New Formulas of Feedback Capacity for AGN Channels with Memory: A Time-Domain Sufficient Statistic Approach †

Charalambos D. Charalambous *, Christos Kourtellaris and Stelios Louka ‡

Department of Electrical and Computer Engineering, University of Cyprus, P.O. Box 20537, Nicosia CY-1678, Cyprus; kourtellaris.christos@ucy.ac.cy (C.K.)
* Correspondence: chadcha@ucy.ac.cy
† Preliminary results are presented in part at the 2021 IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021. This paper appeared in arXiv document https://arxiv.org/abs/2010.06226, 13 October 2020.

Abstract: Recently, several papers identified technical issues related to equivalent time-domain and frequency-domain “characterization of the n–block or transmission” feedback capacity formula and its asymptotic limit, the feedback capacity, of additive Gaussian noise (AGN) channels, first introduce by Cover and Pombra in 1989 (IEEE Transactions on Information Theory). The main objective of this paper is to derive new results on the Cover and Pombra characterization of the n–block feedback capacity formula, and to clarify the main points of confusion regarding the time-domain results that appeared in the literature. The first part of this paper derives new equivalent time-domain sequential characterizations of feedback capacity of AGN channels driven by non-stationary and non-ergodic Gaussian noise. It is shown that the optimal channel input processes of the new equivalent sequential characterizations are expressed as functionals of a sufficient statistic and a Gaussian orthogonal innovations process. Further, the Cover and Pombra n–block capacity formula is expressed as a functional of two generalized matrix difference Riccati equations (DREs) of the filtering theory of Gaussian systems, contrary to results that appeared in the literature and involve only one DRE. It is clarified that prior literature deals with a simpler problem that presupposes the state of the noise is known to the encoder and the decoder. In the second part of this paper, the existence of the asymptotic limit of the n–block feedback capacity formula is shown to be equivalent to the convergence properties of solutions of the two generalized DREs. Further, necessary and or sufficient conditions are identified for the existence of asymptotic limits, for stable and unstable Gaussian noise, when the optimal input distributions are asymptotically time-invariant but not necessarily stationary. This paper contains an in-depth analysis, with various examples, and identifies the technical conditions on the feedback code and state space noise realization, so that the time-domain capacity formulas that appeared in the literature, for AGN channels with stationary noises, are indeed correct.

Keywords: feedback capacity; non-stationary; unstable; Gaussian channels; memory; sufficient statistic

1. Introduction, Motivation, Main Results, Current State of Knowledge

In the recent papers [1–3], concerns are raised whether the time-domain analysis in [4], deals with the Cover and Pombra [5] “characterization of the n–block or transmission” feedback capacity formula and its asymptotic limit, the feedback capacity, of additive Gaussian noise (AGN) channels. Furthermore, the recent comment paper [6] identified
gaps in the proof of the simplified frequency-domain characterization of Theorem 4.1 in [4]. The main objective of this paper is to derive new results on the Cover and Pombra characterization of the $n$–block feedback capacity formula, and to clarify the main points of confusion regarding the time-domain results of [4] and related literature i.e., [7].

1.1. The Problem, Motivation, and Main Results

We consider the additive Gaussian noise (AGN) channel defined by [5]

$$Y_t = X_t + V_t, \quad t = 1, \ldots, n,$$

where

$$X^n \triangleq \{X_1, X_2, \ldots, X_n\}$$ is the sequence of channel input random variables (RVs) $X_t : \Omega \rightarrow \mathbb{R}$;

$$Y^n \triangleq \{Y_1, Y_2, \ldots, Y_n\}$$ is the sequence of channel output RVs $Y_t : \Omega \rightarrow \mathbb{R}$;

$$V^n \triangleq \{V_1, \ldots, V_n\}$$ is the sequence of jointly Gaussian distributed RVs $V_t : \Omega \rightarrow \mathbb{R}$, with which are not necessarily stationary or ergodic.

We wish to examine the feedback capacity of the AGN channel (1) for two distinct formulations of code definition and a noise model, described below under Case (I) and Case (II) formulations. Special cases of these will be related to the existing literature.

Case (I) Formulation. This formulation respects the following two conditions.

(I.1) The feedback code does not assume knowledge of the initial state of the noise at the encoder and the decoder (see Definition 1), and

(I.2) the noise sequence $V^n$ is represented by a partially observable (partially observable means that knowledge of $V^{t-1}$ and the initial state do not specify the state $S^t$, $t = 1, \ldots, n$) state space realization, with state sequence $S^n$ (see Definition 2).

For a formulation that respects (I.1) and (I.2), Cover and Pombra characterized the “$n$–finite transmission” feedback capacity [5] (Equations (10) and (11)), using the information measure $(C_{f b}^n (\kappa))$ is identified using the converse coding theorem [5],

$$C_{f b}^n (\kappa) \triangleq \sup_{P_{X_t|Y^{t-1}, Y^{t-1} = 1, \ldots, n}} \frac{1}{n} \mathbb{E} \left\{ \sum_{t=1}^{n} (X_t)^2 \right\} \leq \kappa, \quad \kappa \in [0, \infty) \quad (2)$$

provided the supremum exists, where $H(\cdot)$ denotes (differential) entropy.

Case (I.a) Formulation. Although not mentioned in [5], if the feedback code assumes knowledge of the initial state of the noise or the channel, $S_1 = s$, at the encoder and the decoder (see Definition 3), it follows directly from [5] (Equations (10) and (11)), that (2) is replaced by the information measure

$$C_{f b}^n (\kappa, s) \triangleq \sup_{P_{X_t|Y^{t-1}, Y^{t-1} = 1, \ldots, n}} \frac{1}{n} \mathbb{E} \left\{ \sum_{t=1}^{n} (X_t)^2 \right\} \mid S_1 = s \leq \kappa \quad (3)$$

Case (II) Formulation. This formulation relaxes Conditions (I.1) and (I.2) to the following two conditions:

(II.1) The feedback code assumes knowledge of the initial state of the noise or the channel, $S_1 = s$, at the encoder and the decoder (see Definition 3);

(II.2) the noise sequence $V^n$ is represented by a fully observable state space realization (fully observable means knowledge of $V^{t-1}$ and initial state specify the state $S^t$, $t = 1, \ldots, n$), with state sequence $S^n$ such that the noise $V^{t-1}$ and the initial state $S_1$ uniquely defines the noise state sequence $S^t$ and vice versa for $t = 1, \ldots, n$.

Thus, Formulation (2), which respects Conditions (I.1) and (I.2), is the most general.
For a formulation that respects Conditions (II.1) and (II.2), Yang, Kavcic, and Tatikonda [8] characterized the \( n \)-finite transmission feedback capacity [8] (Section II (particularly Section IIC, I–III)), using the information measure,

\[
C_n^{fb}(\kappa, s) \triangleq \sup_{p_{X_0|Y^{t-1}, s}, t=1, \ldots, n} \frac{1}{n} \sum_{t=1}^{n} H(Y_t|Y^{t-1}, s) - H(V^n|s). \tag{4}
\]

Compared to \( C_n^{fb}(\kappa) \) and \( C_n^{fb}(\kappa, s) \), the definition of \( C_n^{fb, S}(\kappa, s) \) imposes Condition (II.2) and is fundamentally different from the former, because the input distributions of \( C_n^{fb, S}(\kappa, s) \) are different from \( C_n^{fb}(\kappa) \) and \( C_n^{fb}(\kappa, s) \). Hence, the information rates of the three formulas are generally different. However, for certain Gaussian noise models of \( V^n \), it might be the case that, under Condition (II.1), the information measures \( C_n^{fb}(\kappa, s) \) and \( C_n^{fb, S}(\kappa, s) \) coincide. We provide several examples in the main body of this paper.

**Motivation and Fundamental Differences of Case (I) and Case (II) Formulations.**

At this point, we pause to discuss two technical issues of Case (I) and Case (II) formulations, which are not clarified in [4,7,9]. These technical issues are related to the time-domain characterization of feedback capacity, Theorem 4.1 in [4]; they are first discussed in [1–3,10]. A recent comment paper [6] also identified gaps in the proof of the frequency-domain characterization of capacity, Theorem 6.1 in [4]; they are first discussed in [4,7,9].

To illustrate the technical issues, we consider the autoregressive moving average stable noise denoted by ARMA \((a, c), a \in [-1,1], c \in (-1,1), c \neq a\), studied by many authors [4,7–9,11,12], as a benchmark example.

\[
V_t = cV_{t-1} + W_t - aW_{t-1}, \quad \forall t \in \mathbb{Z}_+ = \{1, 2, \ldots\}, \tag{5}
\]

\[
V_0 \in N(0, K_{V_0}), \quad K_{V_0} \geq 0, \quad W_0 \in N(0, K_{W_0}), \quad K_{W_0} \geq 0, \quad W_t \in N(0, K_{W}), \quad K_{W} > 0, \quad \forall t, \tag{6}
\]

\[
\{W_0, W_1, \ldots, W_n\} \text{ indep. seq. and indep. of } V_0. \tag{7}
\]

As in [4,7], we define the state variable of the noise by

\[
S_t \triangleq \frac{c V_{t-1} - a W_{t-1}}{c - a}, \quad \forall t \in \mathbb{Z}_+. \tag{8}
\]

Then, the state space realization of \( V^n \) is

\[
S_{t+1} = cS_t + W_t, \quad V_t = \left( c - a \right) S_t + W_t, \quad \forall t \in \mathbb{Z}_+, \tag{9}
\]

\[
K_{S_t} = \left( \frac{c}{c-a} \right)^2 K_{V_0} + \left( \frac{a}{c-a} \right)^2 K_{W_0}, \quad K_{V_0} \geq 0, \quad K_{W_0} \geq 0 \quad \text{both given.} \tag{10}
\]

Bounds on feedback capacity, when \( a = 0 \), i.e., corresponding to an autoregressive noise are derived in [13–15] using linear coding schemes.

For the Case (I) formulation, the information measure \( C_n^{fb}(\kappa) \), i.e., (2), corresponds to the supremum over channel input distributions \( P_{X_t|Y^{t-1}, t=1, \ldots, n} \).

For the Case (II) formulation, the information measure \( C_n^{fb, S}(\kappa, s) \), i.e., (4), corresponds to the supremum over channel input distributions \( P_{X_t|Y^{t-1}, s}, t=1, \ldots, n \), and Conditions (II.1) and (II.2) are necessary. Alternatively, the necessary conditions for \( C_n^{fb}(\kappa) \) to reduce to
$C_n^{fbS}(\kappa, s)$ are the conditions stated in (12) and (13) (these also follow independently of the Case (I) formulation from the converse coding theorem).

\[
P_{X_t|X^{t-1}, Y^{t-1}} = P_{X_t|Y^{t-1}} \quad \text{holds by channel def. } Y_k = X_k + V_k, k = 1, \ldots, n; \quad (11)
\]

\[
= P_{X_t|Y^{t-1}, S^t} \quad \text{if } S_1 = s \text{ is known to the feedback code, i.e., Cond. (II.1);} \quad (12)
\]

\[
= P_{X_t|S^t, Y^{t-1}} \quad \text{if } (V^{t-1}, S_1 = s) \text{ uniquely def. } S^t \text{ and vice-versa, i.e., Cond. (II.2).} \quad (13)
\]

Thus, a necessary condition for (13) to hold is $S_1 = (V_0, W_0) = (v_0, w_0) = s$, which is known to the encoder. It follows from [4] (Theorem 6.1; also see Lemma 6.1 and the comments above it, Equation (71)), that Conditions (II.1) and (II.2) are assumed; hence, these results are not developed for the Cover and Pombra [5] formulation. Additional information can be found in Remark 5.

Second, the analysis of the asymptotic per unit time limits of (2)–(4), i.e.,

\[
C_{\text{FB}}(\kappa) \triangleq \lim_{n \to \infty} \sup_{\{P_{X_t|X^{t-1}}\}_{t=1}^n} \frac{1}{n} \mathbb{E} \left\{ \sum_{t=1}^n (X_t)^2 \right\} \leq \kappa,
\]

(14)

is a non-trivial problem, and it requires certain technical necessary and/or sufficient conditions for the limits to exist, as well as for the rates to be independent of the initial distribution $P_{V_1}$ or the initial data, $S_1 = s$ (see [3,10]), even if the noise process $V^n$ is stationary. It is easy to verify that the analysis in the past studies [4,7,9,11,12] considered the simpler problem $C_n^{fbS}(\kappa, s)$, and that the asymptotic limit $\lim_{n \to \infty} \frac{1}{n} C_n^{fbS}(\kappa, s)$ does not correspond to the ergodic capacity. We clarify these points in our examples.

**Main Results.** The main results of this paper are briefly stated below.

1. In the first part of this paper, we derive new equivalent sequential characterizations of the Cover and Pombra “$n$–block or transmission” feedback capacity formula [5] (Equation (11)), $C_n^{fb}(\kappa)$ (this first appeared in [10]). In particular, we derive equivalent realizations to the optimal channel input process $X^n$ [5] (Equation (11)), which are linear functionals of a finite-dimensional sufficient statistic and an orthogonal innovations process. From these new realizations follows the equivalent sequential characterizations of the “$n$–block or transmission” feedback capacity formula [5] (Equation (11)), which will henceforth be called the “$n$–finite transmission feedback information (n–FTFI) capacity”. The new $n$–FTFI capacity is expressed as a functional of two generalized matrix difference Riccati equations (DREs) of the filtering theory of Gaussian systems, instead of the one DRE given in [4,7,8]. In fact, we also show that the $n$–FTFI capacity of [4,7,8] corresponds to $C_n^{fbS}(\kappa, s)$.

2. In the second part of this paper, we analyze the asymptotic per unit time limit of the sequential characterizations of the $n$–FTFI capacity, when the supremum and limit over $n \to \infty$ are interchanged in (14), denoted by $C_{\text{FB}}^{\rho}(\kappa)$. Then, we show $C_{\text{FB}}^{\rho}(\kappa) = C_{\text{FB}}(\kappa)$. We identify necessary and/or sufficient conditions for the asymptotic limit to exist, and for the optimal input process $X_t$, $t = 1, \ldots$, to be asymptotically stationary, in terms of the convergence properties of two generalized matrix difference Riccati equations (DREs) to their corresponding two generalized matrix algebraic Riccati equations (AREs). We make use of the so-called detectability, stabilizability, and unit circle controllability conditions of generalized Kalman filters of Gaussian processes [16,17].

3. From (1) and (2), we derive analogous results for $C_n^{fb}(\kappa, s)$ and its per unit time asymptotic limit, denoted by $C_{\text{FB}}^{\rho}(\kappa, s)$, as degenerate versions of $C_n^{fb}(\kappa)$ and $C_{\text{FB}}^{\rho}(\kappa)$. Further, we show that for certain noise models, and under certain conditions, it holds that $C_{\text{FB}}^{\rho}(\kappa, s) = C_{\text{FB}}(\kappa)$, i.e., these values do not depend on the initial state or initial distributions.
(4) From (1) and (2), we derive analogous results for the Case (II) formulation, i.e., \( C_n^{fb,S}(\kappa, s) \) and its per unit time asymptotic limit denoted by \( C^{fb,S}(\kappa, s) \), and we show these are fundamentally different from the Case (I) formulation, \( C_n^{fb}(\kappa) \) and \( C^{fb}(\kappa) \), as well as \( C_n^{fb}(\kappa, s) \) and \( C^{fb}(\kappa, s) \). In particular, we show that the characterizations of \( n \)-FTFI capacity, \( C_n^{fb,S}(\kappa, s) \), for the Case (II) formulation follow directly from the Case (I) formulation, as a special case (an independent derivation is also presented). Moreover, \( C_n^{fb,S}(\kappa, s) \) is a functional of one generalized DRE, while \( C_n^{fb}(\kappa), C_n^{fb}(\kappa, s) \) are functionals of two generalized DREs.

1.2. The Code Definitions and Noise Models

**Case (I) Feedback Code and Noise Definitions.** For the Case (I) formulation, we consider the code of Definition 1 (due to [5]).

**Definition 1.** Time-varying feedback code [5]
A noiseless time-varying feedback code for the AGN Channel (1) is denoted by \((2^nR, n), n = 1, 2, \ldots \) and consists of the following elements and assumptions:

(i) The uniformly distributed messages \( W : \Omega \to \mathcal{M}_n \equiv \{1, 2, \ldots , 2^nR\} \).

(ii) The time-varying encoder strategies, often called codewords of block length \( n \), defined by \((\text{the superscript } e(\cdot) \text{ on } \mathcal{E}^e \text{ indicates that the distribution depends on the strategy } e(\cdot) \in \mathcal{E}_{[1,n]}(\kappa))\).

\[ E_{[1,n]}(\kappa) \triangleq \{X_1 = e_1(W), X_2 = e_2(W, X_1, Y_1), \ldots , X_n = e_n(W, X_{n-1}, Y_{n-1}) : \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^{n} (X_i)^2 \right) \leq \kappa \}. \]  

(15)

(iii) The average error probability of the decoder functions \( y^n \longmapsto d_n(y^n) \in \mathcal{M}_n \), defined by

\[ P_{\text{error}}^{(n)} = \mathbb{P}\{d_n(Y^n) \neq W\} = \frac{1}{2^{nR}} \sum_{w=1}^{2^nR} \mathbb{P}\{d_n(Y^n) \neq w\} \bigg| W = w \}. \]  

(16)

(iv) The channel input sequence \( \{X^n \triangleq \{X_1, \ldots , X_n\}\} \) is causally related (a notion found in [5], page 39, above Lemma 5) to \( V^n \), which is equivalent to the following decomposition of the joint probability distribution of \((X^n, V^n)\):

\[ P_{X^n, V^n} = P_{V^n|V^{n-1}, X^n} P_{X_n|X^{n-1}, V^{n-1}} P_{V_{n-1}|X_{n-1}, V_{n-1}} \cdots P_{V_1|X_1, X_1} P_{V_1|X_1} P_{X_1} \]  

(17)

\[ = P_{V^n} \prod_{t=1}^{n} P_{X_t|X_{t-1}, V_{t-1}}, \quad \text{that is,} \quad P_{V_t|V_{t-1}, X_t} = P_{V_t|V_{t-1}}. \]  

(18)

That is, \( X^t \leftrightarrow \forall V^{t-1} \leftrightarrow V_t \) is a Markov chain, for \( t = 1, \ldots , n \). As usual, the messages \( W \) are independent of the channel noise \( V^n \).

A rate \( R \) is called an achievable rate with feedback coding, if there exists a sequence of codes \((2^{nR}, n), n = 1, 2, \ldots \) such that \( P_{\text{error}}^{(n)} \longrightarrow 0 \) as \( n \longrightarrow \infty \). The feedback capacity \( C^{fb}(\kappa) \) is defined as the supremum of all achievable rates \( R \).

We note that, in general, \( C^{fb}(\kappa) \) depends on the initial distribution \( P_{V_1} \); the ergodic capacity requires that \( C^{fb}(\kappa) \) is independent of \( P_{V_1} \) (see [18]).

We consider a noise model which is consistent with [5], i.e., \( V^n \) is jointly Gaussian distributed, \( P_{V^n} = \prod_{t=1}^{n} P_{V_t|V^{t-1}} \), and induced by the partially observable state space (PO-SS) realization of Definition 2.
**Definition 2.** A time-varying PO-SS realization of Gaussian noise \( V^n \in N(0, K_{V^n}) \) is defined by

\[
\begin{align*}
S_{t+1} &= A_t S_t + B_t W_t, \quad t = 1, \ldots, n-1 \\
V_t &= C_t S_t + N_t W_t, \quad t = 1, \ldots, n, \\
S_1 &\in N(\mu_{S_1}, K_{S_1}), \quad K_{S_1} \succeq 0, \\
W_t &\in N(0, K_{W_t}), \quad K_{W_t} \succeq 0, \quad t = 1, \ldots, n \quad \text{indep. Gaussian process}, \quad W^t \text{ indep. of } S_1, \\
S_t : \Omega \to \mathbb{R}^{n_1}, \quad W_t : \Omega \to \mathbb{R}^{n_w}, \quad V_t : \Omega \to \mathbb{R}^{n_2}, \quad R_t = N_t K_{W_t} N_t^T > 0, \quad t = 1, \ldots, n
\end{align*}
\]

where \( n_1, n_2, n_w \) are arbitrary positive integers, \((A_t, B_t, C_t, N_t, \mu_{S_t}, K_{S_t}, K_{W_t})\) are nonrandom, measurable in \( t \), with bounded entries, and \( n_w, n_2, n_2 \) are finite positive integers.

The above components \( B_t W_t \) and \( N_t W_t \) can be taken as

\[
B_t W_t = B_t^1 W_t^1 + B_t^2 W_t^2, \quad N_t W_t = N_t^1 W_t^1 + N_t^2 W_t^2, \quad t = 1, \ldots, n,
\]

where \( B_t^1 \) or \( B_t^2 \) can be zero and \( N_t^1 \) or \( N_t^2 \) can be zero \( \forall t \).

A time-invariant PO-SS realization of the Gaussian noise \( V^n \in N(0, K_{V^n}) \) is defined by (19)–(23), with \((A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_{W}), \forall t\).

For the Case (I) formulation, we use the terminology “partially observable”, which is standard in filtering theory [16], because the noise \( V^n \) induces a distribution \( P_{V^n} = \times_{t=1}^{n} P_{V_t|V^{t-1}} \) and \( P_{V_t|V^{t-1}} \) cannot be expressed as a function of the state of the noise, i.e., \( V^{t-1} \) does not uniquely define \( S^t \). However, if \( S_1 = s \) is known to the encoder, then it can be as easily verified from the ARMA \((a, c)\), that \( V^{t-1} \) uniquely defines \( S^t \), recursively. However, for the PO-SS realization, with \( B_t^2 = 0, N_t^1 = 0, \forall t \), even if the initial state \( S_1 = s \) is known to the encoder, \( V^{t-1} \) does not uniquely define \( S^t \) because there are two independent noises \( W^{1,n-1} \) and \( W^{2,n} \) that enter the equations of \( S^n \) and \( V^n \). The PO-SS realization is often adopted in many practical problems of engineering and science to realize jointly Gaussian processes \( V^n \).

We should emphasize that for the Case (I) formulation to be consistent with the Cover and Pombara [5] formulation, both the code of Definition 1 and the PO-SS realization of Definition 2 must respect the following two conditions:

(A1) The initial state \( S_1 \) of the noise is not known at the encoder nor the decoder;

(A2) At each \( t \), the representation of the noise \( V^{t-1} \) by the PO-SS realization of Definition 2 does not uniquely determine the state of the noise \( S^t \) and vice-versa, i.e., it is a partially observable realization.

**Case (II) Formulation of Feedback Code and Noise Definitions.** For the Case (II) formulation, we presuppose the following:

**Condition 1.** The initial state of the noise or channel \( S_1 = s \) is known to the encoder and decoder;

**Condition 2.** Given a fixed initial state \( S_1 = s \), known to the encoder and the decoder, at each \( t \), the channel noise \( V^{t-1} \) uniquely defines the state of the noise \( S^t \) and vice-versa.

Thus, for the Case (II) formulation, the code is that of Definition 3, below (hence different from Definition 1).

**Definition 3.** A code with initial state known at the encoder and the decoder

A variant of the code of Definition 1 is feedback code with the initial state of the noise or channel \( S_1 = s \), known to the encoder and decoder strategies, denoted by \((s, 2^{nR}, n)\), \( n = 1, 2, \ldots \).

The code \((s, 2^{nR}, n)\), \( n = 1, 2, \ldots \) is defined as in Definition 2, with (ii), (iii), and (iv) being replaced by...
\( \mathcal{E}^{[1,n]}(\kappa) \triangleq \left\{ X_1 = e_1(W,S_1), X_2 = e_2(W,S_1,Y_1), \ldots, X_n = e_n(W,S_1,X^{n-1},Y^{n-1}) : \right. \\
\left. \frac{1}{n} \mathbb{E} \left\{ \sum_{i=1}^{n} (X_i)^2 \right\} \leq \kappa \right\}, \quad (y^n,s) \mapsto d^n_s(y^n,s) \in \mathcal{M}_n, \tag{26} \\
P_{X^n,Y^n|S_1} = \mathbb{P}_{V^n|S_1} \prod_{t=1}^{n} \mathbb{P}_{X_t|V^{t-1},Y_t,S_t}, \quad \text{that is,} \quad P_{V_t|V^{t-1},X^t,S_t} = \mathbb{P}_{V_t|V^{t-1},S_t}. \tag{27} \\

The feedback capacity is denoted by \( C^{fb}(\kappa,s) \) and should be distinguished from \( C^{fb}(\kappa) \).

The feedback capacity is denoted by \( C^{fb}(\kappa,s) \) if, in addition, Condition 2 holds.

The initial state may include \( S_1 \triangleq (V_0^{\infty}, Y_0^{\infty}) \), etc.

It will become apparent that past studies \([4,7,8]\) considered feedback capacity, \( C^{fb}(\kappa,s) \), and not \( C^{fb}(\kappa,s) \) or \( C^{fb}(\kappa) \).

For the Case (II) formulation, it is obvious (from the converse to the coding theorem) that the optimal channel input conditional distribution is expressed as a function of the state of the noise, \( S^n \), due to (12) and (13).

1.3. Approach of This Paper

Our approach and analysis of information measures (2)–(4), as well as their per unit time limits, is based on the following two step procedure:

**Step # 1.** We apply a linear transformation to the Cover and Pombra optimal channel input process \([5]\) (Equation (11)), (see (33)–(39) to equivalently represent it by a linear functional of the past channel noise sequence, the past channel output sequence, and an orthogonal Gaussian process, i.e., an innovations process. That is, \( X^n \) is uniquely represented, since it is expressed in terms of the orthogonal process.

**Step # 2.** We express the optimal input process by a functional of a *sufficient statistic*, which satisfies a Markov recursion, and an *orthogonal innovations process*. It then follows that the Cover and Pombra characterization of the “\(n\)–block” formula \([5]\) (Equation (10)) (see (33) and (34)) is equivalently represented by a sequential characterization. The problem of feedback capacity is then expressed as the maximization over two sequences of time-varying strategies of the channel input process of the difference of (differential) entropies of the innovations processes of \( Y^n \) and \( V^n \) (analog of entropies in the right-hand side of (2)).

\[
H(Y^n) - H(V^n) = \sum_{t=1}^{n} \left\{ H(Y_t|Y^{t-1}) - H(V_t|V^{t-1}) \right\} \tag{28} \\
= \sum_{t=1}^{n} \left\{ H(Y_t - \mathbb{E}\{Y_t\}|Y^{t-1}) - H(V_t - \mathbb{E}\{V_t\}|V^{t-1}) \right\} \tag{29} \\
\overset{(a)}{=} \sum_{t=1}^{n} \left\{ H(I_t) - H(\hat{I}_t) \right\}, \quad I_t \overset{\triangle}{=} Y_t - \mathbb{E}\{Y_t\}|Y^{t-1}, \quad \hat{I}_t \overset{\triangle}{=} V_t - \mathbb{E}\{V_t\}|V^{t-1} \tag{30} \\
= \sum_{t=1}^{n} \left\{ H(I_t) - H(\hat{I}_t) \right\}, \quad I_t \overset{\triangle}{=} Y_t - \mathbb{E}\{Y_t\}|Y^{t-1}, \quad \hat{I}_t \overset{\triangle}{=} V_t - \mathbb{E}\{V_t\}|V^{t-1} \tag{31} \\
\]

where \((a)\) is due to the Gaussianity of \( Y^n \) and \( V^n \), as well as due the independence of the innovations processes \( I_t \) and \( Y^{t-1} \) and of innovations processes \( \hat{I}_t \) and \( V^{t-1} \).

The asymptotic analysis of \( \lim_{n \to \infty} \frac{1}{n} C^{fb}(\kappa) \) (or with limit and supremum interchanged) is then addressed from the asymptotic properties of entropy rates and the average power,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \left\{ H(I_t) - H(\hat{I}_t) \right\}, \quad \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left\{ \sum_{i=1}^{n} (X_i)^2 \right\} \tag{32} \\
\]
over the channel input distributions, and where the covariance of the innovations process of \( Y^n \) is a functional of the solutions of two generalized matrix DREs. We identify necessary and/or sufficient conditions for the existence of limits, irrespective of whether the noise \( V^n \) is non-stationary, unstable, or stationary. Further, we show that, in general, the characterizations of feedback capacity for the Case (I) and Case (II) formulations are fundamentally different, and we identify conditions based on the feedback codes and noise to coincide.

1.4. Review of Related Literature

Asymptotic feedback capacity formulas and bounds for AGN channels, driven by stationary and asymptotically stationary (often limited memory) noise, are derived since the early 1970s in an anthology of papers based on information theoretic formulas, under various assumptions [7–9,11–14,19–33], in two directions.

(D1) Characterizations and explicit formulas of asymptotic feedback capacity that correspond to feedback codes, when the initial state of the noise (the initial state \( S_1 \) is any a priori information) \( S_1 = s_1 \) is known or not known to the encoder and the decoder;

(D2) Bounds on asymptotic feedback capacity that correspond to linear feedback coding schemes of communicating Gaussian random variables (RVs), \( \Theta : \Omega \to \mathbb{R} \), and coding schemes of communicating digital messages \( W : \Omega \to \{1, 2, \ldots, 2^{nR}\} \), when the initial state the \( S_1 = s_1 \) is known or not known to the encoder and the decoder.

