Distortion Outage Analysis for Joint Space-Time Coding and Kalman Filtering

Reza Parseh*, Kimmo Kansanen*, Dirk Slock†
*Department of Electronics and Telecommunications, NTNU, Trondheim, Norway
†Mobile Communications Group, EURECOM, Sophia-Antipolis, France
{reza.parseh, kimmo.kansanen}@iet.ntnu.no, dirk.slock@eurecom.fr

***Abstract*—In this paper, we consider the scenario of transmitting a first order Gauss-Markov vector signal over a MIMO Rayleigh non-frequency selective fading channel. The signal is reconstructed at the receiver side with the help of a Kalman filter in order to minimize the mean squared error. Orthogonal space time codes are utilized in order to increase the quality of estimation and mitigate the destructing effects of the fading channel. As a criterion for estimation quality assessment, we use the distortion outage probability. We first obtain upper and lower bounds for the outage probability as a function of system parameters. We then perform high SNR analysis of the bounds, through which we prove the achievability of the maximum diversity order for a $N \times K$ MIMO fading channel. In addition, we obtain upper and lower bounds for the coding gain of the distortion outage probability in the high SNR regime, and outline the relation between system parameters and the coding gain.**

***Index Terms*—Kalman filter, orthogonal space-time coding, MIMO fading channels, diversity order, coding gain**

I. Introduction

Analog (uncoded) transmission of discrete-time sources over fading channels is an alternative to the state-of-the-art digital communication systems due to its simplicity and zero-delay property. Analog transmission is specifically attractive for delay sensitive settings, where state-of-the-art channel coding schemes cannot be used due to their need for large buffers. These settings include, but are not limited to, control over wireless channels, real-time monitoring e.g. in sensor networks, and real-time event detection e.g. in intelligent traffic management systems.

Using the analog scheme, the transmitted signals need to be estimated from the channel outputs at the receiver side. It is not possible to add redundancy to the transmitted signals to mitigate the effects of the communication channel either. However, one can exploit the correlation which already exists among the signal samples. The auto-regressive (AR) signal model is widely used to model the correlation in natural signals. With the AR model, Kalman (-like) estimation algorithms may be used at the receiver side to minimize the distortion.

In order to measure the quality of estimation at the receiver, two main criteria have been considered before, namely end-to-end average distortion and distortion outage probability. The end-to-end average distortion measure was first studied in [1] and later considered in [2]–[4]. For AR Gauss-Markov models, the end-to-end average distortion, corresponding to the mean of the covariance matrix for the Kalman filter, was also considered in [5].

For the delay sensitive settings and when the mean squared error (MSE) is random due to random channels or modeling errors, the distortion outage probability measure is more insightful than end-to-end average distortion. This measure has also been considered before in [6]–[8]. In order to get a reliable estimation at the receiver and combat the effects of fading, one can incorporate diversity schemes. In [6], estimation outage and estimation diversity are considered in the context of distributed sensing, where several sensors observe an i.i.d. process and transmit their measurements over parallel fading channels. A similar system model is considered in [7], where the focus is on distortion outage minimization. Estimation error outage minimization is also considered in [9], where power allocation strategies are developed for outage minimization and for state estimation with multiple sensors. Transmission is assumed over fading channels, where each sensor knows its own channel gain only, while all the channel gains are known to the receiver. In [8] and with an information theoretic approach, a diversity order analysis is presented for distortion outage probability of source transmission over multi-input multi-output (MIMO) block fading channels. Apart from a slight difference between the definition of distortion in [8] and in our work, the main differences are in the practical scheme of this work over the information theoretical analysis of [8] and the existence of source time correlation in our work. Estimation error outage probability was also used in [10], [11] for transmission of AR Gauss-Markov scalar sources over fading channels, with respectively one and several receive antennas and related high signal to noise ratio (SNR) analysis.

As we will see later in the paper, the the random MSE (estimation error covariance matrix) of the Kalman filter propagates in time through the random Riccati equation. With transmission over a fading channel, stability of the Kalman filter might become an issue due to the random fading channel. The random fading channel has varying quality, and measurement updates for the Kalman filter are thus available with varying quality, which at times might be very poor. The stability issue of the Kalman filter with random system parameters has been investigated in the literature before with respect to various stability measures. To name a few of such works, one can mention [5], where boundedness of the expected error covariance matrix (expectation taken over the channel randomness) for estimation over fading channels is proven. In this paper however, the outage probability measure is considered, which is a more suitable performance analysis

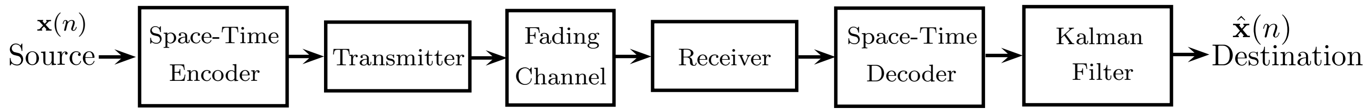

Fig. 1. General system model for joint Kalman filtering and space-time coding

criterion for low latency applications. A Bernoulli packet-drop channel model is considered in [12] and it is shown that the random prediction (rather than estimation) error covariance matrix converges in distribution. While the authors of [12] consider a packet loss channel model, we utilize a specific model for the channel attenuation, allowing a more detailed channel behavior analysis, and the straightforward inclusion of spatial channel characteristics.

Stability of the random Riccati equation for Kalman filtering has also been considered in [13], [14] when several system parameters are random. In [13], almost sure stabilizability and almost sure detectability measures are studied for linear systems with random system parameters, and in [14], the conditions for the weak convergence (in distribution) of the random Riccati equation are presented. These conditions are also used in this work to prove the existence of the outage probability. While more general system models are treated in [13], [14], it is only convergence to a distribution which is proven for the random estimation error and the actual distributions are not provided. We however, study the distribution itself in terms of outage probability, diversity order, and coding gain, but perform this for the special case Rayleigh fading channels.

This work follows the line of work in [10] and [11], which only consider scalar sources, and extends them to vector sources and MIMO fading channels. The vector AR model is more general and more realistic, especially for settings such as wireless control, where the measurements of a plant with arbitrary number of states need to be transmitted to the control center in a delay-free manner. In order to provide the possibility for a transmit diversity gain, we use complex orthogonal space-time codes as originally introduced in [15]. The idea of using space-time codes for estimation has been considered before in [16] for application in MIMO channel estimation. In our work, we adapt the space-time decoding scheme such that it can be used together with the Kalman filter. Figure 1 depicts the general structure of the proposed scheme, consisting mainly of a space-time block encoder at the transmitter side, and a space-time block decoder and a Kalman filter at the receiver side. We consider the Rayleigh fading channels for simplicity of analysis, but the results can be generalized to other channel distributions. We also consider the case of fully observable signals, and concentrate on the characterization of the error behavior of the estimator, and the diversity effects through space-time coding. While limiting the generality of the results, the assumption allows for simpler analysis of the equivalent linear system. It is likely that a complete characterization of the general case would require extended and different analytical tools to be utilized.

The main contributions of this work are the following.

- We jointly incorporate space-time codes and Kalman filtering in a common framework, which allows for extra reliability for delay-free estimation of AR Gauss-Markov sources over fading channels.
- We improve the procedure for decoding the complex orthogonal space-time coding for analog sources, by performing the decoding operation in separate real and imaginary parts. This allows for the use of any general complex orthogonal space-time code and improves over [16], which is only applicable for half-rate codes.
- We provide bounds for the distortion outage probability, which allows for a practical use of the current scheme and facilitates design.
- We perform high SNR analysis for the distortion outage probability, and show that the proposed scheme can achieve the maximum diversity order for transmission over a $N \times K$ MIMO fading channel (K transmit-antennas and N receive antennas), i.e. NK.
- We propose upper and lower bounds for the coding gain of the outage probability in the high SNR regime, which completes the high SNR analysis of the distortion outage probability.

The rest of this paper is organized as follows. We present the system model and the formal problem definition in Sec. II. We then study the details of the joint space-time coding and Kalman filtering scheme in in Sec. III. Furthermore, we focus on the distortion outage probability analysis in Sec. IV, and finally present the simulations and numerical evaluations in Sec. V.

II. System Model and Problem Definition

Consider the following system model

$$\begin{aligned} \mathbf{x}(n) &= A\mathbf{x}(n-1) + \mathbf{u}(n) \\ Y(n) &= \sqrt{P/(KN)}H(n)T\left(\mathbf{x}(n)\right) + V(n), \end{aligned} \tag{1}$$

where $\mathbf{x}(n)$ and $\mathbf{u}(n)$ are column vectors of dimension K and represent the to-be-transmitted signal and the process noise, respectively. In this model, $\mathbf{x}(n)$ is a first order Gauss-Markov process. With respect to that, A is the state-transition matrix and we assume it to be stable and non-singular. Non-singularity is a sufficient condition for existence of the steady-state outage probability function and is explained in details in IV-A. Stability of A is also required, because otherwise the transmission power grows unbounded and the scheme becomes impractical.

The space-time block encoding operation is represented by the operator $T(\cdot)$. The output of the space-time encoding operation, i.e. $T\left(\mathbf{x}(n)\right)$, is a matrix of dimension $K \times N_c$, which corresponds to N_c channel uses by each of the transmit antennas for each new source symbol $\mathbf{x}(n)$. (The details of the space-time coding operation and the structure of $T(\mathbf{x}(n))$ are presented in Sec. III-A). In this work, we assume that the number of transmit antennas is equal to K, i.e. the source dimension.

The MIMO channel matrix of dimension $N \times K$ is denoted by $H(n)$, which consists of i.i.d. complex Gaussian elements with zero mean and unit variance (real and imaginary parts have a variance equal to one half), i.e. non-frequency selective Rayleigh fading. For some of our derivations in the upcoming sections, we need that $H(n)$ are also i.i.d. in time and that $H(1) \neq 0$ (the latter is required in lemma 1). We acknowledge that the i.i.d. assumption is somewhat limiting, as the transmission rate will then be limited as a function of the coherence time of the fading channel. However, in settings such as wireless sensor networks, there is usually no need for constant high rate transmission. In that case, we can have N_c channel uses in a burst mode only when the new channel values can be considered independent from the previous one.

At the receiver, the received signals from the channel and the channel noises for the N_c channel uses are denoted by $Y(n)$ and $V(n)$, which are matrices of dimension $N \times N_c$. The value of P is also selected such that required SNR at the receiver is achieved. We also consider the elements of $V(n)$ to be i.i.d. complex Gaussian random variables. The covariance matrices for $\mathbf{u}(n)$ is denoted by C_u and the elements of $V(n)$ have a variance equal to σ_v^2. We further assume that the $H(n)$ are perfectly known to the receiver. Note that the above system model suggests that the source signal $\mathbf{x}(n)$ is fully observable.

