

# 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 (in 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

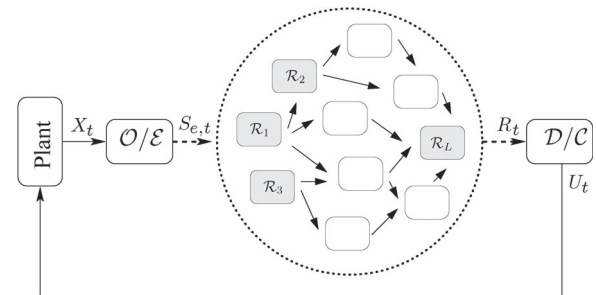


Fig. 1. Unstable plant has to be controlled over a Gaussian relay network.

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

$$X_{t+1} = AX_t + BU_t + W_t \quad (1)$$

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  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  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<sup>1</sup> 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  $\mathcal{E}$  is lumped with the observer  $\mathcal{O}$  and a decoder  $\mathcal{D}$  is lumped with the controller  $\mathcal{C}$ . In addition, there are  $L$  relay nodes  $\{\mathcal{R}_i\}_{i=1}^L$  within the channel to support communication from  $\mathcal{E}$  to  $\mathcal{D}$ . At any time instant  $t$ ,

<sup>1</sup>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.

Manuscript received June 9, 2012; revised February 14, 2013 and September 29, 2013; accepted April 6, 2014. Date of publication May 6, 2014; date of current version August 20, 2014. This paper was presented in part at IEEE ICCA, 2010 and ACC, 2011. This work was supported in part by the Swedish Research Council. Recommended by Associate Editor G. N. Nair.

A. A. Zaidi, T. J. Oechtering, and M. Skoglund are with the School of Electrical Engineering and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm 100 44, Sweden (e-mail: zaidi@kth.se; oech@kth.se; skoglund@kth.se).

S. Yüksel is with the Department of Mathematics and Statistics at Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: yuksel@mast.queensu.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2014.2322213

$S_{e,t}$  and  $R_t$  are the input and the output of the network and  $U_t$  is the control action. Let  $f_t$  denote the observer/encoder policy such that  $S_{e,t} = f_t(X_{[0,t]}, U_{[0,t-1]})$ , where  $X_{[0,t]} := \{X_0, X_1, \dots, X_t\}$  and we have the following average power constraint:  $\lim_{T \rightarrow \infty} (1/T) \sum_{t=0}^{T-1} \mathbb{E}[S_{e,t}^2] \leq P_S$ . Let  $\pi_t$  denote the decoder/controller policy, then  $U_t = \pi_t(R_{[0,t]})$ . 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  $\mathbb{E}[\|X_t\|^2] < M$  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  $m$ -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<sup>2</sup> 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

$$\log(|\det(A)|) \leq \liminf_{T \rightarrow \infty} \frac{1}{T} I(\bar{X}_{[0,T-1]} \rightarrow R_{[0,T-1]}) \quad (2)$$

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

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

is the directed information from the uncontrolled state process  $\bar{X}_{[0,T-1]}$  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].  $\square$

## 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

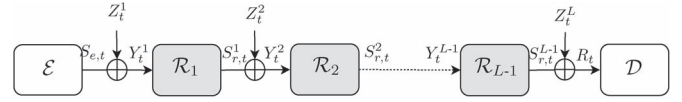


Fig. 2. A cascade Gaussian network model.

state encoder  $\mathcal{E}$  observes the state of the system and transmits its signal to the relay node  $\mathcal{R}_1$ . The relay node  $\mathcal{R}_1$  transmits a signal to the relay node  $\mathcal{R}_2$  and so on. Finally the state information is received at the remote decoder/controller  $\mathcal{D}$  from  $\mathcal{R}_{L-1}$ . 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  $\mathcal{R}_i$  transmits signal to  $\mathcal{R}_{i+1}$ , 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  $\mathcal{E}$  and  $S_{r,t}^i$  is the signal transmitted from  $\mathcal{R}_i$ , which are given by

$$\begin{aligned} S_{e,t} &= f_t(X_{[0,t]}, U_{[0,t-1]}) \quad \forall t : t = 1 + nL, n \in \mathbb{N}, \\ S_{r,t}^i &= g_t^i(Y_{[0,t]}^i) \quad \forall t : t = 1 + i + nL, n \in \mathbb{N}, \\ S_{e,t} &= S_{r,t}^i = 0 \quad \text{otherwise} \end{aligned} \quad (3)$$

where  $\mathbb{N} = \{0, 1, 2, \dots\}$ ,  $f_t : \mathbb{R}^{2t-1} \rightarrow \mathbb{R}$ ,  $g_t^i : \mathbb{R}^t \rightarrow \mathbb{R}$  such that  $\mathbb{E}[f_t^2(X_{[0,t]}, U_{[0,t-1]})] = LP_S$ ,  $\mathbb{E}[(g_t^i(Y_{[0,t]}^i))^2] = LP_r^i$ ,  $\sum_{i=1}^{L-1} P_r^i \leq P_R$ . The signal received by  $\mathcal{R}_i$  is

$$\begin{aligned} Y_t^1 &= S_{e,t} + Z_t^1, Y_t^i = S_{r,t}^{i-1} + Z_t^i \quad \forall t : t = nL + i, n \in \mathbb{N}, \\ Y_t^i &= 0 \quad \text{otherwise.} \end{aligned} \quad (4)$$

Here  $Z_t^i \sim \mathcal{N}(0, N_i)$  denotes mutually independent white Gaussian noise components. Accordingly  $\mathcal{D}$  receives  $R_t = S_{r,t}^{L-1} + Z_t^L$  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

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

*Proof 3.1:* We first derive an outer bound on the directed information  $I(\bar{X}_{[1,LT]} \rightarrow R_{[1,LT]})$  over the given channel and then use Theorem 2.1 to find the necessary condition (5)

$$\begin{aligned} & I(\bar{X}_{[1,LT]} \rightarrow R_{[1,LT]}) \\ & \stackrel{(a)}{\leq} I(\bar{X}_{[1,LT]}; R_{[1,LT]}) \\ & \stackrel{(b)}{\leq} I(\bar{X}_{[1,LT]}; Y_{[1,LT]}^i, R_{[1,LT]}) \\ & = \sum_{t=1}^{LT} I(\bar{X}_{[1,LT]}; R_t, Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i) \\ & \stackrel{(c)}{\leq} \sum_{t=1}^{LT} \left( h(R_t, Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i) \right. \\ & \quad \left. - h(R_t, Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i, \bar{X}_{[1,LT]}) \right) \end{aligned}$$

<sup>2</sup>The definition of information capacity for Gaussian channels can be found on page 263 in [35].

$$\begin{aligned}
&\stackrel{(d)}{=} \sum_{t=1}^{LT} \left( h \left( Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i \right) + h \left( R_t | R_{[1,t-1]}, Y_{[1,t]}^i \right) \right. \\
&\quad - h \left( Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i, \bar{X}_{[1,LT]} \right) \\
&\quad \left. - h \left( R_t | R_{[1,t-1]}, Y_{[1,t]}^i, \bar{X}_{[1,LT]} \right) \right) \\
&\stackrel{(e)}{=} \sum_{t=1}^{LT} \left( h \left( Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i \right) \right. \\
&\quad - h \left( Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i, \bar{X}_{[1,LT]} \right) \\
&\quad \left. + \underbrace{I(R_t; \bar{X}_{[1,LT]} | R_{[1,t-1]}, Y_{[1,t]}^i)}_{=0} \right) \\
&\stackrel{(f)}{\leq} \sum_{t=1}^{LT} \left( h(Y_t^i) - h \left( Y_t^i | R_{[1,t-1]}, Y_{[1,t-1]}^i, \bar{X}_{[1,LT]} \right) \right) \\
&\stackrel{(g)}{\leq} \sum_{t=1}^{LT} \left( h(Y_t^i) - h \left( Y_t^i | S_{r,t}^{i-1}, R_{[1,t-1]}, Y_{[1,t-1]}^i, \bar{X}_{[1,LT]} \right) \right) \\
&\stackrel{(h)}{=} \sum_{t=1}^{LT} I(S_{r,t}^{i-1}; Y_t^i) \stackrel{(i)}{=} \sum_{t=0}^{T-1} I(S_{r,tL+i}^{i-1}; Y_{tL+i}^i) \\
&\stackrel{(j)}{\leq} \frac{1}{2} \sum_{t=0}^{T-1} \log \left( 1 + \frac{LP_r^{i-1}}{N_i} \right) = \frac{T}{2} \log \left( 1 + \frac{LP_r^{i-1}}{N_i} \right) \quad (6)
\end{aligned}$$

where (a) follows from [42, Theorem 1]; (b) follows from the fact that adding side information cannot decrease mutual information; (c), (d) and (e) follow from properties of mutual information and differential entropy; (f) follows from conditioning reduces entropy and the following Markov chain  $\bar{X}_{[1,LT]} - (Y_{[1,t]}^i, R_{[1,t-1]}) - R_t$ ; (g) follow from conditioning reduces entropy; (h) follows from the Markov chain  $Y_t^i - S_{r,t}^{i-1} - (R_{[1,t-1]}, Y_{[1,t-1]}^i, \bar{X}_{[1,LT]})$  due to memoryless channel from  $S_{r,t}^{i-1}$  to  $Y_t^i$ ; (i) follows from (3) and (4); and (j) follows from the fact that mutual information of a Gaussian channel is maximized by the Gaussian input distribution [35, Theorem 8.6.5]. If we replace  $Y_{[1,LT]}^i$  with  $Y_{[1,LT]}^1$  in step (b) of (6) and  $S_{r,t}^{i-1}$  with  $S_{e,t}$  in step (g) of (6), then we get:

$$I(\bar{X}_{[1,LT]} \rightarrow R_{[1,LT]}) \leq \frac{T}{2} \log \left( 1 + \frac{LP_S}{N_1} \right). \quad (7)$$

