Control-theoretic Approach to Communication with Feedback

Ehsan Ardestanizadeh, Member, IEEE and Massimo Franceschetti, Member, IEEE

Abstract—Feedback communication is studied from a control-theoretic perspective, mapping the communication problem to a control problem in which the control signal is received through the same noisy channel as in the communication problem, and the (nonlinear and time-varying) dynamics of the system determine a subclass of encoders available at the transmitter. The MSE exponent is defined to be the exponential decay rate of the mean square decoding error and is used for analysis of the reliable rate of communication. A sufficient condition is provided under which the MMSE capacity, the supremum achievable MSE exponent, is equal to the information-theoretic capacity, the supremum achievable rate. For the special class of stationary Gaussian channels and linear time-invariant systems, a simple application of Bode’s integral formula shows that the feedback capacity, recently characterized by Kim, is equal to the maximum instability that can be tolerated by any linear controller under a given power constraint. Finally, the control mapping is generalized to the \( N \)-sender AWGN multiple access channel. It is shown that Kramer’s code for this channel, which is known to be sum rate optimal in the class of generalized linear feedback codes, can be obtained by solving a linear quadratic Gaussian control problem.

I. INTRODUCTION

Feedback loops are central to many engineering systems. Their study naturally falls at the intersection between communication and control theories. However, the information-theoretic approach and the control-theoretic one have often evolved in isolation, separated by almost philosophical differences. In this paper we attempt one step at bridging the gap, showing how tools from both disciplines can be applied to communication and control theories. However, the information-theoretic approach and the control-theoretic one have often evolved in isolation, separated by almost philosophical differences. In this paper we attempt one step at bridging the gap, showing how tools from both disciplines can be applied to study fundamental limits of feedback systems and to design efficient codes for communication in the presence of feedback.

Consider the feedback communication problem over an arbitrary point-to-point channel depicted in Fig 1a. The encoder, which has access to the channel outputs causally and noislessly, wishes to communicate a continuous message \( M \) to the decoder through \( n \) channel uses. At the end of the transmissions, the decoder forms an estimate \( \hat{M} \) based on the received channel outputs, and the mean square error (MSE) of the estimate \( \hat{M} \) represents the performance metric of the communication.

We map this communication problem to the general (nonlinear and time-varying) controlled dynamical system depicted in Fig. 1b, in which the initial state of the system corresponds to the message \( M \), and the control actions—received through the same noisy channel as in the communication problem—correspond to the transmitted signals by the encoder. In this representation, the set of controllers for a given system corresponds to a subclass of encoders for the communication problem. In fact, the system can be viewed as a filter which determines the information pattern [1], on which the transmitted signals (actions) by the encoder (controller) can depend upon. A similar mapping for the special case of linear time-invariant (LTI) systems and controllers was first presented in [2].

We study three different channel models. First, we consider a general point-to-point channel. The MSE exponent is defined as the exponential decay rate of the MSE with respect to the block length \( n \), and the (feedback) minimum mean square error (MMSE) capacity is defined as the supremum of all achievable MSE exponents with feedback. It is not hard to show that the (feedback) MMSE capacity is upper bounded by the information-theoretic (feedback) capacity, the supremum of all achievable rates at which reliable communication is possible. We present a sufficient condition, under which the information-theoretic capacity coincides with the MMSE capacity. Moreover, we show that the MSE exponent can be a useful tool for analysis of achievable rates. These results provide a step towards the understanding of the connection between estimation and information theory.

Second, we focus on the stationary Gaussian channel with feedback, the capacity of which was recently characterized by Kim [3]. Applying the discrete extension of Bode’s result [4] (cf. [5], [6]), we observe that the capacity of the stationary Gaussian channel with feedback under power constraint \( P \) is equal to the maximum instability of an LTI system which can be tolerated by any linear controller with power at most \( P \), acting over the same stationary Gaussian channel. This follows almost immediately from the previous results [2], [3] and provides a step towards the understanding of the connection between stabilizability over some noisy channel and the capacity of that channel.

Finally, we consider the \( N \)-sender additive white Gaussian noise (AWGN) multiple access channel (MAC) with feedback depicted in Fig. 3a. We show that the linear code proposed by Kramer [7], which is known to be optimal among the class of generalized linear feedback codes [8], can be obtained as the optimal solution of a linear quadratic Gaussian (LQG) problem given by a linear time-invariant (LTI) system controlled over a point-to-point AWGN channel where the asymptotic cost is the average power of the controller. These results provide a step towards the understanding of how control tools can be used to design codes for communication.

We now wish to place our results in the context of the related literature. The results on the \( N \)-sender AWGN-MAC...
generalize the previous ones of Elia [2]. He considers a 2-sender AWGN-MAC and maps it to a control problem also over a 2-sender MAC. His results recover Ozarow’s code [9]—a special case of Kramer’s code—using the technique of Youla parametrization. In contrast, our reduction is to a control problem over a single point-to-point channel for any \( N \geq 2 \) number of senders, our analysis is based on the theory of LQG control, and we recover the more general code of Kramer’s.

The connection between the MMSE and capacity has been investigated extensively and from different perspectives in the literature. For example, in a classic paper Duncan [10] expresses the mutual information between a continuous random process and its noisy version corrupted by white noise, in terms of the causal MMSE. More recently, Forney [11] explains the role of the MMSE in the context of capacity achieving lattice codes over AWGN channels. Guo et al. [12] and Zakai [13] showed that for a discrete random vector observed through an AWGN channel, the derivative of the mutual information between input and output sequences with respect to the signal-to-noise ratio (SNR), is half the (noncausal) MMSE. We point out that these authors study the average MMSE of a vector observed over a noisy channel without feedback as a function of the SNR. In contrast, we consider the estimation of a single random variable (message), given the observation of a whole block of length \( n \), and we look at the exponential decay rate of the MMSE with \( n \), at fixed SNR. Of more relevance to us is the recent work of Liu and Elia [14], who study linear codes over Gaussian channels obtained using a Kalman filter (KF) approach. For this class of codes, they show that the decay rate of the MMSE equals the mutual information between the message and the output sequence. In contrast, our results for the MSE exponent are derived based on an information-theoretic approach and hold for all codes over general channels.

Additional works in the literature revealed connections between control theory and information theory. We distinguish between those who use information theory to study control systems and those who use control theory to study communication systems. Within the first group, Mitter and Newton [15, 16] studied estimation and filtering in terms of information and entropy flows. Bode-like fundamental limitations in controlled systems have been analyzed with success from an information-theoretic perspective in [17, 18, 19, 20, 21]. Several works focused on connections between stabilizability over a communication channel and the capacity of that channel [22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. They consider a finite-capacity communication channel either without noise or with a specific noise distribution (e.g., i.i.d. erasures, AWGN, first-order moving average Gaussian noise), and derive data-rate theorems quantifying how much rate is needed to construct a stabilizing quantizer-controller pair.

