Feedback Capacity of Stationary Sources over Gaussian Intersymbol Interference Channels

Shaohua Yang*, Aleksandar Kavčić+, Sekhar Tatikonda°
*Marvell Semiconductor Inc, Santa Clara, CA
+Information Engineering Department, Chinese University of Hong Kong (on leave from Harvard University)
°Department of Electrical Engineering, Yale University, New Haven, CT

***Abstract*— We consider discrete-time channels with finite-length intersymbol interference and additive Gaussian noise. The channel noise is considered to be a stationary ARMA (autoregressive and/or moving average) Gaussian process. We assume that the channel is used with feedback and compute the maximal feedback information rate achievable by stationary power-constrained sources. We show that feedback-dependent Gauss-Markov sources achieve the feedback capacity, and a Kalman-Bucy filter is optimal for processing the feedback. We then formulate the capacity computation into a system of equations whose solution gives the optimal signaling and the feedback capacity for stationary sources. In general, the equations are solved numerically, but for first-order channels, the problem admits a closed form solution.**

I. Introduction

Intersymbol interference (ISI) channels with Gaussian noise are standard models for data storage channels. Most data storage channels are used with binary (bipolar) inputs. For binary inputs, the information rate can be evaluated using the methods in [1], [2], [3]. Tight lower bounds on the capacity of binary-input channels can be computed using the method in [4], while upper bounds can be evaluated by computing the (delayed) feedback capacity [5]. For channels used with multiple (non-binary) levels, the feed-forward capacity is computed by the waterfilling method [6].

In this paper, we consider continuous-input power-constrained ISI Gaussian noise channels used with feedback, and compute their capacities for stationary channel inputs. For memoryless channels, Shannon [7] showed that feedback does not increase the capacity, and Schalkwijk and Kalaith [8] proposed a capacity achieving feedback code. For channels with memory, bounds have been developed for the feedback capacity [9], [10], [11], [12], [13]. Butman [9] devised a feedback coding scheme for autoregressive (AR) noise channels and showed that feedback increases the capacity.

Notation: Random variables are denoted by upper-case letters, e.g., X_t, and their realizations are denoted using lower case letters, e.g., x_t. A sequence $x_i, x_{i+1}, \dots, x_j$ is shortly denoted by x_i^j. The letter E stands for the expectation. The differential entropy of a random variable X_t is denoted by $h(X_t)$. Bold uppercase letters stand for matrices (e.g., $\mathbf{K}$), while underlined letters stand for column vectors (e.g., $\underline{a}$).

Paper organization: We express the discrete-time ISI Gaussian channel using the state-space notation in Section II. In Section III, we manipulate information rate into a form suitable for computing the feedback capacity. We introduce the optimal source structure in Section IV. In Section V the feedback capacity formulas are derived. Section VI shows the capacity curves for the dicode channel. Section VII concludes the paper. Longer proofs can be found in the Appendix.

II. Channel Model

We start with a general ISI channel corrupted by additive Gaussian (possibly correlated) noise, and transform it into an equivalent ISI channel corrupted by *white* Gaussian noise, for which we then find the feedback capacity.

The channel input X_t, for $t \in \mathbb{Z}$, is passed through a Gaussian-noise intersymbol interference channel whose channel coefficients are $g_0 = 1, g_1, g_2, \dots, g_I$, and

$$G(z) = \sum_{i=0}^{I} g_i z^{-i} \tag{1}$$

is the channel transfer function. The channel output is

$$R_t = \sum_{i=0}^{I} g_i X_{t-i} + N_t, \tag{2}$$

where N_t is the Gaussain channel noise. It is assumed that the power spectral density $S_N(\omega)$ of N_t is known. Further, we assume that N_t is an ARMA (autoregressive moving average process), which can be modeled by filtering a white Gaussian noise process W_t (with a power spectral density σ_W^2) through a linear time invariant filter. Hence, $S_N(\omega)$ may be written as

$$S_N(\omega) = \sigma_W^2 \frac{\left(1 - \sum\limits_{m=1}^{M} p_m \mathrm{e}^{-jm\omega}\right)\left(1 - \sum\limits_{m=1}^{M} p_m \mathrm{e}^{jm\omega}\right)}{\left(1 + \sum\limits_{k=1}^{K} q_k \mathrm{e}^{-jk\omega}\right)\left(1 + \sum\limits_{k=1}^{K} q_k \mathrm{e}^{jk\omega}\right)}, \tag{3}$$

where p_m and q_k are the filter's poles and zeros. Since the poles and zeros of (3) appear in pairs symmetric with respect to the unit circle [14], without loss of generality, we may assume that $|p_m| < 1$ and $|q_k| < 1$. Hence, the filter

$$H(z) = \left(1 - \sum_{m=1}^{M} p_m z^{-m}\right) \Bigg/ \left(1 + \sum_{k=1}^{K} q_k z^{-k}\right) \tag{4}$$

and its inverse are both causal, stable and invertible.

We make the following assumptions on the channel usage:

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

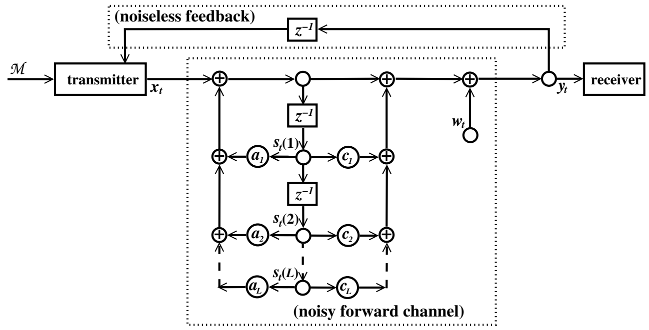

Fig. 1. The equivalent state-space model for a linear ISI channel with stationary ARMA noise and a noiseless feedback link.

1) The power of the channel input process is constrained[1] $\lim_{n\to\infty} \mathrm{E}\left[\sum_{t=1}^{n} X_t^2\right]/n = P$.
2) The prior channel outputs $R_{-\infty}^{t-1}$ are known by feedback to the transmitter before the transmission of X_t.
3) Transmission starts at time $t = 1$, i.e., $X_t = 0$ for $t \leq 0$. Thus, the noise history $N_{-\infty}^{0}$ is known to both the transmitter and the receiver.

