Stabilization of Linear Systems Over Gaussian Networks

Ali A. Zaidi, Tobias J. Oechtering, Member, IEEE, Serdar Yüksel, and Mikael Skoglund, Senior Member, IEEE

Abstract—The problem of remotely stabilizing a noisy linear time invariant plant over a Gaussian relay network is addressed. The network is comprised of a sensor node, a group of relay nodes and a remote controller. The sensor and the relay nodes operate subject to an average transmit power constraint and they can cooperate to communicate the observations of the plant’s state to the remote controller. The communication links between all nodes are modeled as Gaussian channels. Necessary as well as sufficient conditions for mean-square stabilization over various network topologies are derived. The sufficient conditions are in general obtained using delay-free linear policies and the necessary conditions are arrived at using information theoretic tools. Different settings where linear policies are optimal, asymptotically optimal (for certain parameters of the system) and suboptimal have been identified. For the case with noisy multidimensional sources controlled over scalar channels, it is shown that linear time varying policies lead to minimum capacity requirements, meeting the fundamental lower bound. For the case with noiseless sources and parallel channels, nonlinear policies which meet the lower bound have been identified.

Index Terms—Gaussian relay channels, linear systems, moment stabilization, networked control systems.

I. INTRODUCTION

THE emerging area of networked control systems has gained significant attention in recent years due to its potential applications in many fields such as machine-to-machine communication for security, surveillance, production, building management, and traffic control. The idea of controlling dynamical systems over communication networks is supported by the rapid advance of wireless technology and the development of cost-effective and energy efficient devices (sensors), capable of sensing, computing, and transmitting. This paper considers a setup in which a sensor node communicates the observations of a linear dynamical system (plant) over a network of wireless nodes to a remote controller in order to stabilize the system in closed-loop. The wireless nodes have transmit and receive capability and we call them relays, as they relay the plant’s state information to the remote controller. We assume a transmit power constraint on the sensor and relays, and the wireless links between all agents (sensor, relays, and controller) are modeled as Gaussian channels. The objective is to study stabilizability of the plant over Gaussian networks.

A. Problem Formulation

Consider a discrete linear time invariant system

\[ X_{t+1} = AX_t + BU_t + W_t \]  

(1)

where \( X_t \in \mathbb{R}^n \), \( U_t \in \mathbb{R}^m \), and \( W_t \in \mathbb{R}^n \) are state, control, and plant noise variables. The initial state \( X_0 \) is a random variable with bounded differential entropy \( h(X_0) < \infty \) and a covariance matrix \( \Lambda_0 \). The plant noise \( \{W_t\} \) is a zero mean white Gaussian sequence with variance \( K_W \) and it is assumed to be independent of the initial state \( X_0 \). The matrices \( A \) and \( B \) are of appropriate dimensions and the pair \((A, B)\) is controllable. Let \( \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \) denote the eigenvalues of \( A \). Without loss of generality we assume that all the eigenvalues of \( A \) are outside the unit disc, i.e., \(|\lambda_i| \geq 1\). The unstable modes can be decoupled from the stable modes by a similarity transformation. If the system in (1) is one-dimensional then \( A \) is scalar and we use the notation \( A = \lambda \). We consider a remote control setup where a sensor observes the state process and transmits it to a remotely situated controller over a network of relay\(^1\) nodes as shown in Fig. 1. The communication links between nodes are modeled as white Gaussian channels, which is why we refer to the network as Gaussian. In order to communicate the observed state value \( X_t \), an encoder \( E \) is lumped with the observer \( O \) and a decoder \( D \) is lumped with the controller \( C \). In addition, there are \( L \) relay nodes \( \{R_i\}_{i=1}^L \) within the channel to support communication from \( E \) to \( D \). At any time instant \( t \),

\(^1\)A relay is a communication device whose sole purpose is to support communication from the information source to the destination. In our setup the relay nodes cooperate to communicate the state process from sensor to the remote controller. If the design objective is to replace wired connections, then relaying is a vital approach to communicate over longer distances.
there exists a constant constraint:

\[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[S_{x_t}^2] \leq P_S. \]

Let \( \pi_t \) denote the decoder/encoder policy, then \( U_t = \pi_t(R_{t-1}) \). The objective in this paper is to find conditions on \( A \) so that the plant in (1) can be mean square stabilized over a given Gaussian network.

**Definition 1.1:** A system is said to be mean square stable if there exists a constant \( M < \infty \) such that \( E[\| X_t \|^2] < M \) for all \( t \).

**B. Literature Review**

Important contributions to control over communication channels include [1]–[20]. The problem of remotely controlling dynamical systems over communication channels is studied with methods from stochastic control theory and information theory. The seminal paper by Bansal and Başar [1] used fundamental information theoretic arguments to obtain optimal policies for LQG control of a first order plant over a point to point Gaussian channel. Minimum rate requirements for stabilizability of a noiseless scalar plant were first established in [2], [3] followed by [4]. Further rate theorems for stabilization of linear plants over some discrete and continuous alphabet channels can be found in [10], [14]–[16], [18], [21]–[27]. The papers [1], [9], [10], [14], [16]–[18], [22]–[24], [26], [27] addressing control over Gaussian channels are more relevant to our work. In [1] linear sensing and control policies are shown to be optimal for the LQG control of a first order linear plant over a point-to-point Gaussian channel. A necessary condition for stabilization relating eigenvalues of the plant to the capacity of the Gaussian channel first appeared in [9], [10]. Some important contributions on stabilization over Gaussian channels with average transmit power constraints have been made in [14], [16], [22]–[24], [27], [28]. In [14] sufficient conditions for stabilization of both continuous time and discrete time multi-dimensional plants over a scalar white Gaussian channel were obtained using linear time invariant (LTI) sensing and control schemes. It was shown in [14], [23] that under some assumptions there is no loss in using LTI schemes for stabilization, that is the use of non-linear time varying schemes does not allow stabilization over channels with lower signal-to-noise ratio. The stability results were extended to a colored Gaussian channel in [16]. In [18] the authors considered noisy communication links between both sensor—controller and controller—actuator and presented necessary and sufficient conditions for mean square stability. Stabilization of noiseless LTI plants over parallel white Gaussian channels subject to transmit power constraint has been studied in [22], [24], [27], [28]. The paper [22] considers output feedback stabilization and [24] considers state feedback stabilization, and they both derive necessary and sufficient conditions for stability under a total transmit power constraint. The necessary condition derived in [24] for mean-square stabilization of discrete time LTI plants over parallel Gaussian channels is not tight in general and its achievability is not guaranteed by LTI schemes. The paper [27] focuses on mean-square stabilization of two-input two-output system over two parallel Gaussian channels. By restricting the study to LTI schemes and assuming individual power constraint on each channel, the authors derive tight necessary and sufficient conditions for both state feedback and output feedback architectures. Realizing that LTI schemes are not optimal in general for stabilization over parallel channels [24], the paper [28] proposes a non-linear time invariant scheme for stabilization of a scalar noiseless plant over parallel Gaussian channels using the idea that independent information should be transmitted on different channels [17], [29]. The problem of finding a tight necessary and sufficient condition for stabilization of an \( n \)-dimensional plant over an \( n \)-dimensional parallel Gaussian channel is still open, which we investigate in this paper. For a detailed literature review, we refer the reader to [30]–[32].