1.4.1. The Cover and Pombra Characterizations of Capacity Pombra [5]

Cover and Pombra characterized the \( n \)-FTFI capacity for non-stationary and non-ergodic noise \( V^n \), [5] (Equation (10)), by (we use \( H(X) \) to denote differential entropy of a continuous-valued RV \( X \); hence, we indirectly assume the probability density functions exist)

\[
C_{ftfi}^n(k) \triangleq \max_{(B^n,K_{XX})} \frac{1}{n} \log \left[ \frac{\det(K_{YY}^{-1}(B^n + I_{n \times n})K_{VV}(B^n + I_{n \times n})^T + K_{ZZ})}{\det(K_{VV})} \right] \leq \kappa \tag{33}
\]

where the distribution \( P_{Y^n} \) is induced by a jointly Gaussian channel input process \( X^n \) [5] (Equation (11)):

\[
X_t = \sum_{j=1}^{t-1} B_t V_j + Z_t, \quad t = 1, \ldots, n, \tag{35}
\]

\[
X^n = B^n V^n + Z^n, \quad Y^n = (B^n + I_{n \times n}) V^n + Z^n, \tag{36}
\]

\( Z^n \) is jointly Gaussian, \( N(0,K_{ZZ}) \), \( Z^n \) is independent of \( V^n \),

\[
X^n \triangleq \left[ \begin{array}{c} X_1 \ X_2 \ \ldots \ \ X_n \end{array} \right]^T \text{ similarly for the rest, } B^n \text{ lower diagonal matrix,} \tag{37}
\]

\[
\frac{1}{n} \mathbf{E} \left\{ \sum_{t=1}^{n} (X_t)^2 \right\} = \frac{1}{n} \mathbf{tr} \left\{ \mathbf{E} \left( X^n (X^n)^T \right) \right\} \leq \kappa. \tag{38}
\]

The notation \( N(0,K_{ZZ}) \) means that the random variable \( Z^n \) is jointly Gaussian with mean \( \mathbf{E}(Z^n) = 0 \) and covariance matrix \( K_{ZZ} = \mathbf{E}(Z^n (Z^n)^T) \), and \( I_{n \times n} \) denotes an \( n \) by the \( n \) identity matrix. Feedback capacity, \( C_{ftfi}(k) \), is characterized by the per unit time limit of the \( n \)-FTFI capacity [5] (Theorem 1).
\[ C_f^b(\kappa) \triangleq \lim_{n \to \infty} \frac{1}{n} C_n^f(\kappa). \] (40)

Over the years, considerable efforts have been devoted to compute \( C_n^f(\kappa) \) and \( C_f^b(\kappa) \) [4,7–9,11,33], often under simplified assumptions on the channel noise. In addition, bounds are described in [27–29], while numerical methods are developed in [31] for time-invariant AGN channels, driven by stationary noise. In [4,7,11,33,34], the authors considered a variant of (40) by interchanging the per unit time limit and maximization operations under the following assumption: the joint process \((X^n, Y^n), n = 1, 2, \ldots\) is either jointly stationary or asymptotically stationary, and the joint distribution of the joint process \((X^n, Y^n), n = 1, 2, \ldots\) is time-invariant. We describe [4,7,8] below.

1.4.2. The Yang, Kavcic and Tatikonda [8] Characterization of Maximal Information Rate

In [8], the authors analyzed the feedback capacity of the AGN channel (1), driven by a stationary noise, described the power spectral density (PSD) functions \( S_V(e^{j\theta}), \theta \in [-\pi, \pi] \):

\[ S_V(e^{j\theta}) \triangleq K_W \left( \frac{1 - \sum_{k=1}^{L} a(k)e^{jk\theta}}{1 - \sum_{k=1}^{L} a(k)e^{-jk\theta}} \right) \left( \frac{1 - \sum_{k=1}^{L} c(k)e^{jk\theta}}{1 - \sum_{k=1}^{L} c(k)e^{-jk\theta}} \right), \quad |c(k)| < 1, \quad |a(k)| < 1, \quad c(k) \neq a(k). \] (41)

The analysis in [8] is based on time-domain methods and corresponds to the Case (II) formulation (see [8] (Section II; in particular, Section I.I.C, I–III, Theorem 1, Section III) by considering a specific time-invariant, stable, state space realization of the noise PSD (41), such that Conditions (II.1), and (II.2) hold, i.e.,

The initial state of the noise, \( S_1 = s \), is known to the encoder and the decoder, and the initial state and noise \((s, V^{t-1})\) uniquely define the noise state \( S_1 \), and vice-versa, for all \( t \).

The time-domain characterization of feedback capacity, called the maximal information rate [8] (Theorem 6), corresponds to the supremum and limit being interchanged, and involves only one matrix Riccati equation of the linear filtering theory. However, ref. [8] (Theorem 6) does not state the conditions based on which the maximal information rate is valid (i.e., existence of asymptotic limit).

1.4.3. The Kim [4] Characterization of Feedback Capacity

The author in [4] also analyzed the feedback capacity of the AGN channel (1), driven by stationary noise described by the PSD (41) and by a time-invariant, stable, state space realization of the noise \( V^n \) (see [4] (Section VI)). A major point of confusion is that the characterization of feedback capacity in the time domain [4] (Theorem 6.1) does not state whether this corresponds to Case (I) or Case (II) formulations. The reader, however, can verify from [4] (Lemma 6.1 and comments above it) that the time-domain characterization of feedback capacity [4] (Theorem 6.1) corresponds to the Case (II) formulation, as stated in the study by Yang, Kavcic and Tatikonda [8]. In fact, the characterization of feedback capacity [4] (Theorem 6.1) involves only one Riccati equation of the linear filtering theory, as in [8]. We reconfirm this point at various parts of this paper (see, for example, Section 2.6).

1.4.4. The Gattami [7] Characterization of Feedback Capacity and Semi-Definite Programming Formulation

The authors in [7] revisited the feedback capacity of the AGN channel (1) driven by a stationary noise described by the PSD (41) and with a time-invariant, stable, state space realization of the noise \( V^n \). One of the main results of [7] is the feedback capacity characterization for the Case (II) formulation, i.e., that involves only one matrix Riccati
equation of the filtering theory, precisely as in [4,8]. Another main result of [7] is the
re-formulation of the optimization problem using semi-definite programming.

In the following remark, we will provide additional insights into the results of [4,7,8].

**Remark 1.** On the formulas in [4,7,8].

Refs. [4,7] considered the stable, time-invariant PO-SS realization of Definition 2, with $W_t : \Omega \to \mathbb{R}, N_t = 1, \forall t$, i.e., $n_w = 1$.

In [4], (Theorem 6.1) (see [4] (above and below Equation (70))), the asymptotic characterization of feedback capacity involves one filtering matrix ARE and is achieved by a channel input, which is different at times $n = 1, 2, \ldots, n_s$ from subsequent time $n = n_s + 1, \ldots$, given by

$$X_n = Z_n \in N(0, K_Z), K_Z > 0, \quad n = 1, \ldots, n_s,$$

$$X_n = \chi \left( S_n - E \left\{ S_n | Y^{n-1} \right\} \right), \quad n = n_s + 1, n_s + 2, \ldots,$$

where $\chi$ is a constant vector, and $Z_n, n = 1, 2, \ldots, n_s$ are IID Gaussian RVs.

In [7] (Theorem 3), the asymptotic characterization of feedback capacity involves one filtering matrix ARE and is incurred by the time-invariant channel input

$$X_n = \chi \left( S_n - E \left\{ S_n | Y^{n-1} \right\} \right) + Z_n, \quad Z_n \in N(0, K_Z), \quad K_Z \geq 0, \quad n = 1, 2, \ldots$$

$Z_n, n = 1, 2, \ldots$, are IID Gaussian RVs.

In [8], a state space realization of the PSD is considered, and the maximal information rate is presented, (Equation (125)) and [8] (Theorem 6, Equation (138), $l_{max}$), which involves one filtering matrix ARE. It is achieved by

$$X_n = \chi \left( S_n - E \left\{ S_n | Y^{n-1} \right\} \right) + Z_n, \quad Z_n \in N(0, K_Z), \quad K_Z \geq 0, \quad n = 1, 2, \ldots,$$

where $Z_n, n = 1, 2, \ldots$ are IID Gaussian RVs.

The above references computed the feedback capacity and maximal information rate for the ARMA $(a, c), a \in [-1, 1], c \in (-1, 1), c \neq a$ and arrived at the conclusion that it is precisely Butman’s [13] and Wolfowitz’s [14] lower bound.

It will become apparent that [4,7,8] arrived at the above expressions by considering problem $C^{fb,S}(s, \kappa)$, and not $C^{fb}(s, \kappa)$ or $C^{fb}(\kappa, s)$.

Sequential equivalent characterizations of the Cover and Pombra [5] n–FTFI capacity, $C_n^{fb}(\kappa)$ and $C^{fb}(\kappa)$ capacity of the AGN channel (1) driven by a non-stationary and non-ergodic Gaussian noise $Y^n$, as well as by time-varying, unstable (and stable) state space realizations of Definition 2, are derived in [1–3,10], including relations between Case (I) and Case (II) formulations. In particular, ref. [10] proved that $C_n^{fb}(\kappa)$ of the AGN channel with state space noise of Definition 2 involves two matrix Riccati equations of the linear filtering theory and that, under certain conditions, $C_n^{fb}(\kappa)$ reduces to $C_n^{fb,S}(\kappa, s)$, which involves only one matrix Riccati equation. Corresponding expressions are obtained for their per unit time asymptotic limits. The methods and results of [10] are generalized to multiple-input multiple-output (MIMO) Gaussian channels in [3]. Further, ref. [2] derived closed-form expressions of $C^{fb}(\kappa)$ for AGN channels driven by the stable and unstable ARMA $(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty), c \neq a$ noise, showing the connection between Case (I) and Case (II) formulations. Ref. [3] generalized the earlier investigation in [1], which considered the autoregressive unit memory stable and unstable noise. An investigation of nonfeedback capacity of stable and unstable noise is outlined in [35]. The
connection of ergodic theory and feedback capacity of unstable channels is discussed in [36,37].

MIMO Gaussian channels are also investigated in [38]. However, many of the expressions in [38] are previously obtained in [3,10]. The analysis in [38] does not include a derivation of the optimal channel input that achieves $C_{n}^{fb}(\kappa)$ and $C^{fb}(\kappa)$, and does not discuss the connection between Case (I) and Case (II) formulations. The closed-form expressions of capacity of the examples in [38], are special cases of expressions in [1] and in [2], which treated the noise ARMA $(a,c), \forall a \in (-\infty,\infty), \forall c \in (-\infty,\infty), c \neq a$.

We structure this paper as follows.

In Section 2, we derive the new sequential characterizations of the $n$–FTFI capacity for the Cover and Pombra formulation of feedback capacity of the AGN channel (1), i.e., for the Case (I) formulation, $C_{n}^{fb}(\kappa)$. We also derive analogous sequential characterizations for $C_{n}^{fb}(\kappa, s)$ and for the Case (II) formulation, $C_{n}^{fb,S}(\kappa, s)$, i.e., when Conditions 1 and 2 hold, to illustrate their fundamental differences.

In Section 3, we present the asymptotic analysis of feedback capacity for the Case (I) formulation. In Section 4, we treat the Case (II) formulation.

This paper contains several examples and makes comparisons to the existing literature.

### 2. Sequential Characterizations of $n$–FTFI Capacity for Case (I) Formulation

In this section, we derive equivalent sequential characterizations for the following:

(i) $C_{n}^{fb}(\kappa)$ defined by (2) of the Case (I) formulation, i.e., for the Cover and Pombra $n$–FTFI capacity characterization (34);

(ii) $C_{n}^{fb}(\kappa, s)$ defined by (3), as a degenerated case of $C_{n}^{fb}(\kappa)$;

(iii) $C_{n}^{fb,S}(\kappa, s)$ defined by (4) of the Case (II) formulation, as a degenerated case of $C_{n}^{fb}(\kappa)$.

We organize the presentation of the material as follows.

(1) Section 2.1. Here, we introduce our notation.

(2) Section 2.2. The main result is Theorem 1, which provides an equivalent sequential characterization of the Cover and Pombra characterization of the $n$–FTFI capacity, $C_{n}^{fb}(\kappa)$, i.e., of (33) and (34). Our derivation proceeds as follows. We apply a linear transformation to the Cover and Pombra Gaussian optimal channel input $X^{n}$ (35), to represent $X_{t}$, with a linear function of $(V^{t-1}, Y^{t-1})$ or equivalently $(X^{t-1}, Y^{t-1})$ and an orthogonal Gaussian innovations process $Z_{t}$, which is independent of $(Z^{t-1}, X^{t-1}, V^{t-1}, Y^{t-1})$ for $t = 1, \ldots, n$. Subsequently, we apply Theorem 1 to the time-varying PO-SS$(a_{c}, c_{t}, b_{t}^{1}, b_{t}^{2}, d_{t}^{1}, d_{t}^{2})$ noise (see Example 1) to the non-stationary autoregressive moving average, ARMA $(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise, and to the stationary ARMA $(a,c), a \in (-1,1), c \in (-1,1)$ noise (see Example 2), which is found in many references, such as [4]. It will become apparent that our characterizations of $n$–FTFI capacity are fundamentally different from past studies.

(3) Section 2.3. The main result is Theorem 3, which provides a simplified characterization of the sequential characterization of the $n$–FTFI capacity, $C_{n}^{fb}(\kappa)$, given in Theorem 1 (i.e., the equivalent of (34)), for the time-varying AGN channel (1) driven by the PO-SS realization of Definition 2, for the code of Definition 1. The $n$–FTFI capacity of Theorem 3 is expressed in terms of solutions to two DREs. Our derivation is based on identifying a finite-dimensional sufficient statistic to express $X_{t}$ as a functional of the sufficient statistic, instead of $(V^{t-1}, Y^{t-1})$ or $(X^{t-1}, Y^{t-1})$, and an orthogonal Gaussian innovations process.

(4) Section 2.4. The main results is Corollary 6, which is an application of Theorem 3 (i.e., the sufficient statistic representation), to the ARMA $(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Example 2. This example shows that the $n$–FTFI capacity is expressed in terms of
solutions to two DREs.

From Corollary 6, the following will become apparent:

(i) Neither the time-domain characterization [4] (Theorem 6.1) (see [4]) (Theorem 5.3) nor the frequency domain characterization [4] (Theorem 4.1) correspond to the Cover and Pombra characterization of feedback capacity.

(ii) Notation

Throughout this paper, we use the following notation:

- \( \mathbb{Z} \triangleq \{ \ldots, -1, 0, 1, \ldots \} \)
- \( \mathbb{Z}_+ \triangleq \{ 1, 2, 3, \ldots \} \)
- \( \mathbb{R}^n_+ \triangleq \{ x \in \mathbb{R}^n : x \geq 0 \} \)
- \( \mathbb{D}_o \triangleq \{ c \in \mathbb{C} : |c| < 1 \} \)
- \( \mathbb{S}_+^{n \times n} \triangleq \{ A \in \mathbb{R}^{n \times n} : A \succeq 0 \} \)
- \( \mathbb{S}_+^{n \times n} \triangleq \{ A \in \mathbb{R}^{n \times n} : A \succeq 0 \} \)
- \( \mathbb{S}_+^{n \times n} \triangleq \{ A \in \mathbb{R}^{n \times n} : A \succeq 0 \} \)

where the last equality is due to a property of jointly Gaussian distributed RVs.

(iii) Positive definiteness by \( A \succeq 0 \) if for all \( w \in \mathbb{R}^n \), \( w^T A w \geq 0 \). Positive semidefiniteness is denoted by \( A \succeq 0 \) and (strict) positive definiteness by \( A \succ 0 \).

(iv) \( I_{n \times n} \in \mathbb{S}_+^{n \times n} \), \( n \in \mathbb{Z}_+ \) denotes the identity matrix, \( tr(A) \) denotes the trace of any matrix \( A \in \mathbb{R}^{n \times n} \).

(v) If \( A \in \mathbb{S}_+^{n \times n} \), then \( \text{spec}(A) \subset \mathbb{C} \) is the Spectrum of a matrix \( A \in \mathbb{R}^{n \times n} \), \( q \in \mathbb{Z}_+ \) (the set of all its eigenvalues).

A matrix \( A \in \mathbb{R}^{n \times n} \) is called exponentially stable if all its eigenvalues are within the unit open disc, i.e., \( \text{spec}(A) \subset \mathbb{D}_o \).

(vi) \( (\mathbb{O}, \mathcal{F}, \mathbb{P}) \) denotes a probability space. Given a random variable \( X : \mathbb{O} \rightarrow \mathbb{R}^n \), its induced distribution on \( \mathbb{R}^n \) is denoted by \( \mathbb{P}_X \).

\( \mathbb{P}_X \in \mathcal{N}(\mu_X, K_X) \), where \( K_X \succeq 0 \) denotes a Gaussian distributed RV \( X \), with mean value \( \mu_X \) and covariance matrix \( K_X = \text{cov}(X, X) \succeq 0 \), defined by

\[
\mu_X \triangleq \mathbb{E}(X), \quad K_X = \text{cov}(X, X) \triangleq \mathbb{E}\left\{ (X - \mathbb{E}\{X\})(X - \mathbb{E}\{X\})^T \right\}.
\] (47)

Given another Gaussian random variable \( Y : \mathbb{O} \rightarrow \mathbb{R}^n \), \( n_Y \in \mathbb{Z}_+ \), which is jointly Gaussian distributed with \( X \), i.e., the joint distribution is \( \mathbb{P}_{X,Y} \), the conditional covariance of \( X \) given \( Y \) is defined by

\[
K_{X|Y} = \text{cov}(X, X|Y) \triangleq \mathbb{E}\left\{ (X - \mathbb{E}\{X|Y\})(X - \mathbb{E}\{X|Y\})^T \right\}_{|Y} \] (48)

\[
= \mathbb{E}\left\{ (X - \mathbb{E}\{X|Y\})(X - \mathbb{E}\{X|Y\})^T \right\}_{|Y}
\] (49)

where the last equality is due to a property of jointly Gaussian distributed RVs.

Given three arbitrary RVs \( (X, Y, Z) \) with induced distribution \( \mathbb{P}_{X,Y,Z} \), the RVs \( (X, Z) \) are called conditionally independent given the RV \( Y \) if \( \mathbb{P}_{Z|X,Y} = \mathbb{P}_{Z|Y} \). This conditional independence is often denoted by \( X \leftrightarrow Y \leftrightarrow Z \), which is a Markov chain.
2.2. Preliminary Characterizations of n–FTTI Capacity of AGN Channels Driven by Correlated Noise

We start with preliminary calculations for the feedback code of Definition 1, which we use to prove Theorem 1. These calculations are introduced for the sake of clarity and to establish our notation.

For the feedback code of Definition 1, by the channel definition (1), i.e., (18), the conditional distribution of $Y_t$, given $Y^{t-1} = y^{t-1}, X^t = x^t$, is

$$
\mathbb{P}\{Y_t \in dy \mid Y^{t-1} = y^{t-1}, X^t = x^t\} \overset{(a)}{=} \mathbb{P}\{Y_t \in dy \mid Y^{t-1} = y^{t-1}, X^t = x^t, V^{t-1} = v^{t-1}\}
$$

$$
\overset{(b)}{=} \mathbb{P}_{W_i|V^{t-1}}(v_t : x_t + v_t \in dy), \ t = 2, \ldots, n
$$

$$
= \mathbb{P}_{Y_t|X_t, V^{t-1}}
$$

$$
= \mathbb{P}_t(dy|x_t, v^{t-1})
$$

$$
\mathbb{P}\{Y_t \in dy \mid Y^0 = y^0, X^1 = x^1\} = \mathbb{P}_{Y_t|X_t} \equiv \mathbb{P}_t(dy|x_1).
$$

where (a) is due to (1) and (b) is due to (18). We introduce the set of channel input distributions with feedback, which are consistent with the code of Definition 1, not necessarily generated by the messages $W$, as follows:

$$
P_{[1,n]}(\kappa) \triangleq \left\{ \mathbb{P}_t(dx_t|x^{t-1}, y^{t-1}) \overset{\Delta}{=} \mathbb{P}_{X_t|x^{t-1}, y^{t-1}}, t = 1, \ldots, n : \frac{1}{n} \mathbb{E}\left( \sum_{i=1}^{n} (X_t)^2 \right) \leq \kappa \right\}.
$$

By Definition 1, we have $E_{[1,n]}(\kappa) \subseteq P_{[1,n]}(\kappa)$. Moreover, by the channel definition, any pair of the sequence triple $(V^t, X^t, Y^t)$ uniquely defines the remaining sequence. Thus, the identity holds:

$$
P_{[1,n]}(\kappa) \overset{\Delta}{=} \left\{ \mathbb{P}_t(dx_t|v^{t-1}, y^{t-1}), t = 1, \ldots, n : \frac{1}{n} \mathbb{E}\left( \sum_{i=1}^{n} (X_t)^2 \right) \leq \kappa \right\} = P_{[1,n]}(\kappa).
$$

We also emphasize that, by Definition 1, for a given feedback encoder strategy $e(\cdot) \in E_{[1,n]}(\kappa)$, i.e., $x_1 = e_1(w), x_2 = e_2(w, x_1, y_1), \ldots, x_n = e_n(w, x^{n-1}, y^{n-1})$ the conditional distributions of $Y_t$ given $(Y^{t-1}, W) = (y^{t-1}, w)$ is obtained as follows:

$$
P_{Y_t|W, Y^{t-1}}(dy_t, y^{t-1}, w) \overset{(a)}{=} \mathbb{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, y^{t-1}, w)
$$

$$
\overset{(b)}{=} \mathbb{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, y^{t-1}, v^{t-1}, w)
$$

$$
\overset{(c)}{=} \mathbb{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, v^{t-1}, w)
$$

$$
\overset{(d)}{=} \mathbb{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, v^{t-1})
$$

$$
\overset{(e)}{=} \mathbb{P}_t(dy_t|e_1(w, x^{t-1}, y^{t-1}), v^{t-1})
$$

(a) is due to knowledge of the distribution of the strategies $e_j(\cdot), j = 1, \ldots, t$, the code definition, and the recursive substitution, $x_1 = e_1(w), x_2 = e_2(w, x_1, y_1), \ldots, e_t(w, x^{t-1}, y^{t-1})$, where $x^{t-1}$ is specified by the knowledge of the strategies, $e_j(\cdot), j = 1, \ldots, t-1$ and the knowledge of $(y^{t-2}, w)$,

(b) is due to knowing $x_j = e_j(w, x^{j-1}, y^{j-1}), y_j, j = 1, \ldots, t-1$ specifies $v_j = y_j - x_j, j = 1, \ldots, t-1$,

(c) is due to the fact that, any pair of the triple $(x^t, y^t, v^t)$ specifies the remaining sequence, i.e., knowing $(x^{t-1}, v^{t-1})$ specifies $y^{t-1}$, and $y^{t-1}$ is thus redundant,
(d) is due to the conditional independence \( P_{V_l|Y_{l-1},X_l'} = P_{V_l|Y_{l-1}} \).

(e) is due to (18), i.e., \( P_{V_l|Y_{l-1},X_l'} = P_{V_l|Y_{l-1}} \) and the channel definition.

By the channel definition \( Y_t = X_t + V_t, t = 1, \ldots, n \), each \( e(\cdot) \in \mathcal{E}_{[0,n]}(\kappa) \) is also expressed as

\[
\begin{align*}
  x_1 &= e_1(w) = \tilde{r}_1(w), \quad x_2 = e_2(w, x_1, y_1) = \tilde{e}_2(w, x_1, v_1, y_1), \ldots, \\
  x_n &= e_n(w, x^{n-1}, y^{n-1}) = \tilde{e}_n(w, x^{n-1}, v^{n-1}, y^{n-1}), \quad w \in \mathcal{M}^{(n)}.
\end{align*}
\]

(61)

where (a) is due to the channel definition—i.e., the presence of \( x^t \) in \( \tilde{e}_i(\cdot, v^{t-1}, \cdot) \) can be removed, since it is redundant—and specified by \( (v^{t-1}, y^{t-1}) \). Consequently, we have the identity

\[
\mathcal{E}_{[0,n]}(\kappa) \triangleq \left\{ x_1 = \tilde{r}_1(w), x_2 = \tilde{e}_2(w, v_1, y_1), \ldots, x_n = \tilde{e}_n(w, v^{n-1}, y^{n-1}) : \right. \\
\left. \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^{n} (X_i)^2 \right] \leq \kappa \right\} = \mathcal{E}_{[0,n]}(\kappa).
\]

(62)

**Notation 1.** For the feedback code of Definition 3, with initial state \( S_1 = s \), known to the encoder and the decoder, the above sets \( \mathcal{P}_{[0,n]}(\kappa), \mathcal{E}_{[0,n]}(\kappa), \mathcal{E}_{[1,n]}(\kappa), \mathcal{E}_{[1,n]}^s(x) \) are replaced by \( \mathcal{P}_{[0,n]}^s(\kappa), \mathcal{P}_{[1,n]}^s(\kappa), \mathcal{E}_{[1,n]}^s(\kappa), \mathcal{E}_{[1,n]}^{s}\left(x_{1:n}\right) \) to indicate that the distributions and codes are \( \mathcal{P}_s(dx_t|v^{t-1}, y^{t-1}, s), t = 1, \ldots, x_1 = \tilde{r}_1(w, s), x_2 = \tilde{e}_2(w, v_1, y_1), \ldots, x_n = \tilde{e}_n(w, v^{n-1}, y^{n-1}, s), etc., and these depend on \( s \).

In the next theorem, we present our preliminary equivalent sequential characterization of the Cover and Pombra characterization \( \mathcal{C}_{[0,n]}(\kappa) \), i.e., of (33), under encoder strategies \( \mathcal{E}_{[0,n]}(\kappa) = \mathcal{E}_{[0,n]}(\kappa) \) and channel input distributions \( \mathcal{P}_{[0,n]}(\kappa) = \mathcal{P}_{[0,n]}(\kappa) \). Unlike the Cover and Pombra [5] realization of \( X^n \), given by (35), at each time \( t \), \( X_t \) is driven by an orthogonal Gaussian process \( Z_t \).

**Theorem 1.** Information structures of maximizing distributions for AGN Channels

Consider the AGN channel (1), i.e., with noise distribution \( P_{V_n} \), and the code of Definition 1. Then, the following hold:

(a) The following inequality holds:

\[
\sup_{\mathcal{E}_{[0,n]}(\kappa)} \sum_{t=1}^{n} H^F(Y_t|Y^t-1) \leq \sup_{\mathcal{P}_{[0,n]}(\kappa)} \sum_{t=1}^{n} H^F(Y_t|Y^t-1)
\]

(63)

where the conditional (differential) entropy \( H^F(Y_t|Y^t-1) \) is evaluated with respect to the probability distribution \( P^F_t(dy_t|y^{t-1}) \), defined by

\[
P^F_t(dy_t|y^{t-1}) = \int P_t(dy_t|x_t, v^{t-1}, y^{t-1}) P^F_t(dw|y^{t-1}), \quad t = 0, \ldots, n.
\]

(64)

and \( H^F(Y_t|Y^t-1) \) is evaluated with respect to the probability distribution \( P^F_t(dy_t|y^{t-1}) \), defined by

\[
P^F_t(dy_t|y^{t-1}) = \int P_t(dy_t|x_t, v^{t-1}) P^F_t(dx_t|v^{t-1}, y^{t-1}) P^F_t(dw|y^{t-1}), \quad t = 0, \ldots, n.
\]

(65)

(b) The optimal channel input distribution \( \{ P_t(dx_t|v^{t-1}, y^{t-1}), t = 1, \ldots, n \} \in \mathcal{P}_{[1,n]}(\kappa) \), which maximizes \( \sum_{t=1}^{n} H^F(Y_t|Y^t-1) \) of part (a), i.e., the right-hand side of (63), is induced by an input
process $X^n$, which is conditionally Gaussian, with linear conditional mean, nonrandom conditional covariance and given by

$$
\mathbb{E}^{\mathbb{P}} \left\{ X_t \mid \mathcal{Y}^{t-1} \right\} = \left\{ \begin{array}{ll}
\Gamma_1^n \mathcal{V}^{t-1} + \Gamma_2^n \mathcal{Y}^{t-1}, & \text{for } t = 2, \ldots, n \\
0, & \text{for } t = 1,
\end{array} \right. \tag{66}
$$

and such that the average constraint holds and (18) is respected.

(c) The optimal channel input distribution $\{\mathbb{P}(dx_t|\mathcal{Y}^{t-1}), t = 1, \ldots, n\} \in \mathbb{P}_{[1,n]}(\kappa)$ of part (b) is induced by a jointly Gaussian process $X^n$, with a realization given by

$$
X_t = \sum_{j=1}^{t-1} \Gamma_{tj} V_j + \sum_{j=1}^{t-1} \Gamma_{tj}^2 Y_j + Z_t, \quad X_1 = Z_1, \quad t = 2, \ldots, n, \tag{68}
$$

$$
= \Gamma_1^n \mathcal{V}^{t-1} + \Gamma_2^n \mathcal{Y}^{t-1} + Z_t, \tag{69}
$$

$Z_t \in N(0, K_{Z_t}), \quad t = 1, \ldots, n$ a Gaussian sequence,

$Z_t$ independent of $(\mathcal{V}^{t-1}, \mathcal{Y}^{t-1}), \quad t = 1, \ldots, n,$ \tag{70}

$Z^n$ independent of $\mathcal{V}^n$, \tag{71}

$$
\frac{1}{n} \mathbb{E}^{\mathbb{P}} \left\{ \sum_{i=1}^{n} (X_t)^2 \right\} \leq \kappa, \tag{73}
$$

$$(\Gamma_1^n, \Gamma_2^n, K_{Z_t}) \in (-\infty, \infty) \times (-\infty, \infty) \times [0, \infty)$ nonrandom. \tag{74}

(d) An equivalent characterization of the $n$–FTFI capacity $C_n^{fb}(\kappa)$, defined by (33) and (34), is given by

$$
C_n^{fb}(\kappa) = \sup_{\kappa} \mathbb{E}^{\mathbb{P}} \left\{ \sum_{i=1}^{n} H^{\mathbb{P}}(Y_i | \mathcal{Y}^{i-1}) - H(\mathcal{V}^{n}) \right\} \tag{75}
$$

$$
= \sup_{\kappa} \mathbb{E}^{\mathbb{P}} \left\{ \sum_{i=1}^{n} \left\{ H^{\mathbb{P}}(I_i) - H(\mathcal{I}_i) \right\} \right\} \tag{76}
$$

where $I_t$ and $\mathcal{I}_t$ are innovations processes defined by

$$
I_t \overset{\triangle}{=} Y_t - \mathbb{E}^{\mathbb{P}} \left\{ Y_t \mid \mathcal{Y}^{t-1} \right\}, \quad \mathcal{I}_t \overset{\triangle}{=} V_t - \mathbb{E} \{ V_t \mid \mathcal{Y}^{t-1} \} \tag{77}
$$

where parts (b) and (c) hold, and where the supremum is over all $(\Gamma_1^n, \Gamma_2^n, K_{Z_t}), t = 1, \ldots, n$ of the realization of part (c), which induces the distribution $\mathbb{P}_t(dx_t|\mathcal{Y}^{t-1}, y^{t-1}), t = 1, \ldots, n$.

