix B-B. □

Remark 8. Note that the sufficient condition in Prop. 2 is implicit. Yet, Prop. 2 can be used to formulate explicit sufficient conditions (on the sources) for the JSCC-LQG scheme to satisfy the per-symbol average power constraint in (5). For example, if $\alpha_j(\sigma_1,\sigma_2,\rho_s) < \eta_j(\sigma_{z,1},\sigma_{z,2},\rho_z), j = 1,2,3$, and $\text{sgn}(\lambda_1\lambda_2) = 1$, then $P_k \le P, \forall k$.

2) Analysis of the Termination Time: Let K_{LQG} denote the minimal number of channel uses required to achieve an average MSE pair (D_1, D_2) with the JSCC-LQG scheme using the decoder (24). In this subsection we present upper and lower bounds on K_{LQG}. An *explicit* characterization of K_{LQG} for the symmetric setting is provided in Thm. 5, see Subsection IV-E3.[8]

We begin with the following definitions:

$$\begin{aligned}
\tau_1 &\triangleq \frac{\sigma_1\left(|v_1v_4\lambda_1| + |v_2v_3\lambda_2|\right) + |\rho_s\sigma_2v_1v_2|\left(|\lambda_2| + |\lambda_1|\right)}{|\det(\mathsf{V})|}\\
\tau_2 &\triangleq \frac{\sigma_2\sqrt{1-\rho_s^2}\left(|v_1v_2|\left(|\lambda_2| + |\lambda_1|\right)\right)}{|\det(\mathsf{V})|}\\
\tau_3 &\triangleq \frac{|\sigma_1v_3v_4|\left(|\lambda_1| + |\lambda_2|\right) + |\rho_s\sigma_2|\left(|v_1v_4\lambda_2| + |v_2v_3\lambda_1|\right)}{|\det(\mathsf{V})|}\\
\tau_4 &\triangleq \frac{\sigma_2\sqrt{1-\rho_s^2}\left(|v_1v_4\lambda_2| + |v_2v_3\lambda_1|\right)}{|\det(\mathsf{V})|}\\
\vartheta_1 &\triangleq \tau_1^2 + \tau_2^2 + [Q_u]_{1,1}\\
\vartheta_2 &\triangleq \tau_3^2 + \tau_4^2 + [Q_u]_{2,2}\\
\beta_1 &\triangleq \frac{\sigma_1^2\left(|v_1v_4\lambda_1| + |v_2v_3\lambda_2|\right) + |\rho_s\sigma_1\sigma_2v_1v_2|\left(|\lambda_2| + |\lambda_1|\right)}{|\det(\mathsf{V})|}\\
\beta_2 &\triangleq \frac{\sigma_2^2\left(|v_1v_2\lambda_2| + |v_2v_3\lambda_1|\right) + |\rho_s\sigma_1\sigma_2v_3v_4|\left(|\lambda_2| + |\lambda_1|\right)}{|\det(\mathsf{V})|},
\end{aligned}$$

[8]Note that using the approach of Thm. 5 for the general setting results only in an *implicit* characterization of K_{LQG}.

where Q_u is the unique positive semidefinite solution of (20). The following theorem states upper and lower bounds on K_{LQG}:

Theorem 3. The JSCC-LQG scheme with the MMSE decoder in (24) and target MSE values D_1 and D_2 terminates within time $K_{\text{LQG}}^{\text{lb}} \le K_{\text{LQG}} \le K_{\text{LQG}}^{\text{ub}}$, where $K_{\text{LQG}}^{\text{ub}}$ and $K_{\text{LQG}}^{\text{lb}}$ are given in (30) at the top if the page.

Proof. The proof is provided in Appendix B-C. □

E. Finite Horizon Analysis of JSCC-LQG for the Symmetric GBCF

In this subsection we study the JSCC-LQG scheme in the symmetric setting, for different combinations of initialization parameters F and $\mathbf{W}$, in (17). We first consider initialization based on the assignment $\mathsf{F} = \sqrt{\gamma}\cdot\mathsf{I}, \gamma > 0$, and $\mathbf{W} = \mathbf{0}$, for which we explicitly derive the value of γ, which minimizes the MSE subject to the per-symbol average power constraint (5) for *all* time indices k. We show that this optimal γ also minimizes the distance $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$ among all scaling coefficients which satisfy (5).

Next, we consider the general initialization framework of (17). We show that when $\mathbf{W} \neq 0$, or when the off-diagonal elements of F are non-zero, then JSCC-LQG has an MSE floor. A numerical comparison of the different initialization methods indicates that the lowest MSEs are achieved when the optimal scaling $\mathbf{U}_1 = \sqrt{\gamma}\mathbf{S}$ is applied.

Lastly, for $\mathbf{U}_1 = \sqrt{\gamma}\mathbf{S}$, we present an explicit characterization of K_{LQG} in terms of the roots of a quadratic polynomial. Note that since $a_1 \neq a_2$ (see Subsection IV-A) in the symmetric setting we have $a_1 = -a_2$, and the components of the eigenvectors of the matrix M satisfy $v_1 = v_4$, and $v_2 = v_3$.

1) Initialization to Satisfy the Per-Symbol Average Power Constraint: In this subsection we study initialization based on the parameters $\mathsf{F} = \sqrt{\gamma}\cdot\mathsf{I}$ and $\mathbf{W} = \mathbf{0}$. To find the scaling value γ which minimizes the MSE subject to the per symbol average power constraint (5), we first derive necessary and sufficient conditions for (5) to be satisfied for the JSCC-LQG scheme when $\gamma = 1$. Then, from these conditions, we find the maximal γ for which (5) is satisfied for *all* time indices k. Finally, we show that this maximal γ simultaneously minimizes the MSE and the distance $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$, for $\mathbf{U}_1 = \sqrt{\gamma}\mathbf{S}$, subject to (5). We first define the following quantities:

$$\mu_0 = 2c_1^2\sigma_s^2(1-\rho_s) \tag{31a}$$

$$\mu_1 = \frac{2c_1^2\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{1-\lambda_1^4} \tag{31b}$$

$$\mu_2 = 2c_1^2\sigma_s^2(1+\rho_s)a_1^4 \tag{31c}$$

$$\mu_3 = \frac{2c_1^2\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{1-\lambda_1^4}. \tag{31d}$$

Necessary and sufficient conditions for the power constraint to be satisfied for the JSCC-LQG scheme, with the initialization in (18), are stated in the following theorem.

Theorem 4. In the symmetric GBCF, the JSCC-LQG scheme satisfies the per-symbol average power constraint (5) if and only if $\mu_0 \leq \mu_1$ and $\mu_2 \leq \mu_3$.

Proof outline. In Appendix C-A we show that:

$$P_k = \begin{cases} P + (\mu_0 - \mu_1)\lambda_1^{2(k-1)}, & k \text{ is odd}, \\ P + (\mu_2 - \mu_3)\lambda_1^{2(k-1)}, & k \text{ is even}. \end{cases} \tag{32}$$

Since $|\lambda_1| < 1$, it follows that (5) is satisfied if and only if $\mu_0 \leq \mu_1$ and $\mu_2 \leq \mu_3$. □

From Eqn. (32) and from the fact that $|\lambda_1| < 1$, it follows that if (5) is satisfied for some odd k, then it is satisfied for *every* odd k. The same observation holds for even values of k. Thus, using (32) we can characterize the range of γ for which (5) is satisfied. We further note that scaling the sources at the transmitter can be beneficial even if (5) is satisfied for the initialization $\mathbf{U}_1 = \mathbf{S}$. As we show next, by scaling the sources we obtain that P_k is equal to P in at least (approximately) half of the time indices. Consequently, the available transmission power is used more efficiently. In the following proposition we characterize the scaling factor which minimizes the MSE, for the decoder (24), while satisfying the constraint (5). Before stating the proposition we define ν to be:

$$\nu = \min\left\{\frac{\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{(1-\lambda_1^4)(1-\rho_s)}, \frac{\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{(1-\lambda_1^4)(1+\rho_s)a_1^4}\right\}, \tag{33}$$

and let $\mathbf{U}_k(\gamma)$ denote the system state vector at time index k, when $\mathbf{U}_1 = \sqrt{\gamma}\cdot\mathbf{S}$, for some $\gamma > 0$. In a similar manner we also define $\hat{U}_{i,k}(\gamma)$ and $\mathsf{Q}_{u,k}(\gamma)$.

Proposition 3. The optimal scaling factor, in the MMSE sense, is $\sqrt{\gamma} = \sqrt{\frac{\nu}{\sigma_s^2}}$. Furthermore, when $\mathbf{U}_1 = \sqrt{\gamma}\cdot\mathbf{S}$, the MMSE estimator of $S_i, i = 1, 2$, at time k, based on the observation $\hat{U}_{i,k+1}(\gamma)$ is:

$$\hat{S}_{i,k} = \frac{\sqrt{\gamma}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i} - \sigma_s^2 a_i^k\right)}{[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i} - 2\gamma a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \gamma\sigma_s^2 a_i^{2k}}\hat{U}_{i,k+1}(\gamma), \tag{34}$$

and the MSE of $\hat{S}_{i,k}$ is given by:

$$\mathbb{E}\left\{(S_i - \hat{S}_{i,k})^2\right\} = \frac{\sigma_s^2[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i} - \gamma\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i} - 2\gamma a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \gamma\sigma_s^2 a_i^{2k}}. \tag{35}$$

Proof outline. First, we show that (33) constitutes an upper bound on the variance of the sources transmitted via a JSCC-LQG scheme with the initialization $\mathbf{U}_1 = \mathbf{S}$, which also satisfies (5). Explicitly writing the conditions of Thm. 4, i.e., $\mu_0 \leq \mu_1$ and $\mu_2 \leq \mu_3$, we obtain:

$$\sigma_s^2(1-\rho_s) \leq \frac{\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{1-\lambda_1^4},$$

$$\sigma_s^2(1+\rho_s)a_1^4 \leq \frac{\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{1-\lambda_1^4}.$$

This implies that:

$$\sigma_s^2 \leq \min\left\{\frac{\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{(1-\lambda_1^4)(1-\rho_s)}, \frac{\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{(1-\lambda_1^4)(1+\rho_s)a_1^4}\right\}.$$

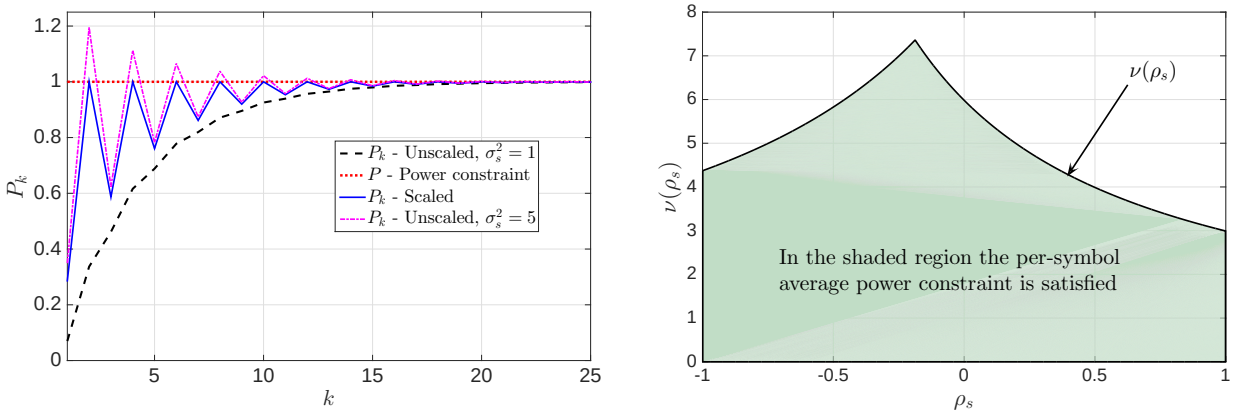
Therefore, the maximal possible scaling factor which satisfies (5) is $\sqrt{\frac{\nu}{\sigma_s^2}}$. In Appendix C-B we also derive the MMSE estimator for scaled transmission, stated in (34), and obtain its MSE, given by (35). Furthermore, we show that scaling by $\sqrt{\frac{\nu}{\sigma_s^2}}$ minimizes the MSE. The detailed proof is provided in Appendix C-B. □

Remark 9. As shown in the proof of Prop. 3, the MSE decreases when the scaling factor increases. Therefore, the optimal scaling factor is determined by the per-symbol average power constraint. This implies that when the optimal scaling factor is used, at least one of the following statements hold: 1) $P_k = P$ for every *odd* k, and $P_k \leq P$ for every *even* k; 2) $P_k = P$ for every *even* k, and $P_k \leq P$ for every *odd* k.

Next, we demonstrate the results of Thm. 4 and Prop. 3.

Example 2. Consider the transmission of a pair of Gaussian sources with variance σ_s^2 and correlation coefficient $\rho_s = 0.4$, over a GBCF with $\sigma_z^2 = 1.5$ and $\rho_z = 0.3$. We further set $P = 1$. Fig. (4a) depicts P_k vs. k for the JSCC-LQG scheme without scaling for $\sigma_s^2 = 1$ and $\sigma_s^2 = 5$, and for the JSCC-LQG scheme with optimal scaling factor specified by Prop. 3. As $\rho_s = 0.4$, it follows that the optimal scaling factor for $\sigma_s^2 = 1$ is $\sqrt{\gamma} = 2.0227$, while the optimal scaling factor for $\sigma_s^2 = 5$ is $\sqrt{\gamma} = 0.9046$. In both cases P_k is the same, illustrated by the blue solid line in Fig. (4a). It can be observed that both the non-scaled JSCC-LQG scheme with $\sigma_s^2 = 1$ and the scaled JSCC-LQG scheme satisfy (5); yet, the scaled scheme uses the available power more efficiently. On the other hand, when $\sigma_s^2 = 5$, then the non-scaled JSCC-LQG scheme violates the per-symbol average power constraint (5). It can further be observed that in the scaled scheme, $P_k = P$ for all even values of k, as stated in Remark 9. Finally, Fig. 4b illustrates ν, computed using (33), as a function of ρ_s. Following the result of Prop. 3, in order to maximize the MSE, one should use ν values that lie on the boundary of the shaded area in Fig. 4b.

The following proposition states that the scaling presented in Prop. 3 also minimizes the distance $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$ among all scaling coefficients which satisfy (5).



(a) (b)

Fig. 4: Satisfying the per-symbol average power constraint for $\sigma_s^2 = 1, \rho_s = 0.4, \rho_z = 0.3, \sigma_z^2 = 1.5$ and $P = 1$. (a) P_k vs. k for scaled and non-scaled JSCC-LQG schemes. (b) ν as a function of ρ_s, see (33).

Proposition 4. Let $\gamma > 0$ and write the solution of (20), Q_u, as $\mathsf{Q}_u = \sigma_u^2 \begin{bmatrix} 1 & \rho_u \\ \rho_u & 1 \end{bmatrix}$. Under the average per-symbol power constraint (5), the γ which minimizes the distance $\mathbb{D}(\mathsf{Q}_u, \gamma \mathsf{Q}_s)$ is given by $\gamma = \frac{\nu}{\sigma_s^2}$, where ν is given in (33). Moreover, the scaling which minimizes $\mathbb{D}(\mathsf{Q}_u, \gamma \mathsf{Q}_s)$, *regardless* of (5), is given by:

$$\gamma^* = \frac{\sigma_u^2(1+\rho_u\rho_s)}{\sigma_s^2(1+\rho_s^2)}.$$

Proof. The proof is provided in Appendix C-C. □

Clearly, if $\rho_s \neq \rho_u$, then $\mathbb{D}(\mathsf{Q}_u, \gamma \mathsf{Q}_s) > 0, \forall \gamma$. Therefore, in order to achieve a distance smaller than $\mathbb{D}(\mathsf{Q}_u, \frac{\nu}{\sigma_s^2}\mathsf{Q}_s)$, it is necessary to either add noise or rotate the sources, which is facilitated in the general initialization (17). In the next subsection we discuss two special instances of the general initialization.

2) Investigation of the General Initialization via (17)*:* We first consider initialization via scaling and noise addition:

Scaling and noise addition: Let $\mathsf{F} \triangleq \sqrt{\xi_0} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \xi_0 > 0$, and $\mathsf{Q}_w \triangleq \sigma_w^2 \begin{bmatrix} 1 & \rho_w \\ \rho_w & 1 \end{bmatrix}$. Recall that the objective of the initialization in the JSCC-LQG scheme is to minimize $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$. Since $\mathbf{S}$ and $\mathbf{W}$ are independent, $\mathsf{Q}_{u,1} = \mathsf{Q}_u$ is achieved by setting:

$$\mathsf{Q}_w = \begin{bmatrix} \sigma_u^2 - \xi_0\sigma_s^2 & \rho_u\sigma_u^2 - \xi_0\rho_s\sigma_s^2 \\ \rho_u\sigma_u^2 - \xi_0\rho_s\sigma_s^2 & \sigma_u^2 - \xi_0\sigma_s^2 \end{bmatrix}. \tag{36}$$

As Q_w is a covariance matrix, its eigenvalues must be non-negative. Thus, by explicitly expressing the eigenvalues using $\sigma_u^2, \rho_u, \sigma_s^2$, and ρ_s, we obtain that ξ_0 must lie in the following range:

$$0 < \xi_0 \leq \frac{\sigma_u^2}{\sigma_s^2} \min\left\{\frac{1+\rho_u}{1+\rho_s}, \frac{1-\rho_u}{1-\rho_s}\right\}.$$

From (36) it directly follows that increasing ξ_0 increases the component of the sources, i.e., FS, in $\mathbf{U}_1$, and decreases the noise component $\mathbf{W}$. As our objective is to convey $\mathbf{S}$, we set $\xi_0 = \frac{\sigma_u^2}{\sigma_s^2}\min\{\frac{1+\rho_u}{1+\rho_s}, \frac{1-\rho_u}{1-\rho_s}\}$. Letting $\xi_1 \triangleq \min\left\{\frac{1+\rho_u}{1+\rho_s}, \frac{1-\rho_u}{1-\rho_s}\right\} \leq 1$, the expression for Q_w in (36) can be written as:

$$\mathsf{Q}_w = \sigma_u^2 \begin{bmatrix} 1-\xi_1 & \rho_u - \xi_1\rho_s \\ \rho_u - \xi_1\rho_s & 1-\xi_1 \end{bmatrix}.$$

The MMSE estimator and its associated MSE, for the initialization $\mathbf{U}_1 = \sqrt{\xi_0} \cdot \mathbf{S} + \mathbf{W}$ can be found by following similar steps to those leading to (34)–(35). In particular, we let $\mathbf{U}_k(\mathbf{U}_1)$ denote the system state vector at time index k, for a given initial state $\mathbf{U}_1$, and let $\hat{\mathbf{U}}_k(\mathbf{U}_1)$ be the estimate of $\mathbf{U}_1$ at time k. The MMSE estimator for S_i at time k can be expressed as:

$$\hat{S}_{i,k} = \frac{\sqrt{\xi_0}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i} - \sigma_s^2 a_i^k\right)}{[\mathsf{Q}_{u,k+1}(\mathbf{U}_1)]_{i,i} - 2\xi_0 a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_u^2 a_i^{2k}} \hat{U}_{i,k+1}(\mathbf{U}_1),$$

and the achievable MSE is given by:

$$\mathbb{E}\left\{(S_i - \hat{S}_{i,k})^2\right\} = \frac{\sigma_s^2[\mathsf{Q}_{u,k+1}(\mathbf{U}_1)]_{i,i} - \xi_0\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2 + \sigma_s^2\sigma_w^2 a_i^{2k}}{[\mathsf{Q}_{u,k+1}(\mathbf{U}_1)]_{i,i} - 2\xi_0 a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_u^2 a_i^{2k}}.$$

Therefore, the MSE in the infinite horizon is given by:

$$\lim_{k\to\infty} \mathbb{E}\left\{(S_i - \hat{S}_{i,k})^2\right\} = \frac{\sigma_s^2\sigma_w^2}{\sigma_u^2} = \sigma_s^2(1-\xi_1).$$

Comparing the asymptotic MSE obtained with the initialization $\mathbf{U}_1 = \sqrt{\xi_0}\mathbf{S} + \mathbf{W}$ to the asymptotic MSE obtained with the initialization of Prop. 3 (for which $\lim_{k\to\infty}\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\} = 0$), we observe that adding the noise $\mathbf{W}$ results in an MSE floor. Note that while $1-\xi_1$ reflects the difference between the initial state and the JSCC-LQG steady state, it is independent of σ_s^2 and σ_u^2, and therefore it does not reflect the finite horizon performance of JSCC-LQG with the initialization $\mathbf{U}_1 = \sqrt{\xi_0}\mathbf{S} + \mathbf{W}$. Thus, $1-\xi_1$ can be interpreted only as a measure of the MSE floor due to the addition of the noise $\mathbf{W}$ in $\mathbf{U}_1$.

Extensive numerical study indicates that using noise addition to generate $\mathbf{U}_1$ results in higher MSEs than those achieved

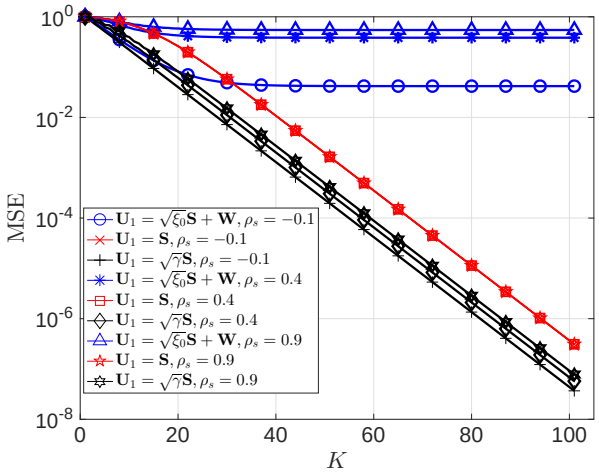


Fig. 5: MSE vs. number of channel uses for different values of ρ_s. $\sigma_s^2=1,\rho_z=0.3,\sigma_z^2=1$, and $P=0.5$.

with the optimal scaling presented in Prop. 3. On the other hand, this initialization can achieve a lower MSE compared to setting $\mathbf{U}_1=\mathbf{S}$.

Example 3. Fig. 5 depicts MSE vs. K for three different values of $\rho_s,\rho_s\in\{-0.1,0.4,0.9\}$, for three different initialization approaches: $\mathbf{U}_1=\sqrt{\xi_0}\mathbf{S}+\mathbf{W}$, $\mathbf{U}_1=\mathbf{S}$, and $\mathbf{U}_1=\sqrt{\gamma}\mathbf{S}$, where γ is the optimal scaling value specified in Prop. 3. Here, $\sigma_s^2=1,\rho_z=0.3,\sigma_z^2=1$, $P=0.5$, and Q_u is computed to be:

$$\mathsf{Q}_u\approx 8.77299\begin{bmatrix}1 & -0.1377\\ -0.1377 & 1\end{bmatrix}.$$

It can be observed that as ρ_s is closer to ρ_u, the MSE floor becomes smaller. It can further be observed that for high target MSEs, the initialization $\mathbf{U}_1=\sqrt{\xi_0}\mathbf{S}+\mathbf{W}$ is superior compared to $\mathbf{U}_1=\mathbf{S}$. On the other hand, for all three values of ρ_s, initialization via $\mathbf{U}_1=\sqrt{\gamma}\mathbf{S}$ achieves the lowest MSEs for all values of K. These performance gaps are also reflected in the distances $\mathbb{D}(\mathsf{Q}_u,\mathsf{Q}_{u,1})$ for the different initializations: First, note that for all considered values of ρ_s, $\mathbb{D}(\mathsf{Q}_u,\mathsf{Q}_s)\approx 11.25$, and, indeed, the respective MSE curves are indistinguishable. For $\rho_s=-0.1$, $\mathbb{D}(\mathsf{Q}_u,\gamma\mathsf{Q}_s)=0.735$, while for $\rho_s=0.9$, $\mathbb{D}(\mathsf{Q}_u,\gamma\mathsf{Q}_s)=9.582$, which explains the gap in performance in favor of the lower ρ_s. Finally, we note that while the distance in the case of $\mathbf{U}_1=\sqrt{\xi_0}\mathbf{S}+\mathbf{W}$ is zero by construction, the addition of noise increases the MSE which results in an inferior performance compared to initializing via $\mathbf{U}_1=\sqrt{\gamma}\mathbf{S}$.

Remark 10. When the encoder and the decoders share a common source of randomness [38], then the noise $\mathbf{W}$ can be removed, thus, eliminating the MSE floor observed in Example 3.

Next, we briefly discuss generating $\mathbf{U}_1$ from $\mathbf{S}$ via multiplication by a non-diagonal F.

Generating $\mathbf{U}_1$ via Multiplying $\mathbf{S}$ by F: Fixing $\mathbf{W}=\mathbf{0}$, the matrix F with which we achieve $\mathsf{Q}_{u,1}=\mathsf{Q}_u$ can be found by applying the Cholesky decomposition to Q_s and Q_u: $\mathsf{Q}_s=\mathsf{L}_s\mathsf{L}_s^T,\mathsf{Q}_u=\mathsf{L}_u\mathsf{L}_u^T$. The matrix F is given by $\mathsf{F}=\mathsf{L}_u\mathsf{L}_s^{-1}$. Since in this case F is lower triangular, $U_{1,1}$ is a scaled version of S_1, while $U_{2,1}$ is a linear combination of both S_1 and S_2. Now, recall that in the LQG scheme the state is first estimated via (21), and then, the sources are estimated from the estimated states via, for example, (34). With this decoding sequence, as $U_{1,1}$ is a scaled version of S_1, Rx_1 first estimates $U_{1,1}$, and then estimates S_1 from $U_{1,k}$. In fact, in this case S_1 enjoys the optimal rate of decrease of the MSE as the covariance matrix of the initial state is identical to the steady state covariance matrix. However, at Rx_2, since $U_{2,1}$ is a linear combination of S_1 and S_2 and since S_1 is not known and not required at Rx_2, the MSE for estimating S_2 from $U_{2,k}$ is generally higher than the MSE achieved in estimating S_2 out of $U_{2,k}$ generated when $U_{2,1}$ does not depend in S_1. Thus, setting $\mathsf{Q}_{u,1}=\mathsf{Q}_u$ via multiplying by F improves the performance at one receiver while degrading the performance at the other receiver.