As summarized above, the previous works on control over Gaussian channels have mostly focused on situations where there is no intermediate node between the sensor and the remote controller. Problems related to control over Gaussian networks with relay nodes have so far been open. Such problems are hard because a relay network can have an arbitrary topology and every node within the network can have memory and can employ any transmit strategy. The papers [33] and [34] have derived conditions for stabilization over networks with digital noiseless channels and analog erasure channels respectively, however those results do not apply to noisy networks. In [12], [25] moment stability conditions in terms of error exponents have been established. However, even a single letter expression for channel capacity of the basic three-node Gaussian relay channel [35] is not known in general. In [36] Gastpar and Vetterli determined the capacity of a large Gaussian relay network in the limit as the number of relays tends to infinity. The problem of control over Gaussian relay channels was introduced in [37], [38] and further studied in [39], [40]. The papers [37]–[40] derived sufficient conditions for mean square stability of a scalar plant by employing linear schemes over Gaussian channels with single relay nodes. In this paper we consider more general setups with multiple relays and multi-dimensional plants. We also derive necessary conditions along with sufficient conditions and further discuss how good linear policies are for various network topologies. In particular this paper makes the following contributions:

**C. Main Contributions**

- In Section II we obtain a necessary condition for mean square stabilization of the linear system in (1) over the general relay network depicted in Fig. 1.
- In Sections III–V we derive necessary as well as sufficient conditions for stabilization over some fundamental network topologies such as cascade network, parallel network, and non-orthogonal network, which serve as building blocks for a large class of Gaussian networks (see Figs. 2, 3, 4, pp. 3, 5, 6). Necessary conditions are obtained using information theoretic tools. Sufficient conditions are obtained using linear schemes. We study these fundamental topologies individually so that the proof techniques and the intuitions gained from this paper are rich enough to address more general networks.
• Sub-optimality of linear policies is discussed and some insights on optimal schemes are presented. In some cases linear schemes can be asymptotically optimal and in some cases exactly optimal.
• A linear time varying scheme is proposed in Section VI, which is optimal for stabilization of noisy multidimensional plants over point-to-point scalar Gaussian channels.
• The information capacity\(^2\) required for stabilization of multidimensional plants over parallel Gaussian channels is established in Section IV, which is achievable by a nonlinear time varying scheme for noiseless plants.

II. NECESSARY CONDITION FOR STABILIZATION

In the literature [7], [15], [25], [41], there exist a variety of information rate inequalities characterizing fundamental limits on the performance of linear systems controlled over communication channels. In the following we state a relationship which gives a necessary condition for mean square stabilization over the general network depicted in Fig. 1.

**Theorem 2.1:** If the linear system in (1) is mean square stable over the Gaussian relay network, then

\[
\log |\det(A)| \leq \liminf_{T \to \infty} \frac{1}{T} I \left( X_{[0,T-1]} \rightarrow R_{[0,T-1]} \right) \tag{2}
\]

where \( \{X_t\} \) denotes an uncontrolled state process obtained by substituting \( U_t = 0 \) in (1), i.e., \( X_{t+1} = AX_t + W_t \). The notation \( |\det(A)| \) represents the absolute value of determinant of matrix \( A \) and

\[
I \left( X_{[0,T-1]} \rightarrow R_{[0,T-1]} \right) = \sum_{t=0}^{T-1} I \left( X_{[0,t]} | R_{[0,t-1]} \right)
\]

is the directed information from the uncontrolled state process \( X_{[0,T-1]} \) to the sequence of variables \( R_{[0,T-1]} \) received by the controller over the network of relay nodes.

**Proof 2.1:** The proof is given in Appendix A, which essentially follows from the same steps as in the proof of Theorem 4.1 in [25], however, with some differences due to the network structure. Similar constructions can also be found in [15], [41].

III. CASCADE (SERIAL) NETWORK

In this section we consider a cascade network of half-duplex relay nodes. A node which is capable of transmitting and receiving signals simultaneously using the same frequency band is known as full-duplex while a half-duplex node cannot simultaneously receive and transmit signals. In practice it is expensive and hard to build a communication device which can transmit and receive signals at the same time using the same frequency, due to the self-interference created by the transmitted signal to the received signal. Therefore half-duplex systems are mostly used in practice. Consider a cascade network comprised of \( L - 1 \) half-duplex relay nodes depicted in Fig. 2, where the state encoder \( E \) observes the state of the system and transmits its signal to the relay node \( R_1 \). The relay node \( R_1 \) transmits a signal to the relay node \( R_2 \) and so on. Finally the state information is received at the remote decoder/controller \( D \) from \( R_{L-1} \). The communication within the network takes place such that only one node is allowed to transmit at every time step. That is, if in a time slot \( R_t \) transmits signal to \( R_{t+1} \), then all the remaining nodes in the network are considered to be silent in that time slot. At any time step \( t \), \( S_{e,t} \) is the signal transmitted from \( E \) and \( S_{r,t} \) is the signal transmitted from \( R_t \), which are given by

\[
S_{e,t} = f_t \left( X_{[0,t-1],U_{[0,t-1]}} \right) \quad \forall t : t = 1 + nL, n \in \mathbb{N},
\]

\[
S_{e,t} = 0 \quad \text{otherwise}
\]

\[ S_{r,t} = S_{r,t}^* = 0 \]

\[
S_{r,t} = 1 + nL, n \in \mathbb{N},
\]

\[ S_{r,t} = 0 \quad \text{otherwise}. \tag{3} \]

Here \( Z_i^1 \sim \mathcal{N}(0,N_i) \) denotes mutually independent white Gaussian noise components. Accordingly \( D \) receives \( R_t = S_{r,t}^* + Z_t^1 \) at \( t = nL \) and zero otherwise.

We now present a necessary condition for mean square stability over the given channel.

**Theorem 3.1:** If the system (1) is mean square stable over the cascade network, then

\[
\log |\det(A)| < \frac{1}{2L} \log \left( 1 + L \min \left\{ \frac{P_S}{N_1}, \frac{P_R}{\sum_{i=2}^{L} N_i} \right\} \right). \tag{5} \]

**Proof 3.1:** We first derive an outer bound on the directed information \( I(X_{[1,L]} \rightarrow R_{[1,L]}) \) over the given channel and then use Theorem 2.1 to find the necessary condition (5).
where (a) follows from the Markov chain $\bar{X}_{1,[L,T]} - (S_{r,[i]}^{-1} - R_{1,[i-1]} - R_i)$; (b) follows from [42, Theorem 1]; (c) follows from (3) and (4); and (d) follows from the fact that mutual information of a Gaussian channel is maximized by the Gaussian input distribution [35, Theorem 8.6.5]. Finally using (6), (7), and (8), we have the following bound:

\[
I(\bar{X}_{1,[L,T]} \rightarrow R_{1,[L,T]}) \leq \frac{T}{2} \min \left\{ \log \left( 1 + \frac{LP_S}{N_1} \right), \ldots, \log \left( 1 + \frac{LP_L^{L-1}}{N_L} \right) \right\}
\]

(a) follows from the fact that $\log(1+x)$ is a monotonically increasing function of $x$; and (b) follows from the optimal power allocation choice $P_{r,i} = N_{i+1} P_R / \sum_{i=2}^L N_i$. Finally dividing (9) by $LT$ and letting $T \rightarrow \infty$ according to Theorem 2.1, we get the necessary condition (5).

We now present a sufficient condition for mean-square stability over the given network.

**Theorem 3.2:** The scalar linear time invariant system in (1) with $A = \lambda$ can be mean square stabilized using a linear scheme over a cascade network of $L$ relay nodes if

\[
\log(|\lambda|) \leq \frac{1}{2L} \log \left( 1 + \frac{LS_P}{LP_S + N_1} \sum_{i=2}^L \frac{LP_{r,i}}{LP_{r,i} + N_{i-1}} \right)
\]

(10) where the optimal power allocation is given by $P_{r,i} = (N_{i+1} + (N_{i+1}^2 - (4N_{i+1}/\gamma))/2$ and $\gamma < 0$ is chosen such that $\sum_{i=1}^{L-1} P_{r,i} \leq P_R$. When all $N_i$ are equal, the optimal choice is $P_{r,i} = P_R / (L - 1)$.

**Outline of Proof:** The result can be derived by using a memoryless linear sensing and control scheme. Under linear policies, the overall mapping from the encoder to the controller becomes a scalar Gaussian channel, which has been well studied in the literature (see for example [1]). Due to space constraints, we refer the reader to the proof of Theorem 5.2, which contains a detailed derivation for the non-orthogonal network and the proof for this setting is similar. The optimal power allocation follows from the concavity of $\prod_{i=1}^{L-1} (LP_{r,i} / (LP_{r,i} + N_{i-1}))$ in $\{P_{r,i}\}_{i=1}^{L-1}$ and by using the Lagrange multiplier method.

**Remark 3.1:** For fixed power allocations, as the number of relays $L$ approaches infinity in (5), the RHS converges to zero and stabilization becomes impossible. We also note that the ratio between the sufficiency and necessity bounds converges to zero as the number of relays goes to infinity.
If the goal is not stabilization, but optimization with a minimum mean-square estimation error or minimum second moment, it is shown in [43], [44] (see also [30] and [32]) that linear sensing policies may or may not be optimal depending on the topology of the system and in particular such schemes are suboptimal for the general relay channels.