Within the second group, Elia [2] was the first to map linear codes for additive white Gaussian noise channels to an LTI system controlled over an AWGN channel. Subsequently, Wu et al. [32] studied the Gaussian interference channel in terms of estimation and control. Tatikonda and Mitter [33] used a Markov decision problem (MDP) formulation to study the capacity of Markov channels with feedback, and recently Coleman [34] considered the posterior matching scheme [35] for feedback communication from a stochastic control perspective. Finally, we also refer the reader to [36], [37], [38] for good overviews of this interdisciplinary field and for additional references.

The rest of the paper is organized as follows. Section II presents the definitions and the mapping between the feedback communication and the control problem. Section III provides the connection between the achievable MSE exponent and rate for a general point-to-point channel. The point-to-point stationary Gaussian channel and the AWGN multiple access channel are considered in Section IV and Section V, respectively. Finally, Section VI concludes the paper.

**Notation:** A random variable is denoted by an upper case letter (e.g. \( X, Y, Z \)) and its realization is denoted by a lower case letter (e.g. \( x, y, z \)). Similarly, a random column vector and its realization are denoted by bold face symbols (e.g. \( \mathbf{X} \) and \( \mathbf{x} \), respectively). Uppercase letters (e.g. \( A, B, C \)) also denote matrices, which can be differentiated from a random variable based on the context. The \((i,j)\)-th element of \( A \) is denoted by \( A_{ij} \), and notation \( AT \) and \( A' \) denote the transpose and complex transpose of matrix \( A \), respectively. We use the following short notation for covariance matrices
\[
K_{XY} := E(\mathbf{XY}') - E(\mathbf{X})E(\mathbf{Y}') \quad \text{and} \quad K_X := K_{XX}.
\]

**II. DEFINITIONS AND CONTROL APPROACH**

Consider the communication problem depicted in Fig. 1a, where the sender wishes to convey a message \( M \) to the receiver through \( n \) uses of a stochastic channel\(^1\),
\[
Y_i = h_i(X_i, Z_i), \quad i = 1, \ldots, n.
\]

where \( X_i \in \mathcal{X} \) and \( Y_i \in \mathcal{Y} \) denote the input and output of the channel, respectively, and \( Z_i \in \mathcal{Z} \) denote the noise at time \( i \). The set of mappings \( h_i : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y} \), \( i = 1, \ldots, n \), and the distribution of the noise sequence \( \{Z_i\}_{i=1}^n \) determine the channel. The noise process \( \{Z_i\} \) is assumed to be independent of the message \( M \). The channel is called memoryless if \( \{Z_i\} \) is independently and identically distributed (i.i.d), in which case, the channel can be described by \( p(y_i|x_i) \).

We assume that the output symbols are causally fed back to the sender and the transmitted symbol \( X_i \) at time \( i \) can depend on both the previous channel output sequence \( Y_{i-1} := (Y_1, Y_2, \ldots, Y_{i-1}) \) and the message \( M \).

**Definition 1:** A (feedback) \( n \)-code consists of

1) a continuous message \( M \in (0, 1) \) in the unit interval\(^2\)
2) an encoder that assigns a symbol
\[
X_i = f_i(M, Y_i^{i-1})
\]

to the message \( M \) and the previous channel output sequence \( Y_{i-1} \) for each \( i \in \{1, \ldots, n\} \), and
3) a decoder that assigns an estimate \( \hat{M} \in (0, 1) \),
\[
\hat{M} = \phi(Y^n)
\]

\(^1\)The presented results can be immediately generalized to any causal channel of the form \( Y_i = h_i(X_i, Z_i) \).
\(^2\)The choice of the unit interval is for simplicity and the results can be immediately extended to any interval.
to each sequence $Y^n = (Y_1, \ldots, Y_n)$. The encoding and decoding mappings, $\{f_i\}_{i=1}^n$ and $\phi$, are determined prior to the communication and are known to both the transmitter and the receiver. Let the message $M \sim \text{Unif}(0, 1)$ be a random variable uniformly distributed over the unit interval\(^3\). As the performance measure of the communication, we consider the MSE,$$
abla(n) := \mathbb{E} \left( (M - \phi(Y^n))^2 \right).$$where the expectation is with respect to randomness of both the message and the channel. In Section III, we show that the MSE is related to the achievable rate which is the information-theoretic performance measure for communication systems (see Lemma 2). Note that the decoder does not affect the joint distribution of $(M, X^n, Y^n, Z^n)$ and simply estimates the message at the end of the block. Hence, without loss of generality, we pick the optimal decoder, namely, the MMSE estimator of the message given $Y^n$, and we call an encoder optimal if it minimizes $\nabla(n)$. Let
\[
E(n) := \frac{1}{2n} \log \nabla(n)
\]
be the exponential decay rate of the MSE with respect to $n$.

Note that the MSE $\nabla(n)$ is different form the traditional distortion considered in the rate-distortion function [39]. The MSE $\nabla(n)$ is defined based on the error of the estimate for a “single” random variable $M$ after $n$ transmissions over a channel (channel coding set-up) while the rate-distortion function is defined in the source coding set-up based on a “sequence” of random variables $X^n = (X_1, \ldots, X_n)$ as the “average” of the symbol by symbol distortion $d(X^n, \hat{X}^n) = 1/n \sum_{i=1}^n d(X_i, \hat{X}_i)$ for a given distortion measure $d(\cdot, \cdot)$.

Definition 2: MSE exponent $E$ is called achievable (with feedback) if there exists a sequence of $n$-codes such that
\[
\liminf_{n \to \infty} E(n) \geq E.
\]
The MMSE capacity $C_{MMSE}$ is the supremum of all achievable MSE exponents.

A. Information-Theoretic Definitions

Next, we review basic information-theoretic definitions [39] and their connection to the MSE exponent is considered in Section III.

Definition 3: A $(2^n R, n)$ (feedback) code consists of
1) a discrete message $M \in \{1, \ldots, 2^n R\}$,
2) an encoder that assigns a symbol $X_i(M, Y_i^{-1})$ to the message $M$ and the previous channel output sequence $Y_i^{-1}$ for each $i \in \{1, \ldots, n\}$, and
3) a decoder that assigns an index $\hat{M} \in \{1, \ldots, 2^n R\}$ to each sequence $Y^n := (Y_1, \ldots, Y_n)$.

Note that the message set depends on the block length $n$ whereas in Definition 1 the message set was independent of $n$. Let $M \sim \text{Unif} \{1, \ldots, 2^n R\}$ and the probability of error be defined as $P_e(n) = P(M \neq \hat{M})$.

Definition 4: A rate $R$ is called achievable (with feedback) if there exists a sequence of $(2^n R, n)$ codes such that $\lim_{n \to \infty} P_e(n) = 0$.