Additionally, note that letting Rx_2 track (estimate) $U_{1,1}$, and then use it to estimate S_1 at Rx_2, results in a sub-optimal estimate of $U_{1,1}$ at Rx_2 since $U_{1,k}$ is updated based only on $Y_{1,k-1}$ and $U_{1,k-1}$. Therefore, as Rx_2 does not observe $Y_{1,k-1}$, it cannot efficiently track it.

3) An Explicit Characterization of the Termination Time K_{LQG}: Lastly, we explicitly characterize K_{LQG} for the scaled JSCC-LQG, i.e., with $\mathbf{U}_1=\sqrt{\gamma}\cdot\mathbf{S}$. We first define the following quantities:

$$\Phi(\varsigma,\rho)\triangleq\frac{\varsigma^2\left(\left(v_1^2+v_2^2-2\rho v_1v_2\right)^2+4(1-\rho^2)v_1^2v_2^2\right)}{\det^2(\mathsf{V})} \tag{37a}$$

$$\Psi_0\triangleq\frac{\sigma_z^2+\lambda_1^2\Phi(\sigma_z,\rho_z)}{1-\lambda_1^4}, \tag{37b}$$

$$\Psi_1\triangleq\frac{\Phi(\sigma_z,\rho_z)+\sigma_z^2\lambda_1^2}{1-\lambda_1^4}. \tag{37c}$$

$$\Gamma_s\triangleq\frac{\sigma_s^2(v_1^2+v_2^2-2\rho_s v_1v_2)}{v_1^2-v_2^2} \tag{37d}$$

$$\Upsilon_0\triangleq\Psi_0(D-\sigma_s^2)-D\gamma\sigma_s^2 \tag{37e}$$

$$\Upsilon_1\triangleq\Psi_0(\sigma_s^2-D)+2D\gamma\sigma_s^2 \tag{37f}$$

$$\Upsilon_2\triangleq(\gamma\Phi(\sigma_s,\rho_s)-\Psi_1)(\sigma_s^2-D)-\gamma\Gamma_s^2 \tag{37g}$$

$$\Upsilon_3\triangleq\Psi_0(\sigma_s^2-D)+2D\gamma\Gamma_s, \tag{37h}$$

Furthermore, let y be a positive integer, and define the functions $f^{(\mathrm{e})}(y)\triangleq 2\cdot\lceil\frac{y}{2}\rceil$, and $f^{(\mathrm{o})}(y)\triangleq 2\cdot\lceil\frac{y-1}{2}\rceil+1$.[9] K_{LQG} is explicitly characterized in the following theorem:

Theorem 5. Let $\left(x_1^{(\mathrm{e})},x_2^{(\mathrm{e})}\right)$ and $\left(x_1^{(\mathrm{o})},x_2^{(\mathrm{o})}\right)$ denote the roots of the polynomials $P^{(\mathrm{e})}(x)\triangleq\Upsilon_0x^2+\Upsilon_1x-D\gamma\sigma_s^2$, and $P^{(\mathrm{o})}(x)\triangleq\Upsilon_2x^2+\Upsilon_3x-D\gamma\sigma_s^2$, respectively. Furthermore, define:

$$x_0^{(\mathrm{e})}\triangleq\begin{cases}\min\{x_1^{(\mathrm{e})},x_2^{(\mathrm{e})}\}, & \frac{-\Upsilon_1^2}{4D\gamma\sigma_s^2}\le\Upsilon_0<0\\ a^{-4}, & \text{otherwise.}\end{cases}$$

[9] $f^{(\mathrm{e})}(y)$ is "round up to the nearest even integer", while $f^{(\mathrm{o})}(y)$ is "round up to the nearest odd integer".

and

$$x_0^{(\mathrm{o})} \triangleq \begin{cases} a_1^{-2}, & \Upsilon_2 < \frac{-\Upsilon_3^2}{4D_1\gamma\sigma_s^2}, \\ \min\left\{x_1^{(\mathrm{o})}, x_2^{(\mathrm{o})}\right\}, & \frac{-\Upsilon_3^2}{4D_1\gamma\sigma_s^2} \leq \Upsilon_2 < 0, \\ \frac{D\gamma\sigma_s^2}{\Upsilon_3}, & \Upsilon_2 = 0, \\ \max\left\{x_1^{(\mathrm{o})}, x_2^{(\mathrm{o})}\right\}, & \text{otherwise.} \end{cases}$$

Then, K_{LQG} is given by:

$$K_{\mathrm{LQG}} = \min\left\{ f^{(\mathrm{e})}\left(\left\lceil -\frac{\log x_0^{(\mathrm{e})}}{2\log|a_1|}\right\rceil\right), f^{(\mathrm{o})}\left(\left\lceil -\frac{\log x_0^{(\mathrm{o})}}{2\log|a_1|}\right\rceil\right)\right\}. \quad (38)$$

Proof outline. The detailed proof is provided in Appendix C-D. We first note that the result of Thm. 5 holds if γ is replaced by any arbitrary constant, regardless of whether (5) is satisfied or not. The proof consists of the following steps:

1) From (35) we conclude that the decoder terminates when:
$$\frac{\sigma_s^2[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i} - \gamma\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i} - 2\gamma a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \gamma\sigma_s^2 a_i^{2k}} \leq D. \quad (39)$$
Expressing $[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i}$ and $[\mathsf{M}^k\mathsf{Q}_s]_{i,i}$ in terms of $v_1, v_2, \lambda_1, \gamma$, and k, we observe that since $\lambda_1 = -\lambda_2$, then a different analysis needs to be applied for even and for odd values of k.
2) We let $x = a_1^{-2k}$, and recall that $\lambda_1 = \frac{1}{a_1}$ (see [42, Lemma 2.4]).
3) Using the definitions in step 2 we write (39) as a quadratic polynomial in x. As even values and odd values of k are analyzed separately, we use $P^{(\mathrm{e})}(x)$ to denote the quadratic polynomial for even values of k, and $P^{(\mathrm{o})}(x)$ to denote the quadratic polynomial for odd values of k.
4) For even values of k, we find the minimal k for which $P^{(\mathrm{e})}(x) \leq 0$ (for odd values of k we find the minimal k for which $P^{(\mathrm{o})}(x) \leq 0$). K_{LQG} is therefore a function of the roots of the polynomials $P^{(\mathrm{e})}(x)$ or $P^{(\mathrm{o})}(x)$.
5) Explicitly computing the roots of the polynomials $P^{(\mathrm{e})}(x)$ and $P^{(\mathrm{o})}(x)$ we obtain (38). □

Remark 11. Consider the expression for $x_0^{(\mathrm{o})}$. It can be observed that if $\Upsilon_2 \geq \frac{-\Upsilon_3^2}{4D\gamma\sigma_s^2}$ and $\Upsilon_2 \neq 0$, then $x_0^{(\mathrm{o})}$ is one of two real roots of $P^{(\mathrm{o})}(x)$, selected as follows: If $\frac{-\Upsilon_3^2}{4D\gamma\sigma_s^2} \leq \Upsilon_2 < 0$ then it is shown in Appendix C-D that $P^{(\mathrm{o})}(x)$ is concave with two positive roots: one smaller then 1 and one larger than 1. For this case, we choose the minimal root. On the other hand, if $0 < \Upsilon_2$ then it is shown in Appendix C-D that $P^{(\mathrm{o})}(x)$ is convex with one negative root and one positive root smaller than 1. For this case, we choose the maximal root. Note that when $\Upsilon_2 < \frac{-\Upsilon_3^2}{4D\gamma\sigma_s^2}$ then $P^{(\mathrm{o})}(x)$ is concave and does not have any real roots. This implies that the condition $P^{(\mathrm{o})}(x) \leq 0$ always holds, and the target MSE is obtained for every k. Therefore, in this case we set $x_0^{(\mathrm{o})} = a^{-2}$, which results in $K_{\mathrm{LQG}} = 1$. Finally, when $\Upsilon_2 = 0$, then $x_0^{(\mathrm{o})}$ is a solution of a simple linear equation. For $x_0^{(\mathrm{e})}$ we follow similar steps while noting that $\Upsilon_0 < 0$. Hence, for even values of k we only need to analyze the case of a concave polynomial.

Recall that both the JSCC-OL scheme considered in Section III and the JSCC-LQG scheme studied in this section are linear and memoryless JSCC transmission schemes, see Subsection II-C. In the next section, we use DP to formulate a linear and memoryless JSCC transmission scheme which outperforms both JSCC-OL and JSCC-LQG in the symmetric setting.

V. Linear and Memoryless JSCC Transmission Scheme Via Dynamic Programming

A fundamental difference between the JSCC-OL and the JSCC-LQG schemes discussed in Sections III and IV, receptively, is the fact that the first is time-varying while the second is time-invariant. Therefore, as stated in Subsection II-C, the JSCC-OL scheme can better exploit the available power and the correlation between the sources to achieve MSEs lower than JSCC-LQG in the finite horizon regime for some GBCF scenarios. On the other hand, the MSE exponent of the JSCC-LQG scheme is larger than the MSE exponent of the JSCC-OL scheme. Thus, for large enough number of channel uses, the JSCC-LQG scheme achieves MSE lower than JSCC-OL. As none of the two schemes, JSCC-OL and JSCC-LQG, dominates the other in the finite horizon regime, we utilize the the DP approach for solving finite horizon control problems [10, Sec. 4.1] to derive the JSCC-DP scheme which achieves MSE at least as low as the smallest MSE among the JSCC-OL and JSCC-LQG schemes, for any a-priori fixed *finite* number of channel uses.

A. Problem Formulation - Revisited

In this section we consider a complimentary problem to the one formulated in Section II: Let D_K denote the MSE after K channel uses. Our objective in this section is to find a linear and memoryless transmission scheme which, for a given K, achieves the minimal MSE at each receiver, denoted by $D_{K,\min}$.

In the following we adopt (most of) the notations used in Sections II and III, and denote the estimation error at Rx_i after k transmissions by $\epsilon_{i,k-1} = \hat{S}_{i,k-1} - S_i, i = 1, 2$. As we focus on linear and memoryless schemes, we let $\hat{\epsilon}_{i,k-1} = b_{i,k}Y_{i,k}$ be the estimator of $\epsilon_{i,k-1}$, and write $\epsilon_{i,k}$ as:

$$\epsilon_{i,k} = \left(\epsilon_{i,k-1} - b_{i,k}Y_{i,k}\right), \quad b_{i,k} \in \mathfrak{R}, \quad (40)$$

Similarly to (10): $\hat{S}_{i,k} = -\sum_{m=1}^{K}\hat{\epsilon}_{i,m-1}$. To simplify the analysis, we limit our focus to the symmetric setting, set $|b_{1,k}| = |b_{2,k}|$, and let $b_k \triangleq b_{1,k}$. Furthermore, following [7] we let $m_k \in \{1, -1\}$ be a modulation coefficient. We now have the following structure of X_k, the transmitted signal:

$$\begin{aligned} X_k &\overset{(a)}{=} d_{k-1}\left(\epsilon_{1,k-1} + m_{k-1}\epsilon_{2,k-1}\right) \\ &\overset{(b)}{=} d_{k-1}\left(\left(\epsilon_{1,k-2} - b_{k-1}Y_{1,k-1}\right)\right. \\ &\qquad \left. + m_{k-1}\left(\epsilon_{2,k-2} - m_{k-2}b_{k-1}Y_{2,k-1}\right)\right), \end{aligned} \quad (41)$$

where in (a) $d_k > 0$ is a gain factor chosen to minimize D_K under the constraint $P_k \leq P$; in (b) we used $b_{2,k} = m_{k-1}b_k$.

Next, similarly to Section III, we let $\alpha_{i,k} \triangleq \mathbb{E}\{\epsilon_{i,k}^2\}$, and note that since $b_{2,k} = m_{k-1}b_k$, then $\alpha_{1,k} = \alpha_{2,k} \triangleq$

$$b_k = \begin{cases} \sqrt{\frac{P(\alpha_{k-1}+m_{k-1}r_{k-1})}{2}}\frac{\eta_k+\theta_k m_k m_{k-1}}{\eta_k(P+\sigma_z^2)+\theta_k m_k m_{k-1}(P+\rho_z\sigma_z^2)}, & k=1,2,\ldots,K-1 \\ \sqrt{\frac{P(\alpha_{K-1}+m_{K-1}r_{K-1})}{2(P+\sigma_z^2)^2}}, & k=K. \end{cases} \tag{48}$$

$\alpha_k \forall k$. With this formulation, similarly to Section III, α_k is the MSE after k channel uses. Furthermore, we define $r_k \triangleq \mathbb{E}\{\epsilon_{1,k}\epsilon_{2,k}\}$. In Appendix D-B we show that the optimal choice of d_k for the proposed JSCC-DP scheme is such that the instantaneous average transmission power obeys $P_k = P$. This results in the following expression for d_k:

$$d_k = \sqrt{\frac{P}{2\left(\alpha_k + m_k r_k\right)}}. \tag{42}$$

Finally, similarly to Section III, we initialize the scheme by setting $\hat{S}_{i,0} = 0, \epsilon_{i,0} = -S_i, \alpha_0 = \sigma_s^2$ and $r_0 = \rho_s \sigma_s^2$.

Similarly to the JSCC-OL scheme, the JSCC-DP scheme described above can be stated within the class of linear and memoryless schemes defined in Subsection II-C: The encoder state update for the JSCC-DP scheme can be written in the from of (7) by setting $\mathbf{U}_k = [\epsilon_{1,k-1}, \epsilon_{2,k-1}]^T$, the transmitted signal X_k is a linear function of the encoder states as given in (41), and the state evolves via $U_{i,k+1} \equiv \epsilon_{i,k} = \epsilon_{i,k-1} - \hat{\epsilon}_{i,k-1}, \hat{\epsilon}_{i,k-1} = b_{i,k} Y_{i,k}$.

Our objective is to minimize the MSE after K channel uses, over all possible vectors of estimation coefficients $\mathbf{b} = [b_1, b_2, \ldots, b_K] \in \mathfrak{R}^K$, and over all possible vectors of modulation coefficients $\mathbf{m} = [m_0, m_1, \ldots, m_{K-1}] \in \{1, -1\}^K$. We denote this minimal MSE by $D_{K,\min}$:

$$D_{K,\min} = \min_{\mathbf{b}\in\mathfrak{R}^K, \mathbf{m}\in\{1,-1\}^K} \alpha_K(\mathbf{m}, \mathbf{b}). \tag{43}$$

As the joint minimization in (43) is computationally very intensive, we define $\alpha_{K,\min}(\mathbf{m})$ to be the minimal achievable MSE after K channel uses, given a specific modulation vector $\mathbf{m}$:

$$\alpha_{K,\min}(\mathbf{m}) = \min_{\mathbf{b}\in\mathfrak{R}^K} \alpha_K(\mathbf{m}, \mathbf{b}). \tag{44}$$

We use DP to *calculate* $\alpha_{K,\min}(\mathbf{m})$, thereby arriving at the optimization problem:

$$D_{K,\min} = \min_{\mathbf{m}\in\{1,-1\}^K} \alpha_{K,\min}(\mathbf{m}), \tag{45}$$

which can be solved by searching over the possible 2^K modulation vectors. In Remark 13 we comment on the practical implementation of this search. In the sequel we refer to the transmission scheme (40)–(41) which uses the optimal $\mathbf{b}$ and $\mathbf{m}$ as the *DP scheme*. Next, we present the algorithm for finding the minimizing $\mathbf{b}$ and the minimal $\alpha_{K,\min}(\mathbf{m})$ for a given $\mathbf{m}$.

B. The Minimizing $\mathbf{b}$ and the Minimal $\alpha_{K,min}(\mathbf{m})$

Let $\mathbf{m}$ be a given modulation vector. Then, from (1) and (40)–(42) we obtain the following recursive expressions for α_k and r_k (see Appendix D-A for the details):

$$\begin{aligned} \alpha_k &= \alpha_{k-1} + b_k^2 \cdot (P+\sigma_z^2) \\ &\quad - b_k\sqrt{2P\left(\alpha_{k-1}+m_{k-1}r_{k-1}\right)} \qquad (46a) \\ r_k &= r_{k-1} + b_k^2 m_{k-1} \cdot (P+\rho_z\sigma_z^2) \\ &\quad - b_k m_{k-1}\sqrt{2P\left(\alpha_{k-1}+m_{k-1}r_{k-1}\right)}. \qquad (46b) \end{aligned}$$

Observe that (α_{k-1}, r_{k-1}) can be treated as a state variable, which, given b_k and $\mathbf{m}$, evolves deterministically at time k. Thus, finding $\alpha_{K,\min}(\mathbf{m})$ can be cast as a DP with state (α_{k-1}, r_{k-1}), actions b_k, and cost function $\alpha_K(\mathbf{m})$. Note that with this formulation, given $\mathbf{m}$, b_k is a function of the constants P, σ_z^2, ρ_z, and of (α_{k-1}, r_{k-1}). Hence, the last action b_K is the linear MMSE estimation coefficient for estimating $\epsilon_{1,K-1}$ from $Y_{1,K}$.[10] Finally, the DP solution [10, Ch. 1.3] implies that α_k can be written as $\alpha_k = \eta_{k-1}\alpha_{k-1} + \theta_{k-1}m_{k-1}r_{k-1}$, where the sequences η_k and $\theta_k, k = 1, 2, \ldots, K-1$, are obtained using backwards recursion (in time). The minimizing $\mathbf{b}$ and the sequences η_k and θ_k are given in the following theorem:

Theorem 6. For a fixed $\mathbf{m}$, the sequences η_k and $\theta_k, k = 1, 2, \ldots, K-1$, are defined through the backwards recursions (in time):

$$\eta_{k-1} = \eta_k - \frac{P\left(\eta_k + \theta_k m_k m_{k-1}\right)^2}{2(\eta_k(P+\sigma_z^2)+\theta_k m_k m_{k-1}(P+\rho_z\sigma_z^2))} \tag{47a}$$

$$\theta_{k-1} = \theta_k m_k m_{k-1} - \frac{P\left(\eta_k + \theta_k m_k m_{k-1}\right)^2}{2(\eta_k(P+\sigma_z^2)+\theta_k m_k m_{k-1}(P+\rho_z\sigma_z^2))}, \tag{47b}$$

where $\eta_{K-1} = \left(1 - \frac{P}{2(P+\sigma_z^2)}\right)$ and $\theta_{K-1} = -\frac{P}{2(P+\sigma_z^2)}$. Furthermore, the coefficients $b_k, k = 1, 2, \ldots, K$, are given by (48) at the top of the page. The corresponding MSE at time K is the minimal MSE given $\mathbf{m}$.

Proof. The proof is provided in Appendix D-A. □

Thm. 6 can be used for calculating the optimal $\mathbf{b}$ for a given $\mathbf{m}$. The procedure is summarized in Alg. 1.

Remark 12. As we aim at minimizing $\alpha_K(\mathbf{m}, \mathbf{b})$ for a given $\mathbf{m}$, then b_K is the MMSE estimation coefficient for estimating $\epsilon_{1,K-1}$ from $Y_{1,K}$, given $\mathbf{m}$. It should be noted that for $k < K$, setting the b_k's to be the MMSE estimation coefficients is not necessarily optimal as the b_k's affect the future time indices. With this observation, it is clear why the JSCC-OL scheme, which applies the MMSE estimator *for all* k's, is not optimal, even among the memoryless linear transmission schemes.

Remark 13. Note that any choice of $\mathbf{m}$ will result in an upper bound on $D_{K,\min}$. While finding $D_{K,\min}$ requires searching

[10] Note that since $\epsilon_{1,k-1}$ and $Y_{1,k}$ are jointly Gaussian, then in this case the linear MMSE is the full MMSE.

Algorithm 1 Calculating the Minimizing $\mathbf{b}$ and $\alpha_{K,\min}(\mathbf{m})$

1: Initialization: $\eta_{K-1} \leftarrow \left(1-\frac{P}{2(P+\sigma_z^2)}\right)$, $\theta_{K-1} \leftarrow -\frac{P}{2(P+\sigma_z^2)}$
2: Compute the sequences η_k and θ_k using the backwards recursions (47)
3: $\alpha_0 \leftarrow \sigma_s^2$, $r_0 \leftarrow \rho_s$
4: **for** $k = 1, 2, \ldots, K$ **do**
5: Calculate b_k as in (48)
6: Calculate α_k and r_k as in (46)
7: **end for**
8: Output: $\mathbf{b}, \alpha_{K,\min}(\mathbf{m})$

over all 2^K possible $\mathbf{m}$ sequences, in practice, the search can be shortened at the expense of possibly achieving a larger MSE. One approach for reducing the search space is motivated by the alternating sign of ρ_k in the JSCC-OL and JSCC-LQG schemes, for for asymptotic large values of k, (see [9, Eqs. (23), (36)–(37)]): We can enforce such a behavior one $\mathbf{m}$ by setting $\mathbf{m}$ to be a sequence with alternating signs after some $L \ll K$ channel uses, thereby searching only over the first possible 2^L sequences. Numerical simulations show that when the SNR is not too low, then this approach performs well, as shown in the next section.

Remark 14. The results of [10, Ch. 1.3] imply that for the symmetric setting in which the MSEs at both receivers are restricted to be the same for every k, the JSCC-DP scheme, described in Alg. 1, is optimal. Hence, for this scenario, the MSE of the JSCC-DP scheme constitutes a lower bound and a benchmark on the MSE of any linear and memoryless scheme for the GBCF in the finite horizon regime. For the general setting, e.g., if the MSEs are not required to be the same at every k, the optimal transmission coefficients can be obtained via a numerical search. Moreover, a simple (and non-tight) lower bound on the MSE achieved by any linear and memoryless scheme is obtained by treating each transmitter-receiver pair as a Gaussian PtP channel with FB and using the expression for the optimal MSE given in [37, Eq. (9)]. Using this bounding technique, the MSE at $\mathrm{Rx}_i, i = 1, 2$, after k channel uses is lower bounded by $\frac{\sigma_i^2}{(P+\sigma_{z,i}^2)^k}$.

VI. Comparative Discussion and Numerical Examples

In this section we compare the different JSCC transmission schemes, i.e., JSCC-OL, JSCC-LQG and JSCC-DP, and demonstrate our results via numerical examples.

A. JSCC-DP Outperforms JSCC-OL and JSCC-LQG

The following proposition formally states that the JSCC-DP scheme outperforms both the JSCC-OL and the JSCC-LQG schemes:

Proposition 5. For any fixed number of channel uses K, the JSCC-DP scheme achieves MSE at least as low as the MSEs achieved by the JSCC-OL and JSCC-LQG schemes.

Proof outline. As stated in Remark 12, JSCC-DP outperforms JSCC-OL. Now, recall that in Appendix D-B it is shown that choosing $P_k = P$ in the JSCC-DP scheme is optimal. Thus, the JSCC-DP scheme is the optimal scheme (in the sense of minimizing the MSE after K channel uses) among the class of linear JSCC schemes which can be formulated via (40)–(41), and satisfy the constraint $P_k \leq P$. In Appendix E we explicitly show that the JSCC-LQG scheme can be written in the form of (40)–(41). Furthermore, we show that all three JSCC-LQG decoders considered in this work, (22), (24), and (35), have the same structure as the decoder applied by the JSCC-DP scheme. We conclude that any JSCC-LQG scheme which satisfies the per-symbol average power constraint (5) is within the search range of the JSCC-DP scheme, and therefore JSCC-DP achieves MSE at least as low as JSCC-LQG. □

B. Numerical Examples

We first consider the low SNR regime as it facilitates demonstrating different characteristics of the JSCC schemes studied in the paper. Consider the transmission of a pair of Gaussian sources with $\sigma_s^2 = 1$ and $\rho_s = 0.4$, over a GBCF, with noise parameters $\sigma_z^2 = 1.5, \rho_z = 0.3$, and power constraint $P = 0.03$. Fig. 6 depicts the MSE values corresponding to (11), (23), (35), and the approximation of (45) described in Remark 13 for $L = 15$ and $L = 25$. The JSCC-OL scheme is initialized via $\epsilon_{i,0} = -S_i$, while for the JSCC-LQG scheme we consider two initializations: either $\mathbf{U}_1 = \mathbf{S}$ or $\mathbf{U}_1 = \sqrt{\gamma}\mathbf{S}$ with the optimal γ. The line marked by `ZT LQG` refers to the original LQG scheme applied as in [9], namely using a zero trajectory (ZT) decoder at the receivers and using the initialization $\mathbf{U}_1 = \mathbf{S}$. It can be observed that for low values of K the new decoder (24) significantly improves upon the ZT decoder, while for large values of K the two decoders achieve approximately the same MSE. It can be further observed from the figure that scaling can significantly improve the performance in the low SNR regime.

Fig. 6 also shows the importance of the parameter L in the approximated DP solution: When $L = 15$, JSCC-OL outperforms JSCC-DP, while when $L = 25$, JSCC-DP outperforms JSCC-OL. Our simulations indicate that for the current scenario parameters the optimal $\mathbf{m}$ sequence starts alternating for $L \geq 25$, and setting $L = 25$ does not result in any difference in the MSE compared to the exact solution of (45).[11] The numerical results also support the conclusion of the discussion on lower bounding the MSE in Subsection V-B as JSCC-DP with a proper value of L outperforms both JSCC-OL and JSCC-LQG. Note that, while the gap between JSCC-DP and JSCC-OL in Fig. 6 is very small for the scenario parameters specified above, this gap becomes larger as the total number of steps K increases. Fig. 6 also shows that, there is a relatively large gap between JSCC-LQG and JSCC-OL, in particular when scaling is not applied for JSCC-LQG. This gap follows from the fact that the distance between the covariance matrix of the sources and the covariance matrix of the JSCC-LQG steady state is large (using the terminology of Subsection IV-B). Explicitly calculating this distance we have $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_s) \approx 213.515$, while

[11] This was verified for $25 \leq K \leq 30$.