IV. PARALLEL NETWORK

Consider the network in Fig. 3, where the signal transmitted by a node does not interfere with the signals transmitted by other nodes, i.e., there are L parallel channels from \( \{R_i\}_{i=1}^L \) to \( D \). We call this setup a parallel network, modeling a scenario over the given parallel network.

![Fig. 3. Parallel relay network.](image)

Theorem 4.1: Consider the network in Fig. 3, where the signal transmitted to the relays and in the second phase all relays simultaneously transmit to the remote controller with an average power \( P_r \). We call this setup a parallel network, modeling a scenario over the Gaussian parallel network.

Proof 4.1: Following the same steps as in proof of Theorem 3.1, we can bound \( I(\hat{X}_{[1,2T]}; R_{[1,2T]}^L) \) as,

\[
I(\hat{X}_{[1,2T]}; R_{[1,2T]}^L) \overset{(a)}{=} \min \left\{ \sum_{i=1}^T \left( S_{r_i,2t-1}; Y_{2t-1}^i \right)_i^L, \sum_{i=1}^T \left( S_{r_i,2t}; R_{2t}^i \right)_i^L \right\}
\]

where (a) follows from the same steps as in (6) and (8); (b) follows from (11); and (c) follows from the fact that Gaussian input distribution maximizes mutual information for a Gaussian channel. The function \( \sum_{i=1}^L \log(1 + (2P_r^i/N_d^i)) \) is jointly concave in \( \{P_r^i\}_{i=1}^L \). The optimal power allocation is given by \( P_r^i = \max\{\gamma - N_d^i/2, 0\} \), where \( \gamma \) is chosen such that \( \sum_{i=1}^L P_r^i = P_r \), which is the well-known water-filling solution [42, pp. 204–205]. We obtain (12) by using (13) in Theorem 2.1.

Proof 4.2: The above result can be obtained by using a memoryless linear sensing and control scheme and as discussed in the proof of Theorem 3.2.

Proposition 4.1: The gap between the necessary and sufficient conditions for a symmetric parallel network with \( P_r^i = P_r \) is a non-decreasing function of \( L \) and approaches \( (1/4) \log(1 + (N_d(2P_r N_r + N_d)/(2P_r N_r))) \) as \( L \to \infty \).

Proof 4.3: For \( P_r^i = P_r \) and \( N_r^i = N_r^i \), the RHS of (14) is evaluated as \( \Gamma_{\text{suf}} := (1/4) \log(1 + (4LP_r P_r^i/(2P_r N_d + 2P_r N_r + N_d N_r))) \) and the RHS of (12) can be bounded as \( \Gamma_{\text{nec}} := (1/4) \log(1 + (2LP_r N_r)) \). The gap is given by

\[
\Gamma_{\text{nec}} - \Gamma_{\text{suf}} = \frac{1}{4} \log \left( \frac{2LP_r N_d (2P_r + N_r)}{4P_r P_r^i N_r + 2LP_r N_d N_r + 2P_r N_r + N_d N_r} \right)
\]

which is an increasing function of \( L \), approaching \( (1/4) \log(1 + (N_d(2P_r N_r + N_d)/(2P_r N_r))) \) as \( L \to \infty \).

Remark 4.1: If \( N_r^i = 0 \), then \( \Gamma_{\text{nec}} - \Gamma_{\text{suf}} = 0 \) and the linear scheme is exactly optimal. For \( N_r^i = 0 \), \( \Gamma_{\text{suf}} := (1/4) \log(1 + (2P_r N_d)) \) according to (12). Clearly \( \lim_{L \to \infty} (\Gamma_{\text{nec}} - \Gamma_{\text{suf}}) = \infty \), showing the inefficiency of the LTI scheme for parallel channels.

It is known that linear schemes can be sub-optimal for transmission over parallel channels [29], [46]. A distributed joint source—channel code is optimal in the sense of minimizing.
mean-square distortion if the following two conditions hold [47]: i) All channels from the source to the destination send independent information; ii) All channels utilize the capacity, i.e., the source and channel need to be matched. If we use linear policies at the relay nodes, then the first condition is not fulfilled because all nodes would be transmitting correlated information. In [17] the authors proposed a non-linear scheme for a parallel network of two sensors without relays, in which one sensor transmits the magnitude of the observed state and the other sensor transmits the phase of the observed state. The magnitude and phase of the state are shown to be independent information. In [17] the authors proposed a non-linear scheme for the LQG control problem in the absence of measurement noise, although the second condition of source-channel matching is not fulfilled. We can use this non-linear scheme together with the initialization step of the Schalkwijk Kailath (SK) type scheme described in Appendix B for the non-orthogonal network, which will ensure source-channel matching by making the outputs of the two sensors Gaussian distributed after the initial transmissions. In [48] it is shown that linear sensing policies may not be even person-by-person optimal for LQG control over parallel network without relays.

For a special case with noiseless $E - R_i$ links, we have the following necessary and sufficient condition.

**Theorem 4.3:** The system (1) in absence of process noise ($W_t = 0$) can be mean square stabilized over the Gaussian parallel network with $Z_{r,t} = 0$ for all $i$, only if

$$\log(|\det(A)|) \leq \frac{1}{4} \sum_{i=1}^{L} \log \left( 1 + \frac{2P_i}{N_d} \right).$$

(16)

If the inequality is strict, then there exists a non-linear policy leading to mean-square stability.

**Proof 4.4:** The necessity follows from Theorem 4.1. The sufficiency part for scalar systems follows from [28, Theorem 6], which is derived using a non-linear scheme. This scheme can be extended to vector systems using a time sharing scheme presented in Section VI.

**Remark 4.2:** The term on RHS of (16) is equal to information capacity of the parallel Gaussian channel. It was shown by Shu and Middleton in [24] that for some first order noiseless plants, linear time invariant encoders/decoders cannot achieve the information capacity of parallel Gaussian channels. However, if the channel capacity for stabilization can always be achieved by a non-linear time varying scheme as discussed in the proof of Theorem 4.3.

V. NON-ORTHOGONAL NETWORK

A communication network is said to be non-orthogonal if all nodes transmit signals in overlapping time slots using the same frequency bands. In this section, we study non-orthogonal networks with half-duplex and full-duplex configurations.

A. Non-Orthogonal Half-Duplex Network

A non-orthogonal half-duplex Gaussian network with $L$ relay nodes $\{R_i\}_{i=1}^{L}$ is illustrated in Fig. 4. The variables $S_{e,t}$ and $S_{r,t}$ denote the transmitted signals from the state encoder $E$ and relay $R_i$ at any discrete time step $t$. The variables $Z_{r,t}$ and $Z_{d,t}$ denote the mutually independent white Gaussian noise components at the relay node $i$ and $D$ of the remote control unit, with $Z_{r,t} \sim N(0,N_{r})$ and $Z_{d,t} \sim N(0,N_{d})$. The noise components $\{Z_{r,t}\}_{i=1}^{L}$ are independent across the relays, i.e., $E[Z_{r,t}Z_{r,t}^\dagger] = 0$ for all $i \neq k$. The information transmission from the state encoder consists of two phases as shown in Fig. 4. In the first phase the encoder $E$ transmits a signal with an average power $2 \beta P_S$, where $0 < \beta \leq 1$ is a parameter that adjusts power between the two transmission phases. In this transmission phase all the relay nodes listen but remain silent.

In the second transmission phase, the encoder $E$ and relay nodes $\{R_i\}_{i=1}^{L}$ transmit simultaneously. In this second transmission phase, the encoder transmits with an average power $2(1 - \beta)P_S$ and the $i$-th relay node transmits with an average power $2P_i$ such that $\sum_{i=1}^{L} P_i \leq P_R$. The input and output of the $i$-th relay are given by

$$Y_t^i = S_{e,t} + Z_{r,t}^i, \quad S_{r,t}^i = 0, \quad t = 1, 3, 5, \ldots$$

$$Y_t^i = 0, \quad S_{r,t}^i = g_t^i \left( Y_{t-1}^{i+1} \right), \quad t = 2, 4, 6$$

(17)

where $g_t^i : \mathbb{R}^{i+1} \to \mathbb{R}$ is the $i$-th relay encoding policy such that $E[(g_t^i(Y_{t-1}^{i+1}))^2] = 2P_i$ and $\sum_{i=1}^{L} P_i \leq P_R$. The signal received at the decoder/controller is given by

$$R_t = hS_{e,t} + \sum_{i=1}^{L} h_i S_{r,t}^i + Z_{d,t}$$

where $h, h_i \in \mathbb{R}$ denote the channel gains of $E - D$ and $R_i - D$ links respectively.