The term  $I(\bar{X}_{[1,LT]} \rightarrow R_{[1,LT]})$  can also be bounded as

$$\begin{aligned}
&I(\bar{X}_{[1,LT]} \rightarrow R_{[1,LT]}) \\
&= \sum_{t=1}^{LT} I(\bar{X}_{[1,t]}; R_t | R_{[1,t-1]}) \\
&\stackrel{(a)}{\leq} \sum_{t=1}^{LT} I(S_{r,[1,t]}^{L-1}; R_t | R_{[1,t-1]}) = I(S_{r,[1,LT]}^{L-1} \rightarrow R_{[1,LT]}) \\
&\stackrel{(b)}{\leq} \sum_{t=1}^{LT} I(S_{r,t}^{L-1}; R_t) \stackrel{(c)}{=} \sum_{t=0}^{T-1} I(S_{r,tL+L}^{L-1}; R_{tL+L}) \\
&\stackrel{(d)}{\leq} \frac{T}{2} \log \left( 1 + \frac{LP_r^{L-1}}{N_L} \right) \quad (8)
\end{aligned}$$

where (a) follows from the Markov chain  $\bar{X}_{[1,LT]} - (S_{r,[1,t]}^{L-1}, R_{[1,t-1]}) - R_{[1,t]}$ ; (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:

$$\begin{aligned}
&I(\bar{X}_{[1,LT]} \rightarrow R_{[1,LT]}) \\
&\leq \frac{T}{2} \min \left\{ \log \left( 1 + \frac{LP_S}{N_1} \right), \right. \\
&\quad \left. \log \left( 1 + \frac{LP_r^1}{N_2} \right), \dots, \log \left( 1 + \frac{LP_r^{L-1}}{N_L} \right) \right\} \\
&\stackrel{(a)}{=} \frac{T}{2} \log \left( 1 + L \min \left\{ \frac{P_S}{N_1}, \frac{P_r^1}{N_2}, \dots, \frac{P_r^{L-1}}{N_L} \right\} \right) \\
&\leq \frac{T}{2} \log \left( 1 + L \min \left\{ \frac{P_S}{N_1}, \max_i \min \left\{ \frac{P_r^1}{N_2}, \dots, \frac{P_r^{L-1}}{N_L} \right\} \right\} \right) \\
&\stackrel{(b)}{=} \frac{T}{2} \log \left( 1 + L \min \left\{ \frac{P_S}{N_1}, \frac{P_R}{\sum_{i=2}^L N_i} \right\} \right) \quad (9)
\end{aligned}$$

(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 let  $T \rightarrow \infty$  according to Theorem 2.1, we get the necessary condition (5).  $\square$

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

$$\log(|\lambda|) < \max_{P_r^i} \frac{1}{2L} \log \left( 1 + \frac{LP_S}{LP_S + N_1} \prod_{i=1}^{L-1} \left( \frac{LP_r^i}{LP_r^i + N_{i+1}} \right) \right) \quad (10)$$

where the optimal power allocation is given by  $P_r^i = (-N_{i+1} + \sqrt{N_{i+1}^2 - (4N_{i+1}/\gamma)})/2$  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.

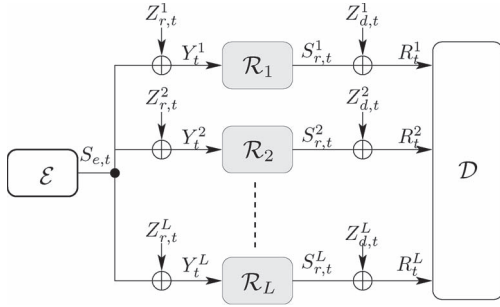


Fig. 3. Parallel relay network.

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  $\{\mathcal{R}_i\}_{i=1}^L$  to  $\mathcal{D}$ . We call this setup a *parallel network*, modeling a scenario where the signal spaces of the relay nodes are mutually orthogonal. For example the signals may be transmitted in either disjoint frequency bands or in disjoint time slots. In the first transmission phase, the sensor transmits  $S_{e,t}$  with an average power  $\mathbb{E}[S_{e,t}^2] = 2P_S$  to the relays and in the second phase all relays simultaneously transmit to the remote controller with average powers  $2P_r^i$  such that  $\sum_{i=1}^L P_r^i \leq P_R$ . Accordingly, the received signals are given by

$$\begin{aligned} Y_t^i &= S_{e,t} + Z_{r,t}^i, & R_t^i &= S_{r,t}^i = 0, & t &= 1, 3, 5, \dots \\ R_t^i &= S_{r,t}^i + Z_{d,t}^i, & Y_t^i &= S_{e,t} = 0, & t &= 2, 4, 6, \dots \end{aligned} \quad (11)$$

where  $Z_{r,t}^i \sim \mathcal{N}(0, N_r^i)$ ,  $Z_{d,t}^i \sim \mathcal{N}(0, N_d^i)$  denote mutually independent white Gaussian noise variables. In the following we present conditions for mean square stability of the system in (1) over the given parallel network.

**Theorem 4.1:** If the system (1) is mean square stable over the *parallel network*, then

$$\begin{aligned} &\log(|\det(A)|) \\ &\leq \frac{1}{4} \min \left\{ \log \left( 1 + 2 \sum_{i=1}^L \frac{P_S}{N_r^i} \right), \sum_{i=1}^L \log \left( 1 + \frac{2P_r^i}{N_d^i} \right) \right\} \end{aligned} \quad (12)$$

where  $P_r^i = \max\{\gamma - N_d^i, 0\}$  and  $\gamma$  is chosen such that  $\sum_{i=1}^L P_r^i = P_R$ .

**Proof 4.1:** Following the same steps as in proof of Theorem 3.1, we can bound  $I(\bar{X}_{[1,2T]} \rightarrow R_{[1,2T]})$  as,

$$\begin{aligned} &I(\bar{X}_{[1,2T]} \rightarrow \{R_{[1,2T]}^i\}_{i=1}^L) \\ &\stackrel{(a)}{\leq} \min \left\{ \sum_{t=1}^{2T} I(S_{e,t}; \{Y_t^i\}_{i=1}^L), \right. \\ &\quad \left. \sum_{t=1}^{2T} I(\{S_{r,t}^i\}_{i=1}^L; \{R_t^i\}_{i=1}^L) \right\} \end{aligned}$$

$$\begin{aligned} &\stackrel{(b)}{=} \min \left\{ \sum_{t=1}^T I(S_{e,2t-1}; \{Y_{2t-1}^i\}_{i=1}^L), \right. \\ &\quad \left. \sum_{t=1}^T I(\{S_{r,2t}^i\}_{i=1}^L; \{R_{2t}^i\}_{i=1}^L) \right\} \\ &\stackrel{(c)}{\leq} \frac{T}{2} \min \left\{ \log \left( 1 + 2 \sum_{i=1}^L \frac{P_S}{N_r^i} \right), \right. \\ &\quad \left. \max_{P_r^i} \sum_{i=1}^L \log \left( 1 + \frac{2P_r^i}{N_d^i} \right) \right\} \end{aligned} \quad (13)$$

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.  $\square$

We now state a sufficient condition for mean square stability under linear policies.

**Theorem 4.2:** The scalar linear time invariant system in (1) with  $A = \lambda$  can be mean square stabilized using a linear scheme over the Gaussian *parallel network* if

$$\log(|\lambda|) < \frac{1}{4} \log \left( 1 + \sum_{i=1}^L \frac{4P_S P_r^i}{2P_S N_d + 2P_r^i N_r^i + N_d N_r^i} \right). \quad (14)$$

**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.  $\square$

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

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

$$\begin{aligned} &\Gamma_{\text{sec}} - \Gamma_{\text{suf}} \\ &= \frac{1}{4} \log \left( 1 + \frac{2P_S N_d (2P_S + N_r)}{4P_S P_r N_r + \frac{N_r (2P_S N_d + 2P_r N_r + N_d N_r)}{L}} \right) \end{aligned} \quad (15)$$

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

**Remark 4.1:** If  $N_d^i = 0$ , then  $\Gamma_{\text{sec}} - \Gamma_{\text{suf}} = 0$  and the linear scheme is exactly optimal. For  $N_r^i = 0$ ,  $\Gamma_{\text{suf}} := (1/4) \log(1 + (2LP_r / N_d))$  and  $\Gamma_{\text{sec}} := (L/4) \log(1 + (2P_r / N_d))$  according to (12). Clearly  $\lim_{L \rightarrow \infty} (\Gamma_{\text{sec}} - \Gamma_{\text{suf}}) = \infty$ , 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 and thus the scheme fulfills the first condition of optimality. This nonlinear sensing scheme is shown to outperform the best 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  $\mathcal{E} - \mathcal{R}_i$  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  $Z_{r,t}^i = 0$  for all  $i$ , only if

$$\log(|\det(A)|) \leq \frac{1}{4} \max_{P_r^i: \sum_i P_r^i \leq P_R} \sum_{i=1}^L \log \left( 1 + \frac{2P_r^i}{N_d^i} \right). \quad (16)$$

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.  $\square$

*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 information capacity required 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  $\{\mathcal{R}_i\}_{i=1}^L$  is illustrated in Fig. 4. The variables  $S_{e,t}$

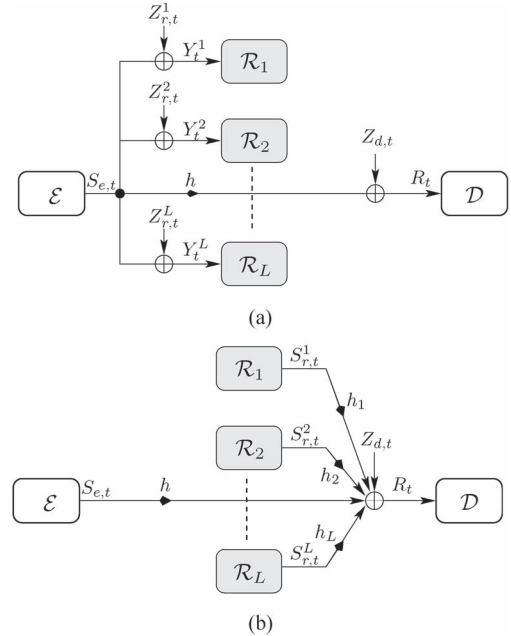


Fig. 4. A non-orthogonal half-duplex Gaussian network model. (a) First transmission phase. (b) Second transmission phase.

and  $S_{r,t}^i$  denote the transmitted signals from the state encoder  $\mathcal{E}$  and relay  $\mathcal{R}_i$  at any discrete time step  $t$ . The variables  $Z_{r,t}^i$  and  $Z_{d,t}$  denote the mutually independent white Gaussian noise components at the relay node  $i$  and  $\mathcal{D}$  of the remote control unit, with  $Z_{r,t}^i \sim \mathcal{N}(0, N_r^i)$  and  $Z_{d,t} \sim \mathcal{N}(0, N_d)$ . The noise components  $\{Z_{r,t}^i\}_{i=1}^L$  are independent across the relays, i.e.,  $\mathbb{E}[Z_{r,t}^k Z_{r,t}^i] = 0$  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  $\mathcal{E}$  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  $\mathcal{E}$  and relay nodes  $\{\mathcal{R}_i\}_{i=1}^L$  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  $2P_r^i$  such that  $\sum_{i=1}^L P_r^i \leq P_R$ . The input and output of the  $i$ -th relay are given by,

$$\begin{aligned} Y_t^i &= S_{e,t} + Z_{r,t}^i, & S_{r,t}^i &= 0, & t &= 1, 3, 5, \dots \\ Y_t^i &= 0, & S_{r,t}^i &= g_t^i(Y_{[0,t-1]}^i), & t &= 2, 4, 6 \end{aligned} \quad (17)$$

where  $g_t^i: \mathbb{R}^{t+1} \rightarrow \mathbb{R}$  is the  $i$ -th relay encoding policy such that  $\mathbb{E}[(g_t^i(Y_{[0,t-1]}^i))^2] = 2P_r^i$  and  $\sum_{i=1}^L P_r^i \leq P_R$ . 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  $\mathcal{E} - \mathcal{D}$  and  $\mathcal{R}_i - \mathcal{D}$  links respectively.