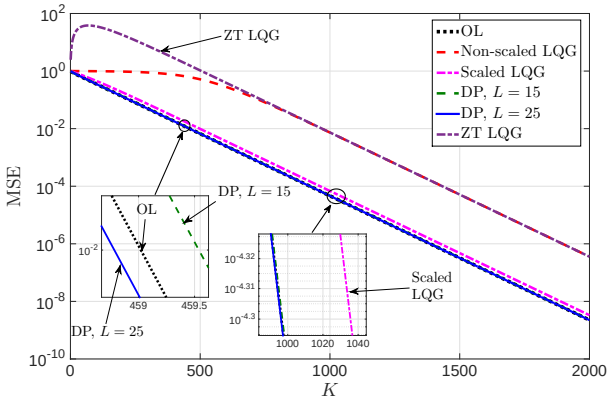


Fig. 6: MSE vs. time for $\sigma_s^2 = 1, \rho_s = 0.4, \rho_z = 0.3, \sigma_z^2 = 1.5$ and $P = 0.03$. m_k is set to an alternating sequence starting from $L = 25$ (solid line), and from $L = 15$ (dashed line).

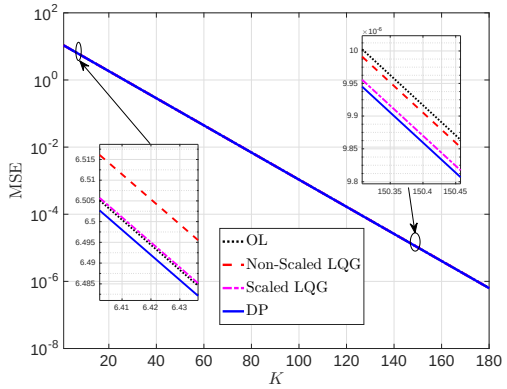


Fig. 7: MSE vs. time for $\sigma_s^2 = 11.8, \rho_s = -0.066, \sigma_z^2 = 1, \rho_z = 0.5$ and $P = 0.2$. For this setting $\mathsf{Q}_s \approx \mathsf{Q}_u$.

the distance $\mathbb{D}(\mathsf{Q}_u, \gamma\mathsf{Q}_s) \approx 88.77$ is much smaller, and indeed, a significant performance improvement is observed.

We next focus on the setting in which the sources and the channel are almost matched. For this setting, following the discussion in Subsection IV-B, it is expected that the JSCC-LQG scheme will have a fast convergence. Furthermore, for this scenario, due to this fast convergence, all the JSCC-LQG versions should perform roughly the same. These expectations are indeed confirmed in Fig. 7. The scenario parameters in Fig. 7 are $\sigma_z^2 = 1, \rho_z = 0.5$ and $P = 0.2$, which leads to $\sigma_u^2 \approx 11.85$ and $\rho_u \approx -0.067$. We set $\sigma_s^2 = 11.8$ and $\rho_s = -0.066$, which results in $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_s) \approx 0.0722$. The scaled JSCC-LQG decreases this distance to $\mathbb{D}(\mathsf{Q}_u, \gamma\mathsf{Q}_s) \approx 0.015$. It can be observed that the plots corresponding to the different schemes are almost indistinguishable. This follows as for small values of K JSCC-OL is close to JSCC-DP, while the JSCC-LQG versions are very close to their steady state, thus, no slow-start is observed. Furthermore, the correlation between the sources is very low which eliminates the most significant advantage of JSCC-OL over JSCC-LQG. A closer look in the "zoom-in" plots in Fig. 7 shows that for very small K's JSCC-OL achieves MSEs smaller than the JSCC-LQG schemes, yet, *this relationship changes when K increases*. This is a consequence of *JSCC-LQG achieving a higher MSE exponent*. It can also be observed that, as expected, the performance improvement of the JSCC-LQG scaling when Q_s is close to Q_u is minor, and that JSCC-DP outperforms both JSCC-OL and JSCC-LQG and has the same slope as the LQG schemes (the best known MSE exponent).[12] Finally, Figs. 6 and 7 indicate that while the JSCC-LQG schemes have the best known MSE exponent, their finite horizon performance can sometimes be hindered by the issue of initialization.

The last numerical example shows the impact of having a source of common randomness at the transmitter and the receivers. Using this common randomness, the transmitter can generate the initial state by transforming the pair of sources without causing an MSE floor. However, as stated in Subsection IV-B, the initializations of the JSCC-OL and the JSCC-LQG schemes aims at achieving different objectives. While in the JSCC-OL scheme it is desirable to have $|\rho_0|$ as high as possible, the JSCC-LQG scheme aims at matching the sources and the channel. To demonstrate this idea we consider a setting in which the sources and the channel are perfectly matched, and a common source of randomness is available. For such a source, the JSCC-LQG scheme simply initializes $\mathbf{U}_1 = \mathbf{S}$, without using the common randomness. On the other hand, the OL scheme uses the common randomness to set $|\rho_0|$ to some value close to 1.

To highlight the difference between the JSCC-OL and JSCC-LQG, we consider a low SNR scenario, by setting $\sigma_z = 1, \rho_z = 0$ and $P = 0.1$. Consequently, the achievable MSE exponents are low. The stationary covariance matrix of the JSCC-LQG scheme for this setting is given by:

$$\mathsf{Q}_u \approx \begin{bmatrix} 21.0606721498435 & -1.00118909550747 \\ -1.00118909550747 & 21.0606721498435 \end{bmatrix},$$

which implies that $\rho_u \approx -0.04754$. Let the sources to be transmitted be distributed according to $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathsf{Q}_u)$. Thus, in this scenario, the JSCC-LQG scheme minimizes $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{U,1})$ by setting $\mathbf{U}_1 = \mathbf{S}$, *regardless* if common source of randomness is available or not. On the other hand, if the transmitter and the receivers share a common source of randomness, the JSCC-OL scheme can use it to set $\rho_0 = 0.95$.[13] Figure 8 depicts the MSE achieved by the two schemes. It can be observed that, for the considered values of K, the JSCC-OL scheme strictly outperforms the JSCC-LQG scheme even though the sources and channel are matched. This gain is achieved by the JSCC-OL as it takes advantage of the common source of randomness to increase $|\rho_0|$, while the common source of randomness is not used by the JSCC-LQG scheme as the sources and the channel are already matched.. It should be noted that if a common source of randomness *is not available*, then the JSCC-LQG scheme indeed outperforms the JSCC-OL scheme in this scenario, and in fact, for this setting, the JSCC-LQG scheme achieves the same MSE as the JSCC-DP scheme.[14]

C. When Does JSCC-OL Outperform JSCC-LQG?

Recall that for channel coding in the infinite horizon regime JSCC-LQG outperforms JSCC-OL. Yet, in the finite horizon regime, Figs. 6 and 7 demonstrate that JSCC-OL can outperform JSCC-LQG. This leads to the question: *When does*

[12] In this setting we used $L = 15$.

[13] Note that this example can be adapted to any $|\rho_0| < 1$.

[14] In this setting we used $L = 15$.

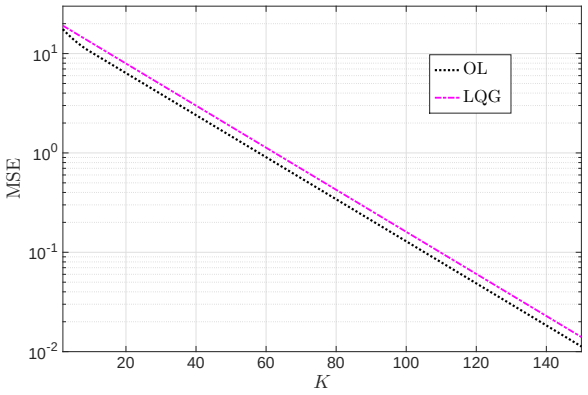


Fig. 8: Matched sources and channel. $\sigma_z = 1, \rho_z = 0$ and $P = 0.1$. The LQG scheme sets $\mathbf{U}_1 = \mathbf{S}$ while the OL scheme sends $\bar{\mathbf{S}}$.

JSCC-OL outperform JSCC-LQG? To answer this question we focus on the symmetric setting, and note that Figs. 6 and 7 imply that the answer changes for different target MSEs. More precisely, using Thm. 1 and Thm. 5 one can answer the question which scheme (JSCC-OL or JSCC-LQG) achieves the target MSE with the least number of channel uses. For instance, consider the scenario illustrated in Fig. 6, i.e., $\sigma_s^2 = 1, \rho_s = 0.4, \sigma_z^2 = 1.5, \rho_z = 0.3, P = 0.03$ and let $D = 10^{-2}$. Here, $K_{\text{LQG}} = 498$ and $K_{\text{OL}}^{\text{ub}} = 470$. Thus, JSCC-OL outperforms JSCC-LQG. In fact, using the upper bound presented in Thm. 1, it can be shown that for $D = 10^{-2}$, JSCC-OL outperforms JSCC-LQG for all $P < 0.1978$.

VII. Conclusion

In this work we studied the transmission of a pair of correlated Gaussian sources over the two-user GBCF focusing on linear and memoryless transmission schemes in the *finite horizon regime*. We characterized the minimal number of channel uses required to achieve a non-zero pair of MSEs for three finite horizon JSCC schemes: An adaptation of the OL scheme of [8], an adaptation of the LQG scheme of [9], and a novel scheme derived in this work designed using the DP approach.

For the JSCC-OL scheme, we first demonstrated that the initialization which takes advantage of the correlation between the sources is superior to the one suggested in [8]. Then, for the proposed JSCC-OL scheme we derived upper and lower bounds on the number of channel uses required to achieve a target pair of MSEs. For the symmetric setting with independent sources and independent noise components, we showed that, even though JSCC-OL does not encode over blocks of source symbols, in the low SNR regime, it achieves approximately the same source-channel bandwidth ratio as the best known SSCC scheme, which applies a source code and a channel code with an asymptotically large blocklength. More precisely, the gap between the source-channel bandwidth ratios achieved by the JSCC-OL and the SSCC schemes is bounded by a quantity independent of the SNR.

For the JSCC-LQG scheme, we first introduced a new decoder based on the MMSE criterion, which achieves MSE values smaller than or equal to those achieved by the original decoder proposed in [9]. For the general setting, we presented lower and upper bounds on the number of channel uses required to achieve a target pair of MSEs, while for the symmetric setting we explicitly characterized this number of channel uses. For the symmetric setting we also showed that, by properly scaling the transmitted sources, it is possible to arrive at the target MSE much faster than with the original initialization of [9]. This optimal scaling leads to a linear and memoryless transmission scheme with very good finite horizon performance and with the best known infinite horizon performance.

Lastly, we used DP to derive the optimal linear and memoryless JSCC scheme in the symmetric setting. This scheme requires finding a vector of modulation coefficients and a vector of estimation coefficients which minimize the MSE after K channel uses. We showed that this minimization problem can be simplified into the problem of searching only over the possible modulation vectors, while the optimal vector of estimation coefficients, per modulation vector, was formulated as a DP problem whose solution is obtained using a recursive deterministic relationship. For a finite number of channel uses, the JSCC-DP scheme achieves MSE values lower than both JSCC-OL and JSCC-LQG. Since finding the optimal modulation vector becomes computationally infeasible as the number of channel uses increases, we proposed a computationally feasible approximate solution, which performs well for moderate to high SNR values.

The comparison of the three JSCC schemes indicate their differences: JSCC-OL is time-varying, and reaches steady state relatively *quickly*, but it is suboptimal in the infinite horizon regime. JSCC-LQG is time-invariant, reaches steady state relatively *slowly*, but has the best known performance in the infinite horizon regime. JSCC-DP is time-varying, and outperforms both JSCC-OL and JSCC-LQG in the finite horizon regime. As JSCC-DP applies backwards recursion, it can be applied only in the finite horizon regime. While JSCC-LQG reaches steady state relatively slowly, by applying the proposed scaling, and using the improved MMSE decoder, its finite horizon performance can be significantly improved. However, even with these improvements JSCC-OL can outperform JSCC-LQG in the finite horizon regime. Finally, the initialization objectives in the finite horizon regime for JSCC-LQG and for JSCC-OL are fundamentally different: The JSCC-OL scheme aims at increasing the correlation between the transmitted signals, while the JSCC-LQG scheme aims at matching the covariance matrices of the the initial and steady states.

We remark that the results presented in this work are important in identifying simple yet efficient coding schemes for the transmission of correlated Gaussian sources over multiuser channels with FB when strict delay constraints are imposed.

Acknowledgment

The authors would like to thank the associate editor and the anonymous reviewers for their comments, which greatly improved the results of the manuscript.

APPENDIX A
JSCC-OL IN THE FINITE HORIZON REGIME - PROOFS

A. Deriving the MSE for the Case $\epsilon_0 = \mathbf{S} + \mathbf{W}$ in Eq. (13)

Recall that we focus here on the symmetric setting, and define $\bar{\mathbf{S}} \triangleq \mathbf{S} + \mathbf{W}$. Rx_i estimates S_i in two steps: First Rx_i estimates $\bar{S}_i$ via (10). Let $\hat{\bar{S}}_{i,k}$ denote this estimation. Then, Rx_i computes the MMSE estimate of S_i from $\hat{\bar{S}}_{i,k}$. The scheme is initialized via $\hat{\bar{S}}_{i,0} = 0$ and $\epsilon_{i,0} = -\bar{S}_i, i = 1, 2$.

From the theory of MMSE estimation [36, Ch. 11.4], the MSE in estimating S_i from $\hat{\bar{S}}_{i,k}$ is given by:

$$\mathbb{E}\left\{(S_i - \hat{\bar{S}}_{i,k})^2\right\} = \sigma_s^2 - \frac{(\mathbb{E}\{S_i \hat{\bar{S}}_{i,k}\})^2}{\mathbb{E}\{\hat{\bar{S}}_{i,k}^2\}}. \tag{A.1}$$

From OL decoding (10) we write $\hat{\bar{S}}_{i,k} = -\sum_{m=1}^{k} \hat{\epsilon}_{i,m-1}$, where the terms $\hat{\epsilon}_{i,m-1}$ denote the estimates of $\epsilon_{i,m-1}$ during the transmission of $\bar{S}_i$. Since $\epsilon_{i,k} = \epsilon_{i,k-1} - \hat{\epsilon}_{i,k-1}$, it follows that $\hat{\bar{S}}_{i,k} = \epsilon_{i,k} - \epsilon_{i,0}$. Therefore, the MSE can be written as:

$$\begin{aligned}&\mathbb{E}\left\{(S_i - \hat{\bar{S}}_{i,k})^2\right\}\\ &\quad= \sigma_s^2 - \frac{(\mathbb{E}\{S_i\epsilon_{i,k}\} + \sigma_s^2)^2}{\alpha_{i,k} + \sigma_s^2 + \sigma_w^2 + 2\mathbb{E}\{(S_i + W_i)\epsilon_{i,k}\}}.\end{aligned} \tag{A.2}$$

Next, to explicitly evaluate the MSE (A.2) we derive recursive expressions for $\mathbb{E}\{S_i\epsilon_{i,k}\}$ and $\mathbb{E}\{(S_i + W_i)\epsilon_{i,k}\}$. First, plugging the relationship $\epsilon_{i,k} = \epsilon_{i,k-1} - \hat{\epsilon}_{i,k-1}$ into the expectation $\mathbb{E}\{S_i\epsilon_{i,k}\}$ we obtain:

$$\mathbb{E}\{S_i\epsilon_{i,k}\} = \mathbb{E}\{S_i\epsilon_{i,k-1}\} - \mathbb{E}\{S_i\hat{\epsilon}_{i,k-1}\}. \tag{A.3}$$

Using (9), letting $\alpha_{i,k} \triangleq \alpha_k$, and considering $i = 1$, we write:

$$\begin{aligned}&\mathbb{E}\{S_1\hat{\epsilon}_{1,k-1}\}\\ &\quad= \frac{\mathbb{E}\{\epsilon_{1,k-1}Y_{1,k}\}}{\mathbb{E}\left\{Y_{1,k}^2\right\}}\mathbb{E}\{S_1Y_{1,k}\}\\ &\quad= \frac{\Psi_{k-1}\sqrt{\alpha_{k-1}}(1 + |\rho_{k-1}|)}{P + \sigma_z^2}\mathbb{E}\{S_1Y_{1,k}\}\\ &\quad= \frac{\Psi_{k-1}^2(1 + |\rho_{k-1}|)}{P + \sigma_z^2}\mathbb{E}\{S_1(\epsilon_{1,k-1} + \epsilon_{2,k-1}\mathrm{sgn}(\rho_{k-1}))\}.\end{aligned} \tag{A.4}$$

Let $\lambda \triangleq \frac{P}{2(P+\sigma_z^2)}$. Combining (A.3) and (A.4), and noting that $\Psi_k^2 = \frac{P}{2(1+|\rho_k|)}$, we have:

$$\begin{aligned}\mathbb{E}\{S_1\epsilon_{1,k}\} &= \mathbb{E}\{S_1\epsilon_{1,k-1}\}\\ &\quad- \frac{P}{2(P+\sigma_z^2)}\mathbb{E}\{S_1(\epsilon_{1,k-1}+\epsilon_{2,k-1}\mathrm{sgn}(\rho_{k-1}))\}\\ &= \left(1 - \frac{P}{2(P + \sigma_z^2)}\right)\mathbb{E}\{S_1\epsilon_{1,k-1}\}\\ &\quad- \frac{P \cdot \mathrm{sgn}(\rho_{k-1})}{2(P + \sigma_z^2)}\mathbb{E}\{S_1\epsilon_{2,k-1}\}\\ &= (1 - \lambda)\mathbb{E}\{S_1\epsilon_{1,k-1}\}\\ &\quad- \lambda \cdot \mathrm{sgn}(\rho_{k-1})\mathbb{E}\{S_1\epsilon_{2,k-1}\},\end{aligned} \tag{A.5}$$

Following similar steps, we write $\mathbb{E}\{S_1\epsilon_{2,k}\}$ as:

$$\begin{aligned}\mathbb{E}\{S_1\epsilon_{2,k}\} &= (1 - \lambda)\mathbb{E}\{S_1\epsilon_{2,k-1}\}\\ &\quad- \lambda \cdot \mathrm{sgn}(\rho_{k-1}) \cdot \mathbb{E}\{S_1\epsilon_{1,k-1}\}.\end{aligned} \tag{A.6}$$

Finally, since $\mathbb{E}\{S_1\epsilon_{1,0}\} = -\sigma_s^2$, and $\mathbb{E}\{S_1\epsilon_{2,0}\} = -\rho_s\sigma_s^2$, we can recursively calculate $\mathbb{E}\{S_1\epsilon_{1,k}\}$.

To evaluate $\mathbb{E}\{(S_i + W_i)\epsilon_{i,k}\}$, we recall that $\bar{S}_i = S_i + W_i$, and follow the steps leading to (A.5)–(A.6) to obtain:

$$\begin{aligned}\mathbb{E}\{\bar{S}_1\epsilon_{1,k}\} &= (1 - \lambda)\mathbb{E}\{\bar{S}_1\epsilon_{1,k-1}\}\\ &\quad- \lambda \cdot \mathrm{sgn}(\rho_{k-1})\mathbb{E}\{\bar{S}_1\epsilon_{2,k-1}\},\end{aligned} \tag{A.7}$$

$$\begin{aligned}\mathbb{E}\{\bar{S}_1\epsilon_{2,k}\} &= (1 - \lambda)\mathbb{E}\{\bar{S}_1\epsilon_{2,k-1}\}\\ &\quad- \lambda \cdot \mathrm{sgn}(\rho_{k-1}) \cdot \mathbb{E}\{\bar{S}_1\epsilon_{1,k-1}\},\end{aligned} \tag{A.8}$$

with the initial conditions $\mathbb{E}\{\bar{S}_1\epsilon_{1,0}\} = -\sigma_s^2 - \sigma_w^2$, and $\mathbb{E}\{\bar{S}_1\epsilon_{2,0}\} = -\rho_s\sigma_s^2 - \rho_w\sigma_w^2$. Therefore, using (A.2) and the recursive relationships (A.5)–(A.8) we recursively obtain the MSEs of the two-step estimator.

B. Proof of Theorem 1

Recall that $\alpha_{i,0} = \sigma_i^2, i = 1, 2$, and that $\alpha_{i,k}$ is the MSE at Rx_i after the k'th transmission. From (11) we have:

$$\log\left(\frac{\alpha_{1,K}}{\sigma_1^2}\right) = \sum_{k=1}^{K}\log\left(\frac{\sigma_{z,1}^2 + \Psi_{k-1}^2 g^2(1 - \rho_{k-1}^2)}{\pi_1}\right).$$

As $|\rho_k| \in [0, 1]$, it follows that:

$$\Psi_{k-1}^2 g^2(1 - \rho_{k-1}^2) = \frac{Pg^2(1 - \rho_{k-1}^2)}{1 + g^2 + 2g|\rho_k - 1|} \le \frac{Pg^2}{1 + g^2}.$$

Thus, we obtain the upper bound $\frac{\sigma_{z,1}^2 + \Psi_{k-1}^2 g^2(1-\rho_{k-1}^2)}{\pi_1} \le \frac{\sigma_{z,1}^2 + \pi_1 g^2}{\pi_1 + \pi_1 g^2}$. Next, we use the fact that $\log(x) \le x - 1$ and write:

$$\log\left(\frac{\sigma_{z,1}^2 + \pi_1 g^2}{\pi_1 + \pi_1 g^2}\right) \le \frac{\sigma_{z,1}^2 + \pi_1 g^2}{\pi_1 + \pi_1 g^2} - 1 = -\frac{P}{\pi_1 + \pi_1 g^2}.$$

Thus, it follows that $\log\left(\frac{\alpha_{1,K}}{\sigma_1^2}\right) = \log\left(\frac{D_1}{\sigma_1^2}\right) \le -\frac{KP}{\pi_1 + \pi_1 g^2}$, which implies that:

$$K_{\mathrm{OL}}^{\mathrm{ub}} = \left\lceil \frac{(1 + g^2)}{P}\max\left\{\pi_1\log\left(\frac{\sigma_1^2}{D_1}\right), \frac{\pi_2}{g^2}\log\left(\frac{\sigma_2^2}{D_2}\right)\right\}\right\rceil.$$

To obtain $K_{\mathrm{OL}}^{\mathrm{lb}}$ we note that $0 \le \Psi_{k-1}^2 g^2(1 - \rho_{k-1}^2)$ where equality is obtained by setting $\rho_{k-1} = 1$. Then, we use the inequality $1 - \frac{1}{x} \le \log x$ to obtain:

$$\log\left(\frac{\sigma_{z,1}^2}{\sigma_{z,1}^2 + P}\right) \ge 1 - \frac{\sigma_{z,1}^2 + P}{\sigma_{z,1}^2} = -\frac{P}{\sigma_{z,1}^2}.$$

Thus, we have $\log\left(\frac{D_1}{\sigma_1^2}\right) \ge -\frac{KP}{\sigma_{z,1}^2}$, which results in the following lower bound:

$$K_{\mathrm{OL}}^{\mathrm{lb}} = \left\lceil \max\left\{\frac{\sigma_{z,1}^2}{P}\log\left(\frac{\sigma_1^2}{D_1}\right), \frac{\sigma_{z,2}^2}{P}\log\left(\frac{\sigma_2^2}{D_2}\right)\right\}\right\rceil.$$

C. Proof of (15)

From [8, pg. 671] it follows that if R is an achievable symmetric rate for the GBCF, and $\rho_z = 0$, then $R < \frac{1}{2}\log\left(1 + \frac{2\chi_0 P}{\sigma_z^2}\right)$, where χ_0 is the unique positive root of the polynomial (in χ): $\chi^2 + \frac{3\sigma_z^2}{2P}\chi - \frac{\sigma_z^2}{2P} = \chi^2 + \frac{3}{2\mathrm{SNR}} - \frac{1}{2\mathrm{SNR}}$. The roots of this polynomial are given by:

$$\chi_{1,2} = \frac{1}{2}\left(-\frac{3}{2\mathrm{SNR}} \pm \sqrt{\frac{9}{4\mathrm{SNR}^2} + \frac{2}{\mathrm{SNR}}}\right).$$

Hence, χ_0 is given by $\chi_0 = \frac{1}{2}\left(-\frac{3}{2\text{SNR}} + \sqrt{\frac{9}{4\text{SNR}^2} + \frac{2}{\text{SNR}}}\right)$. Plugging χ_0 into the upper bound on R we write:

$$\begin{aligned} R &< \frac{1}{2}\log\left(1 + 2\chi_0 \text{SNR}\right) \\ &= \frac{1}{2}\log\left(1 + \text{SNR}\left(-\frac{3}{2\text{SNR}} + \sqrt{\frac{9}{4\text{SNR}^2} + \frac{2}{\text{SNR}}}\right)\right) \\ &= \frac{1}{2}\log\left(\sqrt{\frac{9}{4} + 2\text{SNR}} - \frac{1}{2}\right). \end{aligned}$$

D. Proof of (16a)

First, we obtain an upper bound on $2\log|a_1|$. Following steps similar to those described in [9, Section IV.C] for the symmetric GBCF with independent noises, we conclude that $a_1^2 = x_0$, where x_0 is the unique real positive root[15] of the equation:

$$\sigma_z^2 x^3 + \sigma_z^2 x^2 - (\sigma_z^2 + 2P)x - \sigma_z^2 = 0.$$

Rewriting this equation equivalently as:

$$x^3 + x^2 - \left(1 + \frac{2P}{\sigma_z^2}\right)x - 1 = 0, \tag{A.9}$$

we upper bound x_0 using Budan's theorem [43]:

Theorem. (Budan's theorem) Let $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ be a polynomial of degree n, and let $p^{(j)}(x)$ be its j'th derivative. Define the function $V(\alpha)$ as the number of sign variations in the sequence $p(\alpha), p^{(1)}(\alpha), \ldots, p^{(n)}(\alpha)$. Then, the number of roots of the polynomial $p(x)$ in the open interval (a, b) is either equal to $V(a) - V(b)$, or less by an even number.