![Fig. 4.](image-url)
Theorem 5.1: If the linear system in (1) is mean-square stable over the non-orthogonal half-duplex relay network, then
\[
\log(\det(A)) \leq -\frac{1}{4} \min_{0 < \delta \leq 1} \left( \max_{0 \leq \beta \leq 1} \left( \log \left( 1 + \frac{2h^2(1 - \beta)P_S}{N_d} \right) \right) + \log \left( 1 + 2\beta P_S \left( \sum_{i=1}^{L+1} \frac{1}{N_i} + \frac{h^2}{N_d} \right) \right) \right) + \log \left( 1 + \frac{1 + 2h^2\beta P_S}{N_d} \left( \sum_{i=1}^{L+1} \delta_i \rho \delta_i \delta_k \sqrt{P_i P_k} \right) \right)
\]

where \( \rho_{i,k}^* = 2(1 - \beta)P_S / \sqrt{(2(1 - \beta)P_S + N_i)(2(1 - \beta)P_S + N_k)} \), \( P_{t+1} = 2(1 - \beta)P_S \), \( N_{t+1} = 0 \), \( \delta_{t+1} = h \), \( P_t = 2P_r \), \( \delta_t = h \), \( N_i = N_i^* \) for all \( i = \{1, 2, \ldots, L\} \).

Proof 5.1: We first derive an outer bound on the directed information \( I(X_{[1,LT]} \to Y_{[1,LT]}) \) over the given channel and then use Theorem 2.1 to find the necessary condition (18).

\[
I \left( X_{[1,2T]} \to R_{[1,2T]} \right) \leq I \left( X_{[1,2T]} \to \left\{ Y_i^{\infty}_{t=1} \right\}_{i=1}^{L} \middle| R_{[1,2T]} \right) \\
= \sum_{t=1}^{2T} \left[ I \left( X_{t,T} \to \left\{ Y_i^{\infty}_{t=1} \right\}_{i=1}^{L} \middle| R_{[1,2T]} \right) \right. \\
= \left. \sum_{t=1}^{2T} \sum_{i=1}^{L} \left( S_{c,t} \to \left\{ Y_i^{\infty}_{t=1} \right\}_{i=1}^{L} \middle| R_{[1,2T]} \right) \right] \\
= \sum_{t=1}^{2T} \sum_{i=1}^{L} \left( \sum_{i=1}^{L} \delta_i \rho \delta_i \delta_k \sqrt{P_i P_k} \right)
\]

where \( \rho_{i,k}^* = 2(1 - \beta)P_S / \sqrt{(2(1 - \beta)P_S + N_i)(2(1 - \beta)P_S + N_k)} \), \( P_{t+1} = 2(1 - \beta)P_S \), \( N_{t+1} = 0 \), \( \delta_{t+1} = h \), \( P_t = 2P_r \), \( \delta_t = h \), \( N_i = N_i^* \) for all \( i = \{1, 2, \ldots, L\} \).

The inequality (a) follows from the Markov chain \( X_{[0,t]} \to S_{c,t} \to \left\{ S_{r,t}^{L}_{i=1} \right\} \to R_t \) due to the memoryless channel between \( S_{c,[1,2T]} \) and \( S_{r,[1,2T]} \) and \( R_t \) due to the fact that the first addend on the RHS of (c) is maximized by a centered Gaussian distributed \( S_{c,t} \) and the second addend is bounded using a bound presented in [49], where the author studied the problem of transmitting a Gaussian source over a simple sensor network. In order to apply the upper bound given in (48) of [49] to our setup, we consider encoder \( E \) to be a sensor node with zero observation noise and make the following change of variables so that our system model becomes equivalent to the one discussed in [49]:

\( \sigma_Z^2 := \alpha \), \( \delta_i := h \), \( M := L + 1 \), \( P_t := 2P_r \), \( \delta^Z := N_d \), \( \rho_{i,k}^* := N_i^* \), \( \alpha := \sqrt{(1 - \beta)P_S / \alpha} \).

We finally obtain (18) by dividing (19) and (20) by 2T and let \( T \to \infty \) according to Theorem 2.1.\]
We now present a sufficient condition for stability.

**Theorem 5.2:** The scalar linear time invariant system in (1) with \( A = \lambda \) can be mean square stabilized using a linear scheme over the non-orthogonal half-duplex network if

\[
\log (|\lambda|) < \frac{1}{4} \max_{0 < \beta \leq 1} \left\{ \log \left( 1 + \frac{2h^2 \beta P_S}{N_d} \right) \right\} + \log \left( 1 + \frac{M(\beta, \{ P_r^i \}_{i=1}^L)}{N(\beta, \{ P_r^i \}_{i=1}^L)} \right),
\]

where

\[
M(\beta, \{ P_r^i \}_{i=1}^L) = \sqrt{2\beta P_SN_d/(2h^2(2\beta P_S + N_d)))} \sum_{i=1}^L \frac{P_r^i}{(2h^2 P_r^i/2\beta P_S + N_r^i)} + N_d.
\]

and

\[
N(\beta, \{ P_r^i \}_{i=1}^L) = \sum_{i=1}^L (2h^2 P_r^i N_r^i/2\beta P_S + N_r^i) + N_d.
\]

**Proof 5.2:** The proof is given in Appendix B. \( \square \)

**Remark 5.1:** An optimal choice of the power allocation parameter \( \beta \) at the state encoder and an optimal power allocation at the relay nodes \( \{ P_r^i \}_{i=1}^L \) which maximizes the term on the right hand side of (21) depend on the quality of the \( \mathcal{E} - \mathcal{D}, \mathcal{E} - \mathcal{R}_i, \) and \( \mathcal{R}_i - \mathcal{D} \) links. This is a non-convex optimization problem, however it can be transformed into an equivalent convex problem by using the approach in [50, Appendix A]. This equivalent convex problem can be efficiently solved for optimal \( \{ P_r^i \}_{i=1}^L \) using the interior point method. For \( \beta = 1 \), we can analytically obtain the following optimal power allocation method using the Lagrangian method:

\[
P_r^i = P_R \left( \frac{h_r^2 (2P_S + N_r^i)}{(2P_S N_d + N_r^i N_d + P_R h_r^2 N_r^i)^2} \right) \times \left( \sum_{i=1}^L \frac{h_r^2 (2P_S + N_r^i)}{(2P_SN_d + N_r^i N_d + P_R h_r^2 N_r^i)^2} \right)^{-1}. \tag{22}
\]

**Remark 5.2:** For channels with feedback, directed information is a useful quantity [42], [51]. It is shown in Appendix C that the term on the right hand side of (21) is the information rate over the half-duplex network with noiseless feedback, obtained when running the described closed-loop protocol. Further we show that the directed information rate is also equal to the term on the right hand side of (21). By following the same steps as in Appendix C, one can also obtain relationships between sufficient conditions and information rates under linear policies for the cascade and the parallel relay networks considered in Section III and Section IV respectively. It can be shown that the directed information rate between the sequence of channel inputs and the sequence of channel outputs also gives an outer bound for the term on the right hand side of (2).