The vector source $\mathbf{x}(n)$ is space-time encoded at the transmitter side and sent over the channel. There are then two major operations which should be performed at the receiver. The first operation is the space-time decoding, i.e. the inverse operation for $T(\mathbf{x}(n))$, which in turn leads to an equivalent channel and received signal model. The next operation is the estimation of $\mathbf{x}(n)$ from the received signal. The optimal causal minimum mean square error (MMSE) estimator for this setting is the Kalman filter. The Kalman filter provides us with an optimal estimate of the source at the receiver, namely $\hat{\mathbf{x}}(n)$ which minimizes the (normalized) random instantaneous distortion at time n, i.e.

$$d(n) = \frac{1}{K} E\left(\|\mathbf{x}(n) - \hat{\mathbf{x}}(n)\|^2\right). \tag{2}$$

We then define the distortion outage probability as

$$P_{\text{out}}(d_{\text{th}}) = \Pr\left(d(n) \geqslant d_{\text{th}}\right), \tag{3}$$

which we are interested to characterize. In addition, we are also interested in the asymptotic behavior of $P_{\text{out}}(d_{\text{th}})$ as a function of

$$\text{SNR} = \frac{P}{\sigma_v^2 KN} E\left(\|H(n)\|^2\right) E\left(\|T(\mathbf{x}(n)\|^2\right),$$

when SNR $\to \infty$, i.e. in the high SNR regime. We define the diversity order for the distortion outage probability as

$$\text{d}_{\text{ord}} = -\lim_{\text{SNR}\to\infty} \frac{\log(P_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})}. \tag{4}$$

Assuming that the diversity order is a finite and positive number, the outage probability may be written as follows

$$P_{\text{out}}(d_{\text{th}}) = (G \cdot \text{SNR})^{-d_{\text{ord}}} + o(\text{SNR}^{-d_{\text{ord}}}). \tag{5}$$

In this formulation, d_{ord} is the diversity order and the value G may be called the coding gain, parallel to the terms used in outage analysis in digital communication [17]. While the diversity order is the slope of the outage probability function vs. SNR in the log-log scale, G denotes the average relative asymptotic power gain. As we will see later, the maximum diversity order is only dependent on the number of available independent individual channel branches, and its achievability only depends on the space-time code. The coding gain however is a function of the source structure and the selected threshold. Although the diversity order provides a useful rule of thumb for evaluating the quality of estimation at high SNR, the coding gain provides a more complete characterization of performance, which allows for comparison of systems which have the same diversity order, but different coding gains.

III. Joint Space-time Coding and Kalman Filtering

In this section, we first describe the space-time coding scheme used in this paper and then the Kalman filter which is used in order to estimate the transmitted signal. We then describe how these two parts should interact with one another.

A. General Space-time Coding for Analog Communication

A space-time block code based on orthogonal designs as defined in [15], is used for transmission of $\mathbf{x} = \{x_1, x_2, \ldots, x_K\}$ over the channel $H_{N\times K}$ (we drop the time index n in this section). The encoding is adopted (and slightly modified) from [15] as follows. We form a matrix $X = T(\mathbf{x})$ of dimension $K \times N_c$, as instructed in [15] and which consists of elements $\pm x_1, \pm x_2, \ldots, \pm x_K$, their conjugates $\pm x_1^*, \pm x_2^*, \ldots, \pm x_K^*$ or multiples of these elements by $\pm \boldsymbol{i}$ (with $\boldsymbol{i} = \sqrt{-1}$) or, if necessary, other scaling factors. The first column of X can without loss of generality be assumed to be $[x_1, x_2, \ldots, x_K]^T$, where T denotes the transpose operation. The space-time code rate can be defined as

$$r = K/N_c \tag{6}$$

source dimensions per channel use. The code rate needs to be maximized in order to minimize the extra incurred channel uses. This is however not the focus of this work and we mainly refer to the current literature on space-time codes for that matter. The code design is such that $XX^\dagger$ ($X^\dagger$ being the conjugate transpose of X) is a diagonal matrix. It is also shown in [15] that if at least one orthogonal design exists, one can always find another design such that

$$XX^\dagger = c\|\mathbf{x}\|^2 I_K,$$

where c is some constant depending on the code. It is also possible to normalize the codewords such that $c = 1$, as we will assume in the rest of this paper.

Referring to the system model in (1), we note that with this structure, each row of Y and V corresponds to a particular receiver antenna, comprising a total number of N receive antennas, and each column of Y and V corresponds to one channel use, comprising a total number of N_c channel uses for each source symbol transmission, indexed by n. Note that H, although random, is fixed for the transmission of each source symbol.

At the receiver side, the space-time coded signal Y should be decoded first before it is directed to the Kalman filter in order to estimate $\mathbf{x}$. The objective of the space-time decoder, appearing before the Kalman filter as in Fig. 1, is to provide an equivalent orthogonal channel and as we will see later, allow for a spatial diversity gain. The number of such orthogonal branches for the equivalent channel is at most NK. It is worthwhile mentioning that N has no effect on the code selection as long as the code is orthogonal.

The decoding we suggest here is different from what is proposed in [15] due to the different nature of estimation and detection. We use an approach similar to the one used in [16], where orthogonal space-time block codes are used for analog channel state information feedback. The basic idea is to convert the channel into an equivalent orthogonal channel and then perform decoding by simply multiplying the received vector by the transpose of the equivalent channel (matched filtering). The method in [16] is applicable for the transmission of real signals only, and uses the 1/2 rate codes based on real orthogonal designs proposed in [15]. While the same design can be used for our purpose as well (by alternating the transmission of real and imaginary parts of the signal $\mathbf{x}(n)$ and therefore having a code of rate 1/4), we propose a different approach which allows for the incorporation of the available complex orthogonal space-time codes, and thus operating at a better rate (e.g. full rate for $K = 2$). This is made possible by converting all the complex vectors into equivalent real vectors with twice the size and finding the equivalent real channel. The proposed space-time decoding can be performed as follows.

Consider the l-th row of H corresponding to the l-th receiver ($l = 1, 2, \ldots, N$), and call that row $\mathbf{h}_l$. We then take the corresponding rows in Y and V to be $\mathbf{y}_l$ and $\mathbf{v}_l$. The received signal for that receiver is $\mathbf{y}_l = \sqrt{P/(KN)}\mathbf{h}_l X + \mathbf{v}_l$. Note that it is only enough to analyze the space-time code for one receiver antenna. Through that, we are able to show that each receiver is able to provide K orthogonal channels, involving the corresponding row in H. With N independent rows, a number of NK orthogonal channels can be created by simply summing all the results of the space-time decoding for each receiver.

The next step is to convert all the complex operations to real ones. First extend the source vector $\mathbf{x}$ into a real vector of dimension $2K$ by replacing each complex element by a 2×1 real vector of the real and imaginary part of that element and call this new vector $\mathbf{x}_r$, i.e.

$$\mathbf{x}_r = [x_1^r, x_1^i, x_2^r, x_2^i, \ldots, x_K^r, x_K^i]^T,$$

where the superscripts r and i for each element x_k, $k = 1, 2, \ldots, K$ indicate the real and imaginary part of that element. Then perform the same procedure for X to provide the matrix X_r of dimension $2K \times N_c$. Vectorize (column-wise reshape) the matrix X_r into a real vector of size $2KN_c \times 1$, in the same manner as we created $\mathbf{x}_r$, and call it $\tilde{\mathbf{x}}$. It is now possible to create a mapping matrix T with size $2KN_c \times 2K$ which maps $\mathbf{x}^r$ to $\tilde{\mathbf{x}}$ and only consists of real numbers. In other words, we must find a T such that it satisfies $\tilde{\mathbf{x}} = T\mathbf{x}^r$. This can be done by considering that each element in $\tilde{\mathbf{x}}$ can be found in $\mathbf{x}^r$, possibly with a different sign and scaling factor.

We convert $\mathbf{v}_k$ to the equivalent real vector $\tilde{\mathbf{v}}$ in the same manner.

Next, we consider converting the channel into the real and imaginary parts. For this reason, each channel tap in $\mathbf{h}_l$, i.e. $h_{l,k}$, $k = 1, 2, \ldots, K$ is converted to the following 2×2 matrix

$$\tilde{H}_{l,k} = \begin{bmatrix} h_{l,k}^r & -h_{l,k}^i \\ h_{l,k}^i & h_{l,k}^r \end{bmatrix}.$$

The complex-valued $1 \times K$ vector $\mathbf{h}_l$ is then expanded to the real-valued $2 \times 2K$ channel $\tilde{H} = [\tilde{H}_{l,1}|\tilde{H}_{l,2}, \ldots, |\tilde{H}_{l,K}]$. With these definitions, it can be easily shown that the operation $\mathbf{y}_l = \sqrt{P/(KN)}\mathbf{h}_l X + \mathbf{v}_l$ in the domain of complex numbers can be represented (using the Kronecker product $\otimes$) by the following operation over the domain of real numbers

$$\begin{aligned} \tilde{\mathbf{y}} &= \sqrt{P/(KN)}(I_{N_c} \otimes \tilde{H})\tilde{\mathbf{x}} + \tilde{\mathbf{v}} \\ &= \sqrt{P/(KN)}(I_{N_c} \otimes \tilde{H})T\mathbf{x}^r + \tilde{\mathbf{v}} \\ &= \sqrt{P/(KN)}H_{\text{eq}}\mathbf{x}^r + \tilde{\mathbf{v}}, \end{aligned} \tag{7}$$

where $H_{\text{eq}} = (I_{N_c} \otimes \tilde{H})T$ is the equivalent real channel which acts on the equivalent real source vector $\mathbf{x}_r$ (note the conversion from complex row vectors to equivalent real column vectors). We show in Appendix A that

$$H_{\text{eq}}^T H_{\text{eq}} = \|\mathbf{h}_l\|^2 I_{2K},$$

i.e. the equivalent real channel can be orthogonalized by a simple matched filtering operation constructed as follows

$$\begin{aligned} H_{\text{eq}}^T\tilde{\mathbf{y}} &= H_{\text{eq}}^T\left(\sqrt{P/(KN)}H_{\text{eq}}\mathbf{x}^r + \tilde{\mathbf{v}}\right) \\ &= \sqrt{P/(KN)}H_{\text{eq}}^T H_{\text{eq}}\mathbf{x}^r + H_{\text{eq}}^T\tilde{\mathbf{v}} \\ &= \sqrt{P/(KN)}\|\mathbf{h}_l\|^2\mathbf{x}^r + H_{\text{eq}}^T\tilde{\mathbf{v}}. \end{aligned} \tag{8}$$

Due to the independence of channel noises for each dimension, it can be shown that the variance of each element of $H_{\text{eq}}^T\tilde{\mathbf{v}}$ is equal to $\|\mathbf{h}_l\|^2\sigma_v^2$.