Let $p(x)$ be the polynomial in (A.9). Then we have:

$$p^{(0)}(x) = x^3 + x^2 - \left(1 + \frac{2P}{\sigma_z^2}\right)x - 1, \tag{A.10a}$$

$$p^{(1)}(x) = 3x^2 + 2x - \left(1 + \frac{2P}{\sigma_z^2}\right), \tag{A.10b}$$

$$p^{(2)}(x) = 6x + 2, \tag{A.10c}$$

$$p^{(3)}(x) = 6. \tag{A.10d}$$

For $x = 1$ we have $V(1) = 1$. Note that $\text{sgn}(p^{(1)}(1))$ depends on the term $\frac{2P}{\sigma_z^2}$, however, since $\text{sgn}(p^{(0)}(1)) = -1$ and $\text{sgn}(p^{(2)}(1)) = 1$, in both cases we have $V(1) = 1$. Next, we let $\chi = \frac{P}{2\sigma_z^2}$, and set $x = 1 + \chi$ to obtain:

$$\begin{aligned} p^{(0)}(1+\chi) &= \chi^3 > 0, \\ p^{(1)}(1+\chi) &= 3\chi^2 + 4\chi + 4 > 0, \\ p^{(2)}(1+\chi) &= 6\chi + 8 > 0, \\ p^{(3)}(1+\chi) &= 6 > 0. \end{aligned}$$

all larger than zero. Therefore, $V(1+\chi) = 0$. Thus, Budan's theorem implies that the number of roots of (A.9) in the interval $(1, 1+\chi)$ is 1. From Descartes' rule we know that there is a unique positive root, therefore $1 + \chi$ is an upper bound on x_0: $x_0 < 1 + \frac{P}{2\sigma_z^2}$.

[15]The uniqueness of a real positive root follows from Descartes' rule [39, Subsection 1.6.3.2].

Next, recall that $a_1^2 = x_0$, which implies that $2\log(|a_1|) = \log(x_0) \le \log\left(1 + \frac{P}{2\sigma_z^2}\right)$. Using the fact that $\log(x) \le x - 1$ we have the following bound on $2\log(|a_1|)$:

$$2\log(|a_1|) \le \frac{P}{2\sigma_z^2}. \tag{A.11}$$

Next, we explicitly upper bound $K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}}$ in the symmetric setting (we set $g = 1$ in (14a)):

$$\begin{aligned} K_{\text{OL}} - \kappa_{\text{sep}}^{\text{ub}} &\le K_{\text{OL}}^{\text{ub}} - \kappa_{\text{sep}}^{\text{ub}} \\ &\overset{(a)}{=} \frac{2(P+\sigma_z^2)}{P}\log\left(\frac{\sigma_s^2}{D}\right) - \frac{1}{2\log|a_1|}\log\left(\frac{\sigma_s^2}{D}\right) \\ &\overset{(b)}{\le} \log\left(\frac{\sigma_s^2}{D}\right)\left(\frac{2(P+\sigma_z^2)}{P} - \frac{2\sigma_z^2}{P}\right) \\ &\le \left\lceil 2\log\left(\frac{\sigma_s^2}{D}\right)\right\rceil, \end{aligned} \tag{A.12}$$

where (a) follows from specializing Thm. 1 to the symmetric setting, and (b) follows from the bound $2\log(|a_1|) \le \frac{P}{2\sigma_z^2}$.

E. Proof of (16b)

Recall that $\kappa_{\text{sep}}^{\text{lb}} \triangleq \frac{\log\left(\frac{\sigma_s^2}{D}\right)}{\log\left(\sqrt{\frac{9}{4}+2\text{SNR}} - \frac{1}{2}\right)}$. Thus, we write:

$$\begin{aligned} K_{\text{OL}} - \kappa_{\text{sep}}^{\text{lb}} &\le K_{\text{OL}}^{\text{ub}} - \kappa_{\text{sep}}^{\text{lb}} \\ &= \frac{2(P+\sigma_z^2)}{P}\log\left(\frac{\sigma_s^2}{D}\right) \\ &\quad - \frac{1}{\log\left(\sqrt{\frac{9}{4}+2\text{SNR}} - \frac{1}{2}\right)}\log\left(\frac{\sigma_s^2}{D}\right) \\ &= \log\left(\frac{\sigma_s^2}{D}\right)\left(2 + \frac{2}{\text{SNR}} - \frac{1}{\log\left(\sqrt{\frac{9}{4}+2\text{SNR}} - \frac{1}{2}\right)}\right) \\ &\overset{(a)}{\le} \log\left(\frac{\sigma_s^2}{D}\right)\left(2 + \frac{2}{\text{SNR}} - \frac{1}{\sqrt{2\text{SNR}}}\right) \\ &\le \left\lceil \left(2 + \frac{2}{\text{SNR}} - \frac{1}{\sqrt{2\text{SNR}}}\right)\log\left(\frac{\sigma_s^2}{D}\right)\right\rceil, \end{aligned}$$

where (a) follows from the fact that $\log\left(\sqrt{\frac{9}{4}+2\text{SNR}} - \frac{1}{2}\right) \le \sqrt{\frac{9}{4}+2\text{SNR}} - \frac{3}{2} \le \sqrt{2\text{SNR}}$.

APPENDIX B
JSCC-LQG IN THE FINITE HORIZON REGIME - PROOFS

A. Proof of Theorem 2

The MMSE estimator of S_i based on $\hat{U}_{i,k}$ is the conditional expectation $\mathbb{E}\{S_i|\hat{U}_{i,k}\}$, [36, Eqn. (11.10)]. Now, from (18) we can write:

$$\begin{aligned} \mathbf{U}_k &= \mathsf{A}\mathbf{U}_{k-1} + \mathbf{Y}_{k-1} \\ &= \mathsf{A}\mathbf{U}_{k-1} - \mathbf{B}\mathbf{C}^T\mathbf{U}_{k-1} + \mathbf{Z}_{k-1} \\ &= (\mathsf{A} - \mathbf{B}\mathbf{C}^T)\mathbf{U}_{k-1} + \mathbf{Z}_{k-1}, \end{aligned} \tag{B.1}$$

and from (21) we have:

$$\begin{aligned}\hat{\mathbf{U}}_k &= \mathsf{A}\hat{\mathbf{U}}_{k-1} + Y_{k-1} \\ &= \mathsf{A}^{k-1}\hat{\mathbf{U}}_1 + \sum_{m=1}^{k-1}\mathsf{A}^{k-m-1}\mathbf{Y}_m \\ &= \sum_{m=1}^{k-1}\mathsf{A}^{k-m-1}(-\mathbf{B}\mathbf{C}^T\mathbf{U}_{k-1} + \mathbf{Z}_{k-1}). \end{aligned} \tag{B.2}$$

From the fact that $\mathbf{Z}_k$ is a zero-mean Gaussian vector, from the linear relationship in (B.2), and from the fact that $\mathbf{U}_1 = \mathbf{S}$, it follows that for $i = 1, 2$, $\hat{U}_{i,k+1}$ and S_i are jointly Gaussian, both with zero mean. From [36, Eqn. (10.16)] it follows that $\mathbb{E}\left\{S_i \middle| \hat{U}_{i,k+1}\right\} = \frac{\mathbb{E}\{S_i\hat{U}_{i,k+1}\}}{\mathbb{E}\{\hat{U}^2_{i,k+1}\}}\hat{U}_{i,k+1}$. Next, we expand (18) as:

$$\begin{aligned}\mathbf{U}_k &= \mathsf{A}\mathbf{U}_{k-1} + \mathbf{Y}_{k-1} \\ &= \mathsf{A}^{k-1}\mathbf{S} + \sum_{m=1}^{k-1}\mathsf{A}^{k-m-1}\mathbf{Y}_m. \end{aligned} \tag{B.3}$$

Therefore, combining (B.3) and (B.2) we have $\mathbf{U}_{k+1} - \hat{\mathbf{U}}_{k+1} = \mathsf{A}^k\mathbf{S} \Rightarrow \hat{\mathbf{U}}_{k+1} = \mathbf{U}_{k+1} - \mathsf{A}^k\mathbf{S}$, and since A is a diagonal matrix it follow that $\hat{U}_{i,k+1} = U_{i,k+1} - a_i^k S_i$. At time $k+1$, the MMSE estimate of S_i based on $\hat{U}_{i,k+1}$ is given by:

$$\begin{aligned}\hat{S}_{i,k} &= \frac{\mathbb{E}\left\{S_i(U_{i,k+1} - a_i^k S_i)\right\}}{\mathbb{E}\left\{(U_{i,k+1} - a_i^k S_i)^2\right\}}\hat{U}_{i,k+1} \\ &= \frac{\mathbb{E}\left\{S_i U_{i,k+1}\right\} - a_i^k\sigma_i^2}{\mathbb{E}\{U^2_{i,k+1}\} - 2a_i^k\mathbb{E}\left\{S_i U_{i,k+1}\right\} + a_i^{2k}\sigma_i^2}\hat{U}_{i,k+1}. \end{aligned} \tag{B.4}$$

From the independence of $\mathbf{S}$ and $\mathbf{Z}_k$ we have $\mathbb{E}\left\{\mathbf{U}_{k+1}\mathbf{S}^T\right\} = (\mathsf{A} - \mathbf{B}\mathbf{C}^T)\mathbb{E}\left\{\mathbf{U}_k\mathbf{S}^T\right\}$, and since $\mathbf{U}_1 = \mathbf{S}$ it follows that $\mathbb{E}\left\{\mathbf{U}_{k+1}\mathbf{S}^T\right\} = (\mathsf{A} - \mathbf{B}\mathbf{C}^T)^k\mathsf{Q}_s$. Recalling the definition $\mathsf{M} \triangleq \mathsf{A} - \mathbf{B}\mathbf{C}^T$ we conclude that:

$$\mathbb{E}\left\{S_i U_{i,k+1}\right\} = [\mathsf{M}^k\mathsf{Q}_s]_{i,i}. \tag{B.5}$$

Using the definition of $\mathsf{Q}_{u,k}$ in Subsection IV-C and plugging (B.5) into (B.4) we obtain (24). Next, we use (24) to obtain a recursive expression for the MSE. By plugging the expression for $\hat{S}_{i,k}$ in (B.4) into $\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\}$ we obtain that:

$$\begin{aligned}\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\} &= \sigma_i^2 - \frac{\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i} - \sigma_i^2 a_i^k\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2 a_i^{2k}} \\ &= \frac{\sigma_i^2[\mathsf{Q}_{u,k+1}]_{i,i} - \left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2 a_i^{2k}}, \end{aligned} \tag{B.6}$$

which is Eqn. (25). Finally, we consider (B.6) for $k \to \infty$. As the magnitudes of eigenvalues of the matrix M are smaller than unity it follows that $\lim_{k\to\infty}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2 = 0$ and $\lim_{k\to\infty}[\mathsf{M}^k\mathsf{Q}_s]_{i,i} = 0$. Furthermore, since $|a_i| > 1$ and since $\lim_{k\to\infty}\mathsf{Q}_{u,k} = \mathsf{Q}_u$ it follows that:

$$\lim_{k\to\infty} a_i^{2k}\left(\frac{[\mathsf{Q}_{u,k+1}]_{i,i}}{a_i^{2k}} - 2\frac{[\mathsf{M}^k\mathsf{Q}_s]_{i,i}}{a_i^k} + \sigma_i^2\right) = \sigma_i^2 a_i^{2k}.$$

Therefore, for k large enough we have:

$$\begin{aligned}\frac{\sigma_i^2[\mathsf{Q}_{u,k+1}]_{i,i} - \left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2 a_i^{2k}} &\approx a_i^{-2k}[\mathsf{Q}_{u,k+1}]_{i,i} \\ &= a_i^{-2k}\mathbb{E}\left\{U^2_{i,k+1}\right\}. \end{aligned}$$

B. Proof of Proposition 2

We begin with explicitly writing P_k using $\mathbf{U}_k$:

$$P_k = \mathbb{E}\left\{X_k^2\right\} \overset{(a)}{=} \mathbb{E}\left\{\mathbf{C}^T\mathbf{U}_k\mathbf{U}_k^T\mathbf{C}\right\} = \mathbf{C}^T\mathbb{E}\left\{\mathbf{U}_k\mathbf{U}_k^T\right\}\mathbf{C},$$

where (a) follows from the structure of the controller. Now, recalling that $\mathsf{M} = (\mathsf{A} - \mathbf{B}\mathbf{C}^T)$, we use (B.1) and the fact that $\mathbf{U}_k$ and $\mathbf{Z}_k$ are independent and write:

$$\begin{aligned}\mathbb{E}\left\{\mathbf{U}_k\mathbf{U}_k^T\right\} &= \mathsf{M}\mathbb{E}\left\{\mathbf{U}_{k-1}\mathbf{U}_{k-1}^T\right\}\mathsf{M}^T + \mathsf{Q}_z \\ &= \mathsf{M}\left(\mathsf{M}\mathbb{E}\left\{\mathbf{U}_{k-2}\mathbf{U}_{k-2}^T\right\}\mathsf{M}^T\right)\mathsf{M}^T \\ &\quad + \mathsf{M}\mathsf{Q}_z\mathsf{M}^T + \mathsf{Q}_z \\ &= \mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1} + \sum_{l=0}^{k-2}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l. \end{aligned} \tag{B.7}$$

Therefore, we have:

$$\begin{aligned}P_k &= \mathbf{C}^T\mathbb{E}\left\{\mathbf{U}_k\mathbf{U}_k^T\right\}\mathbf{C} \\ &= \mathbf{C}^T\mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1}\mathbf{C} + \sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C}. \end{aligned} \tag{B.8}$$

Next, we focus on the term $\mathbf{C}^T\mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1}\mathbf{C}$. Since $|\rho_s| < 1$ we can apply Cholesky decomposition [39, Subsection 19.2.1.2] on Q_s and obtain:

$$\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^k)^T = \mathsf{M}^k\mathsf{L}\mathsf{L}^T(\mathsf{M}^k)^T, \quad \mathsf{L} = \begin{bmatrix}\sigma_1 & 0 \\ \rho_s\sigma_2 & \sigma_2\sqrt{1-\rho_s^2}\end{bmatrix}.$$

We now write M^k in terms of the eigenvalues and eigenvectors of M, see [39, Subsection 4.5.2.2]. Let $\mathsf{D} = \mathrm{diag}(\lambda_1, \lambda_2)$ be the diagonal matrix of the eigenvalues of M, while $\mathsf{V} = \begin{bmatrix}v_1 & v_2 \\ v_3 & v_4\end{bmatrix}$ is the matrix whose columns are the corresponding eigenvectors of M. Thus, we have:

$$\mathsf{M} = \mathsf{V}\mathsf{D}\mathsf{V}^{-1} \Rightarrow \mathsf{M}^k = \mathsf{V}\mathsf{D}^k\mathsf{V}^{-1}. \tag{B.9}$$

Next, we define $\mathsf{R} \triangleq \mathsf{V}\mathsf{D}^k\mathsf{V}^{-1}\mathsf{L} = \begin{bmatrix}r_1 & r_2 \\ r_3 & r_4\end{bmatrix}$. Note that R is a function of k, yet, to reduce clutter we omit this notation. This implies that:

$$\begin{aligned}\mathbf{C}^T\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\mathbf{C} &= \mathbf{C}^T\mathsf{R}\mathsf{R}^T\mathbf{C} \\ &= (c_1r_1 + c_2r_3)^2 + (c_1r_2 + c_2r_4)^2. \end{aligned} \tag{B.10}$$

Writing $\mathsf{V}\mathsf{D}^k\mathsf{V}^{-1}$ explicitly we have:

$$\begin{aligned}\mathsf{V}\mathsf{D}^k\mathsf{V}^{-1} &= \begin{bmatrix}v_1 & v_2 \\ v_3 & v_4\end{bmatrix}\begin{bmatrix}\lambda_1^k & 0 \\ 0 & \lambda_2^k\end{bmatrix}\begin{bmatrix}v_1 & v_2 \\ v_3 & v_4\end{bmatrix}^{-1} \\ &= \frac{1}{\det(\mathsf{V})}\begin{bmatrix}v_1v_4\lambda_1^k - v_2v_3\lambda_2^k & v_1v_2(\lambda_2^k - \lambda_1^k) \\ v_3v_4(\lambda_1^k - \lambda_2^k) & v_1v_4\lambda_2^k - v_2v_3\lambda_1^k\end{bmatrix}. \end{aligned} \tag{B.11}$$

Therefore, it follows that:

$$\begin{aligned}\mathsf{R} &= \mathsf{V}\mathsf{D}^k\mathsf{V}^{-1}\mathsf{L}\\ &= \frac{1}{\det(\mathsf{V})}\begin{bmatrix} v_1v_4\lambda_1^k - v_2v_3\lambda_2^k & v_1v_2(\lambda_2^k-\lambda_1^k)\\ v_3v_4(\lambda_1^k-\lambda_2^k) & v_1v_4\lambda_2^k - v_2v_3\lambda_1^k\end{bmatrix}\\ &\qquad\times\begin{bmatrix}\sigma_1 & 0\\ \rho_s\sigma_2 & \sigma_2\sqrt{1-\rho_s^2}\end{bmatrix},\end{aligned}$$

which implies that:

$$r_1 = \frac{\sigma_1(v_1v_4\lambda_1^k - v_2v_3\lambda_2^k) + \rho_s\sigma_2v_1v_2(\lambda_2^k-\lambda_1^k)}{\det(\mathsf{V})} \quad \text{(B.12a)}$$

$$r_2 = \frac{\sigma_2\sqrt{1-\rho_s^2}\cdot v_1v_2\cdot(\lambda_2^k-\lambda_1^k)}{\det(\mathsf{V})} \quad \text{(B.12b)}$$

$$r_3 = \frac{\sigma_1v_3v_4(\lambda_1^k-\lambda_2^k) + \rho_s\sigma_2\left(v_1v_4\lambda_2^k - v_2v_3\lambda_1^k\right)}{\det(\mathsf{V})} \quad \text{(B.12c)}$$

$$r_4 = \frac{\sigma_2\sqrt{1-\rho_s^2}\left(v_1v_4\lambda_2^k - v_2v_3\lambda_1^k\right)}{\det(\mathsf{V})}. \quad \text{(B.12d)}$$

Next, we explicitly write $c_1r_1 + c_2r_3$:

$$\begin{aligned}&c_1r_1+c_2r_3\\ &\quad= \frac{c_1}{\det(\mathsf{V})}\left(\sigma_1(v_1v_4\lambda_1^k - v_2v_3\lambda_2^k)+\rho_s\sigma_2v_1v_2(\lambda_2^k-\lambda_1^k)\right)\\ &\qquad+\frac{c_2}{\det(\mathsf{V})}\left(\sigma_1v_3v_4(\lambda_1^k-\lambda_2^k)+\rho_s\sigma_2\left(v_1v_4\lambda_2^k - v_2v_3\lambda_1^k\right)\right)\\ &= \lambda_1^k\frac{c_1(\sigma_1v_1v_4-\rho_s\sigma_2v_1v_2)+c_2(\sigma_1v_3v_4-\rho_s\sigma_2v_2v_3)}{\det(\mathsf{V})}\\ &\qquad+\lambda_2^k\frac{c_1(\rho_s\sigma_2v_1v_2-\sigma_1v_2v_3)+c_2(\rho_s\sigma_2v_1v_4-\sigma_1v_3v_4)}{\det(\mathsf{V})}\\ &= \lambda_1^k\omega_1(\sigma_1,\sigma_2,\rho_s)+\lambda_2^k\omega_2(\sigma_1,\sigma_2,\rho_s). \quad \text{(B.13a)}\end{aligned}$$

Similarly, we explicitly write $c_1r_2 + c_2r_4$:

$$\begin{aligned}c_1r_2 + c_2r_4 =&\lambda_1^k\frac{-\sigma_2\sqrt{1-\rho_s^2}(c_1v_1v_2 + c_2v_2v_3)}{\det(\mathsf{V})}\\ &+\lambda_2^k\frac{\sigma_2\sqrt{1-\rho_s^2}(c_1v_1v_2 + c_2v_1v_4)}{\det(\mathsf{V})}\\ =&\lambda_1^k\omega_3(\sigma_1,\sigma_2,\rho_s)+\lambda_2^k\omega_4(\sigma_1,\sigma_2,\rho_s), \quad \text{(B.13b)}\end{aligned}$$

where $\omega_j(\varsigma_1,\varsigma_2,\rho), j=1,2,3,4$, are defined in (26). Hence, squaring (B.13a) and (B.13b), summing and using the expressions $\alpha_j, j=1,2,3$, defined in (27) we obtain:

$$\begin{aligned}&(c_1r_1 + c_2r_3)^2 + (c_1r_2 + c_2r_4)^2\\ &\quad= (\lambda_1^k\omega_1(\sigma_1,\sigma_2,\rho_s) + \lambda_2^k\omega_2(\sigma_1,\sigma_2,\rho_s))^2\\ &\qquad+ (\lambda_1^k\omega_3(\sigma_1,\sigma_2,\rho_s) + \lambda_2^k\omega_4(\sigma_1,\sigma_2,\rho_s))^2\\ &\quad= \lambda_1^{2k}\omega_1^2(\sigma_1,\sigma_2,\rho_s)+\lambda_2^{2k}\omega_2^2(\sigma_1,\sigma_2,\rho_s)\\ &\qquad+2\lambda_1^k\lambda_2^k\omega_1(\sigma_1,\sigma_2,\rho_s)\omega_2(\sigma_1,\sigma_2,\rho_s)+\lambda_1^{2k}\omega_3^2(\sigma_1,\sigma_2,\rho_s)\\ &\qquad+\lambda_2^{2k}\omega_4^2(\sigma_1,\sigma_2,\rho_s)+2\lambda_1^k\lambda_2^k\omega_3(\sigma_1,\sigma_2,\rho_s)\omega_4(\sigma_1,\sigma_2,\rho_s)\\ &\quad= \lambda_1^{2k}\left(\omega_1^2(\sigma_1,\sigma_2,\rho_s) + \omega_3^2(\sigma_1,\sigma_2,\rho_s)\right)\\ &\qquad+ \lambda_2^{2k}\left(\omega_2^2(\sigma_1,\sigma_2,\rho_s) + \omega_4^2(\sigma_1,\sigma_2,\rho_s)\right)\\ &\qquad+ \lambda_1^k\lambda_2^k\left(2\omega_1(\sigma_1,\sigma_2,\rho_s)\omega_2(\sigma_1,\sigma_2,\rho_s)\right.\\ &\qquad\qquad\left.+2\omega_3(\sigma_1,\sigma_2,\rho_s)\omega_4(\sigma_1,\sigma_2,\rho_s)\right)\\ &\quad= \lambda_1^{2k}\alpha_1(\sigma_1,\sigma_2,\rho_s)\\ &\qquad+ \lambda_2^{2k}\alpha_2(\sigma_1,\sigma_2,\rho_s) + \lambda_1^k\lambda_2^k\alpha_3(\sigma_1,\sigma_2,\rho_s), \quad \text{(B.14)}\end{aligned}$$

We conclude that:

$$\begin{aligned}&\mathbf{C}^T\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\mathbf{C}\\ &\quad= \lambda_1^{2k}\alpha_1(\sigma_1,\sigma_2,\rho_s)\\ &\qquad+ \lambda_2^{2k}\alpha_2(\sigma_1,\sigma_2,\rho_s) + \lambda_1^k\lambda_2^k\alpha_3(\sigma_1,\sigma_2,\rho_s). \quad \text{(B.15)}\end{aligned}$$

Next, we focus on the second term in (B.8): $\sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C}$. Following identical steps to those leading to (B.15), and recalling that $|\rho_z|<1$, we write:

$$\begin{aligned}&\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C}\\ &\quad= \lambda_1^{2l}\alpha_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\qquad+ \lambda_2^{2l}\alpha_2(\sigma_{z,1},\sigma_{z,2},\rho_z) + \lambda_1^l\lambda_2^l\alpha_3(\sigma_{z,1},\sigma_{z,2},\rho_z).\end{aligned}$$

Therefore, summing over l we obtain:

$$\begin{aligned}&\sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C}\\ &\quad= \sum_{l=0}^{k-2}\lambda_1^{2l}\alpha_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\qquad\qquad+ \lambda_2^{2l}\alpha_2(\sigma_{z,1},\sigma_{z,2},\rho_z) + \lambda_1^l\lambda_2^l\alpha_3(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\quad= \frac{1-\lambda_1^{2(k-1)}}{1-\lambda_1^2}\alpha_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\qquad+ \frac{1-\lambda_2^{2(k-1)}}{1-\lambda_2^2}\alpha_2(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\qquad\qquad+ \frac{1-\lambda_1^{k-1}\lambda_2^{k-1}}{1-\lambda_1\lambda_2}\alpha_3(\sigma_{z,1},\sigma_{z,2},\rho_z). \quad \text{(B.16)}\end{aligned}$$

Combining (B.15) and (B.16) and using the expressions $\eta_j(\varsigma_1,\varsigma_2,\rho), j=1,2,3$, defined in (28), results in:

$$\begin{aligned}P_k &= \mathbf{C}^T\mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1}\mathbf{C} + \sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C}\\ &= \lambda_1^{2(k-1)}\alpha_1(\sigma_1,\sigma_2,\rho_s) + \lambda_2^{2(k-1)}\alpha_2(\sigma_1,\sigma_2,\rho_s)\\ &\quad+ (\lambda_1\lambda_2)^{k-1}\alpha_3(\sigma_1,\sigma_2,\rho_s)\\ &\quad+ \frac{1-\lambda_1^{2(k-1)}}{1-\lambda_1^2}\alpha_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\quad+ \frac{1-\lambda_2^{2(k-1)}}{1-\lambda_2^2}\alpha_2(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\quad+ \frac{1-(\lambda_1\lambda_2)^{k-1}}{1-\lambda_1\lambda_2}\alpha_3(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &= \eta_1(\sigma_{z,1},\sigma_{z,2},\rho_z)+\eta_2(\sigma_{z,1},\sigma_{z,2},\rho_z)+\eta_3(\sigma_{z,1},\sigma_{z,2},\rho_z)\\ &\quad+\lambda_1^{2(k-1)}\left(\alpha_1(\sigma_1,\sigma_2,\rho_s)-\eta_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\right)\\ &\quad+\lambda_2^{2(k-1)}\left(\alpha_2(\sigma_1,\sigma_2,\rho_s)-\eta_2(\sigma_{z,1},\sigma_{z,2},\rho_z)\right)\\ &\quad+(\lambda_1\lambda_2)^{k-1}\left(\alpha_3(\sigma_1,\sigma_2,\rho_s)-\eta_3(\sigma_{z,1},\sigma_{z,2},\rho_z)\right), \quad \text{(B.17)}\end{aligned}$$

From [42, Lemma 2.4] we have $|\lambda_i|<1, i=1,2$,[16] which

[16]Recall that $\mathbf{C} = (\mathbf{B}^T\mathsf{G}\mathbf{B}+1)^{-1}\mathsf{A}\mathsf{G}^T\mathbf{B}$ where G is the unique positive-definite solution of the DARE (19). Now, from [42, Lemma 2.4, item (iv)] it follows that the eigenvalues of the closed-loop matrix $\mathsf{M} = \mathsf{A}-\mathbf{B}\mathbf{C}^T$ are given by $\lambda_i = \frac{1}{a_i}$. Note that [42, Lemma 2.4] assumes a DARE of the form (19) and studies the properties of the matrix $\mathsf{A}-\mathbf{B}\mathbf{C}^T$, for $\mathbf{C} = (\mathbf{B}^T\mathsf{G}\mathbf{B}+1)^{-1}\mathsf{A}\mathsf{G}^T\mathbf{B}$, see [42, Equation below (11)]. Therefore, it follows that $\lambda_1 = \frac{1}{a_1}$.

implies that:

$$\lim_{k\to\infty} \mathbf{C}^T\mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1}\mathbf{C} + \sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C} = \eta_1(\sigma_{z,1},\sigma_{z,2},\rho_z) + \eta_2(\sigma_{z,1},\sigma_{z,2},\rho_z) + \eta_3(\sigma_{z,1},\sigma_{z,2},\rho_z). \quad \text{(B.18)}$$

Recall that the JSCC-LQG scheme is designed such that the asymptotic average transmit power is P. This implies that:

$$\eta_1(\sigma_{z,1},\sigma_{z,2},\rho_z)+\eta_2(\sigma_{z,1},\sigma_{z,2},\rho_z)+\eta_3(\sigma_{z,1},\sigma_{z,2},\rho_z) = P.$$

For the power constraint in (5) to be satisfied for every $k = 1,2,3,\ldots$, we should have $P_k \le P$. From (B.17) we conclude that this condition can be equivalently stated as follows:

$$\lambda_1^{2(k-1)}\big(\alpha_1(\sigma_1,\sigma_2,\rho_s) - \eta_1(\sigma_{z,1},\sigma_{z,2},\rho_z)\big) + \lambda_2^{2(k-1)}\big(\alpha_2(\sigma_1,\sigma_2,\rho_s) - \eta_2(\sigma_{z,1},\sigma_{z,2},\rho_z)\big) + (\lambda_1\lambda_2)^{k-1}\big(\alpha_3(\sigma_1,\sigma_2,\rho_s) - \eta_3(\sigma_{z,1},\sigma_{z,2},\rho_z)\big) \le 0.$$

C. Proof of Theorem 3

We begin with $K_{\text{LQG}}^{\text{ub}}$. Since (24) is the optimal estimator based on the observation $\hat{U}_{i,k+1}$, it follows that we can upper bound K_{LQG} by upper bounding the number of channel uses required to achieve a target MSE pair using the decoder in (22). Recall that the MSE of the decoder in (22) is given by (23): $\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\} = a_i^{-2k}\mathbb{E}\{U_{i,k+1}^2\}$. Let $\mathbb{E}\{(S_i - \hat{S}_{i,k})^2\} \triangleq D_{i,k}$ be the MSE after k channel uses, i.e., at time instance $k+1$. We upper bound $D_{i,k}$ via upper bounding $\mathbb{E}\{U_{i,k+1}^2\}$.