Since the filter $H(z)$ is invertible, we may apply $H^{-1}(z)$ to the channel output R_t without changing the channel capacity. Let $L = \max(I + K, M)$, and let

$$H^{-1}(z)G(z) = \left(1 + \sum_{l=1}^{L} c_l z^{-l}\right) \Bigg/ \left(1 - \sum_{l=1}^{L} a_l z^{-l}\right). \quad (5)$$

The coefficients c_l are obtained by multiplying $G(z)$ with the denominator in (4), and $a_l = p_l$ for $l \leq M$. The equivalent channel has X_t as the channel input, Y_t as the channel output, and white Gaussian noise W_t (with power σ_W^2)

$$Y_t - \sum_{l=1}^{L} a_l Y_{t-l} = X_t + \sum_{l=1}^{L} c_l X_{t-l} + W_t - \sum_{l=1}^{L} a_l W_{t-l}. \quad (6)$$

Notice that the channel is completely characterized by its order L and the two vectors of tap coefficients

$$\underline{a} \triangleq [a_1, a_2, \cdots, a_L]^{\mathrm{T}}, \quad \underline{c} \triangleq [c_1, c_2, \cdots, c_L]^{\mathrm{T}}. \quad (7)$$

The channel is used with feedback as depicted in Figure 1, i.e., the channel outputs $Y_{-\infty}^{t-1}$ are known to the transmitter at the time of transmitting symbol X_t.

A channel depicted in Figure 1 has a state-space representation. Let the vector of values stored in the shift registers in Figure 1, i.e., $\underline{S}_t \triangleq [S_t(1), S_t(2), \ldots, S_t(L)]^{\mathrm{T}}$, be the channel *state* vector. The state space channel equations are

$$\underline{S}_t = \mathbf{A}\underline{S}_{t-1} + \underline{b}X_t \quad (8)$$
$$Y_t = (\underline{a} + \underline{c})^{\mathrm{T}} \underline{S}_{t-1} + X_t + W_t, \quad (9)$$

where W_t is white Gaussian noise with variance σ_W^2. The constant square matrix $\mathbf{A}$ and vector $\underline{b}$ are defined as

$$\mathbf{A} \triangleq \begin{bmatrix} a_1 & a_2 & \ldots & a_{L-1} & a_L \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}, \quad \underline{b} \triangleq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (10)$$

From channel assumptions 1)-3), we have the following:

I) Since $X_t = 0$ for $t \leq 0$, the initial channel state $\underline{s}_0 = \underline{0}$ is known to both the transmitter and the receiver.

II) The sequences $\underline{S}_1^t$ and X_1^t determine each other uniquely according to equation (8). Hence, we interchangeably use the input sequence X for the state sequence $\underline{S}$.

III) Given the channel states $\underline{S}_{t-1}^t = (\underline{S}_{t-1}, \underline{S}_t)$, the channel output Y_t is statistically independent of previous channel states $\underline{S}_0^{t-2}$ and outputs Y_1^{t-1}, that is

$$P_{Y_t|\underline{S}_0^t, Y_1^{t-1}}\left(y_t \left|\underline{s}_0^t, y_1^{t-1}\right.\right) = P_{Y_t|\underline{S}_{t-1}^t}\left(y_t \left|\underline{s}_{t-1}^t\right.\right). \quad (11)$$

Since the variance of the process W_t is σ_W^2, the conditional differential entropy of the channel output equals

$$h\left(Y_t \left|\underline{S}_0^t, Y_1^{t-1}\right.\right) = h\left(Y_t \left|\underline{S}_{t-1}^t\right.\right) = \frac{1}{2}\log(2\pi e \sigma_W^2). \quad (12)$$

IV) The channel input X_t is causally dependent on all previous channel states $\underline{S}_0^{t-1}$ and channel outputs Y_1^{t-1}. We characterize the source distribution as

$$P_t\left(x_t \middle| \underline{s}_0^{t-1}, y_1^{t-1}\right) \triangleq P_{X_t|\underline{S}_0^{t-1}, Y_1^{t-1}}\left(x_t \middle| \underline{s}_0^{t-1}, y_1^{t-1}\right), \quad (13)$$

or equivalently in terms of the channel states (since the input and state sequences are interchangeable) as

$$P_t\left(\underline{s}_t \middle| \underline{s}_0^{t-1}, y_1^{t-1}\right) \triangleq P_{\underline{S}_t|\underline{S}_0^{t-1}, Y_1^{t-1}}\left(\underline{s}_t \middle| \underline{s}_0^{t-1}, y_1^{t-1}\right). \quad (14)$$

III. Directed Information and Feedback Capacity

We start by reviewing the notion of *directed information* [16], [17] between a vector of channel inputs X_1^n and outputs Y_1^n, conditioned on the initial channel state $\underline{s}_0$, as

$$\begin{aligned} I\left(X_1^n \to Y_1^n \left|\underline{s}_0\right.\right) &\triangleq \sum_{t=1}^{n} I\left(\underline{X}_1^t; Y_t \left|Y_1^{t-1}, \underline{s}_0\right.\right) & (15) \\ &= \sum_{t=1}^{n} I\left(\underline{S}_1^t; Y_t \left|Y_1^{t-1}, \underline{s}_0\right.\right) & (16) \\ &= \sum_{t=1}^{n} I\left(\underline{S}_{t-1}^t; Y_t \left|Y_1^{t-1}, \underline{s}_0\right.\right) & (17) \\ &= \sum_{t=1}^{n} \left[h\left(Y_t \middle| Y_1^{t-1}, \underline{s}_0\right) - \frac{1}{2}\log\left(2\pi e \sigma_W^2\right)\right] & (18) \end{aligned}$$

where equality in (16) holds since channel state and input sequences determine each other uniquely, and equalities in (17), (18) come from the channel assumptions, see (11) and (12).

Tatikonda [18] showed that the feedback capacity can be expressed as maximal achievable directed information rate[2]

$$C^{\mathrm{fb}} = \lim_{n\to\infty} \max_{\mathcal{P}} \frac{1}{n} I\left(X_1^n \to Y_1^n \left|\underline{s}_0\right.\right), \quad (19)$$

[1]Since it has been shown [15] that the feedback capacity is a concave function of P, it is not necessary to consider the inequality constraint $\lim_{n\to\infty} \mathrm{E}\left[\sum_{t=1}^{n} X_t^2\right]/n \leq P$.