The error exponent $E_e(R)$ [40] is defined as the decay rate of the minimum probability of error corresponding to the optimal sequence of $(2^n R, n)$,$$
E_e(R) = \limsup_{n \to \infty} \frac{1}{n} \log P_{e, \text{opt}}(n).
\]
The capacity $C$ is the supremum of all achievable rates.
B. Control Approach

Consider the control problem in Fig. 1b where the state at time $i$ is

$$S_i = g_i(S_{i-1}, Y_{i-1}), \quad i = 1, \ldots, n$$

with initial state $S_0 = M$ and $Y_0 = \emptyset$. We refer to the mappings $g_i : S_{i-1} \times \mathcal{Y} \rightarrow S_i$, $i = 1, \ldots, n$ as the system. The controller, which observes the current state $S_i$, picks an action (symbol) $X_i \in \mathcal{X}$,

$$X_i = \pi_i(S_i), \quad i = 1, \ldots, n$$

according to a set of (stochastic) mappings $\pi_i : S_i \rightarrow \mathcal{X}$, $i = 1, \ldots, n$. We refer to the set $\{\pi_i\}_{i=1}^n$ as the controller.

The communication problem in Fig. 1a can be represented as the control problem in Fig. 1b as follows. Let the system $\{g_i\}_{i=1}^n$ be such that the state $S_i$ at time $i$ is the collection of the initial state $S_0 = M$ and the past observations $Y^{i-1}$, namely,

$$S_i = (M, Y^{i-1}) \quad i = 1, \ldots, n.$$  \hfill (8)

Also, let the encoder $\{f_i\}_{i=1}^n$ in the communication problem be picked according to the controller $\{\pi_i\}_{i=1}^n$ such that

$$f_i(M, Y^{i-1}) = \pi_i(S_i), \quad i = 1, \ldots, n.$$  \hfill (7)

Then the joint distribution of all the random variables $(M, X^n, Y^n, Z^n)$ in the control problem is the same as that in the communication problem.

To complete the representation, let $\hat{S}_0(Y^n)$ be the MMSE estimate of the initial state $S_0$ based on $Y^n$, and the final cost be $c_n(S_n) := (S_n - \hat{S}_0(Y^n))^2$. We call a controller optimal if it minimizes the final expected cost

$$E(c_n(S_n)) = E((M - \hat{M}(Y^n))^2) = D^{(n)}.$$  

Thus, the optimal controller represents the optimal encoder for the communication problem.

The system in (8) is the most general system which can represent all the encoders for the communication system. However, if the system $\{g_i\}_{i=1}^n$ is more restricted such that the state $S_i$ is a filtered version of $(S_0, Y^{i-1})$, then the controller $\{\pi_i\}_{i=1}^n$ represents only a subclass of encoders $\{f_i\}_{i=1}^n$ where the transmitted symbol $X_i$ depends on $(M, Y^{i-1})$ only through $S_i$ (see (7)). In that case, we can view the system as a filter which determines the information pattern available at the controller (encoder).

For memoryless channels, we show a subclass of encoders for which the state $S_i$ does not include all the past output $Y^{i-1}$ as in (8), yet it contains all the optimal encoders. Let $F_M(\cdot|Y^{i-1})$ be the conditional distribution of the message $M$ given the previous channel output sequence $Y^{i-1}$.

Lemma 1: An optimal encoder which minimizes the MSE $D^{(n)}$ for a discrete memoryless channel $p(y_i|x_i)$, can be found in the subclass of encoders which is determined by the system with state of the form $S_i = (M, F_M(\cdot|Y^{i-1}))$.

Proof: See Appendix A.

This lemma, which is based on a well-known result in stochastic control, provides a sufficient information pattern such that the optimal feedback encoders for the communication over a general channel can be built upon. Shayelevitz and Feder [35] proposed an explicit encoder which uses the information pattern described in Lemma 1, and showed that it is optimal in terms of the achievable rates.

III. MSE Exponent vs Rate

In this section we present the relationship between the achievable MSE exponent and the information-theoretic notions of achievable rate and error exponent.

Lemma 2: If MSE exponent $E$ is achievable, then any rate $0 < R < E$ is achievable. Moreover, if rate $R$ is achievable then MSE exponent

$$E = \min \left( R, \frac{E_n(R)}{2} \right)$$

is achievable.

Proof: See Appendix B.

Remark 1: Lemma 2 holds without any assumption on the noise sequence, input, or output of the channel.

The first part of Lemma 2 provides a useful tool for analyzing achievable rates when considering MSE exponent is more natural, e.g., Gaussian channels. It states that the achievability of an MSE exponent is sufficient to show the achievability of any rate below that MSE exponent. This in turn implies that the (feedback) MMSE capacity is upper bounded by the (feedback) capacity for a general channel\(^4\). We use an extension of Lemma 2 in Section V for the analysis of the achievable rate of the AWGN-MAC. From the second part of Lemma 2, we have the following sufficient condition for equality of the MMSE capacity and the information-theoretic capacity.

Corollary 1: $C_{MSE} = C$ if

$$\liminf_{R \to C} E_n(R) \geq 2C.$$  \hfill (9)

In the next section we show that condition (9) holds for some special cases of Gaussian channels including the AWGN channel.

IV. Point-to-Point Gaussian Channels

In this section, we turn our attention to additive Gaussian channels,

$$Y_i = X_i + Z_i$$  \hfill (10)

where the noise sequence $\{Z_i\}$ is a (colored) stationary Gaussian process with power spectral density $S_Z(\omega)$. The transmitted symbols are assumed to satisfy the (block) power constraint $P$, i.e., $\sum_{i=1}^n E(X_i^2) \leq nP$. For the rest of the paper, we consider only continuous random variables. Let $h(X) := \lim_{n \to \infty} h(X^n)/n$ denote the entropy rate [39] of a stationary continuous random process $X = \{X_i\}_{i=1}^\infty$ where $h(X^n)$ is the differential entropy of the sequence $(X_1, \ldots, X_n)$.

\(^4\)For information stable channels this fact simply follows form the maximum entropy theorem.
Consider the feedback controlled system in Fig. 2 where
the system and the controller are both LTI and combined can
be represented as the open-loop transfer function
\[ L(z) = \frac{X(z)}{Y(z)} \]  (11)
where \( X(z) \) and \( Y(z) \) are the Z-transform of the input and
the output sequences, respectively. Since the channel (10) is
additive, a negative sign is included in (11) to comply with
the convention of the linear unit-feedback control systems [41]
in which the feedback signal is subtracted from the reference
input \( Z \) and the goal is to design the control such that the
output of the system \( X \) follows the reference input \( Z \). Let
\( S(z) \) and \( T(z) \) denote the sensitivity and the complementary
sensitivity functions [41], respectively,
\[ S(z) = \frac{1}{1 + L(z)}, \quad T(z) = \frac{L(z)}{1 + L(z)}. \]  (12)
The instability \( U \) is defined as
\[ U = \sum_{j=1}^{m} \log(|\beta_j|) \]  (13)
where \( \beta_j \) are the \( m \) unstable poles of the open-loop transfer
function \( L(z) \). By the discrete version of Bode’s integral [4]
we know that if the closed loop is stable then
\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(|S(e^{j\omega})|) d\omega = \sum_{j=1}^{m} \log(|\beta_j|) = U. \]  (14)

**Theorem 1**: The capacity of the stationary Gaussian channel
with feedback, under power constraint \( P \), is
\[ C(P) = \sup_{L} U \]
where \( U \) is the instability in (13) and \( L \) is the set of all causal
LTI open-loop transfer functions \( L(z) \) in (11) which can be
stabilized by a unit-feedback over the same channel as shown in
Fig. 2 such that
\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |T(e^{j\omega})|^2 S_Z(\omega) d\omega \leq P. \]  (15)
Here, \( S_Z(\omega) \) is the power spectral density of the noise and \( T \)
is the complementary sensitivity function in (12).