For decoding the whole received signals, one should perform the same procedure for all the receiver antennas, i.e. all the rows of H, and sum the results. If we call the resulting sum $\mathbf{y}_{\text{eq}}$ and convert the real vectors back to the complex domain again, we may finally write

$$\begin{aligned} \mathbf{y}_{\text{eq}} &= \sqrt{P/(KN)}\sum_{l=1}^{N}\|\mathbf{h}_l\|^2\mathbf{x} + \mathbf{v}_{\text{eq}} \\ &= \sqrt{P/(KN)}\|H\|_F^2\mathbf{x} + \mathbf{v}_{\text{eq}}, \end{aligned} \tag{10}$$

where each element in $\mathbf{v}_{\text{eq}}$ has the variance $\|H\|_F^2\sigma_v^2$. The SNR is then equal to SNR $= PP_x/\sigma_v^2$, where $P_x = E(\|\mathbf{x}(n)\|^2)$. After space-time decoding for each time step n, $\mathbf{y}_{\text{eq}}(n)$ is delivered to the Kalman filter in order to estimate $\mathbf{x}(n)$. This is reviewed in the next section.

B. Kalman Filtering of Space-time Coded Analog Sources

Given that the correct initialization and equivalent channel and received signal model are used, the general equations for the Kalman filter (the estimator) adapted from [18] are

$$\begin{aligned}
&\hat{\mathbf{x}}(n|n-1) = A\hat{\mathbf{x}}(n-1|n-1)\\
&P(n) = AM(n-1)A^T + C_u\\
&K(n) = \sqrt{P/(KN)}P(n)\|\mathbf{H}(n)\|^2\times\\
&\left(\|\mathbf{H}(n)\|^2\sigma_v^2 + P/(KN)P(n)\|\mathbf{H}(n)\|^4\right)^{-1}\\
&\hat{\mathbf{x}}(n|n) = \hat{\mathbf{x}}(n|n-1)\\
&\qquad + K(n)\left(\mathbf{y}_{\text{eq}}(n) - \sqrt{P/(KN)}\|\mathbf{H}(n)\|^2\hat{\mathbf{x}}(n|n-1)\right)\\
&M(n) = \left(I - K(n)\sqrt{P/(KN)}\|\mathbf{H}(n)\|^2\right)P(n). \qquad (11)
\end{aligned}$$

The second step in (11), i.e. $P(n) = AM(n-1)A^T + C_u$ is called prediction and it is known that the prediction error covariance matrix $P(n)$ propagates through the random Riccati equation given in (12).

We can also simplify (12) to

$$\begin{aligned}
&P(n+1)\\
&= A\left(P^{-1}(n) + P/(KN)\|\mathbf{H}(n)\|^2/\sigma_v^2 I\right)^{-1}A^T + C_u\\
&= AM(n)A^T + C_u \qquad (13)
\end{aligned}$$

by invoking the Woodbury matrix identity on (13). Comparing (13) with the second line in (11) necessitates that

$$M(n) = \left(P^{-1}(n) + P/(KN)\|\mathbf{H}(n)\|^2/\sigma_v^2 I\right)^{-1}. \qquad (14)$$

We may then rewrite (14) as

$$\begin{aligned}
M(n) &= \left(P^{-1}(n) + P/(KN)\|\mathbf{H}(n)\|^2/\sigma_v^2 I\right)^{-1}\\
&= \frac{KN}{P\|\mathbf{H}(n)\|^2/\sigma_v^2}\left(\left(P/(KN)\|\mathbf{H}(n)\|^2/\sigma_v^2 P(n)\right)^{-1} + I\right)^{-1}\\
&= \frac{1}{\gamma_n}\left(\frac{1}{\gamma_n}P^{-1}(n) + I\right)^{-1}, \qquad (15)
\end{aligned}$$

with

$$\gamma_n = \frac{P\|\mathbf{H}(n)\|^2}{\sigma_v^2 KN} \qquad (16)$$

denoting the instantaneous channel SNR at time instant n. One can also rewrite (13) (while setting $n-1$ instead of n) as

$$\begin{aligned}
P(n) &= AM(n-1)A^T + C_u\\
&= \frac{A}{\gamma_n}\left(\frac{1}{\gamma_n}P^{-1}(n-1) + I\right)^{-1}A^T + C_u. \qquad (17)
\end{aligned}$$

If we denote the k'th diagonal element of $M(n)$ by $M_{kk}(n)$ and define the distortion as $d(n) = \frac{1}{K}\text{tr}\left(M(n)\right)$, the distortion outage probability at time n is equal to

$$\begin{aligned}
\text{P}_{\text{out}}(\text{d}_{\text{th}}) &= \Pr\left(d(n) \geqslant \text{d}_{\text{th}}\right)\\
&= \Pr\left(\frac{1}{K}\sum_{k=1}^{K} M_{kk}(n) \geqslant \text{d}_{\text{th}}\right), \qquad (18)
\end{aligned}$$

where d_{th} is an arbitrary threshold value. The analysis of this outage probability as a function of SNR and other system parameters is the topic of the next section.

IV. Outage Probability Analysis

We begin this section by first proving that $d(n)$ converges in distribution and consequently that the outage probability as defined in (18) exists. After that, we study the achievable diversity order and coding gain for distortion outage probability when the orthogonal space-time codes are used in conjunction with the Kalman filter. We also develop upper and lower bounds on the distortion outage probability, which are used for both obtaining numerical values with application in practical systems and also a prerequisite tool in obtaining the asymptotic results.

A. Existence of the Outage Probability

The proof presented here for the existence of the outage probability is very similar to the proof provided in [10] for the scalar case. We start by lemma 1 to prove that the process $M(n)$ converges in distribution. We then use this fact to show the existence of a stationary outage probability.

Lemma 1. *The random process $M(n)$ converges in distribution.*

Proof. Convergence of the estimation error covariance matrix of the Kalman filter (denoted by $M(n)$ in this paper) with stochastic system parameters, is studied in [14, Theorem 2.4]. For convergence in distribution, it is first required that a hypothesis $\mathcal{H}$ (defined in [14, Section 2]) is satisfied. Secondly, it is necessary that the system is weakly observable and weakly controllable as defined in [14, Definition 2.1]. Thirdly, certain system parameters which we name later, must be integrable.

For hypothesis $\mathcal{H}$ to hold, it is mentioned in [14, Section 2] that a conditionally Gaussian system satisfies such a requirement. In our equivalent system model, $\mathbf{u}(n)$ and $\mathbf{v}_{\text{eq}}(n)$ are i.i.d. Gaussian random processes and also independent of $H(n)$. Therefore, the system in this paper is also conditionally Gaussian and satisfies the aforementioned hypothesis $\mathcal{H}$.

Then, it must be shown that the system is weakly controllable and weakly observable. The exact definitions for weak controllability and weak observability are provided in [14,

$$P(n+1) = AP(n)A^T - P/(KN)AP(n)\|\mathbf{H}(n)\|^2\left(\|\mathbf{H}(n)\|^2\sigma_v^2 + P/(KN)\|\mathbf{H}(n)\|^4 P(n)\right)^{-1}\|\mathbf{H}(n)\|^2 P(n)A^T + C_u \qquad (12)$$

Definition 2.1]. From there and in order to define weak controllability and weak observability, the following probabilities are defined first.

$$\epsilon_o = \Pr\left(\text{Det}\left(A^T\|\gamma(1)\|^2A + (A^T)^2\|\gamma(2)\|^2A^2 + \ldots + (A^T)^n\|\gamma(n)\|^2A^n\right) > 0\right)$$
$$\epsilon_c = \Pr\left(\text{Det}\left(C_u + AC_uA^T + A^2C_u(A^T)^2 + \ldots + A^nC_u(A^T)^n\right) > 0\right),$$

where Det(.) represents the determinant of a matrix.

For weak observability, it must hold that ϵ_o is non-zero and for weak controllability, it must hold that ϵ_c is non-zero. It is easy to show that ϵ_c is non-zero as the determinant of the sum of positive definite matrices is non-zero. The same argument holds for ϵ_o as well, as long as the non-existent channel ($h(n) = 0$ for all n) does not occur. The non-existent channel may only happen with zero-probability for Rayleigh fading. Therefore, the system is weakly controllable.

For the third condition to hold, it should be such that the (random) variables $\log\log^+(A)$, $\log\log^+(A^{-1})$, $\log\log^+(C_u)$ and $\log\log^+(\gamma(1))$ are integrable, where

$$\log^+(x) = \max\left(\log(x), 0\right), \tag{19}$$

i.e. they have a well-defined expectation value (see e.g. [19] Chapter 13 for a definition of integrable random variables). Obviously, A and $C_u > 0$ are deterministic parameters. Therefore, they are integrable. $\log\log^+(\gamma(1))$, is also integrable, given that $H(n)$ is defined as in Sec. 1. As a result, our system model satisfies all the prerequisites of Theorem 2.4 in [14]. The consequence of the aforementioned theorem is that $M(n)$ converges in distribution (law). □

When $M(n)$ converges in distribution, then $d(n)$ which is the normalized sum of the diagonal elements of $M(n)$ also converges in distribution. As a result, $\text{P}_{\text{out}}(\text{d}_{\text{th}})$ as defined in (18) exists.

B. Bounds for the Outage Probability

As the distortion at each time step n is obtained from $M(n)$, we should first try to develop an equation for $M(n)$ which makes distortion calculation possible. One easy way to find the diversity order is finding a closed-form equation for $M(n)$ which is independent of $P^{-1}(n)$ such that it allows for diversity order calculation. This however, proves to be rather complicated. Previously, some efforts had been carried out to characterize the single-input single-output (SISO) and single-input multi-output (SIMO) cases in [10] and [11], respectively. While some initial results were obtained, even for those simpler cases one had to resort to finding bounds and approximates for characterization of the outage probability and then its high SNR behavior. We follow the same approach in this paper as well. In this section, we first establish upper and lower bounds for the outage probability and then in the following section obtain the diversity order and coding gain via the high SNR analysis of the bounds.

In order to get upper and lower bounds for $P_{\text{out}}(d_{\text{th}})$, we use the following fact. It is straightforward to show that if for two random variables X and Y, we have that $X \leqslant Y$ (meaning that $X(\omega) \leq Y(\omega)$ for all ω), then as a result, we obtain that $\Pr(X \geqslant T) \leqslant \Pr(Y \geqslant T)$. As $d(n) = \frac{1}{K}\text{tr}(M(n))$, if we can find $d^l(n)$ and $d^u(n)$ such that we would have $d^l(n) < d(n) < d^u(n)$, then we may bound the outage probability as $\Pr(d^l(n) \geqslant d_{\text{th}}) < \Pr(d(n) \geqslant d_{\text{th}}) < \Pr(d^u(n) \geqslant d_{\text{th}})$, i.e. upper and lower bounds on the outage probability may be established. We would prefer random variables whose cdf have a Taylor series with the first non-zero term equal to that of the original distortion variable. This is for the diversity analysis to be successful and will be explained later in the section.