**B. Two-Hop Network**

Consider the half-duplex network illustrated in Fig. 4 with \( h = 0 \). The state information is communicated to the remote controller only via the relays. We call this setup a two-hop relay network, where the communication from the state encoder to the controller takes place in two hops. In the first hop, the relay nodes receive the state information from the state encoder, which then communicate the state information to the controller in the second hop. The controller then can control the state information to the remote controller in the second hop. Further by restricting the state encoder to be linear, the relay network becomes equivalent to the Gaussian network studied in [49], [52], where it is shown that linear policies are optimal if the network is symmetric.
C. Non-Orthogonal Full-Duplex Network

We now consider a non-orthogonal network of $L$ full-duplex relay nodes, where all the nodes receive and transmit their signals in every time step, i.e., at any time instant $t \in \mathbb{N}$

$$S_{e,t} = f_t(X_{[0,t]}, U_{[0,t-1]}), \quad S_{i,t}^i = g_i^i(Y_{i,[0,t-1]}),$$

$$Y_t^i = S_{e,t} + Z_{i,t}^i, \quad R_t = h S_{e,t} + \sum_{i=1}^L S_{i,t}^i + Z_{d,t}, \quad (24)$$

where $E[(S_{e,t})^2] = P_S$, $E[(S_{i,t}^i)^2] = P_{i,t}^i$, $\sum_{i=1}^L P_{i,t}^i \leq P_R$.

**Theorem 5.3:** If the linear system in (1) is mean-square stable over the non-orthogonal full-duplex relay network, then

$$\log(|\det(A)|) \leq \frac{1}{2} \min\left\{ \log \left( 1 + P_S \left( \sum_{i=1}^L \frac{1}{N_r} + h^2 N_d \right) \right), \right.$$  

$$\max_{P_{i,t}^i} \left\{ \log \left( 1 + \frac{1}{N_d} \sum_{i=1}^{L+i} \delta_i^2 P_i \right) + 2 \sum_{i=1}^{L+i} \sum_{k=1} P_{i,t}^i \delta_i \delta_k \sqrt{P_{i,t}^i P_{k,t}^k} \right\} \right\}, \quad (25)$$

where $\rho_{i,k}^* = P_S \sqrt{(P_S + N_i)(P_S + N_k)}$, $P_{L+1,i} := P_S$, $N_{L+1,i} := 0$, $\delta_{L+1,i} := h$, $P_t := P_{i,t}$, $\delta_t := h_t$, $N_t := N_{i,t}$ for all $i = 1, \ldots, L$.

**Proof:** The proof follows exactly in the style of the proof of Theorem 5.1, with an exception that odd and even indexed terms are not treated separately because $E[S_{e,t}^2] = P_S^2$ and $E[(S_{i,t}^i)^2] = P_{i,t}^i$ for all $t$.

**Theorem 5.4:** The scalar linear time invariant system in (1) with $A = \lambda$ and $W_t = 0$ can be mean square stabilized using a linear scheme over the non-orthogonal full-duplex Gaussian network if

$$\log(|\lambda|) < \frac{1}{2} \max_{P_{i,t}^i} \left\{ \log \left( 1 + \left( \sqrt{h^2 P_S + \eta^*} \sum_{i=1}^L \frac{h_i^2 P_{i,t}^i P_{i,t}}{P_S + N_i^2} \right)^2 \left( N_d + \sum_{i=1}^L \frac{h_i^2 P_{i,t}^i N_i^2}{P_S + N_i^2} \right)^{-1} \right) \right\}, \quad (26)$$

where $\eta^*$ is the unique solution in the interval $[0,1]$ of

$$\left( \sum_{i=1}^L \frac{h_i^2 P_{i,t}^i P_{i,t}}{(P_S + N_i^2)} \right) \eta^4 + \left( 2 h P_S \sum_{i=1}^L \frac{\sqrt{h_i^2 P_{i,t}^i P_{i,t}}}{(P_S + N_i^2)} \right) \eta^3 + \left( h^2 P_S + N_d + \sum_{i=1}^L \frac{h_i^2 P_{i,t}^i N_i^2}{P_S + N_i^2} \right) \eta^2 = \left( N_d + \sum_{i=1}^L \frac{h_i^2 P_{i,t}^i N_i^2}{P_S + N_i^2} \right).$$

**Proof 5.5:** The proof can be found in [37] for a single relay setup, which can be easily extended for multiple relays.

Although we expect that Theorem 5.4 holds in the presence of process noise like in other setups, we have not shown the convergence of second moment of the state process. Numerical experiments suggest that the result should hold.

**Remark 5.4:** The term on the right hand side of the inequality in (26) is an achievable information rate$^3$ with which information can be transmitted reliably over the non-orthogonal full-duplex relay network in an information theoretic sense, that is, using codes of unbounded block lengths (see [35] for the typical constructions and Chapter V in [30] for the connection with real-time systems of such an operational use of information rate). This result is derived for a network with single relay node in [53, Theorem 5], however it can be easily extended to problems with multiple relays.

VI. Noisy Multi-Dimensional Systems

In this section we investigate stabilization of multi-dimensional systems over multi-dimensional channels. First we state a result for a scalar Gaussian channel.

**Theorem 6.1:** The $n$-dimensional noisy linear system (1) can be mean square stabilized over a scalar Gaussian channel having information capacity $C$, if $\log(|\lambda|) < C$. Furthermore, a linear time varying policy is sufficient through sequential linear encoding of scalar components.

**Proof Outline:** We outline a proof of Theorem 6.1 with the help of a simple example, due to space limitation in the paper. Consider a two-dimensional plant with system matrix $A = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix}$ and an invertible input matrix $B$ has to be stabilized over a Gaussian channel disturbed by a zero mean Gaussian noise with variance $N$. We assume that the sensor transmits with an average $P$. For this channel, the information capacity is obtained as $C := (1/2) \log(1 + (P/N))$ [35, Theorem 9.1.1]. We denote the state and the control variables as $X_t := [x_{1,t}, x_{2,t}]^T$ and $U_t := [u_{1,t}, u_{2,t}]^T$ respectively. Consider the following scheme for stabilization. The sensor observes state vector $X_t$ in alternate time steps (that is, at $t, t+2, t+4, \ldots$), whose elements are sequentially transmitted. The sensor linearly transmits $x_{2,t}$ at time $t$ and $x_{1,t}$ at time $t+1$ with an average transmit power constraint. The control actions for the two modes are also taken in alternate time steps, that is, $u_{1,t} = 0$ and $u_{2,t+1} = 0$. Accordingly the state equations for the two modes at time $t+1$ are given by

$$x_{2,t+1} = \lambda_2 x_{2,t} + u_{2,t} + w_{2,t}$$

where $(a)$ and $(b)$ follow from $u_{2,t} = -\lambda_2 x_{2,t}$ and $u_{1,t} = 0$. The state equations at time $t + 2$ are

$$x_{2,t+2} = \lambda_2 x_{2,t+1} + w_{2,t+1}$$

where $(b)$ follow from $u_{2,t+1} = \lambda_2 x_{2,t+1}$

$$x_{1,t+2} = \lambda_1 x_{1,t+1} + x_{2,t+1} + u_{1,t+1} + w_{1,t+1}$$

where $(b)$ follow from $u_{1,t+1} = \lambda_1 x_{1,t+1}$

$$w_{2,t+1} = \lambda_1 x_{1,t} - \hat{x}_{1,t} + \lambda_1 x_{2,t} - \hat{x}_{2,t}$$

where $(b)$ follow from $u_{2,t} = -\lambda_2 x_{2,t}$ and $u_{1,t} = 0$. The state equations at time $t + 2$ are

$$x_{2,t+2} = \lambda_2 x_{2,t+1} + w_{2,t+1}$$

where $(b)$ follow from $u_{2,t+1} = \lambda_2 x_{2,t+1}$. Numerical experiments suggest that the result should hold.

