Optimal signaling and detector design for *M*-ary communication systems in the presence of multiple additive noise channels ☆

Berkan Dulek [a], Mehmet Emin Tutay [a], Sinan Gezici [a,*], Pramod K. Varshney [b]

[a] *Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey*
[b] *Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244, USA*

ARTICLE INFO

ABSTRACT

An *M*-ary communication system is considered in which the transmitter and the receiver are connected via multiple additive (possibly non-Gaussian) noise channels, any one of which can be utilized for the transmission of a given symbol. Contrary to deterministic signaling (i.e., employing a fixed constellation), a stochastic signaling approach is adopted by treating the signal values transmitted for each information symbol over each channel as random variables. In particular, the joint optimization of the channel switching (i.e., time sharing among different channels) strategy, stochastic signals, and decision rules at the receiver is performed in order to minimize the average probability of error under an average transmit power constraint. It is proved that the solution to this problem involves either one of the following: (i) deterministic signaling over a single channel, (ii) randomizing (time sharing) between two different signal constellations over a single channel, or (iii) switching (time sharing) between two channels with deterministic signaling over each channel. For all cases, the optimal strategies are shown to employ corresponding maximum *a posteriori* probability (MAP) decision rules at the receiver. In addition, sufficient conditions are derived in order to specify whether the proposed strategy can or cannot improve the error performance over the conventional approach, in which a single channel is employed with deterministic signaling at the average power limit. Finally, numerical examples are presented to illustrate the theoretical results.

1. Introduction

In recent studies, the benefits of randomization (time sharing) have been analyzed for various detection problems in an environment of additive and non-varying but otherwise arbitrarily distributed noise [1–16]. In the context of noise enhanced detection, an additive "noise" component that is realized by a randomization between at most two different signal levels can be injected into the input of a suboptimal detector in order to improve its detection performance under a false alarm constraint [1–3]. Similar noise benefits are investigated for detection problems in the Bayesian, minimax, and restricted Bayesian frameworks as well, and it is shown that the optimal additive noise can be characterized by a randomization among a certain number of signal values in each scenario [2,4,5].

Due to the irrelevance theorem of optimal detection [17], it is known that the performance of an optimal receiver cannot be improved if the injected noise is independent of the received signal and the hypotheses. On the other hand, if the signal values transmitted for each information symbol are designed by taking into account the probability density function (PDF) of channel noise, some performance improvement can be obtained even if the receiver is optimal. For example, it is well known that the performance of optimal binary detection in Gaussian noise is improved by selecting deterministic antipodal signals along the eigenvector of the noise covariance matrix corresponding to the minimum eigenvalue [17]. In *stochastic signaling*, a more general approach is adopted by treating the signal values transmitted for each information symbol as random variables, and the optimal signal distribution is obtained by maximizing some performance criterion under certain system constraints [8,9,11,18]. For communication systems that operate over channels with multimodal noise distributions, it is shown in [8] that transmitting a stochastic signal for each symbol instead of a deterministic signal can improve performance of a given receiver in terms of error probability. In particular, it is proved that an optimal stochastic signal can be represented by a randomization of no more than three different signal values under second and fourth moment constraints. In [9], joint optimal design of stochastic
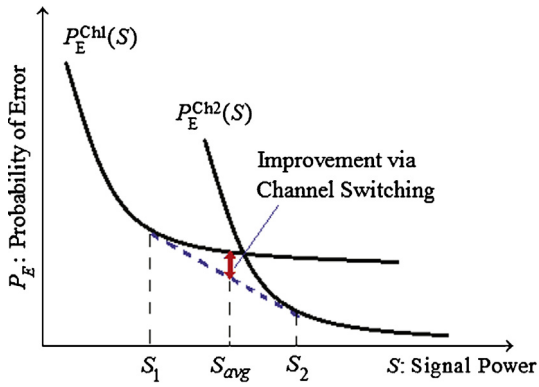
Fig. 1. Illustrative example demonstrating the benefits of switching between two channels under an average power constraint.

signals and a detector is considered under an average transmit power constraint. It is shown that the solution results in a randomization between at most two distinct signal constellations with the corresponding maximum *a posteriori* probability (MAP) detector at the receiver. A similar analysis is conducted under the Neyman–Pearson criterion in [11]. Stochastic signaling in the presence of imperfect channel state information at the transmitter is studied in [18], and various stochastic signal design approaches are proposed for that scenario. In addition, in other studies such as [19–24], time-varying or random signal constellations are utilized in order to enhance error performance or to achieve diversity.

Error performance of some communication systems that operate over an additive time-invariant noise channel can also be improved via detector randomization, which involves the use of multiple detectors at the receiver with certain probabilities [2,3, 12,13,25]. In other words, a receiver can randomize among multiple detectors in order to achieve a lower average probability of error. In [3], an average power constrained binary communication system is considered, and randomization between two antipodal signal pairs and the corresponding MAP detectors is studied. Significant performance improvements are reported as a result of detector randomization in the presence of symmetric Gaussian mixture noise over a range of average power constraint values. In [13], the results in [3] and [9] are generalized by considering an average power constrained *M*-ary communication system that can employ both detector randomization and stochastic signaling over an additive noise channel with some known distribution. It is shown that the joint optimization of the transmitted signals and the detectors at the receiver results in a randomization between at most two MAP detectors corresponding to two deterministic signal constellations. In a related study, the form of the optimal additive noise is determined for variable detectors in the context of noise enhanced detection under both Neyman–Pearson and Bayesian criteria [2].

When multiple channels are available between a transmitter and a receiver, it may be advantageous to perform channel switching; that is, to transmit over one channel for a certain fraction of time, and then switch to another channel during the next transmission period even if the channel statistics are not varying with time [6,26,27]. Fig. 1 illustrates this fact for an average power constrained binary communication system which employs antipodal signaling with $\{-\sqrt{S}, \sqrt{S}\}$ for a given signal power S. It is seen that the average probability of error can be reduced by switching (time sharing) between channel 1 and channel 2 with respective power levels S_1 and S_2 in comparison to the constant power transmission scheme that employs power S_{avg} exclusively over channel 1. More precisely, time sharing exploits the nonconvexity of the plot for the *minimum* of the error probabilities over both channels as a function of the signal power. The resulting strategy yields the convex hull of the individual error probability functions. This observation

is first noted in [6] while studying the convexity properties of error probability with respect to the transmit signal power for the optimal detection of *antipodal* signals corrupted by additive *unimodal* noise. It is shown that the optimum performance under an average power constraint can be achieved by time sharing between at most two channels and power levels.