In the following lemmas (Lemma 2 and 3), we present $d^l(n)$ and $d^u(n)$, used to establish upper and lower bounds on the outage probability.

Lemma 2. *The instantaneous distortion $d(n)$ may be lower bounded by $d^l(n) \leqslant d(n)$ as*

$$d^l(n) = \frac{1}{\gamma_n + \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}},$$

where the $\lambda_l(.)$ function denotes the l-th eigenvalue of its matrix argument.

Proof. Please see Appendix B.

Lemma 3. *The instantaneous distortion $d(n)$ may be upper bounded by $d^u(n) \geqslant d(n)$ as follows*

$$d^u(n) = \frac{1}{\gamma_n + \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l}}$$

with

$$\theta_l = \lambda_l(C_u) + \frac{|\lambda_{max}(A)|^2}{\frac{1}{\alpha_{max}} + \gamma_{n-1}}, \tag{20}$$

where $\lambda_{max}(.)$ denotes the maximum value of the eigenvalues of its matrix argument, and $\alpha_{max} = |\lambda_{max}(A)|^2\lambda_{max}(C_x) + \lambda_{max}(C_u)$, with C_x being the stationary covariance matrix of the source.

Proof. Please see Appendix B.

From Lemmas 2 and 3, and the previous discussion, we can now obtain the bounds for the outage probability as follows. For the lower bound, we have that

$$
\begin{aligned}
P^l_{\text{out}}(d_{\text{th}}) &= \Pr(d^l(n) \geqslant d_{\text{th}}) \\
&= \Pr\left(\frac{1}{\gamma_n + \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}} \geqslant d_{\text{th}}\right) \\
&= \Pr\left(\gamma_n \leqslant (\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)})\right) \\
&= \Pr\left(\frac{P\|\mathbf{H}(n)\|^2}{\sigma_v^2 KN} \leqslant (\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)})\right) \\
&= \Pr\left(\|\mathbf{H}(n)\|^2 \leqslant \frac{\sigma_v^2 KN}{P}(\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)})\right) \\
&= F_{\|\mathbf{H}(n)\|^2}\left(\frac{\sigma_v^2 KN}{P}(\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)})\right),
\end{aligned}
\tag{21}
$$

where $F_{\|\mathbf{H}(n)\|^2}(.)$ is the cdf of the random variable $\|\mathbf{H}(n)\|^2$ and can be obtained from

$$
F_{\|\mathbf{H}(n)\|^2}(z) = \frac{1}{(NK-1)!}\int_0^z \mathrm{e}^{-t} t^{NK-1}\,\mathrm{d}t. \tag{22}
$$

Note that the only difference between (22) and the standard cdf of a χ^2 random variable is a scaling factor of 2 in the integral upper limit z, which is added due to the normalized variance assumption on the individual complex channel paths.

For the upper bound, we similarly have that

$$
\begin{aligned}
P^u_{\text{out}}(d_{\text{th}}) &= \Pr(d^u(n) \geqslant d_{\text{th}}) \\
&= \Pr\left(\frac{1}{\gamma_n + \frac{1}{1/K\cdot\sum_{l=1}^{K}\theta_l}} \geqslant d_{\text{th}}\right) \\
&= \Pr\left(\gamma_n \leqslant \frac{1}{d_{\text{th}}} - \frac{1}{1/K\cdot\sum_{l=1}^{K}\theta_l}\right).
\end{aligned}
\tag{23}
$$

Since θ_l are functions of γ_{n-1}, we may fix the value of γ_{n-1} in order to perform the same procedure as for $P^l_{\text{out}}(d_{\text{th}})$ and then integrate over the pdf of γ_{n-1}, in order to obtain the total outage probability. Note that in order for this procedure to be correct, we need that γ_n and γ_{n-1} are statistically independent. After that, we obtain

$$
\begin{aligned}
&P^u_{\text{out}}(d_{\text{th}}) \\
&= \int_0^\infty \Pr\left(\gamma_n \leqslant \left(\frac{1}{d_{\text{th}}} - \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l}\right)\middle|\gamma_{n-1} = z\right) \\
&\times f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&= \int_0^\infty \Pr\left(\frac{P\|\mathbf{H}(n)\|^2}{\sigma_v^2 KN} \leqslant \left(\frac{1}{d_{\text{th}}} - \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l}\right)\middle|\gamma_{n-1} = z\right) \\
&\times f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&= \int_0^\infty F_{\|\mathbf{H}(n)\|^2}\left(\frac{\sigma_v^2 KN}{P}\left(\frac{1}{d_{\text{th}}} - \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l}\right)\middle|\gamma_{n-1} = z\right) \\
&\times f_{\gamma_{n-1}}(z)\,\mathrm{d}z.
\end{aligned}
\tag{24}
$$

Note that it is only θ_l which is a function of γ_{n-1}.

For the numerical evaluation of the bounds, please see Sec. V.

C. Analysis of Diversity Order and Coding Gain

In order to obtain the diversity order with the help of the bounds as we previously mentioned, we present the following lemma.

Lemma 4. *Assume that the outage probability $P_{out}(d_{th})$ can be lower and upper bounded by $P^l_{out}(d_{th})$ and $P^u_{out}(d_{th})$, respectively, i.e. $P^l_{out}(d_{th}) < P_{out}(d_{th}) < P^u_{out}(d_{th})$ for all system parameters and all n. If $P^l_{out}(d_{th})$ and $P^u_{out}(d_{th})$ have a diversity order of d^0_{ord}, then the outage probability $P_{out}(d_{th})$ has a diversity order of d^0_{ord} as well.*

Proof. Since $P^l_{\text{out}}(d_{\text{th}}) < P_{\text{out}}(d_{\text{th}}) < P^u_{\text{out}}(d_{\text{th}})$ and $\log(.)$ is a monotonic increasing function for all valid (positive) arguments and $\log(\text{SNR})$ is a positive number, then we have that

$$
\frac{\log(P^l_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} < \frac{\log(P_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} < \frac{\log(P^u_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})}. \tag{25}
$$

As we have that $\lim_{\text{SNR}\to\infty}\frac{\log(P^l_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} = -d^0_{\text{ord}}$ and that $\lim_{\text{SNR}\to\infty}\frac{\log(P^u_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} = -d^0_{\text{ord}}$, then according to the well-known squeeze theorem for limits, we obtain that

$$
\lim_{\text{SNR}\to\infty}\frac{\log(P_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} = -d^0_{\text{ord}}, \tag{26}
$$

and the proof is complete. □

At this stage, we only need to find the diversity order for $P^l_{\text{out}}(d_{\text{th}})$ and $P^u_{\text{out}}(d_{\text{th}})$. By a Taylor series expansion of $F_{\|\mathbf{H}(n)\|^2}(z)$ from (22), it is easy to show that the cumulative distribution function (cdf) of this distribution near zero (small z) is of the form

$$
F_{\|\mathbf{H}(n)\|^2}(z) = \frac{1}{(NK)!}z^{NK} + o(z^{NK}). \tag{27}
$$

Then, we obtain that

$$\begin{aligned}
\mathrm{P}^l_{\text{out}}(\mathrm{d}_{\text{th}}) &= F_{\|\mathbf{H}(n)\|^2}\left(\frac{\sigma_v^2 KN}{P}\Big(\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}\Big)\right) \\
&= \frac{1}{(NK)!}\left(\frac{\sigma_v^2 KN}{P}\Big(\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}\Big)\right)^{NK} \\
&+ o(P^{-NK}) \\
&= \frac{(\sigma_v^2 KN)^{KN}}{(NK)!}\left(\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}\right)^{NK} P^{-NK} \\
&+ o(P^{-NK}) \\
&= \frac{\left(\sigma_v^2 KNP_x\right)^{NK}}{(NK)!}\left(\frac{1}{d_{\text{th}}} - \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}\right)^{NK} \text{SNR}^{-NK} \\
&+ o(\text{SNR}^{-NK}),
\end{aligned} \tag{28}$$

with $\lambda_l(.)$ defined in Lemma 2.

Similarly for $P^u_{\text{out}}(d_{\text{th}})$ and taking $\tilde{P} = \dfrac{P}{\sigma_v^2 KN}$ and $\bar{\lambda} = \dfrac{1}{K}\sum_{l=1}^{K}\lambda_l(C_u)$, we have that

$$\begin{aligned}
&\mathrm{P}^u_{\text{out}}(\mathrm{d}_{\text{th}}) \\
&= \int_0^\infty F_{\|\mathbf{H}(n)\|^2}\left(\frac{1}{\tilde{P}}\left(\frac{1}{d_{\text{th}}} - \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l}\right)\middle|\gamma_{n-1} = z\right) \\
&\times f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&= \int_0^\infty F_{\|\mathbf{H}(n)\|^2}\left(\frac{1}{\tilde{P}}\left(\frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda} + \frac{|\lambda_{\max}(A)|^2\alpha_{\max}}{1+z\alpha_{\max}}}\right)\right) \\
&\times f_{\gamma_{n-1}}(z)\,\mathrm{d}z,
\end{aligned}$$

which by substituting the Taylor series expansion results in

$$\begin{aligned}
\mathrm{P}^u_{\text{out}}(\mathrm{d}_{\text{th}}) &= \frac{1}{(NK)!}\tilde{P}^{-NK}\times \\
&\int_0^\infty \left(\frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda} + \frac{|\lambda_{\max}(A)|^2\alpha_{\max}}{1+z\alpha_{\max}}}\right)^{NK} f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&+ \int_0^\infty o(\tilde{P}^{-NK}) f_{\gamma_{n-1}}(z)\,\mathrm{d}z. \\
&= \frac{(\sigma_v^2 NKP_x)^{NK}}{(NK)!}\text{SNR}^{-NK}\times \\
&\int_0^\infty \left(\frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda} + \frac{|\lambda_{\max}(A)|^2\alpha_{\max}}{1+z\alpha_{\max}}}\right)^{NK} f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&+ \int_0^\infty o(\text{SNR}^{-NK}) f_{\gamma_{n-1}}(z)\,\mathrm{d}z.
\end{aligned} \tag{29}$$

We can now calculate the diversity order for the bounds in order to show their equality and thus prove the diversity result for the outage probability function. For the lower bound, we have that

$$\begin{aligned}
\mathrm{d}^l_{\text{ord}} &= -\lim_{\text{SNR}\to\infty}\frac{\log(P^l_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} \\
&= KN.
\end{aligned} \tag{30}$$

This is due to the fact that when $\text{SNR} \to \infty$, $o(\text{SNR}^{-NK}) \ll \text{SNR}^{-NK}$ and thus the $o(\text{SNR}^{-NK})$ term in (28) vanishes before the first term containing SNR^{-NK}, resulting in a diversity order of NK.