$^3$The definition of achievable rate for Gaussian channels is given on page 264 in [35].
where (a) follows (28); and (b) follows from $u_{1,t+1} = -\lambda_1^2\hat{x}_1,t - (\lambda_1 + \lambda_2) \hat{x}_2,t$. We first study the stabilization of the lower mode. According to (29) the second moment of $x_{2,t}$ is

$$\mathbb{E}[x_{2,t}^2] = \lambda_2^2\mathbb{E}[(x_{2,t} - \hat{x}_{2,t})^2] + \tilde{n}_2$$

$$= \lambda_2^2 - 2\sigma^2 \mathbb{E}[x_{2,t}^2] + \tilde{n}_2.$$

(31)

where the last equality follows from the linear mean-square estimation of a Gaussian variable over a scalar Gaussian channel of capacity $C$ and $\tilde{n}_2 := (\lambda_3^2 + 1)u_{w,2}$. We observe that the lower mode is stable if and only if $\lambda_2^2 - 2\sigma^2 < 1 \Rightarrow \log(|\lambda_2|) < (C/2)$. Assuming that $x_{2,t}$ is stable, the second moment of $x_{1,t}$ is given by

$$\mathbb{E}[x_{1,t+1}^2] = \lambda_1^2\mathbb{E}[(x_{1,t} - \hat{x}_{1,t})^2]$$

$$+ 2\lambda_1^2(\lambda_1 + \lambda_2)\mathbb{E}[(x_{1,t} - \hat{x}_{1,t})(x_{2,t} - \hat{x}_{2,t})]$$

$$+ (\lambda_1 + \lambda_2)^2\mathbb{E}[(x_{2,t} - \hat{x}_{2,t})^2] + \tilde{n}_1$$

$$= \lambda_1^2 - 2\sigma^2 \mathbb{E}[x_{1,t}^2] + 2\lambda_1^2(\lambda_1 + \lambda_2)\sqrt{2-2\sigma^2} \mathbb{E}[x_{2,t}^2] + \tilde{n}_1$$

$$+ \sqrt{2-2\sigma^2} \mathbb{E}[x_{2,t}^2] + (\lambda_1 + \lambda_2)^2\mathbb{E}[x_{2,t}^2] + \tilde{n}_1$$

$$\leq k_1\mathbb{E}[x_{2,t}^2] + k_2\sqrt{\mathbb{E}[x_{2,t}^2]} + k_3.$$

(32)

where (a) follows from (30) and $\tilde{n}_1 := (\lambda_1^2 + 1)u_{w,1} + u_{w,2}$; (b) follows from the linear mean-square estimation of a Gaussian variable over a scalar Gaussian channel of capacity $C$; (c) follows from the Cauchy–Schwarz inequality; (d) follows from the fact $\mathbb{E}[x_{1,t}^2] < M$ (assuming that $\lambda_2^2 - 2\sigma^2 < 1$) and by defining $k_1 := \lambda_1^2 - 2\sigma^2$, $k_2 := 2\lambda_1^2(\lambda_1 + \lambda_2)\sqrt{2-2\sigma^2} M$, and $k_3 := (\lambda_1 + \lambda_2)^22^{-2\sigma^2}M + \tilde{n}_1$. We now want to a find condition which ensures convergence of the following sequence:

$$\alpha_{t+1} = k_1\alpha_t + k_2\sqrt{\alpha_t} + k_3.$$

(33)

To show convergence, we make use of the following lemma.

**Lemma 6.1:** Let $T: \mathbb{R} \rightarrow \mathbb{R}$ be a non-decreasing continuous mapping with a unique fixed point $x^* \in \mathbb{R}$. If there exists $u \leq 0 \leq v$ such that $T(u) \geq u$ and $T(v) \leq v$, then the sequence generated by $x_{t+1} = T(x_t), t \in \mathbb{N}$ converges starting from any initial value $x_0 \in \mathbb{R}$.

**Proof 6.1:** The proof is given in Appendix D. □

We observe that the mapping $T(\alpha) = k_1\alpha + k_2\sqrt{\alpha} + k_3$ with $\alpha \geq 0$ is monotonically increasing since $k_1, k_2 > 0$. It will have a unique fixed point $\alpha^*$ if and only if $k_1 < 1$, since $k_2, k_3 > 0$. Assuming $k_1 < 1$, there exists $u < \alpha^* < v$ such that $T(u) \geq u$ and $T(v) \leq v$. Therefore, by Lemma 6.1, $\{\alpha_t\}$ is convergent if $k_1 = \lambda_1^2 - 2\sigma^2 < 1 \Rightarrow \log(|\lambda_1|) < (C/2)$.

**VII. CONCLUSION**

The problem of mean-square stabilization of LTI plants over Gaussian relay networks is analyzed. Necessary and sufficient conditions for stabilization are presented which reveal relationships between stabilizability and communication parameters. These results can serve as a guideline for a system designer. Necessary conditions have been derived using information theoretic cut-set bounds, which are not tight in general due to the real-time nature of the information transmission. Sufficient conditions for stabilization of scalar plants are obtained by employing time invariant communication and control schemes.

We have shown that time invariant schemes are not sufficient in general for stabilization of multi-dimensional plants. However, a simple time variant scheme is always shown to stabilize...
multi-dimensional plants. In this time varying scheme, one component of the state vector is transmitted at a time and the state component corresponding to a more unstable mode is transmitted more often. The sufficient conditions for stabilization of multi-dimensional plants are obtained by using this time varying scheme. We also established minimum signal-to-noise ratio requirement for stabilization of a noiseless multi-dimensional plant over a parallel Gaussian channel. It is observed in some network settings that sufficient conditions do not depend on the plant noise and they can be characterized by the directed information rate between the sequence of channel inputs and the sequence of channel outputs.

APPENDIX

A. Necessary Condition

Consider the following series of inequalities:

\[
I (X_{[0,T-1]} \rightarrow R_{[0,T-1]}) \leq \sum_{t=0}^{T-1} I (X_{[t]}; R_t|R_{[0,t-1]}) \leq I (X_0; R_0) + \sum_{t=1}^{T-1} (h (X_t|R_{[0,t-1]}) - h (X_t|R_{[0,t]}))
\]

where \((a)\) follows from the definition of directed information; \((b)\) follows from the fact that discarding variables cannot increase mutual information; \((c)\) follows by writing mutual information in terms of differential entropies and from the fact that these differential entropies are finite for a mean-square stable system due to finite second moment of the state process; \((d)\) follows from \((1)\); \((e)\) follows by using \(U_{t-1} = \pi_{t-1} (R_{[0,t-1]})\); \((f)\) follows from the fact that conditioning reduces entropy; \((g)\) follows from \(h (AX_{t-1} + W_{t-1}|R_{[0,t-1]}, W_{t-1}) = h (AX_{t-1}|R_{[0,t-1]}, W_{t-1}) = h (AX_{t-1}|R_{[0,t-1]})\) due to mutual independence of \(X_t\) and \(W_t\); \((h)\) follows from \(h (AX_t) = \log (|\det (A)|) + h (X)\) \cite[Theorem 8.6.4]{Cover2012}; \((i)\) follows from conditioning reduces entropy; and \((j)\) follows the fact that for a mean square stable system there exists a matrix \(K > 0\) with \(\det (E[X_t^T X_t]) < \det (K)\) for all \(t\) and further for a given covariance matrix the differential entropy is maximized by the Gaussian distribution. We can also write

\[
I (X_{[0,T-1]} \rightarrow R_{[0,T-1]}) = \sum_{t=0}^{T-1} I (X_{[t]}; R_t|R_{[0,t-1]})
\]

with \((a)\) follows by defining uncontrolled state process \(\hat{X}_{t+1} = A\hat{X}_t + W_t\) and writing \(X_{[t]}\) as a sum of \(\hat{X}_{[t]}\) and a linear function of \(U_{[0,t-1]}\), since the system is linear and control actions are additive; and \((b)\) follows from \(U_t = \pi_t (R_{[0,t]})\). From \((34)\) and \((35)\) we have \(\lim_{T \rightarrow \infty} (1/T) I (X_{[0,T-1]} \rightarrow R_{[0,T-1]}) \geq \log (|\det (A)|)\), since \(h (X_0) < \infty\).

B. Proof of Theorem 5.2

In order to prove Theorem 5.2, we propose a linear communication and control scheme. This scheme is based on the coding scheme given in \cite{Weissman2008} which is an adaptation of the well-known Schalkwijk—Kailath scheme \cite{Schalkwijk1969}. By employing the proposed linear scheme, we find a condition on the system parameters \(\lambda\) which is sufficient to mean square stabilize the system \((1)\). The scheme works as follows: If the initial state \(X_0\) is not Gaussian distributed, then we first make the state process Gaussian distributed by performing the following initialization.