In this manuscript, we study the optimal channel switching, signaling and detection strategy that minimizes the average probability of error for an average power constrained *M*-ary communication system in which the transmitter and the receiver are connected via multiple additive noise channels. Although the channel switching problem is treated in some studies, such as [6], for unimodal noise distributions and deterministic binary antipodal signals, no previous work has considered this problem for generic noise PDFs (i.e., including non-Gaussian or multimodal cases) and in the presence of stochastic signaling (i.e., when the transmitter can perform signal randomization for each information symbol sent over any one of the channels) for *M*-ary communication systems. More specifically, we investigate the *joint* optimization of the channel switching strategy, stochastic signals (employed for the transmission of each symbol over each channel), and decision rules (used for each channel at the receiver) in order to minimize the average probability of error under an average transmit power constraint.

The main contributions of this study can be summarized as follows:

- A novel problem formulation is proposed for the optimal signaling and detection problem in the presence of multiple additive noise channels by considering the *joint* optimization of the channel switching strategy, stochastic signals, and detectors without imposing any restrictions except the continuity of the probability distributions of the channel noise.
- It is proved that the solution to this generic problem corresponds to either (i) deterministic signaling (i.e., employing a fixed constellation) over a single channel with the corresponding MAP detector, (ii) randomizing (time sharing) between two different signal constellations over a single channel with the corresponding MAP detector, or (iii) switching (time sharing) between the MAP detectors of two channels with deterministic signaling over each channel.
- Various sufficient conditions are derived in order to specify whether or not the proposed channel switching strategy can improve the error performance over the conventional approach, in which a single channel is employed with deterministic signaling at the average power limit.

In addition, numerical examples are provided to illustrate the improvements that can be achieved via the optimal signaling and detection strategy. The results in this manuscript generalize some of the previous studies in the literature and cover them as special cases. For example, in the absence of channel switching (i.e., in the presence of a single channel between the transmitter and the receiver) and for binary communications, the results reduce to those in [9]. In addition, in the absence of stochastic signaling and when the channel noise is assumed to have a unimodal differential PDF for a binary communication system, the problem considered in this study covers the one in [6] as a special case.

In a recent conference paper [28], we have presented the optimal channel switching, signaling and detection problem, and provided its solution. The current paper presents a more detailed derivation of this solution. In addition, a number of sufficient conditions, which are presented in Propositions 2–6, are obtained for the improvability and non-improvability of the correct decision performance via stochastic signaling or channel switching over a fixed power transmission scheme that employs MAP detection
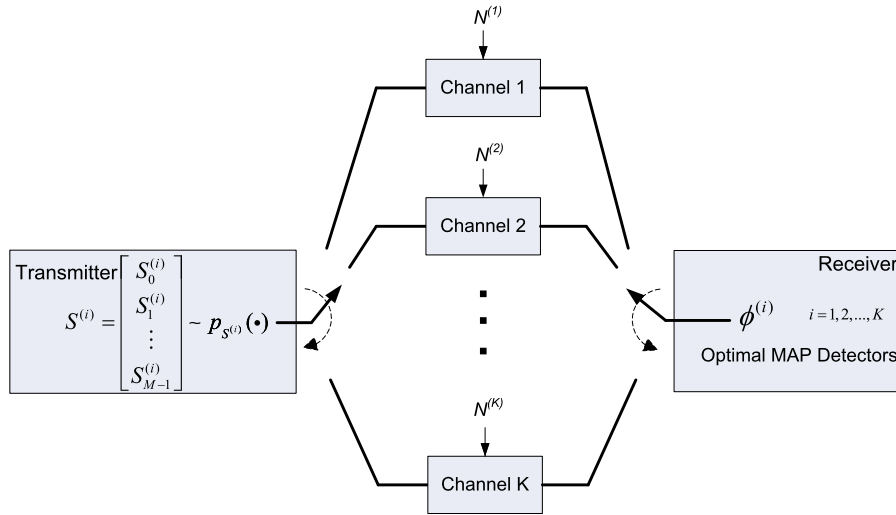
Fig. 2. M-ary communication system that employs stochastic signaling and channel switching.

using the most favorable channel. Since a set of possibly nonconvex optimization problems has to solved in order to obtain the optimal signaling strategy, these conditions can be checked beforehand to determine whether an improvement via stochastic signaling or channel switching is even possible. Numerical examples are also presented to corroborate these results. More specifically, both distinct and identical noise channels are considered, and various performance graphs are presented to explain the benefits of stochastic signaling and channel switching.

The remainder of the manuscript is organized as follows. In Section 2, the optimal signaling and detection problem is formulated in the presence of multiple additive noise channels under an average transmit power constraint, and the form of the solution to this optimization problem is obtained. In Section 3, improvability and non-improvability conditions are provided in order to specify when the proposed channel switching strategy can improve performance over the conventional approach. Numerical examples are presented in Section 4, which is followed by some concluding remarks in Section 5.

2. Stochastic signaling and channel switching

Consider an M-ary communication system, in which the information can be conveyed from the transmitter to the receiver over K additive non-varying and independent noise channels as illustrated in Fig. 2. The transmitter is allowed to switch or time share among these K channels to improve the correct decision performance at the receiver. A relay at the transmitter controls access to the channels so that only one of the channels can be used for symbol transmission at any given time. Furthermore, a *stochastic signaling* approach is adopted by treating the signal transmitted from each channel for each information symbol as a random vector instead of a constant value [8,13]. In other words, the transmitter can perform randomization of signal values for each information symbol, which also corresponds to a form of constellation randomization [9,19,20]. The transmitter and the receiver are assumed to be synchronized so that the receiver knows which channel is currently in use, and employs the optimal decision rule for the corresponding channel and the stochastic signaling scheme. In practice, this assumption can be realized by employing a communications protocol that allocates the first $N_{s,1}$ symbols in the payload for channel 1, the next $N_{s,2}$ symbols in the payload for channel 2, and so on. The information on the number of symbols for different channels can be included in the header of a communications packet [13].