The analysis for the upper bound is also similar. We note that when $\text{SNR} \to \infty$, the term $o(\text{SNR}^{-NK})$ can be upper-bounded by κSNR^{-NK}, with κ being an arbitrary constant. Therefore, it is possible to deduce that

$$\begin{aligned}
&\int_0^\infty o(\text{SNR}^{-NK}) f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&< \int_0^\infty \kappa\text{SNR}^{-NK} f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&= \kappa\text{SNR}^{-NK}\int_0^\infty f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\
&= \kappa\text{SNR}^{-NK}.
\end{aligned}$$

As a result, the second integral term containing $o(\text{SNR}^{-NK})$ in (29) vanishes much faster than SNR^{-NK} and we can say

$$\begin{aligned}
\mathrm{d}^u_{\text{ord}} &= -\lim_{\text{SNR}\to\infty}\frac{\log(P^u_{\text{out}}(d_{\text{th}}))}{\log(\text{SNR})} \\
&= KN.
\end{aligned} \tag{31}$$

As $\mathrm{d}^u_{\text{ord}} = \mathrm{d}^l_{\text{ord}} = NK$, we deduce from Lemma 4 that the diversity order for the outage probability is also equal to NK and the analysis is complete. It is worthwhile mentioning that the maximum diversity order is dependent on the MIMO channel and its achievability only on the space-time code. The source structure does not play any role on the diversity order. However, as we will see next, the source structure plays an important role on the coding gain.

Next, we consider the coding gain for the outage probability. We know that both $P^l_{\text{out}}(d_{\text{th}})$ and $P^u_{\text{out}}(d_{\text{th}})$ have the same diversity order, but possibly different coding gains. We may by the diversity order results deduce that

$$\begin{aligned}
&P^l_{\text{out}}(d_{\text{th}}) < P_{\text{out}}(d_{\text{th}}) < P^u_{\text{out}}(d_{\text{th}}) \\
&(G_1\cdot\text{SNR})^{-KN} + o(\text{SNR}^{-KN}) < (G\cdot\text{SNR})^{-KN} \\
&\qquad + o(\text{SNR}^{-KN}) \\
&\qquad < (G_2\cdot\text{SNR})^{-KN} \\
&\qquad + o(\text{SNR}^{-KN})
\end{aligned} \tag{32}$$

This results in

$$\begin{aligned}
G_1^{-NK} + o(1) &< G^{-NK} + o(1) \\
&< G_2^{-NK} + o(1).
\end{aligned} \tag{33}$$

In the high SNR regime, the term $o(1)$ vanishes quickly compared to the constants G_1, G, G_2. As a result, the relationship in (33) simplifies to

$$G_2 < G < G_1, \quad \text{SNR} \to \infty, \tag{34}$$

which provides upper and lower bounds for the coding gain by setting $G^l = G_2$ and $G^u = G_1$. Note that a higher coding gain means a better SNR performance, i.e. a lower outage probability for a given SNR. That is the reason we obtain the upper bound for coding gain from the lower bound on the outage probability and vice versa. The constants G^u and G^l may in turn be extracted from (28) and (29) as

$$G^u = \frac{(NK)!^{1/(NK)}}{(\sigma_v^2 KNP_x)} \left(\frac{1}{d_{\text{th}}} - \frac{1}{K} \sum_{l=1}^{K} \frac{1}{\lambda_l(C_u)} \right)^{-1} \tag{35}$$

and

$$G_1^l = \frac{(NK)!^{1/(NK)}}{(\sigma_v^2 NKP_x)} \times \left(\int_0^\infty \left(\frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda} + \frac{|\lambda_{\max}(A)|^2 \alpha_{\max}}{1 + z\alpha_{\max}}} \right)^{NK} f_{\gamma_{n-1}}(z)\, \mathrm{d}z \right)^{-1/(NK)}. \tag{36}$$

Remark 1. The value of G^l introduced in (36) (subscripted by 1) is computationally more demanding to calculate than the value for G^u, especially because one should also consider the limit behavior of $f_{\gamma_{n-1}(z)}$ when $P \to \infty$. It is relatively easy to show that G_1^l itself may be lower bounded by the following value

$$G_2^l = \frac{(NK)!^{1/(NK)}}{(\sigma_v^2 NKP_x)} \left(\frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda} + |\lambda_{\max}(A)|^2 \alpha_{\max}} \right)^{-1}, \tag{37}$$

which is less accurate, but is of a much simpler form than G_1^l.

As we can see from (35), (36), and (37), the coding gain depends on the source structure and the threshold. It is the eigenvalues of A and C_u which play a significant role. We observe e.g. that smaller d_{th} leads to smaller coding gain. This is due to the fact that lower thresholds lead to higher outage probabilities and for fixed diversity order, this leads to smaller coding gains. Also, if the eigenvalues of C_u are large, the coding gain decreases, i.e. the outage probabilities increase in the asymptotic limit. Heuristically, such values imply more randomness in the process, resulting in higher distortion for the Kalman filter and consequently higher outage value.

Remark 2. There are limits for d_{th} for which (35), (36) and (37) are valid. In (35), it is required that

$$\frac{1}{d_{\text{th}}} - \frac{1}{K} \sum_{l=1}^{K} \frac{1}{\lambda_l(C_u)} \geqslant 0 \tag{38}$$

so that the lower bound is meaningful. This leads to

$$d_{\text{th}} \leqslant \frac{K}{\sum_{l=1}^{K} \frac{1}{\lambda_l(C_u)}}, \tag{39}$$

which equals the harmonic mean of the eigenvalues of C_u. Similarly, it is also possible to show that a sufficient condition for (36) and (37) to be valid is that

$$d_{\text{th}} \leqslant \frac{1}{K} \sum_{l=1}^{K} \lambda_l(C_u). \tag{40}$$

Therefore, a sufficient condition on d_{th} in order to get valid bounds can be obtain from

$$d_{\text{th}} \leqslant \min \left\{ \frac{K}{\sum_{l=1}^{K} \frac{1}{\lambda_l(C_u)}}, \frac{1}{K} \sum_{l=1}^{K} \lambda_l(C_u) \right\}, \tag{41}$$

which basically requires that d_{th} is smaller than the minimum of mean and harmonic mean of the eigenvalues of the process noise covariance matrix. The limiting regime in both cases is when the eigenvalues of C_u are small. This happens when the randomness in the process from $\mathbf{u}(n)$ is too slow compared to the process memory from A. In realistic applications, this can be solved by adjusting the sampling rate of the original continuous-time process, if necessary. One might argue that lowering the sampling rate in very slow varying processes in order to get better bounds would eventually increase the outage probability. However, if the application is critically sensitive in that regard, the interesting regime is already small d_{th}, because it is the regime which results in higher outage probabilities. For small d_{th} however, the bounds would be functioning. This shows that the limiting behavior in (41) is not a serious issue for the bounds in most practical cases.

Remark 3. For well-conditioned C_u, the bounds perform better than the case when C_u is ill-conditioned. In fact, for the case when $C_u = \sigma_u^2 I$, the bounds are tight. This is elaborated more in Appendix C.

For the numerical evaluation of the accuracy of the coding gain expressions and related discussions, please see Sec. V.

V. Numerical Evaluation of the Bounds and Diversity Results

In this section, we provide simulation results to accompany the presented theory in the previous sections. We begin the numerical evaluations with the following system parameters. We take $K = 2$ and $N = 1$ to keep the simulated outage values practically calculable. This necessitates a maximum diversity order of $KN = 2$. We select

$$A = \begin{bmatrix} 0.6 & -0.8 \\ 0.7 & 0.6 \end{bmatrix}$$

has the eigenvalues $\{0.6 \pm j\sqrt{0.56}\}$. This corresponds to the case where $x_1(n)$ and $x_2(n)$ are relatively highly cross-correlated in time. We select $\sigma_v^2 = 1$ and $d_{\text{th}} = 0.1$. As for the orthogonal space-time code we use the Alamouti code from [20], while for simplicity we calculate P_x from simulations. We consider two cases for C_u, namely C_u^1 and C_u^2 as follows

$$C_u^1 = \begin{bmatrix} 0.25 & 0 \\ 0 & 1.44 \end{bmatrix}, \quad C_u^2 = \begin{bmatrix} 0.53 & 0.28 \\ 0.28 & 0.53 \end{bmatrix}.$$

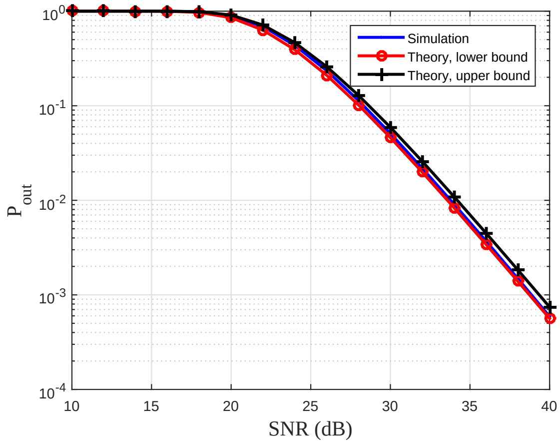

Fig. 2. Outage probability and the corresponding bounds for $K = 2$, $N = 1$, $d_{\text{th}} = 0.1$, $\sigma_v^2 = 1$, and for $\lambda_{1,2}(C_{u,1}) = 0.25, 1.44$.

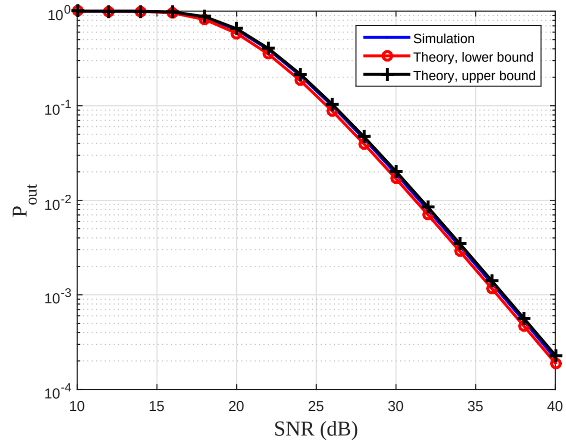

Fig. 3. Outage probability and the corresponding bounds for $K = 2$, $N = 1$, $d_{\text{th}} = 0.1$, $\sigma_v^2 = 1$, and for $\lambda_{1,2}(C_{u,2}) = 0.25, 0.81$.

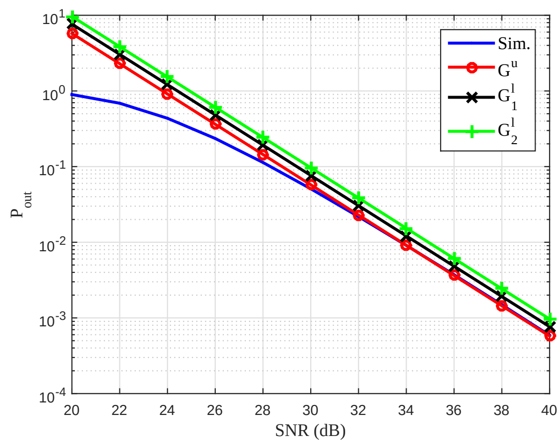

Fig. 4. Comparison of accuracy of the coding gain bounds for $K = 2$, $N = 1$, $d_{\text{th}} = 0.1$, $\sigma_v^2 = 1$, and for $\lambda_{1,2}(C_{u,1}) = 0.25, 1.44$.