**Theorem 5.1:** If the linear system in (1) is mean-square stable over the *non-orthogonal half-duplex relay network*, then

$$\begin{aligned} \log(|\det(A)|) &\leq \frac{1}{4} \min \left\{ \max_{0 < \beta \leq 1} \left( \log \left( 1 + \frac{2h^2(1-\beta)P_S}{N_d} \right) \right. \right. \\ &+ \log \left( 1 + 2\beta P_S \left( \sum_{i=1}^L \frac{1}{N_r^i} + \frac{h^2}{N_d} \right) \right) \left. \right\} \\ &+ \max_{P_r^i: \sum_i P_r^i \leq P_R} \left( \log \left( 1 + \frac{2h^2\beta P_S}{N_d} \right) \right. \\ &+ \left. \left. \log \left( 1 + \frac{1}{N_d} \left( \sum_{i=1}^{L+1} \delta_i^2 P_i + 2 \sum_{i=1}^{L+1} \sum_{k=i+1}^{L+1} \rho_{i,k}^* \delta_i \delta_k \sqrt{P_i P_k} \right) \right) \right) \right\} \end{aligned} \quad (18)$$

where  $\rho_{i,k}^* = 2(1-\beta)P_S/\sqrt{(2(1-\beta)P_S+N_i)(2(1-\beta)P_S+N_k)}$ ,  $P_{L+1} := 2(1-\beta)P_S$ ,  $N_{L+1} := 0$ ,  $\delta_{L+1} := h$ ,  $P_i := 2P_r^i$ ,  $\delta_i := h_i$ ,  $N_i := N_r^i$  for all  $i = \{1, 2, \dots, L\}$ .

*Proof 5.1:* We first derive an outer bound on the directed information  $I(\bar{X}_{[1,L,T]} \rightarrow R_{[1,L,T]})$  over the given channel and then use Theorem 2.1 to find the necessary condition (18)

$$\begin{aligned} I(\bar{X}_{[1,2T]} \rightarrow R_{[1,2T]}) &\stackrel{(a)}{=} I(\bar{X}_{[1,2T]}; R_{[1,2T]}) \\ &\stackrel{(b)}{\leq} I(\bar{X}_{[1,2T]}; \{Y_{[1,2T]}^i\}_{i=1}^L, R_{[1,2T]}) \\ &\stackrel{(c)}{\leq} I(\bar{X}_{[1,2T]}; \tilde{R}_{[1,2T]}, \{Y_{[1,2T]}^i\}_{i=1}^L) \\ &= \sum_{t=1}^{2T} I(\bar{X}_{[1,2T]}; \tilde{R}_t, \{Y_t^i\}_{i=1}^L | \tilde{R}_{[1,t-1]}, \{Y_{[1,t-1]}^i\}_{i=1}^L) \\ &\stackrel{(d)}{\leq} \sum_{t=1}^{2T} I(S_{e,t}; \tilde{R}_t, \{Y_t^i\}_{i=1}^L | \tilde{R}_{[1,t-1]}, \{Y_{[1,t-1]}^i\}_{i=1}^L) \\ &\stackrel{(e)}{\leq} \sum_{t=1}^{2T} I(S_{e,t}; \tilde{R}_t, \{Y_t^i\}_{i=1}^L) \\ &\stackrel{(f)}{=} \sum_{t=1}^T I(S_{e,2t}; \tilde{R}_{2t}) \\ &\quad + \sum_{t=1}^T I(S_{e,2t-1}; \tilde{R}_{2t-1}, \{Y_{2t-1}^i\}_{i=1}^L) \\ &\stackrel{(g)}{\leq} \frac{T}{2} \log \left( 1 + \frac{2h^2(1-\beta)P_S}{N_d} \right) \\ &\quad + \frac{T}{2} \log \left( 1 + 2\beta P_S \left( \sum_{i=1}^L \frac{1}{N_r^i} + \frac{h^2}{N_d} \right) \right) \\ &\leq \frac{T}{2} \max_{0 < \beta \leq 1} \left\{ \log \left( 1 + \frac{2h^2(1-\beta)P_S}{N_d} \right) \right. \\ &\quad \left. + \log \left( 1 + 2\beta P_S \left( \sum_{i=1}^L \frac{1}{N_r^i} + \frac{h^2}{N_d} \right) \right) \right\} \end{aligned} \quad (19)$$

where (a) follows from [42, Theorem 1]; (b) follows from the fact that adding side information cannot decrease mutual information; (c) follows by defining  $\tilde{R}_t := R_t - \sum_{i=1}^L h_i S_{r,t}^i$  and from the fact that  $S_{r,t}^i$  is a function of  $Y_{[1,t-1]}^i$ ; (d) follows from the Markov chain  $\bar{X}_{[1,2T]} - S_{e,t} - (\tilde{R}_t, \{Y_t^i\}_{i=1}^L)$ , since  $\bar{X}_{[0,T]}$  is the uncontrolled state process and the fact

that the channel between  $S_{e,[1,2T]}$  and  $(\tilde{R}_{[1,2T]}, \{Y_{[1,2T]}^i\}_{i=1}^L)$  is memoryless due to  $\tilde{R}_t = R_t - \sum_{i=1}^L h_i S_{r,t}^i$ ; (e) follows from the Markov chain  $(\tilde{R}_{[1,t-1]}, \{Y_{[1,t-1]}^i\}_{i=1}^L) - S_{e,t} - (\tilde{R}_t, \{Y_t^i\}_{i=1}^L)$  and conditioning reduces entropy; (f) follows by separating odd and even indexed terms and  $Y_{2t}^i = 0$  according to (17); (g) follows from  $Y_{2t-1}^i = S_{e,2t-1} + Z_{r,2t-1}^i$ ,  $\tilde{R}_t = S_{e,t} + Z_t$ ,  $\mathbb{E}[S_{e,2t}^2] = 2(1-\beta)P_S$ ,  $\mathbb{E}[S_{e,2t-1}^2] = 2\beta P_S$ , and the fact that mutual information of a Gaussian channel is maximized by centered Gaussian input distribution [45]. The term  $I(\bar{X}_{[1,2T]} \rightarrow R_{[1,2T]})$  can also be bounded as

$$\begin{aligned} I(\bar{X}_{[1,2T]} \rightarrow R_{[1,2T]}) &= \sum_{t=1}^{2T} I(\bar{X}_{[1,t]}; R_t | R_{[1,t-1]}) \\ &\stackrel{(a)}{\leq} \sum_{t=1}^{2T} I(S_{e,t}, \{S_{r,t}^i\}_{i=1}^L; R_t | R_{[1,t-1]}) \\ &\stackrel{(b)}{\leq} \sum_{t=1}^{2T} I(S_{e,t}, \{S_{r,t}^i\}_{i=1}^L; R_t) \\ &\stackrel{(c)}{=} \sum_{t=1}^T \left( I(S_{e,2t-1}; R_{2t-1}) + I(S_{e,2t}, \{S_{r,2t}^i\}_{i=1}^L; R_{2t}) \right) \\ &\stackrel{(d)}{\leq} \frac{T}{2} \log \left( 1 + \frac{2h^2\beta P_S}{N_d} \right) \\ &\quad + \frac{T}{2} \log \left( 1 + \frac{1}{N_d} \left( \sum_{i=1}^{L+1} \delta_i^2 P_i + 2 \sum_{i=1}^{L+1} \sum_{k=i+1}^{L+1} \rho_{i,k}^* \delta_i \delta_k \sqrt{P_i P_k} \right) \right) \\ &\leq \frac{T}{2} \max_{P_r^i: \sum_i P_r^i \leq P_R} \left\{ \log \left( 1 + \frac{2h^2\beta P_S}{N_d} \right) \right. \\ &\quad \left. + \log \left( 1 + \frac{1}{N_d} \left( \sum_{i=1}^{L+1} \delta_i^2 P_i + 2 \sum_{i=1}^{L+1} \sum_{k=i+1}^{L+1} \rho_{i,k}^* \delta_i \delta_k \sqrt{P_i P_k} \right) \right) \right\} \end{aligned} \quad (20)$$

where  $\rho_{i,k}^* = 2(1-\beta)P_S/\sqrt{(2(1-\beta)P_S+N_i)(2(1-\beta)P_S+N_k)}$ ,  $P_{L+1} = 2(1-\beta)P_S$ ,  $N_{L+1} = 0$ ,  $\delta_{L+1} = h$ ,  $P_i = 2P_r^i$ ,  $\delta_i = h_i$ ,  $N_i = N_r^i$  for all  $i = \{1, 2, \dots, L\}$ . The inequality (a) follows from the Markov chain  $\bar{X}_{[0,t]} - (S_{e,t}, \{S_{r,t}^i\}_{i=1}^L) - R_t$  due to the memoryless channel between  $S_{e,[1,2T]}, \{S_{r,[1,2T]}^i\}_{i=1}^L$  and  $R_{[1,2T]}$ ; (b) follows from the Markov chain  $R_{[1,t-1]} - (S_{e,t}, \{S_{r,t}^i\}_{i=1}^L) - R_t$  and conditioning reduces entropy; (c) follows by separating the odd and even indexed terms and  $S_{r,2t-1}^i = 0$  according to (17); (d) follows from the fact that the first addend on the RHS of (c) is maximized by a centered Gaussian distributed  $S_{e,t}$  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  $\mathcal{E}$  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]:  $\sigma_S^2 := \alpha_t$ ,  $\delta_i := h_i$ ,  $M := L+1$ ,  $P_i := 2P_r^i$ ,  $\sigma_Z^2 := N_d$ ,  $\sigma_{W,i}^2 := N_r^i$ ,  $\alpha_i = \sqrt{2(1-\beta)P_S}/\alpha_t$ . We finally obtain (18) by dividing (19) and (20) by  $2T$  and let  $T \rightarrow \infty$  according to Theorem 2.1.  $\square$