Multiple channels can be available between a transmitter and a receiver, for example, in cognitive radio systems, where secondary users sense the spectrum in order to determine available frequency bands for communications [29,30]. In the presence of multiple available frequency bands between a transmitter-receiver pair in a cognitive radio system (see, e.g., [31]), channel switching can be performed in order to improve the error performance of the secondary system. Therefore, one application of the scenario in Fig. 2 can be the communications of secondary users in a cognitive radio system.

As pointed out in [6], for a binary-valued scalar communication system that employs antipodal signaling and the corresponding optimal MAP detector at the receiver, error probability is a non-increasing convex function of the signal-to-noise ratio (SNR) when the channel has a continuously differentiable unimodal noise PDF with a finite variance. The more general case of arbitrary signal constellations is investigated in [7] by concentrating on the maximum likelihood (ML) detection over additive white Gaussian noise (AWGN) channels. The symbol error rate (SER) is shown to be always convex in SNR for 1-D and 2-D constellations, and also for higher dimensional constellations in high SNR regimes. As a result, it is impossible to improve the error performance of an optimal detector via stochastic signaling under an average transmit power constraint in the above mentioned cases due to the convexity of the error probability. On the other hand, nonconvexity can be observed at low to intermediate SNRs in the presence of multimodal noise and even unimodal (including Gaussian) noise for high dimensional constellations.[1] As an example, it is reported in [8] and [9] that employing stochastic signaling; that is, modeling signals for different symbols as random variables instead of deterministic quantities, can provide significant performance improvement under Gaussian mixture noise. Motivated by this observation, we consider additive noise channels with generic PDFs and aim to obtain the optimal signaling and detection strategy when multiple channels are available for symbol transmission and stochastic signaling can be performed over each channel. In this scenario, the noisy observation vector **Y** received by the detector corresponding to the ith channel can be modeled as follows.

$$\mathbf{Y} = \mathbf{S}_j^{(i)} + \mathbf{N}^{(i)}, \quad j \in \{0, 1, \ldots, M-1\} \text{ and } i \in \{1, \ldots, K\}, \qquad (1)$$

---

[1] Non-Gaussian and multimodal noise distributions are observed in some practical systems due to effects such as interference and jamming [32–34].

where $\mathbf{S}_j^{(i)}$ represents the N-dimensional signal vector transmitted for symbol j over channel i, and $\mathbf{N}^{(i)}$ is the noise in channel i with a continuous PDF $p_{\mathbf{N}^{(i)}}$. $\mathbf{N}^{(i)}$ is assumed to be independent of $\mathbf{S}_j^{(i)}$ and all the noise components of the remaining channels. It should be emphasized that $\mathbf{S}_j^{(i)}$ is modeled as a random vector to employ stochastic signaling. Also, the prior probabilities of the symbols, denoted by $\pi_0, \pi_1, \ldots, \pi_{M-1}$, are assumed to be known. The vector channel model given above provides the discrete-time equivalent representation of a continuous-time system that processes the received signal by an orthonormal set of linear filters, samples the output of each filter once per symbol interval and concatenates the sampled values into a vector, thereby capturing the effects of modulator, additive noise channel and receiver front-end processing on the noisy observation signal. The resulting digital signal vector is fed to the designated detector to carry out the demodulation task. In addition, although the signal model in (1) is in the form of a simple additive noise channel, it is sufficient to incorporate various effects such as thermal noise, multiple-access interference, and jamming [6]. It is also valid in the case of flat-fading channels assuming perfect channel estimation [8]. Note that the probability distribution of the noise component in (1) is not necessarily Gaussian since it is modeled to include the effects of interference and jamming as well. Hence, the noise component can have a significantly different probability distribution from the Gaussian distribution [32–34].

The receiver uses the observation in (1) in order to determine the transmitted information symbol. For that purpose, a generic decision rule (detector) is considered for each channel making a total of K detectors getting utilized at the receiver. That is, for a given observation vector $\mathbf{Y} = \mathbf{y}$, the detector of the ith channel $\phi^{(i)}(\mathbf{y})$ can be characterized as

$$\phi^{(i)}(\mathbf{y}) = j, \quad \text{if } \mathbf{y} \in \Gamma_j^{(i)}, \tag{2}$$

for $j \in \{0, 1, \ldots, M-1\}$, where $\Gamma_0^{(i)}, \Gamma_1^{(i)}, \ldots, \Gamma_{M-1}^{(i)}$ are the decision regions (i.e., a partition of the observation space $\mathbb{R}^N$) for the detector of the ith channel [17]. The transmitter and the receiver can switch between these K channels in any manner in order to optimize the probability of error performance. Let v_i denote the probability that channel i is selected for a given symbol transmission by the communication system. In the remainder of this paper, v_i is called the *channel switching factor* for channel i, where $\sum_{i=1}^{K} v_i = 1$ and $v_i \geqslant 0$ for $i = 1, \ldots, K$. In the context of time sharing, the transmitter and the receiver communicate over channel i for $100v_i$ percent of the time.