[2]Note that the limit in (19) exists since we are considering stationary channel input processes.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

where the maximum is taken over all feedback-dependent Gaussian sources defined in (13) or (14)

$$\mathcal{P} \triangleq \{P_t(\underline{s}_t \,|\underline{s}_0^{t-1}, y_1^{t-1}), t = 1, 2, \ldots\}. \quad (20)$$

We shall also consider the finite-horizon (finite block length $n > 0$) version of the optimization problem in (19)

$$C^{\mathrm{fb}(n)} \;=\; \max_{\mathcal{P}} \frac{1}{n} I\left(X_1^n \to Y_1^n \,|\underline{s}_0\right). \quad (21)$$

Correspondingly, the power constraint is $\mathrm{E}\left[\sum_{t=1}^n X_t^2\right] = nP$.

IV. Feedback Capacity Achieving Strategy

In the following analysis, we note that since the initial channel state $\underline{s}_0$ is known according to the channel assumptions in Section II, for notational simplicity, we will not explicitly write the dependence on $\underline{s}_0$ when obvious.

A. Gauss-Markov Sources Achieve the Feedback Capacity

Theorem 1: For the power constrained ISI Gaussian channel, a feedback-dependent Gauss-Markov source

$$\mathcal{P}^{\mathrm{GM}} \triangleq \left\{P_t\left(\underline{s}_t \,|\underline{s}_{t-1}, y_1^{t-1}\right), t = 1, 2, \ldots\right\} \quad (22)$$

achieves the feedback capacity. (proof in Appendix) □

By Theorem 1, without loss of optimality, in the sequel we only consider feedback-dependent *Gauss-Markov* sources as in (22).

B. Optimal Kalman-Bucy Filter for Processing the Feedback

Definition 1: We use $\alpha_t(\cdot)$ as shorthand notation for the posterior pdf of the channel state $\underline{S}_t$, that is

$$\alpha_t(\underline{\mu}) \triangleq P_{\underline{S}_t|\underline{S}_0, Y_1^t}\left(\underline{\mu}\,|\underline{s}_0, y_1^t\right), \quad (23)$$

which is Gaussian due to Gaussian channel inputs. □

For a feedback-dependent Gauss-Markov source $\mathcal{P}^{\mathrm{GM}}$, the functions $\alpha_t(\cdot)$ can be recursively computed as

$$\alpha_t\left(\underline{\mu}\right) = \frac{\int \alpha_{t-1}(\underline{v}) P_t\left(\underline{\mu}|\underline{v}, y_1^{t-1}\right) P_{Y_t|\underline{S}_{t-1},\underline{S}_t}\left(y_t|\underline{v},\underline{\mu}\right) \mathrm{d}\underline{v}}{\iint \alpha_{t-1}(\underline{v}) P_t\left(\underline{u}|\underline{v}, y_1^{t-1}\right) P_{Y_t|\underline{S}_{t-1},\underline{S}_t}\left(y_t|\underline{v},\underline{u}\right) \mathrm{d}\underline{u}\mathrm{d}\underline{v}}. \quad (24)$$

The Gaussian function $\alpha_t(\cdot)$ is completely characterized by the conditional mean $\underline{m}_t$ and conditional covariance matrix $\mathbf{K}_t$

$$\underline{m}_t = \mathrm{E}\left[\underline{S}_t \,|\underline{s}_0, y_1^t\right], \quad (25)$$

$$\mathbf{K}_t = \mathrm{E}\left[\left(\underline{S}_t - \underline{m}_t\right)\left(\underline{S}_t - \underline{m}_t\right)^{\mathrm{T}} |\underline{s}_0, y_1^t\right]. \quad (26)$$

We note that the recursion (24) can be implemented by a Kalman-Bucy filter [19].

Theorem 2: For the power-constrained linear Gaussian channel, the feedback capacity is achieved by a feedback-dependent Gauss-Markov source $\mathcal{P}_\alpha^{\mathrm{GM}}$ defined as

$$\mathcal{P}_\alpha^{\mathrm{GM}} \triangleq \left\{P_t\left(\underline{s}_t \,|\underline{s}_{t-1}, \alpha_{t-1}(\cdot)\right), t = 1, 2, \ldots\right\}, \quad (27)$$

where the Markov transition probability depends only on the posterior distribution function of the channel state $\alpha_t(\cdot)$ instead of all prior channel outputs Y_1^{t-1}. (proof in Appendix) □

Theorem 2 suggests that, for the task of constructing the next signal to be transmitted, all the "knowledge" contained in the vector of prior channel outputs is captured by the posterior distribution $\alpha_{t-1}(\cdot)$ of the channel state.

V. Feedback Capacity Computation

By Theorem 1 and Theorem 2, we only need to consider a feedback-dependent Gauss-Markov source $\mathcal{P}_\alpha^{\mathrm{GM}}$ as defined in (27).

A. Source Parameterization

Without loss of generality, a feedback-dependent Gauss-Markov source $\mathcal{P}_\alpha^{\mathrm{GM}}$ can be expressed as

$$X_t = \underline{d}_t^{\mathrm{T}} \underline{S}_{t-1} + e_t Z_t + g_t, \quad (28)$$

where Z_t is a Gaussian random variable with zero-mean and unit-variance and is independent of Z_1^{t-1}, X_1^{t-1} and Y_1^{t-1}. The coefficients $\underline{d}_t$, e_t and g_t are all dependent on the Gaussian pdf $\alpha_{t-1}(\cdot)$, or alternatively on its mean $\underline{m}_{t-1}$ and covariance matrix $\mathbf{K}_{t-1}$. The set of coefficients $\{\underline{d}_t, e_t, g_t\}$ completely determine the transition probabilities of the feedback-dependent Gauss-Markov source $\mathcal{P}_\alpha^{\mathrm{GM}}$ defined in Theorem 2.