**Remark 2**: Theorem 1 is based on the results of [2], [3].
Loosely speaking, Kim [3] proved that to achieve the capacity
of a stationary Gaussian channel with feedback it is sufficient
to consider stationary linear schemes. On the other hand,
Elia [2] considered a stationary linear scheme over a Gaussian
channel and expressed the directed information between the
input and the output processes in terms of the Bode integral
and instability. Theorem 1 combines the two results and
connects the capacity of the stationary Gaussian channel with
feedback to the Bode integral.

**Proof**: The capacity of the stationary Gaussian channel
under power constraint \( P \) is [3]
\[ C(P) = \sup_{X} h(Y) - h(Z) \]  (16)
where the supremum is over all stationary Gaussian processes
\( X = \{x_i\} \) of the form \( x_i = \sum_{k=1}^{\infty} b_k z_{i-k} \) such that
\( \text{E}(X_i^2) \leq P \).

Next, similar to [2], we express \( h(Y) - h(Z) \) in terms of the
Bode integral when the input process \( X \) is stationary Gaussian.
Note that the entropy rate of a stationary Gaussian process
\( X = \{x_i\}_{i=1}^{\infty} \) is [39] \( h(X) = \frac{1}{T} \int_{-\pi}^{\pi} \log(2\pi e S_X(\omega)) d\omega \)
where \( S_X(\omega) \) is the power spectral density of the process \( X \).
For a stationary Gaussian input process \( X \), the output process
\( Y \) is also stationary Gaussian and we have
\[ h(Y) - h(Z) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left( \frac{S_Y(\omega)}{S_Z(\omega)} \right) d\omega \]  (17)
where \( S_Y(\omega) \) and \( S_Z(\omega) \) are the power spectral densities of
the output and the noise sequence, respectively.

Note that the sensitivity function \( S(z) \) and the complementary
sensitivity function \( T(z) \) in (12) can be written as
\[ S(z) = \frac{Y(z)}{Z(z)}, \quad T(z) = -\frac{X(z)}{Z(z)}. \]  (18)
Therefore,
\[ S_Y(\omega) = |S(e^{j\omega})|^2 S_Z(\omega) \]  (19)
\[ S_X(\omega) = |T(e^{j\omega})|^2 S_Z(\omega) \]  (20)

Plugging \( S_Y(\omega) \) from (18) into (17) and using (14), we have
\[ h(Y) - h(Z) = U. \]  (21)
Moreover, from (19) we have
\[ \text{E}(X_i^2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\omega) d\omega \]
\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} |T(e^{j\omega})|^2 S_Z(\omega) d\omega. \]  (22)
Finally, the proof is complete by plugging (20) into (16) and
noting that the supremum in (16) can be equivalently represented
as the supremum over all causal transfer functions \( L(z) \)
such that \( \text{E}(X_i^2) \) given in (21) satisfies the power constraint.

At a high level it is well-known that in order to stabilize
an open-loop unstable control system over a limited rate
channel, then the communication rate must be greater than
the instability of the system. These type of results are known
in the literature as data-rate theorems [22], [23], [24], [25],
[26], [27], [28]. However, they typically consider error-free
digital links that impose a limit on the rate at which quantized
data can be transmitted. The cases with i.i.d erasures, AWGN,
and first-order moving average Gaussian noise are considered
in [26], [22], [28], respectively.
Theorem 1 is similar in spirit to the above works, but it considers the case of control occurring over a general (colored) stationary Gaussian channel. It states that under a given power constraint, the maximum tolerable instability of an LTI system controlled by a linear controller over a stationary Gaussian channel corresponds exactly to the feedback capacity of that channel.

**Example 1 (ARMA(1) Gaussian Channel):** In this example, we consider the special case of the first-order autoregressive moving average (ARMA(1)) Gaussian noise for which the feedback capacity can be achieved [3, Theorem 5.3] with probability of error going to zero doubly exponentially fast. Therefore, the sufficient condition (9) provided in Section III is satisfied and we have $C_{MMSE}(P) = C(P)$. Next, considering Theorem 1 we have

$$C_{MMSE}(P) = C(P) = \sup_{\mathcal{L}} U$$

where $\mathcal{L}$ is the set of all causal open-loop transfer functions $L(z)$ such that the power constraint (15) is satisfied.

**V. GAUSSIAN MULTIPLE ACCESS CHANNEL**

In this section, we extend the control representation for the communication over point-to-point channels to the feedback communication of $N$ senders and one receiver over the AWGN-MAC depicted in Fig. 3a. Each sender $j \in \{1, \ldots, N\}$ wishes to reliably transmit a message $M_j$ to the receiver. At each time $i$, the output of the channel is

$$Y_i = \sum_{j=1}^{N} X_{ji} + Z_i$$

where $X_{ji} \in \mathbb{R}$ is the transmitted symbol by sender $j$ at time $i$, $Y_i \in \mathbb{R}$ is the output of the channel, and $\{Z_i\}$ is a discrete-time zero-mean white Gaussian noise process with unit average power, i.e., $E(Z_i^2) = 1$, independent of $M_1, \ldots, M_N$. We assume that output symbols are causally fed back to the sender and the transmitted symbol $X_{ji}$ for sender $j$ at time $i$ can depend on both the message $M_j$ and the previous channel output sequence $Y_i^{-1}$. Similar to Definition 1, we define a $n$-code for the AWGN-MAC as

1) $N$ continuous messages $M_j \in (0, 1)$
2) $N$ encoders: each encoder $j$, $j = 1, \ldots, N$, assigns a symbol $X_{ji} = f_{ji}(M_j, Y_{i-1}^{-1})$
3) a decoder that assigns an estimate vector $(\hat{M}_1, \ldots, \hat{M}_N) = \phi(Y^n)$.