The aim of this study is to *jointly* optimize the channel switching strategy $(v_1, \ldots, v_K)$, stochastic signals, and detectors in order to achieve the minimum average probability of error, or equivalently, the maximum average probability of correct decision. The average probability of correct decision can be expressed as $P_c = \sum_{i=1}^{K} v_i P_c^{(i)}$, where $P_c^{(i)}$ represents the corresponding probability of correct decision for channel i under M-ary signaling; that is

$$P_c^{(i)} = \sum_{j=0}^{M-1} \pi_j \int_{\Gamma_j^{(i)}} p_j^{(i)}(\mathbf{y}) \, d\mathbf{y} \tag{3}$$

for $i = 1, 2, \ldots, K$, with $p_j^{(i)}(\mathbf{y})$ denoting the conditional PDF of the observation when the jth symbol is transmitted over the ith channel. Since stochastic signaling is considered, $\mathbf{S}_j^{(i)}$ in (1) is modeled as a random vector. Recalling that the signals and the noise are independent, the conditional PDF of the observation can be obtained as $p_j^{(i)}(\mathbf{y}) = \int_{\mathbb{R}^N} p_{\mathbf{S}_j^{(i)}}(\mathbf{x}) p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{x}) \, d\mathbf{x} = \mathbb{E}\{p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{S}_j^{(i)})\}$, where the

expectation is over the PDF of $\mathbf{S}_j^{(i)}$. Then, the average probability of correct decision can be expressed as

$$P_c = \sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j \mathbb{E}\{p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{S}_j^{(i)})\} \, d\mathbf{y} \right). \tag{4}$$

In practical systems, there is a constraint on the average power emitted from the transmitter. Under the framework of stochastic signaling and channel switching, this constraint on the average power can be expressed in the following form [17].

$$\sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \pi_j \mathbb{E}\{\|\mathbf{S}_j^{(i)}\|_2^2\} \right) \leqslant A, \tag{5}$$

where A denotes the average power limit.

In this study, we primarily concentrate on obtaining the optimal signaling and detection strategy in terms of the correct decision probability for an M-ary communication system in the presence of multiple channels. The novelty of the problem introduced here arises from the following two aspects: (i) signals transmitted over each channel corresponding to different symbols are modeled as random vectors subject to an average power constraint, (ii) the only restriction is the continuity of the noise PDFs of the channels available for switching, and (iii) optimal detectors are designed jointly with the optimal signaling and switching strategies. This formulation, in turn translates into a design problem over the channel switching factors $\{v_i\}_{i=1}^{K}$, channel specific signal PDFs employed at the transmitter $\{p_{\mathbf{S}_0^{(i)}}, p_{\mathbf{S}_1^{(i)}}, \ldots, p_{\mathbf{S}_{M-1}^{(i)}}\}_{i=1}^{K}$, and the corresponding optimal detectors used at the receiver $\{\phi^{(i)}\}_{i=1}^{K}$. Stated more formally, the aim is to solve the following optimization problem.

$$\max_{\{\phi^{(i)}, v_i, p_{\mathbf{S}_0^{(i)}}, p_{\mathbf{S}_1^{(i)}}, \ldots, p_{\mathbf{S}_{M-1}^{(i)}}\}_{i=1}^{K}}$$

$$\sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j \, \mathbb{E}\{p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{S}_j^{(i)})\} \, d\mathbf{y} \right)$$

subject to

$$\sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \pi_j \, \mathbb{E}\{\|\mathbf{S}_j^{(i)}\|_2^2\} \right) \leqslant A,$$

$$\sum_{i=1}^{K} v_i = 1, \quad v_i \geqslant 0, \, \forall i \in \{1, 2, \ldots, K\}. \tag{6}$$

Included in the above statement are the implicit assumptions stating that each $p_{\mathbf{S}_j^{(i)}}(\cdot)$ should represent a PDF. Therefore, $p_{\mathbf{S}_j^{(i)}}(\mathbf{x}) \geqslant 0$, $\forall \mathbf{x} \in \mathbb{R}^N$, and $\int_{\mathbb{R}^N} p_{\mathbf{S}_j^{(i)}}(\mathbf{x}) \, d\mathbf{x} = 1$ are required $\forall j \in \{0, 1, \ldots, M-1\}$ and $\forall i \in \{1, \ldots, K\}$.

The signals for all the M symbols that are transmitted over channel i can be expressed as the elements of a random vector as follows: $\mathbf{S}^{(i)} \triangleq [\mathbf{S}_0^{(i)} \ \mathbf{S}_1^{(i)} \ \cdots \ \mathbf{S}_{M-1}^{(i)}] \in \mathbb{R}^{MN}$, where $\mathbf{S}_j^{(i)}$'s are N-dimensional row vectors $\forall j \in \{0, 1, \ldots, M-1\}$. More explicitly, each realization of $\mathbf{S}^{(i)}$ represents a signal constellation for M-ary symbol transmission in an N-dimensional space. Then, the optimization problem in (6) can be expressed in a more compact form as follows:

$$\max_{\{\phi^{(i)}, v_i, p_{\mathbf{S}^{(i)}}\}_{i=1}^{K}} \quad \sum_{i=1}^{K} v_i \mathbb{E}\big\{G_i\big(\mathbf{S}^{(i)}\big)\big\}$$

$$\text{subject to} \quad \sum_{i=1}^{K} v_i \mathbb{E}\big\{H\big(\mathbf{S}^{(i)}\big)\big\} \leqslant \mathrm{A},$$

$$\sum_{i=1}^{K} v_i = 1, \quad v_i \geqslant 0, \ \forall i \in \{1, 2, \ldots, K\}, \qquad (7)$$

where

$$G_i\big(\mathbf{S}^{(i)}\big) = \sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j \, p_{\mathbf{N}^{(i)}}\big(\mathbf{y} - \mathbf{S}_j^{(i)}\big) \, d\mathbf{y},$$

$$H\big(\mathbf{S}^{(i)}\big) = \sum_{j=0}^{M-1} \pi_j \big\|\mathbf{S}_j^{(i)}\big\|_2^2,$$

and each expectation is taken with respect to $p_{\mathbf{S}^{(i)}}(\cdot)$, which denotes the PDF of the signal constellation employed for symbol transmission over channel i. Specifically, $G_i(\mathbf{s}^{(i)})$ represents the probability of correct decision when the signal constellation represented by the deterministic vector $\mathbf{s}^{(i)}$ is used for the transmission of M symbols over the additive noise channel i and the corresponding detector $\phi^{(i)}$ is employed at the receiver. Then, $\mathbb{E}\{G_i(\mathbf{S}^{(i)})\}$ can be interpreted as the probability of correct decision for a generic stochastic signaling scheme over channel i. The exact number of signal constellations employed by this scheme is determined by the number of distinct values that the random vector $\mathbf{S}^{(i)}$ can take. The expression for $H(\cdot)$ is the same irrespective of which channel is used, and an explicit reference to the channel number as in the subscript of $G_i(\cdot)$ is not necessary.