Since the eigenvalues of M are inside the unit circle, it follows that $\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{i,i} \to 0$ as $k\to\infty$, and therefore, $\lim_{k\to\infty}\left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{i,i} = [\mathsf{Q}_u]_{i,i}$.[17] Since Q_s is a covariance matrix then the diagonal elements of $\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k$ are non-negative and we can write:

$$\mathbb{E}\left\{U_{i,k+1}^2\right\} \le \left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{i,i} + [\mathsf{Q}_u]_{i,i}.$$

Next, we derive an upper bound on $\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{i,i}$. Following the arguments leading to (B.10) we can write $\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^k)^T = \mathsf{R}\mathsf{R}^T$, again omitting the dependence in k from the matrix R, we write:

$$\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{1,1} = r_1^2 + r_2^2,$$

where r_1 and r_2 are given in (B.12). For ease of reference we repeat the expressions for r_1 and r_2:

$$r_1 = \frac{1}{\det(\mathsf{V})}\left(\sigma_1(v_1v_4\lambda_1^k - v_2v_3\lambda_2^k) + \rho_s\sigma_2v_1v_2(\lambda_2^k - \lambda_1^k)\right)$$
$$r_2 = \frac{1}{\det(\mathsf{V})}\sigma_2\sqrt{1-\rho_s^2}\cdot v_1v_2\cdot(\lambda_2^k - \lambda_1^k).$$

[17] Note that $\left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{i,i} \ge 0, k = 1,2,\ldots, i = 1,2$, since the diagonal elements are sum of the variances of the noise.

Next, we upper bound $\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{1,1}$ via upper bounding r_1^2 and r_2^2:

$$\begin{aligned}|r_1| &\le \frac{1}{|\det(\mathsf{V})|}\left(\sigma_1\left(|v_1v_4||\lambda_1|^k + |v_2v_3||\lambda_2|^k\right)\right.\\ &\qquad \left.+|\rho_s|\cdot\sigma_2\cdot|v_1v_2|(|\lambda_2|^k + |\lambda_1|^k)\right)\\ &\overset{(a)}{\le} \frac{\sigma_1\left(|v_1v_4\lambda_1| + |v_2v_3\lambda_2|\right) + |\rho_s\sigma_2v_1v_2|(|\lambda_2| + |\lambda_1|)}{|\det(\mathsf{V})|}\\ &\triangleq \tau_1\end{aligned}$$

where (a) follows from the fact that $|\lambda_i| < 1, i = 1,2$. Using similar arguments we bound $|r_2|$ as follows:

$$|r_2| \le \frac{\sigma_2\sqrt{1-\rho_s^2}|v_1v_2|(|\lambda_2| + |\lambda_1|)}{|\det(\mathsf{V})|} \triangleq \tau_2.$$

Hence, we have $\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{1,1} \le \tau_1^2 + \tau_2^2$, and this implies that:

$$\mathbb{E}\left\{U_{1,k}^2\right\} \le \tau_1^2 + \tau_2^2 + [\mathsf{Q}_u]_{1,1} \triangleq \vartheta_1.$$

Following similar arguments we have $\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{2,2} \le \tau_3^2 + \tau_4^2$, where:

$$\tau_3 \triangleq \frac{|\sigma_1v_3v_4|(|\lambda_1| + |\lambda_2|) + |\rho_s\sigma_2|\left(|v_1v_4\lambda_2| + |v_2v_3\lambda_1|\right)}{|\det(\mathsf{V})|}$$
$$\tau_4 \triangleq \frac{\sigma_2\sqrt{1-\rho_s^2}\left(|v_1v_4\lambda_2| + |v_2v_3\lambda_1|\right)}{|\det(\mathsf{V})|},$$

and therefore:

$$\mathbb{E}\left\{U_{2,k}^2\right\} \le \tau_3^2 + \tau_4^2 + [\mathsf{Q}_u]_{2,2} \triangleq \vartheta_2.$$

To conclude, we have:

$$K_{\text{LQG}} \le \frac{\log\left(\frac{\vartheta_i}{D_i}\right)}{2\log|a_i|}, \quad i = 1,2.$$

To lower bound K_{LQG} we first lower bound the MSE in (25) as follows:

$$\begin{aligned}\mathbb{E}\left\{(S_i - \hat{S}_{i,k})^2\right\} &= \sigma_i^2\frac{[\mathsf{Q}_{u,k+1}]_{i,i} - \frac{1}{\sigma_i^2}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2a_i^{2k}}\\ &\overset{(a)}{\ge} \sigma_i^2\frac{[\mathsf{Q}_z]_{i,i} - \frac{1}{\sigma_i^2}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_z]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2a_i^{2k}}.\end{aligned} \quad \text{(B.19)}$$

To see why step (a) holds we note that:

$$0 \le \frac{[\mathsf{Q}_{u,k+1}]_{i,i} - \frac{1}{\sigma_i^2}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2a_i^{2k}} \le 1.$$

This follows as $\mathbb{E}\left\{(S_i - \hat{S}_{i,k})^2\right\} = \sigma_i^2 - \mathbb{E}\left\{\hat{S}_{i,k}^2\right\}$, and therefore we have:

$$\sigma_i^2\left(1 - \frac{[\mathsf{Q}_{u,k+1}]_{i,i} - \frac{1}{\sigma_i^2}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2a_i^{2k}}\right) = \mathbb{E}\left\{\hat{S}_{i,k}^2\right\} \ge 0,$$

which implies that $\frac{[\mathsf{Q}_{u,k+1}]_{i,i} - \frac{1}{\sigma_i^2}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}]_{i,i} - 2a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i} + \sigma_i^2a_i^{2k}} \le 1$. Next, consider the function $f(x) = \frac{x+a}{x+b}, x > 0, a, b \in \mathfrak{R}$. We now show that if $0 \le f(x) \le 1$, then $f(x)$ is an increasing function.

The derivative of $f(x)$ is given by: $f'(x)=\frac{b-a}{(x+b)^2}$. Consider the following cases:

- $a,b\geq 0$: From the fact that $f(x)\leq 1$, it follows that $b\geq a$, and therefore $f'(x)\geq 0$.
- $b\geq 0, a\leq 0$: For this case, it is clear that $f'(x)\geq 0$.
- $a,b\leq 0$: From the fact that $f(x)\leq 1$, it follows that $a\leq b$, and therefore $f'(x)\geq 0$.
- $b<0, a>0$: This assignment is not valid since $f(x)\leq 1$.

We conclude that for all valid cases $f'(x)\geq 0$ which implies that $f(x)$ is an increasing function. With this in mind we note that $[\mathsf{Q}_{u,k+1}]_{i,i}\geq[\mathsf{Q}_z]_{i,i}$,[18] which concludes the proof of step (a) in (B.19).

Next, we lower bound the numerator of (B.19) and upper bound the denominator of (B.19). Recall that $\mathsf{M}^k=(\mathsf{A}-\mathbf{B}\mathbf{C}^T)^k$ and consider upper bounding $[\mathsf{M}^k\mathsf{Q}_s]_{1,1}$. Similarly to Section B-B we write M^k in terms of the eigenvalues matrix D and the eigenvectors matrix V of M as in (B.9): $\mathsf{M}^k=\mathsf{V}\mathsf{D}^k\mathsf{V}^{-1}$. From (B.11) we have:

$$\begin{aligned}&\mathsf{V}\mathsf{D}^k\mathsf{V}^{-1}\mathsf{Q}_s\\&=\frac{1}{\det(\mathsf{V})}\begin{bmatrix}v_1v_4\lambda_1^k-v_2v_3\lambda_2^k & v_1v_2(\lambda_2^k-\lambda_1^k)\\ v_3v_4(\lambda_1^k-\lambda_2^k) & v_1v_4\lambda_2^k-v_2v_3\lambda_1^k\end{bmatrix}\\&\qquad\times\begin{bmatrix}\sigma_1^2 & \rho_s\sigma_1\sigma_2\\ \rho_s\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix},\end{aligned}$$

from which we compute:

$$[\mathsf{M}^k\mathsf{Q}_s]_{1,1}=\frac{\sigma_1^2\left(v_1v_4\lambda_1^k-v_2v_3\lambda_2^k\right)+\rho_s\sigma_1\sigma_2v_1v_2(\lambda_2^k-\lambda_1^k)}{\det(\mathsf{V})}. \tag{B.20}$$

Using the fact that $|\lambda_i|<1, i=1,2$, we obtain the following upper bound on $[\mathsf{M}^k\mathsf{Q}_s]_{1,1}$, $k\geq 1$:

$$\begin{aligned}&\left|[\mathsf{M}^k\mathsf{Q}_s]_{1,1}\right|\\&\leq\frac{\sigma_1^2\left(|v_1v_4\lambda_1|+|v_2v_3\lambda_2|\right)+|\rho_s\sigma_1\sigma_2v_1v_2|(|\lambda_2|+|\lambda_1|)}{|\det(\mathsf{V})|}\\&\triangleq\beta_1.\end{aligned}$$

Similarly, we also bound:

$$\begin{aligned}&\left|[\mathsf{M}^k\mathsf{Q}_s]_{2,2}\right|\\&\leq\frac{\sigma_2^2\left(|v_1v_4\lambda_2|+|v_2v_3\lambda_1|\right)+|\rho_s\sigma_1\sigma_2v_2v_4|(|\lambda_2|+|\lambda_1|)}{|\det(\mathsf{V})|}\\&\triangleq\beta_2.\end{aligned}$$

Now, for $i=1,2$, plugging β_i into (B.19) and setting $D_{i,k}=D_i$ we write:

$$D_i\geq\frac{\sigma_i^2[\mathsf{Q}_z]_{i,i}-\beta_i^2}{[\mathsf{Q}_z]_{i,i}+2|a_i|^k\beta_i+\sigma_i^2|a_i|^{2k}},$$

which can also be written as:

$$\begin{aligned}&D_i[\mathsf{Q}_z]_{i,i}-\sigma_i^2[\mathsf{Q}_z]_{i,i}+\beta_i^2\geq-D_i\left(2|a_i|^k\beta_i+\sigma_i^2|a_i|^{2k}\right),\\&\Rightarrow\frac{\sigma_i^2[\mathsf{Q}_z]_{i,i}-\beta_i^2-D_i[\mathsf{Q}_z]_{i,i}}{D_i}\leq 2|a_i|^k\beta_i+\sigma_i^2|a_i|^{2k}.\end{aligned}$$

[18]From (B.7) it follows that $\mathbb{E}\{U_{i,k}^2\}\geq[\mathsf{Q}_z]_{i,i}$.

Next, we recall that $|a_i|>1$ and write:

$$\frac{\sigma_i^2[\mathsf{Q}_z]_{i,i}-\beta_i^2-D_i[\mathsf{Q}_z]_{i,i}}{D_i}\leq(2\beta_i+\sigma_i^2)|a_i|^{2k}.$$

Applying the log to both sides we have:

$$\begin{aligned}&\log\left(\frac{\sigma_i^2[\mathsf{Q}_z]_{i,i}-\beta_i^2-D_i[\mathsf{Q}_z]_{i,i}}{D_i}\right)\\&\qquad\leq\log\left((2\beta_i+\sigma_i^2)|a_i|^{2k}\right),\end{aligned}$$

which can be written as:

$$\log\left(\frac{\sigma_i^2[\mathsf{Q}_z]_{i,i}-\beta_i^2-D_i[\mathsf{Q}_z]_{i,i}}{(2\beta_i+\sigma_i^2)D_i}\right)\leq 2k\log|a_i|.$$

Thus, we write:

$$\frac{\log\left(\frac{\sigma_i^2[\mathsf{Q}_z]_{i,i}-\beta_i^2-D_i[\mathsf{Q}_z]_{i,i}}{(2\beta_i+\sigma_i^2)D_i}\right)}{2\log|a_i|}\leq K_{\mathrm{LQG}},$$

which is stated in (30b).

Appendix C
Proofs for the JSCC-LQG Scheme for the Symmetric Setting

A. Proof of Theorem 4

We begin with the following lemma:

Lemma 1. For symmetric GBCFs $c_2=-c_1$.

Proof. We explicitly express c_1 in terms of a_1. Recall the definition of the vector $\mathbf{C}$ in Section IV-A:

$$\mathbf{C}=(\mathbf{B}^T\mathsf{G}\mathbf{B}+1)^{-1}\mathsf{A}\mathsf{G}^T\mathbf{B} \tag{C.1}$$

where G is the unique positive-definite solution of the DARE $\mathsf{G}=\mathsf{A}^T\mathsf{G}\mathsf{A}-\mathsf{A}^T\mathsf{G}\mathbf{B}(\mathbf{B}^T\mathsf{G}\mathbf{B}+1)^{-1}\mathbf{B}^T\mathsf{G}\mathsf{A}$, such that all the eigenvalues of the matrix $\mathsf{A}-\mathbf{B}\mathbf{C}^T$ have magnitudes smaller than 1. Let $\mathsf{G}=\begin{bmatrix}g_1 & g_2\\ g_3 & g_4\end{bmatrix}$. From [44, Prop. 1] we have that for the symmetric case and for $\mathsf{A}=\begin{bmatrix}a_1 & 0\\ 0 & -a_1\end{bmatrix}$ the elements of G are given by:

$$\begin{aligned}g_1=g_4&=\frac{(a_1^2-1)(1+a_1^2)^2}{4a_1^2},\\g_2=g_3&=\frac{(1-a_1^2)^2(1+a_1^2)}{4a_1^2},\end{aligned}$$

and it follows that $\mathsf{G}=\mathsf{G}^T$. Writing $\mathsf{A}\mathsf{G}^T\mathbf{B}=\mathsf{A}\mathsf{G}\mathbf{B}$ explicitly:

$$\begin{aligned}\mathsf{A}\mathsf{G}\mathbf{B}&=\begin{bmatrix}a_1 & 0\\ 0 & -a_1\end{bmatrix}\begin{bmatrix}g_1 & g_2\\ g_2 & g_1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\&=\begin{bmatrix}a_1(g_1+g_2)\\ -a_1(g_1+g_2)\end{bmatrix}.\end{aligned}$$

Using the explicit expressions for g_1 and g_2 we can write $g_1+g_2=\frac{(1+a_1^2)(a_1^2-1)}{2}$. Next, writing $\mathbf{B}^T\mathsf{G}\mathbf{B}+1$ explicitly we obtain:

$$\mathbf{B}^T\mathsf{G}\mathbf{B}+1=2(g_1+g_2)+1=(1+a_1^2)(a_1^2-1)+1.$$

We now can explicitly compute c_1, the first element of $\mathbf{C}$ in (C.1):

$$\begin{aligned} c_1 &= \frac{a_1(g_1+g_2)}{2(g_1+g_2)+1} \\ &= \frac{a_1(1+a_1^2)(a_1^2-1)}{2((1+a_1^2)(a_1^2-1)+1)} \\ &= \frac{a_1^4-1}{2a_1^3} \end{aligned} \tag{C.2}$$

Computing c_2 via similar arguments we find $c_2 = -c_1$. □

Next, we recall (B.8):

$$P_k = \mathbf{C}^T \mathsf{M}^{k-1} \mathsf{Q}_s (\mathsf{M}^T)^{k-1} \mathbf{C} + \sum_{l=0}^{k-2} \mathbf{C}^T \mathsf{M}^l \mathsf{Q}_z (\mathsf{M}^T)^l \mathbf{C},$$

and note that with $c_2 = -c_1$ (B.10) is specialized to:

$$\mathbf{C}^T \mathsf{M}^k \mathsf{Q}_s (\mathsf{M}^T)^k \mathbf{C} = c_1^2 \left((r_1 - r_3)^2 + (r_2 - r_4)^2 \right). \tag{C.3}$$

In the symmetric setting we also have $\sigma_1 = \sigma_2 = \sigma_s$ and $\sigma_{z,1} = \sigma_{z,2} = \sigma_z$. From the expression for the matrix M and from the expression for c_1 it follows that $v_1 = v_4, v_2 = v_3$, and $-\lambda_2 = \lambda_1$. Therefore, (B.12) is specialized to:

$$r_1 = \frac{\lambda_1^k \sigma_s}{\det(\mathsf{V})} \left(v_1^2 - v_2^2(-1)^k + \rho_s v_1 v_2 ((-1)^k - 1) \right) \tag{C.4a}$$

$$r_2 = \frac{\lambda_1^k \sigma_s \sqrt{1-\rho_s^2}}{\det(\mathsf{V})} \left(v_1 v_2 ((-1)^k - 1) \right) \tag{C.4b}$$

$$r_3 = \frac{\lambda_1^k \sigma_s}{\det(\mathsf{V})} \left(v_1 v_2 (1-(-1)^k) + \rho_s \left(v_1^2 (-1)^k - v_2^2 \right) \right) \tag{C.4c}$$

$$r_4 = \frac{\lambda_1^k \sigma_s \sqrt{1-\rho_s^2}}{\det(\mathsf{V})} \left(v_1^2 (-1)^k - v_2^2 \right). \tag{C.4d}$$

Next, we explicitly write $r_1 - r_3$:

$$\begin{aligned} r_1 - r_3 = \frac{\lambda_1^k \sigma_s}{\det(\mathsf{V})} & \left(v_1^2 - v_2^2(-1)^k + \rho_s v_1 v_2 ((-1)^k - 1) \right. \\ & \left. - v_1 v_2 (1-(-1)^k) - \rho_s \left(v_1^2 (-1)^k - v_2^2 \right) \right) \\ = \frac{\lambda_1^k \sigma_s}{\det(\mathsf{V})} & \left(v_1^2 (1-\rho_s(-1)^k) + v_2^2 (\rho_s - (-1)^k) \right. \\ & \left. + v_1 v_2 (\rho_s + 1)((-1)^k - 1) \right), \end{aligned}$$

and by squaring we obtain:

$$\begin{aligned} & (r_1 - r_3)^2 \\ &= \frac{\lambda_1^{2k} \sigma_s^2}{\det^2(\mathsf{V})} \left(v_1^2 (1 - \rho_s(-1)^k) + v_2^2 (\rho_s - (-1)^k) \right. \\ &\qquad \left. + v_1 v_2 (\rho_s + 1)((-1)^k - 1) \right)^2 \\ &= \frac{\lambda_1^{2k} \sigma_s^2}{\det^2(\mathsf{V})} \Bigg(v_1^4 (1-\rho_s(-1)^k)^2 + v_2^4 (\rho_s - (-1)^k)^2 \\ &\qquad + v_1^2 v_2^2 (1+\rho_s)^2 ((-1)^k - 1)^2 \\ &\qquad + 2 v_1^2 v_2^2 (1 - \rho_s(-1)^k)(\rho_s - (-1)^k) \\ &\qquad + 2 v_1^3 v_2 (1 - \rho_s(-1)^k)(\rho_s + 1)((-1)^k - 1) \\ &\qquad + 2 v_1 v_2^3 (\rho_s - (-1)^k)(\rho_s + 1)((-1)^k - 1) \Bigg). \end{aligned}$$

Now, for *even* values of k we have:

$$\begin{aligned} & (r_1 - r_3)^2 \\ &= \frac{\lambda_1^{2k} \sigma_s^2}{\det^2(\mathsf{V})} \left(v_1^4 (1 - \rho_s(-1)^k)^2 \right. \\ &\qquad \left. + v_2^4 (\rho_s - (-1)^k)^2 - 2 v_1^2 v_2^2 (1 - \rho_s)^2 \right) \\ &= \frac{\lambda_1^{2k} \sigma_s^2}{\det^2(\mathsf{V})} \left((1-\rho_s)^2 (v_1^2 - v_2^2)^2 \right) \\ &= \lambda_1^{2k} \sigma_s^2 (1-\rho_s)^2, \end{aligned} \tag{C.5}$$

while for *odd* values of k we have:

$$\begin{aligned} & (r_1 - r_3)^2 \\ &= \frac{\lambda_1^{2k} \sigma_s^2}{\det^2(\mathsf{V})} \left(v_1^4 (1+\rho_s)^2 + v_2^4 (1+\rho_s)^2 \right. \\ &\qquad + 4 v_1^2 v_2^2 (1+\rho_s)^2 + 2 v_1^2 v_2^2 (1+\rho_s)^2 \\ &\qquad \left. - 4 v_1^3 v_2 (1+\rho_s)^2 - 4 v_1 v_2^3 (1+\rho_s)^2 \right) \\ &= \lambda_1^{2k} \sigma_s^2 (1+\rho_s)^2 \frac{v_1^4 + v_2^4 + 6 v_1^2 v_2^2 - 4 v_1 v_2 (v_1^2 + v_2^2)}{\det^2(\mathsf{V})} \\ &\overset{(a)}{=} \lambda_1^{2k} \sigma_s^2 (1+\rho_s)^2 a_1^4, \end{aligned} \tag{C.6}$$

where (a) follows from the following lemma.

Lemma 2. The following equality holds: $\frac{v_1^4 + v_2^4 + 6 v_1^2 v_2^2 - 4 v_1 v_2 (v_1^2 + v_2^2)}{\det^2(\mathsf{V})} = a_1^4$.