**Proof.** See Appendix A.1. \(\square\)

**Remark 2.** For the code of Definition 3 that assumes knowledge of the initial state $S_1 = s$, it is easy to verify that $C_n^{fb}(\kappa, s)$ is directly obtained from Theorem 1, as a degenerate case (an independent derivation is easily produced following the derivation of Corollary 11, with slight variations).

By utilizing Theorem 1, we can derive the converse coding theorems stated below for the feedback codes of Definitions 1 and 3.
Theorem 2. Converse coding theorems for codes of Definitions 1 and 3
Consider the AGN channel (1).
(a) Any achievable rate $R$ for the code of Definition 1 satisfies

$$R \leq C^f_b(\kappa) \triangleq \lim_{n \to \infty} \frac{1}{n} C^f_b(n\kappa),$$

$$C^f_b(n\kappa) = \sup_{P_{Y^n|X^n}} \sum_{t=1}^n H(Y_t|Y^{t-1}) - H(V^n)$$

provided the supremum exists and the limit exists, where the right-hand side of (79) is given in Theorem 1(d).

(b) Any achievable rate $R$ for the code of Definition 3 (with initial state $s = s_1$) satisfies

$$R \leq C^f_b(\kappa, s) \triangleq \lim_{n \to \infty} \frac{1}{n} C^f_b(n\kappa, s),$$

$$C^f_b(n\kappa, s) = \sup_{P_{Y^n|X^n}} \sum_{t=1}^n H(Y_t|Y^{t-1}, s) - H(V^n|s).$$

where $E^n_\mathbb{E}[\cdot]$ means the expectation is for a fixed $s = s_1$, provided the supremum exists and the limit exists, and where the right-hand side of (81) is obtained from Theorem 1, part (d), by replacing all conditional distributions, entropies, etc., for fixed initial state $S_1 = s$ (see Notation 1).

Proof. Following on from standard arguments, we use Fano’s inequality (see also [5]) and Theorem 1. □

In the next remark, we clarify the equivalence of Theorem 1(d) to Cover and Pombra [5].

Remark 3. Relation of Theorem 1 and Cover and Pombra [5]
(a) From the realization of $X^n$ given by (68), we can recover the Cover and Pombra [5] realization (35) by recursive substitution of $Y^{t-1}$ into the right-hand side of (68), as follows:

$$X_t = \sum_{j=1}^{t-1} \Gamma_{t,j} V_j + \sum_{j=1}^{t-1} \Gamma_{t,j}^2 Y_j + Z_t$$

$$= \sum_{j=1}^{t-1} \Gamma_{t,j} V_j + \sum_{j=1}^{t-2} \Gamma_{t,j}^2 Y_j + \Gamma_{t,t-1}^2 (X_{t-1} + Z_{t-1}) + Z_t$$

$$= \sum_{j=1}^{t-1} B_{t,j} V_j + Z_t, \quad \text{by recursive substitution of } X_1, \ldots, X_{t-1}, Y_1, \ldots, Y_{t-2}$$

for some $Z_t \in (0, K^*_Z)$ which is jointly correlated, and some nonrandom $B_{t,j}$, as given by (35) and (36).

(b) Unlike the Cover and Pombra [5] realization of $X^n$, i.e., (35), the realization of $X^n$ given by (68) or in vector form by (69) is such that, at each time $t$, $X_t$ depends on $(V^{t-1}, Y^{t-1}, Z_t)$ or in vector form on $(V^{t-1}, Y^{t-1}, Z_t)$, where $Z_t$ is an innovations or orthogonal process, i.e., (71) holds.

(c) In subsequent parts of the paper, we derive an equivalent sequential characterization of the Cover and Pombra $n$–FTFI capacity (34), which is simplified further with the use of a sufficient statistic (that satisfies a Markov recursion).
To characterize $C^T \mu (k)$ using Theorem 1, part (d), we need to compute the (differential) entropy $H(V^n)$ of $V^n$. The following lemma is useful in this respect:

**Lemma 1.** Entropy $H(V^n)$ calculation from generalized Kalman filter of the PO-SS noise realization. Consider the PO-SS realization of $V^n$ of Definition 2. Define the conditional covariance and conditional mean of $S_t$ given $V^{t-1}$ by

$$\Sigma_t \triangleq \text{cov}(S_t, S_t | V^{t-1}) = E \left\{ (S_t - \hat{S}_t) (S_t - \hat{S}_t)^T | V^{t-1} \right\}, \quad \hat{S}_t \triangleq E \left\{ S_t | V^{t-1} \right\}, \quad t = 2, \ldots, n.$$  \hspace{1cm} (85)

$$\Sigma_1 \triangleq \text{cov}(S_1, S_1) = K_{S_1}, \quad S_1 \triangleq \mu_{S_1}.$$  \hspace{1cm} (86)

Then, the following hold:

(a) The conditional distribution of $V_t$ conditioned on $V^{t-1}$ is Gaussian, i.e.,

$$P_{V_t | V^{t-1}} \in N(\mu_{V_t | V^{t-1}}, K_{V_t | V^{t-1}}), \quad t = 1, \ldots, n.$$  \hspace{1cm} (87)

where $\mu_{V_t | V^{t-1}} \triangleq E \{ V_t | V^{t-1} \}, K_{V_t | V^{t-1}} \triangleq \text{cov} \{ V_t, V_t | V^{t-1} \}$.

(b) The optimal mean and covariance $\mu_{V_t | V^{t-1}}, K_{V_t | V^{t-1}}$ are given by the generalized Kalman filter recursions, as follows:

(i) The optimal mean square error estimate $\hat{S}_t$ satisfies the generalized Kalman filter recursion

$$\hat{S}_{t+1} = A_t \hat{S}_t + M_t (\Sigma_t) \hat{I}_t, \quad \hat{S}_1 = \mu_{S_1},$$  \hspace{1cm} (88)

$$M_t (\Sigma_t) \triangleq \left( A_t \Sigma_t C_t^T + B_t K_{W_t} N_t^T \right) \left( N_t K_{W_t} N_t^T + C_t \Sigma_t C_t^T \right)^{-1},$$  \hspace{1cm} (89)

$$\hat{I}_t \triangleq V_t - E \{ V_t | V^{t-1} \} = V_t - C_t \hat{S}_t = C_t (S_t - \hat{S}_t) + N_t W_t, \quad t = 1, \ldots, n,$$  \hspace{1cm} (90)

$$\hat{I}_t \in N(0, K_{I_t}), \quad t = 1, \ldots, n \text{ is an orthogonal innovations process, i.e., } \hat{I}_t \text{ is independent of } \hat{I}_s \text{ for all } t \neq s, \text{ and } \hat{I}_t \text{ is independent of } V^{t-1},$$  \hspace{1cm} (91)

$$K_{I_t} \triangleq \text{cov} \{ \hat{I}_t, \hat{I}_t \} = C_t \Sigma_t C_t^T + N_t K_{W_t} N_t^T,.$$  \hspace{1cm} (92)

(ii) The error $E_t \triangleq S_t - \hat{S}_t$ satisfies the recursion

$$E_{t+1} = M_t^{CL} (\Sigma_t) E_t + \left( B_t - M(\Sigma_t) N_t \right) W_t, \quad E_1 = S_1 - \hat{S}_1, \quad t = 1, \ldots, n,$$  \hspace{1cm} (93)

$$M_t^{CL} (\Sigma_t) \triangleq A_t - M(\Sigma_t) C_t.$$  \hspace{1cm} (94)

(iii) The covariance of the error is such that $E \{ E_t E_t^T \} = \Sigma_t$ and satisfies the generalized matrix DRE

$$\Sigma_{t+1} = A_t \Sigma_t A_t^T + B_t K_{W_t} B_t^T - \left( A_t \Sigma_t C_t^T + B_t K_{W_t} N_t^T \right) \left( N_t K_{W_t} N_t^T + C_t \Sigma_t C_t^T \right)^{-1}$$

$$\cdot \left( A_t \Sigma_t C_t^T + B_t K_{W_t} N_t^T \right)^T, \quad t = 1, \ldots, n, \quad \Sigma_1 = K_{S_1} \succeq 0, \quad \Sigma_t \succeq 0.$$  \hspace{1cm} (95)

(iv) The conditional mean and covariance $\mu_{V_t | V^{t-1}}, K_{V_t | V^{t-1}}$ are given by

$$\mu_{V_t | V^{t-1}} = C_t \hat{S}_t, \quad t = 1, \ldots, n,$$  \hspace{1cm} (96)

$$K_{V_t | V^{t-1}} = K_{I_t} = C_t \Sigma_t C_t^T + N_t K_{W_t} N_t^T, \quad t = 1, \ldots, n.$$  \hspace{1cm} (97)
(v) The entropy of $V^n$ is given by
\[
H(V^n) = \sum_{i=1}^{n} H(I_i) = \frac{1}{2} \sum_{i=1}^{n} \log \left( 2\pi e \left[ C_i \Sigma_i C_i^T + N_i K_{wi} N_i^T \right] \right)
\]  
(98)

**Proof.** (a,b),(i–iv). The generalized Kalman filter of the PO-SS realization of $V^n$ and accompanied statements can be found in many textbooks, i.e., [16]. However, it is noted that $I_t, t = 2, \ldots, n, I_1 = V_1$ are all independent Gaussian. For example, to show (93), we write the recursion for $E_s = S_t - \hat{S}_t$ using part (i) and the realization of $S_t$, part (b). (v) By the chain rule of joint entropy, we have
\[
H(V^n) = H(V_1) + \sum_{t=2}^{n} H(V_t|V^{t-1})
\]  
(99)
\[
= H(V_1) + \sum_{t=2}^{n} H(V_t - \mathbb{E}\{V_t|V^{t-1}\}|V^{t-1})
\]  
(100)
\[
= H(V_1) + \sum_{t=2}^{n} H(I_t), \text{ by orthogonality of } I_t \overset{\Delta}{=} V_t - \mathbb{E}\{V_t|V^{t-1}\} \text{ and } V^{t-1}
\]  
(101)

From (101) and (97), we have (98) from the entropy formula of Gaussian RVs. 

The next corollary of the entropy $H(V^n|s)$ follows directly from Lemma 1 when $S_1 = s$ is fixed.

**Corollary 1.** Conditional entropy $H(V^n|s), S_1 = s$ of the PO-SS noise realization.
Consider the PO-SS realization of $V^n$ of Definition 2, for fixed $S_1 = s$, and denote the state process generated by recursion (19) via (we often use the notation $S_t = S^s_t$ to emphasize that the $S_t$ process is generated for $S_1 = S^s_1 = s$ fixed) $S_t = S^s_t, t = 2, \ldots, n, S_1 = S^s_1 = s$. Replace the conditional covariance and conditional mean (85) and (86) by
\[
\Sigma^s_t \overset{\Delta}{=} \text{cov}(S^s_t, S^s_t|V^{t-1}, S^s_1), \quad t = 2, \ldots, n, \quad S^s_1 = s, \quad \hat{S}^s_t \overset{\Delta}{=} \mathbb{E}\{S^s_t|V^{t-1}, S^s_1\},
\]  
(102)
\[
\Sigma^s_t \overset{\Delta}{=} \text{cov}(S^s_t, S^s_t|V^{t-1}, S^s_1), \quad t = 2, \ldots, n, \quad S_1 = s, \quad \hat{S}^s_t \overset{\Delta}{=} S^s_t, \quad \Sigma^s_t \overset{\Delta}{=} \text{cov}(S^s_t, S^s_t|S^s_1) = 0.
\]  
(103)

Then, all statements of Lemma 1 hold, with the following changes:
\[
\Sigma_t \mapsto \Sigma^s_t, \quad \Sigma^s_1 = 0, \quad P_{V_t|V^{t-1}} \mapsto P_{V_t|V^{t-1}, S^s_t}, \quad \hat{S}_t \mapsto \hat{S}^s_t, \quad \hat{S}^s_t = s, \quad \text{etc, } t = 1, \ldots, n.
\]  
(104)

In particular, the conditional entropy of the $V^n$ conditioned on $S_1 = S^s_1 = s$ is given by
\[
H(V^n|s) = \frac{1}{2} \sum_{i=1}^{n} \log \left( 2\pi e \left[ C_i \Sigma^s_i C_i^T + N_i K_{wi} N_i^T \right] \right)
\]  
(105)
where $\Sigma^s_t, t = 2, \ldots, n$ satisfies the generalized DRE (95) with initial condition $\Sigma^s_1 = 0$.

**Proof.** We continue directly from Lemma 1 and (102), (103). 

Next, we introduce an example of a PO-SS realization of the noise that we often use in this paper.
Example 1. A time-varying PO-SS \((a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)\) noise realization is defined by

\[
S_{t+1} = a_t S_t + b_t^1 W_t^1 + b_t^2 W_t^2, \quad t = 1, 2, \ldots, n - 1
\]

\[
V_t = c_t S_t + d_t^1 W_t^1 + d_t^2 W_t^2, \quad t = 1, \ldots, n,
\]

\[
S_1 \in N(\mu_{S_1}, K_{S_1}), \quad K_{S_1} \geq 0, \quad W_t^i \in N(0, K_W^i), \quad K_W^i \geq 0, \quad i = 1, 2, \quad t = 1, \ldots, n, \quad (108)
\]

\[
W_1^{1,n} \text{ and } W_2^{2,n} \text{ indep. seq. and indep. of } S_1, \quad (109)
\]

\[
a_t \in \mathbb{R}, \quad c_t \in \mathbb{R}, \quad b_t^1 \in \mathbb{R}, \quad d_t^1 \in \mathbb{R}, \quad i = 1, 2, \forall t \text{ are nonrandom}, \quad (110)
\]

\[
b_t \circ b_t \triangleq (b_t^1)^2 K_{W_t^1} + (b_t^2)^2 K_{W_t^2}, \quad b_t \circ d_t \triangleq b_t^1 K_{W_t^1} d_t^1 + b_t^2 K_{W_t^2} d_t^2,
\]

\[
d_t \circ d_t \triangleq (d_t^1)^2 K_{W_t^1} + (d_t^2)^2 K_{W_t^2} > 0, \forall t. \quad (112)
\]

The next corollary is an application of Lemma 1 to the time-varying PO-SS noise of Example 1.

Corollary 2. The entropy \(H(V^n)\) of the PO-SS \((a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)\) noise of Example 1 is computed from Lemma 1 with the following changes:

\[
C_t \rightarrow c_t, \quad A_t \rightarrow a_t, \quad B_t K_{W_t^1} N_t^T \rightarrow b_t \circ d_t,
\]

\[
B_t K_{W_t^1} B_t^T \rightarrow b_t \circ b_t, \quad N_t K_{W_t^1} N_t^T \rightarrow d_t \circ d_t, \quad t = 1, \ldots, n. \quad (113)
\]

Proof. This is easily verified. \(\square\)

From Corollary 2, we have the following observations:

Remark 4. Consider the PO-SS \((a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)\) noise of Example 1. Then, the following hold.

(a) Consider the code of Definition 2. At each time \(t\), the optimal channel input process \(X^n\) is either realized by \((35)\), or equivalently by \((68)\), i.e.,

\[
X_t = \sum_{j=1}^{t-1} B_{t,j} V_j + Z_t = \sum_{j=1}^{t-1} \Gamma_{t,j}^1 V_j + \sum_{j=1}^{t-1} \Gamma_{t,j}^2 Y_j + Z_t, \quad t = 1, 2, \ldots, n.
\]

Moreover, \(X_t\) cannot be expressed in terms of the state \(S^t\) because, by \((106)\) and \((107)\), the noise sequence \(V^{t-1}\) does not specify \(S^t\) for \(t = 1, \ldots, n\).

(b) Consider the code of Definition 3, i.e., with a fixed initial state \(S_1 = S_1^t = s\). With Corollary 1 using \((113)\) \(H(V^n|s)\) is computed from Lemma 1, with \(\Sigma_1 = \Sigma_1^t = 0, \text{ and } (105)\) reduces to

\[
H(V^n|s) = \frac{1}{2} \sum_{t=1}^{n} \log \left( 2\pi e \left[ (c_t^2)^2 \Sigma_t^2 + d_t \circ d_t \right] \right)
\]

\[
(115)
\]

where \(\Sigma_t^s\) is the solution of \((95)\) with \(\Sigma_1 = \Sigma_1^s = 0\) (using \((113)\)).

We also apply our results to various versions of the autoregressive moving average (ARMA) noise model, such as the double-side and single-sided, stationary version of the ARMA noise, previously analyzed in [4] and in many other papers, to illustrate fundamental differences of Case (I) and Case (II) formulations.
Example 2. The time-invariant ARMA$(a, c)$ noise

(a) The time-invariant one-sided, stable or unstable, autoregressive moving average (ARMA$(a, c)$, $a \in (-\infty, \infty)$, $c \in (-\infty, \infty)$) noise is defined by

\[
V_t = cV_{t-1} + W_t - aW_{t-1}, \quad \forall t \in \mathbb{Z}_+, \quad V_0 \xrightarrow{\triangle} \{1, 2, \ldots \},
\]

(116)

\[
V_0 \in N(0, K_{V_0}), \quad K_{V_0} \geq 0, \quad W_0 \in N(0, K_{W_0}), \quad K_{W_0} \geq 0,
\]

(117)

\[
W_t \in N(0, K_W), \quad K_W > 0, \quad \forall t \in \mathbb{Z}_+, \quad \{W_0, W_1, \ldots, W_n\} \text{ indep. seq. and indep. of } V_0,
\]

(118)

\[
c \in (-\infty, \infty), \quad a \in (-\infty, \infty), \quad c \neq a.
\]

(119)

To express the AR$(a, c)$ in state space form, we define the state variable of the noise by

\[
S_t \equiv \frac{cV_{t-1} - aW_{t-1}}{c - a}, \quad \forall t \in \mathbb{Z}_+
\]

(120)

Then, the state space realization of $V^n$ is

\[
S_{t+1} = cS_t + W_t, \quad \forall t \in \mathbb{Z}_+, \quad V_0 \xrightarrow{\triangle} \{1, 2, \ldots \},
\]

(121)

\[
V_t = (c - a)S_t + W_t, \quad \forall t \in \mathbb{Z}_+, \quad V_0 \xrightarrow{\triangle} \{1, 2, \ldots \},
\]

(122)

\[
K_{S_1} = \frac{(c - a)^2 lies in \mathbb{Z}_+. \quad \text{indep. seq. and indep. of } V_0
\]

(123)

We note that the AR$(a, c)$ is not necessarily stationary or asymptotically stationary. A special case of the AR$(a, c)$ is the AR$(c)$ noise (i.e., with $a = 0$) defined by

\[
V_t = cV_{t-1} + W_t, \quad t = 1, 2, \ldots, \quad K_{V_0} \geq 0, \quad K_W > 0.
\]

(124)

(b) Double-sided wide-sense stationary ARMA$(a, c)$, $a \in [-1, 1], c \in (-1, 1)$ noise. A double-sided wide-sense stationary ARMA$(a, c)$ noise is defined by

\[
V_t = cV_{t-1} + W_t - aW_{t-1}, \quad \forall t \in \mathbb{Z} \triangleq \{\ldots, -1, 0, 1, \ldots \}, \quad |a| \leq 1, \quad |c| < 1.
\]

(125)

where $W_t, \forall t \in \mathbb{Z}$ is an independent and identically distributed Gaussian sequence, i.e., $W_t \in N(0, K_W), \forall t$. The power spectral density (PSD) of the wide-sense stationary noise (this corresponds to [4] (Equation (43) with $L = 1$)) is given by

\[
S_V(e^{\text{j}\theta}) \overset{\triangle}{=} K_W \frac{1 - ae^{\text{j}\theta}}{1 - ce^{\text{j}\theta}} \frac{1 - ae^{-\text{j}\theta}}{1 - ce^{-\text{j}\theta}}, \quad |c| < 1, \quad |a| \leq 1, \quad c \neq a, \quad K_W > 0.
\]

(126)

We define the state process by

\[
S_t \overset{\triangle}{=} \frac{cV_{t-1} - aW_{t-1}}{c - a}, \quad \forall t \in \mathbb{Z}.
\]

(127)

Then, the stationary state space realization of $V_t, \forall t \in \mathbb{Z}$ is

\[
S_{t+1} = cS_t + W_t, \quad \forall t \in \mathbb{Z},
\]

(128)

\[
V_t = (c - a)S_t + W_t, \quad \forall t \in \mathbb{Z}
\]

(129)

provided that the initial covariances $\text{cov}(S_t, S_t), \text{cov}(S_t, V_t), \text{cov}(V_t, V_t)$ are chosen appropriately to ensure stationarity (see Proposition 1).
For the AR(a,c) noise, we clarify differences of the feedback codes of Definitions 1 and 3 in the next remark, as well as of the Case (I) formulation versus the Case (II) formulation (and we discuss implication to results in [4,7,9,11,12]).

**Remark 5. ARMA(a,c) noise of Example 2**

(a) Consider any of the AR(a,c) of Example 2. For the code of Definition 2, the channel input process X^n cannot be expressed in terms of the state S^n (also see Remark 4(a)).

(b) Consider the non-stationary AR(a,c), a ∈ (−∞,∞), c ∈ (−∞,∞) of Example 2(a).

(i) Assume the code of Definition 3, with initial state V_0 = v_0 known to the encoder. By (120),

\[ S_1 = S_1^{v_0} = \frac{ce_0 - aW_0}{c - a}, \quad V_0 = v_0; \]

hence, knowledge of V_0 = v_0 at the encoder does not determine S_1^{v_0} because the encoder requires knowledge of W_0 for this to hold. It then follows that H(V^n|v_0) is computed from Corollary 1 as follows:

\[ \Sigma_1 = \Sigma_1^{v_0} = \frac{(a)^2K_0}{(c - a)^2}, \quad (105) \text{ reduces to } H(V^n|v_0) = \frac{1}{2} \sum_{i=1}^{n} \log \left( 2\pi e \left[ (c)^2\Sigma_1^{v_0} + K_W \right] \right) \]

where \( \Sigma_1^{v_0} \) is the solution of (95) with initial data \( \Sigma_1 = K_S = \Sigma_1^{v_0} \), \( K_W \geq 0 \).

(ii) Assume the code of Definition 3, with initial state \( S_1 = s \) or \( (V_0, W_0) = (v_0, w_0) \), known to the encoder. Then, by Corollary 1,

\[ H(V^n|v_0, w_0) = \frac{1}{2} \sum_{i=1}^{n} \log \left( 2\pi e K_W \right). \]

By (120), \( S_1 \triangleq \frac{ce_0 - aW_0}{c - a} \), and a necessary condition for Conditions 1 of Section 1.1 to hold is the following: both \( (V_0, W_0) = (v_0, w_0) \) are known to the encoder and the decoder.

(c) The statements of parts (a) and (b) also hold for the double-sided and the one-sided wide-sense stationary AR(a,c), a ∈ [−1,1], c ∈ (−1,1) of Example 2(b,c).

(d) The Case (II) formulation discussed in Section 1.1 requires Conditions 1 and 2 to hold. For any of the AR(a,c) noise models, Conditions 1 and 2 hold if and only if \( S_1 = s_1 \) or \( (V_0, W_0) = (v_0, w_0) \) are known to the encoder. Clearly, the values of H(V^n) under the Case (I) formulation are fundamentally different from the values of H(V^n|s), \( S_1 = s \) under the Case (II) formulation. Consequently, in general, \( C_{FB}^b(\kappa) \) given by (75) is fundamentally different from \( C_{FB}^b(\kappa, s) \), i.e., it corresponds to a fixed initial state \( S_1 = s \), known to the encoder and the decoder, and to the channel input distribution.

(e) From parts (a–d), the characterization of feedback capacity for the stationary ARMA(a,c), a ∈ [−1,1], c ∈ (−1,1) given in [4] (Theorem 6.1, C_{FB}^b) (which is derived based on [4] (Lemma 6.1)) presupposed the encoder and the decoder assumed knowledge of \( S_1 = S_1^s = s \).

In fact, the formulas of capacity in [4,7,8] use \( H(V^n|v_0, w_0) = \frac{1}{2} \sum_{i=1}^{n} \log \left( 2\pi e K_W \right). \)

In the next proposition, we state conditions for the stable realizations of Example 2(a), i.e., AR(a,c), a ∈ [−1,1], c ∈ (−1,1) to be asymptotically stationary, and for the realizations of Example 2(b,c) to be stationary. We should emphasize that for stationary noise, we
need to determine the initial conditions of the generalized Kalman filter of Lemma 1 to correspond to the stationary noise.

**Proposition 1.** Asymptotically stationary and stationary ARMA\((a,c)\) noises of Example 2
(a) The realization of the double-sided ARMA\((a,c)\), \(a \in [-1,1], c \in (-1,1)\) noise of Example 2(b) is stationary if the following conditions hold:
\[
d_{11} \triangleq \text{cov}(S_t, S_t) = K_{S_t}, \quad d_{12} \triangleq \text{cov}(S_t, V_t) = K_{S_t,V_t},
\]
\[
d_{22} \triangleq \text{cov}(V_t, V_t) = K_{V_t}, \quad \text{are constant } \forall t \in \mathbb{Z}.
\]
(133)

where the constants \((d_{11}, d_{12}, d_{22})\) are given by
\[
d_{11} = \frac{K_W}{1 - c^2}, \quad d_{12} = \frac{(c - a)K_W}{1 - c^2}, \quad d_{22} = \frac{(c - a)^2K_W}{1 - c^2} + K_W.
\]
(134)

Similarly, the one-sided ARMA\((a,c)\), \(a \in [-1,1], c \in (-1,1)\) noise of Example 2(c) is stationary if the above equations hold \(\forall t \in \mathbb{Z}_+ \triangleq \{1, 2, \ldots\}\).

(b) The realization of the ARMA\((a,c)\) noise of Example 2(a) is asymptotically stationary if \(a \in [-1,1], c \in (-1,1)\).

(c) For the stationary realization of part (a), the optimal conditional variance and conditional mean of \(S_t\) from \((V_0, V_1, V_2, \ldots, V_{t-1})\), i.e., \(\Sigma_t \triangleq \text{cov}(S_t, S_t^T, V_0), \hat{S}_t \triangleq \mathbb{E}\{S_t|V^T, V_0\}\) are defined by the generalized Kalman filter given by
\[
\hat{S}_{t+1} = c\hat{S}_t + \left( (c)^2\Sigma_t + K_W \right) \left( K_W + (c - a)^2\Sigma_t \right)^{-1} \left( V_t - (c - a)\hat{S}_t \right), \quad t = 1, 2, \ldots, (135)
\]
\[
\Sigma_{t+1} = (c)^2\Sigma_t + K_W - \left( (c(a-c))\Sigma_t + K_W \right) \left( K_W + (c(a-c))^2\Sigma_t \right)^{-1}, (136)
\]
initialized at the initial data
\[
\hat{S}_1 \triangleq \mathbb{E}\{S_1|V_0\} = \frac{cd_{12} + K_W}{d_{22}} V_0, \quad (137)
\]
\[
\Sigma_1 \triangleq \text{cov}(S_1, S_1|V_0) = d_{11} - \frac{(d_{12})^2}{d_{22}}. \quad (138)
\]

(i) If the conditioning information is \((V_{-N}, \ldots, V_0, V_1, V_2, \ldots, V_{t-1})\), then the generalized Kalman filters (135) and (136) still hold and are initialized at the initial data
\[
\hat{S}_{-N+1} = \frac{cd_{12} + K_W}{d_{22}} V_{-N}, \quad (139)
\]
\[
\Sigma_{-N+1} = d_{11} - \frac{(d_{12})^2}{d_{22}}. \quad (140)
\]

(ii) If the initial data \(V_0\) are not available, then the generalized Kalman filter is initialized at initial data \(\hat{S}_1 = 0, \Sigma_1 = \text{cov}(S_1, S_1) = d_{11}\).

**Proof.** See Appendix A.2. \(\square\)

**Remark 6.** Consider the stationary double-sided or one-sided ARMA\((a,c)\), \(a \in [-1,1], c \in (-1,1)\) of Example 2. From Proposition 1 and, in particular, the initial data \(\hat{S}_1, \Sigma_1\) stated in (137) and (138), it is clear that even if the encoder and the decoder know the initial state \(V_0\). Thus, \(H(V^n|v_0) \neq \frac{1}{2} \sum_{t=1}^n \log \left( 2\pi eK_W \right)\). In this case, the value of \(C_n^b(\kappa, v_0)\) defined by (81) is
fundamentally different from the formulation in [4,7,8], leading to the characterization of feedback capacity [4] (Theorem 6.1).