Lemma 1: For the feedback-dependent Gauss-Markov source as parameterized in (28), we have

$$h\left(Y_t \,|\underline{s}_0, y_1^{t-1}\right) - \frac{1}{2}\log\left(2\pi\mathrm{e}\sigma_W^2\right) \quad (29)$$
$$= \frac{1}{2}\log\left(\frac{\sigma_W^2 + \left(\underline{a}+\underline{c}+\underline{d}_t\right)^{\mathrm{T}} \mathbf{K}_{t-1}\left(\underline{a}+\underline{c}+\underline{d}_t\right) + \left(e_t\right)^2}{\sigma_W^2}\right)$$

and

$$\mathrm{E}\left[(X_t)^2 \,|\underline{s}_0, y_1^{t-1}\right] = \left(\underline{d}_t^{\mathrm{T}} \underline{m}_{t-1} + g_t\right)^2 + \underline{d}_t^{\mathrm{T}} \mathbf{K}_{t-1} \underline{d}_t + (e_t)^2 \quad (30)$$

where the values of $\underline{d}_t$, e_t, g_t depend on $\underline{m}_{t-1}$ and $\mathbf{K}_{t-1}$. □

Proof: The first and second order moments of the channel input X_t and output Y_t can be computed as

$$\mathrm{E}\left[X_t \,|\underline{s}_0, y_1^{t-1}\right] = \underline{d}_t^{\mathrm{T}} \underline{m}_{t-1} + g_t \quad (31)$$

$$\mathrm{E}\left[Y_t \,|\underline{s}_0, y_1^{t-1}\right] = \left(\underline{a} + \underline{c} + \underline{d}_t\right)^{\mathrm{T}} \underline{m}_{t-1} + g_t \quad (32)$$

$$\mathrm{E}\left[(X_t)^2 \,|\underline{s}_0, y_1^{t-1}\right] = \left(\underline{d}_t^{\mathrm{T}} \underline{m}_{t-1} + g_t\right)^2 + \underline{d}_t^{\mathrm{T}} \mathbf{K}_{t-1} \underline{d}_t + (e_t)^2 \quad (33)$$

$$\mathrm{E}\left[\left(Y_t - \mathrm{E}\left[Y_t \,|\underline{s}_0, y_1^{t-1}\right]\right)^2 |\underline{s}_0, y_1^{t-1}\right] = \left(\underline{a} + \underline{c} + \underline{d}_t\right)^{\mathrm{T}} \mathbf{K}_{t-1}\left(\underline{a} + \underline{c} + \underline{d}_t\right) + (e_t)^2 + \sigma_W^2 \quad (34)$$

Conditioned on $\underline{s}_0$ and y_1^{t-1}, the variable Y_t has a Gaussian distribution with variance (34), thus we obtain (29). ■

Lemma 2: The parameters of the optimal feedback-dependent Gauss-Markov source must satisfy

$$g_t = -\underline{d}_t^{\mathrm{T}} \underline{m}_{t-1}. \quad (35)$$

Proof: By Lemma 1 and equation (18), the value of g_t does not affect the information rate, but choosing g_t as in (35) minimizes the average input power for given $\underline{d}_t$ and e_t. ■

By Lemma 2, the finite-horizon feedback capacity (21) thus can be expressed as

$$C^{\mathrm{fb}(n)} = \max_{\mathcal{P}_\alpha^{\mathrm{GM}}} \sum_{t=1}^{n} \frac{1}{2n} \log \frac{\sigma_W^2 + \left(\underline{a}+\underline{c}+\underline{d}_t\right)^{\mathrm{T}} \mathbf{K}_{t-1}\left(\underline{a}+\underline{c}+\underline{d}_t\right) + e_t^2}{\sigma_W^2}. \quad (36)$$

B. Feedback Capacity for Stationary Sources

Definition 2 (Stationary sources): A *stationary* feedback-dependent (Gauss-Markov) source is a source that induces stationary channel input and output processes. An *asymptotically* stationary feedback-dependent (Gauss-Markov) source, in its limit as $t \to \infty$, induces stationary channel input and output processes. $\square$

The following analysis holds for both stationary and asymptotically stationary sources.

Lemma 3: For a stationary (or asymptotically stationary) feedback-dependent Gauss-Markov source, the covariance matrix $\mathbf{K}_t$ and source coefficients $\underline{d}_t$ and e_t converge, i.e.,

$$\lim_{t\to\infty} \mathbf{K}_t = \mathbf{K}, \qquad \lim_{t\to\infty} \underline{d}_t = \underline{d}, \qquad \lim_{t\to\infty} e_t = e. \tag{37}$$

Here, the matrix $\mathbf{K}$ satisfies the stationary Kalman-Bucy filter equation (the algebraic Riccati equation)

$$\begin{aligned}\mathbf{K} &= \mathbf{QKQ}^{\mathrm{T}} + \underline{b}\,\underline{b}^{\mathrm{T}} e^2 \qquad (38)\\ &- \frac{\left(\mathbf{QK}\left(\underline{a}+\underline{c}+\underline{d}\right)+\underline{b}\,e^2\right)\left(\mathbf{QK}\left(\underline{a}+\underline{c}+\underline{d}\right)+\underline{b}\,e^2\right)^{\mathrm{T}}}{\left(\underline{a}+\underline{c}+\underline{d}\right)^{\mathrm{T}}\mathbf{K}\left(\underline{a}+\underline{c}+\underline{d}\right)+e^2+\sigma_W^2}\end{aligned}$$

where the matrix $\mathbf{Q}$ is defined as $\mathbf{Q} \stackrel{\triangle}{=} \mathbf{A} + \underline{b}\,\underline{d}^{\mathrm{T}}$. $\square$

Proof: Since the (asymptotically) stationary source induces, in its limit as $t \to \infty$, stationary channel input and output processes, the Kalman-Bucy filter has a steady state, and thus the sequences $\mathbf{K}_t$, $\underline{d}_t$ and e_t converge. The Riccati equation (38) is obtained as the stationary form of the covariance matrix of the Kalman-Bucy filter [20]. ■