Let $\mathrm{P}_c^{\dagger}$ denote the maximum average probability of correct decision obtained as the solution of the optimization problem in (7). To provide a simpler formulation of this problem, an upper bound on $\mathrm{P}_c^{\dagger}$ will be derived first, and then the achievability of that bound will be investigated.

Suppose that $G(\mathbf{x})$ denotes the maximum of the probabilities of correct decision when the deterministic signal constellation $\mathbf{x}$ is used for the transmission of M symbols over the additive noise channels $i = 1, 2, \ldots, K$ and the corresponding detectors for all K channels are employed at the receiver. That is, $G(\mathbf{x}) \triangleq \max_{i \in \{1, 2, \ldots, K\}} G_i(\mathbf{x})$, from which $G(\mathbf{x}) \geqslant G_i(\mathbf{x})$ follows $\forall i \in \{1, 2, \ldots, K\}$ and $\forall \mathbf{x} \in \mathbb{R}^{MN}$. This inequality can be applied to the objective function in (7) to obtain a new optimization problem that provides an upper bound on the solution of the optimization problem in (7) as follows.

$$\max_{\{\phi^{(i)}, v_i, p_{\mathbf{S}^{(i)}}\}_{i=1}^{K}} \quad \sum_{i=1}^{K} v_i \mathbb{E}\big\{G\big(\mathbf{S}^{(i)}\big)\big\}$$

$$\text{subject to} \quad \sum_{i=1}^{K} v_i \mathbb{E}\big\{H\big(\mathbf{S}^{(i)}\big)\big\} \leqslant \mathrm{A},$$

$$\sum_{i=1}^{K} v_i = 1, \quad v_i \geqslant 0, \ \forall i \in \{1, 2, \ldots, K\}, \qquad (8)$$

where the expectations are taken with respect to $p_{\mathbf{S}^{(i)}}(\cdot)$'s. Note that by replacing $G_i(\mathbf{S}^{(i)})$ with $G(\mathbf{S}^{(i)})$, the reference to individual channels inside the expectation operator is dropped which will prove useful in the foregoing analysis.

Let $\mathrm{P}_c^{\star}$ denote the maximum average probability of correct decision obtained as the solution to the optimization problem in (8). From the definition of function $G(\cdot)$, $\mathrm{P}_c^{\star} \geqslant \mathrm{P}_c^{\dagger}$ is always satisfied.

In order to achieve further simplification of the problem in (8), define $p_{\mathbf{S}}(\mathbf{s}) \triangleq \sum_{i=1}^{K} v_i p_{\mathbf{S}^{(i)}}(\mathbf{s})$, where $\mathbf{s} \triangleq [\mathbf{s}_0 \, \mathbf{s}_1 \, \cdots \, \mathbf{s}_{M-1}] \in \mathbb{R}^{MN}$, and $\mathbf{s}_j$'s are N-dimensional row vectors $\forall j \in \{0, 1, \ldots, M-1\}$. Since $\sum_{i=1}^{K} v_i = 1$, $v_i \geqslant 0 \forall i$, and $p_{\mathbf{S}^{(i)}}(\cdot)$'s are valid PDFs on $\mathbb{R}^{MN}$, $p_{\mathbf{S}}(\mathbf{s})$ satisfies the conditions to be a PDF. Then, the optimization problem in (8) can be written in the following equivalent form.

$$\max_{p_{\mathbf{S}}, \{\phi^{(i)}\}_{i=1}^{K}} \quad \mathbb{E}\big\{G(\mathbf{S})\big\} \qquad \text{subject to} \quad \mathbb{E}\big\{H(\mathbf{S})\big\} \leqslant \mathrm{A}, \qquad (9)$$

where $G(\mathbf{s}) = \max_{i \in \{1, 2, \ldots, K\}} G_i(\mathbf{s})$ for all $\mathbf{s} \in \mathbb{R}^{MN}$, and the expectations are taken with respect to $p_{\mathbf{S}}(\cdot)$, which denotes the PDF of the signal constellation employed for transmission of symbols $\{0, 1, \ldots, M-1\}$.

In (9), $G(\mathbf{s})$ represents the maximum of the probabilities of correct decision when the deterministic signal constellation $\mathbf{s}$ is used for the transmission of M symbols over the additive noise channels $i = 1, 2, \ldots, K$ and the corresponding detectors are employed at the receiver. Then, $\mathbb{E}\{G(\mathbf{S})\}$ can be interpreted as a randomization among channels with respect to the PDF $p_{\mathbf{S}}(\cdot)$, where the probability of correct decision corresponding to each component of $p_{\mathbf{S}}$ (i.e., for each signal constellation $\mathbf{s}$ in the support of $p_{\mathbf{S}}$) is maximized by transmitting it over the *most favorable channel* (i.e., the channel with the highest probability of correct decision for the given signal constellation $\mathbf{s}$), and altogether they maximize the average probability of correct decision.

Optimization problems in the form of (9) have been investigated in various studies in the literature [1,5,13]. Assuming that the signal values specified by the signal constellation $\mathbf{s} \in \mathbb{R}^{MN}$ are bounded, i.e., $\boldsymbol{a} \preccurlyeq \mathbf{s} \preccurlyeq \boldsymbol{b}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are finite real vectors in $\mathbb{R}^{MN}$, and $\preccurlyeq$ denotes element-wise inequality; an optimal solution to (9) can be represented by a randomization of at most two signal constellations, that is, $p_{\mathbf{S}}(\mathbf{s}) = \lambda \delta(\mathbf{s} - \mathbf{s}_1) + (1 - \lambda)\delta(\mathbf{s} - \mathbf{s}_2)$, where $\lambda \in [0, 1]$ and $\delta(\cdot)$ is the Dirac delta function. This result follows from Carathéodory's theorem [35], and can be derived using a similar approach to those in [1, Theorem 3] and [5, Theorem 4]. Substituting this result in (9), the following optimization problem is obtained:

$$\max_{\{\lambda, \mathbf{s}_1, \mathbf{s}_2, \{\phi^{(i)}\}_{i=1}^{K}\}} \quad \lambda G(\mathbf{s}_1) + (1 - \lambda) G(\mathbf{s}_2)$$

$$\text{subject to} \quad \lambda H(\mathbf{s}_1) + (1 - \lambda) H(\mathbf{s}_2) \leqslant \mathrm{A}, \quad \lambda \in [0, 1], \qquad (10)$$

where

$$G(\mathbf{s}_k) = \max_{i \in \{1, 2, \ldots, K\}} G_i(\mathbf{s}_k),$$

$$G_i(\mathbf{s}_k) = \sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j \, p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_{k,j}) \, d\mathbf{y} \quad \forall i \in \{1, 2, \ldots, K\},$$

$$H(\mathbf{s}_k) = \sum_{j=0}^{M-1} \pi_j \big\|\mathbf{s}_{k,j}\big\|_2^2, \quad \text{and}$$

$$\mathbf{s}_k = [\mathbf{s}_{k,0} \, \mathbf{s}_{k,1} \, \cdots \, \mathbf{s}_{k,M-1}] \in \mathbb{R}^{MN}$$

with $\mathbf{s}_{k,j}$ denoting the N-dimensional vector representing the jth symbol in constellation $\mathbf{s}_k$. Therefore, the solution to the optimization problem given in (9), which is an upper bound on the solution of the original problem presented in (6), is achieved by randomizing between at most two signal constellations, $\mathbf{s}_1$ and $\mathbf{s}_2$.

In order to understand the possible implications of the representation given in (10), we consider the following scenario. Let λ, $\mathbf{s}_1$ and $\mathbf{s}_2$ be the optimal parameters obtained from the solution of (10). Evidently, if either $\lambda = 0$ or $\mathbf{s}_1 = \mathbf{s}_2$, this would

imply that the optimal performance under the average power constraint is achieved by transmitting over a single channel with deterministic signaling. Next, we investigate the possible cases for $\lambda \neq 0$ and $\mathbf{s}_1 \neq \mathbf{s}_2$. By construction, $G(\mathbf{s}_k)$ selects the channel with the largest average probability of correct decision for the transmission of the symbols in the constellation $\mathbf{s}_k$. Therefore, it may either be that $\mathbf{s}_1$ and $\mathbf{s}_2$ are transmitted over the same channel (i.e., stochastic signaling over a single channel) or over distinct channels (i.e., channel switching with deterministic signals over each channel). It should be noted that channel switching between two channels while stochastic signaling over each channel is overruled by the form of the optimization problem given in (10). Nevertheless, an intuitive explanation for this fact can be given as follows. Suppose that the optimal strategy results in switching between channels 1 and 2 with probability λ, and it is found that randomization between two signal constellations, represented with $\mathbf{s}_1^{(i)}$ and $\mathbf{s}_2^{(i)}$, is optimal with probability α_i for channel $i \in \{1, 2\}$, where $\lambda, \alpha_1, \alpha_2 \in (0, 1)$. Let $(g_1^{(i)}, h_1^{(i)})$ denote the point for the average probability of correct decision and average signal power corresponding to the signal constellation $\mathbf{s}_1^{(i)}$. Similarly, let $(g_2^{(i)}, h_2^{(i)})$ denote the corresponding point for the signal constellation $\mathbf{s}_2^{(i)}$. It is easy to see that the assumed strategy results in a convex combination of the four points in the following set $S \triangleq \{(g_1^{(1)}, h_1^{(1)}), (g_2^{(1)}, h_2^{(1)}), (g_1^{(2)}, h_1^{(2)}), (g_2^{(2)}, h_2^{(2)})\}$. This convex combination is determined by the parameters λ, α_1 and α_2. For different values of these parameters, any point in the convex hull of the set S can be attained. However, since an optimal strategy should maximize the average probability of correct decision under the average transmit power constraint, the optimal point should lie on the boundary of the convex hull of the set S. But any point on the boundary of the convex hull can be represented by a convex combination of at most two points in the set S, which implies that the optimal strategy is, in fact, either stochastic signaling over a single channel or switching between two channels with deterministic signals over each channel. All in all, it is concluded that the objective function in (10) is maximized under the specified constraints by either one of the following strategies:

1. transmitting exclusively over a single channel via deterministic signaling, i.e., $\lambda \in \{0, 1\}$,
2. randomizing (time sharing) between two signal constellations over a single channel, i.e., $\lambda \in (0, 1)$ and $\arg\max_{i \in \{1,2,\dots,K\}} G_i(\mathbf{s}_1) = \arg\max_{i \in \{1,2,\dots,K\}} G_i(\mathbf{s}_2)$,
3. switching (time sharing) between two channels and deterministic signaling over each channel, i.e., $\lambda \in (0, 1)$ and $\arg\max_{i \in \{1,2,\dots,K\}} G_i(\mathbf{s}_1) \neq \arg\max_{i \in \{1,2,\dots,K\}} G_i(\mathbf{s}_2)$.

Three distinct cases mentioned above can also be grouped under two overlapping cases as follows:

1. randomizing between *at most* two signal constellations over a single channel,
2. switching between *at most* two channels and deterministic signaling over each channel.

It is noted that randomizing between *at most* two signal constellations over a single channel covers deterministic signaling since the former reduces to the latter for $\lambda \in \{0, 1\}$. Similarly, switching between *at most* two channels and deterministic signaling over each channel also reduces to deterministic signaling over a single channel when $\lambda \in \{0, 1\}$. This form is introduced because it provides an ease of notation in the following analysis.