In the next corollary, we further clarify the difference between Case (I) formulation and Case (II) formulation, by stating the analog of Theorem 1 for the code of Definition 3, i.e., when \( S_1 = S_1^* = s \) is fixed.

**Corollary 3.** \( n \)-FTFI capacity for feedback code of Definition 3

Consider the time-varying AGN channel defined by (1), driven by a noise with the PO-SS realization of Definition 2, and the code of Definition 3, with initial state \( S_1 = S_1^* = s \) being fixed. Then, the following hold.

(a) The \( n \)-FTFI capacity \( C_{n}^{fb}(\kappa, s) \) is given by

\[
C_{n}^{fb}(\kappa, s) \triangleq \sup_{\frac{1}{n} \sum_{t=1}^{n} (X_t^2) \leq \kappa} H^\mathcal{W}(Y^n|s) - H(V^n|s).
\]

(b) A necessary condition for Condition 2 of Section 1.1 to hold is as follows:

(i) The initial state of the noise \( S_1 \) is known, i.e., \( S_1 = S_1^* = s \) or \( S_1 = s_1^* = s = 0 \), and \( N_t W_t \) uniquely defines \( C_{t+1} B_t W_t, \forall t \).

Moreover, if (i) holds, then the entropy \( H(V^n|s) \) of part (a) is given by

\[
H(V^n|s) = \frac{1}{2} \sum_{t=1}^{n} \log \left( 2 \pi e N_t W_t N_t^2 \right).
\]

The stable, time-invariant PO-SS realization of Definition 2, which is considered in [4,7], satisfies \( W_t : \Omega \to \mathbb{R}, N_1 = 1, \forall t, \) i.e., \( n_w = 1 \). Moreover, for this realization, (i) always holds.

**Proof.** See Appendix A.3. \( \square \)

In the next remark, we illustrate that \( H(V^n|s) \) given by (143) follows directly from Lemma 1 by fixing \( S_1 = S_1^* = s \) and assuming that \( N_t W_t \) uniquely defines \( C_{t+1} B_t W_t, \forall t \).

**Remark 7.** The \( n \)-FTFI capacity for code of Definition 1 versus code of Definition 3.

Consider the generalized Kalman filter of the PO-SS noise realization, of Lemma 1, and assume the following:

(i) The initial state of the noise \( S_1 \) is known, i.e., \( S_1 = S_1^* = s \) or \( S_1 = s_1^* = s = 0 \), and \( N_t W_t \) uniquely defines \( C_{t+1} B_t W_t, \forall t \).

Then, all statements of Lemma 1 hold, by replacing \((\Sigma_t, S_t)\) with \((\Sigma_t^*, S_t^*)\) for \( t = 1, 2, \ldots \).

Since \( \Sigma_t^* \) satisfies the generalized DRE (95) with initial condition \( \Sigma_1^* = 0 \), then it is easy to deduce that \( \Sigma_t^* = 0 \) for \( t = 1, 2, \ldots, n \) is a solution. By substituting \( \Sigma_t^* = 0, t = 1, 2, \ldots, n \) in (98), we obtain (143), as expected, which is precisely the entropy of the noise that appeared in [4,7].

On the other hand, for the code of Definition 1, by Theorem 1(d), the right-hand side of the \( n \)-FTFI capacity \( C_{n}^{fb}(\kappa) \) involves \( H(V^n) \), which is computed using the generalized Kalman filter of Lemma 1.
2.3. A Sufficient Statistic Approach to the Characterization of \(n\)–FTFI Capacity of AGN Channels Driven by PO-SS Noise Realizations

The characterization of the \(n\)–FTFI capacity via (34) (which is equivalently given in Theorem 1(d)), although compactly represented, is not very practical because the input process \(X^n\) is not expressed in terms of a sufficient statistic that summarizes the information of the channel input strategy [39].

In this section, we wish to identify a sufficient statistic for the input process \(X_t\), given by (68), called the state of the input, which summarizes the information contained in \((V_t^{t-1}, Y_t^{t-1})\). It will then become apparent that the characterization of the \(n\)–FTFI capacity for the Cover and Pombra formulation and code of Definition 1 can be expressed as a functional of two generalized matrix DREs.

First, we invoke Theorem 1 and Lemma 1 to show that for each time \(t\), \(X_t\) is expressed as

\[
X_t = \Lambda_t \left( \hat{S}_t - \mathbf{E} \left\{ \hat{S}_t \mid Y_{t-1} \right\} \right) + Z_t, \quad t = 1, \ldots, n, \tag{144}
\]

\[
\hat{S}_t \overset{\Delta}{=} \mathbf{E} \left\{ \hat{S}_t \mid V_{t-1} \right\}, \quad \hat{S}_t \overset{\Delta}{=} \mathbf{E} \left\{ \hat{S}_t \mid Y_{t-1} \right\} \tag{145}
\]

which means, at each time \(t\), the state of the channel input process \(X_t\) is \((\hat{S}_t, \hat{S}_t)\). We show that \(\hat{S}_t\) satisfies another generalized Kalman filter recursion.

Now, we prepare to prove (144) and the main theorem. We start with preliminary calculations.

\[
\mathbb{P} \{ Y_t \in dy \mid Y_{t-1}, X^t_1 \} = \mathbf{P}_t(dy \mid X_t, V_t^{t-1}), \quad t = 2, \ldots, n, \quad \text{by channel definition} \tag{146}
\]

\[
= \mathbf{P}_t(dy \mid X_t, V_t^{t-1}, \hat{S}_t), \quad \text{by channel definition} \tag{147}
\]

\[
= \mathbf{P}_t(dy \mid X_t, V_t^{t-1}, \hat{S}_t, \hat{S}_t^{t-1}), \quad \text{by (90), i.e.,} \ V_t = C_t \hat{S}_t + \hat{I}_t \tag{148}
\]

\[
= \mathbf{P}_t(dy \mid X_t, \hat{S}_t), \quad \text{by} \ Y_t = X_t + \hat{I}_t = X_t + C_t \hat{S}_t + \hat{I}_t, \tag{91} \tag{149}
\]

At \(t = 1\), we also have \(\mathbb{P} \{ Y_1 \in dy \mid X_1 \} = \mathbf{P}_1(dy \mid X_1)\). By (149), it follows that the conditional distribution of \(Y_t\) given \(Y_{t-1} = y_{t-1}\) is

\[
\mathbf{P}_t(dy_t \mid y_{t-1}) = \int \mathbf{P}_t(dy_t \mid x_t, \hat{S}_t) \mathbf{P}_t(dx_t \mid \hat{S}_t, y_{t-1}) \mathbf{P}_t(dy_t \mid y_{t-1}), \quad t = 2, \ldots, n, \tag{150}
\]

\[
\mathbf{P}_1(dy_1) = \int \mathbf{P}_1(dy_1 \mid \hat{S}_t) \mathbf{P}_1(dx_1 \mid \hat{S}_t) \mathbf{P}_1(dy_1). \tag{151}
\]

From the above distributions, at each time \(t\), the distribution of \(X_t\) conditioned on \((V_t^{t-1}, Y_t^{t-1})\), given in Theorem 1, is also expressed as a linear functional of \((\hat{S}_t, Y_t^{t-1})\), for \(t = 1, \ldots, n\).

The next theorem further shows that for each \(t\), the dependence of \(X_t\) on \(Y_{t-1}\) is expressed in terms of \(\mathbf{E} \left\{ \hat{S}_t \mid Y_{t-1} \right\}\) for \(t = 1, \ldots, n\), and this dependence gives rise to an equivalent sequential characterization of the Cover and Pombra \(n\)–FTFI capacity, \(C_n^{fb}(\kappa)\).

**Theorem 3.** Equivalent characterization of \(n\)–FTFI capacity \(C_n^{fb}(\kappa)\) for PO-SS noise realizations Consider the time-varying AGN channel defined by (1), driven by a noise with the PO-SS realization.
of Definition 2, and the code of Definition 1. Also consider the generalized Kalman filter of Lemma 1. Define the conditional covariance and conditional mean of $\hat{S}_t$, given $Y^{t-1}$, by

$$K_t \triangleq \text{cov} \left( \hat{S}_t, \hat{S}_t | Y^{t-1} \right) = E \left\{ (\hat{S}_t - \hat{S}_t)^T | Y^{t-1} \right\}, \quad t = 2, \ldots, n,$$

$$\hat{S}_t \triangleq E \left\{ \hat{S}_t | Y^{t-1} \right\}, \quad \hat{S}_1 \triangleq \mu_{S_1}, \quad K_1 \triangleq 0. \quad (153)$$

Then, the following hold.

(a) An equivalent characterization of the n–FTFI capacity $C_{\bar{b}}(n)$, defined by (34) and (35), is

$$C_{\bar{b}}(n) = \sup_{P_{\bar{b}}(X)} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V^n) \quad (154)$$

where $(X^n, Y^n)$ is jointly Gaussian, and

$$H(V^n) \text{ is the entropy of } V^n \text{ given in Lemma 1, i.e., } (98),$$

$I^n \text{ is the innovations process of } V^n \text{ given in Lemma 1};$

$$Y_t = X_t + V_t, \quad t = 1, \ldots, n;$$

$$V_t = C_t \hat{S}_t + \hat{I}_t;$$

$$P_t(dy_t | y^t) = \int P_t(dy|x_t, \hat{s}_t) P_t(dx_t | x_t, y^t) P_t(d\hat{s}_t | y^t), \quad t = 2, \ldots, n; \quad (159)$$

$$P_1(dy_1) = \int P_1(dy|x_1, \hat{s}_1) P_1(dx_1 | x_1) P_1(d\hat{s}_1); \quad (160)$$

$$P_1(dy_1) \in \mathcal{N}(\mu_{Y_1}, K_{Y_1});$$

$$\mu_{Y_t} | Y^{t-1} \text{ is linear in } Y^{t-1} \text{ and } K_{Y_t} | Y^{t-1} \text{ is nonrandom};$$

$$P_t(dx_t | \hat{s}_t, y^t) \in \mathcal{N}(\mu_{X_t}, K_{X_t});$$

$$\mu_{X_t} | S_t, Y^{t-1} \text{ is linear in } (S_t, Y^{t-1}) \text{ and } K_{X_t} | S_t, Y^{t-1} \text{ is nonrandom};$$

$$P_{\bar{b}}(X) \triangleq \left\{ \mathcal{P}_t(dx_t | \hat{s}_t, y^t), t = 1, \ldots, n: \frac{1}{n} E \left( \sum_{i=1}^n (X_i)^2 \right) \leq \kappa \right\}. \quad (165)$$

(b) The optimal jointly Gaussian process $(X^n, Y^n)$ of part (a) is represented as a function of a sufficient statistic by

$$X_t = \Lambda_t \left( \hat{S}_t - \hat{S}_t \right) + Z_t, \quad t = 1, \ldots, n, \quad (166)$$

$$Z_t \in \mathcal{N}(0, K_{Z_t}) \text{ independent of } (X^{t-1}, V^{t-1}, S^t, \hat{S}^t, \hat{I}^t, Y^{t-1}), \quad t = 1, \ldots, n; \quad (167)$$

$$I_t \in \mathcal{N}(0, K_{I_t}) \text{ independent of } (X^{t-1}, V^{t-1}, S^t, \hat{S}^t, Y^{t-1}), \quad t = 1, \ldots, n; \quad (168)$$

$$Y_t = \Lambda_t \left( \hat{S}_t - \hat{S}_t \right) + Z_t + V_t, \quad t = 1, \ldots, n; \quad (169)$$

$$= \Lambda_t \left( \hat{S}_t - \hat{S}_t \right) + C_t \hat{S}_t + \hat{I}_t + Z_t; \quad (170)$$

$$\frac{1}{n} E \left\{ \sum_{i=1}^n (X_i)^2 \right\} = \frac{1}{n} \sum_{i=1}^n \left( \Lambda_t K_t \Lambda_t^T + K_{Z_t} \right). \quad (171)$$

where $\Lambda_t$ is nonrandom.

The conditional mean and covariance, $\hat{S}_t$ and $K_t$, are given by generalized Kalman filter equations, as follows:

(i) $\hat{S}_t$ satisfies the Kalman filter recursion
Entropy 2025, 27, 207

\[ S_{t+1} = A_t \hat{S}_t + F_t(\Sigma_t, K_t) I_t, \quad \hat{S}_1 = \mu S_1; \]  

(172)

\[ F_t(\Sigma_t, K_t) \triangleq \left( A_t K_t (\Lambda_t + C_t)^T + M_t(\Sigma_t) K_{t_{1}} \right) \left\{ K_{t_1} + K_{Z_1} + (\Lambda_t + C_t) K_t (\Lambda_t + C_t)^T \right\}^{-1}; \]  

(173)

\[ I_t \triangleq Y_t - E\left\{ Y_t | Y_t^{t-1} \right\} = Y_t - C_t \hat{S}_t = \left( \Lambda_t + C_t \right) \left( \hat{S}_t - \hat{\hat{S}}_t \right) + \hat{I}_t + Z_t, \quad t = 1, \ldots, n; \]  

(174)

\[ I_t \in N(0, K_{t_{1}}), \quad t = 1, \ldots, n \text{ is an orthogonal innovations process,} \]  

(175)

\[ K_{Y_t | Y_t^{t-1}} = K_{I_t} \triangleq \text{cov} \left( I_t, I_t \right) = \left( \Lambda_t + C_t \right) K_t \left( \Lambda_t + C_t \right)^T + K_{t_t} + K_{Z_t}; \]  

(176)

\[ K_{t_t} \text{ given by (92).} \]  

(177)

(ii) The error \( \hat{E}_t \triangleq \hat{S}_t - \tilde{S}_t \) satisfies the recursion

\[
\hat{E}_{t+1} = F_t(\Sigma_t, K_t) \hat{E}_t + \left( M_t(\Sigma_t) - F_t(\Sigma_t, K_t) \right) \hat{I}_t \\
- F_t(\Sigma_t, K_t) Z_t, \quad \hat{E}_1 = \tilde{S}_1 - \hat{\hat{S}}_1 = 0, \quad t = 1, \ldots, n;
\]  

(178)

\[
F_t(\Sigma_t, K_t) \triangleq A_t - F_t(\Sigma_t, K_t) \left( \Lambda_t + C_t \right). \]  

(179)

(iii) \( K_t = E \left\{ \hat{E}_t \hat{E}_t^T \right\} \) satisfies the generalized DRE

\[
K_{t+1} = A_t K_t A_t^T - \left( A_t K_t (\Lambda_t + C_t)^T + M_t(\Sigma_t) K_{t_1} \right) \left\{ K_{t_1} + K_{Z_1} + (\Lambda_t + C_t) K_t (\Lambda_t + C_t)^T \right\}^{-1} \left( A_t K_t (\Lambda_t + C_t)^T + M_t(\Sigma_t) K_{t_1} \right)^T + M_t(\Sigma_t) K_{t_1} \left( M_t(\Sigma_t) \right)^T, \quad K_t \succeq 0, \quad t = 1, \ldots, n, \quad K_1 = 0. \]  

(180)

(c) An equivalent characterization of the \( n \)--FTFI capacity \( C_{n}^{fb}(\kappa) \), defined by (33) and (34), using the sufficient statistics of part (b) is

\[
C_{n}^{fb}(\kappa) = \sup \left\{ \frac{1}{2} \sum_{t=1}^{n} \log \frac{K_{Y_t | Y_t^{t-1}}}{K_{I_t | Y_t^{t-1}}} \right\} \]  

(181)

\[
= \sup \left\{ \frac{1}{2} \sum_{t=1}^{n} \log \frac{K_t}{K_{I_t}} \right\} \]  

(182)

\[
= \sup \left\{ \frac{1}{2} \sum_{t=1}^{n} \log \frac{\left( A_t + C_t \right) K_t \left( \Lambda_t + C_t \right)^T + K_{t_t} + K_{Z_t}}{K_{t_{1}}} \right\} \]  

(183)

Proof. See Appendix A.4. \( \Box \)

Remark 8. On the characterization of \( n \)--FTFI capacity of Theorem 3

The characterization of the \( n \)--FTFI capacity \( C_{n}^{fb}(\kappa) \) given by (183) involves the generalized matrix DRE \( K_t \), which is also a functional of the generalized matrix DRE \( \Sigma_t \) of the error covariance of the state \( S^n \) from the noise output \( V^n \). This feature is not part of the analysis in [4] and recent studies [4,7,9,11,12].

The next corollary follows directly from Theorem 3 as a degenerate case.
Corollary 4. Equivalent characterization of $n$–FTFI capacity $C_{fb}^n(\kappa, s)$ for PO-SS noise realizations
Consider the time-varying AGN channel defined by (1), driven by a noise with the PO-SS realization of Definition 2, and the code of Definition 3, with initial state $S_1 = S_1^b = s$ fixed, and replace (152) and (153) by

$$K_1 = K_1^* \triangleq \text{cov} \left( \hat{S}_t^b, \hat{S}_t^b \right| Y^{t-1}, S_1 = s) = \mathbb{E} \left\{ \left( \hat{S}_t^b - \hat{S}_t^b \right) \left( \hat{S}_t^b - \hat{S}_t^b \right)^T \right\},$$

(184)

$$S_1 = \hat{S}_1^b \triangleq \mathbb{E} \left\{ \hat{S}_t^b \right| Y^{t-1}, S_1 = s \right\}, \quad t = 2, \ldots, n, \quad \hat{S}_1 = \hat{S}_1^b \triangleq s, \quad K_1 = K_1^* = 0.$$ (185)

Then, the characterization of $n$–FTFI capacity, (3), is

$$C_{fb}^n(\kappa, s) = \sup_{Q_{\mathcal{Y}t}^n(\kappa)} \sum_{t=1}^n H(Y_t|Y^{t-1}, s) - H(V^n|s),$$

(186)

$$P_{\mathcal{Y}t}^S(\kappa) \triangleq \left\{ P_t(dx_t|\hat{S}_t^b, Y^{t-1}, s), t = 1, \ldots, n : \frac{1}{n} \mathbb{E} \left\{ \sum_{t=1}^n (X_t)^2 \right\} | S_1^b = s \leq \kappa \right\}$$

(187)

where $H(V^n|s)$ is given by Corollary 1, and the statements of Theorem 3 hold with the above changes, i.e., (184), (185), and all conditional entropies, distributions, expectations, etc., are defined for fixed $S_1^b = S_1^b = s$.

Proof. It is easily verified from the derivation of Theorem 3 by fixing $S_1 = S_1^b = s$. □

Remark 9. On the characterization of $n$–FTFI capacity of Corollary 4
The characterization of the $n$–FTFI capacity $C_{fb}^n(\kappa, s)$ given in Corollary 4 (similar to Theorem 3) involves two generalized matrix DREs, because it does not assume that Conditions 1 and 2 hold. This distinction is not part of the analysis in [4,7,9,11,12].

2.4. Application Examples
In this section, we apply Theorem 3 to specific examples.

First, we consider the application example of the AGN channel driven by the PO-SS$(a_i, c_i, b_i^1, b_i^2, d_i^1, d_i^2)$ noise.

Corollary 5. The $n$–FTFI capacity $C_{fb}^n(\kappa)$ of the AGN channel driven by the PO-SS$(a_i, c_i, b_i^1, b_i^2, d_i^1, d_i^2)$ noise is obtained from Lemma 1 and Theorem 3 by using (113).

Proof. This is easily verified, as in Corollary 2. □

In the next corollary, we apply Theorem 3 to the stable and unstable ARMA$(a, c)$ noise to obtain the characterization of the $n$–FTFI capacity $C_{fb}^n(\kappa)$ and $C_{fb}^n(\kappa, s)$. It is then obvious that for the stable ARMA$(a, c), a \in [-1, 1], c \in (-1, 1)$ noise, the characterization of $C_{fb}^n(\kappa)$ involves two generalized DREs, contrary to the analysis in [4,7,9,11,12], for the same noise model.

Corollary 6. Characterization of $n$–FTFI capacity $C_{fb}^n(\kappa)$ for the ARMA$(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$
Consider the time-varying AGN channel defined by (1) and the code of Definition 1.
(a) For the non-stationary ARMA$(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise of Example 2(a), the characterization of the $n$–FTFI capacity, $C_{fb}^n(\kappa)$, is
\[ C_{n}^{fb}(\kappa) = \sup_{(\Lambda_t, K_{Z_t}), t=1, \ldots, n: \sum_{t=1}^{n} \left( (\Lambda_t)^2 K_t + K_{Z_t} \right) \leq \kappa} \frac{1}{2} \sum_{t=1}^{n} \log \left( \frac{(\Lambda_t + c - a)^2 K_t + K_{I_t} + K_{Z_t}}{K_{I_t}} \right) \] (188)

subject to the constraints

\[ K_{t+1} = (c)^2 K_t + \left( M_t(\Sigma_t) \right)^2 K_{I_t} - \left( c K_t (\Lambda_t + c - a) + M_t(\Sigma_t) K_{I_t} \right)^2 \cdot (K_{I_t} + K_{Z_t} + (\Lambda_t + c - a)^2 K_t)^{-1}, \quad K_1 = 0, \quad t = 1, \ldots, n, \] (189)

\[ K_{Z_t} \geq 0, \quad K_t \geq 0, \quad c \neq a, \quad K_W > 0, \quad t = 1, \ldots, n \] (190)

and where

\[ M_t(\Sigma_t) = (c \Sigma_t (c - a) + K_W) \left( K_W + (c - a)^2 \Sigma_t \right)^{-1}, \] (191)

\[ K_{I_t} = (c - a)^2 \Sigma_t + K_W, \quad t = 1, \ldots, n, \] (192)

\[ \Sigma_{t+1} = (c)^2 \Sigma_t + K_W - (c \Sigma_t (c - a) + K_W)^2 \left( K_W + (c - a)^2 \Sigma_t \right)^{-1}, \quad t = 1, \ldots, n, \] (193)

\[ \Sigma_1 = K_{S_1} = \frac{(c_0)^2 K_{S_0} + (a_0)^2 K_{W_0}}{(c_0 - a_0)^2}. \] (194)

The optimal jointly Gaussian process \((X^n, Y^n)\) is obtained from Theorem 3(b) by invoking

\[ A_t \mapsto c, \quad C_t \mapsto c - a, \quad B_t \mapsto 1, \quad N_t \mapsto 1, \quad t = 1, 2, \ldots, n. \] (195)

Special Case. If \(\Sigma_1 = 0\) or the initial state is fixed, \(S_1 = S_1^0 = s\), then

\[ \Sigma_t = \Sigma_t^0 = 0, \quad K_{I_t} = K_W, \quad M_t(\Sigma_t) = M_t(\Sigma_t^0) = 1, \quad t = 1, 2, \ldots \] (196)

and \(C_{n}^{fb}(\kappa)\) reduces to

\[ C_{n}^{fb}(\kappa) = C_{n}^{fb}(\kappa, s) = \sup_{(\Lambda_t, K_{Z_t}), t=1, \ldots, n: \sum_{t=1}^{n} \left( (\Lambda_t)^2 K_t + K_{Z_t} \right) \leq \kappa} \frac{1}{2} \sum_{t=1}^{n} \log \left( \frac{(\Lambda_t + c - a)^2 K_t + K_W + K_{Z_t}}{K_W} \right) \] (197)

subject to the constraints

\[ K_{t+1}^s = (c)^2 K_t^s + K_W - \left( c K_t^s (\Lambda_t + c - a) + K_W \right)^2 \cdot (K_{Z_t} + (\Lambda_t + c - a)^2 K_t^s + K_W)^{-1}, \quad K_1^s = 0, \quad K_t^s \geq 0, \quad K_{Z_t} \geq 0, \quad t = 1, \ldots, n. \] (198)

(This special case is precisely the application example analyzed in [4,7,8]).

(b) For the non-stationary AR(c), \(c \in (-\infty, \infty)\) noise of Example 2(c), the characterization of the \(n\)-FTFI capacity \(C_{n}^{fb}(\kappa)\) is obtained from part (a) by setting \(a = 0\), i.e.,
\[ C_n^{fb}(\kappa) = \sup_{(\Lambda, K_{Z_t}), \sum_{i=1}^n} \left\{ \frac{1}{2} \sum_{t=1}^n \log \left( \frac{\left(\Lambda_i + c\right)^2 K_i + (c)^2 \Sigma + K_W + K_{Z_t}}{(c)^2 \Sigma + K_W} \right) \right\} \]

subject to the constraints of \( K_t, \Sigma \) are the non-negative solutions of the generalized RDEs:

\[ K_{t+1} = (c)^2 K_t + (c)^2 \Sigma + K_W - \left( cK_t (\Lambda_i + c) + (c)^2 \Sigma + K_W \right) \]

\[ \left( (c)^2 \Sigma + K_W + K_{Z_t} + (\Lambda_i + c)^2 K_t \right)^{-1}, \quad K_t = 0, \quad t = 1, \ldots, n, \]  

\[ \Sigma_{t+1} = (c)^2 \Sigma_t + K_W - \left( (c)^2 \Sigma_t + K_W \right) \left( K_W + (c)^2 \Sigma_t \right)^{-1}, \quad \Sigma_1 = K_{S_1} = K_{S_0} \geq 0. \]

(c) For the non-stationary AR(c), \( c \in (-\infty, \infty) \) noise of Example 2(c), with \( \Sigma_1 = 0 \) or a fixed initial state \( S_1 = S_t^1 = s \), then (196) holds, i.e., \( \Sigma_t = \Sigma_t^1 = 0, K_{Z_t} = K_W, M_t(\Sigma_t) = M_t(\Sigma_t^1) = 1, t = 1, 2, \ldots, \) and \( C_n^{fb}(\kappa) \) reduces to

\[ C_n^{fb}(\kappa) = C_n^{fb}(\kappa, s) \]

subject to the constraint

\[ K_{t+1} = (c)^2 K_t + K_W \]

\[ - \left( cK_t (\Lambda_i + c) + K_W \right) \left( K_W + (c)^2 \Sigma_t \right)^{-1}, \quad K_{Z_t} = 0, \quad t = 1, \ldots, n. \]

Proof. (a) The first part follows directly from Theorem 3 by using (195). The last part is obtained as follows. If \( \Sigma_1 = 0 \) or \( S_1 = S_t^1 = s \) is fixed, then, by (193), it follows that \( \Sigma_t = \Sigma_t^1 = 0, \forall t = 2, \ldots \). Moreover, by (191) and (192), it follows that \( M_t(\Sigma_t) = M_t(\Sigma_t^1) = 1, K_{Z_t} = (c - a)^2 \Sigma_t^1 + K_W = K_W, t = 1, 2, \ldots \). By substituting into (188) and (189), we obtain (197) and (180). (b) From part (a), let \( a = 0 \). Then,

\[ M_t(\Sigma_t) = \left( (c)^2 \Sigma_t + K_W \right) \left( K_W + (c)^2 \Sigma_t \right)^{-1}, \quad K_{Z_t} = (c)^2 \Sigma_t + K_W, \quad t = 1, \ldots, n. \]

By substituting into the equations of part (a), we obtain (200) and (201). (c) This is a special case of parts (a) and (b). \( \square \)

Remark 10. By Corollary 6(a), it is obvious that if \( \Sigma_1 = 0 \), i.e., \( K_{S_0} = K_{W_0} = 0 \), which means \( S_1 = S_t^1 = s \) is fixed, \( (V_0, W_0) = (v_0, w_0) \) is fixed (and known to the encoder and the decoder)—see (120). Then, \( \Sigma_1 = \Sigma_1^1 = 0 \) and \( C_n^{fb}(\kappa) = C_n^{fb}(\kappa, s) \), which depends on the initial state \( S_1 = S_t^1 = s \). To ensure that we obtain a large enough \( n \), the rate \( \frac{1}{n} C_n(\kappa, s) \) is independent of \( s \). Moreover, it is necessary to identify conditions for convergence of solutions \( K_t^1, t = 1, 2, \ldots \) of the generalized DRE (180) to a unique limit, \( \lim_{n \to \infty} K_t^1 = K^\infty \geq 0 \), which does not depend on the initial data \( K_1^1 = 0 \). We address this problem in Section 3.

2.5. Case (II) Formulation: A Degenerate of Case (I) Formulation

Theorem 3 gives the \( n \)–FTF capacity for the Case (I) formulation. However, since the Case (II) formulation is a special case of the Case (I) formulation, we expect that...
we can recover the characterization of the n–FTFI capacity for the Case (II) formulation from Theorem 3, i.e., when the code is \((s, 2^{nR}, n), n = 1, 2, \ldots\), and Conditions 1 and 2 of Section 1.1 hold. We show this in the next corollary.

**Corollary 7.** The degenerate n–FTFI capacity \(C_{n}^{fb}(\kappa)\) of Theorem 3 for the Case (II) formulation Consider the time-varying AGN channel defined by (1), driven by a noise with PO-SS realization of Definition 2, and suppose that the following hold:

1. The code is \((s, 2^{nR}, n), n = 1, 2, \ldots\);
2. Conditions 1 and 2 of Section 1.1 hold.

Then, the following hold:

(a) Corollary 1 holds, i.e., all statements of Lemma 1 hold with \((\Sigma_t, \hat{S}_t)\) replaced by \((\Sigma_t^b, \hat{S}_t^b)\) as defined by (102) and (103). In particular, \((\Sigma_t^b, \hat{S}_t^b) = (0, \hat{S}_t^b)\) for \(t = 1, 2, \ldots\), and \(H(V^n) = H(V^n|s)\) is given by (143).