The set of **MSE exponents** $(E_1, \ldots, E_N)$ is called achievable if there exists a sequence of $n$-codes such that for $j = 1, \ldots, N$,

$$\liminf_{n \to \infty} \frac{1}{2n} \log(D_j^{(n)}) \geq E_j.$$  

We say that a sequence of $n$-codes has **asymptotic powers** $(\bar{P}_1, \ldots, \bar{P}_N)$ if

$$\lim_{n \to \infty} E(X_{ji}^2) = \bar{P}_j, \quad j \in \{1, \ldots, N\}.$$  

Similar to the point-to-point case, the information-theoretic achievable rates [39] are defined based on a sequence of message sets $M_j \sim \text{Unif}(\{1, \ldots, 2^{|R|}\})$ (see Definition 3). Let the probability of error be

$$P_e^{(n)} := \mathbb{P}\{(M_1, \ldots, M_N) \neq (\hat{M}_1, \ldots, \hat{M}_N)\}.$$  

The set of rates $(R_1, \ldots, R_N)$ is called achievable under block power constraints $(\bar{P}_1, \ldots, \bar{P}_N)$ if there exists a sequence of codes such that

$$\sum_{i=1}^{n} E(X_{ji}^2) \leq nP_j, \quad j \in \{1, \ldots, N\}$$

and $P_e^{(n)} \to 0$ as $n \to \infty$. We refer to $R = \sum_{j=1}^{N} R_j$ as the **sum rate** of a given code.

The following lemma presents the connection between the achievable MSE exponents and the achievable rates.

**Lemma 3:** Let MSE exponents $(E_1, \ldots, E_N)$ with asymptotic powers $(\bar{P}_1, \ldots, \bar{P}_N)$ be achievable and

$$R_j < E_j, \quad \bar{P}_j < P_j$$

for $j = 1, \ldots, N$. Then the rate-tuple $(R_1, \ldots, R_N)$ is also achievable and the block power constraints $(\bar{P}_1, \ldots, \bar{P}_N)$ are asymptotically satisfied.

**Proof:** Let the MSE exponents $(E_1, \ldots, E_N)$ be achievable. By a similar argument as in Lemma 2 we can see that the rates $(R_1, \ldots, R_N)$ are also achievable if

$$R_j < E_j, \quad j = 1, \ldots, N.$$  

Moreover, if $\bar{P}_j = \lim_{n \to \infty} E(X_{ji}^2)$, $j = 1, \ldots, N$ and $\bar{P}_j < P_j$, then using the Cesàro sum we can see that the constraint (25) is satisfied for sufficiently large $n$.  

**A. Control Approach for the AWGN-MAC**

In the following we show that codes can be designed for the AWGN-MAC with feedback, based on the LTI system controlled over a point-to-point AWGN channel depicted in Fig. 3b. We refer to these codes as LTI codes for the AWGN-MAC, which generalizes the approach introduced by Elia [2].

Let the state $S_i \in \mathbb{C}^{N \times 1}$ be a complex column vector of length $N$, and the system dynamics be

$$S_i = AS_{i-1} + BY_{i-1} = AS_{i-1} + BX_{i-1} + BZ_{i-1} \quad i = 1, \ldots, n$$

(26)
Note that one transmission over this complex channel can be complex Gaussian noise process with identity covariance. \( X \), respectively, and \( \{ Z_i \} \) is a discrete-time zero-mean white complex Gaussian noise process with identity covariance. Note that one transmission over this complex channel can be viewed as two transmissions over the real channel. Hence, the achievable rates per each complex dimension are also achievable over the real channel.

For the purpose of analysis we assume complex transmissions over a complex channel,

\[
Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{CN}(0, 1) \text{ i.i.d.} \tag{31}
\]

where \( X_i, Y_i \in \mathbb{C} \) are input and output of the channel, respectively, and \( \{ Z_i \} \) is a discrete-time zero-mean white complex Gaussian noise process with identity covariance. We assume the complex messages \( M_j \in \mathbb{C}, \ j = 1, \ldots, N \) in the AWGN-MAC are uniform over \((0, 1) \times (0, 1)\) and we set the initial state \( S_0 \in \mathbb{C}^N \) of the system as

\[
S_0 = M = (M_1, \ldots, M_N)^T.
\]

Given the system (26) and the controller (30), we derive the LTI \( n \)-code for the AWGN-MAC as follows.

**Definition 5:** An LTI \( n \)-code for the AWGN-MAC based on LTI system (26) and (30) consists of

1) Encoding mappings: The encoder \( j \) recursively forms

\[
S_i(j) = \beta_j \omega_i S_{i-1}(j) + Y_{i-1}, \quad i = 1, \ldots, n \tag{32}
\]

and at time \( i \) transmits \( X_{ji} = \pi_{ji}(S_i(j)) \).

2) Decoding: At the end of the block, the decoder forms the estimate vector \( \hat{\text{M}}_n := -A^{-n}S_n \) where

\[
\hat{\text{M}}_i = A\hat{\text{S}}_{i-1} + BY_{i-1}, \quad \hat{\text{S}}_0 = 0, \ Y_0 = 0 \tag{33}
\]

and picks \( \hat{\text{M}}_n(j) \) as the estimate of the message \( M_j \).

Note that to design codes for the MAC it is necessary that the matrix \( A \) is diagonal as in (27) and that the control signal is of the form (30) since the encoders in the MAC are decentralized and do not have access to each other’s messages. In the code described above, the dynamics of the \( j \)-th mode of the system \( S_{i}(j) \) represents the information based on which the encoder \( j \) picks the transmitted signal \( X_{ji} \) at time \( i \). We say that the controller stabilizes the system if

\[
\lim_{n \to \infty} \sup E(|S_{n}(j)|^2) < \infty, \quad j = 1, \ldots, N.
\]
Lemma 4: If the controller of the form (30) stabilizes the linear system in (26), then the corresponding sequence of \( n \)-codes for the AWGN-MAC with feedback described in (32) and (33) achieves MSE exponents
\[
E_j = \log(\beta_j), \quad j = 1, \ldots, N.
\]

Proof: See Appendix C.

Remark 3: Note that the set \( \{\beta_j\} \) depends only on the system\(^5\) and not on the controller. This means that any stabilizing controller for the system (26) can be used to design a code for the AWGN-MAC, which achieves the same set of MSE exponents \( \{\log(\beta_1), \ldots, \log(\beta_N)\} \) or, according to Lemma 3, the same set of rates.

B. Optimal Power Based on the LQG Theory

In this section, for a given system as in (26), we find the stationary controller which minimizes the asymptotic power
\[
\bar{P} := \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E(|X_i|^2). \tag{34}
\]

As we show below, the code for the AWGN-MAC which is based on this optimal controller is also optimal in terms of sum rate among the LTI codes for the AWGN-MAC under equal power constraints.

For a controller of the form (30), let \( K_{X_i} := \text{Cov}(X_i) \) be the covariance of \( X_i \) given in (28). We say that \( K_X \) is the asymptotic covariance if
\[
K_X := \lim_{n \to \infty} K_{X_n}.
\]

Note that by (30), the asymptotic power of the controller can be written as \( \bar{P} = \sum_{j,k=1}^{N} K_{X_{j,k}} \). Hence, \( \bar{P} \) represents the asymptotic combined power of all the senders in the corresponding code for the AWGN-MAC, which also captures the correlation between the signals by each sender. Whereas, the asymptotic power of each sender in the AWGN-MAC is determined by \( \bar{P}_j = (K_{X_{j,j}}) \), \( j = 1, \ldots, N \).