Proof. We begin with expressing λ_1, v_1, and v_2. From [42, Lemma 2.4] it follows that $\lambda_1 = \frac{1}{a_1}$, see Footnote 16 for a detailed explanation. Next, we explicitly write $\mathsf{M} = \begin{bmatrix} a_1 - c_1 & c_1 \\ -c_1 & -(a_1 - c_1) \end{bmatrix}$, and note that an eiegenvector $\mathbf{V}_0$ of M, corresponding to the eigenvalue λ_1, obeys $\mathsf{M}\mathbf{V}_0 = \lambda_1 \mathbf{V}_0$. This equation can also be written using a matrix form:

$$(\mathsf{M} - \lambda_1 \mathsf{I}) \mathbf{V}_0 = \begin{bmatrix} a_1 - c_1 - \lambda_1 & c_1 \\ -c_1 & -(a_1 - c_1) - \lambda_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0}.$$

Recalling that eigenvectors have unit norm, we obtain an explicit expression for $\mathbf{V}_0$:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{c_1}{\sqrt{c_1^2 + (a_1 - c_1 - \lambda_1)^2}} \\ -\frac{a_1 - c_1 - \lambda_1}{\sqrt{c_1^2 + (a_1 - c_1 - \lambda_1)^2}} \end{bmatrix}.$$

Substituting $\lambda_1 = \frac{1}{a_1}$ we obtain:

$$\begin{aligned} v_1 &= \frac{c_1}{\sqrt{c_1^2 + (a_1 - c_1 - \frac{1}{a_1})^2}} \\ &= \frac{a_1 c_1}{\sqrt{a_1^2 c_1^2 + ((a_1 - c_1) a_1 - 1)^2}} \end{aligned} \tag{C.7a}$$

$$\begin{aligned} v_2 &= -\frac{a_1 - c_1 - \frac{1}{a_1}}{\sqrt{c_1^2 + (a_1 - c_1 - \frac{1}{a_1})^2}} \\ &= \frac{1 - (a_1 - c_1) a_1}{\sqrt{a_1^2 c_1^2 + ((a_1 - c_1) a_1 - 1)^2}}. \end{aligned} \tag{C.7b}$$

$$\frac{v_1^4+v_2^4+6v_1^2v_2^2-4v_1v_2(v_1^2+v_2^2)}{\det^2(\mathsf{V})} = \frac{(a_1^4-1)^4+(a_1^2-1)^8+6(a_1^4-1)^2(a_1^2-1)^4+4(a_1^4-1)(a_1^2-1)^2((a_1^4-1)^2+(a_1^2-1)^4)}{((a_1^4-1)^2-(a_1^2-1)^4)^2}. \tag{C.9}$$

Note that (C.2) implies that $0 \le c_1 \le a_1$. Using the expression for c_1 we now write $a_1(a_1-c_1)-1$ in terms of a_1:

$$\begin{aligned} a_1(a_1-c_1)-1 &= a_1\left(a_1-\frac{(a_1^4-1)}{2a_1^3}\right)-1 \\ &= \frac{a_1^5+a_1}{2a_1^3}-1 \\ &= \frac{(a_1^2-1)^2}{2a_1^2}. \end{aligned} \tag{C.8}$$

Thus, the numerator of (C.7a) equals $a_1c_1 = \frac{a_1^4-1}{2a_1^2}$, while the numerator of (C.7b) equals $1-(a_1-c_1)a_1 = -\frac{(a_1^2-1)^2}{2a_1^2}$. We further note that the denominators of (C.7a) and (C.7b) are the same. Therefore, we write (C.9) at the top of the page. The denominator of (C.9) can be written as:

$$\begin{aligned} ((a_1^4-1)^2-(a_1^2-1)^4)^2 &= ((a_1^2-1)^2(a_1^2+1)^2-(a_1^2-1)^4)^2 \\ &= 16a_1^4(a_1^2-1)^4. \end{aligned} \tag{C.10a}$$

The numerator of (C.9) can be written as:

$$\begin{aligned} &(a_1^4-1)^4+(a_1^2-1)^8+6(a_1^4-1)^2(a_1^2-1)^4 \\ &\quad+4(a_1^4-1)(a_1^2-1)^2((a_1^4-1)^2+(a_1^2-1)^4) \\ &\quad= (a_1^2-1)^4\left(8a_1^8+8+8(a_1^8-1)\right) \\ &\quad= 16a_1^8(a_1^2-1)^4. \end{aligned} \tag{C.10b}$$

Thus, by combining (C.10a) and (C.10b) we obtain:

$$\begin{aligned} \frac{v_1^4+v_2^4+6v_1^2v_2^2-4v_1v_2(v_1^2+v_2^2)}{\det^2(\mathsf{V})} &= \frac{16a_1^8(a^2-1)^4}{16a_1^4(a^2-1)^4} \\ &= a_1^4. \end{aligned} \tag{C.11}$$

This concludes the proof of the lemma. □

Similarly to (C.6), we write:

$$\begin{aligned} &(r_2-r_4)^2 \\ &\quad= \frac{\lambda_1^{2k}\sigma_s^2(1-\rho_s^2)}{\det^2(\mathsf{V})}\left(v_1v_2((-1)^k-1)-v_1^2(-1)^k+v_2^2\right)^2 \\ &\quad= \frac{\lambda_1^{2k}\sigma_s^2(1-\rho_s^2)}{\det^2(\mathsf{V})}\big(v_1^2v_2^2((-1)^k-1)^2+v_1^4+v_2^4 \\ &\qquad\qquad -2v_1^3v_2(-1)^k((-1)^k-1) \\ &\qquad\qquad +2v_1v_2^3((-1)^k-1)-2v_1^2v_2^2(-1)^k\big) \\ &\quad= \begin{cases} \lambda_1^{2k}\sigma_s^2(1-\rho_s^2), & k \text{ is even}, \\ \lambda_1^{2k}\sigma_s^2(1-\rho_s^2)a_1^4, & k \text{ is odd}. \end{cases} \end{aligned} \tag{C.12}$$

Hence, combining (C.3) and (C.5)–(C.12) we obtain:

$$\begin{aligned} &\mathbf{C}^T\mathsf{M}^{k-1}\mathsf{Q}_s(\mathsf{M}^T)^{k-1}\mathbf{C} \\ &\quad= \begin{cases} 2c_1^2\lambda_1^{2(k-1)}\sigma_s^2(1-\rho_s), & k-1 \text{ is even}, \\ 2c_1^2\lambda_1^{2(k-1)}\sigma_s^2(1+\rho_s)a_1^4, & k-1 \text{ is odd}. \end{cases} \end{aligned} \tag{C.13}$$

Next, we focus on $\sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C}$. Following the steps leading to (C.13) we write:

$$\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C} = \begin{cases} 2c_1^2\lambda_1^{2l}\sigma_z^2(1-\rho_z), & l \text{ is even}, \\ 2c_1^2\lambda_1^{2l}\sigma_z^2(1+\rho_z)a_1^4, & l \text{ is odd}. \end{cases}$$

For even values of $k-1$ we have:

$$\begin{aligned} &\sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C} \\ &\quad= \sum_{m=0}^{\frac{k-1}{2}-1}2c_1^2\sigma_z^2(1-\rho_z)\lambda_1^{4m}+\sum_{m=0}^{\frac{k-1}{2}-1}2c_1^2\sigma_z^2(1+\rho_z)a_1^4\lambda_1^{4m+2} \\ &\quad= 2c_1^2\sigma_z^2((1-\rho_z)+(1+\rho_z)a_1^4\lambda_1^2)\sum_{m=0}^{\frac{k-1}{2}-1}\lambda_1^{4m} \\ &\quad\overset{(a)}{=} \frac{2c_1^2\sigma_z^2((1-\rho_z)+(1+\rho_z)a_1^2)}{1-\lambda_1^4}(1-\lambda_1^{2(k-1)}) \\ &\quad= \mu_1\cdot(1-\lambda_1^{2(k-1)}), \end{aligned} \tag{C.14}$$

where (a) follows from the fact that $a_1 = \frac{1}{\lambda_1}$. For odd values of $k-1$ we have:

$$\begin{aligned} &\sum_{l=0}^{k-2}\mathbf{C}^T\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\mathbf{C} \\ &\quad= \sum_{m=0}^{\frac{k-2}{2}}2c_1^2\sigma_z^2(1-\rho_z)\lambda_1^{4m}+\sum_{m=0}^{\frac{k-2}{2}-1}2c_1^2\sigma_z^2(1+\rho_z)a_1^4\lambda_1^{4m+2} \\ &\quad= 2c_1^2\sigma_z^2(1-\rho_z)\lambda_1^{2(k-2)}+2c_1^2\sigma_z^2((1-\rho_z) \\ &\qquad\quad+(1+\rho_z)a_1^4\lambda_1^2)\sum_{m=0}^{\frac{k-2}{2}-1}\lambda_1^{4m} \\ &\quad= \frac{(1-\lambda_1^4)2c_1^2\sigma_z^2(1-\rho_z)\lambda_1^{2(k-2)}+2c_1^2\sigma_z^2((1-\rho_z)}{1-\lambda_1^4} \\ &\qquad\quad+\frac{(1+\rho_z)a_1^4\lambda_1^2)(1-\lambda_1^{2(k-2)})}{1-\lambda_1^4} \\ &\quad= \frac{2c_1^2\sigma_z^2(1-\rho_z)\lambda_1^{2(k-2)}}{1-\lambda_1^4}-\frac{2c_1^2\sigma_z^2(1-\rho_z)\lambda_1^{2k}}{1-\lambda_1^4} \\ &\qquad\quad+\frac{2c_1^2\sigma_z^2((1-\rho_z)+(1+\rho_z)a_1^4\lambda_1^2)}{1-\lambda_1^4} \\ &\qquad\quad-\frac{2c_1^2\sigma_z^2((1-\rho_z)+(1+\rho_z)a_1^4\lambda_1^2)\lambda_1^{2(k-2)}}{1-\lambda_1^4} \\ &\quad= \frac{2c_1^2\sigma_z^2((1-\rho_z)+(1+\rho_z)a_1^2)}{1-\lambda_1^4} \\ &\qquad\quad-\frac{2c_1^2\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{1-\lambda_1^4}\lambda_1^{2(k-1)} \\ &\quad= \mu_1-\mu_3\cdot\lambda_1^{2(k-1)}. \end{aligned} \tag{C.15}$$

Recalling that $\mu_0 = 2c_1^2\sigma_s^2(1-\rho_s)$ and $\mu_2 = 2c_1^2\sigma_s^2(1+\rho_s)a_1^4$, we combine (C.13)–(C.15) to obtain (32). Similarly to (B.18) we have that $\lim_{k\to\infty} P_k = P$, and since $|\lambda_1| < 1$ it follows that $\mu_1 = P$. Therefore, the power constraint (5) is satisfied *if and only if* $\mu_0 \le \mu_1$ and $\mu_2 \le \mu_3$.

B. Proof of Proposition 3

First, we show that the maximal possible scaling which satisfies (5) is: $\sqrt{\frac{\nu}{\sigma_s^2}}$. Then, we prove that the optimal estimator and the obtained MSE are given in (34) and (35), respectively. Finally, we show that setting $\gamma = \frac{\nu}{\sigma_s^2}$ indeed minimizes the MSE.

1) Maximal Scaling: Recall that (33) constitutes an upper bound on the variance of the sources transmitted via a JSCC-LQG scheme initialized with $\mathbf{U}_1 = \mathbf{S}$, which satisfy (5). Explicitly writing the conditions of Thm. 4, i.e., $\mu_0 \le \mu_1$ and $\mu_2 \le \mu_3$, where $\mu_j, j = 0,\ldots,3$ are defined in (31), we obtain:

$$\sigma_s^2(1-\rho_s) \le \frac{\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{1-\lambda_1^4},$$
$$\sigma_s^2(1+\rho_s)a_1^4 \le \frac{\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{1-\lambda_1^4}.$$

This implies that:

$$\sigma_s^2 \le \min\left\{\frac{\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{(1-\lambda_1^4)(1-\rho_s)}, \frac{\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{(1-\lambda_1^4)(1+\rho_s)a_1^4}\right\}, \quad \text{(C.16)}$$

and therefore, the maximal possible scaling which satisfies (5) is: $\sqrt{\frac{\nu}{\sigma_s^2}}$.

2) Optimal Estimator and Resulting MSE: Following the same arguments as those applied in Appendix B-A the optimal estimator of S_i based on the observation $\hat{U}_{i,k+1}(\gamma)$ is given by $\mathbb{E}\left\{S_i|\hat{U}_{i,k}(\gamma)\right\} = \frac{\mathbb{E}\{S_i\hat{U}_{i,k}(\gamma)\}}{\mathbb{E}\{\hat{U}_{i,k}^2(\gamma)\}}\hat{U}_{i,k}$. Letting $\tilde{\mathbf{S}} = \sqrt{\gamma}\cdot\mathbf{S}$ we can write:

$$\mathbb{E}\left\{S_i|\hat{U}_{i,k}(\gamma)\right\} = \frac{1}{\sqrt{\gamma}}\mathbb{E}\left\{\tilde{S}_i|\hat{U}_{i,k}(\gamma)\right\}.$$

Note that $\mathbb{E}\left\{\tilde{S}_i|\hat{U}_{i,k}(\gamma)\right\}$ can be obtained from (24) by setting $\sigma_i^2 = \gamma\cdot\sigma_s^2$. Let $\mathsf{Q}_{\tilde{s}} \triangleq \mathbb{E}\{\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T\}$. Following the arguments leading to (B.4) we write:

$$\begin{aligned}\mathbb{E}&\left\{S_i|\hat{U}_{i,k}(\gamma)\right\}\\ &= \frac{1}{\sqrt{\gamma}}\cdot\frac{[\mathsf{M}^k\mathsf{Q}_{\tilde{s}}]_{i,i}-\sigma_{\tilde{s}}^2a_i^k}{[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i}-2a_i^k[\mathsf{M}^k\mathsf{Q}_{\tilde{s}}]_{i,i}+\sigma_{\tilde{s}}^2a_i^{2k}}\hat{U}_{i,k+1}(\gamma)\\ &= \frac{\sqrt{\gamma}\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}-\sigma_s^2a_i^k\right)}{[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i}-2\gamma a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i}+\gamma\sigma_s^2a_i^{2k}}\hat{U}_{i,k+1}(\gamma).\end{aligned}$$

Moreover, by following the arguments leading to (B.6) we obtain (35):

$$\mathbb{E}\left\{(S_i-\hat{S}_{i,k})^2\right\} = \frac{\sigma_s^2[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i}-\gamma\left([\mathsf{M}^k\mathsf{Q}_s]_{i,i}\right)^2}{[\mathsf{Q}_{u,k+1}(\gamma)]_{i,i}-2\gamma a_i^k[\mathsf{M}^k\mathsf{Q}_s]_{i,i}+\gamma\sigma_s^2a_i^{2k}}. \quad \text{(C.17)}$$

3) Explicit Expression of the MSE: We now derive an explicit expression for the MSE. From (C.17) it follows that we need to characterize $[\mathsf{Q}_{u,k+1}(\gamma)]_{1,1}$ and $[\mathsf{M}^k\mathsf{Q}_s]_{1,1}$. Next, we explicitly characterize $[\mathsf{Q}_{u,k+1}(\gamma)]_{1,1}$ as a function of k.

a) Analysis of $[\mathsf{Q}_{u,k+1}(\gamma)]_{1,1}$*:* From (B.7) we have:

$$[\mathsf{Q}_{u,k+1}(\gamma)]_{1,1} = \gamma\cdot\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{1,1} + \left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{1,1}. \quad \text{(C.18)}$$

We now separately analyze the two terms on the RHS of (C.18).

Analysis of $\gamma\cdot\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{1,1}$***:*** Following arguments similar to those leading to (B.10) we have $\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_1 = r_1^2+r_2^2$, where r_1 and r_2, specialized to the symmetric setting, are given in (C.4). Further simplifying the expressions we obtain:

$$r_1 = \begin{cases}\lambda_1^k\sigma_s, & k \text{ is even},\\ \frac{\sigma_s\lambda_1^k}{\det(\mathsf{V})}\left(v_1^2+v_2^2-2\rho_sv_1v_2\right) & k \text{ is odd}.\end{cases}$$
$$r_2 = \begin{cases}0, & k \text{ is even},\\ \frac{-2\lambda_1^k\sigma_s\sqrt{1-\rho_s^2}v_1v_2}{\det(\mathsf{V})} & k \text{ is odd}.\end{cases}$$

Thus, as $\Phi(\varsigma,\rho)\triangleq\frac{\varsigma^2\left(\left(v_1^2+v_2^2-2\rho v_1v_2\right)^2+4(1-\rho^2)v_1^2v_2^2\right)}{\det^2(\mathsf{V})}$, see (37a), we obtain:

$$\gamma\cdot\left[\mathsf{M}^k\mathsf{Q}_s(\mathsf{M}^T)^k\right]_{1,1} = \begin{cases}\lambda_1^{2k}\gamma\sigma_s^2, & k \text{ is even},\\ \lambda_1^{2k}\gamma\Phi(\sigma_s,\rho_s) & k \text{ is odd},\end{cases} \quad \text{(C.19)}$$

where we note that $\Phi(\sigma_s,\rho_s) \ge 0$. Next, we analyze the second term on the RHS of (C.18).

Analysis of $\left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{1,1}$***:*** Following the same arguments used for deriving (C.19), we write:

$$\left[\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{1,1} = \begin{cases}\lambda_1^{2l}\sigma_z^2, & l \text{ is even},\\ \lambda_1^{2l}\Phi(\sigma_z,\rho_z) & l \text{ is odd}.\end{cases}$$

Now, for even k, following similar arguments that led to to (C.14), we obtain:

$$\begin{aligned}\left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{1,1} &= \left(\sigma_z^2+\Phi(\sigma_z,\rho_z)\lambda_1^2\right)\sum_{m=0}^{\frac{k}{2}-1}\lambda_1^{4m}\\ &= \frac{\sigma_z^2+\lambda_1^2\Phi(\sigma_z,\rho_z)}{1-\lambda_1^4}(1-\lambda_1^{2k})\\ &= \Psi_0(1-\lambda_1^{2k}),\end{aligned}$$

where $\Psi_0 \triangleq \frac{\sigma_z^2+\lambda_1^2\Phi(\sigma_z,\rho_z)}{1-\lambda_1^4}$ is defined in (37b). Since $\Phi(\sigma_1,\rho_s) \ge 0$ and $0 < \lambda_1 < 1$, it follows that $\Psi_0 > 0$.

$$\mathbb{E}\left\{(S_1-\hat{S}_{1,k})^2\right\}=\begin{cases}\dfrac{\sigma_s^2(\lambda_1^{2k}(\gamma\sigma_s^2-\Psi_0)+\Psi_0)-\gamma\left(\lambda_1^k\sigma_s^2\right)^2}{\lambda_1^{2k}(\gamma\sigma_s^2-\Psi_0)+\Psi_0-2\gamma\sigma_s^2+\gamma\sigma_s^2\lambda_1^{-2k}}, & k \text{ is even},\\ \dfrac{\sigma_s^2(\lambda_1^{2k}(\gamma\Phi(\sigma_s,\rho_s)-\Psi_1)+\Psi_0)-\gamma\lambda_1^{2k}\Gamma_s^2}{\lambda_1^{2k}(\gamma\Phi(\sigma_s,\rho_s)-\Psi_1)+\Psi_0-2\gamma\Gamma_s+\gamma\sigma_s^2\lambda_1^{-2k}}, & k \text{ is odd}.\end{cases}$$
$$=\begin{cases}\dfrac{\sigma_s^2\Psi_0(1-\lambda_1^{2k})}{\gamma\sigma_s^2(\lambda_1^{2k}-2+\lambda_1^{-2k})+\Psi_0(1-\lambda_1^{2k})}, & k \text{ is even},\\ \dfrac{\gamma(\lambda_1^{2k}\sigma_s^2\Phi(\sigma_s,\rho_s)-\lambda_1^{2k}\Gamma_s^2)+\sigma_s^2(\Psi_0-\lambda_1^{2k}\Psi_1)}{\gamma(\lambda_1^{2k}\Phi(\sigma_s,\rho_s)-2\Gamma_s+\sigma_s^2\lambda_1^{-2k})+\Psi_0-\lambda_1^{2k}\Psi_1}, & k \text{ is odd}.\end{cases} \tag{C.23}$$

For odd k, we follow steps similar to those leading to (C.15) to obtain:

$$\begin{aligned}&\left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{1,1}\\ &=\left(\sigma_z^2+\Phi(\sigma_z,\rho_z)\lambda_1^2\right)\left(\sum_{m=0}^{\frac{k-1}{2}-1}\lambda_1^{4m}\right)+\sigma_z^2\lambda_1^{2(k-1)}\\ &=\frac{\left(\sigma_z^2+\Phi(\sigma_z,\rho_z)\lambda_1^2\right)\left(1-\lambda_1^{2(k-1)}\right)+\left(1-\lambda_1^4\right)\sigma_z^2\lambda_1^{2(k-1)}}{1-\lambda_1^4}\\ &=\frac{\sigma_z^2+\lambda_1^2\Phi(\sigma_z,\rho_z)}{1-\lambda_1^4}-\lambda_1^{2k}\frac{\Phi(\sigma_z,\rho_z)+\lambda_1^2\sigma_z^2}{1-\lambda_1^4}\\ &=\Psi_0-\lambda_1^{2k}\Psi_1.\end{aligned}$$

Hence, we have:

$$\left[\sum_{l=0}^{k-1}\mathsf{M}^l\mathsf{Q}_z(\mathsf{M}^T)^l\right]_{1,1}=\begin{cases}\Psi_0(1-\lambda_1^{2k}), & k \text{ is even},\\ \Psi_0-\lambda_1^{2k}\Psi_1 & k \text{ is odd}.\end{cases} \tag{C.20}$$

Next, we combine (C.19) and (C.20) to obtain:

$$[\mathsf{Q}_{u,k+1}(\gamma)]_{1,1}=\begin{cases}\lambda_1^{2k}(\gamma\sigma_s^2-\Psi_0)+\Psi_0, & k \text{ is even},\\ \lambda_1^{2k}(\gamma\Phi(\sigma_s,\rho_s)-\Psi_1)+\Psi_0 & k \text{ is odd}.\end{cases} \tag{C.21}$$

b) Analysis of $[\mathsf{M}^k\mathsf{Q}_s]_{1,1}$*:* Recall the definition of $\Gamma_s \triangleq \frac{\sigma_s^2(v_1^2+v_2^2-2\rho_s v_1 v_2)}{v_1^2-v_2^2}$ in (37d). For the symmetric setting, we rewrite $[\mathsf{M}^k\mathsf{Q}_s]_{1,1}$, given in (B.20), as follows:

$$\begin{aligned}[\mathsf{M}^k\mathsf{Q}_s]_{1,1}&=\frac{\sigma_s^2\lambda_1^k\left(v_1^2-v_2^2(-1)^k+\rho_s v_1 v_2((-1)^k-1)\right)}{v_1^2-v_2^2}\\ &=\begin{cases}\lambda_1^k\sigma_s^2, & k \text{ is even},\\ \lambda_1^k\frac{\sigma_s^2(v_1^2+v_2^2-2\rho_s v_1 v_2)}{v_1^2-v_2^2} & k \text{ is odd}.\end{cases}\\ &=\begin{cases}\lambda_1^k\sigma_s^2, & k \text{ is even},\\ \lambda_1^k\Gamma_s & k \text{ is odd}.\end{cases}\end{aligned} \tag{C.22}$$

c) An Explicit Expression: By plugging (C.21) and (C.22) into (C.17) we obtain an explicit expression for the MSE given in (C.23) at the top of the page. Next, we show that (C.23) decreases when γ increases.

4) The MSE Decrease with γ*:* We begin with the case of even values of k:

a) Even values of k*:* Note that $\Psi_0>0$. Thus, as $\lambda_1^{2k}<1$, we have that if $\lambda_1^{2k}-2+\lambda_1^{-2k}>0$ then the MSE decreases when γ increases:

$$\lambda_1^{2k}-2+\lambda_1^{-2k}=\frac{\lambda_1^{4k}-2\lambda_1^{2k}+1}{\lambda_1^{2k}}=\frac{(\lambda_1^{2k}-1)^2}{\lambda_1^{2k}}>0,\quad k>0.$$

Thus, for even values of k, the MSE decreases with γ.

b) Odd values of k*:* Recalling the definitions of $\Phi(\sigma_s,\rho_s)$ and Γ_S in (37a) and (37d), respectively, we write:

$$C_0\triangleq\frac{\left(v_1^2+v_2^2-2\rho_s v_1 v_2\right)^2+4(1-\rho_s^2)v_1^2v_2^2}{\det^2(\mathsf{V})},$$
$$C_1\triangleq\frac{v_1^2+v_2^2-2\rho_s v_1 v_2}{\det(\mathsf{V})}.$$

which implies that $\Phi(\sigma_s,\rho_s)=\sigma_s^2\cdot C_0$ and $\Gamma_s=\sigma_s^2\cdot C_1$. Thus, for odd values of k, we write the MSE as follows:

$$\mathbb{E}\left\{(S_1-\hat{S}_{1,k})^2\right\}=\frac{\gamma\sigma_s^4\lambda_1^{2k}(C_0-C_1^2)+\sigma_s^2(\Psi_0-\lambda_1^{2k}\Psi_1)}{\gamma\sigma_s^2(\lambda_1^{2k}C_0-2C_1+\lambda_1^{-2k})+\Psi_0-\lambda_1^{2k}\Psi_1}.$$

Defining $\theta_1\triangleq\sigma_s^4\lambda_1^{2k}(C_0-C_1^2),\theta_2\triangleq\sigma_s^2(\Psi_0-\lambda_1^{2k}\Psi_1),\theta_3\triangleq\sigma_s^2(\lambda_1^{2k}C_0-2C_1+\lambda_1^{-2k})$ and $\theta_4\triangleq\Psi_0-\lambda_1^{2k}\Psi_1$, the MSE is of the form: $\text{MSE}(\gamma)=\frac{\gamma\cdot\theta_1+\theta_2}{\gamma\cdot\theta_3+\theta_4}$. Clearly, if $\theta_j>0, j=1,2,3,4$, and $\theta_3>\theta_1$, then $\text{MSE}(\gamma)$ decreases with γ. Thus, we now show that these conditions hold.

Positivity of θ_1***:*** The positivity of θ_1 follows directly from the definitions of C_0 and C_1.

Positivity of θ_2 ***and*** θ_4***:*** Note that both Ψ_0 and Ψ_1 are positive. Furthermore, since $\lambda_1^2<1$, it is enough to show that $\Psi_0-\lambda_1^2\Psi_1>0$. We have:

$$\Psi_0-\lambda_1^2\Psi_1=\frac{\sigma_z^2+\lambda_1^2\Phi(\sigma_z,\rho_z)}{1-\lambda_1^4}-\lambda_1^2\frac{\Phi(\sigma_z,\rho_z)+\sigma_z^2\lambda_1^2}{1-\lambda_1^4}=\sigma_z^2>0.$$

Positivity of θ_3***:*** Let $\chi=\lambda_1^{2k}$, and write $\theta_3=\sigma_s^2(\chi C_0-2C_1+\chi^{-1})=\frac{\sigma_s^2}{\chi}(\chi^2C_0-2C_1\chi+1)$. Therefore, as $\chi>0$ and $C_0>0$, $\theta_3>0$ if the discriminant of $\chi^2C_0-2C_1\chi+1$ is negative:

$$\Delta=4C_1^2-4C_0=\frac{-16(1-\rho_s^2)v_1^2v_2^2}{\det^2(\mathsf{V})}<0. \tag{C.24}$$

Thus, we conclude that $\theta_3>0$.