Theorem 3 (Feedback capacity): The feedback capacity for stationary sources with the input power constraint P equals

$$C^{\mathrm{fb}} = \max_{\underline{d},e} \frac{1}{2}\log\left(\frac{\sigma_W^2 + \left(\underline{a}+\underline{c}+\underline{d}\right)^{\mathrm{T}}\mathbf{K}\left(\underline{a}+\underline{c}+\underline{d}\right)+e^2}{\sigma_W^2}\right) \tag{39}$$

where the maximization in (39) is taken under constraints

$$\underline{d}^{\mathrm{T}}\mathbf{K}\underline{d} + e^2 = P \tag{40}$$

$$\begin{aligned}\mathbf{K} &= \mathbf{QKQ}^{\mathrm{T}} + \underline{b}\,\underline{b}^{\mathrm{T}} e^2 \qquad (41)\\ &- \frac{\left(\mathbf{QK}\left(\underline{a}+\underline{c}+\underline{d}\right)+\underline{b}\,e^2\right)\left(\mathbf{QK}\left(\underline{a}+\underline{c}+\underline{d}\right)+\underline{b}\,e^2\right)^{\mathrm{T}}}{\left(\underline{a}+\underline{c}+\underline{d}\right)^{\mathrm{T}}\mathbf{K}\left(\underline{a}+\underline{c}+\underline{d}\right)+e^2+\sigma_W^2}.\end{aligned}$$

The matrix $\mathbf{Q}$ is defined as $\mathbf{Q} \stackrel{\triangle}{=} \mathbf{A} + \underline{b}\,\underline{d}^{\mathrm{T}}$, and the matrix $\mathbf{K}$ is constrained to be non-negative definite. $\square$

Proof: By Lemma 3, for any (asymptotically) stationary Gauss-Markov source, the sequences $\mathbf{K}_t$, $\underline{d}_t$ and e_t converge as $t \to \infty$, so (36) turns into (39) as $n \to \infty$. Constraint (41) is the algebraic Riccati equation (38). Constraint (40) is obtained by setting $\mathrm{E}\left[(X_t)^2 \left| \underline{s}_0, y_1^{t-1}\right.\right] = P$ in (30), and subsequently utilizing Lemmas 2 and 3. ■

In general, the optimization problem in Theorem 3 can be easily solved numerically. For first-order channels we next give an analytic solution.

C. Feedback Capacity of First-Order Channels

First-order channels are channels for which both the vectors $\underline{a} = a$ and $\underline{c} = c$ in (7) are scalars, and $a+c \neq 0$. Consequently, the channel state covariance matrix is also a scalar $\mathbf{K} = K$.

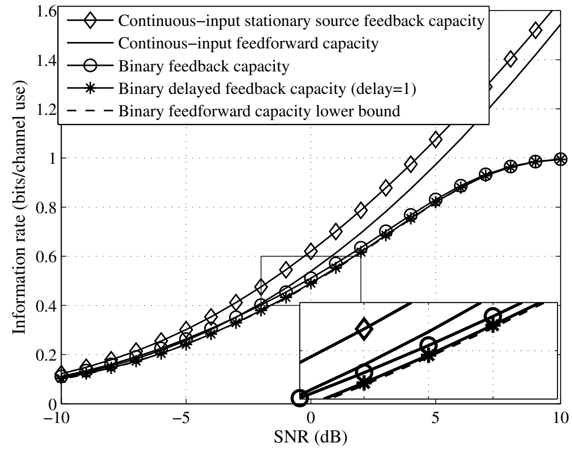

Fig. 2. Channel Capacity Curves for the dicode channel.

Theorem 4: For a first-order Gaussian noise channel defined by coefficients a and c, the feedback capacity of stationary sources under the average input power constraint equals

$$C^{\mathrm{fb}} = \frac{1}{2}\log\left(1 + \frac{(1+\eta)^2 P}{\sigma_W^2}\right). \tag{42}$$

Here, the parameter η is the largest positive root of

$$\frac{P}{\sigma_W^2}\eta^4 + 2\frac{P}{\sigma_W^2}\eta^3 + \left(\frac{P}{\sigma_W^2} + 1 - a^2\right)\eta^2 - 2a(a+c)\eta = (a+c)^2. \tag{43}$$

The feedback-dependent Gauss-Markov source that solves (42) has coefficients $d = \frac{(a+c)}{\eta}$ and $e = 0$. (proof in Appendix) $\square$

When the noise is a first-order autoregressive process, formula (42) becomes the same as Butman's achievable rate [9].

VI. Capacity plots for the dicode channel

Fig. 2 depicts the channel capacity curves for the dicode $1 - D$ channel. The feedback capacity of continuous-input stationary sources is compared to the waterfilling feedforward capacity [6], the binary-input feedback capacity [5], the binary-input 1-tap delayed feedback capacity [5], and the maximal feed-forward information rate achieved by a binary-input 6-th order Markov source [4].

VII. Conclusion

In this paper, we derived the feedback capacity of power-constrained stationary sources over ISI channels with ARMA Gaussian noise. We first showed that the feedback-capacity-achieving source can be a feedback-dependent Gauss-Markov source. Then we showed that the Kalman-Bucy filter is optimal for processing the feedback. The feedback capacity is expressible as an optimization problem with constraints on the conditional state covariance matrix of the Kalman-Bucy filter. The optimization problem can be solved using numerical non-linear programming techniques. For first-order channels (such as the dicode channel), we derived the closed-form solution.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

ACKNOWLEDGMENT

We thank an anonymous reviewer, who pointed out the Ph.D. thesis [21] (which became available after the submission of this work) that seems to have proved that stationary sources achieve the feedback capacity. The thesis rederives formulas (39) and (42). It further shows that in (39), the constant e should be $e = 0$, which we were able to show only for the first-order channel in (42).

APPENDIX

A. Proof of Theorem 1

Let $\mathcal{P}_1$ be any valid feedback-dependent Gaussian source distribution (not necessarily Markov) defined as

$$\mathcal{P}_1 \stackrel{\triangle}{=} \left\{ P_t\left(\underline{s}_t \left| \underline{s}_0^{t-1}, y_1^{t-1}\right.\right), t = 1, 2, \cdots \right\}. \tag{44}$$

From $\mathcal{P}_1$, we construct a Markov (not necessarily stationary) source distribution $\mathcal{P}_2$ as

$$\mathcal{P}_2 = \left\{ Q_t\left(\underline{s}_t \left| \underline{s}_{t-1}, y_1^{t-1}\right.\right), t = 1, 2, \cdots \right\}. \tag{45}$$

where the functions $Q_t\left(\underline{s}_t \left| \underline{s}_{t-1}, y_1^{t-1}\right.\right)$ are defined as the conditional marginal pdf's computed from $\mathcal{P}_1$

$$Q_t\left(\underline{s}_t \left| \underline{s}_{t-1}, y_1^{t-1}\right.\right) \stackrel{\triangle}{=} P^{(\mathcal{P}_1)}_{\underline{S}_t|\underline{S}_{t-1}, Y_1^{t-1}}\left(\underline{s}_t \left| \underline{s}_{t-1}, y_1^{t-1}\right.\right) \tag{46}$$