(b) All statements of Theorem 3 hold with \((\Sigma_t, \hat{S}_t)\) replaced by \((\Sigma_t^b, \hat{S}_t^b)\), as in part (a), and \((K_t, \hat{S}_t)\) defined by (152) and (153) reduces to

\[
K_t = K_t^b = \operatorname{cov}\left(S_t^b, S_t^b | Y^{t-1}, S_t^b = s\right), \tag{205}
\]

\[
\hat{S}_t = \hat{S}_t^b = \mathbf{E}\left(S_t^b | Y^{t-1}, S_t^b = s\right), \quad K_t^b = 0, \hat{S}_t^b = s, \quad t = 2, \ldots, n \tag{206}
\]

In particular, the optimal input process \(X^n\) of Theorem 3(c) degenerates to

\[
X_t = \Lambda_t \left(S_t^b - \hat{S}_t^b\right) + Z_t, \quad X_1 = Z_1, \quad t = 2, \ldots, n. \tag{207}
\]

(c) The characterization of the n–FTFI capacity, \(C_{n}^{fb}(\kappa, s)\) of Theorem 3, degenerates to \(C_{n}^{fb,b}(\kappa, s)\). It is defined by

\[
C_{n}^{fb}(\kappa) = \sup_{\Lambda, K_t} \frac{1}{2}\sum_{t=1}^{n} \log \frac{K_{Y_t|V^{t-1},s}}{K_{V_t|V^{t-1},s}} \tag{208}
\]

\[
C_{n}^{fb,b}(\kappa, s) = \sup_{\Lambda, K_t} \left\{ \frac{1}{2}\sum_{t=1}^{n} \left( N_t K_{W_t N_t^T} + K_{Z_t} \right) \right\}. \tag{209}
\]

\[
K_t = K_t^b = \mathbb{E}\left\{ E_t^b \left(E_t^b\right)^T \right\} \text{ satisfies the generalized DRE}
\]

\[
K_{t+1}^b = \Lambda_t K_t A_t^T + B_t K_{W t} N_t^T + \left( B_t K_{W t} N_t^T + A_t K_t^b (\Lambda_t + C_t)^T \right) \left( N_t K_{W_t N_t^T} + K_{Z_t} \right)^{-1} \left( B_t K_{W_t N_t^T} + A_t K_t^b (\Lambda_t + C_t)^T \right)^T, \quad K_t^b \geq 0, \quad K_1^b = 0, \quad t = 1, \ldots, n. \tag{210}
\]

and the statements of parts (a) and (b) hold.

**Proof.** (a) The statements about Lemma 1 follow from Remark 7. (b) The statements about Theorem 3 are easily verified by replacing all conditional expectations, distributions, etc., for a fixed initial state \(S_1 = S_t^b = s\), and by using part (a), i.e., \((\Sigma_t^b, \hat{S}_t^b) = (0, \hat{S}_t^b), t = 1, 2, \ldots\). Part (c) follows from parts (a) and (b).
2.6. Comments on Past Studies

It is easily verified that Yang, Kavcic and Tatikonda [8] analyzed $C_{fb}^n(\kappa, s)$, which is defined by (81), under the Case (II) formulation, i.e., Conditions 1 and 2 of Section 1.1 hold, as discussed in the next remark.

**Remark 11.** Prior studies on the time-invariant stationary noise of PSD (41)
Yang, Kavcic and Tatikonda [8] analyzed the AGN channel driven by a stationary noise with PSD defined by (41) (see [8] (Theorem 1)). The special case of (126) is found in [8] (Section VI.B, Theorem 7).
The analysis in [8] presupposed the following formulation:
(i) The code is $(s, 2^n R, n), n = 1, 2, \ldots$, where $S_1 = S_1^* = s$ is the initial state of the noise, known to the encoder and the decoder, as discussed in Definition 3;
(ii) Conditions 1 and 2 of Section 1.1 hold;
(iii) The $n$–FTFI capacity formula is $C_{fb}^n(\kappa, s)$, defined by (81).
We emphasize that in [8] (Section II.C), a specific realization of the PSD is considered to ensure that Conditions 1 and 2 hold, i.e., the analysis in [8] presupposed a stationary noise and the Case (II) formulation.

Now, we ask the following: Given the PSD of the noise defined by (41), and the double-sided realization [4] (Equation (58)), i.e., the analog of time-invariant version of the PO-SS realization of Definition 2, or its analogous one-sided realization, what are the necessary conditions for the feedback capacity of [4] (Theorem 6.1) to be valid?
The answer to this question is as follows: Conditions 1 and 2 of Section 1.1 are necessary conditions. We show this in the next proposition.

**Proposition 2.** Conditions for validity of the feedback capacity characterization of [4] (Theorem 6.1)
Consider the AGN channel (1) driven by a stationary noise with PSD defined by (41) with the double-sided or one-sided realization [4] (Equation (58)), (i.e., analog of time invariant of Definition 2).
Then, a necessary condition for [4] (Theorem 6.1) to hold is

$$
P_{X_t|X_{t-1}, Y_{t-1}^{\infty}} = P_{X_t|S_t, Y_{t-1}^{\infty}}, \quad t = 1, \ldots, (211)$$

Further, Conditions 1 and 2 of Section 1.1 are necessary and sufficient for Equality (211) to hold.

**Proof.** See Section A.5. □

The next remark is our final observation on prior studies.

**Remark 12.** Comparison of Cover and Pombra Characterization and current literature
From Corollary 7 and Proposition 2, we have the following:
The characterization of feedback capacity given in [4] (Theorem 6.1, $C_{FB}$) corresponds to the Case (II) formulation and not to the Case (I) formulation. Further, the optimization problem of [4] (Theorem 6.1, $C_{FB}$) is precisely the optimization problem investigated in [8] (Section VI), with the additional restriction that the innovations’ part of the channel input is taken to be asymptotically zero in [4] (Theorem 6.1, $C_{FB}$), i.e., see [4] (Lemma 6.1 and comments above it). Recent studies [7,9,11,12] should be read with caution because the results therein often build on [4] (Theorems 4.1 and 6.1).

3. Asymptotic Analysis for Case (I) Formulation

In this section, we address the asymptotic per unit time limit of the $n$–FTFI capacity. Our analysis includes the following:
(1) Fundamental differences of entropy rates of jointly Gaussian stable versus unstable noise processes.

(2) Necessary and/or sufficient conditions for existence of entropy rates of unstable (and stable) \( Y^n, n = 1, 2, \ldots \), and \( V^n, n = 1, 2, \ldots \), expressed in terms of detectability and stabilizability or unit circle controllability conditions of generalized DREs [16,17], and asymptotic stationarity of the optimal input process \( X^n, n = 1, 2, \ldots \) (and output process \( Y^n, n = 1, 2, \ldots \), if the noise is stable).

This section also reconfirms that, in general, the asymptotic analysis of the \( n \)-FTFI capacity of a feedback code that depends on the initial state of the channel, i.e., \( S_1 = S_1^n = s \) is fundamentally different from code that does not depend on the initial state.

Closed-form expressions of the asymptotic per unit time limit of \( C_{f b}^n(k, s) \) of AGN channels driven by AR\((c)\), \( c \in (-\infty, \infty) \) noise, i.e., stable and unstable, are found in [1].

Closed-form expressions of the asymptotic per unit time limit of \( C_{n}^f(k) \) of AGN channels driven by ARMA\((a, c)\), \( a \in (-\infty, \infty), c \in (-\infty, \infty) \) noise are found in [3].

We consider the following definition of rate, often used for nonfeedback capacity of stationary processes; however, our formulations does not assume stationarity.

**Definition 4.** Per unit time limit of \( C_{f b}^n(k) \) and \( C_{n}^f(k, s) \)

Consider the AGN channel defined by (1), driven by the time-invariant PO-SS realization of Definition 2.

(a) For the code of Definition 1, define the per unit time limit

\[
C_{f b}^n(k) \triangleq \sup_{\lim_{n \to \infty} \frac{1}{n} E \left\{ \sum_{i=1}^{n} (x_i)^2 \right\} \leq k} \lim_{n \to \infty} \frac{1}{n} \left\{ H(Y^n) - H(V^n) \right\}
\]  

where, for problem \( C_{f b}^n(k) \), the supremum is taken over all asymptotically time-invariant distributions with feedback \( P_{X|Y^{-1},Y^{t-1}} = P_{X|Y^{-1},Y^{t-1}}, t = 1, 2, \ldots \), such that the limit exists and the supremum exists and it is finite.

(b) For the code of Definition 3, i.e., \( (s, 2^n R, n), n = 1, 2, \ldots \), with initial state \( S_1 = S_1^n = s \), \( C_{f b}^n(k) \) is replaced by \( C_{f b}^n(k, s) \), defined by (212), with differential entropies, conditional expectations, and conditional distributions, defined for fixed \( S_1^n = s \).

The rate definition, \( C_{f b}^n(k) \), i.e., the interchange of limit and supremum, is consistent with the definition of rates considered in [4,7,9,11,12]. However, unlike [4,7,9,11,12], we treat the general time-invariant, stable and unstable PO-SS noise realization of Definition 2, which is not necessarily stationary or asymptotically stationary.

We should emphasize that, in general, and irrespective of whether the noise is stable or unstable, the entropy rates that appear in (212) and (213) may not exist. To show existence of the limits \( C_{f b}^n(k), C_{f b}^n(k) \) and \( C_{f b}^n(k, s) \), we identify necessary and/or sufficient conditions, using the characterization of Theorem 3, when the channel input strategies are restricted to asymptotically time-invariant strategies \( \lim_{n \to \infty} \Lambda_n = \Lambda^\infty, \lim_{n \to \infty} K_{Z_n} = K_Z^\infty \). Clearly, by (212), whether the limit as, \( n \to \infty \) and the supremum over channel input distributions exist, depend on the convergence properties of the coupled generalized matrix DREs, \( \Sigma_n, K_n \), as \( n \to \infty \).

3.1. Entropy Rates of Gaussian Processes

First, we recall the following definition, which is standard and found in many textbooks:
Definition 5. Entropy rate of continuous-valued random processes

Let \( X_t : \Omega \to \mathbb{R}^{n_x}, n_x \in \mathbb{Z}_+ \) be a random process defined on some probability space \( (\Omega, \mathcal{F}, P) \). The entropy rate (differential) is defined by

\[ H_R(X^\infty) \triangleq \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \]  

when the limit exists.

The next theorem quantifies the existence of entropy rates of stationary Gaussian processes [16].

Theorem 4. The entropy rate of stationary zero-mean full-rank Gaussian process [16]

Let \( X_t : \Omega \to \mathbb{R}^{n_x}, n_x \in \mathbb{Z}_+, \forall t \in \mathbb{Z}_+ \) be a stationary Gaussian process, with a zero mean, and full-rank covariance of \( X^n \). Let \( \mathcal{H}_t^X \) denote the Hilbert space of RVs generated by \( \{ X_t : s \leq t, s, t \in \mathbb{Z}_+ \} \), and define the innovations process by

\[ \Sigma_t \triangleq \mathbb{E} \left\{ \left( X_t - \mathbb{E} \{ X_t \} \right) \left( X_t - \mathbb{E} \{ X_t \} \right)^T \right\} > 0 \]  

and its limit

\[ \Sigma \triangleq \lim_{n \to \infty} \Sigma_n \]  

Then, the entropy rate is given by

\[ H_R(X^\infty) = \frac{n_x}{2} \log(2\pi e) + \frac{1}{2} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \log |\Sigma_t| \]  

when it exists.

An application of Theorem 4 is given in the next proposition [15].

Proposition 3. Entropy rate of Gaussian process described by PSD (41)

Let \( V_t, \forall t \in \mathbb{Z}_+ \) be a real, scalar-valued, stationary Gaussian noise with PSD (41), with a corresponding time-invariant stationary realization (similar to Definition 2). Then, the entropy rate is given by

\[ H_R(V^\infty) = \frac{1}{2} \log(2\pi e K_W). \]  

Proof. This is shown in [15] by using the Szego formula and Poisson’s integral formula.

The next remark is trivial; it is introduced for a subsequent comparison.

Remark 13. Let \( V_t, \forall t \in \mathbb{Z}_+ \) be the non-stationary ARMA\((a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)\) noise of Example 2. Then, the conditional entropy of \( V^n \) for fixed initial state \( S_1 = S_1^s = s \) is given by

\[ H_R(V^\infty | s) \triangleq \lim_{n \to \infty} \frac{1}{n} H(V^n | s) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{2} \log(2\pi e K_W) = \frac{1}{2} \log(2\pi e K_W). \]  

The next lemma identifies fundamental conditions for the existence of the entropy rate of the time-varying PO-SS noise realization of Definition 2 (if \( S_1 = S_1^s = s \) is not fixed)
and includes the entropy rate $H_R(V^\infty)$ of the non-stationary ARMA $(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise of Remark 13.

**Lemma 2.** Entropy rate of the time-varying PO-SS noise realization of Definition 2  
Consider the time-varying PO-SS noise realization of Definition 2. Then, the following hold:
(a) The joint entropy of $V^n$, when it exists, is given by

$$H(V^n) = \sum_{t=1}^n H(\hat{I}_t) = \frac{1}{2} \sum_{t=1}^n \log (2\pi e K_n)$$  \hspace{1cm} (221)

where $\hat{I}_t, t = 1, \ldots, n$ is a zero-mean covariance $K_n \triangleright \text{cov}(\hat{I}_t, \hat{I}_t)$, Gaussian orthogonal innovation process of $V^n$ that is defined by

$$\hat{I}_t \triangleq V_t - \mathbb{E}\{V_t \mid V^{t-1}\}, \quad t = 1, \ldots, n$$  \hspace{1cm} (222)

that is, $\hat{I}_t$ is independent of $\hat{I}_k, \forall k \neq t$.  
(b) Suppose that the sequence $K_n, t = 1, 2, \ldots, n$ is such that

$$\lim_{n \to \infty} K_n = K^\infty_1 > 0.$$  \hspace{1cm} (223)

Then, the entropy rate of $V_t, \forall t \in \mathbb{Z}_+$ is given by

$$H_R(V^\infty) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n H(\hat{I}_t) = \frac{1}{2} \log (2\pi e K^\infty_1).$$  \hspace{1cm} (224)

**Proof.** See Appendix A.6. \hfill $\Box$

**Remark 14.** Entropy rate of non-stationary Gaussian noise  
By Lemma 2, a necessary condition for existence of the entropy rate of non-stationary Gaussian process $V^n$ is the convergence of the covariance of the Gaussian orthogonal innovations process of $V^n$, i.e., of $K_n \triangleright \text{cov}(\hat{I}_t, \hat{I}_t)$, since $\lim_{n \to \infty} \frac{1}{n} H(V^n) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n H(\hat{I}_t)$. We can determine such necessary and/or sufficient conditions from the convergence properties of the generalized Kalman filter equations [16,17] of Lemma 1.

3.2. Convergence Properties of Generalized Matrix DREs to AREs  
To address the asymptotic properties of estimation errors generated by the recursions of generalized Kalman filters, such as $\hat{E}_t, t = 1, 2, \ldots$ of Theorem 3, generated by (178), we need to introduce the stabilizing solutions of generalized AREs. The next definition is useful in this respect.

**Definition 6.** Stabilizing solutions of generalized matrix AREs  
Let $(A, C, Q, S, R, C) \in \mathbb{R}^{q \times q} \times \mathbb{R}^{q \times k} \times \mathbb{R}^{k \times k} \times \mathbb{R}^{k \times p} \times \mathbb{R}^{p \times p} \times \mathbb{R}^{p \times q}$. Define the generalized time-invariant matrix DRE

$$P_{t+1} = AP_tC^T + GQC^T - \left( AP_tC^T + GS \right) \left( R + CP_tC^T \right)^{-1} \left( AP_tC^T + GS \right)^T, \quad P_1 = \text{given},$$  \hspace{1cm} (225)

$$P_t \in \mathbb{S}^{n \times q}_+, t = 1, \ldots, \quad R = R^T > 0,$$

$$F^{CL}(P) \triangleq A - \left( AP_tC^T + GS \right) \left( R + CP_tC^T \right)^{-1} C.$$
Moreover, define the corresponding generalized matrix ARE as follows:

$$P = APA^T + GQG^T - \left( APC^T + GS \right) \left( R + CPC^T \right)^{-1} \left( APC^T + GS \right)^T, \quad P \in S_+^{n \times n}. \quad (226)$$

A solution $P = P^T \succeq 0$ to the generalized matrix ARE (226), assuming it exists, is called stabilizing if $\text{spec}(F^e(P)) \in \mathbb{D}_o$. In this case, we say that $F^e(P)$ is asymptotically stable, i.e., the eigenvalues of $F^e(P)$ are stable.

With respect to any of the above generalized matrices DRE and ARE, we introduce the important notions of detectability, unit circle controllability, and stabilizability. We use these notions to characterize the convergence properties of solutions of generalized matrix DREs, $P_n$, as $n \to \infty$, to a unique symmetric, non-negative, stabilizing solution $P$ of the generalized matrix ARE. These notions are used to identify necessary and/or sufficient conditions for the error recursions of generalized Kalman filters, such as $\tilde{E}_t, \quad t = 1, 2, \ldots$ of Theorem 3, generated by (178), to converge in a mean square sense, to a unique limit. However, we should distinguish whether the convergence is uniform for all initial conditions, $P_1 \succeq 0$ or only for $P_1 \succ 0$.

Definition 7. Detectability, Stabilizability, and Unit Circle Controllability

Consider the generalized matrix ARE of Definition 6 and introduce the matrices

$$A^* \triangleq A - GSR^{-1}C, \quad B^* \triangleq Q - SR^{-1}S^T, \quad B^* = B^* + \frac{1}{2} \left( B^* \right)^T. \quad (227)$$

(a) The pair $\{ A, C \}$ is called detectable if there exists a matrix $K \in \mathbb{R}^{1 \times p}$ such that $\text{spec}(A - KC) \in \mathbb{D}_o$, i.e., the eigenvalues $\lambda$ of $A - KC$ lie in $\mathbb{D}_o$ (stable).
(b) The pair $\{ A^*, GB^* \}$ is called unit circle controllable if there exists a $K \in \mathbb{R}^{k \times q}$ such that $\text{spec}(A^* - GB^* K) \notin \{ c \in \mathbb{C} : |c| = 1 \}$, i.e., all eigenvalues $\lambda$ of $A^* - GB^* K$ are such that $|\lambda| \neq 1$.
(c) The pair $\{ A^*, GB^* \}$ is called stabilizable if there exists a $K \in \mathbb{R}^{k \times q}$ such that $\text{spec}(A^* - GB^* K) \in \mathbb{D}_o$, i.e., all eigenvalues $\lambda$ of $A^* - GB^* K$ lie in $\mathbb{D}_o$.
(d) The pair $\{ A, C \}$ is called observable if the rank condition holds:

$$\text{rank}(O) = q, \quad O \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}. \quad (228)$$

(e) The pair $\{ A^*, GB^* \}$ is called controllable if the rank condition holds:

$$\text{rank}(C) = q, \quad C \triangleq \begin{bmatrix} GB^* \\ A^*GB^* \\ \vdots \\ (A^*)^{q-1}GB^* \end{bmatrix}. \quad (229)$$

Remark 15. The following are well known [116]. If the pair $\{ A, C \}$ is observable, then it is detectable. If the pair $\{ A^*, GB^* \}$ is controllable, then it is stabilizable.

The next theorem characterizes detectability, unit circle controllability, and stabilizability [17,40].

Lemma 3 ([17,40]). Necessary and sufficient conditions for detectability, unit circle controllability, and stabilizability

(a) The pair $\{ A, C \}$ is detectable if and only if there exists no eigenvalue and eigenvector $\{ \lambda, x \}$,
Ax = λx, such that |λ| ≥ 1 and such that Cx = 0.

(b) The pair \( \{ A^*, GB^{*\frac{1}{2}} \} \) is unit circle controllable if and only if there exists no eigenvalue and eigenvector \( (λ, x^*) \), \( x^T(A^*)^T = x^Tλ \), such that \( |λ| = 1 \) and such that \( x^T GB^{*\frac{1}{2}} = 0 \).

(c) The pair \( \{ A^*, GB^{*\frac{1}{2}} \} \) is stabilizable if and only there exists no eigenvalue and eigenvector \( (λ, x^*) \), \( x^T(A^*)^T = x^Tλ \), such that \( |λ| ≥ 1 \) and such that \( x^T GB^{*\frac{1}{2}} = 0 \).

In the next theorem, we summarize known results on sufficient and/or necessary conditions for the convergence of solutions \( \{ P_t, t = 1, 2, \ldots, n \} \) of the general time-invariant DRE (225), as \( n \to ∞ \), to a symmetric, non-negative \( P > 0 \), stabilizing solution of the corresponding generalized ARE (226). We recall that the pair \( \{ A, C \} \) is detectable, which is a necessary condition for the convergence of the sequence \( \{ P_t, t = 1, 2, \ldots, n \} \) as \( n \to ∞ \) to a non-negative \( P \), which is a stabilizing solution of a corresponding generalized ARE. However, it is not sufficient. To have a sufficient condition, it is necessary that the pair \( \{ A^*, GB^{*\frac{1}{2}} \} \) is unit circle controllable; however, the limiting \( P \) is necessarily the unique solution of \( P > 0 \) of the generalized ARE. There may be multiple solutions depending on the initial condition \( P_1 ≥ 0 \).

**Theorem 5** ([16,17]). **Convergence of time-invariant generalized DRE**

Let \( \{ P_t, t = 1, 2, \ldots, n \} \) denote a sequence that satisfies the time-invariant generalized DRE (225) with arbitrary initial condition \( P_1 ≥ 0 \). The following hold:

1. Consider the generalized DRE (225) with a zero initial condition, i.e., \( P_1 = 0 \), and assume that the pair \( \{ A, C \} \) is detectable and that the pair \( \{ A^*, GB^{*\frac{1}{2}} \} \) is unit circle controllable. Then, the sequence \( \{ P_t: t = 1, 2, \ldots, n \} \) that satisfies the generalized DRE (225), with a zero initial condition \( P_1 = 0 \), converges to \( P \), i.e., \( \lim_{n→∞} P_n = P \), where \( P \) satisfies the generalized matrix ARE (226) if and only if the pair \( \{ A^*, GB^{*\frac{1}{2}} \} \) is stabilizable.

2. Assume that the pair \( \{ A, C \} \) is detectable and that the pair \( \{ A^*, GB^{*\frac{1}{2}} \} \) is unit circle controllable. Then there exists a unique stabilizing solution \( P > 0 \) to the generalized ARE (226), i.e., such that \( \text{spec}(F^C(P)) ∈ \mathbb{D}_0 \) if and only if \( \{ A^*, GB^{*\frac{1}{2}} \} \) is stabilizable.

3. If \( \{ A, C \} \) is detectable and \( \{ A^*, GB^{*\frac{1}{2}} \} \) is stabilizable, then any solution \( P_t, t = 1, 2, \ldots, n \) to the generalized matrix DRE (225) with arbitrary initial condition \( P_1 ≥ 0 \) is such that \( \lim_{n→∞} P_n = P \), where \( P ≥ 0 \) is the unique solution of the generalized matrix ARE (226) with \( \text{spec}(F^C(P)) ∈ \mathbb{D}_0 \), i.e., it is stabilizing.

4. \( \{ A, C \} \) is detectable and \( \{ A^*, GB^{*\frac{1}{2}} \} \) unit circle controllable, which are necessary and sufficient conditions for any solution \( P_t, t = 2, \ldots, n \) to the generalized DRE (225) to converge, \( \lim_{n→∞} P_n = P \), from some initial condition \( P_1 ≥ 0 \), where \( P ≥ 0 \) is a stabilizing solution of the generalized ARE (226), but it may not be unique (i.e., (226) may have multiple solutions \( P ≥ 0 \)).

**Proposition 4.** Generalizations to asymptotic-time invariant coefficients

Suppose that the coefficients of the generalized matrix DRE (225), \( \{ A, C, G, Q, S \} \) are replaced by \( \{ A_t, C_t, G_t, Q_t, S_t \} \), \( t = 1, 2, \ldots, n \), and they are asymptotically time-invariant, i.e.,

\[
\lim_{n→∞} (A_n, C_n, G_n, Q_n, S_n) = (A, C, G, Q, S).
\]  

(230)

Then, Theorem 5 remains valid.

**Proof.** This is due to the well-known continuity properties of matrix DREs with respect to their coefficients, i.e., the convergence properties are characterized by the limiting pairs, \( \{ A, C \} \) and \( \{ A^*, GB^{*\frac{1}{2}} \} \). □
3.3. Feedback Rates

Now, we return to the feedback rates of Definition 4. The next corollary is an application of Theorem 5 to the generalized Kalman filter of Lemma 1 (for the time-invariant PO-SS realization); it identifies conditions for existence of the entropy rate $H_R(V^\infty)$, irrespective of whether the noise is stable or unstable.

**Corollary 8.** The entropy rate of PO-SS noise realization based on the generalized Kalman filter

Let $\Sigma^o_0 = \Sigma_0$, $t = 1, 2, \ldots$ denote the solution of the generalized matrix DRE (95) of the generalized Kalman filter of Lemma 1 of the time-invariant PO-SS realization of $V^n$ of Definition 2, i.e., $(A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_W), \forall t$ generated by

$$
\Sigma^o_{t+1} = A \Sigma^o_{t} A^T + BK_W B^T - \left( A \Sigma^o_{t} C^T + BK_W N^T \right) \left( NK_W N^T + C \Sigma^o_{t} C^T \right)^{-1} \left( A \Sigma^o_{t} C^T + BK_W N^T \right)^T, \quad \Sigma^o_0 \succeq 0, \quad t = 1, \ldots, n, \quad \Sigma^o_t = K_t \succeq 0. \quad (231)
$$

$$
M^{CL}(\Sigma^o) \triangleq A - M(\Sigma^o) C, \quad M(\Sigma^o) \triangleq \left( A \Sigma^o C^T + B_t K_W N^T \right) \left( NK_W N^T + C \Sigma^o C^T \right)^{-1}. \quad (232)
$$

Let $\Sigma^\infty = \Sigma^\infty_t \succeq 0$ be a solution of the corresponding generalized ARE

$$
\Sigma^\infty = A \Sigma^\infty A^T + BK_W B^T - \left( A \Sigma^\infty C^T + BK_W N^T \right) \left( NK_W N^T + C \Sigma^\infty C^T \right)^{-1} \left( A \Sigma^\infty C^T + BK_W N^T \right)^T. \quad (233)
$$

Define the matrices

$$
G Q G^T \triangleq BK_W B^T, \quad G S \triangleq BK_W N^T, \quad R \triangleq NK_W N^T \implies G \triangleq B, \quad Q \triangleq K_W, \quad S \triangleq K_W N^T, \quad (234)
$$

$$
A^* \triangleq A - BK_W N^T \left( NK_W N^T \right)^{-1} C, \quad B^* \triangleq K_W - K_W N^T \left( NK_W N^T \right)^{-1} \left( K_W N^T \right)^T. \quad (235)
$$

(a) All statements of Theorem 5 hold with $(G, Q, S, R)$ as defined by (234), (235).

In particular, suppose the following:

(i) \{A, C\} is detectable;

(ii) \{A^*, GB^{+\frac{1}{2}}\} is stabilizable.

Then, any solution $\Sigma^o_t$, $t = 1, 2, \ldots, n$ to the generalized matrix DRE (231) with arbitrary initial condition $\Sigma^o_1 \succeq 0$ is such that $\lim_{n \to \infty} \Sigma^o_n = \Sigma^\infty$, where $\Sigma^\infty \succeq 0$ is the unique and stabilizing solution of the generalized matrix ARE (233), i.e., with $\text{spec}(M^{CL}(\Sigma^\infty)) \in \mathbb{D}_0$.

(b) Suppose that (i) and (ii) hold. The entropy rate of $V^n$ is given by

$$
H_R(V^n) = \lim_{n \to \infty} \frac{1}{2n} \sum_{t=1}^{n} \log \left( 2 \pi e \left[ C \Sigma^o_t C^T + NK_W N^T \right] \right) \quad (236)
$$

$$
= \overline{H}\left( R^o_t \right) \triangleq \frac{1}{2} \log \left( 2 \pi e \left[ C \Sigma^\infty C^T + NK_W N^T \right] \right), \quad \forall \Sigma^o_1 \succeq 0, \quad \forall t \quad (237)
$$

where

$$
\overline{I}^o_t \triangleq C(S_t - S^T_t) + NW_t \in N(0, C \Sigma^\infty C^T + NK_W N^T), \quad t = 1, 2, \ldots. \quad (238)
$$

is the stationary Gaussian innovations process, i.e., with $\Sigma^o_t$ replaced by $\Sigma^\infty$, and the entropy rate $H_R(V^n)$ is independent of the initial data $\Sigma^o_1 \succeq 0$.

(c) Suppose in parts (a) and (b), the condition \{A^*, GB^{+\frac{1}{2}}\} is stabilizable is replaced by

(iii) \{A^*, GB^{+\frac{1}{2}}\} is unit circle controllable.
Then, the statements of parts (a) and (b) hold, for some \( \Sigma_t^0 \geq 0 \), but not for all \( \Sigma_t^0 \geq 0 \). Moreover, \( \hat{\Sigma}^0 \) is not necessarily a stationary process, i.e., it depends on the value of \( \Sigma_t^0 \geq 0 \).

**Proof.** (a) These are direct applications of Theorem 5. (b) This follows from Lemma 2. (c) This is due to Theorem 5(4). \( \square \)

Next, we apply Corollary 8 to the non-stationary AR(a,c), \( a \in (-\infty, \infty), c \in (-\infty, \infty) \) noise.