Initial Time Step, \(t = 0\): At time step \(t = 0\), the state encoder \(E\) observes \(X_0\) and it transmits \(S_{e,0} = \sqrt{P_S/\alpha_0} X_0\). The decoder \(D\) receives \(R_0 = h S_{e,0} + Z_{d,0}\). It estimates \(X_0\) as \(\hat{X}_0 = (1/h) \sqrt{\alpha_0/P_S} R_0 = X_0 + (1/h) \sqrt{\alpha_0/P_S} Z_{d,0}\). The controller \(C\) then takes an action \(U_0 = -\lambda \hat{X}_0\) which results in

\[
X_1 = \lambda X_0 + U_0 + W_0 = \lambda (X_0 - \hat{X}_0) + W_0 = -\lambda \hat{X}_0 + (1/h) \sqrt{\alpha_0/P_S} Z_{d,0} + W_0.
\]

The new plant state \(X_1 \sim \mathcal{N} (0, \alpha_1)\) where \(\alpha_1 = (\lambda^2 N_d/h^2 P_S) \alpha_0 + n_w\).
\[ \hat{X}_t = \mathbb{E}[X_t | R_{[1,t]}] = \mathbb{E}[X_t | R_t] = \mathbb{E}[X_t R_t | R_t^2] R_t, \]

where \((a)\) follows from the orthogonality principle of MMSE estimation (that is \(\mathbb{E}[X_t | R_{t-j}] = 0\) for \(j \geq 1\)) \([55]\); \((b)\) follows from the fact that the optimum MMSE estimator for a Gaussian variable is linear \([55]\); and \((c)\) follows from \(\mathbb{E}[X_t R_t] = \sqrt{2h^2 \beta P_S \alpha_t} \) and \(\mathbb{E}[R_t^2] = 2h^2 \beta P_S + N_d\).

The controller \(C\) takes an action \(U_t = -\lambda \hat{X}_t\) which results in \(X_{t+1} = \lambda (X_t - \hat{X}_t) + W_t\). The new state \(X_{t+1}\) is a linear combination of zero mean Gaussian variables \(\{X_t, \hat{X}_t, W_t\}\), therefore it is also zero mean Gaussian with variance

\[ \alpha_{t+1} := \mathbb{E}[X_{t+1}^2] = \lambda^2 \mathbb{E}[(X_t - \hat{X}_t)^2] + \mathbb{E}[W_t^2] = \lambda^2 \left( \frac{N_d}{2h^2 \beta P_S + N_d} \right) \alpha_t + N_w. \]  

where the last equality follows from \(\mathbb{E}[X_t \hat{X}_t] = \mathbb{E}[\hat{X}_t^2] = 2h^2 \beta P_S \alpha_t/(2h^2 \beta P_S + N_d)\) by computation.

Second transmission phase, \(t = 2, 4, 6, \ldots\): The encoder \(E\) observes \(X_t\) and transmits \(S_{c,t} = \sqrt{2(1 - \beta)} P_S / \alpha_t X_t\). In this phase the relay nodes choose to transmit their own signal to the decoder \(D\) and thus they cannot listen to the signal transmitted from the state encoder due to the half-duplex assumption. Each relay node amplifies the signal that it had received in the previous time step (first transmission phase) and transmits it to the decoder \(D\). The signal transmitted from the \(i\)-th relay node is thus given by \(S_{r,i} = \sqrt{2P_f / (2P_f + N_i)} (S_{c,t-1} + Z_{r,i-1})\). The decoder \(D\) accordingly receives

\[ R_t = h S_{c,t} + \sum_{i=1}^{L} h_i S_{r,i} + Z_{d,t} = L_1 X_t + L_2 X_{t-1} + \tilde{Z}_t \]

where \(L_1 = \sqrt{2(1 - \beta) h^2 P_S / \alpha_t}\), \(L_2 = \sum_{i=1}^{L} \sqrt{2h^2 \beta P_S / (2P_f + N_i)} \alpha_{t-1}\), and \(\tilde{Z}_t = Z_{d,t} + \sum_{i=1}^{L} \sqrt{2h^2 \beta P_S / (2P_f + N_i)} Z_{r,i-1}\) is a white Gaussian noise sequence with zero mean and variance \(N(\beta, \{P_{f,i}\}_{i=1}^{L}) = N_d + \sum_{i=1}^{L} (2h^2 \beta P_f / (2P_f + N_i))\). The decoder then computes the MMSE estimate of \(X_t\) given all previous channel outputs \(\{R_1, R_2, \ldots, R_t\}\) in the following three steps:

1) Compute the MMSE prediction of \(R_t\) from \(\{R_1, R_2, \ldots, R_{t-1}\}\), which is given by \(\hat{R}_t = L_2 \hat{X}_{t-1}\), where \(\hat{X}_{t-1}\) is the MMSE estimate of \(X_{t-1}\).

2) Compute the innovation:

\[ I_t = R_t - \hat{R}_t = L_1 X_t + L_2 (X_{t-1} - \hat{X}_{t-1}) + \tilde{Z}_t \]

\[ = \left( \frac{L_1 + L_2}{\lambda} \right) X_t - \frac{L_2}{\lambda} W_{t-1} + \tilde{Z}_t \]

where \((a)\) follows from \(X_t = \lambda (X_{t-1} - \hat{X}_{t-1}) + W_t\).

3) Compute the MMSE estimate of \(X_t\) given \(\{R_1, R_2, \ldots, R_{t-1}\}\). The state \(X_t\) is independent of \(\{R_1, R_2, \ldots, R_{t-1}\}\) given \(I_t\), therefore, without any loss of optimality we compute \(\hat{X}_t\) based only on \(I_t\) as

\[ \hat{X}_t = \mathbb{E}[X_t | I_t] = \mathbb{E}[X_t I_t] \mathbb{E}[I_t^2]^{-1} I_t \]

\[ = \lambda (L_1 + L_2)^2 \alpha_t + \lambda^2 \tilde{N}(\beta, P_r) \]

where \((a)\) follows from an MMSE estimation of a Gaussian variable; and \((b)\) follows from \(\mathbb{E}[X_t I_t] = (L_1 + L_2) / \lambda \alpha_t\) and \(\mathbb{E}[I_t^2] = ((L_1 + L_2) / \lambda)^2 \alpha_t + (L_2^2 \tilde{N}(\beta, P_r))\). The controller \(C\) takes action \(U_t = -\lambda \hat{X}_t\) which results in \(X_{t+1} = \lambda (X_t - \hat{X}_t) + W_t\). The new state \(X_{t+1}\) is a linear combination of zero mean Gaussian random variables \(\{X_t, \hat{X}_t, W_t\}\), therefore it is also zero mean Gaussian with variance

\[ \alpha_{t+1} = \lambda^2 \mathbb{E}[X_{t+1}^2] = \lambda^2 \left( \frac{N_d}{2h^2 \beta P_S + N_d} \right) \alpha_t + N_w. \]  

where the last equality follows from \(\mathbb{E}[X_t \hat{X}_t] = \mathbb{E}[\hat{X}_t^2] = 2h^2 \beta P_S \alpha_t/(2h^2 \beta P_S + N_d)\) by computation.

First Transmission Phase, \(t = 1, 3, 5, \ldots\): The state encoder \(E\) observes \(X_t\) and transmits \(S_{c,t} = \sqrt{2h^2 \beta P_S / \alpha_t} X_t\). The relay nodes \(\{R_{1,i}\}_{i=1}^{L}\) receive this signal over the Gaussian links and do not transmit any signal in this transmission phase due to half-duplex restriction. The decoder \(D\) observes \(R_t = h S_{c,t} + Z_{d,t}\) and computes the MMSE estimate of \(X_t\) as

\[ \hat{X}_t = \mathbb{E}[X_t | R_{[1,t]}] = \mathbb{E}[X_t | R_t] = \mathbb{E}[X_t R_t | R_t^2] R_t, \]
\[
\alpha_{t+1} = \lambda^2 \mathbb{E} \left[ (X_t - \tilde{X}_t)^2 \right] + \mathbb{E} \left[ W_t^2 \right] \stackrel{(a)}{=} \lambda^2 \alpha_t \left( \frac{L^2 n_w + \lambda^2 \tilde{N}(\beta, P_r)}{(\lambda L_1 + L_2)^2} \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_r) \right) + n_w
\]

\[
= \lambda^2 \left( \lambda^2 k \alpha_{t-1} + n_w \right) \left( \frac{(n_w k_1)^{1/\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_r)}{k^{1/\alpha_{t-1}} + \tilde{N}(\beta, P_r)} + n_w \right) + n_w
\]