Proving that $\theta_1 > \theta_3$**:** We have:

$$\begin{aligned}\theta_3 - \theta_1 &= \sigma_s^2(\lambda_1^{2k}C_0 - 2C_1 + \lambda_1^{-2k}) - \sigma_s^2\lambda_1^{2k}(C_0 - C_1^2) \\ &= \sigma_s^2\lambda_1^{2k}(\lambda_1^{-4k} - 2\lambda_1^{-2k}C_1 + C_1^2) \\ &= \sigma_s^2\lambda_1^{2k}(\lambda_1^{-2k} - C_1)^2 > 0.\end{aligned}$$

Thus, for odd values of k, the MSE decreases with γ. We conclude that the MSE decreases with γ, for all values of k. Hence, the optimal γ is determined by the per-symbol average power constraint, and is given by $\sqrt{\frac{\nu}{\sigma_s^2}}$, where ν is specified in (33).

C. Proof of Proposition 4

We first show that the global minimizer (regardless of the per-symbol average power constraint) of the distance between th initial and steady state covariance matrices $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$ is given by γ^*. Recalling that $\mathbf{U}_1 = \sqrt{\gamma}\mathbf{S}$, we write $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$ as:

$$\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1}) = \sqrt{2(\sigma_u^2 - \gamma\sigma_s^2)^2 + 2(\rho_u\sigma_u^2 - \gamma\rho_s\sigma_s^2)}. \quad \text{(C.25)}$$

Therefore, to find the minimizing γ we minimize following polynomial in γ:

$$q(\gamma) = \gamma^2\sigma_s^4(1+\rho_s^2) - 2\gamma\sigma_u^2\sigma_s^2(1+\rho_u\rho_s) + \sigma_u^4(1+\rho_u^2).$$

Differentiating $q(\gamma)$ and equating the result to zero we obtain the minimizing γ as:

$$\gamma^* = \frac{\sigma_u^2(1+\rho_u\rho_s)}{\sigma_s^2(1+\rho_s^2)}. \quad \text{(C.26)}$$

Now, to show that $\gamma = \frac{\nu}{\sigma_s^2}$ minimizes (C.25) under (5), we first recall that in the proof of Prop. 3 we showed that the MSE monotonically decreases when the scaling coefficient increases, and also showed that $\gamma = \frac{\nu}{\sigma_s^2}$ is the maximal scaling such that (5) is satisfied. In the following we show that $\frac{\nu}{\sigma_s^2} \leq \gamma^*$, and therefore, since $q(\gamma)$ is convex, $\gamma = \frac{\nu}{\sigma_s^2}$ minimizes $\mathbb{D}(\mathsf{Q}_u, \mathsf{Q}_{u,1})$ under (5).

To prove that $\frac{\nu}{\sigma_s^2} \leq \gamma^*$ we first derive explicit expressions for σ_u^2 and ρ_u. Recall that $\sigma_u^2 = [\mathsf{Q}_u]_{1,1}$, and that $\rho_u\sigma_u^2 = [\mathsf{Q}_u]_{1,2}$, where Q_u is the solution of (20). The following lemma provides explicit expressions for σ_u^2 and ρ_u:

Lemma 3. σ_u^2 and ρ_u are given by:

$$\sigma_u^2 = \frac{\sigma_z^2(1-a_1^2+a_1^4+a_1^6+\rho_z(-1+a_1^2-a_1^4+a_1^6))}{2a_1^2(a_1^2-1)}, \quad \text{(C.27a)}$$

$$\rho_u = \frac{1-a_1^8-\rho_z(1-2a_1^6+a_1^8)}{1+2a_1^6+a_1^8+\rho_z(a_1^8-1)}. \quad \text{(C.27b)}$$

Proof. The matrix Q_u is the solution of the discrete algebraic Lyapunov equation:

$$\mathsf{Q}_u = \mathsf{M}\mathsf{Q}_u\mathsf{M}^T + \mathsf{Q}_z, \quad \mathsf{M} = \mathsf{A} - \mathbf{B}\mathbf{C}^T. \quad \text{(C.28)}$$

To prove the lemma we show that σ_u^2 and ρ_u satisfy the following equations:

$$\mathsf{M}_0\begin{bmatrix}\sigma_u^2 \\ \rho_u\sigma_u^2\end{bmatrix} = \begin{bmatrix}\sigma_z^2 \\ \rho_z\sigma_z^2\end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix}\sigma_u^2 \\ \rho_u\sigma_u^2\end{bmatrix} = \mathsf{M}_0^{-1}\begin{bmatrix}\sigma_z^2 \\ \rho_z\sigma_z^2\end{bmatrix},$$

where the matrix M_0 is a full rank matrix derived as follows: First, recall the proof of Lemma 2 which states that $\mathsf{M} = \begin{bmatrix}a_1 - c_1 & c_1 \\ -c_1 & -(a_1 - c_1)\end{bmatrix}$. Using the relationship between c_1 and a_1 in (C.2): $c_1 = \frac{a_1^4-1}{2a_1^3}$, the matrix M can be written as:

$$\mathsf{M} = \begin{bmatrix}\frac{a_1^4+1}{2a_1^3} & \frac{a_1^4-1}{2a_1^3} \\ -\frac{a_1^4-1}{2a_1^3} & -\frac{a_1^4+1}{2a_1^3}\end{bmatrix}. \quad \text{(C.29)}$$

Writing $\mathsf{Q}_u = \begin{bmatrix}\sigma_u^2 & \rho_u\sigma_u^2 \\ \rho_u\sigma_u^2 & \sigma_u^2\end{bmatrix} = \begin{bmatrix}q_1 & q_2 \\ q_2 & q_1\end{bmatrix}$ and plugging the expression for M from (C.29) into (C.28) we have:

$$\begin{aligned}\begin{bmatrix}q_1 & q_2 \\ q_2 & q_1\end{bmatrix} = &\begin{bmatrix}\frac{q_1+a_1^8q_1-(1-a_1^8)q_2}{2a_1^6} & \frac{q_1-a_1^8q_1-(1+a_1^8)q_2}{2a_1^6} \\ \frac{q_1-a_1^8q_1-(1+a_1^8)q_2}{2a_1^6} & \frac{q_1+a_1^8q_1-(1-a_1^8)q_2}{2a_1^6}\end{bmatrix} \\ &+ \begin{bmatrix}\sigma_z^2 & \rho_z\sigma_z^2 \\ \rho_z\sigma_z^2 & \sigma_z^2\end{bmatrix},\end{aligned}$$

which can also be written as:

$$\begin{aligned}&\begin{bmatrix}-\frac{(1-2a_1^6+a_1^8)q_1+(-1+a_1^8)q_2}{2a_1^6} & \frac{(-1+a_1^8)q_1+(1+2a_1^6+a_1^8)q_2}{2a_1^6} \\ \frac{(-1+a_1^8)q_1+(1+2a_1^6+a_1^8)q_2}{2a_1^6} & -\frac{(1-2a_1^6+a_1^8)q_1+(-1+a_1^8)q_2}{2a_1^6}\end{bmatrix} \\ &\qquad = \begin{bmatrix}\sigma_z^2 & \rho_z\sigma_z^2 \\ \rho_z\sigma_z^2 & \sigma_z^2\end{bmatrix}. \quad \text{(C.30)}\end{aligned}$$

Note that the two rows of (C.30) are in fact the same equation. Therefore, we obtain the following system of equations for $[q_1, q_2]^T$:

$$\begin{aligned}&\begin{bmatrix}-\frac{1-2a_1^6+a_1^8}{2a_1^6} & \frac{a_1^8-1}{2a_1^6} \\ \frac{a_1^8-1}{2a_1^6} & \frac{1+2a_1^6+a_1^8}{2a_1^6}\end{bmatrix}\begin{bmatrix}q_1 \\ q_2\end{bmatrix} = \begin{bmatrix}\sigma_z^2 \\ \rho_z\sigma_z^2\end{bmatrix} \\ &\Rightarrow \quad \mathsf{M}_0 = \begin{bmatrix}-\frac{1-2a_1^6+a_1^8}{2a_1^6} & \frac{a_1^8-1}{2a_1^6} \\ \frac{a_1^8-1}{2a_1^6} & \frac{1+2a_1^6+a_1^8}{2a_1^6}\end{bmatrix}.\end{aligned}$$

Lastly, we note that since $a_1 > 1$, the matrix M_0 is invertible. Thus, we obtain:

$$\begin{aligned}q_1 &= \frac{\sigma_z^2(1-a_1^2+a_1^4+a_1^6+\rho_z(-1+a_1^2-a_1^4+a_1^6))}{2a_1^2(a_1^2-1)}, \\ q_2 &= \frac{\sigma_z^2(-(1-a_1^2+a_1^4+a_1^6)+\rho_z(1+a_1^2+a_1^4-a_1^6))}{2a_1^2(a_1^2+1)}.\end{aligned}$$

Finally, recalling that $\rho_u = \frac{q_2}{q_1}$ we conclude the proof of the lemma. □

Next, we use the explicit expressions for σ_u^2 and ρ_u to show that $\frac{\nu}{\sigma_s^2} \leq \gamma^*$, or equivalently $\nu \leq \frac{\sigma_u^2(1+\rho_u\rho_s)}{1+\rho_s^2}$. Recall the definition of ν in (33), repeated here for ease of reference:

$$\nu = \min\left\{\frac{\sigma_z^2(1-\rho_z+(1+\rho_z)a_1^2)}{(1-\lambda_1^4)(1-\rho_s)}, \frac{\sigma_z^2((1-\rho_z)\lambda_1^2+(1+\rho_z)a_1^4)}{(1-\lambda_1^4)(1+\rho_s)a_1^4}\right\}.$$

Using the relationship $\lambda_1 = \frac{1}{a_1}$, we write $\nu = \min\{\nu_1, \nu_2\}$, where:

$$\begin{aligned}\nu_1 &\triangleq \frac{\sigma_z^2a_1^4(1+a_1^2+\rho_z(a_1^2-1))}{(a_1^4-1)(1-\rho_s)} \\ \nu_2 &\triangleq \frac{\sigma_z^2(1+a_1^6+\rho_z(a_1^6-1))}{a_1^2(a_1^4-1)(1+\rho_s)}.\end{aligned}$$

$$\Delta_{\gamma,1}=\frac{\sigma_z^2(1+\rho_s)}{2a_1^2(a_1^4-1)(1+\rho_s^2)}\cdot\frac{-1+\rho_s+2a_1^6\rho_s+a_1^8(1+\rho_s)+\rho_z(1-2a_1^6-\rho_s+a_1^8(1+\rho_s))}{\rho_s-1}, \tag{C.31a}$$

$$\Delta_{\gamma,2}=\frac{\sigma_z^2(1-\rho_s)}{2a_1^2(a_1^4-1)(1+\rho_s^2)}\cdot\frac{-1+\rho_s+2a_1^6\rho_s+a_1^8(1+\rho_s)+\rho_z(1-2a_1^6-\rho_s+a_1^8(1+\rho_s))}{\rho_s+1}. \tag{C.31b}$$

We now show that either $\Delta_{\gamma,1}\triangleq\frac{\sigma_u^2(1+\rho_u\rho_s)}{1+\rho_s^2}-\nu_1\geq 0$ or $\Delta_{\gamma,2}\triangleq\frac{\sigma_u^2(1+\rho_u\rho_s)}{1+\rho_s^2}-\nu_2\geq 0$, which implies that $\nu\leq\frac{\sigma_u^2(1+\rho_u\rho_s)}{1+\rho_s^2}$. Plugging the expressions for σ_u^2 and ρ_u in (C.27) into (C.26) we write $\frac{\sigma_u^2(1+\rho_u\rho_s)}{1+\rho_s^2}$ as:

$$\frac{\sigma_u^2(1+\rho_u\rho_s)}{1+\rho_s^2}=\frac{\sigma_z^2(1+2a_1^6-a_1^8(\rho_s-1)+\rho_s)}{2a_1^2(a_1^4-1)(1+\rho_s^2)}+\frac{\sigma_z^2(-\rho_z(1+a_1^8(\rho_s-1)+\rho_s-2a_1^6\rho_s))}{2a_1^2(a_1^4-1)(1+\rho_s^2)}.$$

Thus, $\Delta_{\gamma,1}$ and $\Delta_{\gamma,2}$ are given in (C.31) at the top of the page. Note that since $|\rho_s|<1$, the first terms on the RHSs of (C.31a) and (C.31b) have the same sign. Therefore, since $|\rho_s|<1$, we conclude that $\text{sgn}(\Delta_{\gamma,1})=-\text{sgn}(\Delta_{\gamma,2})$. This implies that either $\Delta_{\gamma,1}\geq 0$ or $\Delta_{\gamma,2}\geq 0$. Finally, since $\nu=\min\{\nu_1,\nu_2\}$ we conclude that $\frac{\nu}{\sigma_s^2}\leq\gamma^*$.

D. Proof of Theorem 5

To explicitly characterize K_{LQG} we solve the inequality $\mathbb{E}\{(S_1-\hat{S}_{1,k})^2\}\leq D$, where $\mathbb{E}\{(S_1-\hat{S}_{1,k})^2\}$ is given in (C.23). We begin with even values of k.

1) Analysis for even values of k: For even values of k we are interested in the minimal even k such that:

$$\frac{\sigma_s^2\Psi_0(1-\lambda_1^{2k})}{\gamma\sigma_s^2(\lambda_1^{2k}-2+\lambda_1^{-2k})+\Psi_0(1-\lambda_1^{2k})}\leq D, \tag{C.32}$$

where $\sqrt{\gamma}$ is a scaling factor of the transmitted sources, i.e., $\mathbf{U}_1=\sqrt{\gamma}\cdot\mathbf{S}$. This inequality can also be written as:[19]

$$\sigma_s^2\Psi_0(1-\lambda_1^{2k})\leq D\left(\gamma\sigma_s^2(\lambda_1^{2k}-2+\lambda_1^{-2k})+\Psi_0(1-\lambda_1^{2k})\right).$$

Next, we multiply both sides of the inequality by λ_1^{2k}, and collect common terms to obtain the inequality:

$$\lambda_1^{4k}\left(\Psi_0(D-\sigma_s^2)-D\gamma\sigma_s^2\right)+\lambda_1^{2k}\left(\Psi_0(\sigma_s^2-D)+2D\gamma\sigma_s^2\right)-D\gamma\sigma_s^2\leq 0,$$

which, by using the definitions of Υ_0 and Υ_1, can also be written as $\lambda_1^{4k}\Upsilon_0+\lambda_1^{2k}\Upsilon_1-D\gamma\sigma_s^2\leq 0$. Next, we set $x=\lambda_1^{2k}$ and note that $\Upsilon_0<0$.[20] Thus, we obtain the following monic polynomial inequalities:

$$x^2+x\frac{\Upsilon_1}{\Upsilon_0}-\frac{D\gamma\sigma_s^2}{\Upsilon_0}\geq 0 \tag{C.33}$$

The discriminant of $x^2+x\frac{\Upsilon_1}{\Upsilon_0}-\frac{D\gamma\sigma_s^2}{\Upsilon_0}$ is equal to $\frac{1}{\Upsilon_0^2}(\Upsilon_1^2+4D\gamma\sigma_s^2\Upsilon_0)$. Therefore, if $\Upsilon_0<\frac{-\Upsilon_1^2}{4D\gamma\sigma_s^2}$ then $P^{(\text{e})}(x)$ has no real roots. Since $x^2+x\frac{\Upsilon_1}{\Upsilon_0}-\frac{D\gamma\sigma_s^2}{\Upsilon_0}$ is convex, if it has no real roots then it is strictly positive. Hence, *in this case the required distortion is achieved for every even k*. Therefore, we set $k=2$ and obtain $x_0^{(\text{e})}=a_1^{-4}$.
On the other hand, if $\Upsilon_0>\frac{-\Upsilon_1^2}{4D\gamma\sigma_s^2}$ then $P^{(\text{e})}(x)$ has two real roots. We write $x^2+x\frac{\Upsilon_1}{\Upsilon_0}-\frac{D\gamma\sigma_s^2}{\Upsilon_0}$ as:

$$x^2+px+q,\quad p=\frac{\Upsilon_1}{\Upsilon_0},\quad q=-\frac{D\gamma\sigma_s^2}{\Upsilon_0}.$$

The roots of this polynomial are given by $-\frac{p}{2}\pm\sqrt{\frac{p^2}{4}-q}$. Now, since $\Upsilon_1>0$ and $\Upsilon_0<0$ we have $-\frac{p}{2}>0$. As $\Upsilon_0<0$, then $q>0$. Therefore, $P^{(\text{e})}(x)$ has two positive roots; one is smaller than $-\frac{p}{2}$, and the other is larger. In fact, it is easy to see that $P^{(\text{e})}(1)=0$, thus, the larger root equals 1. Let $(x_1^{(\text{e})},x_2^{(\text{e})})$ denote the real roots of $P^{(\text{e})}(x)$. Since $x=\lambda_1^{2k}$, and since $k\geq 1$, then the root we seek is $\min\{x_1^{(\text{e})},x_2^{(\text{e})}\}$. Let $x_0^{(\text{e})}$ denote the required root. Then we have:

$$x_0^{(\text{e})}\triangleq\begin{cases}\min\left\{x_1^{(\text{e})},x_2^{(\text{e})}\right\}, & \frac{-\Upsilon_1^2}{4D\gamma\sigma_s^2}\leq\Upsilon_0<0\\ a_1^{-4}, & \text{otherwise.}\end{cases}$$

2) Analysis of Odd k's: For odd k's we use (C.23) to write:

$$\frac{\gamma(\lambda_1^{2k}\sigma_s^2\Phi(\sigma_s,\rho_s)-\lambda_1^{2k}\Gamma_s^2)+\sigma_s^2(\Psi_0-\lambda_1^{2k}\Psi_1)}{\gamma(\lambda_1^{2k}\Phi(\sigma_s,\rho_s)-2\Gamma_s+\sigma_s^2\lambda_1^{-2k})+\Psi_0-\lambda_1^{2k}\Psi_1}\leq D. \tag{C.34}$$

Based on Subsection C-B3c, this inequality can also be written as:

$$\lambda_1^{4k}\left((\gamma\Phi(\sigma_s,\rho_s)-\Psi_1)(\sigma_s^2-D)-\gamma\Gamma_s^2\right)+\lambda_1^{2k}\left(\Psi_0(\sigma_s^2-D)+2D\gamma\Gamma_s\right)-D\gamma\sigma_s^2\leq 0,$$

which, by using the definitions of Υ_2 and Υ_3, can also be written as $\lambda_1^{4k}\Upsilon_2+\lambda_1^{2k}\Upsilon_3-D\gamma\sigma_s^2\leq 0$. Now, similarly to (C.33), we set $x=\lambda_1^{2k}$ to obtain the following monic polynomial inequalities:

$$\begin{cases}x^2+x\frac{\Upsilon_3}{\Upsilon_2}-\frac{D\gamma\sigma_s^2}{\Upsilon_2}\leq 0, & \Upsilon_2>0,\\ x^2+x\frac{\Upsilon_3}{\Upsilon_2}-\frac{D\gamma\sigma_s^2}{\Upsilon_2}\geq 0 & \Upsilon_2<0,\\ x\Upsilon_3-D\gamma\sigma_s^2\leq 0, & \Upsilon_2=0.\end{cases} \tag{C.35}$$

The discriminant of $P^{(\text{o})}(x)=x^2+x\frac{\Upsilon_3}{\Upsilon_2}-\frac{D_1\gamma\sigma_s^2}{\Upsilon_2}$ is given by $\frac{1}{\Upsilon_2^2}(\Upsilon_3^2+4D_1\gamma\sigma_s^2\Upsilon_2)$. Therefore, by applying arguments similar to those applied in Subsection C-D1 we have:

$$x_0^{(\text{o})}\triangleq\begin{cases}a_1^{-2}, & \Upsilon_2<\frac{-\Upsilon_3^2}{4D_1\gamma\sigma_s^2},\\ \min\left\{x_1^{(\text{o})},x_2^{(\text{o})}\right\}, & \frac{-\Upsilon_3^2}{4D_1\gamma\sigma_s^2}\leq\Upsilon_2<0,\\ \frac{D\gamma\sigma_s^2}{\Upsilon_3}, & \Upsilon_2=0,\\ \max\left\{x_1^{(\text{o})},x_2^{(\text{o})}\right\}, & \text{otherwise.}\end{cases}$$

[19] Note that $\gamma\sigma_s^2(\lambda_1^{2k}-2+\lambda_1^{-2k})+\Psi_0(1-\lambda_1^{2k})>0$.

[20] From the fact that $\Phi(\sigma_z,\rho_z)\geq 0$ it follows that $\Psi_0\geq 0$. Furthermore from (6) we have that $D_1\leq\sigma_s^2$. Therefore, as $D_1,\gamma,\sigma_s^2>0$ it follows that $\Upsilon_0=\Psi_0(D_1-\sigma_s^2)-D_1\gamma\sigma_s^2<0$.

Note that if $\Upsilon_2 > 0$, then one of the roots is negative while the other is positive and smaller than 1. In such case we choose the positive root. Furthermore, if $P^{(\mathrm{o})}(x)$ has no real roots, then we set $x_0^{(\mathrm{o})} = a_1^{-2}$ which results in the minimal possible odd k, i.e., $k = 1$.

Next, we focus on the case of $\Upsilon_2 = 0$ and note that for $\Upsilon_2 = 0$ equation (C.35) can be written as: $x \leq \frac{D\gamma\sigma_s^2}{\Upsilon_3}$. This follows from the fact that $\Upsilon_3 = \Psi_0(\sigma_s^2 - D) + 2D\gamma\Gamma_s > 0$. To see this note that $\Psi_0 \geq 0, \sigma_s^2 \geq D$ and $D, \gamma > 0$. Furthermore, the numerator of Γ_s in (37d) is positive, while the positivity of the denominator of Γ_s follows from (C.11) which implies that $v_1^2 - v_2^2 > 0$. Next, we show that $0 < \frac{D\gamma\sigma_s^2}{\Upsilon_3} < 1$. The first inequality follows from the fact that $\Upsilon_3, D, \gamma, \sigma_s^2 > 0$. The second inequality follows from the fact that $\Gamma_s = \frac{\sigma_s^2(v_1^2+v_2^2-2\rho_s v_1 v_2)}{v_1^2-v_2^2} > \sigma_s^2$. We conclude that when $\Upsilon_2 = 0$ we can set $x_0^{(\mathrm{o})} = \frac{D\gamma\sigma_s^2}{\Upsilon_3}$.

Lastly, as $x = \lambda_1^{2k} = a_1^{-2k}$ then $k = -\frac{\log(x)}{2\log|a_1|}$. This implies that we have two candidates for the required K: $K^{(\mathrm{e})}$ obtained from $x_0^{(\mathrm{e})}$ and $K^{(\mathrm{o})}$ obtained from $x_0^{(\mathrm{o})}$. Since K is an integer, we use $\lceil\cdot\rceil$, and the functions $f^{(\mathrm{e})}(\cdot)$ and $f^{(\mathrm{o})}(\cdot)$ to round up to the nearest even and odd integers, respectively.

APPENDIX D
PROOFS FOR THE DYNAMIC PROGRAMMING SCHEME

A. Proof of Theorem 6

Let $W_k = \{\alpha_k, r_k\}$ be a "state" variable, and let $\mathbf{m} = [m_0, m_1, \ldots, m_{K-1}]$ be a given modulation vector. In the following we show that there exists a deterministic function f_{DP} such that $W_k = f_{\mathrm{DP}}(W_{k-1}, b_k, \mathbf{m}), k = 1, 2, \ldots, K$. Namely, given the action b_k and the modulation vector $\mathbf{m}$, the state evolves deterministically as in (46). In appendix D-B we show that $P_k = P$ is the optimal assignment for the JSCC-DP scheme which implies that the JSCC-DP scheme exploits all the available instantaneous average transmission power. For α_k we write:

$$\begin{aligned}\alpha_k &= \mathbb{E}\left\{(\epsilon_{1,k-1} - b_k Y_{1,k})^2\right\}\\ &\overset{(a)}{=} \alpha_{k-1} + b_k^2(P+\sigma_z^2) - 2b_k\mathbb{E}\{\epsilon_{1,k-1}Y_{1,k}\}\\ &\overset{(b)}{=} \alpha_{k-1} + b_k^2(P+\sigma_z^2)\\ &\quad - 2b_k d_{k-1}\mathbb{E}\{\epsilon_{1,k-1}(\epsilon_{1,k-1} + m_{k-1}\epsilon_{2,k-1})\} \quad \text{(D.1a)}\\ &= \alpha_{k-1} + b_k^2(P+\sigma_z^2) - b_k\sqrt{2P(\alpha_{k-1} + m_{k-1}r_{k-1})}, \quad \text{(D.1b)}\end{aligned}$$

where (a) follows from the fact that since the transmitted signal and the noises are independent, then when $P_k = P$ we have $\mathbb{E}\{Y_{1,k}^2\} = P + \sigma_z^2$; (b) follows by noting that $\mathbb{E}\{\epsilon_{1,k-1}Y_{1,k}\} = \mathbb{E}\{\epsilon_{1,k-1}X_k\} = d_k\mathbb{E}\{\epsilon_{1,k-1}(\epsilon_{1,k-1} + m_{k-1}\epsilon_{2,k-1})\}$. For r_k we write:

$$\begin{aligned}r_k &= \mathbb{E}\left\{(\epsilon_{1,k-1} - b_k Y_{1,k})(\epsilon_{2,k-1} - b_k m_{k-1} Y_{2,k})\right\}\\ &= r_{k-1} + b_k^2 m_{k-1}(P + \rho_z\sigma_z^2)\\ &\quad - b_k d_{k-1} m_{k-1}(\alpha_{k-1} + m_{k-1}r_{k-1})\\ &\quad - b_k d_{k-1}(m_{k-1}\alpha_{k-1} + r_{k-1}) \quad \text{(D.2a)}\\ &= r_{k-1} + b_k^2 m_{k-1}(P + \rho_z\sigma_z^2)\\ &\quad - b_k m_{k-1}\sqrt{2P(\alpha_{k-1} + m_{k-1}r_{k-1})}. \quad \text{(D.2b)}\end{aligned}$$

Therefore, the optimization problem in (45) can be cast as a dynamic program with state W_k, actions $\{b_k\}_{k=1}^K$ and cost function $\alpha_{K,\min}(\mathbf{m})$, namely, a cost function that takes into account only the MSE at time K and ignores all the MSEs at times $k < K$.