Lemma 5: The controller of the form (30) which stabilizes the system (26) and minimizes the cost (34) is stationary and linear, i.e.,
\[
X_i = -HS_i, \tag{35}
\]
where
\[
H = (B'GB + 1)^{-1}B'GA \tag{36}
\]
and \( G \) is the unique positive definite solutions to the discrete algebraic Riccati equation (DARE)
\[
G = A'GA - A'GB(B'GB + 1)^{-1}B'GA \tag{37}
\]
such that \( A-BH \) is stable, that is, every eigenvalue of \( A-BH \) lies inside the unit circle.

Proof: The proof is established in three steps:
1) Let \( \bar{P}^* \) be the minimum power required by a controller of the form (30) to stabilize the system (26). Then, we have
\[
\bar{P}^* \geq \bar{P}_{\min}
\]
where \( \bar{P}_{\min} \) is the minimum power required by a more general controller of the form
\[
X_i = \pi_i(S_i) \tag{39}
\]
to stabilize the system. The controller in (39) is not necessarily separated for different modes as in (30) and picks the scalar action \( X_i \) based on the complete state \( S_i = (S_{i1}, \ldots, S_{iN}) \).

2) Next, we find \( \bar{P}_{\min} \). The problem of finding a controller of the form (39) which stabilizes the system (26) and minimizes the power (34) is similar to the standard LQG problem [42] in the special case where the cost function does not depend on the state. For this problem, we can derive the Riccati equation (37) and the stationary linear control (36), similar to the solution to the LQG problem, to establish a sufficient condition for optimality in terms of the asymptotic power.

Unlike the standard LQG problem, though, here we require the control to stabilize the system (see Lemma 4). Next, we show that there exists a unique solution to (37) such that the control \( H \) in (36) is stabilizing, that is, \( A-BH \) is stable. Since the eigenvalues of \( A \) are all distinct points outside of the unit circle and the elements of \( B \) are nonzero we know \( (A, B) \) is detectable [42], that is, there exists \( H \in \mathbb{C}^{1 \times k} \) such that \( A-BH \) is stable. Then, there exists [42] a unique positive definite solution to (37) for which \( A-BH \) is stable.

3) Note that the optimal linear control \( H = (h_1, \ldots, h_N) \), which has power \( \bar{P}_{\min} \), is of the form (30), i.e.,
\[
X_i = -HS_i = \sum_{j=1}^{N} h_j S_i(j)
\]
Moreover, \( \bar{P}^* \) is by assumption the minimum power that a controller of the form (30) can stabilize the system (26). Therefore, we have
\[
\bar{P}^* \leq \bar{P}_{\min} \tag{40}
\]
Comparing with (38) and (40), we conclude that
\[
\bar{P}^* = \bar{P}_{\min}
\]
and the optimal controller of the form (30) is the same as in (36).

By Lemma 5, the optimal controller of the form (30) is linear. The following theorem provides rate and power analysis for the AWGN-MAC code based on a general stationary linear control of the form
\[
X_i = -HS_i, \quad H = (h_1h_2 \ldots h_N). \tag{41}
\]

Theorem 2: Consider the stationary linear controller (41) for the system (26). If \( A-BH \) is stable, i.e., the eigenvalues of \( A-BH \) are inside the unit circle, then the corresponding code for AWGN-MAC designed based on this linear control achieves the rates
\[
R_j < E_j = \log(\beta_j), \quad j = 1, \ldots, N
\]
with asymptoticpowers
\[
\bar{P}_j = |h_j|^2 K_{jj} = G_{jj}, \quad j = 1, \ldots, N \tag{42}
\]
\[\text{For the rest of the paper, \textit{system} refers to the one given in (26).} \]
Combining (47) and (49) completes the proof.

\[ \bar{K} = (A - BH)\bar{K}(A - BH)' + Q, \quad Q = BB'. \]  \hspace{1cm} (43)

**Remark 4:** Theorem 2 generalizes the results for \( N = 2 \) in [2] to \( N \geq 2 \).

**Proof:** First note that by Lemma 4, MSE exponents \( E_j = \log(\beta_j), j = 1, \ldots, N \), are achievable and hence, by Lemma 3, any rate \( R_j < E_j, j = 1, \ldots, N \) is achievable in each complex dimension.

For the asymptotic power, first we show \( P_j = |h_j|^2 \bar{K}_{jj} \).

Consider the code corresponding to the linear control (41), where the encoder \( j \) transmits

\[ X_{ji} = -h_j S_i(j) \]

at time \( i \) and let \( K_i := \text{Cov}(S_i) \). Then, the asymptotic power \( \bar{P}_j \) is

\[ \bar{P}_j = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E(X_{ji}^2) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |h_j|^2 (K_i)_{jj}. \]  \hspace{1cm} (45)

From the closed loop dynamic given in (44) we have the following Lyapunov recursion,

\[ K_n = (A - BH)K_{n-1}(A - BH)' + Q, \quad Q = BB'. \]

If \( A - BH \) is stable, we know [42] that

\[ \lim_{n \to \infty} K_n = \bar{K} > 0, \]  \hspace{1cm} (46)

where \( \bar{K} \) is the unique positive-definite solution of the corresponding discrete algebraic Lyapunov equation (DALE)

\[ \bar{K} = (A - BH)\bar{K}(A - BH)' + Q. \]

Therefore, by (46) and the Cesàro mean theorem, the asymptotic power in (45) becomes

\[ \bar{P}_j = |h_j|^2 \bar{K}_{jj}. \]  \hspace{1cm} (47)

Finally, note that the DARE given in (37) can be equivalently written as the following discrete algebraic Lyapunov equation (DALE)

\[ G = (A - BH)'G(A - BH) + H'H. \]  \hspace{1cm} (48)

Comparing (43) and (48), the diagonal elements of \( G \) and \( \bar{K} \) can be related as follows

\[ G_{jj} = |h_j|^2 \bar{K}_{jj}. \]  \hspace{1cm} (49)

Combining (47) and (49) completes the proof.

**C. Kramer Code vs. Optimal LTI Code**

Kramer [7] presented a linear code for the Gaussian multiple access channel with feedback which provides the best known lower bound on the sum rate. This code, under symmetric block power constraints \( P \), achieves any sum rate \( R < R(P) \) where

\[ R(P) := \frac{1}{2} \log(1 + NP\phi(P)) \]  \hspace{1cm} (50)

and \( \phi(P) \) is the unique solution to

\[ (1 + NP\phi)^{N-1} = (1 + P\phi(N - \phi))^N \]  \hspace{1cm} (51)

in the interval \([1, N]\). An alternative representation of Kramer’s code is provided in [8] where it is also shown that this code is optimal among the class of generalized linear feedback codes.