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

$$\log(|\lambda|) < \frac{1}{4} \max_{\substack{0 < \beta \leq 1 \\ P_r^i: \sum_i P_r^i \leq P_R}} \left\{ \log \left( 1 + \frac{2h^2 \beta P_S}{N_d} \right) + \log \left( 1 + \frac{\tilde{M}(\beta, \{P_r^i\}_{i=1}^L)}{\tilde{N}(\beta, \{P_r^i\}_{i=1}^L)} \right) \right\}, \quad (21)$$

where  $\tilde{M}(\beta, \{P_r^i\}_{i=1}^L) = \frac{(\sqrt{2h^2(1-\beta)P_S} + \sqrt{2\beta P_S N_d / (2h^2 \beta P_S + N_d)}) (\sum_{i=1}^L \sqrt{2h_i^2 P_r^i / (2\beta P_S + N_r^i)})^2}{\sum_{i=1}^L (2h_i^2 P_r^i N_r^i / (2\beta P_S + N_r^i)) + N_d}$  and  $\tilde{N}(\beta, \{P_r^i\}_{i=1}^L) = \sum_{i=1}^L (2h_i^2 P_r^i N_r^i / (2\beta P_S + N_r^i)) + N_d$ .

*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 using the Lagrangian method:

$$P_r^i = P_R \left( \frac{h_i^2 (2P_S + N_r^i)}{(2P_S N_d + N_r^i N_d + P_R h_i^2 N_r^i)^2} \right) \times \left[ \sum_{l=1}^L \frac{h_l^2 (2P_S + N_r^l)}{(2P_S N_d + N_r^l N_d + P_R h_l^2 N_r^l)^2} \right]^{-1}. \quad (22)$$

**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,

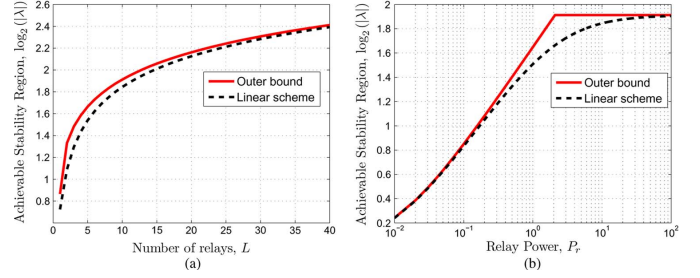


Fig. 5. Comparison of necessary and sufficient conditions for a symmetric two-hop relay network. (a)  $P_S = 2P_r^i = 10$ ,  $N_r^i = N_d = 1$ ,  $h_i = 1$ ;  $L = 10$ ,  $P_S = 10$ ,  $N_r^i = N_d = 1$ ,  $h_i = 1$ .

which then communicate the state information to the controller in the second hop. The controller takes action in alternate time steps upon receiving the state information. We can obtain a sufficient condition for stability over this network by substituting  $h = 0$ ,  $\beta = 1$  in Theorem 5.2. Similarly, a necessary condition can be obtained from (18), where  $\beta = 1$  is the maximizer of the first term and  $\beta = 0$  is the maximizer of the second term. In the following we evaluate the gap between the sufficient and necessary conditions for a symmetric two hop network.

**Proposition 5.1:** For a symmetric two-hop network with  $P_r^i = P_r$ ,  $N_r^i = N_r$ ,  $h_i = c$ ,  $h = 0$ ,  $\beta = 1$ , the gap between necessary and sufficient conditions approaches zero as the number of relays  $L$  goes to infinity. The gap also monotonically approaches zero as  $P_r$  goes to infinity.

*Proof 5.3:* For  $P_r^i = P_r$ ,  $N_r^i = N_r$ ,  $h_i = c$ ,  $h = 0$ ,  $\beta = 1$  for all  $i$ , the RHS of (21) is evaluated as  $\Gamma_{\text{suf}} := (1/4) \log(1 + (4L^2 c^2 P_S P_r / (2Lc^2 P_r N_r + N_d(2P_S + N_r))))$  and the RHS of (18) can be bounded as  $\Gamma_{\text{sec}} := (1/4) \log(1 + (2LP_S / N_r))$ . The gap between  $\Gamma_{\text{suf}}$  and  $\Gamma_{\text{sec}}$  is given by

$$\Gamma_{\text{sec}} - \Gamma_{\text{suf}} = \frac{1}{4} \log \left( 1 + \frac{4P_S^2 N_d + 2P_S N_r N_d}{4c^2 P_S P_r N_r + \frac{L}{2c^2 P_r N_r^2} + \frac{N_d N_r (2P_S + N_r)}{L^2}} \right) \quad (23)$$

which approaches zero as  $L \rightarrow \infty$ . The gap also monotonically approaches zero as  $P_r \rightarrow \infty$ .  $\square$

In Fig. 5(b) we have plotted  $\Gamma_{\text{sec}}$  and  $\Gamma_{\text{suf}}$  as functions of  $L$  and  $P_r$ . These figures show that linear schemes are quite efficient in some regimes.

**Remark 5.3:** Linear policies can even be exactly optimal in the following special cases: i) If we fix all relaying policies to be linear, then the channel becomes equivalent to a point-point scalar Gaussian channel, for which linear sensing is known to be optimal for LQG control [1]. ii) If we fix the state encoder to be linear and assume noiseless causal feedback links from the controller to the relay nodes, then linear policies are optimal for mean-square stabilization over a symmetric *two-hop* relay network, by the following arguments. Since the control actions are available at the relay nodes via noiseless feedback links, there is no dual effect of control, i.e., the separation of estimation and control holds, see e.g., [30]. 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  $t \in \mathbb{N}$

$$\begin{aligned} S_{e,t} &= f_t(X_{[0,t]}, U_{[0,t-1]}), \quad S_{r,t}^i = g_t^i(Y_{[0,t-1]}^i), \\ Y_t^i &= S_{e,t} + Z_{r,t}^i, \quad R_t = hS_{e,t} + \sum_{i=1}^L S_{r,t}^i + Z_{d,t}, \end{aligned} \quad (24)$$

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

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

$$\begin{aligned} \log(|\det(A)|) &\leq \frac{1}{2} \min \left\{ \log \left( 1 + P_S \left( \sum_{i=1}^L \frac{1}{N_r^i} + \frac{h^2}{N_d} \right) \right), \right. \\ &\quad \max_{P_r^i} \left( \log \left( 1 + \frac{1}{N_d} \left( \sum_{i=1}^{L+1} \delta_i^2 P_i \right. \right. \right. \\ &\quad \left. \left. \left. + 2 \sum_{i=1}^{L+1} \sum_{k=i+1}^{L+1} \rho_{i,k}^* \delta_i \delta_k \sqrt{P_i P_k} \right) \right) \right) \right\}, \end{aligned} \quad (25)$$

where  $\rho_{i,k}^* = P_S / \sqrt{(P_S + N_i)(P_S + N_k)}$ ,  $P_{L+1} := P_S$ ,  $N_{L+1} := 0$ ,  $\delta_{L+1} := h$ ,  $P_i := P_r^i$ ,  $\delta_i := h_i$ ,  $N_i := N_r^i$  for all  $i = \{1, 2, \dots, L\}$ .

*Proof 5.4:* The proof follows exactly in the steps of the proof of Theorem 5.1, with an exception that odd and even indexed terms are not treated separately because  $\mathbb{E}[S_{e,t}^2] = P_S$  and  $\mathbb{E}[(S_{r,t}^i)^2] = P_r^i$  for all  $t$ .  $\square$

*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

$$\begin{aligned} \log(|\lambda|) &< \frac{1}{2} \max_{P_r^i: \sum_{i=1}^L P_r^i \leq P_R} \left\{ \log \left( 1 + \right. \right. \\ &\quad \left. \left. \left( \sqrt{h^2 P_S + \eta^*} \sum_{i=1}^L \sqrt{\frac{h_i^2 P_S P_r^i}{P_S + N_r^i}} \right)^2 \left( N_d + \sum_{i=1}^L \frac{h_i^2 P_r^i N_r^i}{P_S + N_r^i} \right)^{-1} \right) \right\}, \end{aligned} \quad (26)$$

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

$$\begin{aligned} &\left( \sum_{i=1}^L \sqrt{\frac{h_i^2 P_S P_r^i}{P_S + N_r^i}} \right) \eta^4 + \left( 2hP_S \sum_{i=1}^L \sqrt{\frac{h_i^2 P_r^i}{P_S + N_r^i}} \right) \eta^3 \\ &+ \left( h^2 P_S + N_d + \sum_{i=1}^L \frac{h_i^2 P_r^i N_r^i}{P_S + N_r^i} \right) \eta^2 = \left( N_d + \sum_{i=1}^L \frac{h_i^2 P_r^i N_r^i}{P_S + N_r^i} \right). \end{aligned}$$