The last step in the simplification of the optimization problem in (10) comes from an observation about the structure of optimal detectors. For a given channel i and the corresponding signaling

scheme over the channel (deterministic or randomization between two signal constellations), the conditional probability of the observation $\mathbf{y}$ given that symbol j is transmitted can be expressed as

$$p_j^{(i)}(\mathbf{y}) = \mathbb{E}\{p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{S}_j^{(i)})\}$$

$$= \begin{cases} p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_j^{(i)}), \\ \quad \text{if deterministic}, \\ \lambda p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_{1,j}^{(i)}) + (1-\lambda)p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_{2,j}^{(i)}), \\ \quad \text{if randomized}. \end{cases} \quad (11)$$

When deciding among M symbols based on observation $\mathbf{y}$ at detector i, the MAP decision rule selects symbol j if $j = \arg\max_{l \in \{0,1,\dots,M-1\}} \pi_l p_l^{(i)}(\mathbf{y})$, and it maximizes the probability of correct decision [17]. Therefore, it is not necessary to search over all decision rules in (10); only the MAP decision rule should be determined for the detector of each channel and its corresponding probability of correct decision should be considered. The probability of correct decision for a generic decision rule is given in (3). Using the decision regions corresponding to the MAP detector, i.e., $\Gamma_j^{(i)} = \{\mathbf{y} \in \mathbb{R}^N \mid \pi_j p_j^{(i)}(\mathbf{y}) \geqslant \pi_l p_l^{(i)}(\mathbf{y}), \forall l \neq j\}$, the average probability of correct decision for ith channel becomes

$$P_{c,\text{MAP}}^{(i)} = \int_{\mathbb{R}^N} \max_{j \in \{0,1,\dots,M-1\}} \{\pi_j \, p_j^{(i)}(\mathbf{y})\} \, d\mathbf{y}, \quad (12)$$

where $p_j^{(i)}(\mathbf{y})$ is as in (11).

Below, more explicit forms of the optimization problem stated in (10) are given for all possible scenarios mentioned previously.

(i) Case 1. Transmitting exclusively over a single channel via deterministic signaling:

In this case, a single channel is utilized exclusively, and the transmitted signal for each symbol is deterministic, i.e., a fixed signal constellation is employed for symbol transmission over the channel. Without loss of generality, channel i is considered. The optimization problem in (10) becomes

$$\max_{\{\mathbf{s}^{(i)}, \phi^{(i)}\}} \quad \sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_j^{(i)}) \, d\mathbf{y}$$

$$\text{subject to} \quad \sum_{j=0}^{M-1} \pi_j \|\mathbf{s}_j^{(i)}\|_2^2 \leqslant A. \quad (13)$$

Using the result given in (12) for the deterministic case, the equivalent optimization problem can be written as follows.

$$\max_{\mathbf{s}^{(i)}} \int_{\mathbb{R}^N} \max_{j \in \{0,1,\dots,M-1\}} \{\pi_j \, p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_j^{(i)})\} \, d\mathbf{y}$$

$$\text{subject to} \quad \sum_{j=0}^{M-1} \pi_j \|\mathbf{s}_j^{(i)}\|_2^2 \leqslant A. \quad (14)$$

(ii) Case 2. Randomizing (time sharing) between at most two signal constellations over a single channel:

Similarly to the previous case, the transmission occurs over a single channel exclusively, but in this case the transmitted signal for each symbol is a randomization between at most two different signal vectors. Without loss of generality, channel i is considered. The optimization problem in (10) is expressed as follows.

$$\max_{\{\lambda, \mathbf{s}_1^{(i)}, \mathbf{s}_2^{(i)}, \phi^{(i)}\}} \quad \lambda G_i(\mathbf{s}_1^{(i)}) + (1-\lambda)G_i(\mathbf{s}_2^{(i)})$$

$$\text{subject to} \quad \lambda H(\mathbf{s}_1^{(i)}) + (1-\lambda)H(\mathbf{s}_2^{(i)}) \leqslant A, \quad \lambda \in [0, 1] \quad (15)$$

where

$$G_i\big(\mathbf{s}_k^{(i)}\big) = \sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j\, p_{\mathbf{N}^{(i)}}\big(\mathbf{y} - \mathbf{s}_{k,j}^{(i)}\big)\, d\mathbf{y},$$

$$H(\mathbf{s}_k) = \sum_{j=0}^{M-1} \pi_j \big\|\mathbf{s}_{k,j}^{(i)}\big\|_2^2,$$

and $k \in \{1, 2\}$. As stated earlier, it is assumed that a single detector is employed for each channel at the receiver. Using the result for randomized signaling case given in (12), the equivalent optimization problem can be written as

$$\max_{\{\lambda,\mathbf{s}_1^{(i)},\mathbf{s}_2^{(i)}\}} \int_{\mathbb{R}^N} \max_{j\in\{0,1,\ldots,M-1\}} \big\{\pi_j\, p_j^{(i)}(\mathbf{y})\big\}\, d\mathbf{y}$$

subject to $\quad \lambda\left(\sum_{j=0}^{M-1} \pi_j \big\|\mathbf{s}_{1,j}^{(i)}\big\|_2^2\right) + (1-\lambda)\left(\sum_{j=0}^{M-1} \pi_j \big\|\mathbf{s}_{2,j}^{(i)}\big\|_2^2\right)$

$$\leqslant A, \quad \lambda \in [0, 1] \qquad (16)$$

where $p_j^{(i)}(\mathbf{y}) = \lambda p_{\mathbf{N}^{(i)}}\big(\mathbf{y} - \mathbf{s}_{1,j}^{(i)}\big) + (1-\lambda) p_{\mathbf{N}^{(i)}}\big(\mathbf{y} - \mathbf{s}_{2,j}^{(i)}\big)$. It is recalled that the optimization problem in (16) reduces to that of (14) when $\lambda \in \{0, 1\}$.