Let the optimal LTI code for the AWGN-MAC be the code based on the optimal control given in Lemma 5. We show that the optimal LTI code for the system \( A \) with symmetric parameters

\[ \omega_j = e^{2\pi\sqrt{\beta_j}}, \quad \beta_j = \beta, \quad j = 1, \ldots, N. \]  \hspace{1cm} (52)

has the same performance as Kramer’s code under equal power constraints for all the senders.

The following lemma characterizes the matrix \( G \) for the symmetric choice of matrix \( A \) given in (52).

**Lemma 6:** [8] Let \( \beta_j = \beta \) and \( \omega_j = e^{2\pi\sqrt{\beta_j}} \). The unique positive-definite solution \( G \geq 0 \) to the DARE (37) is circulant with real eigenvalues satisfying \( \lambda_i = \frac{\beta}{2\pi}\lambda_i - 1 \) for \( i = 2, \ldots, N \). The largest eigenvalue \( \lambda_1 \) satisfies

\[ 1 + N\lambda_1 = \beta^{2N}. \]  \hspace{1cm} (53)

\[ \left( 1 + \lambda_1(\frac{N - \lambda_1}{G_{11}}) \right) = \beta^{2(N-1)}. \]  \hspace{1cm} (54)

**Remark 5:** A closed form solution to the DARE (37) for \( (A, B) \) in (27) with non-symmetric real eigenvalues is provided in [43]. By generalizing the result of [43] to complex eigenvalues, the powers \( P_j, j = 1, \ldots, N \), can be determined, in a closed form, as a function of achievable MSE exponents (or achievable rates) \( E_j = \log(\beta_j, j = 1, \ldots, N \), for the general case.

**Theorem 3:** Given \( R < R(P) \), there exists \( \beta > 1 \) such that the optimal LTI code based on the symmetric system (52) achieves sum rate \( R \) while satisfying the symmetric power constraints \( P \) asymptotically, that is, it has the same performance as Kramer’s code.

**Proof:** Since \( R < R(P) \) we can always find \( \beta > 1 \) such that

\[ R < N\log(\beta) < R(P) \]  \hspace{1cm} (55)

By Theorem 2, the optimal LTI code based on the symmetric system (52) and the optimal controller in (36) achieves the symmetric MSE exponent \( E, i.e., E_j = E = \log(\beta), j = 1, \ldots, N \) with asymptotic powers \( P_j = h_j^2 \bar{K}_{jj}, j = 1, \ldots, N \) where \( \bar{K} \) is given in (43). Moreover, by Lemma 3, rates \( R_j < \log(\beta_j), j = 1, \ldots, N \) are achievable and therefore the sum rate

\[ R < N \log(\beta) \]  \hspace{1cm} (56)
is achievable.

It remains to show that the asymptotic powers \( \hat{P}_j \) in the optimal LTI code satisfy

\[
\hat{P}_j < P, \quad j = 1, \ldots, N. \tag{57}
\]

Since \( \hat{P}_j = G_{jj} \) (see (42)) it is sufficient to show that if \( \beta \) satisfies (55) then \( G_{jj} < P \) for the corresponding \( G \). First, note that from (53) and (54) we have

\[
\left(1 + N \lambda_1 \right)^{-1} = \left(1 + \lambda_1 (N - \frac{\lambda_1}{G_{11}}) \right)^N. \tag{58}
\]

Comparing with (51) and noting that \( G \) is circulant we get

\[
\lambda_1 = G_{jj} \phi(G_{jj}), \quad j = 1, \ldots, N. \tag{59}
\]

Now if (55) holds, from (53) we know

\[
\frac{1}{2} \log(1 + N \lambda_1) = N \log(\beta) < \frac{1}{2} \log(1 + N P \phi(P))
\]

and hence

\[
\lambda_1 < P \phi(P). \tag{60}
\]

From (59) and (60) we have \( G_{jj} \phi(G_{jj}) < P \phi(P) \), and \( G_{jj} < P \) follows from monotonicity of \( \phi(P) \) in \( P \). Combined with (42), then (57) follows and the proof is complete.  

\[\Box\]

\section*{VI. Conclusion}

Communication and control problems have different goals, but they both deal with information dynamics and in many cases they share similar formulations that can be tackled by tools and techniques developed in both fields. Understanding the interface between these two theories has become even more important in the last decade, as we have witnessed technological advancements leading to the convergence of computing, communication, and control over networked platforms of embedded systems.

We considered the problem of communication in presence of feedback and presented three contributions. First, we introduced the MSE exponent as a tool for analyzing the reliable rate of communication and also presented a sufficient condition under which the MMSE capacity is equal to the information-theoretic capacity. Second, we showed that the capacity of the stationary Gaussian channel with feedback subject to a certain power constraint is equal to the maximum instability of an LTI system that can be tolerated by any linear controller with the same power constraint. This result is obtained combining the works of Elia [2], who provided the relationship between the entropy rates restricted to stationary linear schemes over Gaussian channels and the Bode integral; and Kim [3] who showed that stationary linear schemes are sufficient to achieve capacity. Finally, we showed how LQG control tools can be used to design codes for communication. In this case, we focused on the AWGN-MAC with feedback and showed that our technique can generalize results in [2] to an arbitrary number of senders.

\section*{Appendix A \quad Proof of Lemma 1}

The communication problem with feedback can be viewed as a partially observable Markov decision problem (POMDP) where the controller has access only to \( Y_{i-1} \) but not the complete state \( (M, Y_{i-1}) \) as follows. Note that without loss of generality we can decompose the encoding functions \( f_i \) in (2) into two steps as follows. First, based on the past output \( Y_{i-1} \), the encoder (controller) picks a function (action) \( a_i \),

\[
a_i : \mathcal{M} \rightarrow \mathcal{X}. \tag{61}
\]

Then it transmits

\[
x_i = a_i(m). \tag{62}
\]

With this separation, the encoder (controller) can be defined by the set of mappings (actions)

\[
\pi_i : \mathcal{Y}_{i-1} \rightarrow \mathcal{A}_i
\]

where \( \mathcal{A}_i = \{ a_i : \mathcal{M} \rightarrow \mathcal{X} \} \) is the set of all functions (actions) given in (61). Note that this new controller \( \{ \pi_i \}_{i=1}^n \) only observes the past output \( Y_{i-1} \) and not the message \( M \).