**Lemma 4. Properties of solutions of DREs and AREs of AR(a,c), \( a \in (-\infty, \infty), c \in (-\infty, \infty) \) noise and entropy rate \( H_R(V^\infty) \).**

Consider the AR(a,c), \( a \in (-\infty, \infty), c \in (-\infty, \infty) \) noise of Example 2(a), and the DRE \( \Sigma_t^0 = \Sigma_t, t = 1, \ldots, n \), generated by Corollary 6(a), i.e.,

\[
\Sigma_{t+1}^0 = (c)^2 \Sigma_t^0 + K_W - \left( c \Sigma_t^0 (c-a) + K_W \right)^2 \left( K_W + (c-a)^2 \Sigma_t^0 \right)^{-1}, \quad t = 1, \ldots, n, \tag{239}
\]

\[
\Sigma_0^0 = K_{S_1} = \frac{\left( c_0^0 \right)^2 K_{S_0} + a_0^2 K_{W_0}}{(c_0-a_0)^2} \geq 0. \tag{240}
\]

where \( K_W > 0, c \neq a, K_{S_0} \geq 0, K_{W_0} \geq 0. \) Let \( \Sigma^\infty \geq 0 \) be a solution of the corresponding generalized ARE, as follows:

\[
\Sigma^\infty = \left( c \right)^2 \Sigma^\infty + K_W - \left( \left( \Sigma^\infty \right) (c-a) + K_W \right)^2 \left( K_W + (c-a)^2 \Sigma^\infty \right)^{-1}. \tag{241}
\]

Then, the detectability and stabilizability pairs are

\[
\{ A, C \} = \{ c, c-a \}, \quad \{ A^*, GB^{b+\frac{1}{2}} \} = \{ a, 0 \}. \tag{242}
\]

and the following hold:

1. The pair \( \{ A, C \} = \{ c, c-a \} \) is detectable \( \forall c \in (-\infty, \infty), a \in (-\infty, \infty) \) (the restriction \( c \neq a \) is always assumed).
2. The pair \( \{ A^*, GB^{b+\frac{1}{2}} \} = \{ a, 0 \} \) is unit circle controllable if and only if \( |a| \neq 1 (\forall c \in (-\infty, \infty)) \).
3. The pair \( \{ A^*, GB^{b+\frac{1}{2}} \} = \{ a, 0 \} \) is stabilizable if and only if \( a \in (-1,1) (\forall c \in (-\infty, \infty)) \).
4. Suppose \( c \in (-\infty, \infty) \) and \( a \in (-\infty, \infty) \). The sequence \( \{ \Sigma_t^0, t = 1, 2, \ldots, n \} \) that satisfies the generalized DRE with any initial condition, \( \Sigma_1^0 \geq 0 \), converges to \( \Sigma^\infty \), i.e., \( \lim_{n \to \infty} \Sigma_n^0 = \Sigma^\infty \), where \( \Sigma^\infty \geq 0 \) satisfies the ARE (241) if and only if the \( \{ A^*, GB^{b+\frac{1}{2}} \} = \{ a, 0 \} \) is unit circle controllable, equivalently, \( |a| \neq 1 \). Moreover, the solutions of the quadratic Equation (241), without imposing \( \Sigma^\infty \geq 0 \) are

\[
\Sigma^\infty = \begin{cases} 0 & \text{unique, stabilizing, } \Sigma^\infty \geq 0 \text{ sol. of (241) for } c \in (-\infty, \infty), |a| < 1 \\ \frac{K_w (a^2-1)}{(c-a)^2} & > 0 \text{ maximal, stabilizing, } \Sigma^\infty > 0 \text{ sol. of (241) for } c \in (-\infty, \infty), |a| > 1 \\ \frac{K_w (a^2-1)}{(c-a)^2} & < 0 \text{ non-stabilizing, } \Sigma^\infty < 0 \text{ sol. of (241) for } c \in (-\infty, \infty), |a| < 1. \end{cases} \tag{243}
\]

That is, \( \lim_{n \to \infty} \Sigma_n^0 = \Sigma^\infty = 0, \forall \Sigma_0^0 \geq 0 \), is the unique and stabilizing solution \( \Sigma^\infty \geq 0 \) of (241), i.e., such that \( |M^{CL}(\Sigma^\infty)| < 1, \text{ if and only if } |a| < 1, \text{ and} \)

\[
\lim_{n \to \infty} \Sigma_n^0 = \Sigma^\infty = \frac{K_w (a^2-1)}{(c-a)^2} > 0, \forall \Sigma_0^0 > 0 \text{ is the maximal and stabilizing solution } \Sigma^\infty \text{ of (241), i.e., such that } |M^{CL}(\Sigma^\infty)| < 1, \text{ if and only if } |a| > 1. \]

5. Suppose \( c \in (-\infty, \infty) \) and \( |a| < 1 \). Then, any solution \( \Sigma_t^0, t = 1, 2, \ldots, n \) to the generalized
DRE \((239)\) with an arbitrary initial condition, \(\Sigma_0^o \geq 0\), is such that \(\lim_{n \to \infty} \Sigma_n^o = \Sigma^\infty\), where \(\Sigma^\infty \geq 0\) is the unique solution of the generalized ARE \((241)\) with \(M^C(\Sigma^\infty) \in (-1, 1)\), i.e., it is stabilizing. Moreover, \(\Sigma^\infty = 0\).

(6) (i) Suppose \(c \in (-\infty, \infty)\) and \(|a| < 1\). The entropy rate of \(V_t\), \(\forall t \in \mathbb{Z}_+\) is given by

\[
H_R(V^\infty) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n \frac{1}{2} \log \left(2\pi e \left[(c-a)^2 \Sigma_t^o + K_W\right]\right)
\]

\[
= \frac{1}{2} \log (2\pi e K_W), \quad \forall \Sigma_t^o \geq 0.
\]

(ii) Suppose \(c \in (-\infty, \infty)\) and \(|a| > 1\). The entropy rate of \(V_t\), \(\forall t \in \mathbb{Z}_+\) is given by

\[
H_R(V^\infty) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n \frac{1}{2} \log \left(2\pi e \left[(c-a)^2 \Sigma_t^o + K_W\right]\right)
\]

\[
= \frac{1}{2} \log (2\pi e K_W a^2), \quad \forall \Sigma_t^o > 0.
\]

Proof. See Appendix A.7.

Remark 16. Lemma 4(4) emphasizes the fact that in the asymptotic analysis of \(\{\Sigma_t^o, t = 1, 2, \ldots\}\), which satisfies the DREs \((239)\) and \((240)\), its limiting value, \(\lim_{n \to \infty} \Sigma_n^o = \Sigma^\infty\), where \(\Sigma^\infty \geq 0\) satisfies the ARE \((241)\), with two solutions \(\Sigma^\infty = 0\) and \(\Sigma^\infty = \frac{K_W(a^2-1)}{(c-a)^2}\). For any \(c \in (-\infty, \infty)\), it is clear that for \(|a| < 1\), the unique and stabilizing solution is \(\Sigma^\infty = 0\), \(\forall \Sigma_t^o \geq 0\), since the other solution \(\Sigma^\infty = \frac{K_W(a^2-1)}{(c-a)^2} < 0\) is negative. On the other hand, for any \(c \in (-\infty, \infty)\) and \(|a| > 1\), the stabilizing solution is the maximal solution, \(\Sigma^\infty = \frac{K_W(a^2-1)}{(c-a)^2} > 0\), provided \(\Sigma_t^o > 0\).

To gain additional insights, we discuss the application of Lemma 4 to the AR(c), \(c \in (-\infty, \infty)\) noise in the next remark.

Remark 17. Entropy rate \(H_R(V^\infty)\) of the AR(c), \(c \in (-\infty, \infty)\) noise

Consider the non-stationary AR(c), \(c \in (-\infty, \infty)\) noise defined by \((124)\). Then, from Lemma 4, \(\Sigma_t^o, t = 1, \ldots, n\) is the solution of \((239)\) and \((240)\), with \(a = 0\) (see Corollary 6(b), \((201)\)), and \((241)\) degenerates to the ARE, as follows:

\[
\Sigma^\infty = (c)^2 \Sigma^\infty + K_W - \left((c)^2 \Sigma^\infty + K_W\right)^2 \left(K_W + (c)^2 \Sigma^\infty\right)^{-1}
\]

\[
(248)
\]

For \(a = 0\) by \((242)\), the pair \(\{A, C\} = \{c, c\}\) is detectable, and the pair \(\{A^*, GB^{+\frac{1}{2}}\} = \{0, 0\}\) is stabilizable. The two solutions of the ARE \((248)\), without imposing \(\Sigma^\infty \geq 0\), are

\[
\Sigma^\infty = \begin{cases} 
0 & \text{the unique, stabilizing, non-negative solution of the ARE} \\
-\frac{K_W}{c^2} < 0 & \text{the non-stabilizing, negative solution of the ARE.}
\end{cases}
\]

\[
(249)
\]

That is, \(\lim_{n \to \infty} \Sigma_n^o = \Sigma^\infty \geq 0\), where \(\Sigma^\infty = 0\) is the unique (stabilizing) solution of the ARE and corresponds to the stable eigenvalue of the error equation (see \((93)\)), i.e., \(M^C(\Sigma^\infty) = c - \frac{K_W}{K_W}c = 0\).

Next, we compute the entropy rate \(H_R(V^\infty)\) of the time-invariant non-stationary PO-SS(a, c, b\(^1\), b\(^2\), d\(^1\), d\(^2\)) noise of Corollary 2 to show fundamental differences from the entropy rate \(H_R(V^\infty)\) of the AR(a, c) noise of Lemma 4.

Lemma 5. Properties of solutions of DREs and AREs of PO-SS(a, c, b\(^1\), b\(^2\), d\(^1\), d\(^2\)) noise and entropy rate \(H_R(V^\infty)\)
Consider the time-invariant non-stationary PO-SS \((a, c, b^1, b^2 = 0, d^1 = 0, d^2 = d)\) noise of Example 1, i.e., given by

\[
\begin{align*}
S_{t+1} &= aS_t + bW_t^1, \quad t = 1, 2, \ldots, n - 1 \\
V_t &= cS_t + dW_t^2, \quad t = 1, \ldots, n,
\end{align*}
\]

and the sequence \(\Sigma^n_t \triangleright \Sigma_t, t = 1, \ldots, n\), generated by the DRE of Lemma 1 (see (113), i.e.,

\[
\begin{align*}
\Sigma^n_{t+1} &= (a^2\Sigma^n_t + b^2K_{W^1}) \\
&\quad - \left(a\Sigma^n_t c^2 \left((d^2\Sigma^n_t + c^2\Sigma^n_t)^{-1} + 1, \quad t = 1, \ldots, n, \quad \Sigma^n_0 = K_S, \quad \Sigma^n_0 \geq 0
\end{align*}
\]

where \((b^2K_{W^1} \geq 0, (d^2K_{W^1} > 0). Let \(\Sigma^n_0 \geq 0\) be the corresponding solution of generalized ARE:

\[
\Sigma^n_0 = (a^2\Sigma^n_0 + b^2K_{W^1} - (a\Sigma^n c^2) \left((d^2\Sigma^n_t + c^2\Sigma^n_0)^{-1}
\]

Then, the detectability and stabilizability pairs are

\[
\{A, C\} = \{a, c\}, \quad \{A^*, GB^{+\frac{1}{2}}\} = \{a, b(K_{W^1})^{\frac{1}{2}}\}
\]

and the following hold:

1. The pair \(\{A, C\} = \{a, c\}\) is detectable \(\forall c \in (-\infty, \infty), a \in (-\infty, \infty), c \neq 0\). If \(c = 0\) the pair \(\{A, C\} = \{a, 0\}\) is detectable if and only if \(|a| < 1\).

2. The pair \(\{A^*, GB^{+\frac{1}{2}}\} = \{a, b(K_{W^1})^{\frac{1}{2}}\}\) is unit circle controllable if and only if \(|b(K_{W^1})^{\frac{1}{2}}| \neq 1, \forall a \in (-\infty, \infty), c \in (-\infty, \infty)\).

3. The pair \(\{A^*, GB^{+\frac{1}{2}}\} = \{a, b(K_{W^1})^{\frac{1}{2}}\}\) is stabilizable if \(b(K_{W^1})^{\frac{1}{2}} \neq 0, \forall a \in \infty, c \in (-\infty, \infty). If \(b(K_{W^1})^{\frac{1}{2}} = 0\) the pair \(\{A^*, GB^{+\frac{1}{2}}\} = \{a, 0\}\) is stabilizable if and only if \(|a| < 1\).

4. Define the set

\[
\mathcal{L}^\infty \triangleq \{(a, c, (b^2K_{W^1}) \in (-\infty, \infty)^2 \times [0, \infty) : \ (i) \ the \ pair \ \{A, C\} = \{a, c\} \ is \ detectable, \ and \\
(ii) \ the \ pair \ \{A^*, GB^{+\frac{1}{2}}\} = \{a, b(K_{W^1})^{\frac{1}{2}}\} \ is \ stabilizable\}.
\]

For any \((a, c, b(K_{W^1})^{\frac{1}{2}}) \in \mathcal{L}^\infty\), any solution \(\Sigma^n_t, t = 1, 2, \ldots, n\) to the (classical) DRE (252) with an arbitrary initial condition, \(\Sigma^n_0 \geq 0\) is such that \(\lim_{n \to \infty} \Sigma^n_0 = \Sigma^n_0\), where \(\Sigma^n \geq 0\) is the unique solution of the (classical) ARE (253) with \(M^{CL}(\Sigma^n) \in (-1, 1)\), i.e., it is stabilizing.

5. For any \((a, c, b^2K_{W^1}) \in \mathcal{L}^\infty\) of part (4), the entropy rate of \(V_t, \forall t \in \mathbb{Z}_+\) is given by

\[
H_R(Y^n) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{2} \log \left(2\pi e \left[(c^2\Sigma^n_t + (d^2K_{W^1}) - \left(\frac{1}{2} \log \left(2\pi e \left[(c^2\Sigma^n_0 + (d^2K_{W^1})\right), \quad \forall \Sigma^n_0 \geq 0.
\]

Proof. Follows from Theorem 5. □

Next, we turn our attention to the convergence properties of the entropy rate \(H_R(Y^n)\), which is needed for the characterization of \(C^{b, b}(x)\) of Definition 4.

**Theorem 6. Asymptotic properties of entropy rate** \(H_R(Y^n)\) **of Theorem 3**

Let \(K^n_t, t = 1, \ldots, \) be the solution of the generalized DRE (180) of the generalized Kalman filter of Theorem 3, corresponding to the time-invariant PO-SS realization of \(V^n\) of Defini-
tion 2, \((A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_W), \forall t,\) with time-invariant strategies \((A_t, K_{Z_t}) = (A^\infty, K_{Z^\infty}), \forall t,\) generated by

\[
K_{t+1}^0 = A K_t^0 A^T + M(\Sigma_t^0) K_{t+1}^0 (M(\Sigma_t^0))^T
- \left( A K_t^0 (A^\infty + C)^T + M(\Sigma_t^0) K_{t+1}^0 \right) \left( K_{t+1}^0 + K_{Z^\infty}^0 + (A^\infty + C) K_{Z^\infty} (A^\infty + C)^T \right)^{-1}
- \left( A K_t^0 (A^\infty + C)^T + M(\Sigma_t^0) K_{t+1}^0 \right) K_{t+1}^0 = K_{t+1}^0 T, \quad t = 1, \ldots, n, \quad K_{t+1}^0 = 0
\]

where

\[
K_{t+1}^0 = C \Sigma_t^0 C^T + N K_W N^T, \quad \Sigma_t^0 \text{ is a solution of (231), } M(\Sigma_t^0) \text{ is given by (232)},
\]

\[
F^{CL}(\Sigma_t^0, K_t^0) \triangleq A - F(\Sigma_t^0, K_t^0) \left( \Lambda^\infty + C \right),
\]

\[
F(\Sigma_t^0, K_t^0) \triangleq \left( A K_t^0 (A^\infty + C)^T + M(\Sigma_t^0) K_{t+1}^0 \right) \left( K_{t+1}^0 + K_{Z^\infty}^0 + (A^\infty + C) K_{Z^\infty} (A^\infty + C)^T \right)^{-1},
\]

Define the corresponding generalized ARE by

\[
K^\infty = A K^\infty A^T + M(\Sigma^\infty) K_{f_0} (M(\Sigma^\infty))^T - \left( A K^\infty (A^\infty + C)^T + M(\Sigma^\infty) K_{f_0} \right) \left( K_{f_0} + K_{Z^\infty}^\infty \right.
+ (A^\infty + C) K^\infty (A^\infty + C)^T \left. \right)^{-1} \left( A K^\infty (A^\infty + C)^T + M(\Sigma^\infty) K_{f_0} \right)^T, \quad K^\infty = K^\infty T \succeq 0.
\]

where

\[
K_{f_0} = C \Sigma^\infty C^T + N K_W N^T, \quad \Sigma^\infty \text{ is a solution of (233), } M(\Sigma^\infty) \text{ is given by (232)}.
\]

Introduce the matrices

\[
C(\Lambda^\infty) \triangleq \Lambda^\infty + C, \quad G Q C^T \triangleq M(\Sigma^\infty) K_{f_0} (M(\Sigma^\infty))^T,
\]

\[
G S \triangleq M(\Sigma^\infty) K_{f_0}, \quad R(K_{Z^\infty}^\infty) \triangleq K_{f_0} + K_{Z^\infty}^\infty.
\]

\[
\implies G \triangleq M(\Sigma^\infty), \quad Q \triangleq K_{f_0}, \quad S \triangleq K_{f_0},
\]

\[
A^*(\Lambda^\infty, K_{Z^\infty}^\infty) \triangleq A - M(\Sigma^\infty) K_{f_0} \left( K_{f_0} + K_{Z^\infty}^\infty \right)^{-1} (\Lambda^\infty + C),
\]

\[
B^*(K_{Z^\infty}^\infty) \triangleq K_{f_0} - K_{f_0} \left( K_{f_0} + K_{Z^\infty}^\infty \right)^{-1} K_{f_0}.
\]

Suppose that the detectability and stabilizability conditions of Corollary 8(i, ii) hold.
Then, all statements of Theorem 5 hold with \((C(\Lambda^\infty), G, Q, S, R(K_{Z^\infty}^\infty))\) as defined by (265).
In particular, suppose
(i) \(\{A, C(\Lambda^\infty)\}\) is detectable;
(ii) \(A^*(\Lambda^\infty, K_{Z^\infty}^\infty), GB^*(K_{Z^\infty}^\infty)\) is stabilizable.
Then, any solution \(K_{t+1}^0, t = 1, 2, \ldots, n\) to the generalized matrix DRE (258) with an arbitrary initial condition \(K_1^0 \succeq 0\) is such that \(\lim_{t \to \infty} K_{t+1}^0 = K^\infty\), where \(K^\infty \succeq 0\) is the unique solution of the generalized matrix ARE (262) with \(\text{spec}(F^{CL}(K^\infty, \Sigma^\infty)) \in \mathbb{D}_0\), i.e., it is stabilizing.
Moreover, the entropy rate of \(Y^n\) is given by
Define the set $DRE(233)$, then the statements of Theorem 6 hold, as stated. In particular, under the detectability

$$K \in \mathbb{R}^n \times \mathbb{R}^n,$$

Lemma 6. Consider the AR $(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise of Example 2(a) using Lemma 4.

Proof. Since the detectability and stabilizability conditions of Corollary 8 hold, then the statements of Corollary 8 hold. By the continuity property of the solutions of generalized difference Riccati equations, with respect to its coefficients (see [16]), and the convergence of the sequence $\lim_{n \to \infty} \Sigma_n = \Sigma^\infty$, where $\Sigma^\infty \geq 0$ is the unique stabilizing solution of (233), then the statements of Theorem 6 hold, as stated. In particular, under the detectability and stabilizability Conditions (i) and (ii), $\lim_{n \to \infty} K_n = K^\infty$, where $K^\infty \geq 0$ is the unique and stabilizing solution of (262).

In the next lemma, we apply Theorem 6 to the AR $(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise of Example 2(a) using Lemma 4.

Lemma 6. Consider the AR $(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise of Example 2(a), and the DRE $\Sigma_t = \Sigma_t, t = 1, \ldots, n$ and ARE of Lemma 4, (239)–(242).

Let $K_t, t = 1, \ldots, n$ denote the solution of the DRE of Corollary 6(a), when $\Lambda_t = \Lambda^\infty, KZ_t = K^\infty, K_t = K^\infty_t, t \geq 0$ and $I_t \in \mathbb{R}^n$, $t = 1, \ldots, n$.

\begin{equation}
K_{t+1} = (c^2 - 1) K_t + (M(\Sigma_t)) K_{t-1} - \left( c K_t \right)^2 K_{t-1} K_{t-2}^{-1}, \quad t = 1, \ldots, n,
\end{equation}

\begin{equation}
K_t \geq 0, \quad t = 1, \ldots, n
\end{equation}

and where

\begin{equation}
M(\Sigma_t) = (c^2 - 1) K_t + K_{t-1} - \left( c K_t \right)^2 K_{t-1} K_{t-2}^{-1},
\end{equation}

\begin{equation}
K_{t-1} = (c - 1) K_{t-1} + K_{t-2}, \quad t = 1, \ldots, n.
\end{equation}

Define the set

\begin{equation}
\mathcal{L}^\infty = \left\{ (a, c) \in (-\infty, \infty)^2, a \neq c : \begin{array}{l}
(i) \text{the pair } \{A, C\} = \{a, c - a\} \text{ is detectable, and} \\
(ii) \text{the pair } \{A^*, GB^{+1}\} = \{a, 0\} \text{ is stabilizable}
\end{array} \right\}.
\end{equation}

For any $(a, c) \in \mathcal{L}^\infty$, let $K^\infty \geq 0$ be a corresponding solution of the ARE (evaluated at

\begin{equation}
0 \to \infty \Sigma_n = \Sigma^\infty = 0,
\end{equation}

\begin{equation}
H_R(Y^\infty) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} H(I_t)
\end{equation}

\begin{equation}
= \lim_{n \to \infty} \frac{1}{2n} \sum_{t=1}^{n} \log \left( 2 \pi e \left( \Lambda^\infty \Sigma^\infty + C \right)^{T} \right)
\end{equation}

\begin{equation}
\frac{1}{2} \left( \Lambda^\infty + C \right) K^\infty \left( \Lambda^\infty + C \right)^{T} + K_{t-1} + K^\infty_{t-2} \right), \quad \forall K^\infty \geq 0, \quad \forall t
\end{equation}

where $I_t, t = 1, \ldots, n$ is the innovations process of Theorem 3 (with indicated changes of time-invariant strategies) and where

\begin{equation}
I_t = (\Lambda^\infty + C) \left( \hat{S}_t^\infty - \hat{S}_t^\infty + Z_t \right)
\end{equation}

\begin{equation}
I_t \in \mathbb{N}(0, (\Lambda^\infty + C) K^\infty, (\Lambda^\infty + C)^{T} + K_{t-1} + K^\infty_{t-2}), \quad t = 1, 2, \ldots,
\end{equation}

is the stationary Gaussian innovations process, i.e., with $(K^\infty_t, \Sigma^\infty_t)$ replaced by $(K^\infty, \Sigma^\infty)$.
\[
K_\infty = (c_2 K_\infty + K_W - (c K_\infty (\Lambda_\infty + c - a) + K_W))^2 (K_W + K_\infty (\Lambda_\infty + c - a)^2 K_\infty)^{-1}.
\]  \hspace{1cm} (279)

\[
K_\infty^2 \geq 0, \quad K_W > 0.
\]  \hspace{1cm} (280)

and define the pairs

\[
\{A, C(\Lambda_\infty)\} = \{c, \Lambda_\infty + c - a\},
\]  \hspace{1cm} (281)

\[
\{A^*(\Lambda_\infty, K_\infty^2), GB^{\frac{1}{2}}(K_\infty^2)\} = \{c - K_W (K_W + K_\infty^2)^{-1} (\Lambda_\infty + c - a), (K_W - (K_W)^2 (K_W + K_\infty^2)^{-1})^{-1}\}.
\]  \hspace{1cm} (282)

Then, the following hold:

1. Suppose \(\Lambda_\infty + c - a \neq 0\). Then, \(\{A, C(\Lambda_\infty)\} = \{c, \Lambda_\infty + c - a\}\) is detectable \(\forall (a, c) \in (-\infty, \infty)^2\).
2. Suppose \(\Lambda_\infty + c - a = 0\). Then, \(\{A, C(\Lambda_\infty)\} = \{c, 0\}\) is detectable if and only if \(|c| < 1\) \(\forall a \in (-\infty, \infty)\).
3. Suppose \(K_\infty^2 = 0\). Then, the pair \(\{A^*, GB^{\frac{1}{2}}\} = \{-\Lambda_\infty + a, 0\}\) is unit circle controllable if and only if \(|\Lambda - a| \neq 1\) \(\forall a \in (-\infty, \infty)\).
4. Suppose \(K_\infty^2 = 0\). Then, the pair \(\{A^*, GB^{\frac{1}{2}}\} = \{-\Lambda_\infty + a, 0\}\) is stabilizable if and only if \(|\Lambda - a| < 1\) \(\forall a \in (-\infty, \infty)\).
5. Suppose \(\Lambda_\infty + c - a \neq 0, |\Lambda - a| \neq 1\) \(\forall (a, c) \in (-\infty, \infty)^2\), and \(K_\infty^2 = 0, \Sigma_n^\infty = 0\). The sequence \(K_t^G, t = 1, 2, \ldots, n\) that satisfies the generalized DRE (274) with a zero initial condition, \(K_0^G\), converges to \(K_\infty^G\), i.e., \(\lim_{n \to \infty} K_n^G = K_\infty^G\), where \(K_\infty^G\) satisfies the generalized ARE,

\[
K_\infty = (c_2 K_\infty + K_W - (c K_\infty (\Lambda_\infty + c - a) + K_W))^2 (K_W + (\Lambda_\infty + c - a)^2 K_\infty)^{-1}, \quad K_\infty \geq 0
\]  \hspace{1cm} (283)

if and only if \(|a| < 1\) (by Lemma 4(4)), and the pair \(\{A^*, GB^{\frac{1}{2}}\} = \{-\Lambda_\infty + a, 0\}\) is stabilizable, equivalently, \(|\Lambda_\infty - a| < 1\).

Moreover, the solutions of the ARE (283), under the stabilizability condition, i.e., \(|\Lambda_\infty - a| < 1\), are

\[
K_\infty^G = \begin{cases} 
0 & \text{the unique, stabilizing, } K_\infty^G \geq 0 \text{ sol. of (283) for } |\Lambda_\infty - a| < 1 \\
K_0^G \left( (\Lambda_\infty - a)^{2-1} \right)^{\frac{1}{\gamma}} & \text{the non-stabilizing, } K_\infty^G < 0 \text{ sol. of (283) for } |\Lambda_\infty - a| < 1.
\end{cases}
\]  \hspace{1cm} (284)

That is, \(\lim_{n \to \infty} M_\Sigma_n = \Sigma_\infty = 0\) is the unique and stabilizing solution \(\Sigma_\infty \geq 0\) of (283), i.e., such that \(|M^{\text{col}}(\Sigma_\infty)| < 1\), if and only if \(|\Lambda_\infty - a| < 1, |a| < 1\).

\textbf{Proof.} The statements follow from Lemma 4, Theorem 6 (and general properties of Theorem 5). \(\square\)

\textbf{Remark 18.} From Lemma 6(5), it follows that if \(K_\infty^2 = 0, \Sigma_1^\infty = 0\), then the unique and stabilizing solution is \(K_\infty^G = 0\) and corresponds to \(|\Lambda_\infty - a| < 1, a \in (-1, 1)\). This is an application of Theorem 5(1).

In the next theorem, we characterize the asymptotic limit of Definition 4 by invoking Theorems 3 and 6 and Corollary 8.

\textbf{Theorem 7.} Feedback capacity \(C^{\text{fb}, \rho}(k)\) of Theorem 3 for time-invariant strategies

Consider \(C^{\text{fb}, \rho}(k)\) of Definition 4 corresponding to Theorem 3, i.e., the PO-SS realization of \(V_n^\rho\) of
Definition 2 is time-invariant, \((A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_W), \forall t\), and the strategies are time-invariant, \((A_t, K_{Z_t}) = (\Lambda^\infty, K_Z^\infty), \forall t\).

Define the set
\[
P^\infty = \left\{ (\Lambda^\infty, K_Z^\infty) \in \mathbb{R}^2 : (i) \{A, C\} \text{ of Corollary 8 is detectable; } \\
(ii) \{A^*, GB^{\frac{1}{2}}\} \text{ of Corollary 8 is stabilizable, where } (A^*, B^*) \text{ is defined by (235); } \\
(iii) \{A, C(\Lambda^\infty)\} \in \{A, \Lambda^\infty + C\} \text{ of Theorem 6 is detectable; } \\
(iv) \{A^*(\Lambda^\infty, K_Z^\infty), GB^{\frac{1}{2}}(K_Z^\infty)\} \text{ of Theorem 6 is stabilizable; } \\
A^*(\Lambda^\infty, K_Z^\infty, B^*(K_Z^\infty)) \text{ is defined by (268)} \right\}.
\]

Then,
\[
C^{f, p}(\kappa) = \sup_{(\Lambda^\infty, K_Z^\infty) \in P^\infty} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \log \left( \frac{(\Lambda^\infty + C) K_Z^\infty (\Lambda^\infty + C)^T + \Sigma^\infty C^T + N K_W N^T + K_Z^\infty}{\Sigma^\infty C^T + N K_W N^T} \right) \right\}
\]
\[
\lim_{n \to \infty} \frac{1}{2n} \sum_{i=1}^{n} \log \left( \frac{(\Lambda^\infty + C) K_Z^\infty (\Lambda^\infty + C)^T + \Sigma^\infty C^T + N K_W N^T + K_Z^\infty}{\Sigma^\infty C^T + N K_W N^T} \right)
\]
where
\[
P^\infty(\kappa) = \left\{ (\Lambda^\infty, K_Z^\infty) \in P^\infty : K_Z^\infty \geq 0, \Lambda^\infty K_Z^\infty (\Lambda^\infty)^T + K_Z^\infty \leq \kappa, \right\}
\]
\[
\Sigma^\infty \text{ is the unique and stabilizing solution of (262), i.e., } |F^{CL}(\Sigma^\infty, K^\infty)| < 1
\]
\[
\Sigma^\infty \text{ is the unique stabilizing solution of (233), i.e., } |M^{CL}(\Sigma^\infty)| < 1
\]

provided there exists \(\kappa \in [0, \infty)\) such that the set \(P^\infty(\kappa)\) is non-empty.