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

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<sup>3</sup> 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  $\mathcal{C}$ , if  $\log(|A|) < \mathcal{C}$ . 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 that 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  $\mathcal{C} := (1/2) \log(1 + (P/N))$  [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  $t, t+2, t+4, \dots$ ), 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

$$\begin{aligned} x_{2,t+1} &= \lambda_2 x_{2,t} + u_{2,t} + w_{2,t} \\ &\stackrel{(a)}{=} \lambda_2 (x_{2,t} - \hat{x}_{2,t}) + w_{2,t} \end{aligned} \quad (27)$$

$$x_{1,t+1} \stackrel{(b)}{=} \lambda_1 x_{1,t} + x_{2,t} + w_{1,t} \quad (28)$$

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

$$\begin{aligned} x_{2,t+2} &= \lambda_2 x_{2,t+1} + w_{2,t+1} \\ &= \lambda_2^2 (x_{2,t} - \hat{x}_{2,t}) + \lambda_2 w_{2,t} + w_{2,t+1} \end{aligned} \quad (29)$$

$$\begin{aligned} x_{1,t+2} &= \lambda_1 x_{1,t+1} + x_{2,t+1} + u_{1,t+1} + w_{1,t+1} \\ &\stackrel{(a)}{=} \lambda_1^2 x_{1,t} + (\lambda_1 + \lambda_2) x_{2,t} + u_{1,t+1} \\ &\quad + \lambda_1 w_{1,t} + w_{2,t} + w_{1,t+1} \\ &\stackrel{(b)}{=} \lambda_1^2 (x_{1,t} - \hat{x}_{1,t}) + (\lambda_1 + \lambda_2) (x_{2,t} - \hat{x}_{2,t}) + \lambda_1 w_{1,t} \\ &\quad + w_{2,t} + w_{1,t+1} \end{aligned} \quad (30)$$

<sup>3</sup>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

$$\begin{aligned} \mathbb{E}[x_{2,t+2}^2] &= \lambda_2^4 \mathbb{E}[(x_{2,t} - \hat{x}_{2,t})^2] + \tilde{n}_2 \\ &= \lambda_2^4 2^{-2C} \mathbb{E}[x_{2,t}^2] + \tilde{n}_2. \end{aligned} \quad (31)$$

where the last equality follows from the linear mean-square estimation of a Gaussian variable over a scalar Gaussian channel of capacity  $C$  and  $\tilde{n}_2 := (\lambda_2^2 + 1)n_{w,2}$ . We observe that the lower mode is stable if and only if  $\lambda_2^4 2^{-2C} < 1 \Rightarrow \log(|\lambda_2|) < (C/2)$ . Assuming that  $x_{2,t}$  is stable, the second moment of  $x_{1,t}$  is given by

$$\begin{aligned} \mathbb{E}[x_{1,t+2}^2] &\stackrel{(a)}{=} \lambda_1^4 \mathbb{E}[(x_{1,t} - \hat{x}_{1,t})^2] \\ &\quad + 2\lambda_1^2(\lambda_1 + \lambda_2) \mathbb{E}[(x_{1,t} - \hat{x}_{1,t})(x_{2,t} - \hat{x}_{2,t})] \\ &\quad + (\lambda_1 + \lambda_2)^2 \mathbb{E}[(x_{2,t} - \hat{x}_{2,t})^2] + \tilde{n}_1 \\ &\stackrel{(b)}{=} \lambda_1^4 2^{-2C} \mathbb{E}[x_{1,t}^2] \\ &\quad + 2\lambda_1^2(\lambda_1 + \lambda_2) \mathbb{E}[(x_{1,t} - \hat{x}_{1,t})(x_{2,t} - \hat{x}_{2,t})] \\ &\quad + (\lambda_1 + \lambda_2)^2 2^{-2C} \mathbb{E}[x_{2,t}^2] + \tilde{n}_1 \\ &\stackrel{(c)}{\leq} \lambda_1^4 2^{-2C} \mathbb{E}[x_{1,t}^2] \\ &\quad + 2\lambda_1^2(\lambda_1 + \lambda_2) \sqrt{\mathbb{E}[(x_{1,t} - \hat{x}_{1,t})^2] \mathbb{E}[(x_{2,t} - \hat{x}_{2,t})^2]} \\ &\quad + (\lambda_1 + \lambda_2)^2 2^{-2C} \mathbb{E}[x_{2,t}^2] + \tilde{n}_1 \\ &= \lambda_1^4 2^{-2C} \mathbb{E}[x_{1,t}^2] + 2\lambda_1^2(\lambda_1 + \lambda_2) \sqrt{2^{-2C} \mathbb{E}[x_{1,t}^2]} \\ &\quad \times \sqrt{2^{-2C} \mathbb{E}[x_{2,t}^2]} + (\lambda_1 + \lambda_2)^2 2^{-2C} \mathbb{E}[x_{2,t}^2] + \tilde{n}_1 \\ &\stackrel{(d)}{\leq} k_1 \mathbb{E}[x_{1,t}^2] + k_2 \sqrt{\mathbb{E}[x_{1,t}^2]} + k_3 \end{aligned} \quad (32)$$

where (a) follows from (30) and  $\tilde{n}_1 := (\lambda_1^2 + 1)n_{w,1} + n_{w,2}$ ; (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  $\lambda_1^4 2^{-2C} < 1$ ) and by defining  $k_1 := \lambda_1^4 2^{-2C}$ ,  $k_2 := 2\lambda_1^2(\lambda_1 + \lambda_2) 2^{-2C} \sqrt{M}$ , and  $k_3 := (\lambda_1 + \lambda_2)^2 2^{-2C} M + \tilde{n}_1$ . We now want to find a condition which ensures convergence of the following sequence:

$$\alpha_{t+1} = k_1 \alpha_t + k_2 \sqrt{\alpha_t} + k_3. \quad (33)$$

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

*Lemma 6.1:* Let  $T: \mathbb{R} \mapsto \mathbb{R}$  be a non-decreasing continuous mapping with a unique fixed point  $x^* \in \mathbb{R}$ . If there exists  $u \leq x^* \leq v$  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.  $\square$

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  $k_1 = \lambda_1^4 2^{-2C} < 1 \Rightarrow \log(|\lambda_1|) < (C/2)$ .

The time sharing scheme illustrated above can be generalized to any  $n$ -dimensional plant and the stability conditions can be easily obtained using Lemma 6.1. We know that any system matrix  $A$  can be written in the Jordan form by a similarity matrix transformation. We can then use the following scheme for stabilization. The encoder chooses to send only one component of the observed state vector at each time  $t$  over a Gaussian channel of capacity  $C$ . Assume that for a fraction  $\log(|\lambda_m|)/\sum_{i=1}^K \log(|\lambda_i|)$  of the total available time the encoder transmits the  $m$ -th component of the state vector. Thus the rate available for the transmission of the  $m$ -th state component is  $(\log(|\lambda_m|)/\sum_{i=1}^K \log(|\lambda_i|))C$ . The system will be stable if and only if  $\log(|\lambda_m|) < (\log(|\lambda_m|)/\sum_{i=1}^K \log(|\lambda_i|))C$  for all  $m \in \{1, 2, \dots, K\}$ , which implies  $\sum_{i=1}^K \log(|\lambda_i|) = \log(|\det(A)|) < C$ . For a multi-dimensional system with a controllable  $(A, B)$  pair, any input (control action) can be realized in  $n$  time steps. If the encoder has access to the channel output, then it can refine estimate of the state using noiseless feedback channel (SK coding scheme) during these  $n$  time steps and observe the new state periodically after every  $n$  time steps.  $\square$

*Remark 6.1:* The sufficiency results in Sections III–V for scalar systems can be extended to multi-dimensional systems using the proposed time varying scheme. The sufficient conditions for vector systems will be identical to scalar systems except that  $\log(|\lambda|)$  is replaced with  $\log(|\det(A)|)$  everywhere.

*Remark 6.2:* In [14] the authors studied stabilization of a noiseless multi-dimensional system over a point-to-point scalar Gaussian channel using a linear time invariant scheme, that is the state encoder transmits  $S_t = EX_t$ , where  $E$  is a row vector. This LTI scheme cannot stabilize if the pair  $(A, E)$  is not observable. For example consider a diagonal system matrix  $A$  with two equal eigenvalues. This system cannot be stabilized by any choice of the encoding matrix  $E$ , irrespective of how much power the state encoder is allowed to spend. However our linear time varying scheme can always stabilize the system, even in the presence of process noise.

*Remark 6.3:* As mentioned in Theorem 4.3 and Remark 4.2, the proposed time varying scheme can be used with the non-linear scheme of [28] to achieve the minimum power required for stabilization of noiseless multi-dimensional plants over vector Gaussian channels.

## 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:

$$\begin{aligned}
 & I(X_{[0,T-1]} \rightarrow R_{[0,T-1]}) \\
 & \stackrel{(a)}{=} \sum_{t=0}^{T-1} I(X_{[0,t]}; R_t | R_{[0,t-1]}) \\
 & \stackrel{(b)}{\geq} \sum_{t=0}^{T-1} I(X_t; R_t | R_{[0,t-1]}) \\
 & = I(X_0; R_0) + \sum_{t=1}^{T-1} I(X_t; R_t | R_{[0,t-1]}) \\
 & \stackrel{(c)}{=} I(X_0; R_0) + \sum_{t=1}^{T-1} (h(X_t | R_{[0,t-1]}) - h(X_t | R_{[0,t]})) \\
 & \stackrel{(d)}{=} I(X_0; R_0) \\
 & + \sum_{t=1}^{T-1} (h(AX_{t-1} + BU_{t-1} + W_{t-1} | R_{[0,t-1]}) - h(X_t | R_{[0,t]})) \\
 & \stackrel{(e)}{=} I(X_0; R_0) \\
 & + \sum_{t=1}^{T-1} (h(AX_{t-1} + W_{t-1} | R_{[0,t-1]}) - h(X_t | R_{[0,t]})) \\
 & \stackrel{(f)}{\geq} I(X_0; R_0) \\
 & + \sum_{t=1}^{T-1} (h(AX_{t-1} + W_{t-1} | R_{[0,t-1]}, W_{t-1}) - h(X_t | R_{[0,t]})) \\
 & \stackrel{(g)}{=} I(X_0; R_0) + \sum_{t=1}^{T-1} (h(AX_{t-1} | R_{[0,t-1]}) - h(X_t | R_{[0,t]})) \\
 & \stackrel{(h)}{=} I(X_0; R_0) \\
 & + \sum_{t=1}^{T-1} (\log(|\det(A)|) + h(X_{t-1} | R_{[0,t-1]}) - h(X_t | R_{[0,t]})) \\
 & = I(X_0; R_0) \\
 & + (T-1) \log(|\det(A)|) + h(X_0 | R_0) - h(X_{T-1} | R_{[0,T-1]}) \\
 & = h(X_0) + (T-1) \log(|\det(A)|) - h(X_{T-1} | R_{[0,T-1]}) \\
 & \stackrel{(i)}{\geq} h(X_0) + (T-1) \log(|\det(A)|) - h(X_{T-1}) \\
 & \stackrel{(j)}{\geq} h(X_0) + (T-1) \log(|\det(A)|) - \log((2\pi e)^n |\det(K)|), \tag{34}
 \end{aligned}$$