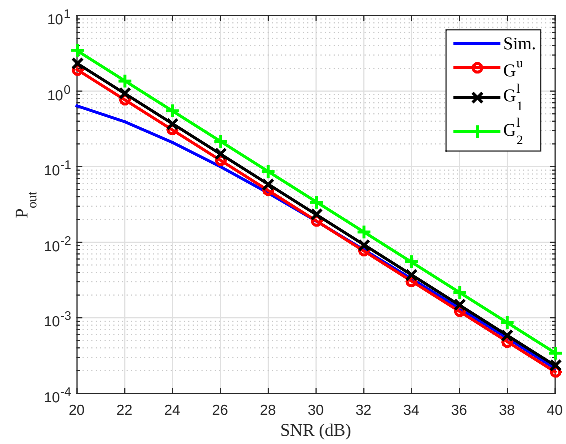

Fig. 5. Comparison of accuracy of the coding gain bounds for $K = 2$, $N = 1$, $d_{\text{th}} = 0.1$, $\sigma_v^2 = 1$, and for $\lambda_{1,2}(C_{u,2}) = 0.25, 0.81$.

The choice is mainly to show how the accuracy of the bounds will change as we use different values for C_u, and also that different values for C_u will result in different coding gains. Also, the eigenvalues of C_u^1 are equal $\{1.44, 0.25\}$ and the eigenvalues of C_u^2 are equal to $\{0.81, 0.25\}$. The P_{out} vs. SNR graph is depicted in Figures 2 and 3. We simulate the Kalman filter for $n \leqslant 10^7$ and numerically calculate the outage probabilities after discarding the first 400 samples. The upper and lower bounds are visibly good for both cases and seem to be quite accurate for a large range of SNR values, compared to the simulated result from the Kalman filter. The numerical evaluation for the coding gain bounds is also depicted in Figures 4 and 5. We note we plot the simulated outage probabilities along with a linear function with a slope of 2 and with calculated values for G^l, G_1^u, G_2^u. The upper and lower bounds for the coding gain become visibly accurate from SNR's close to 30 dB. This shows that for the high SNR analysis to be correct, one needs at least an SNR of the same value or higher. A slope of 2 corresponding to the diversity order div $= 2$ is quite visible in both cases as well. One can also notice in Figures 4 and 5 that the lower bound for coding gain is more accurate than the upper bound and that G_1^l is a much better lower bound that G_2^l, as expected.

In order to observe the performance of the system for higher dimensions, we now take $K = 3$, $N = 1$ and use the space-time code construction from [15, Eq. 39]. This leads to $N_c = 4$ and $r = 3/4$ as well. We also modify other system parameters to $C_u = \text{diag}\{0.5, 0.75, 0.65\}$ and $A = D\,\text{diag}\{0.95, 0.9, 0.98\}\,D^{-1}$, with D representing the normalized discrete cosine transform matrix, and then simulate the system for $n \leqslant 10^8$. The results for outage probability bounds and the coding gain bounds and the diversity order are presented in Figures 6 and 7. A diversity order of 3 is visible in both figure and the bounds are visibly very accurate.

VI. Conclusion

In this paper, we propose a new method for analog transmission of Gauss-Markov sources over MIMO fading channels, which incorporates the use of complex orthogonal space-time codes. By decoding the real and imaginary parts of the code separately, we allow any complex orthogonal space-time code with arbitrary rate to be used for analog transmission. We then show that for Rayleigh fading channels, the distortion outage probability can achieve the maximum diversity order allowed by the number of antennas. By considering process memory only limited to two previous steps, we are able to provide bounds for the distortion outage probability which are applicable for any SNR, and also present bounds for the coding gain in the high SNR regime. We also outline how the coding gain depends on the eigenvalues of the state transition and the process noise covariance matrices, and the outage threshold.

Appendix A
Orthogonality for the Equivalent Real Channel Matrix

In this section, we denote the elements of the arbitrary matrix W as $W[i,j]$, where i denotes the row position and j the column position. The i-th row of W from columns j_1 to j_2 is denoted by $W[i, j_1 : j_2]$ and the whole row is denoted by $W[i,:]$. Similar rules hold for the j-th column.

We first define $T_l = T[2(l-1)K+1 : 2lK, 1 : 2K]$, $l = 1, 2, \dots, N_c$, i.e.

$$T = [T_1^T | T_2^T | \cdots | T_{N_c}^T]^T. \tag{42}$$

Note that each block T_l then maps the variable $\mathbf{x}_r$ to the l–th column of the matrix X_r, i.e. $X_r[:,l] = T_l \mathbf{x}_r$. In order to show that H_{eq} can in fact be orthogonalized, we first see that

$$\begin{aligned}\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}} \mathbf{x}_r &= \mathbf{x}_r^T T^T (I_{N_c} \otimes \tilde{H})^T (I_{N_c} \otimes \tilde{H}) T \mathbf{x}_r \\ &= \mathbf{x}_r^T T^T \left(I_{N_c} \otimes \left(\tilde{H}^T \tilde{H} \right) \right) T \mathbf{x}_r \\ &= \sum_{l=1}^{N_c} \mathbf{x}_r^T T_l^T \left(\tilde{H}^T \tilde{H} \right) T_l \mathbf{x}_r \end{aligned} \tag{43}$$

The matrix $\tilde{H}^T \tilde{H}$ is of dimension $2K \times 2K$, whereas the matrix $\tilde{H}$ is of dimension $2 \times 2K$. For that reason, the rank of $\tilde{H}^T \tilde{H}$ is equal to the rank of $\tilde{H}$, which for i.i.d. Rayleigh fading is equal to 2 with probability one. In addition, the eigenvalues of $\tilde{H}^T \tilde{H}$ are the same as those of $\tilde{H} \tilde{H}^T$, with additional $2K-2$ zeros. It is easy to see that

$$\tilde{H} \tilde{H}^T = \|\mathbf{h}_k\|^2 I_2. \tag{44}$$

Therefore, we can write the eigenvalue decomposition of $\tilde{H}^T \tilde{H}$ as

$$\begin{aligned}\tilde{H}^T \tilde{H} &= Q \,\text{diag}\{0, 0, \dots, 0, \|\mathbf{h}_k\|^2, \|\mathbf{h}_k\|^2\} Q^T \\ &= \|\mathbf{h}_k\|^2 Q Z Q^T, \end{aligned} \tag{45}$$

with $QQ^T = I_{2K}$ and $Z = \text{diag}\{0, 0, \dots, 0, 1, 1\}$. Note that the position of ones in Z is only for the simplification of the proof, as the equivalent matrix QZQ^T will be the same for any positioning. Inserting (45) into (43), we obtain that

$$\begin{aligned}\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}} \mathbf{x}_r &= \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T T_l^T Q Z Q^T T \mathbf{x}_r \\ &= \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T (Q^T T_l)^T Z (Q^T T_l) \mathbf{x}_r \\ &= \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T T_l^{'^T} Z T_l^{'} \mathbf{x}_r \\ &\overset{(a)}{=} \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T (T_l^{'^T} Z)(Z T_l^{'}) \mathbf{x}_r \end{aligned} \tag{46}$$

where $T_l^{'}, l = 1, 2, \dots, N_c$ may be assumed to be the building blocks of a matrix $T^{'}$(same as for T), which maps $\mathbf{x}_r$ to the matrix $Q^T X_r$ in the same way as T maps $\mathbf{x}_r$ to X_r. In addition, (a) holds because $Z^2 = Z$. The positioning of the ones in Z is such that it is only the last two columns of $T_l^{'^T}$ which remain non-zero after multiplication by Z. The last two columns of $T_l^{'^T}$ correspond to the last two rows of $T_l^{'}$ (also visible in the structure of $ZT_l^{'}$ in (46)). In order to provide better intuition into (46), we see that (46) can be rewritten as

$$\begin{aligned}\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}} \mathbf{x}_r &= \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T T_l^{'^T} (Z_{2K-1} + Z_{2K}) T_l^{'} \mathbf{x}_r \\ &= \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T T_l^{'^T} Z_{2K-1} T_l^{'} \mathbf{x}_r \\ &\quad + \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T T_l^{'^T} Z_{2K} T_l^{'} \mathbf{x}_r, \end{aligned} \tag{47}$$

where Z_{2K-1} and Z_{2K} are all-zero matrices, except for $Z_{2K-1}[2K-1, 2K-1] = Z_{2K}[2K, 2K] = 1$. We can then rewrite (47) as

$$\begin{aligned}\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}} \mathbf{x}_r &= \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T (T_l^{'^T} Z_{2K-1})(Z_{2K-1} T_l^{'}) \mathbf{x}_r \\ &\quad + \|\mathbf{h}_k\|^2 \sum_{l=1}^{N_c} \mathbf{x}_r^T (T_l^{'^T} Z_{2K})(Z_{2K} T_l^{'}) \mathbf{x}_r. \end{aligned} \tag{48}$$

In that sense the result of (48) and equivalently (46)is basically the sum of the squares of the last two rows (row $2K-1$ and $2K$) of $X_r^{'} = Q^T X_r$. This can be written as

$$\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}} \mathbf{x}_r = \|\mathbf{h}_k\|^2 \left(\sum_{l=1}^{N_c} X_r^{'^2}[2K-1, l] + X_r^{'^2}[2K, l] \right). \tag{49}$$

It is easy to see that $X_r^{'} = Q_r X_r$, where Q_r is a matrix of dimension $2 \times 2K$, which consists of the last two rows of the matrix Q^T (transpose of the last two columns of Q). From the definition of the eigenvalue decomposition for $\tilde{H}^T \tilde{H}$ and given that $\tilde{H}^T \tilde{H}$ only has two non-zero eigenvalues (with the corresponding eigenvectors as the last two columns of Q, equal to Q_r^T), it is readily established that

$$\tilde{H}^T \tilde{H} = \|\mathbf{h}_k\|^2 Q_r^T Q_r, \tag{50}$$

which basically means that $\tilde{H} = \pm\|\mathbf{h}_k\|Q_r$, i.e. Q_r is a normalized version of $\tilde{H}$. This is a property we will use later on in the proof.