We next show by induction that the the sources $\mathcal{P}_1$ and $\mathcal{P}_2$ induce the same distribution of $\underline{S}_{t-1}^t$ and Y_1^t, i.e.,

$$P^{(\mathcal{P}_1)}_{\underline{S}_{t-1}^t, Y_1^t|\underline{S}_0}\left(\underline{s}_{t-1}^t, y_1^t \left| \underline{s}_0\right.\right) = P^{(\mathcal{P}_2)}_{\underline{S}_{t-1}^t, Y_1^t|\underline{S}_0}\left(\underline{s}_{t-1}^t, y_1^t \left| \underline{s}_0\right.\right). \tag{47}$$

For $t = 1$, by the definition of source $\mathcal{P}_2$ we have

$$\begin{aligned} P^{(\mathcal{P}_2)}_{\underline{S}_1, Y_1|\underline{S}_0}\left(\underline{s}_1, y_1 \left| \underline{s}_0\right.\right) &= Q_1(\underline{s}_1 \left| \underline{s}_0\right.) P_{Y_1|\underline{S}_0^1}\left(y_1 \left| \underline{s}_0^1\right.\right) &(48)\\ &= P_1(\underline{s}_1 \left| \underline{s}_0\right.) P_{Y_1|\underline{S}_0^1}\left(y_1 \left| \underline{s}_0^1\right.\right) &(49)\\ &= P^{(\mathcal{P}_1)}_{\underline{S}_1, Y_1|\underline{S}_0}\left(\underline{s}_1, y_1 \left| \underline{s}_0\right.\right). &(50) \end{aligned}$$

Since $\underline{s}_0$ is known, this directly implies

$$P^{(\mathcal{P}_2)}_{\underline{S}_0^1, Y_1|\underline{S}_0}\left(\underline{s}_0^1, y_1 \left| \underline{s}_0\right.\right) = P^{(\mathcal{P}_1)}_{\underline{S}_0^1, Y_1|\underline{S}_0}\left(\underline{s}_0^1, y_1 \left| \underline{s}_0\right.\right). \tag{51}$$

Now, assume that the equality (47) holds for up to time $t-1$, where $t > 1$, particularly,

$$\begin{aligned} P^{(\mathcal{P}_2)}_{\underline{S}_{t-2}^{t-1}, Y^{t-1}|\underline{S}_0}\left(\underline{s}_{t-2}^{t-1}, y_1^{t-1} | \underline{s}_0\right) &= P^{(\mathcal{P}_1)}_{\underline{S}_{t-2}^{t-1}, Y^{t-1}|\underline{S}_0}\left(\underline{s}_{t-2}^{t-1}, y_1^{t-1} | \underline{s}_0\right) &(52)\\ &= \int \prod_{\tau=1}^{t-1} P_\tau\left(\underline{s}_\tau \left| \underline{s}_0^{\tau-1}, y_1^{\tau-1}\right.\right) P_{Y_\tau|\underline{S}_{\tau-1}^\tau}\left(y_\tau \left| \underline{s}_{\tau-1}^\tau\right.\right) \mathrm{d}\underline{s}_1^{t-3}. &(53) \end{aligned}$$

The induction step for time t is simply shown as follows

$$\begin{aligned} &P^{(\mathcal{P}_2)}_{\underline{S}_{t-1}^t, Y_1^t|\underline{S}_0}\left(\underline{s}_{t-1}^t, y_1^t \left| \underline{s}_0\right.\right)\\ &= Q_t\left(\underline{s}_t \left| \underline{s}_{t-1}, y_1^{t-1}\right.\right) \times P_{Y_t|\underline{S}_{t-1}^t}\left(y_t \left| \underline{s}_{t-1}^t\right.\right) \times\\ &\qquad \int P^{(\mathcal{P}_2)}_{\underline{S}_{t-2}^{t-1}, Y_1^{t-1}|\underline{S}_0}\left(\underline{s}_{t-2}^{t-1}, y_1^{t-1} \left| \underline{s}_0\right.\right) \mathrm{d}\underline{s}_{t-2} &(54)\\ &\stackrel{(a)}{=} \frac{\int\left[\prod_{\tau=1}^{t-1} P_\tau\left(\underline{s}_\tau | \underline{s}_0^{\tau-1}, y_1^{\tau-1}\right) f_{Y_\tau|\underline{S}_{\tau-1}^\tau}\left(y_\tau \left| \underline{s}_{\tau-1}^\tau\right.\right)\right] P_t\left(\underline{s}_t \left| \underline{s}_0^{t-1}, y_1^{t-1}\right.\right) \mathrm{d}\underline{s}_1^{t-2}}{\int\left[\prod_{\tau=1}^{t-1} P_\tau\left(\underline{s}_\tau | \underline{s}_0^{\tau-1}, y_1^{\tau-1}\right) f_{Y_\tau|\underline{S}_{\tau-1}^\tau}\left(y_\tau \left| \underline{s}_{\tau-1}^\tau\right.\right)\right] \mathrm{d}\underline{s}_1^{t-2}} \times\\ &\quad P_{Y_t|\underline{S}_{t-1}^t}\left(y_t \left| \underline{s}_{t-1}^t\right.\right) \times \int\left[\prod_{\tau=1}^{t-1} P_\tau\left(\underline{s}_\tau | \underline{s}_0^{\tau-1}, y_1^{\tau-1}\right) f_{Y_\tau|\underline{S}_{\tau-1}^\tau}\left(y_\tau \left| \underline{s}_{\tau-1}^\tau\right.\right)\right] \mathrm{d}\underline{s}_1^{t-2} &(55)\\ &\stackrel{(b)}{=} \int \prod_{\tau=1}^{t} P_\tau\left(\underline{s}_\tau \left| \underline{s}_0^{\tau-1}, y_1^{\tau-1}\right.\right) P_{Y_\tau|\underline{S}_{\tau-1}^\tau}\left(y_\tau \left| \underline{s}_{\tau-1}^\tau\right.\right) \mathrm{d}\underline{s}_1^{t-2} &(56)\\ &= P^{(\mathcal{P}_1)}_{\underline{S}_{t-1}^t, Y_1^t|\underline{S}_0}\left(\underline{s}_{t-1}^t, y_1^t \left| \underline{s}_0\right.\right), &(57) \end{aligned}$$

where (a) is the result of expanding the definition in (46) for source $\mathcal{P}_2$ and the induction assumption (53) using the Bayes rule and substituting them into (54), and (b) is obtained by simplifying the expression in (55).