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 from  $U_{t-1} = \pi_{t-1}(R_{[0,t-1]})$ ; (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  $h(AX) = \log(|\det(A)|) + h(X)$  [35, Theorem 8.6.4]; (i) follows from conditioning reduces entropy; and (j) follows the fact that for a mean square stable system there exists a matrix  $K \succ 0$  with  $\det(\mathbb{E}[X_t^T X_t]) < \det(K)$  for all  $t$  and further for a given covariance matrix the differential entropy is maximized by the Gaussian distribution. We can also write

$$\begin{aligned}
 I(X_{[0,T-1]} \rightarrow R_{[0,t-1]}) & = \sum_{t=0}^{T-1} I(X_{[0,t]}; R_t | R_{[0,t-1]}) \\
 & \stackrel{(a)}{=} \sum_{t=0}^{T-1} I(\bar{X}_{[0,t]} + \bar{f}(U_{[0,t-1]}) ; R_t | R_{[0,t-1]}) \\
 & \stackrel{(b)}{=} \sum_{t=0}^{T-1} I(\bar{X}_{[0,t]}; R_t | R_{[0,t-1]}) = I(\bar{X}_{[0,T-1]} \rightarrow R_{[0,T-1]}), \tag{35}
 \end{aligned}$$

where (a) follows by defining uncontrolled state process  $\bar{X}_{t+1} = A\bar{X}_t + W_t$  and writing  $X_{[0,t]}$  as a sum of  $\bar{X}_{[0,t]}$  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  $\liminf_{T \rightarrow \infty} (1/T) I(\bar{X}_{[0,T-1]} \rightarrow R_{[0,T-1]}) \geq \log(|\det(A)|)$ , 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 [53] which is an adaptation of the well-known Schalkwijk—Kailath scheme [54]. 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  $\mathcal{E}$  observes  $X_0$  and it transmits  $S_{e,0} = \sqrt{P_S/\alpha_0} X_0$ . The decoder  $\mathcal{D}$  receives  $R_0 = hS_{e,0} + Z_{d,0}$ . 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  $\mathcal{C}$  then takes an action  $U_0 = -\lambda \hat{X}_0$  which results in

$$\begin{aligned}
 X_1 & = \lambda X_0 + U_0 + W_0 = \lambda(X_0 - \hat{X}_0) + W_0 \\
 & = -\frac{\lambda}{h} \sqrt{\frac{\alpha_0}{P_S}} Z_{d,0} + W_0. \tag{36}
 \end{aligned}$$

The new plant state  $X_1 \sim \mathcal{N}(0, \alpha_1)$  where  $\alpha_1 = (\lambda^2 N_d / h^2 P_S) \alpha_0 + n_w$ .

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

$$\begin{aligned} \hat{X}_t &= \mathbb{E}[X_t | R_{[1,t]}] \stackrel{(a)}{=} \mathbb{E}[X_t | R_t] \stackrel{(b)}{=} \frac{\mathbb{E}[X_t R_t]}{\mathbb{E}[R_t^2]} R_t \\ &\stackrel{(c)}{=} \left( \frac{h\sqrt{2\beta P_S \alpha_t}}{2h^2\beta P_S + N_d} \right) R_t, \end{aligned}$$

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  $\mathcal{C}$  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 plant 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

$$\begin{aligned} \alpha_{t+1} &:= \mathbb{E}[X_{t+1}^2] = \lambda^2 \mathbb{E}[(X_t - \hat{X}_t)^2] + \mathbb{E}[W_t^2] \\ &= \lambda^2 \left( \frac{N_d}{2h^2\beta P_S + N_d} \right) \alpha_t + n_w, \quad (37) \end{aligned}$$

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,  $t = 2, 4, 6, \dots$ :* The encoder  $\mathcal{E}$  observes  $X_t$  and transmits  $S_{e,t} = \sqrt{2(1-\beta)P_S/\alpha_t} X_t$ . In this phase the relay nodes choose to transmit their own signal to the decoder  $\mathcal{D}$  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  $\mathcal{D}$ . The signal transmitted from the  $i$ -th relay node is thus given by  $S_{r,t}^i = \sqrt{2P_r^i/(2\beta P_S + N_r^i)}(S_{e,t-1} + Z_{r,t-1}^i)$ . The decoder  $\mathcal{D}$  accordingly receives

$$\begin{aligned} R_t &= hS_{e,t} + \sum_{i=1}^L h_i S_{r,t}^i + Z_{d,t} \\ &= L_1 X_t + L_2 X_{t-1} + \tilde{Z}_t \quad (38) \end{aligned}$$

where  $L_1 = \sqrt{2(1-\beta)h^2 P_S/\alpha_t}$ ,  $L_2 = \sum_{i=1}^L \sqrt{4\beta h_i^2 P_r^i/(2\beta P_S + N_r^i)} \alpha_{t-1}$ , and  $\tilde{Z}_t = Z_{d,t} + \sum_{i=1}^L \sqrt{2h_i^2 P_r^i/2\beta P_S + N_r^i} Z_{r,t-1}^i$  is a white Gaussian noise sequence with zero mean and variance  $\tilde{N}(\beta, \{P_r^i\}_{i=1}^L) = N_d + \sum_{i=1}^L (2h_i^2 P_r^i N_r^i / 2\beta P_S + N_r^i)$ . The decoder then computes the MMSE estimate of  $X_t$  given all previous channel outputs  $\{R_1, R_2, \dots, R_t\}$  in the following three steps:

- 1) Compute the MMSE prediction of  $R_t$  from  $\{R_1, R_2, \dots, R_{t-1}\}$ , 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:

$$\begin{aligned} I_t &= R_t - \hat{R}_t = L_1 X_t + L_2 (X_{t-1} - \hat{X}_{t-1}) + \tilde{Z}_t \\ &\stackrel{(a)}{=} \left( \frac{\lambda L_1 + L_2}{\lambda} \right) X_t - \frac{L_2}{\lambda} W_{t-1} + \tilde{Z}_t \quad (39) \end{aligned}$$

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

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

$$\begin{aligned} \hat{X}_t &= \mathbb{E}[X_t | I_t] \stackrel{(a)}{=} \frac{\mathbb{E}[X_t I_t]}{\mathbb{E}[I_t^2]} I_t \\ &\stackrel{(b)}{=} \frac{\lambda(\lambda L_1 + L_2) \alpha_t}{(\lambda L_1 + L_2)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_r)} I_t \quad (40) \end{aligned}$$

where (a) follows from an MMSE estimation of a Gaussian variable; and (b) follows from  $\mathbb{E}[X_t I_t] = ((\lambda L_1 + L_2)/\lambda) \alpha_t$  and  $\mathbb{E}[I_t^2] = ((\lambda L_1 + L_2)/\lambda)^2 \alpha_t + (L_2^2 n_w / \lambda^2) + \tilde{N}(\beta, P_r)$ . The controller  $\mathcal{C}$  takes action  $U_t = -\lambda \hat{X}_t$  which results in  $X_{t+1} = \lambda(X_t - \hat{X}_t) + W_t$ . The new plant 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 distributed with variance given in (41) on the top of the next page. In (41), (a) follows from  $\mathbb{E}[X_t \hat{X}_t] = \mathbb{E}[\hat{X}_t^2] = (\lambda L_1 + L_2)^2 \alpha_t / ((\lambda L_1 + L_2)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_r))$ ; (b) follows by substituting the values of  $L_1$  and  $L_2$ ; and (c) by substituting  $\alpha_t/\alpha_{t-1}$  using (37) and by defining  $k := N/(2h^2\beta P_S + N)$ ,  $k_1 := (\sum_{i=1}^L \sqrt{4h_i^2 \beta P_r^i / (2\beta P_S + N_r^i)})^2$ ,  $k_2 := \sqrt{2h^2(1-\beta)P_S}$ .

We want to find the values of the parameter  $\lambda$  for which the second moment of the state remains bounded. Rewriting (37) and (41), the variance of the state is given by

$$\begin{aligned} \alpha_t &= \lambda^2 (\lambda^2 k \alpha_{t-2} + n_w) \\ &\times \frac{\left( \frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-2}} + \tilde{N}(\beta, P_r)}{\underbrace{\left( k_2 + \sqrt{k_1 k + \frac{n_w k_1}{\lambda^2} \frac{1}{\alpha_{t-2}}} \right)^2 + \left( \frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-2}} + \tilde{N}(\beta, P_r)}_{\triangleq f(\alpha_{t-2})}} + n_w \\ &= \lambda^2 (\lambda^2 k \alpha_{t-2} + n_w) f(\alpha_{t-2}) + n_w, \quad t = 3, 5, 7, \dots \quad (42) \end{aligned}$$

$$\alpha_t = \lambda^2 \left( \frac{N}{2h^2\beta P_S + N} \right) \alpha_{t-1} + n_w, \quad t = 2, 4, 6, \dots \quad (43)$$

where  $\alpha_1 = (\lambda^2 N/h^2 P_S) \alpha_0 + n_w$ . If the odd indexed sub-sequence  $\{\alpha_{2t+1}\}$  in (42) is bounded, then the even indexed sub-sequence  $\{\alpha_{2t}\}$  in (43) is also bounded. Thus it is sufficient to consider the odd indexed sub-sequence  $\{\alpha_{2t+1}\}$ . We will now construct a sequence  $\{\alpha'_t\}$  which upper bounds the sub-sequence  $\{\alpha_{2t+1}\}$ . Then we will derive conditions on the system parameter  $\lambda$  for which  $\{\alpha'_t\}$  stays bounded and consequently the boundedness of  $\{\alpha_{2t+1}\}$  will be guaranteed. In order to construct the upper sequence  $\{\alpha'_t\}$ , we work on the term  $f(\alpha_{t-2})$  in (42) and make use of the following lemma.