\[
\leq \lambda^2 \left( \lambda^2 k \alpha_{t-1} + n_w \right) \left( f_{\infty} + \frac{m}{\alpha_{t-2}} \right) + n_w
\]

\[
= \lambda^4 k^m \alpha_{t-2} + \lambda^2 n_w f_{\infty} + \lambda^4 m k + n_w
\]

\[
\Rightarrow \lambda^4 k^m \alpha_{t-2} + \lambda^2 n_w f_{\infty} + \lambda^4 m k + n_w = g(\alpha_{t-2})
\]

\[
\alpha_{t+1} = \lambda^2 \mathbb{E} \left[ (X_t - \tilde{X}_t)^2 \right] + \mathbb{E} \left[ W_t^2 \right] \stackrel{(a)}{=} \lambda^2 \alpha_t \left( \frac{L^2 n_w + \lambda^2 \tilde{N}(\beta, P_r)}{(\lambda L_1 + L_2)^2} \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_r) \right) + n_w
\]

\[
\Rightarrow \log(\lambda) < \log \left( \frac{1 + \frac{2h^2 \beta P_S}{Nd}}{\tilde{N}(\beta, P_r)} \right)
\]

\[I^4 < \frac{(k^2 + \sqrt{k^2})^2 + \tilde{N}(\beta, P_r)}{\tilde{N}(\beta, P_r)}
\]

\[
\Rightarrow \log(\lambda) < \log \left( \frac{1 + \frac{2h^2 \beta P_S}{Nd}}{\tilde{N}(\beta, P_r)} \right)
\]

C. Remark 5.2 on Information Rate

The given scheme can be seen as a point-to-point communication channel, where \(R_{2t-1}\) is the channel output corresponding to the input \(S_{2t-1}\) and \(I_{2t}\) is the channel output corresponding to the input \(S_{2t}\) for \(t = 1, 2, 3, \ldots\). The information rate is given by

\[
\lim_{T \to \infty} \frac{1}{2T} I \left( \{S_{2t-1}\}_{t=1}^T, \{S_{2t}\}_{t=1}^T; \{R_{2t-1}\}_{t=1}^T, \{I_{2t}\}_{t=1}^T \right)
\]

\[= \lim_{T \to \infty} \frac{1}{2T} \left[ h \left( \{R_{2t-1}\}_{t=1}^T, \{I_{2t}\}_{t=1}^T \right) - h \left( \{R_{2t-1}\}_{t=1}^T, \{I_{2t}\}_{t=1}^T \right| \{S_{2t-1}\}_{t=1}^T, \{S_{2t}\}_{t=1}^T \right) \]

where \(h\) is the channel entropy function.
between the transmitted and the innovation variables is fixed point $x$ converges. There are three possibilities: i)

$$P(I_2, R_{2t-1} \mid S_{2t-1}, S_{2t-1}) = P(I_2 \mid S_{2t}) P(R_{2t-1} \mid S_{2t-1}).$$

Since the channel is memoryless, the random variables are Gaussian and $E[R_{2t-1} R_{2k-1}] = E[I_2 I_2] = 0$ for $k \neq l$, and $E[R_{2t-1} I_{2k}] = 0$ for all $l, k = 1, 2, 3, \ldots$; and (b) follows from the fact that $R_{2t-1}$ and $I_{2t}$ are both sequences of i.i.d. variables. For the first transmission phase, we have

$$I(S_{2t-1}; R_{2t-1}) = h(I_{2t}) - h(U_{2t} \mid S_{2t}) = h(I_{2t}) - h(Z_{2t}).$$

where (a) follows from $R_{2t-1} \sim N(0, 2h^2 \beta P_s + N)$ and $Z_{2t} \sim N(0, N)$. In the second phase, the decoder computes the innovation $I_t$ according to (39). The mutual information between the transmitted and the innovation variables is

$$I(S_{2t}; I_{2t}) = h(I_{2t}) - h(I_{2t} \mid S_{2t}) = h(I_{2t}) - h(Z_{2t})$$

where (a) follows from $I_{2t} \sim N(0, M(\beta, P_r) + \bar{N}(\beta, P_r))$ and $Z_{2t} \sim N(0, \bar{N}(\beta, P_r))$. From (49), (50), and (48) the corresponding information rate is equal to

$$\frac{1}{4} \left( \log \left( 1 + \frac{2h^2 \beta P_s}{N} \right) + \log \left( 1 + \frac{\bar{M}(\beta, P_r)}{\bar{N}(\beta, P_r)} \right) \right).$$

For the given channel, the directed information rate is equal to information rate due to mutual independence of the channel output sequence [42, Theorem 2].

D. Proof of Lemma 6.1

Assume that $T(x)$ is a non-decreasing mapping with a unique fixed point $x^*$. Further assume that there exist $u \leq x \leq v$ such that $T(u) \geq u$ and $T(v) \leq v$. Consider a sequence generated by the following iterations: $x_{t+1} = T(x_t)$ with $x_n \in \mathbb{R}$. We want to show that starting from any $x_0 \in \mathbb{R}$, the sequence $\{x_t\}$ converges. There are three possibilities: i) $x_0 = x^*$, ii) $x_0 > x^*$, and iii) $x_0 < x^*$. For $x_0 \in [x^*, \infty)$ we have $T(x) \leq x$, therefore $x_1 = T(x_0) \leq x_0$. Since $T(x)$ is non-decreasing, $x_2 = T(x_1) \leq T(x_0) = x_1$. Thus for any $t \in \mathbb{N}$ we have $x_{t+1} = T(x_t) \leq T(x_1) = x_1$. Further this sequence is lower bounded by $x^*$ because for any $x_t \in [x^*, \infty)$, $x^* = T(x^*) \leq T(x_t) = x_{t+1}$ due to non-decreasing $T(x)$. Thus the sequence $\{x_t\}$ converges since it is monotonically decreasing and lower bounded by $x^*$ [56, Theorem 3.14]. For $x \in (-\infty, x^*]$ we have $T(x) \geq x$, therefore $x_1 = T(x_0) \geq x_0$. Since $T(x)$ is non-decreasing, we have $x_2 = T(x_1) \geq T(x_0) = x_1$. Thus for any $t \in \mathbb{N}$ we have $x_{t+1} = T(x_t) \geq T(x_1) = x_1$. Further this sequence is upper bounded by $x^*$ because for any $x_t \in (-\infty, x^*)$, we have $x_{t+1} = T(x_t) \leq T(x^*) = x^*$ due to non-decreasing $T(x)$. Since $\{x_t\}$ is strictly increasing and upper bounded by $x^*$ for $x_0 \in [x^*, \infty)$, it converges [56, Theorem 3.14].

REFERENCES


Serdar Yüksel received the B.Sc. degree in electrical and electronics engineering from Bilkent University, Ankara 06800, Turkey, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 2003 and 2006, respectively.

He was a Post-Doctoral Researcher at Yale University, New Haven, CT, for a year before joining Queens University, Kingston, ON, Canada, as an Assistant Professor of mathematics and engineering in the Department of Mathematics and Statistics, where he is now an Associate Professor. He has held visiting positions at the KTH Royal Institute of Technology, Sweden and Bilkent University. He is a co-author of *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints* (Birkhäuser, 2013). His research interests are on control theory, networked control, and information theory.

Dr. Yüksel received the 2013 CAIMS/PIMS Early Career Award in Applied Mathematics.

Mikael Skoglund (S’93–M’97–SM’04) received the Ph.D. degree from Chalmers University of Technology, Göteborg, Sweden, in 1997.

In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Communication Theory Division and he is the Assistant Dean for Electrical Engineering. He is also a founding faculty member of the ACCESS Linnaeus Center and director for the Center Graduate School. He has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory and statistical signal processing. He has authored and co-authored more than 120 journal and 275 conference papers, and he holds six patents.

Dr. Skoglund has served on numerous technical program committees for IEEE sponsored conferences. From 2003 to 2008 he was an Associate Editor with the IEEE TRANSACTIONS ON COMMUNICATIONS and from 2008 to 2012 he was on the editorial board for the IEEE TRANSACTIONS ON INFORMATION THEORY.