Thus, we have shown that the channel states $\underline{S}_{t-1}^t$ and outputs Y_1^t induced by sources $\mathcal{P}_1$ and $\mathcal{P}_2$ have the same distribution. It is therefore clear that the non-Markov source $\mathcal{P}_1$ and Markov source $\mathcal{P}_2$ induce the same directed information according to equality (17).

B. Proof of Theorem 2

Suppose that two different feedback vectors $\tilde{y}_1^{t-1}$ and y_1^{t-1} ($\tilde{y}_1^{t-1} \neq y_1^{t-1}$) induce the same posterior channel state pdf $\alpha_{t-1}(\cdot)$, i.e., for any possible state value $\underline{s}_{t-1} = \underline{\mu}$ we have

$$P_{\underline{S}_{t-1}|\underline{S}_0, Y_1^{t-1}}\left(\underline{\mu} \left| \underline{s}_0, \tilde{y}_1^{t-1}\right.\right) = P_{\underline{S}_{t-1}|\underline{S}_0, Y_1^{t-1}}\left(\underline{\mu} \left| \underline{s}_0, y_1^{t-1}\right.\right). \tag{58}$$

Now consider two distributions for the source S_τ, for $\tau \geq t$, the first distribution conditioned on y_1^{t-1}, and the second conditioned on $\tilde{y}_1^{t-1}$. If we let these two distributions be equal to each other for $\tau \geq t$, that is, if

$$\begin{aligned} &\left\{ P_\tau\left(\underline{s}_\tau \left| \underline{s}_{\tau-1}, \tilde{y}_1^{t-1}, y_t^{\tau-1}\right.\right), \tau \geq t \right\}\\ &\qquad = \left\{ P_\tau\left(\underline{s}_\tau \left| \underline{s}_{\tau-1}, y_1^{t-1}, y_t^{\tau-1}\right.\right), \tau \geq t \right\}, \end{aligned} \tag{59}$$

then we have for any $k \geq t$

$$\begin{aligned} &P_{Y_t^k, \underline{S}_{t-1}^k|\underline{S}_0, Y_1^{t-1}}\left(y_t^k, \underline{s}_{t-1}^k \left| \underline{s}_0, \tilde{y}_1^{t-1}\right.\right)\\ &\quad = \alpha_{t-1}(\underline{s}_{t-1}) \prod_{\tau=t}^{k} P_\tau\left(\underline{s}_\tau | \underline{s}_{\tau-1}, y_1^{\tau-1}\right) P_{Y_\tau|\underline{S}_{\tau-1}^\tau}\left(y_\tau | \underline{s}_{\tau-1}^\tau\right)\\ &\quad = P_{Y_t^k, \underline{S}_{t-1}^k|\underline{S}_0, Y_1^{t-1}}\left(y_t^k, \underline{s}_{t-1}^k \left| \underline{s}_0, y_1^{t-1}\right.\right). \end{aligned} \tag{60}$$

This shows that for any $k \geq t$ the entropies are equal

$$h\left(Y_t^k \left| \underline{s}_0, \tilde{y}_1^{t-1}\right.\right) = h\left(Y_t^k \left| \underline{s}_0, y_1^{t-1}\right.\right), \tag{61}$$

and for any $\tau \geq t$ the powers are equal

$$\mathrm{E}\left[(X_\tau)^2 \left| \underline{s}_0, \tilde{y}_1^{t-1}\right.\right] = \mathrm{E}\left[(X_\tau)^2 \left| \underline{s}_0, y_1^{t-1}\right.\right]. \tag{62}$$

Therefore, the optimal source distribution for time $\tau \geq t$ when y_1^{t-1} is the feedback vector, must also be optimal when $\tilde{y}_1^{t-1}$ is the feedback vector, and vice versa. Since time t is arbitrary, we conclude that, for any $t > 0$, the function $\alpha_{t-1}(\cdot)$ extracts from y_1^{t-1} all that is necessary for formulating the optimal source distribution functions $P_t\left(\underline{s}_t \left| \underline{s}_{t-1}, y_1^{t-1}\right.\right)$.

C. Proof of Theorem 4

By Theorem 3, the feedback capacity C^{fb} is determined by solving the following optimization problem

$$\max_{d,e} \frac{1}{2} \log\left(\frac{\sigma_W^2 + (a+c+d)^2 K + e^2}{\sigma_W^2}\right), \tag{63}$$

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

under the following constraints

$$d^2K + e^2 = P, \tag{64}$$

$$K^2(a+c+d)^2 - e^2\sigma_W^2 + K\left(e^2 + \sigma_W^2 - \sigma_W^2(a+d)^2 - c^2e^2\right) = 0. \tag{65}$$

Substituting the constraint (64) into the objective function (63) and noting that the function $\log(\cdot)$ is strictly monotonic, we obtain that the optimization problem (63) is equivalent to the following optimization problem

$$\max_{d,e}\left[(a+c+d)^2 K - d^2K\right]. \tag{66}$$

By zeroing the derivative of the Lagrangian (with auxiliary Lagrange variables λ and ρ) for the optimization problem (66), (64) and (65), we obtain the following necessary conditions:

$$2K\left(a+c+\lambda d+\rho\left(K(a+c+d) - \sigma_W^2\,(a+d)\right)\right) = 0, \tag{67}$$

$$2e\left[\lambda + \rho\left(K\left(1-c^2\right) - \sigma_W^2\right)\right] = 0, \tag{68}$$

$$(a+c+d)^2 - d^2 + \lambda d^2 + \rho\Big(2K(a+c+d)^2 + e^2 + \sigma_W^2 - \sigma_W^2(a+d)^2 - c^2e^2\Big) = 0, \tag{69}$$

$$d^2K + e^2 - P = 0, \tag{70}$$

$$K^2(a+c+d)^2 + K\left(e^2 + \sigma_W^2 - \sigma_W^2(a+d)^2 - c^2e^2\right) - e^2\sigma_W^2 = 0. \tag{71}$$