$$\begin{aligned}
 \alpha_{t+1} &= \lambda^2 \mathbb{E} \left[ (X_t - \hat{X}_t)^2 \right] + \mathbb{E} [W_t^2] \stackrel{(a)}{=} \lambda^2 \alpha_t \left( \frac{L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_r)}{(\lambda L_1 + L_2)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_r)} \right) + n_w \\
 &\stackrel{(b)}{=} \lambda^2 \alpha_t \left( \frac{\left( \sum_{i=1}^L \sqrt{\frac{4h_i^2 \beta P_S P_r^i}{2\beta P_S + N_r^i}} \right)^2 \frac{n_w}{\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_r)}{\left( \lambda \sqrt{2h^2(1-\beta P_S)} + \sum_{i=1}^L \sqrt{\frac{4h_i^2 \beta P_S P_r^i}{2\beta P_S + N_r^i}} \frac{\alpha_t}{\alpha_{t-1}} \right)^2 + \left( \sum_{i=1}^L \sqrt{\frac{4h_i^2 \beta P_S P_r^i}{2\beta P_S + N_r^i}} \right)^2 \frac{n_w}{\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_r)} \right) + n_w \\
 &\stackrel{(c)}{=} \lambda^2 (\lambda^2 k \alpha_{t-1} + n_w) \left( \frac{(n_w k_1) \frac{1}{\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_r)}{\left( \lambda k_2 + \sqrt{\frac{k_1}{\lambda^2}} (\lambda^2 k + n_w \frac{1}{\alpha_{t-1}}) \right)^2 + (n_w k_1) \frac{1}{\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_r)} \right) + n_w \\
 &= \lambda^2 (\lambda^2 k \alpha_{t-1} + n_w) \left( \frac{\left( \frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-1}} + \tilde{N}(\beta, P_r)}{\left( k_2 + \sqrt{k_1 k} + \frac{n_w k_1}{\lambda^2} \frac{1}{\alpha_{t-1}} \right)^2 + \left( \frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-1}} + \tilde{N}(\beta, P_r)} \right) + n_w \tag{41}
 \end{aligned}$$

1) *Lemma A. 1 ([39, Lemma 4.1]):* Consider a function  $f(x) = a + (b/x)/((c + \sqrt{d + (b/x)})^2 + a + (b/x))$  defined over the interval  $[0, \infty)$ , where  $0 \leq a, b, c, d < \infty$ . The function  $f(x)$  can be upper bounded as  $f(x) \leq f_\infty + (m/x)$  for some  $0 < m < \infty$ , where  $f_\infty = \lim_{x \rightarrow \infty} f(x) = a/((c + \sqrt{d})^2 + a)$ .

Starting from (42) and by using the above lemma, we write the following series of inequalities

$$\begin{aligned}
 \alpha_t &= \lambda^2 (\lambda^2 k \alpha_{t-2} + n_w) f(\alpha_{t-2}) + n_w \\
 &\stackrel{(a)}{\leq} \lambda^2 (\lambda^2 k \alpha_{t-2} + n_w) \left( f_\infty + \frac{m}{\alpha_{t-2}} \right) + n_w \\
 &= \lambda^4 k f_\infty \alpha_{t-2} + \frac{\lambda^2 n_w m}{\alpha_{t-2}} + \lambda^2 n_w f_\infty + \lambda^4 m k + n_w \\
 &\stackrel{(b)}{\leq} \lambda^4 k f_\infty \alpha_{t-2} + \lambda^2 m + \lambda^2 n_w f_\infty + \lambda^4 m k + n_w =: g(\alpha_{t-2}), \tag{44}
 \end{aligned}$$

where (a) follows from Lemma A.1 and  $f_\infty = \lim_{\alpha \rightarrow \infty} f(\alpha) = (\tilde{N}(\alpha, P_r))/((k_2 + \sqrt{k_1 k})^2 + \tilde{N}(\beta, P_r))$ ; and (b) follows from the fact that  $\alpha_t \geq n_w$  for all  $t$  according to (43) and (42). Since  $g(\alpha)$  in (44) is a linearly increasing function, it can be used to construct the sequence  $\{\alpha'_t\}$ , which upper bounds the odd indexed sub-sequence  $\{\alpha_{2t+1}\}$  given in (42). We construct the sequence  $\{\alpha'_t\}$  for all  $t \geq 1$  as

$$\begin{aligned}
 \alpha_{2t+1} &\leq \alpha'_{t+1} := g(\alpha'_t) \\
 &\stackrel{(a)}{=} \lambda^4 k f_\infty \alpha'_t + \lambda^2 m + \lambda^2 n_w f_\infty + \lambda^4 m k + n_w \\
 &\stackrel{(b)}{=} (\lambda^4 k f_\infty)^t \alpha'_0 \\
 &\quad + (\lambda^2 m + \lambda^2 n_w f_\infty + \lambda^4 m k + n_w) \sum_{i=0}^{t-1} (\lambda^4 k f_\infty)^i, \tag{45}
 \end{aligned}$$

where (a) follows from (44) and (b) follows by recursively applying (a).

We observe from (45) that if  $(\lambda^4 k f_\infty) = (\lambda^4 k \tilde{N}(\beta, P_r))/((k_2 + \sqrt{k_1 k})^2 + \tilde{N}(\beta, P_r)) < 1$ , then the

sequence  $\{\alpha'_t\}$  converges as  $t \rightarrow \infty$  and consequently the original sequence  $\{\alpha_t\}$  is guaranteed to stay bounded. Thus the system in (1) can be mean square stabilized if

$$\begin{aligned}
 \lambda^4 &< \left( \frac{(k_2 + \sqrt{k_1 k})^2 + \tilde{N}(\beta, \{P_r^i\}_{i=1}^L)}{k \tilde{N}(\beta, \{P_r^i\}_{i=1}^L)} \right) \tag{46} \\
 &\Rightarrow \log(\lambda) < \\
 &\frac{1}{4} \left( \log \left( 1 + \frac{2h^2 \beta P_S}{N_d} \right) + \log \left( 1 + \frac{\tilde{M}(\beta, \{P_r^i\}_{i=1}^L)}{\tilde{N}(\beta, \{P_r^i\}_{i=1}^L)} \right) \right), \tag{47}
 \end{aligned}$$

where the last equality follows from  $k = N/(2h^2 \beta P_S + N)$  and  $M(\beta, \{P_r^i\}_{i=1}^L) := (k_2 + \sqrt{k_1 k})^2$ . Since the relay nodes amplify the desired signal as well as the noise, which is then superimposed at the decoder to the signal coming directly from the state encoder, the optimal choice of the transmit powers  $\{P_r^i\}_{i=1}^L : \sum_{i=1}^L P_r^i \leq P_R$  depends on the parameters  $\{P_S, \{N_r^i\}_{i=1}^L, N_d, h, h_i, \beta\}$ . Moreover, the optimal choice of the power allocation factor  $\beta$  at the state encoder also depends on these parameters. Therefore, we rewrite (47) as (21).  $\square$

### C. Remark 5.2 on Information Rate

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

$$\begin{aligned}
 &\lim_{T \rightarrow \infty} \frac{1}{2T} I \left( \{S_{e,2t-1}\}_{t=1}^T, \{S_{e,2t}\}_{t=1}^T; \{R_{2t-1}\}_{t=1}^T, \{I_{2t}\}_{t=1}^T \right) \\
 &= \lim_{T \rightarrow \infty} \frac{1}{2T} \left[ h \left( \{R_{2t-1}\}_{t=1}^T, \{I_{2t}\}_{t=1}^T \right) \right. \\
 &\quad \left. - h \left( \{R_{2t-1}\}_{t=1}^T, \{I_{2t}\}_{t=1}^T | \{S_{e,2t-1}\}_{t=1}^T, \{S_{e,2t}\}_{t=1}^T \right) \right]
 \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{=} \lim_{T \rightarrow \infty} \frac{1}{2T} \left[ \sum_{t=1}^T (h(R_{2t-1}) \right. \\
&\quad \left. + h(I_{2t}) - h(R_{2t-1}|S_{e,2t-1}) - h(I_{2t}|S_{e,2t})) \right] \\
&\stackrel{(b)}{=} \lim_{T \rightarrow \infty} \frac{1}{2T} [T(h(R_{2t-1}) + h(I_{2t}) - h(R_{2t-1}|S_{e,2t-1}) \\
&\quad - h(I_{2t}|S_{e,2t}))] = \frac{1}{2} (I(S_{e,2t-1}; R_{2t-1}) + I(S_{e,2t}; I_{2t})), \quad (48)
\end{aligned}$$

where (a) follows from the fact that  $P(I_{2t}, R_{2t-1}|S_{e,2t}, S_{e,2t-1}) = P(I_{2t}|S_{e,2t})P(R_{2t-1}|S_{e,2t-1})$ , the channel is memoryless, the random variables are Gaussian and  $\mathbb{E}[R_{2l-1}R_{2k-1}] = \mathbb{E}[I_{2l}I_{2k}] = 0$  for  $k \neq l$ , and  $\mathbb{E}[R_{2l-1}I_{2k}] = 0$  for all  $l, k = 1, 2, 3, \dots$ ; 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

$$\begin{aligned}
I(S_{e,2t-1}; R_{2t-1}) &= h(R_{2t-1}) - h(R_{2t-1}|S_{e,2t-1}) \\
&= h(R_{2t-1}) - h(Z_{2t-1}) \stackrel{(a)}{=} \frac{1}{2} \log \left( 1 + \frac{2h^2\beta P_S}{N} \right), \quad (49)
\end{aligned}$$

where (a) follows from  $R_{2t-1} \sim \mathcal{N}(0, 2h^2\beta P_S + N)$  and  $Z_{2t-1} \sim \mathcal{N}(0, N)$ . 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

$$\begin{aligned}
I(S_{e,2t}; I_{2t}) &= h(I_{2t}) - h(I_{2t}|S_{e,2t}) = h(I_{2t}) - h(\tilde{Z}_{2t}) \\
&\stackrel{(a)}{=} \frac{1}{2} \log \left( 1 + \frac{\tilde{M}(\beta, P_r)}{\tilde{N}(\beta, P_r)} \right) \quad (50)
\end{aligned}$$

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

$$\frac{1}{4} \left( \log \left( 1 + \frac{2h^2\beta P_S}{N} \right) + \log \left( 1 + \frac{\tilde{M}(\beta, P_r)}{\tilde{N}(\beta, P_r)} \right) \right). \quad (51)$$