Moreover, the maximum element \((\Lambda^\infty, K_Z^\infty) \in P^\infty(\kappa)\) is such that
(1) It induces asymptotic stationarity of the corresponding input and innovations processes (see Theorem 3 for specification);
(2) If \(V^n, n = 1, 2, \ldots, n\) is asymptotically stationary, then it induces asymptotic stationarity of the corresponding, input and output processes;
(3) For (i) and (ii), \(C^{f, p}(\kappa)\) is independent of the initial conditions \(K_t^0 \geq 0, \Sigma_t^0 \geq 0\).

Furthermore, if the set \(P^\infty(\kappa)\) is empty, replace stabilizability of \(\{A^*, GB^{\frac{1}{2}}\}\) and \(\{A^*(\Lambda^\infty, K_Z^\infty), GB^{\frac{1}{2}}(K_Z^\infty)\}\) by unit circle controllability, i.e., the maximal and stabilizing solutions \(K^\infty, \Sigma^\infty\) of the AREs are utilized.

**Proof.** By Definition 4, Theorems 3 and 6 and Corollary 8, (286) follows. We defined the set \(P^\infty\) using the detectability and stabilizability conditions of Corollary 8 and Theorem 6 to ensure the convergence of solutions \(\{(K_t, \Sigma_t) : t = 1, 2, \ldots, n\}\) of the generalized matrix DREs to unique non-negative, stabilizing solutions of the corresponding generalized matrix AREs. Then, for any element \((\Lambda^\infty, K_Z^\infty) \in P^\infty\), both summands in (286) converge. This establishes the characterization of the right-hand side of (287). Parts (1)–(3) follow from the asymptotic properties of the Kalman filter (due to the stabilizability and detectability conditions). The last statement follows due to the relaxation, Theorem 5(4).
Then, Theorem 8. Feedback capacity $C_{\nu}^f$ of Theorem 3 for asymptotically time-invariant strategies Consider the problem statement of Theorem 7 with asymptotically time-invariant strategies, $\lim_{n \to \infty} (\Lambda_n, K_{Z_n}) = (\Lambda^\infty, K^\infty_Z)$ and corresponding $K_n, n = 1, 2, \ldots$. Then,

$$C_{\nu}^f(\kappa) \triangleq \sup_{\lim_{n \to \infty} (\Lambda_n, K_n) = (\Lambda^\infty, K^\infty_Z) \in \mathcal{P}^\infty: \lim_{n \to \infty} \frac{1}{\kappa} \sum_{i=1}^{n} \left( A_i + C \right) K_i \left( A_i + C \right)^T + C \Sigma_i C^T + NK_W N^T + K_{Z_i}} \frac{1}{2n} \sum_{i=1}^{n} \log \left( \frac{\left( A_i + C \right) K_i \left( A_i + C \right)^T + C \Sigma_i C^T + NK_W N^T + K_{Z_i}}{C \Sigma_i C^T + NK_W N^T} \right)$$ (287)

and the statements of Theorem 7 hold.

Proof. This follows from Proposition 4 and Theorem 7.

Remark 19. Explicit closed-form expressions of $C_{\nu}^{f, o}(\kappa)$ are given in [1, 2] for the stable and unstable AR and ARMA noise processes $V^n$. The expressions $C_{\nu}^{f, o}(\kappa)$ consist of multiple regimes that depend on the parameters of noise, i.e., $(c, a, K_W)$ for the ARMA noise and value of $\kappa$. Moreover, for some regimes, $C_{\nu}^{f, o}(\kappa)$ is achieved by an optimal $K^\infty_Z > 0$, while for other regimes, it is achieved by $\lim_{n \to \infty} K_{Z_n} = K^\infty_Z = 0$, such that $K_{Z_1} > 0, K_{Z_i} = 0, i = 2, 3, \ldots, n$.

Remark 20. Explicit closed-form expressions of $C_{\nu}^{f, o}(\kappa)$ are given in [1, 2] for the stable and unstable AR and ARMA noise processes $V^n$. The expressions $C_{\nu}^{f, o}(\kappa)$ consist of multiple regimes that depend on the parameters of noise, i.e., $(c, a, K_W)$ for the ARMA noise and value of $\kappa$. Moreover, for some regimes, $C_{\nu}^{f, o}(\kappa)$ is achieved by an optimal $K^\infty_Z > 0$, while for other regimes, it is achieved by $\lim_{n \to \infty} K_{Z_n} = K^\infty_Z = 0$, such that $K_{Z_1} > 0, K_{Z_i} = 0, i = 2, 3, \ldots, n$.

Next, we give the expression of feedback capacity $C_{\nu}^f(\kappa)$, which is generally an upper bound on the expressions of Theorem 7 and Corollary 9.

Theorem 8. Feedback capacity $C_{\nu}^f(\kappa)$ of Theorem 3 for asymptotically time-invariant noise and strategies Consider $C_{\nu}^f(\kappa)$ of Definition 4 corresponding to Theorem 3, where $(\Lambda_t, K_{Z_t}), t = 1, \ldots, n$ and the coefficients of the PO-SS realization of $V^n$ of Definition 2 are asymptotically time-invariant, i.e.,

$$\begin{align*}
(a) \quad & \lim_{n \to \infty} (\Lambda_n, K_{Z_n}) = (\Lambda^\infty, K^\infty_Z); \\
(b) \quad & \lim_{n \to \infty} (A_n, B_n, C_n, N_n, K_{W_n}) = (A, B, C, N, K_W).
\end{align*}$$ (291)

Let $\mathcal{P}^{\infty, u\infty}$ correspond to $\mathcal{P}^{\infty} = (285)$ with $\{ A^*, GB^{1/2} \}$ and $\{ A^*(\Lambda^\infty, K^\infty_Z), GB^{1/2}(K^\infty_Z) \}$ being replaced by unit circle controllability. Let $\mathcal{P}^{\infty, u\infty}(\kappa)$ correspond to $\mathcal{P}^{\infty}(\kappa) = (288)$, with $K^\infty$ and $\Sigma^\infty$ being the maximal and stabilizing solution of (233), respectively. Then,
\[
C^{fb}(\kappa) \triangleq \lim_{n \to \infty} \sup_{\{\Lambda_1, K_2\}} \left\{ \frac{1}{2n} \sum_{i=1}^{2^n} \log \left( \frac{\left( \Lambda_1 + C_i \right) \Lambda_i^T + C_i \Sigma_i C_i^T + N_i K_W N_i^T + K_Z}{\Sigma_i^\infty C_i^T + N_i K_W N_i^T} \right) \right\}
\]

Proof. First, we note that Theorem 7 continuous to hold, if we consider asymptotically time-invariant strategies and coefficients, i.e., (a) and (b) (by Proposition 4), and the stabilizability conditions are replaced by unit circle controllability conditions. Hence, (288) remains valid with the set \( P^\infty(\kappa) \) replaced by the larger set \( P^{\infty,ucc}(\kappa) \), giving the higher value (293) \( \geq (288) \). For the derivation of (292) = (293), it suffices to show we can interchange the \( \lim_{n \to \infty} \) and the supremum in (292). This can be completed by using the definition of the set \( P^{\infty,ucc}(\kappa) \) and Conditions (a) and (b). The procedure, although lengthy, is similar to the one described in [41]; hence, we omit it. \[\square\]

Conclusion 1. Degenerate versions of Theorem 7, Theorem 8 for feedback code of Definition 3, i.e., \((s, 2^n R, n), n = 1, 2, \ldots\)

The characterizations of feedback capacity \( C^{fb}(\kappa, s) \) of the AGN channel (1) driven by a noise \( V^n \) of Definition 2, for the code of Definition 3, i.e., \((s, 2^n R, n), n = 1, 2, \ldots\), are degenerate cases of Theorem 7 and Theorem 8, corresponding to \( \Sigma_1 = \Sigma_2 = 0 \). In particular, since Theorem 7 characterizes \( C^{fb}(\kappa) \) for all initial data \( \Sigma_1 \geq 0 \), then it includes \( \Sigma_1 = \Sigma_2 = 0 \). Moreover, it follows that \( C^{fb}(\kappa) = C^{fb}(\kappa, s) \), where \( C^{fb}(\kappa, s) \) is independent of the initial state \( S_1^0 = s \).

We apply Theorem 7 to obtain \( C^{fb}(\kappa) \) of the AR\((a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)\) noise.

Corollary 10. Consider the AR\((a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)\) noise of Example 2(a).

Define the set

\[
P^\infty \triangleq \left\{ (\Lambda^\infty, K_2^\infty) \in (-\infty, \infty) \times [0, \infty) : \right\}
\]

(i) \( c \in (-\infty, \infty), a \in (-1, 1), c \neq a, \)

(ii) the pair \( \{A, C(\Lambda^\infty)\} \triangleq \{c, \Lambda^\infty + c - a\} \) is detectable,

(iii) the pair \( \{A^*, (\Lambda^\infty, K_2^\infty), GB^{1/2}(K_2^\infty)\} \) is stabilizable, where

\[
\Lambda^*(\Lambda^\infty, K_2^\infty) \triangleq c - K_W (K_W + K_2^\infty)^{-1} (\Lambda^\infty + c - a),
\]

\[
GB^{1/2}(K_2^\infty) \triangleq \left(K_W - (K_W)^2 (K_W + K_2^\infty)^{-1}\right)^{1/2}
\]

Then,

\[
C^{fb}(\kappa) = \sup_{(\Lambda^\infty, K_2^\infty) \in P^\infty(\kappa)} \frac{1}{2} \log \left( \frac{\left(\Lambda^\infty + c - a\right)^2 K_W + K_2^\infty}{K_W} \right) = C^{fb}(\kappa, s), \forall s
\]
where
\[ \mathcal{P}^\infty(\kappa) \triangleq \left\{ (\Lambda^\infty, K_2^\infty) \in \mathcal{P}^\infty : (\Lambda^\infty)^2 K_2^\infty + K_2^\infty \leq \kappa, \right. \\
K_2^\infty \geq 0 \quad \text{is the unique and stabilizing solution of} \\
K_2^\infty = (c)^2 K_2^\infty + K_W \\
- \left( cK_2^\infty (\Lambda^\infty + c - a) + K_W \right)^2 \left( K_W + K_2^\infty + (\Lambda^\infty + c - a)^2 K_2^\infty \right)^{-1} \right\} \] (297)

provided that there exists \( \kappa \in [0, \infty) \) such that the set \( \mathcal{P}^\infty(\kappa) \) is non-empty.
Moreover, the maximum element \( (\Lambda^\infty, K_2^\infty) \in \mathcal{P}^\infty(\kappa) \), is such that,
1. It induces asymptotic stationarity of the corresponding, input and innovations processes;
2. If \( V^n, n = 1, 2, \ldots \) is asymptotically stationary, then it induces asymptotic stationarity of the corresponding, input and output processes;
3. For (i) and (ii), \( C^{fb, o}(\kappa) \) and \( C^{fb, s}(\kappa, s) \) are independent of \( \Sigma_1 \geq 0 \) and \( s \), respectively, and the following identities hold.
\[ C^{fb, o}(\kappa) = C^{fb, o}(\kappa, s) = C^{fb, s}(\kappa, s), \quad \forall s \] (298)

Furthermore, if the set \( \mathcal{P}^\infty(\kappa) \) is empty, replace stabilizability of \( \{ A^o, GB^{-1} \} \) by unit circle controllability, i.e., so that \( K_2^\infty \) is the maximal and stabilizing solutions of the ARE.

**Proof.** The first part is an application of Theorem 7, Lemmas 4 and 6. Parts (1)–(3) are due to the convergence properties of the Kalman filter (due to the stabilizability and detectability conditions). It remains to show (298). The equality \( C^{fb, o}(\kappa) = C^{fb, o}(\kappa, s), \forall s \) holds by Conclusion 1(a). The last equality holds due to the AR\( (a, c), a \in (-\infty, \infty), c \in (-\infty, \infty) \) noise. If the initial state \( S_1 = S_1^o = s \) is known to the encoder and the decoder, then Condition 1 of Section 1.1 holds. In addition, Condition 2 also holds, as can be easily verified from Equations (121) and (122). \( \square \)

**Remark 21.** From Corollary 10, we obtain the degenerate cases, AR\( (c), c \in (-\infty, \infty) \), noise, i.e., setting \( a = 0 \). The various implications of the detectability and stabilizability conditions for the AR\( (c), c \in (-\infty, \infty) \) noise are found in [1]. The corresponding \( C^{fb, o}(\kappa, s) \) states that for stable AR\( (c) \) and time-invariant strategies, feedback does not increase capacity (because of the stronger condition of stabilizability). However, if unit circle controllability is imposed instead, then feedback increases capacity.

4. Sequential Characterization of \( n \)-FTFI Capacity for Case (II) Formulation

In this section, we consider the Case (II) formulation, and we derive the characterization of feedback capacity, \( C^{fb, s}(\kappa, s) \), of the AGN channel (1) driven by a noise \( V^n \) of Definition 2, i.e., for the code of Definition 3, \( (s, 2^n R, n), n = 1, 2, \ldots, \) when Conditions 1 and 2 of Section 1.1 hold.

**Definition 8.** AGN channels driven by noise with invertible PO-SS realizations
The PO-SS realization of the noise of Definition 2 is called invertible if it satisfies the following condition:
(A1) Given the initial state \( S_1 = S_1^o = s \), the noise \( V^{t-1} \) uniquely specifies the state \( S_t^o \), for \( t = 1, \ldots, n \), and vice versa.

**Corollary 11.** Characterization of \( n \)-FTFI Capacity for the Case (II) formulation
Consider the AGN channel (1) driven by a noise \( V^n \) of Definition 8, and the code of Definition 3,
(s, 2^{nR}, n), n = 1, 2, \ldots, \text{i.e., Conditions 1 and 2 of Section 1.1 hold.}

Define the n–FTFI capacity for a fixed initial state \( S_1 = S_1^* = s \) by

\[
C^{f_{\text{BS}}} (n, s) = \sup_{P_{[1,n]}(\kappa)} H^P(Y^n | s) - H(V^n | s)
\tag{299}
\]

where the set \( P_{[1,n]}(\kappa) \) is defined by

\[
P_{[1,n]}(\kappa) \triangleq \left\{ P_1(dx_1 | x^{l-1}, y^{l-1}, s), t = 1, \ldots, n : \frac{1}{n} \mathbf{E}_s^P \left( \sum_{i=1}^{n} (X_i)^2 \right) \leq \kappa \right\}
\tag{300}
\]

and where \( \mathbf{E}_s^P \) means \( S_1 = S_1^* = s \) is fixed, and the joint distribution depends on the elements of \( P_{[1,n]}(\kappa) \).

Then, the following hold:

(a) The n–FTFI capacity, for a fixed \( S_1 = s \), is characterized by

\[
C^{f_{\text{BS}}} (\kappa, s) = \sup_{P_{[1,n]}(\kappa)} \sum_{t=1}^{n} \left\{ H^P(Y_t | Y^{t-1}, s) - H(V_t | Y^{t-1}, s) \right\}
\tag{301}
\]

\[
= \sup_{P_{[1,n]}(\kappa)} \sum_{t=1}^{n} \left\{ H^P(Y_t - \mathbf{E}^P \{ Y_t | Y^{t-1}, s \}) - H(V_t - \mathbf{E} \{ V_t | Y^{t-1}, s \} | s) \right\}
\tag{302}
\]

where the \( P_{[1,n]}(\kappa) \) is defined by

\[
P_{[1,n]}(\kappa) \triangleq \left\{ P_1(dx_1 | s_t, y^{l-1}, s), t = 1, \ldots, n : \frac{1}{n+1} \mathbf{E}_s^P \left( \sum_{i=1}^{n} (X_i)^2 \right) \leq \kappa \right\}
\tag{303}
\]

and where (18) is respected, \( P_1(dx_1 | s_t, y^{l-1}, s) \), is conditionally Gaussian, with linear conditional mean and nonrandom conditional covariance, given by (the notation \( S_1 = S_1^t, t = 2, \ldots, n \) means this sequence is generated from (19), when the initial state is fixed, \( S_1 = S_1^* = s \)).

\[
\mathbf{E}^P \{ X_t | S_1^t, Y^{t-1}, S_1^t = s \} = \begin{cases} A_t \left( S_1^t - \mathbf{E}^P \{ S_1^t | Y^{t-1}, S_1^t = s \} \right), & t = 2, \ldots, n \\ 0, & t = 1, \end{cases}
\tag{304}
\]

\[
K_{X_t|S_t^t, Y^{t-1}, S_1^t = s} \triangleq \text{cov}(X_t, X_t | S_t^t, Y^{t-1}, S_1^t = s) = K_{Z_t} \geq 0, \quad t = 1, \ldots, n.
\tag{305}
\]

and \( H^P(Y_t | Y^{t-1}, s) \) is evaluated with respect to the probability distribution \( P_1^M(dy_t | y^{t-1}, s) \), defined by

\[
P_1^M(dy_t | y^{t-1}, s) = \int P_1(dy_t | x_t, s_t) \mathbf{P}_1^M(dx_t | s_t, y^{t-1}, s) \mathbf{P}_1^M(ds_t | y^{t-1}, s), \quad t = 1, \ldots, n.
\tag{306}
\]

(b) Define the conditional means and conditional covariance for a fixed \( S_1 = S_1^* = s \) by

\[
K_t^s \triangleq \text{cov}(S_t^s, S_t^s | Y^{t-1}, S_1^s = s) = \mathbf{E}_s^P \left\{ \left( S_t^s - S_t^s \right) \left( S_t^s - S_t^s \right)^T \right\},
\tag{307}
\]

\[
S_t^s \triangleq \mathbf{E}_s^P \{ S_t^s | Y^{t-1}, S_1^s = s \}, \quad t = 2, \ldots, n, \quad K_t^s \triangleq \text{cov}(S_t^s, S_t^s | S_1^s = s) = 0, \quad S_1^s \triangleq s.
\tag{308}
\]

The optimal channel input distribution of part (a) is induced by a jointly Gaussian process \( X^n \), with a realization given by
\[ X_t = \Lambda_t \left( \hat{S}_t^s - \tilde{S}_t^s \right) + Z_t, \quad X_1 = Z_1, \quad t = 2, \ldots, n, \]  
\[ Z_t \in \mathcal{N}(0, K_{Z_t}) \text{ independent of } (S_t, X_t^{t-1}, V_t^{t-1}, Y_t^{t-1}), \quad t = 1, \ldots, n, \]
\[ Y_t = \Lambda_t \left( S_t^s - \tilde{S}_t^s \right) + Z_t + V_t, \quad t = 1, \ldots, n, \]
\[ = \Lambda_t \left( S_t^s - \tilde{S}_t^s \right) + C_t S_t^s + N_t W_t + Z_t, \]
\[ \frac{1}{n} E_s^m \left\{ \sum_{t=1}^{n} (X_t)^2 \right\} = \frac{1}{n} \sum_{t=1}^{n} \left( \Lambda_t K_t^s \Lambda_t^T + K_{Z_t} \right) \leq \kappa \]

where \( \Lambda_t \) is nonrandom.

The conditional means and conditional covariance \( \tilde{S}_t^s \) and \( K_t^s \) are given by the generalized Kalman filter, as follows:

(i) \( \tilde{S}_t^s \) satisfies the Kalman filter recursion
\[ \tilde{S}_{t+1}^s = A_t \tilde{S}_t^s + F_t(K_t^s) I_t^s, \quad \tilde{S}_1 = s, \]
\[ F_t(K_t^s) \triangleq \left( A_t K_t^s (\Lambda_t + C_t)^T + B_t K_{W_t} N_t^T \right) \cdot \left\{ N_t K_{W_t} N_t^T + K_{Z_t} + (\Lambda_t + C_t) K_t^s (\Lambda_t + C_t)^T \right\}^{-1}, \]
\[ I_t^s \triangleq Y_t - C_t \tilde{S}_t^s = (\Lambda_t + C_t) (S_t^s - \tilde{S}_t^s) + N_t W_t + Z_t, \quad t = 1, \ldots, n, \]
\[ I_t^s \in \mathcal{N}(0, K_{I_t^s}), \quad t = 1, \ldots, n \text{ an orthogonal innovations process, i.e.,} \]
\[ I_t^s \text{ is independent of } I_k^s \text{ for all } t \neq k, \text{ and } I_t^s \text{ is independent of } Y_t^{t-1}, \]
\[ K_{Y_t|Y^{t-1},s} = K_{I_t^s} \triangleq \text{cov}(I_t, I_t^s) = \begin{pmatrix} s \\ \Lambda_t + C_t \end{pmatrix}^T K_t^s (\Lambda_t + C_t)^T + N_t K_{W_t} N_t^T + K_{Z_t}, \]

(ii) The error \( E_t^s \triangleq S_t^s - \tilde{S}_t^s \) satisfies the recursion
\[ E_{t+1}^s = F_t^{CL}(K_t^s) E_t^s + \left( \begin{pmatrix} B_t - F_t(K_t^s) N_t \end{pmatrix} W_t - F_t(K_t^s) Z_t \right), \quad E_1^s = S_1^s - \tilde{S}_1^s = 0, \quad t = 1, \ldots, n, \]
\[ F_t^{CL}(K_t^s) \triangleq A_t - F_t(K_t^s) \left( \Lambda_t + C_t \right). \]

(iii) \( K_t^s \) satisfies the generalized DRE
\[ K_{t+1}^s = A_t K_t A_t^T + B_t K_{W_t} B_t^T - \left( B_t K_{W_t} N_t^T + A_t K_t^s (\Lambda_t + C_t)^T \right) \cdot \left\{ N_t K_{W_t} N_t^T + K_{Z_t} \right\}^{-1} \left( B_t K_{W_t} N_t^T + A_t K_t^s (\Lambda_t + C_t)^T \right)^T, \quad K_1^s \leq 0, \quad K_0^s = 0, \quad t = 1, \ldots, n. \]

(c) The characterization of the \( n \)-FFI capacity of part (a) is
Corollary 12. Characterization of Feedback Capacity for the Case (II) formulation

Consider the statement of Corollary 11 for asymptotically time-invariant noise. The feedback capacity is given by

\[
C_n^{fb,S}(\kappa, s) = \lim_{n \to \infty} \frac{1}{n} C_n^{fb,S}(\kappa, s)
\]

\[
= \sup_{(\Lambda_1, K_1^T) \in \mathcal{P}^{\infty,acc,S}(\kappa)} \frac{1}{2} \log \left( \frac{(\Lambda^\infty + C) K^\infty (\Lambda^\infty + C)^T + N K W N^T + K_2^\infty}{N K W N^T} \right),
\]

(325)

(326)

(327)

(328)

If the optimal solution is such that the stabilizability holds, then the feedback capacity is independent of the initial state \(S_1 = s\).

Proof. This is a special case of Theorem 7, with \(\Sigma_1 = 0\).

Corollary 13. Feedback capacity of [4,7]

Consider the channels studied in [4,7], i.e., with time-invariant and stable realization.

(a) The time-domain \(n\)-FTFI capacity is \(C_n^{fb,S}(\kappa, s)\) given in Corollary 11.

(b) The time-domain feedback capacity is given by Corollary 12.

Proof. Since the noise in [4,7] is time-invariant and stable, according to Definition 8, and the code is that of Definition 3, i.e., \((s, 2^n R, n), n = 1, 2, \ldots\), then their results are special cases of Corollaries 11 and 12.

In the next remark, we clarify the relation of Corollary 11 and the analysis of [4,8].
Remark 23. Relations of Corollary 11 and [4,8]

(a) The state space problem analyzed in [8] is precisely $C_{n}^{fb,S}\left(\kappa,s\right)$, when the noise is stationary and Gaussian, i.e., it corresponds to the Case (II) formulation. Corollary 11 is derived in [8] for the degenerate case of a time-invariant realization of the noise $V^{n}$, i.e., of Definition 8. However, the asymptotic analysis of [8] (Section VI) should be read with caution, because it did not impose the necessary and/or sufficient conditions for convergence of the sequence $K_{n}^{\infty}$, $t = 1, 2, \ldots$ generated by the time-invariant version of the generalized DRE (321), i.e., $\lim_{n \to \infty} K_{n}^{\infty} = K^{\infty} \succeq 0$, where $K^{\infty} \succeq 0$ is either the maximal or the unique and stabilizing solution of a corresponding generalized ARE.

(b) The problem analyzed [4] that led to [4] (Theorem 6.1, $C_{FB}$) is the per unit time limit of $C_{n}^{fb,S}\left(\kappa,s\right)$, when the noise is stationary, two-sided or one-sided (asymptotically stationary) and Gaussian, i.e., it corresponds to the Case (II) formulation. The characterization of feedback capacity presented in [4] (Theorem 6.1, $C_{FB}$) presupposed that the following hold ((i)–(iii) are also assumed in [8] (Section VI)):

(i) The feedback code is Definition 3, i.e., $(s, z^{nit}, n)$.

(ii) The noise is time-invariant and stable, and the PO-SS realization of the noise is invertible, as presented in Definition 8.

(iii) The definition of rate is $C^{fb,S,\omega}\left(\kappa,s\right)$, with supremum and per unit time limit interchanged, and the supremum taken over using time-invariant channel input distributions.

(iv) The innovations covariance of the channel input process is asymptotically zero, i.e., $\lim_{n \to \infty} K_{z_{n}}^{\omega} = K_{z}^{\omega} = 0$. This implies the corresponding $\lim_{n \to \infty} K_{n}^{\omega} = K^{\omega} \succeq 0$ is the maximal and stabilizing solution of the corresponding matrix ARE, since detectability and unit circle controllability conditions hold, but not the stabilizability condition.

Items (i)–(iv) are confirmed from [4] (Lemma 6.1) (and comments above), which is used to derive [4] (Theorem 6.1, $C_{FB}$).

However, the characterization of feedback capacity in [4] (Theorem 6.1, $C_{FB}$) should be read with caution, because the stabilizability condition is violated, due to the requirement of the author, which states that $K_{z}^{\omega} = 0$ is optimal. By Theorem 5(1), for the choice $K_{z}^{\omega} = 0$, the only choice is the maximal and stabilizing solution of the generalized ARE presented in [4] (Theorem 6.1, $C_{FB}$).

However, it is easy to verify that [4, Theorem 6.1, $C_{FB}$] cannot be the capacity of asymptotically stationary noise because $C_{FB}$ depends on the covariance $K_{z} > 0$. Moreover, $K_{z} > 0$ is required. Finally, we emphasize that the treatment of the ARMA noise in [2] clarifies the above issues.

5. Conclusions

New equivalent sequential characterizations of Cover and Pombra [5] “n–block” feedback capacity formulas are derived using time-domain methods for additive Gaussian noise (AGN) channels driven by non-stationary Gaussian noise. New features of the equivalent characterizations encompass the representation of the optimal channel input process by a sufficient statistic and Gaussian orthogonal innovations process. The sequential characterizations of the n–block feedback capacity formula are expressed as a functional of two generalized matrix difference Riccati equations (DREs) of the filtering theory of Gaussian systems. The asymptotic analysis of the per unit time limit of the n–block”, called feedback capacity, is also presented for time-invariant and asymptotically time-invariant channel input distributions, using the tools from the theory of generalized matrix Riccati equations.

Prior analysis and characterizations of feedback capacity follows on from the analysis and derivation of the new sequential characterizations of feedback capacity, such as [4,7,11,12], who do not address the Cover and Pombra [5] feedback capacity problem, as the code definitions and noise assumptions in [4,7,11,12] (even under the restriction of stationary noise) are fundamentally different from those in [5]. This paper clarifies several of these points of confusion.
Author Contributions: C.D.C., C.K. and S.L. contributed to the conceptualization, methodology, and writing of this manuscript. All authors have read and agreed to the published version of this manuscript.

Funding: The work of C.D. Charalambous was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: EXCELLENCE/1216/0296).

Institutional Review Board Statement: Not applicable

Data Availability Statement: No data are contained within this article.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of this manuscript; or in the decision to publish the results.

Appendix A

Appendix A.1. Proof of Theorem 1

(a) Consider an element of \( \mathbb{E}_{[0,1]}(\kappa) \). Then, the conditional entropies \( H^T(Y_t|Y^{t-1}) \), \( t = 1, \ldots, n \) are defined, provided that the conditional distributions of \( Y_t \) conditioned on \( Y^{t-1} \), i.e., \( P_t(dy_t|y^{t-1}) \), for \( t = 1, \ldots, n \), are determined. By the reconditioning, using (60), we have

\[
P_t(dy_t|y^{t-1}) = \int P_t(dy_t|y^{t-1}, w) P_t(dw|y^{t-1}), \quad t = 0, \ldots, n \tag{A1}
\]

Thus, (64) is shown. Similarly, consider an element of \( \mathbb{P}_{[0,1]}(\kappa) \). Then, the conditional entropies \( H^T(Y_t|Y^{t-1}) \), \( t = 1, \ldots, n \) are defined, provided that the conditional distributions of \( Y_t \) conditioned on \( Y^{t-1} \), i.e., \( P_t(dy_t|y^{t-1}) \) for \( t = 1, \ldots, n \), are determined. By (52) and (53), (65) is obtained. Since \( \mathbb{E}_{[0,1]}(\kappa) \subseteq \mathbb{P}_{[0,1]}(\kappa) \), Inequality (63) follows.