Based on the results in [44, Chapter 6], we show that the posterior \( F_M(\cdot | y_{i-1}) \) is an information state for this new controller. Let the new state space be

\[
\mathcal{S} = \{ F : F \text{ is a probability distribution over } \mathcal{M} \},
\]

and the state at time \( i \) be \( s_i = F_M(\cdot | y_{i-1}, a_{i-1}) \). We use \( S_i \) to denote a random distribution. Note that the MSE can be written as \( D^{(n)} = E(c_n(s_n)) \) where

\[
c_n(s_n) = E \left( M - \hat{M}(Y_{i-1} | Y_{i-1}) \right). \tag{63}
\]

At time \( i \), action \( a_i \) generally depends on \( (y_{i-1}, a_{i-1}) \). We use \( A_i \) to denote a random function. Then, it follows that

\[
s_{i+1} = F_M(\cdot | y_i, a_i)
\]

\[
= F_M(\cdot | y_{i-1}, a_i | y_i, y_{i-1}, a_i) \tag{64}
\]

\[
= \frac{f(y_i | m, y_{i-1}, a_i) dF_M(m | y_{i-1}, a_i)}{f(y_i | y_{i-1}, a_i)}
\]

\[
= \frac{f(y_i | m, a_i) dF_M(m | y_{i-1}, a_i)}{f(y_i | y_{i-1}, a_i)}
\]

\[
= \frac{f(y_i | m, a_i) dF_M(m | y_{i-1}, a_i)}{f(y_i | y_{i-1}, a_i)}
\]

where equality (64) comes from the fact that \( Y_i \rightarrow (M, A_i) \rightarrow (Y_{i-1}, A_{i-1}) \) form a Markov chain since \( x_i = a_i(M) \) and the memoryless channel is determined by \( p(y_i | x_i) \). Equality (65) comes from the fact that \( M \rightarrow (A_{i-1}, Y_{i-1}) \rightarrow A_i \) form a Markov chain since the common agent picks \( A_i \) only based on \( (A_{i-1}, Y_{i-1}) \).

From (65) it can be verified that \( S_{i+1} \) is a deterministic function of \( (S_i, A_i, Y_i) \). Moreover, the distribution of \( Y_i \) is determined by \( (S_i, A_i) \) through (62) and the channel distribution \( p(y_i | x_i) \). Therefore \( (Y_{i-1}, A_{i-1}, S_{i-1}) \rightarrow (S_i, A_i) \rightarrow (S_{i+1}, Y_i) \) form a Markov chain and we conclude [44, Chapter 6] that \( S_i \) is an information state for the purpose of control. Hence, there is no loss of optimality if we consider
only Markov policies such that the action at time $i$ depends on $(Y_i, A_i, A_{i-1})$ only through the information state $S_i$ as follows

$$A_i = \tilde{A}_i(S_i).$$

Finally, since $A_i$ is chosen based on $Y_i-1$ we know $A_i \rightarrow Y_i \rightarrow M$ form a Markov chain and $F_M(\cdot | Y_i-1, A_i-1) = F_M(\cdot | Y_i-1)$. Therefore, to find $X_i = A_i(M)$, it is sufficient to have the message $M$ and its posterior distribution $F_M(\cdot | Y_i-1)$.

**APPENDIX B**

**PROOF OF LEMMA 2**

By assumption, there exists a sequence of $n$-codes such that for $M \sim \text{Unif}(0, 1)$

$$\lim \inf_{n \to \infty} -\frac{1}{2n} \log D(n) \geq E. \quad (66)$$

Consider

$$p_n := P\left\{ |\Theta - \hat{\Theta}| > \frac{1}{2} \cdot 2^{-nR} \right\} \leq 4 \cdot 2^{2nR} \cdot D(n) \quad (67)
\leq 4 \cdot 2^{2nR} \cdot 2^{-2n(E-\epsilon_n)} \quad (68)
= 4 \cdot 2^{-2n(E-R-\epsilon_n)} \quad (69)$$

where $\epsilon_n \to 0$ as $n \to \infty$. The inequalities (68) and (69) follow from the Chebyshev inequality and (66), respectively. From (70) and the assumption $R < E$, we have

$$p_n \to 0 \quad \text{as} \quad n \to 0. \quad (71)$$

Now consider message points in the interval $(0, 1)$ such that the distance between any two points is equal to $2^{-nR}$. Then, by (71) we know $\lim_{n \to \infty} p_n(n) = 0$ and hence rate $R < E$ is achievable. (For a rigorous argument see [35, Lemma 3]).

To prove the second part, let $R \to \infty$ be achievable such that $\lim_{n \to \infty} -\frac{1}{2n} \log P_e(n) \leq -2R$. For each $n$, we divide the unit interval $(0, 1)$ into $2^{-nR}$ equal sub-intervals and map the continuous message $M \in (0, 1)$ to the discrete message $M_n \in \{1, \ldots, 2^{2nR}\}$ according to the sub-interval $M$ lies in. To communicate $M$, we send the corresponding $M_n$ using the $(2^{2nR}, n)$ code, which achieves rate $R$, and we pick the middle point of the interval corresponding to the decoded message $\hat{M}_n$ as the estimate of $M \in (0, 1)$. The MSE for $M$ can be upper bounded by $2^{-2nR}$ if the message $M_n$ is decoded correctly, and by 1 in case of an error. Hence,

$$D(n) \leq P_e(n) + (1 - P_e(n))2^{-2nR}$$

and we have

$$-\frac{1}{2n} \log D(n) \geq R - \frac{1}{2n} \log \left(2^{2nR} P_e(n) + (1 - P_e(n))\right) \quad (72)$$

Taking liminf from both sides and noting that if $E_e(R) \geq 2R$ the second term in (72) will be zero we conclude that

$$E = R - \max\left(0, R - \frac{E_e(R)}{2}\right) = \min\left(R, \frac{E_e(R)}{2}\right)$$

is achievable.

**APPENDIX C**

**PROOF OF LEMMA 4**

The system dynamics given in (26) can be rewritten as

$$S_i = AS_{i-1} + BY_{i-1} = A' \tilde{S}_0 + \tilde{S}_i \quad (73)$$

where $\tilde{S}_i$ is given in (33). Plugging $i = n$ and $S_0 = M$ into (73) and multiplying both sides by $A^{-n}$ we have

$$A^{-n}S_n = M + A^{-n}\tilde{S}_n = M - \tilde{M}_n$$

where $M = -A^{-n}\tilde{S}_n$ is the estimate of the message (see Definition 5). The covariance matrix of the error vector $\epsilon_n := M - \tilde{M}_n = A^{-n}S_n$ can be written as

$$\text{Cov}(\epsilon_n) = A^{-n}K_nA^{-n}$$

where $K_n := \text{Cov}(S_n)$ is the covariance matrix of $S_n$. Therefore, the MSE for the sender $j$ is

$$D_j^{(n)} = E\left( |\epsilon_n(j)|^2 \right) = \beta_j^{2n}(K_n)_{jj} \quad (74)$$

By the assumption of stability $\lim sup_{n \to \infty} (K_n)_{jj} < \infty$ and from (74) we have

$$\lim inf_{n \to \infty} -\frac{1}{2n} \log(D_j^{(n)}) = \log(\beta_j).$$

Hence, the MSE exponents $E_j = \log(\beta_j)$ for $j = 1, \ldots, N$ are achievable.

**ACKNOWLEDGMENTS**

The authors would like to thank Tara Javidi, Young-Han Kim, Paolo Minero, and Ofer Shayevitz for valuable discussions.

**REFERENCES**