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  $t \in \mathbb{N}$ ,  $x_0 \in \mathbb{R}$ . 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  $x_{t+1} = T(x_t) \leq T(x_{t-1}) = x_t$ . 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  $T(\cdot)$ . Thus the sequence  $\{x_t\}$  con-

verges 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  $x_{t+1} = T(x_t) \geq T(x_{t-1}) = x_t$ . 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  $T(\cdot)$ . Since  $\{x_t\}$  is strictly increasing and upper bounded by  $x^*$  for  $x_0 \in [x^*, \infty)$ , it converges [56, Theorem 3.14].  $\square$

#### REFERENCES

- [1] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, 1989.
- [2] J. Baillieul, "Feedback designs for controlling device arrays with communication channel bandwidth constraints," in *Proc. ARO Workshop Smart Struct.*, Aug. 1999.
- [3] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints-II: Stabilization with limited information feedback," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 1049–1053, May 1999.
- [4] G. N. Nair and R. J. Evans, "Stabilization with data-rate-limited feedback: Tightest attainable bounds," *Syst. Control Lett.*, vol. 41, no. 1, pp. 49–56, Sep. 2000.
- [5] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE Trans. Autom. Control*, vol. 46, no. 9, pp. 1384–1400, Sep. 2001.
- [6] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM J. Control Optim.*, vol. 43, no. 2, pp. 413–436, 2004.
- [7] N. Elia, "When Bode meets Shannon: Control-oriented feedback communication schemes," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1477–1488, 2004.
- [8] A. S. Matveev and A. V. Savkin, "The problem of LQG optimal control via a limited capacity communication channel," *Syst. Control Lett.*, vol. 53, no. 1, pp. 51–64, Sep. 2004.
- [9] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1549–1561, Sep. 2004.
- [10] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1196–1201, 2004.
- [11] N. C. Martins, M. A. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 438–447, Mar. 2006.
- [12] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over noisy communication links—part I: Scalar systems," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.
- [13] A. S. Matveev and A. V. Savkin, "An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels," *SIAM J. Control Optim.*, vol. 46, no. 4, pp. 1323–1367, 2007.
- [14] J. H. Braslavsky, R. H. Middleton, and J. S. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels," *IEEE Trans. Autom. Control*, vol. 52, no. 8, pp. 1391–1403, Aug. 2007.
- [15] N. C. Martins and M. A. Dahleh, "Feedback control in the presence of noisy channels: 'Bode-like' fundamental limitations of performance," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1604–1615, Aug. 2008.
- [16] R. H. Middleton, A. J. Rojas, J. S. Freudenberg, and J. H. Braslavsky, "Feedback stabilization over a first order moving average Gaussian noise channel," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 163–167, Jan. 2009.
- [17] S. Yüksel and S. Tatikonda, "A counterexample in distributed optimal sensing and control," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 841–844, 2009.
- [18] S. Yüksel and T. Başar, "Control over noisy forward and reverse channels," *IEEE Trans. Autom. Control*, vol. 56, pp. 1014–1029, May 2011.
- [19] L. Bao, M. Skoglund, and K. H. Johansson, "Iterative encoder-controller design for feedback control over noisy channels," *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 265–278, 2011.
- [20] A. Farhadi and N. U. Ahmed, "Suboptimal decentralized control over noisy communication channels," *Syst. Control Lett.*, vol. 60, pp. 285–293, Apr. 2011.

- [21] C. D. Charalambous, A. Farhadi, and S. Z. Denic, "Control of continuous-time linear Gaussian systems over additive Gaussian wireless fading channels: A separation principle," *IEEE Trans. Autom. Control*, vol. 53, no. 4, pp. 1013–1019, May 2008.
- [22] L. Yiçian, E. Tuncel, J. Chen, and S. Weizhou, "Optimal tracking performance of discrete-time systems over an additive white noise channel," in *Proc. IEEE CDC*, 2009, pp. 4026–4031.
- [23] J. S. Freudenberg, R. H. Middleton, and V. Solo, "Stabilization and disturbance attenuation over a Gaussian communication channel," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 795–799, Mar. 2010.
- [24] Z. Shu and R. H. Middleton, "Stabilization over power-constrained parallel Gaussian channels," *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1718–1724, Jul. 2011.
- [25] S. Yüksel, "Characterization of information channels for asymptotic mean stationarity and stochastic stability of non-stationary/unstable linear systems," *IEEE Trans. Inform. Theory*, vol. 58, no. 10, pp. 6332–6354, Oct. 2012.
- [26] E. I. Silva, G. C. Goodwin, and D. E. Quevedo, "Control system design subject to SNR constraints," *Automatica*, vol. 46, no. 2, pp. 428–436, 2010.
- [27] F. J. Vargas, E. I. Silva, and J. Chen, "Stabilization of TITO systems over parallel SNR-constraint AWN channels," in *Proc. 7th IFAC Symp. Robust Control Design*, 2012, pp. 85–90.
- [28] U. Kumar, V. Gupta, and J. N. Laneman, "On stability across a Gaussian product channel," in *Proc. IEEE CDC*, Dec. 2011, pp. 3142–3147.
- [29] V. A. Vaishampayan, "Combined Source-Channel Coding for Bandlimited Waveform Channels," Ph.D. dissertation, University of Maryland, 1989.
- [30] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. Boston, MA: Birkhäuser, 2013.
- [31] A. S. Matveev and A. V. Savkin, *Estimation and Control Over Communication Networks*. Boston, MA: Birkhäuser, 2008.
- [32] A. A. Zaidi, T. J. Oechtering, S. Yüksel, and M. Skoglund, "Stabilization and control over Gaussian networks," in *Information and Control in Networks*, B. Bernhardsson, G. Como, and A. Rantzer, Eds: Springer, 2014.
- [33] S. Tatikonda, "Some scaling properties of large distributed control systems," in *Proc. IEEE CDC*, Dec. 2003, pp. 3142–3147.
- [34] V. Gupta, A. F. Dana, J. P. Hespanha, R. M. Murray, and B. Hassibi, "Data transmission over networks for estimation and control," *IEEE Trans. Autom. Control*, vol. 54, no. 8, pp. 1807–1819, Aug. 2009.
- [35] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.
- [36] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.
- [37] A. A. Zaidi, T. J. Oechtering, and M. Skoglund, "Rate sufficient conditions for closed-loop control over AWGN relay channels," in *Proc. IEEE ICCA*, Jun. 2010, pp. 602–607.
- [38] A. A. Zaidi, T. J. Oechtering, S. Yüksel, and M. Skoglund, "Closed-loop control over half-duplex AWGN relay channels," *Reglermöte*, Jun. 2010.
- [39] A. A. Zaidi, T. J. Oechtering, S. Yüksel, and M. Skoglund, "Sufficient conditions for closed-loop control over a Gaussian relay channel," in *Proc. IEEE ACC*, Jun. 2011, pp. 2240–2245.
- [40] U. Kumar, V. Gupta, and J. N. Laneman, "Sufficient conditions for stabilizability over Gaussian relay channel and cascade channels," in *Proc. IEEE CDC*, Dec. 2010, pp. 4765–4770.
- [41] E. I. Silva, M. S. Derpich, and J. Ostergaard, "A framework for control system design subject to average data-rate constraints," *IEEE Trans. Autom. Control*, vol. 56, no. 8, pp. 1886–1899, Aug. 2011.
- [42] J. L. Massey, "Causality, feedback and directed information," in *Proc. IEEE ISITA*, 1990, pp. 303–305.
- [43] G. M. Lipsa and N. C. Martins, "Optimal memoryless control in Gaussian noise: A simple counterexample," *Automatica*, vol. 47, pp. 552–558, Mar. 2011.
- [44] A. A. Zaidi, S. Yüksel, T. J. Oechtering, and M. Skoglund, "On optimal policies for control and estimation over a Gaussian relay channel," *Automatica*, vol. 49, no. 9, pp. 2892–2897, Sep. 2013.
- [45] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge Univ. Press, 2005.
- [46] N. Wernersson and M. Skoglund, "Nonlinear coding and estimation for correlated data in wireless sensor networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 2932–2939, Oct. 2009.
- [47] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.
- [48] M. Andersson, A. A. Zaidi, N. Wernersson, and M. Skoglund, "Nonlinear distributed sensing for closed-loop control over Gaussian channels," in *Proc. IEEE Swe-CTW*, Oct. 2011, pp. 19–23.
- [49] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," *Inform. Theory Appl. Workshop*, pp. 177–182, Feb. 2007.
- [50] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Processing*, vol. 56, no. 2, pp. 757–770, Feb. 2008.
- [51] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.
- [52] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," *IEEE Trans. Inform. Theory*, vol. 54, pp. 5247–5251, 2008.
- [53] S. Bross and M. Wigger, "On the relay channel with receiver-transmitter feedback," *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 275–291, 2009.
- [54] Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback-I: No bandwidth constraint," *IEEE Trans. Inform. Theory*, vol. 12, no. 2, pp. 172–182, 1966.
- [55] M. Hayes, *Statistical Digital Signal Processing and Modelling*. New York: Wiley, 1996.
- [56] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976.



**Ali A. Zaidi** received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Pakistan, in 2005, the M.Sc. degree in wireless systems and the Ph.D. degree in Telecommunications from KTH Royal Institute of Technology, Stockholm, Sweden, in 2008 and 2013, respectively.

Since February 2014, he has been a Post-Doctoral Researcher in the Department of Signals and Systems, Chalmers University of Technology, Stockholm, Sweden. His research interests include information transmission, information processing, and feedback control in wireless networked systems.



**Tobias J. Oechtering** (S'01–M'08) received the Diplom-Ingenieur degree in electrical engineering and information technology from RWTH Aachen University, Germany, in 2002 and the Doktor-Ingenieur degree in electrical engineering from the Technische Universität Berlin, Germany, in 2007.

In November 2008, he joined the Communication Theory Lab, KTH Royal Institute of Technology, Stockholm, Sweden, as a Post-Doctoral Researcher, became an Assistant Professor in July 2010, and works now as Associate Professor since May 2013.

Previously he was with the Fraunhofer German-Sino Lab for Mobile Communications, Berlin, Germany, from 2007 to 2008. His research interests include information theory, distributed detection, and signal processing, as well as communication for networked control and physical layer security.

Dr. Oechtering received the "Advancement Award" from the Vodafone Foundation in June 2009. He has served on several technical program committees for conferences and is currently an Editor for the *IEEE Communications Letters*.



**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* (Birkhuser, 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, Goteborg, 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.