As we aim at minimizing $\alpha_{K,\min}(\mathbf{m})$, the last action, b_K, should be the MMSE estimator of ϵ_{K-1} based on $Y_{1,K}$, which is given by $b_K = \frac{\mathbb{E}\{\epsilon_{1,K-1}Y_{1,K}\}}{\mathbb{E}\{Y_{1,K}^2\}}$:

$$\begin{aligned}b_K &\overset{(a)}{=} \frac{d_{K-1}(\alpha_{K-1} + m_{K-1}r_{K-1})}{P+\sigma_z^2} \quad \text{(D.3a)}\\ &\overset{(b)}{=} \sqrt{\frac{P}{2(\alpha_{K-1} + m_{K-1}r_{K-1})}}\frac{(\alpha_{K-1} + m_{K-1}r_{K-1})}{P+\sigma_z^2}\\ &= \sqrt{\frac{P(\alpha_{K-1} + m_{K-1}r_{K-1})}{2(P+\sigma_z^2)^2}}, \quad \text{(D.3b)}\end{aligned}$$

where (a) is obtained by assuming that $P_k = P$ in evaluating $\mathbb{E}\{Y_{1,K}^2\}$; and (b) is obtained by plugging the expression for d_{K-1} which is given in (42). In order to find the optimal $\{b_k\}_{k=1}^K$, we first plug (D.3b) into (46a) and write:

$$\begin{aligned}\alpha_K &= \alpha_{K-1} + b_K^2(P+\sigma_z^2) - b_K\sqrt{2P(\alpha_{K-1}+m_{K-1}r_{K-1})}\\ &= \alpha_{K-1} + \frac{P(\alpha_{K-1}+m_{K-1}r_{K-1})}{2(P+\sigma_z^2)}\\ &\quad - \frac{P(\alpha_{K-1}+m_{K-1}r_{K-1})}{P+\sigma_z^2}\\ &= \alpha_{K-1}\left(1 - \frac{P}{2(P+\sigma_z^2)}\right) - \frac{P}{2(P+\sigma_z^2)}m_{K-1}r_{K-1}\\ &\overset{(a)}{=} \eta_{K-1}\alpha_{K-1} + \theta_{K-1}m_{K-1}r_{K-1}, \quad \text{(D.4)}\end{aligned}$$

where (a) follows by defining $\eta_{K-1} \triangleq 1 - \frac{P}{2(P+\sigma_z^2)}$ and $\theta_{K-1} \triangleq -\frac{P}{2(P+\sigma_z^2)}$. Next, plugging (D.1a) and (D.2a) into (D.4) we write:

$$\begin{aligned}\alpha_K &= \eta_{K-1}\alpha_{K-1} + \theta_{K-1}m_{K-1}r_{K-1}\\ &= \eta_{K-1}\left(\alpha_{K-2} + b_{K-1}^2(P+\sigma_z^2)\right.\\ &\qquad \left. -2b_{K-1}d_{K-2}(\alpha_{K-2} + m_{K-2}r_{K-2})\right)\\ &\quad + \theta_{K-1}m_{K-1}\left(r_{K-2} + b_{K-1}^2 m_{K-2}(P+\rho_z\sigma_z^2)\right.\\ &\qquad \left. -2b_{K-1}d_{K-2}m_{K-2}(\alpha_{K-2} + m_{K-2}r_{K-2})\right)\\ &= b_{K-1}^2\left(\eta_{K-1}(P+\sigma_z^2) + \theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2)\right)\\ &\quad - 2b_{K-1}d_{K-2}\left(\eta_{K-1} + \theta_{K-1}m_{K-1}m_{K-2}\right)\\ &\qquad \times (\alpha_{K-2} + m_{K-2}r_{K-2})\\ &\quad + \eta_{K-1}\alpha_{K-2} + \theta_{K-1}m_{K-1}r_{K-2}. \quad \text{(D.5)}\end{aligned}$$

Hence, given $W_{K-2}, \mathbf{m}, \eta_{K-1}$ and θ_{K-1}, α_K is a quadratic function of b_{K-1}. This implies that the optimizing b_{K-1} is given in (D.6b) at the top of the next page. We note that (D.6b) holds with $K-1$ replaced by $k, k \leq K-1$, and $K-2$ replaced by $k-1$. In the following we derive the backwards calculation of η_k and θ_k. Hence, (D.6b) along with (D.3b) constitute (48). Note that given W_{K-2} and $\mathbf{m}$, b_{K-1} is a function of η_{K-1} and θ_{K-1}. Next, we plug (D.6b) back into (D.5) to obtain (D.7) at the top of the next page. Hence, (D.7) implies that the sequences η_k and θ_k, for $k = K-1, K-2, \ldots, 1$ obey the backwards recursive formulation given in (47), where η_{K-1}

$$b_{K-1} = \frac{d_{K-2}\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)\left(\alpha_{K-2}+m_{K-2}r_{K-2}\right)}{\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2)} \tag{D.6a}$$

$$= \sqrt{\frac{P}{2\left(\alpha_{K-2}+m_{K-2}r_{K-2}\right)}}\frac{\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)\left(\alpha_{K-2}+m_{K-2}r_{K-2}\right)}{\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2)}$$

$$= \sqrt{\frac{P\left(\alpha_{K-2}+m_{K-2}r_{K-2}\right)}{2}}\frac{\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}}{\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2)}. \tag{D.6b}$$

$$\begin{aligned}\alpha_K &= -\frac{d_{K-2}^2\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)^2\left(\alpha_{K-2}+m_{K-2}r_{K-2}\right)^2}{\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2)}\\ &\quad+\eta_{K-1}\alpha_{K-2}+\theta_{K-1}m_{K-1}r_{K-2}\\ &= -\frac{P\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)^2\left(\alpha_{K-2}+m_{K-2}r_{K-2}\right)}{2(\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2))}\\ &\quad+\eta_{K-1}\alpha_{K-2}+\theta_{K-1}m_{K-1}r_{K-2}\\ &= \alpha_{K-2}\left(\eta_{K-1}-\frac{P\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)^2}{2(\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2))}\right)\\ &\quad+m_{K-2}r_{K-2}\left(\theta_{K-1}m_{K-1}m_{K-2}-\frac{P\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)^2}{2(\eta_{K-1}(P+\sigma_z^2)+\theta_{K-1}m_{K-1}m_{K-2}(P+\rho_z\sigma_z^2))}\right)\\ &\triangleq \eta_{K-2}\alpha_{K-2}+\theta_{K-2}m_{K-2}r_{K-2}.\end{aligned} \tag{D.7}$$

$$b_{K-1} = \frac{d_{K-2}\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)\zeta_{K-2}}{2d_{K-2}^2\zeta_{K-2}(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2})+\sigma_z^2(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\rho_z)}. \tag{D.12}$$

---

and θ_{K-1} are provided just below (D.4). We conclude that given the sequences η_k and θ_k, and given $\mathbf{m}$ and W_{k-1}, the optimal coefficient b_k can be calculated via (D.6b), and then, the computed b_k can be used in the forward calculation (46). The optimal α_K for the given modulation vector $\mathbf{m}$ is given by (D.4).

B. Optimality of $P_k = P$ in the JSCC-DP Scheme

In this subsection we prove that the optimal scaling d_k, in the MMSE sense, results in $P_k = P$, $\forall k$. We begin our analysis with d_{K-1}, and recall that $\mathbb{E}\{Y_{1,K}^2\} = 2d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1}) + \sigma_z^2$. Thus, rewriting (D.1b) for $k = K$ we obtain:

$$\begin{aligned}\alpha_K = \alpha_{K-1} + b_K^2(2d_{K-1}^2(\alpha_{K-1}+m_{K-1}r_{K-1})+\sigma_z^2)\\ -2b_K d_{K-1}(\alpha_{K-1}+m_{K-1}r_{K-1}).\end{aligned} \tag{D.8}$$

Similarly, (D.3a) becomes:

$$b_K = \frac{d_{K-1}(\alpha_{K-1}+m_{K-1}r_{K-1})}{2d_{K-1}^2(\alpha_{K-1}+m_{K-1}r_{K-1})+\sigma_z^2} \tag{D.9}$$

Next, we plug (D.9) into (D.8) to obtain:

$$\alpha_K = \alpha_{K-1} - \frac{d_{K-1}^2(\alpha_{K-1}+m_{K-1}r_{K-1})^2}{2d_{K-1}^2(\alpha_{K-1}+m_{K-1}r_{K-1})+\sigma_z^2}. \tag{D.10}$$

Now, fix α_{K-1} and let $r_{K-1} = \rho_{K-1}\alpha_{K-1}$, $|\rho_{K-1}| \leq 1$,1 which implies that $\alpha_{K-1} + m_{K-1}r_{K-1} = \alpha_{K-1}(1 + m_{K-1}\rho_{K-1}) \geq 0$. Let $\zeta_k \triangleq \alpha_k + m_k r_k$ and define $x \triangleq d_{K-1}^2$. We write:

$$\frac{d_{K-1}^2(\alpha_{K-1}+m_{K-1}r_{K-1})^2}{2d_{K-1}^2(\alpha_{K-1}+m_{K-1}r_{K-1})+\sigma_z^2} = \frac{\zeta_{K-1}^2\cdot x}{2\zeta_{K-1}\cdot x+\sigma_z^2} = f_0(x),$$

where $x, \zeta_{K-1} \geq 0$. It can be easily shown that $\frac{df_0(x)}{dx} = \frac{\zeta_{K-1}^2\sigma_z^2}{(2\zeta_{K-1}\cdot x+\sigma_z^2)^2} \geq 0$, which implies that $f_0(x)$ is a monotonic non-decreasing function for $x \geq 0$. We conclude that for any given α_{K-1} (D.10) is minimized when d_{K-1}^2 is maximized, i.e., $d_{K-1} = \sqrt{\frac{P}{2(\alpha_{K-1}+m_{K-1}r_{K-1})}}$ and therefore $P_K = P$, thus, satisfying the average per-symbol power constraint with equality.

Next, we consider the case of $k = K-1$, and since setting $P_K = P$ is optimal we can use η_{K-1} and θ_{K-1} given below (D.4). Recall that $\mathbb{E}\{X_k^2\} = d_{k-1}^2(2\alpha_{k-1} + 2m_{k-1}r_{k-1}) = 2d_{k-1}^2\zeta_{k-1}$. Hence, we rewrite (D.5) with P replaced by $2d_{K-2}^2\zeta_{K-2}$ to obtain:

$$\begin{aligned}\alpha_K = b_{K-1}^2\big(&\eta_{K-1}(2d_{K-2}^2\zeta_{K-2}+\sigma_z^2)\\ &+\theta_{K-1}m_{K-1}m_{K-2}(2d_{K-2}^2\zeta_{K-2}+\rho_z\sigma_z^2)\big)\\ &-2b_{K-1}d_{K-2}\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)\zeta_{K-2}\\ &+\eta_{K-1}\alpha_{K-2}+\theta_{K-1}m_{K-1}r_{K-2},\end{aligned} \tag{D.11}$$

where the optimal b_{K-1}, in terms of d_{K-2}, is given in (D.12) at the top of the page, see (D.6a). Plugging (D.12) into (D.11) we write (D.13) at the top of the next page. Again, $\eta_{K-1}, \alpha_{K-2}, \theta_{K-1}, m_{K-1}$, and r_{K-2} can be viewed as constants. Let $\gamma_k \triangleq \eta_k + \theta_k m_k m_{k-1}$, $\xi_k \triangleq \eta_k + \theta_k m_k m_{k-1}\rho_z$.

$$\alpha_K = -\frac{d_{K-2}^2\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)^2\zeta_{K-2}^2}{2d_{K-2}^2\zeta_{K-2}(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2})+\sigma_z^2(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\rho_z)} + \eta_{K-1}\alpha_{K-2}+\theta_{K-1}m_{K-1}r_{K-2}. \tag{D.13}$$

$$-\frac{d_{K-2}^2\left(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\right)^2\zeta_{K-2}^2}{2d_{K-2}^2\zeta_{K-2}(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2})+\sigma_z^2(\eta_{K-1}+\theta_{K-1}m_{K-1}m_{K-2}\rho_z)} = \frac{\zeta_{K-2}^2\gamma_{K-1}^2\cdot x}{2\zeta_{K-2}\gamma_{K-1}\cdot x+\xi_{K-1}\sigma_z^2} = f_1(x). \tag{D.14}$$

We now write the first term on the RHS of (D.13) as (D.14) at the top of the page. Since $\frac{df_1(x)}{dx} = \frac{\zeta_{K-2}^2\gamma_{K-1}^2\xi_{K-1}\sigma_z^2}{(2\zeta_{K-2}\gamma_{K-1}\cdot x+\xi_{K-1}\sigma_z^2)^2}$, we conclude that the sign of $\frac{df_1(x)}{dx}$ does not depend on x, and therefore $f_1(x)$ is monotonic. Now, if $\xi_{K-1} > 0$ then α_K is minimized when d_{K-2}^2 is maximized, i.e., $P_{K-1} = P$. On the other hand, if $\xi_{K-1} < 0$ then α_K is minimized when $d_{K-2}^2 = 0$. Clearly, the case of $d_{K-2} = 0$ implies that $P_{K-1} = 0$, which cannot be optimal as it implies that $\alpha_K = \alpha_{K-1}$. Finally, if $\xi_{K-1} = 0$ we have that α_K is independent of d_{K-2} which clearly cannot hold. We conclude that the optimal choice of d_{K-2} is the one which results in $P_{K-1} = P$. Furthermore, we note that similarly to Subsection D-A, the analysis for $k = K-1$ holds for any $k < K$, which implies that $P_k = P$ is optimal for all values of K.

Appendix E
Proof of Proposition 5

From Remark 12 it is clear that JSCC-DP outperforms JSCC-OL. Next, to compare JSCC-DP with JSCC-LQG we show that both schemes have the same structure of state evolution, transmitted signal, and decoders. Therefore, the JSCC-LQG scheme is in the search space of JSCC-DP.

***The JSCC-DP scheme*:** In the JSCC-DP scheme the transmitted signal is given by (41):

$$X_{k+1} = d_k\left(\epsilon_{1,k}+m_k\epsilon_{2,k}\right), \tag{E.1}$$

where $\epsilon_{i,k}$ evolves as:

$$\epsilon_{1,k} = \epsilon_{1,k-1} - b_{1,k}Y_{1,k}, \quad \epsilon_{2,k} = \epsilon_{2,k-1} - b_{2,k}Y_{2,k}. \tag{E.2}$$

Here, $b_{1,k} = b_k$ and $b_{2,k} = m_{k-1}b_k$. From (10), and similarly to [37, Eq. (7)], it follows that the JSCC-DP scheme estimates the source S_i via $\hat{S}_i = \sum_{m=1}^K b_{i,m}Y_{i,m}$. Note that in the JSCC-DP scheme we optimize over the sequences $\{d_k\}_{k=1}^K, \{b_k\}_{k=1}^K$ and $\{m_k\}_{k=0}^{K-1}$.

***The JSCC-LQG scheme*:** We can write the transmitted signal in the JSCC-LQG scheme as follows, see Subsection IV-A:

$$X_k = \tilde{c}\left(U_{1,k}-U_{2,k}\right), \tag{E.3}$$

where $\tilde{c} = -c_1/a$. Together with (E.3), the states $U_{i,k}$ evolve as (18):

$$U_{1,k} = U_{1,k-1}+\frac{1}{a}Y_{1,k}, \quad U_{2,k} = -U_{2,k-1}+\frac{1}{a}Y_{2,k}. \tag{E.4}$$

Next, recall that the decoding in the JSCC-LQG scheme is applied in two stages. First the state $U_{i,k}$ is estimated as in (21), and then S_i is estimated from the estimated state. From (B.2) we have that the estimated state $\hat{U}_{i,k+1}$ obeys:

$$\hat{U}_{i,k+1} = \sum_{m=1}^K a_i^{k-m-1}Y_{i,m}. \tag{E.5}$$

Now, for any decoder which estimates S_i from $\hat{U}_{i,k+1}$ via: $\tau_{i,k}\hat{U}_{i,k-1}$, where $\{\tau_{i,k}\}_{k=1}^K$ is a sequence which depends on the decoder in use,[21] $\hat{S}_{i,k}$ has the following form:

$$\hat{S}_{i,k} = \sum_{m=1}^K \tau_{i,k}a_i^{k-m-1}Y_{i,m} = \sum_{m=1}^K \tilde{\tau}_{i,m}Y_{i,m} \tag{E.6}$$

for a *known* sequence $\{\tilde{\tau}_{i,k}\}_{k=1}^K$. In particular, this holds for the three decoders (22), (24), and (35). Therefore, as the transmitted signals, the state evolution, and the decoders have the same linear (recursive) structure, and as in the JSCC-DP scheme we optimize over the sequences $\{d_k\}_{k=1}^K, \{b_k\}_{k=1}^K$ and $\{m_k\}_{k=0}^{K-1}$, we conclude that JSCC-DP outperforms JSCC-LQG with each one of the decoders, as long as (5) is satisfied.

References

[1] Y. Murin, Y. Kaspi, R. Dabora and D. Gündüz, "Uncoded transmission of correlated Gaussian sources over broadcast channels with feedback," in *Proc. IEEE GlobalSIP Symp. on Network Theory*, Atlanta, GA, Dec. 2014, pp. 1063–1067.

[2] Y. Murin, Y. Kaspi, R. Dabora and D. Gündüz, "On the transmission of a bivariate Gaussian source over the Gaussian broadcast channel with feedback," *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, Apr. 2015.

[3] I-H. Hou and P. R. Kumar, "Broadcasting delay-constrained traffic over unreliable wireless links with network coding," *Proc. ACM MobiHoc*, Paris, France, May 2011, pp. 33–42.

[4] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Mar.-Apr. 2009.

[5] J. Blessy and A. Anpalagan, "Body area sensor networks: requirements, operations, and challenges," *IEEE Potentials*, vol. 33, no. 2, pp. 21–25, Mar.-Apr. 2014.

[6] Y. Kaspi and N. Merhav, "Zero-delay and causal single-user and multi-user lossy source coding with decoder side information," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6931–6942, Nov. 2014.

[7] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1423–1438, Jun. 2002.

[21] We emphasize that the sequences $\{\tau_{i,k}\}_{k=1}^K$ are given and we do not optimize over their values.

[8] L. H. Ozarow and S. K. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 667–671, Jul. 1984.

[9] E. Ardestanizadeh, P. Minero and M. Franceschetti, "LQG control approach to Gaussian broadcast channels with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5267–5278, Aug. 2012.

[10] D. P. Bertsekas *Dynamic Programing and Optimal Control*. 3rd ed. Nashua, NH: Athena Sceintific. 2007.

[11] C. E. Shannon, "The zero error capacity of a noisy channel", *IRE Trans. Inf. Theory*, vol. 2, pp. 221–238, Sep. 1956.

[12] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive white noise channels with feedback–Part I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 172–182, Apr. 1966.

[13] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, Mar. 2011.

[14] N. T. Gaarder and J .K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 100–102, Jan. 1975.

[15] L. H. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623–629, Jul. 1984.

[16] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 379–381, May 1978.

[17] Y. Murin, Y. Kaspi, and R. Dabora, "On the Ozarow-Leung scheme for the Gaussian broadcast channel with feedback," *IEEE Sig. Proc. Letters*, vol. 22, no. 7, pp. 948–952, Jul. 2015.

[18] A. A. Zaidi, T. J. Oechtering and M. Skoglund, "Sufficient conditions for closed-loop control over multiple-access and broadcast channels," *Proc. IEEE Conf. on Decision and Cont.*, Atlanta, GA, Dec. 2010, pp. 4771–4776.

[19] A. A. Zaidi, T. J. Oechtering and M. Skoglund, "Stabilization of noisy plants over a Gaussian interference channel," *Proc. Iran Workshop on Commun. and Inf. Theory*, Teheran, Iran, May 2014, pp. 1–6.

[20] N. Elia, "When Bode meets Shannon: Control oriented feedback communication schemes," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1477–1488, Sep. 2004.

[21] S. B. Amor, Y. Steinberg and M. Wigger, "MIMO MAC-BC duality with linear-feedback coding schemes," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5976–5998, Nov. 2015.

[22] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger, "Coding schemes and asymptotic capacity for the Gaussian broadcast and interference channels with feedback," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 54–71, Jan. 2014.

[23] Y. Wu, P. Minero and M. Wigger, "Insufficiency of linear-feedback schemes in Gaussian broadcast channels with common message," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4553–4566, Aug. 2014.

[24] P. Minero and S. Jayaraj, "Lossy communication over multiple-access channels with feedback," *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, Jul. 2012, pp. 3003-3007.

[25] P. Minero, S. H. Lim, and Y. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1509–1523, Apr. 2015.

[26] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian source over a Gaussian MAC with feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1852–1864, Apr. 2010.

[27] A. Jain, D. Gündüz, S. R. Kulkarni, V. Poor and S. Verdú, "Energy-distortion tradeoffs in Gaussian joint source-channel coding problems," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3153–3168, May 2012.

[28] Y. Murin, Y. Kaspi, R. Dabora and D. Gündüz, "Energy-distortion tradeoff for the Gaussian broadcast channel with feedback," To be submitted to *Entropy*, available at http://web.stanford.edu/∼moriny.

[29] R. Venkataramanan and S. S. Pradhan, "An achievable rate region for the broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6175–6191, Oct. 2013.

[30] S. R. B. Pillai and V. M. Prabhakaran, "On the noisy feedback capacity of Gaussian broadcast channels," *IEEE Inf. Theory Workshop*, Jerusalem, Israel, Apr. 2015.

[31] A. Ben-Yishai and O. Shayevitz, "Interactive schemes for the AWGN channel with noisy feedback," accepted to *IEEE Trans. Inf. Theory*, 2017.

[32] A. Ben-Yishai and O. Shayevitz, "The AWGN BC with MAC feedback: a reduction to noiseless feedback via interaction," in *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, Apr. 2015.

[33] Y. Murin, R. Dabora, and D. Gündüz, "Source-channel coding theorems for the multiple-access relay channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5446–5465, Sep. 2013.

[34] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[35] S. C. Cripps, *RF power amplifiers for wireless communications*. 2nd ed. Artech House Inc., 2006.

[36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[37] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, Jan. 2010.

[38] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography – part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.

[39] I. N. Bronshtein, K. A. Semendyayev, G. Musiol and H. Muehlig *Handbook of Mathematics*. 5th ed. Springer Berlin Heidelberg New York, 2007.

[40] C. Tian, J. Chen, S. N. Diggavi, and S. Shamai, "Optimality and approximate optimality of source-channel separation in networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 904–918, Feb. 2014.

[41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons Inc., 1991.

[42] Y. H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 57–85, Jan. 2010.

[43] H. S. Wilf, "Budan's theorem for a class of entire functions," *Proc. Amer. Math. Soc.*, vol. 13, no. 1, pp. 122–125, Feb. 1962.

[44] A. J. Rojas, "Explicit solution for a class of discrete-time algebraic Riccati equations," *Asian jour. Cont.*, vol. 15, no. 1, pp. 132–141, Jan. 2013.

Yonathan Murin (M'15) received his BSc degree in Electrical Engineering and Computer Science from Tel-Aviv University, Israel, in 2004, and his MSc and PhD degrees in 2011 and 2015, respectively, from Ben-Gurion University of the Negev, Israel, both in Electrical Engineering. He is currently a postdoctoral scholar in the Wireless Systems Lab at Stanford University. From 2004 to 2010, he worked as a DSP and algorithms engineer, and as a team leader at Comsys Communication and Signal Processing. His research interests include network information theory and wireless communications, coding theory, molecular communications, and signal processing for neuroscience.

Yonatan Kaspi ADD BIO

Ron Dabora (M'07-SM'14) received his BSc and MSc degrees in 1994 and 2000, respectively, from Tel-Aviv University, and his PhD degree in 2007 from Cornell University, all in Electrical Engineering. From 1994 to 2000, he worked as an engineer at the Ministry of Defense of Israel, and from 2000 to 2003, he was with the Algorithms Group at Millimetrix Broadband Networks, Israel. From 2007 to 2009, he was a postdoctoral researcher at the Department of Electrical Engineering, Stanford University. Since 2009, he is an assistant professor at the Department of Electrical and Computer Engineering, Ben-Gurion University, Israel. His research interests include network information theory, wireless communications, and power line communications. Dr. Dabora served as a TPC member in a number of international conferences including WCNC, PIMRC, and ICC. From 2012 to 2014, he served as an associate editor for the IEEE Signal Processing Letters, and he currently serves as a senior area editor for the IEEE Signal Processing Letters.

Deniz Gündüz (M'08-SM'13) received the B.S. degree in electrical and electronics engineering from METU, Turkey in 2002, and the M.S. and Ph.D. degrees in electrical engineering from NYU Polytechnic School of Engineering in 2004 and 2007, respectively. After his PhD, he served as a postdoctoral research associate at Princeton University, and as a consulting assistant professor at Stanford University. He was a research associate at CTTC in Barcelona, Spain until September 2012, when he joined the Electrical and Electronic Engineering Department of Imperial College London, UK, where he is currently a Reader in information theory and communications. His research interests lie in the areas of communications and information theory with special emphasis on joint source-channel coding, multi-user networks, energy efficient communications and privacy in cyber-physical systems. Dr. Gndz is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is the recipient of a Starting Grant of the European Research Council (ERC) in 2016, IEEE Communications Society Best Young Researcher Award for the Europe, Middle East, and Africa Region in 2014, Best Paper Award at the 2016 IEEE Wireless Communications and Networking Conference (WCNC), and the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory (ISIT). He served as the General Co-chair of the 2016 IEEE Information Theory Workshop, and a Co-chair of the PHY and Fundamentals Track of the 2017 IEEE Wireless Communications and Networking Conference.