In order to be able to calculate the value of the sum in (49), we may convert the matrices X_r and Q_r to complex equivalent matrices. In that case, we are able to finally use the orthogonality of the complex space-time code in order to prove the orthogonality of the equivalent channel. This step is necessary, as the code's orthogonality is best described in the domain of the complex numbers. We perform this procedure as follows. For X_r, take all the consecutive odd and even real rows, add the even row multiplied by $\boldsymbol{i}$ to the previous odd row and then remove the even rows. The resulting matrix, of dimension $K \times N_c$ is basically equal to X, the original space-time code. For the matrix Q_r, perform the same procedure with odd and even columns, and call the resulting matrix, of dimension $2 \times K$ as Q_c. We further take the first row of the matrix Q_c to be equal to $\mathbf{q}_1^T$ and the second row as $\mathbf{q}_2^T$. Note that $\mathbf{q}_1$ and $\mathbf{q}_2$ are still orthogonal to one another and we still have that $\mathbf{q}_1^H\mathbf{q}_1 = \mathbf{q}_2^H\mathbf{q}_2 = 1$. With these definitions, the sum in (49) can be rewritten as

$$\begin{aligned}&\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}}\mathbf{x}_r\\&= \|\mathbf{h}_k\|^2 \left(\operatorname{Re}(\mathbf{q}_1^H X)\operatorname{Re}(\mathbf{q}_1^H X)^T + \operatorname{Re}(\mathbf{q}_2^H X)\operatorname{Re}(\mathbf{q}_2^H X)^T\right)\end{aligned}$$

or equivalently

$$\begin{aligned}&\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}}\mathbf{x}_r\\&= \|\mathbf{h}_k\|^2 \left(\operatorname{Re}(\mathbf{q}_1^H X)\operatorname{Re}(X^H\mathbf{q}_1) + \operatorname{Re}(\mathbf{q}_2^H X)\operatorname{Re}(X^H\mathbf{q}_2)\right).\end{aligned} \tag{51}$$

The following equalities are also easy to verify

$$\begin{aligned}\|\mathbf{x}\|^2 &= \mathbf{q}_2^H X X^H \mathbf{q}_2\\&= \mathbf{q}_1^H X X^H \mathbf{q}_1\\&= \operatorname{Re}(\mathbf{q}_1^H X)\operatorname{Re}(X^H\mathbf{q}_1) - \operatorname{Im}(\mathbf{q}_1^H X)\operatorname{Im}(X^H\mathbf{q}_1)\\&= \operatorname{Re}(\mathbf{q}_1^H X)\operatorname{Re}(X^H\mathbf{q}_1) + \operatorname{Im}(\mathbf{q}_1^H X)\operatorname{Im}(\mathbf{q}_1^H X)^T.\end{aligned} \tag{52}$$

If we convert the matrix $\tilde{H}$ to a complex equivalent matrix in the same manner as for Q_r and obtain $\tilde{H}_c$, we see that due to the structure of $\tilde{H}$, we should have

$$\operatorname{Re}(\tilde{H}_c[1,:]) = \operatorname{Im}(\tilde{H}_c[2,:])$$

and

$$\operatorname{Re}(\tilde{H}_c[2,:]) = -\operatorname{Im}(\tilde{H}_c[1,:])$$

As Q_r is only a scaled version of $\tilde{H}$, then Q_c is only a scaled (by a real number) version of $\tilde{H}_c$. Due to that, we should also have for Q_c that

$$\operatorname{Re}(Q_c[1,:]) = \operatorname{Im}(Q_c[2,:]) \tag{53}$$

and

$$\operatorname{Re}(Q_c[2,:]) = -\operatorname{Im}(Q_c[1,:]). \tag{54}$$

It is then straightforward to show using (53) and (54) to that

$$\operatorname{Im}(\mathbf{q}_1^H X) = \operatorname{Re}(\mathbf{q}_2^H X). \tag{55}$$

That can be understood better by considering that if $Q_r[1,:] = [q_{1,1}, q_{1,2}, \ldots, q_{1,2K-1}, q_{1,2K}]^T$ and $Q_r[2,:] = [q_{2,1}, q_{2,2}, \ldots, q_{2,2K-1}, q_{2,2K}]^T$, then we have from (53) and (54) that e.g. $q_{1,1} = q_{2,1}$ and $q_{1,2} = -q_{2,1}$ and so on. This can be considered within the multiplication operations of $\mathbf{q}_1^H X$ and $\mathbf{q}_2^H X$, in order to produce the result in (55). Inserting (55) into (52) results in

$$\|\mathbf{x}\|^2 = \operatorname{Re}(\mathbf{q}_1^H X)\operatorname{Re}(X^H\mathbf{q}_1) + \operatorname{Re}(\mathbf{q}_2^H X)\operatorname{Re}(X^H\mathbf{q}_2), \tag{56}$$

which by comparing to (51) confirms that

$$\begin{aligned}\mathbf{x}_r^T H_{\text{eq}}^T H_{\text{eq}}\mathbf{x}_r &= \|\mathbf{h}_k\|^2\|\mathbf{x}\|^2\\&= \|\mathbf{h}_k\|^2\mathbf{x}_r^T\mathbf{x}_r\\&= \mathbf{x}_r^T\|\mathbf{h}_k\|^2\mathbf{x}_r\end{aligned} \tag{57}$$

for all $\mathbf{x}_r$, which effectively results in

$$H_{\text{eq}}^T H_{\text{eq}} = \|\mathbf{h}_k\|^2 I_{2K} \tag{58}$$

APPENDIX B
BOUNDS FOR THE INSTANTANEOUS DISTORTION

From (15), we have that $M(n) = \frac{1}{\gamma_n}\left(\frac{1}{\gamma_n}P^{-1}(n) + I\right)^{-1}$. If we denote the eigenvalues $M(n)$ by $\lambda_l(M(n))$, $l = 1, 2, \ldots, K$, and the eigenvalues of $P(n)$ by $\lambda_l(P(n))$, $l = 1, 2, \ldots, K$, then we have that

$$\begin{aligned}d(n) &= \frac{1}{K}\text{tr}(M(n))\\&= \frac{1}{K}\sum_{l=1}^{K}\lambda_l(M(n))\\&\overset{(a)}{=} \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\gamma_n}\left(\frac{1}{\gamma_n}\frac{1}{\lambda_l(P(n))} + 1\right)^{-1}\\&= \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\frac{1}{\lambda_l(P(n))} + \gamma_n},\end{aligned} \tag{59}$$

where (a) holds because the eigenvalues of sum of an arbitrary matrix and the identity matrix are equal to the sum of the eigenvalues of that matrix and the identity matrix (this does not hold in general for sum of two arbitrary matrices). As $P(n)$ is a covariance matrix, it is (semi)-positive definite. Therefore, we have that $\lambda_l(P(n)) \geqslant 0, \forall l$. If we denote the ordered eigenvalues of C_u by $\lambda_l(C_u), \forall l$ ($\lambda_1(C_u) \geqslant \lambda_2(C_u) \geqslant \ldots \geqslant \lambda_K(C_u)$) and also order $\lambda_l(P(n))$ such that $\lambda_1(P(n)) \geqslant \lambda_2(P(n)) \geqslant \ldots \geqslant \lambda_K(P(n))$, we know from Weyl's inequalities [21, Ch. 3] that

$$\lambda_l(P(n)) \geqslant \lambda_l(C_u). \tag{60}$$

This is due to the fact that $P(n) = AM(n-1)A^T + C_u$ and $AM(n-1)A^T$ is a positive definite matrix (because $M(n-1)$

is positive-definite). Now combining (59) and (60), we obtain that

$$d(n) = \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\frac{1}{\lambda_l(P(n))}+\gamma_n} \geqslant \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\frac{1}{\lambda_l(C_u)}+\gamma_n}.$$

It is easy to show that the function $f(z)=\frac{1}{z+c}$ is convex in z for any positive c. Now, invoking Jensen's inequality from [22, Ch. 2.6] leads to

$$\sum_l p_l f(z_l) \geqslant f\left(\sum_l p_l z_l\right). \tag{61}$$

Assuming then $p_l = 1/K$ and $z_l = 1/\lambda_l(C_u)$, we would have that

$$\frac{1}{K}\sum_{l=1}^{K}\frac{1}{\frac{1}{\lambda_l(C_u)}+\gamma_n} = \sum_{l=1}^{K}\frac{1}{K}\frac{1}{\frac{1}{\lambda_l(C_u)}+\gamma_n} \geqslant \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}+\gamma_n}, \tag{62}$$

which establishes the lower bound $d^l(n)$ as stated in Lemma 2 as

$$d^l(n) = \frac{1}{\frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}+\gamma_n}. \tag{63}$$

For the upper bound, if we manage to find a series of random variables which are greater than or equal to $\lambda_l(P(n))$, we can then obtain the upper bound $d^u(n)$ in the same manner as we found $d^l(n)$. $\lambda_l(P(n))$ are functions of all filter memory, and it is a cumbersome task to track the filter memory. Instead, we decide to consider only the two previous time steps, i.e. $n-1$ and $n-2$, and show that we are able to find reasonably good bounds.

First, we consider $n-1$. Given that $P(n) = AM(n-1)A^T + C_u$, it is possible to obtain an upper bound on $\lambda_l(P(n))$ based on $\lambda_l(M(n-1))$. Based on Weyl's theorem on eigenvalues of sum of positive definite Hermitian symmetric matrices, we can state that

$$\lambda_l(P(n)) \leqslant \lambda_{\max}(AM(n-1)A^T) + \lambda_l(C_u). \tag{64}$$

It is also easy to show from Weyl's inequalities [21, Ch. 3] that for two symmetric matrices A and B, we have that

$$\lambda_{\max}(AB) \leqslant \lambda_{\max}(A)\lambda_{\max}(B). \tag{65}$$

Based on (65), we may extend (64) to

$$\begin{aligned}\lambda_l(P(n)) &\leqslant \lambda_{\max}(A^T)\lambda_{\max}(AM(n-1)) + \lambda_l(C_u)\\ &\leqslant \lambda_{\max}(A^T)\lambda_{\max}(A)\lambda_{\max}(M(n-1)) + \lambda_l(C_u)\\ &\leqslant |\lambda_{\max}(A)|^2\lambda_{\max}(M(n-1)) + \lambda_l(C_u).\end{aligned} \tag{66}$$

The next step is to find an upper bound for $\lambda_{\max}(M(n-1))$. We know that

$$\lambda_l(M(n-1)) = \frac{1}{\frac{1}{\lambda_l(P(n-1))}+\gamma_{n-1}}. \tag{67}$$

From that we conclude that

$$\lambda_{\max}(M(n-1)) = \frac{1}{\frac{1}{\lambda_{\max}(P(n-1))}+\gamma_{n-1}}. \tag{68}$$

The next step is to find an upper bound for $\lambda_{\max}(P(n-1))$. Now, if we consider one more time step backwards, i.e. $n-2$, we know that $P(n-1) = AM(n-2)A^T + C_u$. Therefore, we have as before that

$$\lambda_{\max}(P(n-1)) \leqslant |\lambda_{\max}(A)|^2\lambda_{\max}(M(n-2)) + \lambda_{\max}(C_u). \tag{69}$$