We first prove that $e = 0$ is necessary for optimality. If $e \neq 0$, then equation (68) is substituted by

$$\lambda + \rho\left(K\left(1-c^2\right) - \sigma_W^2\right) = 0. \tag{72}$$

1) One possible form of solution induced by (72) is

$$K = -\frac{-c^2P + c^2\sigma_W^2 + 2ac\sigma_W^2}{(a+c)^2c^2}, \tag{73}$$

$$e^2 = \frac{a^2\sigma^4 + c^2P\sigma_W^2}{-c^2P + c^2\sigma_W^2 + 2ac\sigma_W^2}. \tag{74}$$

By (73)-(74), K and e^2 cannot both be positive. On the other hand, we must have $e^2 > 0$ and $K > 0$ (since K is a variance), so (73) and (74) cannot be the solution.

2) The only other possible solution induced by (72) is

$$d = \frac{(1+ac)(a+c)^2\sigma_W^2}{\sigma_W^2\left(a^2c^3 - 2a^2c - c - 2a\right) + P\left(c^5 - 2c^3 + c\right)}, \tag{75}$$

$$K = \frac{\sigma_W^2\left(a^2c^3 - 2a^2c - c - 2a\right) + P\left(c^5 - 2c^3 + c\right)}{(a+c)^2\left(c^2-1\right)c}. \tag{76}$$

Now, if we substitute (75) and (76) into the objective function (66), we get a strictly negative value

$$(a+c+d)^2K - d^2K = \frac{P\left(c^2-1\right)^2 + \sigma_W^2\left(ac-1\right)^2}{c^2-1} < 0. \tag{77}$$

Since we will compute a positive objective function (66) when the equality condition (72) is replaced by $e = 0$, this negative value (77) cannot be the maximum of (66).

Therefore, (72) cannot hold and hence (68) implies $e = 0$.

For $e = 0$, equations (67), (69), (70) and (71) can be solved and the solution takes the form $K = P/d^2$, where d satisfies the following 4-th order polynomial equation[3]

$$R(d) = \sigma_W^2 d^4 + 2a\sigma_W^2 d^3 + \left(-P - \sigma_W^2 + a^2\sigma_W^2\right)d^2 - 2P\left(a+c\right)d - P\left(a+c\right)^2 = 0. \tag{78}$$

We let $d = (a+c)/\eta$, substitute it into (78) and thus solve (63) to get (43) and (42), which further implies that the optimal value of η should be the maximal positive root of (43).

References

[1] D. Arnold and H.-A. Loeliger, “On the information rate of binary-input channels with memory,” in *Proc. IEEE ICC 2001*, (Helsinki, Finland), pp. 2692–2695, June 2001.

[2] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, “On the achievable information rates of finite state ISI channels,” in *Proc. IEEE GLOBECOM 2001*, (San Antonio, Texas), pp. 2992–2996, Nov 2001.

[3] V. Sharma and S. K. Singh, “Entropy and channel capacity in the regenerative setup with applications to Markov channels,” in *Proc. IEEE ISIT 2001*, (Washington, DC), p. 283, June 2001.

[4] A. Kavčić, “On the capacity of Markov sources over noisy channels,” in *Proceedings IEEE GLOBECOM 2001*, (San Antonio, Texas), pp. 2997–3001, Nov 2001.

[5] S. Yang, A. Kavčić, and S. Tatikonda, “The feedback capacity of finite-state machine channels,” *IEEE Transactions on Information Theory*, vol. 51, pp. 799–810, March 2005.

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, Inc., 1968.

[7] C. E. Shannon, “The zero error capacity of a noisy channel,” *IRE Trans. on Inform. Theory*, vol. IT-2, pp. 112–124, Sept. 1956.

[8] J. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback–i: No bandwidth constraint,” *IEEE Transactions on Information Theory*, vol. 14, pp. 172–182, Apr. 1966.

[9] S. Butman, “A general formulation of linear feedback communications systems with solutions,” *IEEE Trans. Inform. Theory*, vol. 15, pp. 392–400, May 1969.

[10] T. M. Cover and S. Pombra, “Gaussian feedback capacity,” *IEEE Trans. Inform. Theory*, vol. 35, pp. 37–43, Jan 1989.

[11] M. Pinsker, “Talk delivered at the Soviet Information Theory Meeting.” unpublished, 1969.

[12] P. Ebert, “The capacity of the Gaussian channel with feedback,” *Bell Syst. Tech. J.*, vol. 49, pp. 1705–1712, Oct 1970.

[13] L. H. Ozarow, “Upper bounds on the capacity of Gaussian channels with feedback,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 156–161, Jan 1990.

[14] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewoods Cliffs, NJ: Prentice Hall, 1989.

[15] K. Yanagi, H. W. Chen, and J. W. Yu, “Operator inequality and its application to capacity of Gaussian channel,” *Taiwanese Journal of Mathematics*, vol. 4, Sept. 2000.

[16] H. Marko, “The bidirectional communication theory–a generalization of information theory,” *IEEE Trans. Commun.*, vol. 21, pp. 1345–1351, Dec 1973.

[17] J. L. Massey, “Causality, feedback and directed information,” *Proc. 1990 Intl. Symp. on Info. Th. and its Applications*, pp. 27–30, Nov 1990.

[18] S. Tatikonda, *Control Under Communications Constraints*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, Sept. 2000. available at http://pantheon.yale.edu/~sct29/thesis.html.

[19] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction theory,” *ASME J. Basic Engr.*, Mar 1961.

[20] D. P. Bertsekas, *Dynamic Programming and Optimal Control (2nd Edition)*. Belmont, MA: Athena Scientific, 2001.

[21] Y.-H. Kim, *Gaussian Feedback Capacity*. PhD thesis, Stanford University, Palo Alto, CA, June 2006.

[3] Note that the first and last coefficients of $R(d)$ satisfy $\sigma_W^2 > 0$ and $-P\left(a+c\right)^2 < 0$. Thus, the polynomial $R(d)$ must have at least one positive and one negative root, and we can always select d such that $d(a+c) > 0$ and get a positive objective function to exceed the value in (77).