(b) This part is followed by the maximum entropy principle of Gaussian distributions. That is, under the restriction (18), a conditional Gaussian element of \( \{P_t(dx_t|x^{t-1}, y^{t-1}) \}, \ t = 1, \ldots, n \} \in \mathbb{P}_{[0,1]}(\kappa) \) with linear conditional mean and nonrandom conditional covariance induces a jointly Gaussian distribution of the process \( (X^n, Y^n) \), such that the marginal distribution of \( Y^n \) is jointly Gaussian. Below, we provide alternative proof that uses the Cover and Pombra characterization of the \( n \)-FTFI capacity, given by (34) and (35). Consider (35) and define the process

\[
Z_1 \triangleq Z_1 - E\{Z_1\}, \tag{A3}
\]

\[
Z_t \triangleq Z_t - E^T\{Z_t|X^{t-1}, V^{t-1}, Y^{t-1}\}, \quad t = 2, \ldots, n, \tag{A4}
\]

\[
=Z_t - E^T\{Z_t|V^{t-1}, Y^{t-1}\}, \quad \text{since } X^{t-1} \text{ is uniquely defined by } (V^{t-1}, Y^{t-1}). \tag{A5}
\]

Then, \( Z_t \) is a Gaussian orthogonal innovations process, independent of \( (X^{t-1}, V^{t-1}, Y^{t-1}) \), for \( t = 2, \ldots, n \), and \( E\{Z_t\} = 0 \), for \( t = 1, \ldots, n \). By (35), we re-write \( X_t, t = 1, \ldots, n \) as,
\[ X_t = \sum_{j=1}^{t-1} B_{tj} V_j + Z_{it}, \quad t = 1, \ldots, n, \]  
\[ = \sum_{j=1}^{t-1} B_{tj} V_j + \mathbb{E} \{ Z_t | V^{t-1}, Y^{t-1} \} + Z_{it}, \quad \text{by (A5)} \]  
\[ = \sum_{j=1}^{t-1} B_{tj} V_j + \mathbb{E} \{ Z_t | V^{t-1}, Y^{t-1} \} + Z_{it}, \quad \text{for some } \Gamma_t \text{ nonrandom} \]  
\[ \Rightarrow \quad \sum_{j=1}^{t-1} B_{tj} V_j + \sum_{j=1}^{t-1} \Gamma^1_{tj} Y_j + Z_{it}, \quad \text{for some } (\Gamma^1_t, \Gamma^2_t) \]  
\[ = \Gamma^1_t V^{t-1} + \Gamma^2_t Y^{t-1} + Z_{it}, \quad \text{by definition} \]  

where (a) is due to the by joint Gaussianity of \((Z^n, X^n, Y^n)\). From (A10) and the independence of \(Z_t\) and \((X^{t-1}, V^{t-1}, Y^{t-1})\), for \(t = 2, \ldots, n\), (66) then follows, as well as (67).

(c) The statements follow directly from the representation of part (b), while the independence of \(Z^n\) and \(V^n\) is due to the code definition, i.e., Definition 1(iv).

(d) The statement follows on from parts (a–c) and (28)–(31).

**Appendix A.2. Proof of Proposition 1**

(a) The covariances of the realization of the ARMA\((a,c)\) noise of Example 2(b) satisfy the recursions

\[ K_{S_{t+1}} = c^2 K_{S_t} + K_W, \quad K_{S_t V_t} = (c - a) K_{S_t}, \quad K_{V_t} = (c - a)^2 K_{S_t} + K_W, \quad \forall t \in \mathbb{Z}. \]  

If the recursion \(K_{S_{t+1}} = c^2 K_{S_t} + K_W\) is initiated at the stationary value \(K_{S_1} = d_{11} = \frac{K_W}{1-c^2}\), then \(K_{S_{t+1}} = d_{11}, \forall t = 2, 3, \ldots\); hence, \(S_t, \forall t \in \mathbb{Z}\) is stationary, which then implies the stationarity of \(V_t, \forall t \in \mathbb{Z}\). Hence, if (133) holds, then \((V_t, S_t), \forall t \in \mathbb{Z}\) is stationary. Via simple calculations, (134) then follows. We carry out the same operation for the one-sided ARMA\((a,c)\). (b) By the above covariances, for all \(K_{S_1} \geq 0\), we have \(\lim_{n \to \infty} K_{S_n} = K_S^\infty\), where \(K_S^\infty = c^2 K_S^\infty + K_W\), which then implies \(K_S^\infty = d_{11}\). Similarly, \(\lim_{n \to \infty} K_{S_n V_n} = K_S^\infty, K_{S_n Y_n} = d_{12} \lim_{n \to \infty} K_{V_n} = K_Y^\infty = d_{22}\). (c) (135) and (136) follow on from Lemma 1 by replacing the conditioning information \(V^{t-1}\) with \((V_0, V^{t-1})\) in (85). By mean square estimation, the initial data are

\[ \delta_1 = \mathbb{E} \left\{ S_1 | V_0 \right\} = \mathbb{E} \left\{ cS_0 + W_0 | V_0 \right\} \]  
\[ = \mathbb{E} \left\{ cS_0 + W_0 \right\} + \text{cov}(cS_0 + W_0, V_0) \left\{ \text{cov}(V_0, V_0) \right\}^{-1} (V_0 - \mathbb{E} \left\{ V_0 \right\}) \]  
\[ = \left( cd_{12} + K_W \right) d_{22}^{-1} V_0, \]  
\[ \Sigma_1 = \text{cov}(S_1, S_1 | V_0) = \text{cov}(S_1, S_1) - \left\{ \text{cov}(S_1, S_1) \right\}^2 \left\{ \text{cov}(V_0, V_0) \right\}^{-1} \]  
\[ = d_{11} - d_{11}^2 d_{22}^{-1}. \]

The last part is obvious.

**Appendix A.3. Proof of Corollary 3**

(a) Since we have assumed that \(S_1 = S_1^\infty = s\) is fixed, and that it is known to the encoder and the decoder, then Theorem 1 still holds. This is due to replacing all conditional distributions, expectations and entropies via corresponding expressions with a fixed \(S_1 =\)
$S_1^t = s$. Hence, (75) is replaced by (141), and (68) is replaced by (142) (since the code is allowed to depend on $S_1 = S_1^t = s$). (b) From the PO-SS realization of Definition 2 with $S_1 = S_1^t = s$ fixed, it follows that a necessary condition for Conditions 1 of Section 1.1 to hold is (i). The expression of entropy (143) is easily obtained by invoking Condition (i) and the properties of conditional entropy. That is, $H(V_1|S_1^t = s) = H(C_1S_1^t + N_1W_1|S_1^t = s) = H(N_1W_1|S_1 = s) = H(N_1W_1)$ by independence of $W_1$ and $S_1^t$, and $H(V_2|V_1, s) = H(C_2S_2^t + N_2W_2|C_1S_1^t + N_1W_1, S_1^t = s) = H(C_2S_2^t + N_2W_2|N_1W_1, S_1^t = s) = H(C_2A_1S_1^t + C_2B_1W_1 + N_2W_2|N_1W_1, S_1^t = s) = H(N_2W_2|N_1W_1, S_1^t = s) = H(N_2W_2)$, etc. The last statement is obvious. This completes the proof.

Appendix A.4. Proof of Theorem 3

(a) Clearly, (154)–(165) follow directly from Theorem 1 and the preliminary calculations prior to the statement of the theorem. However, (154)–(165) can also be shown independently of Theorem 1 by invoking the maximum entropy property of Gaussian distributions, as follows: By Lemma 1, we have $H(V^n) = \sum_{t=1}^{n} H(\hat{I}_t)$. By the maximum entropy principle, $H(Y^n)$ is maximized if $P_{Y^n}$ is jointly Gaussian, the average power constraint holds, and (18) is respected. By (149), (159), and (160), if (159)–(165) hold, then $(X^n, Y^n)$ is jointly Gaussian; hence, $H(Y^n)$ is maximized. This shows (a).

(b) Step 1. By (163) and (164), an alternative representation of $X^n$ to the one given in Theorem 1 and induced by (68) is

$$X_t = \Gamma_1^t \hat{S}_t + \Gamma_2^t Y^{t-1} + Z_t, \quad t = 1, \ldots, n$$

(A14)

$$Z_t \text{ satisfies (167).}$$

(A15)

for some nonrandom $(\Gamma_1^t, \Gamma_2^t)$. Upon substituting (A14) into the channel output $Y^n$, we have

$$Y_t = \Gamma_1^t \hat{S}_t + \Gamma_2^t Y^{t-1} + Z_t + V_t, \quad t = 1, \ldots, n$$

(A16)

$$\text{by (167)}$$

$$\text{by (167).}$$

(A17)

The right-hand side of (A17) is driven by two independent processes, $Z_t, t = 1, \ldots, n$ and $\hat{I}_t, t = 1, \ldots, n$, which are also mutually independent. Further, the right-hand side of (A17) is a linear function of a state process $\hat{S}_t, t = 1, \ldots, n$, which satisfies the following recursion (88):

$$\hat{S}_{t+1} = A_t \hat{S}_t + M_t (\Sigma_t) \hat{I}_t, \quad \hat{S}_1 = \mu_{S_1}$$

(A18)

Note that the right-hand side of (A18) is driven by the orthogonal process $\hat{I}_t$, which is independent of $Y^{t-1}$ and hence of $\hat{S}_t$. It is also independent of $Y^{t-1}$. By (167), $Z_t$ is independent of $Y^{t-1}$ and of $\hat{S}_t$. By (A17) and (A18), it follows that $\hat{S}_{t+1}, t = 1, \ldots, n$ satisfies a generalized Kalman filter recursion, similar to that of Lemma 1. Hence, the entropy $H(Y^n)$ can be computed using the innovations process of $Y^n$, as in Lemma 1. Define the orthogonal Gaussian innovations process $I^n$ of $Y^n$ by

$$I_t \triangleq Y_t - \mathbb{E}\left\{Y_t | Y^{t-1}\right\}, \quad t = 1, \ldots, n$$

(A19)

$$= (\Gamma_1^t + C_t) (S_t - \hat{S}_t) + \hat{I}_t - \mathbb{E}\left\{\hat{I}_t | Y^{t-1}\right\} + Z_t, \quad \text{by (A17).}$$

(A20)

$$= (\Gamma_1^t + C_t) (S_t - \hat{S}_t) + \hat{I}_t + Z_t, \quad \text{by } \hat{I}_t \text{ indep. of } Y^{t-1}, \mathbb{E}\left\{\hat{I}_t\right\} = 0$$

(A21)
The entropy of $Y^n$ is computed as follows.

$$H(Y^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1})$$

$$= \sum_{i=1}^{n} H(I_i|Y^{i-1}), \quad \text{by (A19) and a property of conditional entropy;}$$

$$= \sum_{i=1}^{n} H(I_i), \quad \text{by orthogonality of } I_i \text{ and } Y^{i-1}. \quad \text{(A24)}$$

By (A21), the Gaussian innovations process $I^n$ does not depend on the strategy of $\Gamma^2$. Consequently, by (A24), the entropy $H(Y^n)$ does not depend on the strategy of $\Gamma^2$.

Step 2. Let $g_t(Y^{t-1}) \triangleq \Gamma_t^2 Y^{t-1}, t = 1, \ldots, n$. By (A14) and (A15), it then follows that

$$\frac{1}{n} E\left\{ \sum_{i=1}^{n} (X_i)^2 \right\} = \frac{1}{n} E\left\{ \sum_{i=1}^{n} \left( \Gamma_t^2 S_i + g_t(Y^{t-1}) + Z_i \right)^2 \right\}$$

$$= \frac{1}{n} E\left\{ \sum_{i=1}^{n} \left( \Gamma_t^2 S_i + g_t(Y^{t-1}) \right)^2 \right\} + \sum_{i=1}^{n} K_{Z_t}, \quad \text{by indep. of } Z_i \text{ and } (V^{t-1}, S_t, Y^{t-1}). \quad \text{(A26)}$$

By utilizing the mean square estimation theory, then the choice of $g(\cdot)$ that minimizes the right hand of (A26) is

$$g_t(Y^{t-1}) = g_t^*(Y^{t-1}) = -\Gamma_t^2 E\left\{ S_i \big| Y^{t-1} \right\} = -\Gamma_t^2 \hat{S}_t, \quad t = 1, \ldots, n. \quad \text{(A27)}$$

Hence, $\Gamma_t^2 = -\Gamma_t^2, \forall t$. Let $\Lambda_t \triangleq \Gamma_t^2, \forall t$ and substitute into the (A17) recursion to obtain (170) and into the average power (A26) to obtain (171). Hence, the derivation of (166)–(171) is completed.

It then follows that (172)–(180) are generalized Kalman filter recursions of estimating the new state process $\hat{S}_t, t = 1, \ldots, n$ that satisfies recursion (A18) from the channel output process $Y_t$ that satisfies the recursion (170).

(c) By parts (a) and (b), and the entropy of Gaussian RVs, upon substituting (A24) and (98) into (154), we obtain part (c). This completes the proof.

Appendix A.5. Proof of Proposition 2

Since the proof of [4] (Theorem 6.1) is based on [4] (Lemma 6.1), where the channel input $X_t$ is expressed as $X_t = \Lambda \left( S_t - E\left\{ S_i | Y_{t-\infty}^{t-1} \right\} \right), t = 1, \ldots$, where $\Lambda$ is a nonrandom vector, then (211) is necessary for [4] (Theorem 6.1) to hold. Next, we show that Conditions 1 and 2 of Section 1.1 are necessary and sufficient for Equality (211) to hold. To avoid a complex notation, we prove the claim for the realization of Example 2(a). Suppose that the initial state $S_1$ of the noise is $S_1 = S_1 = s$ and that it is known to the encoder and the decoder; without loss of generality, take $s = 0$, which, by (127), implies $V_0 = 0, W_0 = 0$ (as is often carried out in [4]). Then, the following hold:

$$S_1 = 0 \quad \implies \quad V_1 = W_1, \quad S_2 = W_1 = V_1, \quad \text{by (128), (129);} \quad \text{(A28)}$$

$$(S_1 = 0, V_1) \text{ uniquely define } S_2 = c S_1 + W_1 = W_1 = V_1, \quad \text{by (128), (129);} \quad \text{(A29)}$$

$$V_2 = (c - a) S_2 + W_2, \quad S_3 = c S_2 + W_2 = c V_1 + W_2, \quad \text{by (128), (129);} \quad \text{(A30)}$$

$$(S_1 = 0, V_1, V_2) \text{ uniquely define } (S_2, S_3); \quad \text{(A31)}$$

Repeating, then $(S_1 = 0, V_1, \ldots, V_{t-1})$ uniquely define $(S_2, S_3, \ldots, S_t), \forall t = 3, 4, \ldots. \quad \text{(A32)}$
From (A28)–(A32), it then follows that for any \( S_1 = s \), including \( s = 0 \), it is known by the encoder that the following equalities hold:
\[
P_{X_t|X^{t-1},Y^{t-1},s_1} = P_{X_t|Y^{t-1},s_1}, \quad t = 1, \ldots, (A33)
\]
\[
P_{X_t|S_t,Y^{t-1},s_1}, \quad t = 1, \ldots, (A34)
\]
We can go one step further to identify the information structure of optimal channel input distributions using (A34), i.e., to show \( P_{X_t|S_t,Y^{t-1},s_3} = P_{X_t|S_t,Y^{t-1},s_1} \), \( t = 1, \ldots \) by repeating to proof of [8] (Theorem 1). However, for the statement of the proposition, this is not necessary.

Suppose that either \( S_1 = s \) is not known to the encoder, i.e., \( V_0 = v_0, W_0 = w_0 \) are not known to the encoder, and \( S_1 \neq 0 \), while the optimal channel input is expressed as a function of the state of the noise, \( S^n \):
\[
P_{X_t|X^{t-1},Y^{t-1},s_1} = P_{X_t|Y^{t-1},s_1} = P_{X_t|S_t,Y^{t-1},s_1}, \quad t = 1, \ldots, (A35)
\]
Then, by (128) and (129), it follows that \( V_1 = (c-a)S_1 + W_1, S_2 = aS_1 + W_1 \). Hence, knowledge of \( V_1 \) does not specify \( S_2 \). Similarly, \( V^{t-1} \) does not specify \( S^t \) for \( t = 2, 3, \ldots \). Hence, we arrive at a contradiction of the last equality in (A35). This completes the proof.

Appendix A.6. Proof of Lemma 2
(a) This is due to Lemma 1(v).
(b) By taking the per unit time limit (221) and utilizing the hypothesis (223), the continuity of the log(\( \cdot \)) and the fact that, for any convergent sequence \( a_n, n = 1, 2, \ldots \), \( \lim_{n \to \infty} a_n = a \), then \( \frac{1}{n} \sum_{i=1}^n a_i \to a \), as \( n \to \infty \), (224) follows.

Appendix A.7. Proof of Lemma 4
From Corollary 6(a), we deduce that \( S^n_0 \triangleq S_t, t = 1, \ldots, n \) satisfies (239) with initial condition (240). By Definition 7, the corresponding generalized algebraic Riccati equation is (241), and pairs \( \{A, C\} \) and \( \{A^*, GB^* \} \) are given by (242).

1. By Definition 7, for \( \epsilon \neq a \), the pair \( \{A, C\} = \{c, c-a\} \) is observable and hence detectable.
2. By Definition 7, the pair \( \{A^*, GB^* \} = \{a, 0\} \) is unit circle controllable if and only if \( |a| \neq 1 \).
3. By Definition 7, the pair \( \{A^*, GB^* \} = \{a, 0\} \) is stabilizable if and only if \( a \in (-1, 1) \).
4. This follows from Theorem 5(1) and parts (1)–(4). Since (241) is a quadratic equation, we can verify that the two solutions are \( \Sigma_0 = 0 \) and \( \Sigma_\infty = \frac{k_w(s^2-1)}{(c-a)} \), and we consider the statement of (243).
5. For values \( c \in (-\infty, 0) \) and \( |a| < 1 \), the pair \( \{A, C\} = \{c, c-a\} \) is detectable and the pair \( \{A^*, GB^* \} = \{a, 0\} \) is stabilizable, and the statement follows from Theorem 5(3).
6. (245) follows from Lemma 2(b) by invoking Corollary 6(a), i.e., \( K_t = (c-a)^2 \Sigma_t^n + k_w, t = 1, \ldots, n \), where \( \Sigma_t^n \) is generated by (239) and parts (4) and (5).

Appendix A.8. Proof of Corollary 11
First, note that the analog of Theorem 1(a), for the code \( (s, 2n^R, n), n = 1, 2, \ldots \) is (299) and (300) because \( P_t(dx_t|x^{t-1}, y^{t-1}, s) = P_t(dx_t|v^{t-1}, y^{t-1}, s), t = 1, \ldots, n \). Define \( \overline{P}_{[0,\infty]}(\kappa) \) as in (300), with \( x^{t-1} \) being replaced by \( v^{t-1}, t = 1, \ldots, n \).
(a) Then
\[
P_t(dx_t|x^{t-1}, y^{t-1}, s) = P_t(dx_t|v^{t-1}, y^{t-1}, s_0), \quad t = 1, \ldots, n, \quad \text{by } Y_t = X_t + V_t \quad \text{(A36)}
\]
\[
= \overline{P}_t(dx_t|s_0, y^{t-1}, s), \quad \text{Definition 8.} \quad \text{(A37)}
\]
The PO-SS realization, for a fixed $S^t_1 = s$, is then

$$V_t = C_t S^t_i + N_t W_t, \quad S^t_i = s, \quad t = 1, \ldots, n,$$

(A38)

$$S^t_{i+1} = A_t S^t_i + B_t W_t, \quad S^t_1 = s.$$  

(A39)

Then,

$$P_t(dy_i|x^t, y^{t-1}, s) = P_t(dy_i|x^t, y^{t-1}, s)$$

(A40)

$$= P_t(dy_i|x^t, s^t, y^{t-1}, s), \quad \text{by Definition 8, (A1)}$$

(A41)

$$= P_t(dy_i|x_t, s^t), \quad \text{by Definition 2, (A1)}$$

(A42)

$$= P_t(dy_i|x_t, s_t, s), \quad \text{by Definition 8, (A1)}$$

(A43)

The probability distribution $P_t(dy_i|y^{t-1}, s)$ is then given by

$$P_t(dy_i|y^{t-1}, s) = \int P_t(dy_i|x_t, s_t) P_t(dx_i|s_t, y^{t-1}, s)$$

$$\cdot P_t(dx_i|y^{t-1}, s), \quad t = 1, \ldots, n, \quad \text{by reconditioning and (A43).}$$

(A44)

The pay-off is the sum of conditional entropies $\sum_{i=1}^n H(Y_i|Y^{t-1}, s)$, and the constraint is (300). By Definition 2, the state $S^t_i, t = 2, \ldots, n$ is Markov, $P_{S^t_i|S^{t-1}_i} = P_{S^t_i|S^{t-1}_i, 1}, \quad t = 2, \ldots, n$. By (A44) and the Markov property of $S^t_i$, $t = 1, \ldots, n$, at each time $t$, the input distribution $P_t(dy_i|y^{t-1}, s)$ depends on $P_t(dx_i|s_t, y^{t-1}, s)$, $j = 1, \ldots, t - 1$ and not on $P_t(dx_i|s_t, y^{t-1}, s)$, $j = 1, \ldots, t - 1$. By (301) and (303), $P_t(dx_i|s_t, y^{t-1}, s) = P_t(dx_i|s_t, y^{t-1}, s)$, $t = 1, \ldots, n$. It is noted that (301) and (303) also follow from a slight variation of the derivation given in [8] (Theorem 1). By the maximum entropy principle of Gaussian distributions, it then follows that the distribution $P_t(dx_i|s_t, y^{t-1}, s)$ is conditionally Gaussian, with linear conditional mean and nonrandom conditional covariance. Then, (304) follows by repeating Step 2 of the derivation of Theorem 3. This completes the derivation of all statements of part (a).

(b,c). The statements follow from part (a), using the generalized Kalman filter, as in Theorem 3.

References

1. Kourtellaris, C.; Charalambous, C.D.; Sergey, L. New Formulas for Ergodic Feedback Capacity of AGN Channels Driven by Stable and Unstable Autoregressive Noise. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020.

2. Louka, S.; Kourtellaris, C.; Charalambous, C.D. Qualitative Analysis of Feedback Capacity of AGN Channels Driven by Stable and Unstable Autoregressive Moving Average. In Proceedings of the 2021 IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021.

3. Charalambous, C.D.; Kourtellaris, C.; Louka, S. Sequential Characterization of Cover and Pombra Gaussian Feedback Capacity: Generalizations to MIMO Channels via a Sufficient Statistic. In Proceedings of the 2021 IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021.

4. Kim, Y.H. Feedback Capacity of Stationary Gaussian Channels. *IEEE Trans. Inf. Theory* **2010**, *56*, 57–85. [CrossRef]

5. Cover, T.; Pombra, S. Gaussian feedback capacity. *IEEE Trans. Inf. Theory* **1999**, *35*, 37–43. [CrossRef]

6. Derpich, M.S.; Östergaard, J. Comments on “Feedback Capacity of Stationary Gaussian Channels”. *IEEE Trans. Inf. Theory* **2024**, *70*, 1848–1851. [CrossRef]

7. Gattami, A. Feedback Capacity of Gaussian Channels Revisited. *IEEE Trans. Inf. Theory* **2019**, *65*, 1948–1960. [CrossRef]

8. Yang, S.; Kavcic, A.; Tatakonda, S. On Feedback Capacity of Power-Constrained Gaussian Noise Channels with Memory. *Inf. Theory IEEE Trans. Theory* **2007**, *53*, 929–954. [CrossRef]
9. Ihara, S. On the Feedback Capacity of the First-Order Moving Average Gaussian Channel. *Jpn. J. Stat. Data Sci.* 2019, 2, 491–506. [CrossRef]
10. Charalambous, C.D.; Kourtellaris, C.; Louka, S. New Formulas of Feedback Capacity for AGN Channels with Memory: A Time-Domain Sufficient Statistic Approach. *arXiv* 2020, arXiv:2010.06226. [CrossRef]
11. Liu, T.; Han, G. Feedback Capacity of Stationary Gaussian Channels Further Examined. *IEEE Trans. Inf. Theory* 2019, 64, 2494–2506. [CrossRef]
12. Li, C.; Elia, N. Youla coding and computation of Gaussian feedback capacity. *IEEE Trans. Inf. Theory* 2019, 64, 3197–3215. [CrossRef]
13. Butman, S. Linear Feedback Rate Bounds for Regressive Channels. *IEEE Trans. Inf. Theory* 1976, 22, 363–366. [CrossRef]
14. Wolfowitz, J. Signalling over a Gaussian Channel with Feedback and Autoregressive Noise. *J. Appl. Probab.* 1975, 12, 713–723. [CrossRef]
15. Ihara, S. *Information Theory for Continuous Systems*; World Scientific: Singapore, 1993; pp. I–XIII, 1–308.
16. Caines, P.E. *Linear Stochastic Systems*; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1988.
17. Kailath, T.; Sayed, A.; Hassibi, B. *Linear Estimation*; Prentice Hall: Hoboken, NJ, USA, 2000.
18. Gallager, R.T. *Information Theory and Reliable Communication*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.
19. Butman, S. A General Formulation of Linear Feedback Communications systems with Solutions. *IEEE Trans. Inf. Theory* 1969, 15, 392–400. [CrossRef]
20. Ebert, P.M. The capacity of the Gaussian channel with feedback. *Bell Systs. Tech. J.* 1970, 49, 1705–1712. [CrossRef]
21. Tienan, J.; Schalkwijk, J.P.M. An upper bound to the capacity of bandlimitted Gaussian autoregressive channel with noiseless feedback. *IEEE Trans. Inf. Theory* 1974, 20, 311–316. [CrossRef]
22. Dembo, A. On Gaussian Feedback Capacity. *IEEE Trans. Inf. Theory* 1989, 35, 1072–1076. [CrossRef]
23. Ihara, S.; Yanaki, K. Capacity of discrete-time Gaussian channels with and without feedback-II. *Ipn. J. Appl. Math.* 1989, 6, 245–258. [CrossRef]
24. Ozarow, L. Upper Bounds on the Capacity of Gaussian Channel with Feedback. *IEEE Trans. Inf. Theory* 1984, 36, 156–161. [CrossRef]
25. Ozarow, L.H. Random coding for additive Gaussian channels with Feedback. *IEEE Trans. Inf. Theory* 1984, 36, 17–22. [CrossRef]
26. Yanaki, K. Necessary and sufficient conditions for the capacity of the discrete-time Gaussian channel to be increased by feedback. *IEEE Trans. Inf. Theory* 1992, 38, 1788–1791. [CrossRef]
27. Yanaki, K. An upper bound on the discrete-time Gaussian channel with feedback-II. *IEEE Trans. Inf. Theory* 1994, 40, 588–593. [CrossRef]
28. Chen, H.W.; Yanaki, K. Refinements of the half-bit and factor-of-two bounds for capacity in Gaussian channels with feedback. *IEEE Trans. Inf. Theory* 1999, 45, 316–325.
29. Chen, H.W.; Yanaki, K. Upper bounds on the capacity of discrete-time blockwise white Gaussian channels with feedback. *IEEE Trans. Inf. Theory* 2000, 43, 1125–1131. [CrossRef]
30. Gallager, R.G.; Nakiboglu, B. Variations on a Theme by Schalkwijk and Kailath. *IEEE Trans. Inf. Theory* 2010, 56, 6–17. [CrossRef]
31. Ordentlich, E. A Class of Optimal Coding Schemes for Moving Average Additive Gaussian Noise Channels with Feedback. In Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT), Trondheim, Norway, 27 June–1 July 1994; p. 467.
32. Tatikonda, S.; Mitter, S. The Capacity of Channels with Feedback. *IEEE Trans. Inf. Theory* 1999, 55, 323–349. [CrossRef]
33. Liu, T.; Han, G. The ARMAk Gaussian Feedback Capacity. In Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT), Aachen, Germany, 25–30 June 2017; pp. 211–215.
34. Kim, Y.H. Feedback capacity of the first-order moving average Gaussian channel. *IEEE Trans. Inf. Theory* 2006, 52, 3063–3079. [CrossRef]
35. Kourtellaris, C.; Charalambous, C.D.; Sergey, L. From Feedback Capacity to Tight Achievable Bounds without Feedback for AGN Channels with Stable and Unstable Autoregressive Noise. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020.
36. Kourtellaris, C.; Charalambous, C.D. Information Structures of Capacity Achieving Distributions for Feedback Channels with Memory and Transmission Cost: Stochastic Optimal Control & Variational Equalities. *IEEE Trans. Inf. Theory* 2018, 64, 4962–4992.
37. Charalambous, C.D.; Kourtellaris, C.; Loyka, S. Capacity Achieving Distributions and Separation Principle for Feedback Gaussian Channels with Memory: The LQG Theory of Directed Information. *IEEE Trans. Inf. Theory* 2018, 64, 6384–6418. [CrossRef]
38. Sabag, O.; Kostina, V.; Hassibi, B. Feedback Capacity of MIMO Gaussian Channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Melbourne, Australia, 12–20 July 2021.
39. Kumar, P.R.; Varaiya, P. *Stochastic Systems: Estimation, Identification, and Adaptive Control*; Prentice Hall: Hoboken, NJ, USA, 1986.
40. van Schuppen, J.H. *Control and System Theory of Discrete-Time Stochastic Systems*; Springer: Berlin/Heidelberg, Germany, 2021.

41. Charalambous, C.D.; Louka, S. A Riccati-Lyapunov Approach to Nonfeedback Capacity of MIMO Gaussian Channels Driven by Stable and Unstable Noise. In Proceedings of the 2022 IEEE Information Theory Workshop (ITW), Mumbai, India, 1–9 November 2022; pp. 184–189.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.