In order to get an upper bound for $\lambda_{\max}(M(n-2))$, we assume the worst case scenario for $M(n-2)$. Obviously $M(n-2)$ cannot be worse that C_x. That happens when $h(n') = 0, \forall n' < n-2$. As a result, an upper bound for $\lambda_{\max}(M(n-2))$ is $\lambda_{\max}(C_x)$.

$$\lambda_{\max}(P(n-1)) \leqslant |\lambda_{\max}(A)|^2\lambda_{\max}(C_x) + \lambda_{\max}(C_u). \tag{70}$$

So far, we have proven that

$$d(n) \leqslant \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\frac{1}{\theta_l}+\gamma_n} \tag{71}$$

with

$$\theta_l = \lambda_l(C_u) + \frac{|\lambda_{\max}(A)|^2}{\frac{1}{\alpha_{\max}}+\gamma_{n-1}}, \tag{72}$$

and

$$\alpha_{\max} = |\lambda_{\max}(A)|^2\lambda_{\max}(C_x) + \lambda_{\max}(C_u). \tag{73}$$

Now, similar to the approach used to establish the lower bound, we may consider the function $f(z) = \frac{1}{1/z+c}$ for arbitrary positive c. It is easy to show that $f(z)$ is a concave function. Again invoking the Jensen's inequality, we may say that

$$\sum_l p_l f(z_l) \leqslant f\left(\sum_l p_l z_l\right). \tag{74}$$

Assuming then $p_l = 1/K$ and $z_l = \theta_l$, we would have that

$$\frac{1}{K}\sum_{l=1}^{K}\frac{1}{\frac{1}{\theta_l}+\gamma_n} = \sum_{l=1}^{K}\frac{1}{K}\frac{1}{\frac{1}{\theta_l}+\gamma_n} \leqslant \frac{1}{\frac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l}+\gamma_n}, \tag{75}$$

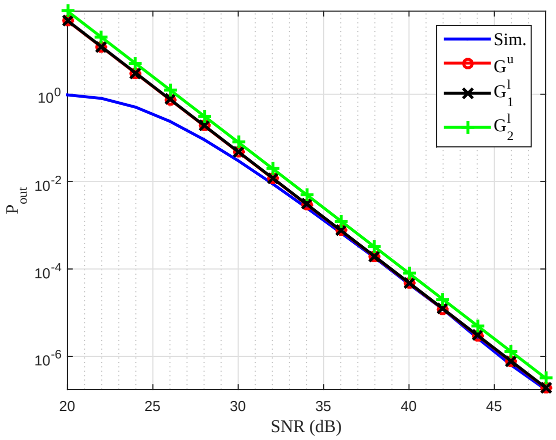

Fig. 6. Comparison of accuracy of the coding gain bounds for $K = 3$, $N = 1$, $d_{\text{th}} = 0.1$, $\sigma_v^2 = 1$.

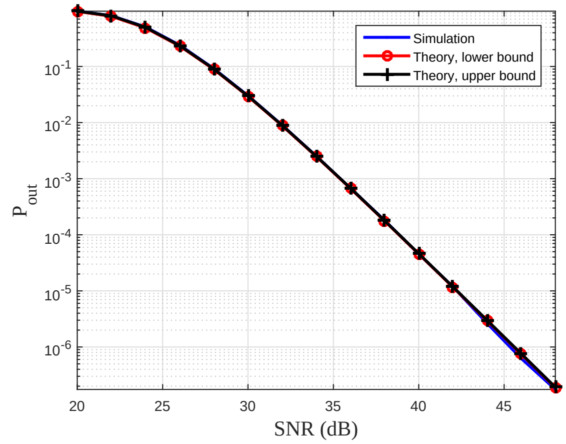

Fig. 7. Outage probability and the corresponding bounds for $K = 3$, $N = 1$, $d_{\text{th}} = 0.1$, $\sigma_v^2 = 1$.

which finally establishes the upper bound $d^u(n)$ as stated in Lemma 3 as

$$d^u(n) = \frac{1}{\dfrac{1}{\frac{1}{K}\sum_{l=1}^{K}\theta_l} + \gamma_n}, \tag{76}$$

with θ_l defined Lemma 3.

Appendix C
Tightness of the Coding Gain Bounds

In this section, we outline how the gap between G^u and G^l behaves as a function of system parameters and if it is eventually tight. We try to evaluate the terms $\log(G_u/G_2^l)$ and $\log(G^u/G_1^l)$, which correspond to the gap (in dB) in the $\log(\text{SNR})$-scale. For the simple lower bound G_2^l, we have that

$$\log(G^u/G_2^l) = \log\left(\frac{\dfrac{1}{d_{\text{th}}} - \dfrac{1}{\bar{\lambda} + |\lambda_{\max}(A)|^2\alpha_{\max}}}{\dfrac{1}{d_{\text{th}}} - \dfrac{1}{K}\sum_{l=1}^{K}\dfrac{1}{\lambda_l(C_u)}}\right). \tag{77}$$

This is a constant gap independent of the average SNR and only a function of system parameters. We take $\left(\frac{1}{\lambda}\right) = \frac{1}{K}\sum_{l=1}^{K}\frac{1}{\lambda_l(C_u)}$. It is possible to show that for $\lambda_l(C_u) > 0$, we have that

$$\frac{1}{d_{\text{th}}} - \left(\frac{1}{\lambda}\right) \leqslant \frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda}} < \frac{1}{d_{\text{th}}} - \frac{1}{\bar{\lambda} + |\lambda_{\max}(A)|^2\alpha_{\max}}. \tag{78}$$

Now for fixed $\bar{\lambda} = 1/K\text{tr}(C_u)$, it is easy to verify that $\log(G^u/G_2^l)$ can be minimized if $\lambda_l(C_u) = \text{const.}$, i.e. the gap is minimized when $C_u = \sigma_u^2 I$, for some $\sigma_u^2 > 0$.

Performing similar analysis for the other lower bound G_1^l, we obtain that

$$\begin{aligned} &NK\log(G^u/G_1^l) = \\ &\lim_{\text{SNR}\to\infty}\log\int_0^\infty\left(\frac{\dfrac{1}{d_{\text{th}}} - \dfrac{1}{\bar{\lambda} + \dfrac{|\lambda_{\max}(A)|^2\alpha_{\max}}{1+z\alpha_{\max}}}}{\dfrac{1}{d_{\text{th}}} - \left(\dfrac{1}{\lambda}\right)}\right)^{NK} f_{\gamma_{n-1}}(z)\,\mathrm{d}z \\ &\leqslant \lim_{\text{SNR}\to\infty}\log\int_0^\infty\left(\frac{\dfrac{1}{d_{\text{th}}} - \dfrac{1}{\bar{\lambda} + |\lambda_{\max}(A)|^2 w}}{\dfrac{1}{d_{\text{th}}} - \left(\dfrac{1}{\lambda}\right)}\right)^{NK} f_{\gamma_{n-1}^{-1}}(w)\,\mathrm{d}w. \end{aligned} \tag{79}$$

It is possible to show (similar to [11, Appendix D]) that when $\text{SNR} \to \infty$, then $f_{\gamma_{n-1}^{-1}}(w) \to \delta(w)$, i.e. Dirac's delta function. Consequently, we will have that

$$\begin{aligned} &\log(G_1^u/G^l) \leqslant \\ &\frac{1}{NK}\lim_{\text{SNR}\to\infty}\log\int_0^\infty\left(\frac{\dfrac{1}{d_{\text{th}}} - \dfrac{1}{\bar{\lambda} + |\lambda_{\max}(A)|^2 w}}{\dfrac{1}{d_{\text{th}}} - \left(\dfrac{1}{\lambda}\right)}\right)^{NK}\delta(w)\,\mathrm{d}w \\ &= \log\left(\frac{\dfrac{1}{d_{\text{th}}} - \dfrac{1}{\bar{\lambda}}}{\dfrac{1}{d_{\text{th}}} - \left(\dfrac{1}{\lambda}\right)}\right), \end{aligned} \tag{80}$$

which is equal to zero iff $\lambda_l(C_u) = \text{const.}$ or equivalently $C_u = \sigma_u^2 I$. The best bounds are thus achieved for well-conditioned C_u.

References

[1] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, "Source-channel diversity for parallel channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3518–3539, 2005.

[2] G. Caire and K. Narayanan, “On the distortion SNR exponent of hybrid digital–analog space–time coding,” *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2867–2878, 2007.

[3] D. Gunduz and E. Erkip, “Source and channel coding for cooperative relaying,” *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3454–3475, 2007.

[4] T. Holliday, A. J. Goldsmith, and H. V. Poor, “Joint source and channel coding for MIMO systems: Is it better to be robust or quick?” *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1393–1405, 2008.

[5] S. Dey, A. S. Leong, and J. S. Evans, “Kalman filtering with faded measurements,” *Automatica*, vol. 45, no. 10, pp. 2223–2233, 2009.

[6] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, “Estimation diversity and energy efficiency in distributed sensing,” *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, 2007.

[7] C.-H. Wang, A. S. Leong, and S. Dey, “Distortion outage minimization and diversity order analysis for coherent multiaccess,” *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6144–6159, 2011.

[8] L. Peng and A. Guillen i Fabregas, “Distortion outage probability in MIMO block-fading channels,” in *IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 2223–2227.

[9] A. S. Leong, S. Dey, G. N. Nair, and P. Sharma, “Power allocation for outage minimization in state estimation over fading channels,” *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3382–3397, 2011.

[10] R. Parseh and K. Kansanen, “On estimation error outage for scalar Gauss–Markov signals sent over fading channels,” *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6225–6234, 2014.

[11] ——, “Diversity effects in Kalman filtering over Rayleigh fading channels,” *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6329–6342, 2015.

[12] S. Kar, B. Sinopoli, and J. M. Moura, “Kalman filtering with intermittent observations: Weak convergence to a stationary distribution,” *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 405–420, 2012.

[13] P. Bougerol, “Almost sure stabilizability and Riccati’s equation of linear systems with random parameters,” *SIAM journal on control and optimization*, vol. 33, no. 3, pp. 702–717, 1995.

[14] ——, “Kalman filtering with random coefficients and contractions,” *SIAM Journal on Control and Optimization*, vol. 31, no. 4, pp. 942–959, 1993.

[15] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[16] J. Chen and D. T. Slock, “Orthogonal space-time block codes for analog channel feedback,” in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2008, pp. 1473–1477.

[17] Z. Wang and G. B. Giannakis, “A simple and general parameterization quantifying performance in fading channels,” *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1398, 2003.

[18] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall New Jersey, 2000, vol. 1.

[19] D. Williams, *Probability with martingales*. Cambridge university press, 1991.

[20] S. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[21] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge university press, 1991.

[22] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.