(iii) Case 3. Switching (time sharing) between at most two channels and deterministic signaling over each channel:

In this case, optimum performance is investigated while transmitting over at most two channels and the transmission over each channel is deterministic, i.e., a fixed signal constellation is employed for symbol transmission over each channel but the channels are switched in time. Without loss of generality, channels i and l are considered ($i \neq l$ and $i, l \in \{1, 2, \ldots, K\}$). The optimization problem in (10) takes the following form.

$$\max_{\{\lambda,\mathbf{s}^{(i)},\mathbf{s}^{(l)},\phi^{(i)},\phi^{(l)}\}} \lambda G_i\big(\mathbf{s}^{(i)}\big) + (1-\lambda)G_l\big(\mathbf{s}^{(l)}\big)$$

subject to $\quad \lambda H\big(\mathbf{s}^{(i)}\big) + (1-\lambda)H\big(\mathbf{s}^{(l)}\big) \leqslant A,$

$$\lambda \in [0, 1] \qquad (17)$$

where

$$G_i\big(\mathbf{s}^{(i)}\big) = \sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j\, p_{\mathbf{N}^{(i)}}\big(\mathbf{y} - \mathbf{s}_j^{(i)}\big)\, d\mathbf{y},$$

$$H\big(\mathbf{s}^{(i)}\big) = \sum_{j=0}^{M-1} \pi_j \big\|\mathbf{s}_j^{(i)}\big\|_2^2,$$

$G_l(\mathbf{s}^{(l)})$ and $H(\mathbf{s}^{(l)})$ are defined similarly by replacing i with l in the preceding equations. Since deterministic signaling is employed in each channel, the result given in (12) for the deterministic case should be applied for each channel. Then, an equivalent optimization problem can be written as

$$\max_{\{\lambda,\mathbf{s}^{(i)},\mathbf{s}^{(l)}\}} \lambda G_{i,\mathrm{MAP}}\big(\mathbf{s}^{(i)}\big) + (1-\lambda)G_{l,\mathrm{MAP}}\big(\mathbf{s}^{(l)}\big)$$

subject to $\quad \lambda H\big(\mathbf{s}^{(i)}\big) + (1-\lambda)H\big(\mathbf{s}^{(l)}\big) \leqslant A, \quad \lambda \in [0, 1] \qquad (18)$

where

$$G_{i,\mathrm{MAP}}\big(\mathbf{s}^{(i)}\big) = \int_{\mathbb{R}^N} \max_{j\in\{0,1,\ldots,M-1\}} \big\{\pi_j\, p_{\mathbf{N}^{(i)}}\big(\mathbf{y} - \mathbf{s}_j^{(i)}\big)\big\}\, d\mathbf{y},$$

$$H\big(\mathbf{s}^{(i)}\big) = \sum_{j=0}^{M-1} \pi_j \big\|\mathbf{s}_j^{(i)}\big\|_2^2,$$

$G_l(\mathbf{s}^{(l)})$ and $H(\mathbf{s}^{(l)})$ are defined similarly by replacing i with l in the respective equations.

It is noted that the optimization space is considerably reduced in (14), (16) and (18) compared to those in (13), (15) and (17), respectively since there is no need to search over the detectors in (14), (16) and (18).

In the rest of the analysis, only the second and third cases will be investigated since they cover deterministic signaling over a single channel as a special case. In view of the above analysis, the solution of the optimization problem in (10) can be decomposed into two parts. First, randomizing between at most two signal constellations over a single channel is considered. Let $P_{c,\mathrm{Opt}}^{(i)}$ be the solution of the optimization problem in (16) for ith channel; that is, $P_{c,\mathrm{Opt}}^{(i)}$ denotes the maximum average probability of correct decision that can be achieved by stochastic signaling over channel i under the average power constraint. Secondly, switching between at most two channels with deterministic signaling over each channel is considered. Let $P_{c,\mathrm{Opt}}^{(i,l)}$ be the solution of the optimization problem in (18) for channels i and l; that is, $P_{c,\mathrm{Opt}}^{(i,l)}$ denotes the maximum average probability of correct decision that can be achieved by switching between channels i and l under the average power constraint. Then, the solution of the optimization problem in (10) can be obtained by solving the following set of optimization problems and computing their maximum.

$$P_c^{\mathrm{Stoc}} = \max_{i\in\{1,2,\ldots,K\}} P_{c,\mathrm{Opt}}^{(i)}, \qquad (19)$$

$$P_c^{\mathrm{CS}} = \max_{i,l\in\{1,2,\ldots,K\} \text{ and } i<l} P_{c,\mathrm{Opt}}^{(i,l)}, \qquad (20)$$

$$P_c^{\star} = \max\big\{P_c^{\mathrm{Stoc}}, P_c^{\mathrm{CS}}\big\} \qquad (21)$$

where the superscript Stoc denotes stochastic signaling over a single channel and CS abbreviates channel switching. When the noise PDFs on all the channels are different, the solution of the optimization problem is given by (21) without any further simplifications. In order to calculate P_c^{Stoc} in (19), the optimal stochastic signaling strategy described by the optimization problem given in (16) should be obtained for all K channels. Likewise, P_c^{CS} in (20) requires that the optimal channel switching strategy characterized by the optimization problem given in (18) should be computed for all channels pairs. Since there are K distinct channels and $K(K-1)/2$ distinct channel pairs, a total of $K(K+1)/2$ optimization problems must be solved to obtain the corresponding performance scores, among which the maximum is selected according to (21) to identify the optimum strategy. In the cases where some channels share the same noise PDF, the results are still valid but the optimization sets given in (19) and (20) over which the maximum values are computed can be refined to avoid repeated computations of the same expressions.[2]

The following proposition states that the expressions in (19)–(21) provides the solution of the generic problem in (7).

Proposition 1. *The maximum average probabilities of correct decision achieved by the solutions of the optimization problems in (7) and (21) are equal, i.e., $P_c^{\dagger} = P_c^{\star}$.*

Proof. First, consider the optimization problem in (7) when $K = 2$ channels are used, and deterministic signaling is employed for each channel, i.e., $p_{\mathbf{S}^{(1)}}(\mathbf{s}^{(1)}) = \delta(\mathbf{s}^{(1)} - \mathbf{s}_1)$ and $p_{\mathbf{S}^{(2)}}(\mathbf{s}^{(2)}) = \delta(\mathbf{s}^{(2)} - \mathbf{s}_2)$.

---

[2] Detector randomization as discussed in [3,13] can also be analyzed using our framework. Specifically, it can be modeled by assuming that some channels have identical noise distributions. That is, each channel appears in the system model with a certain multiplicity.

Suppose also that the symbols transmitted over each channel are decoded using the MAP detector corresponding to that channel. In that case, (7) reduces to the optimization problem in (18); hence, (7) covers (18) as a special case. Secondly, consider the optimization problem in (7) when $K = 1$ channel is used, and a randomization between at most two signal constellations is employed, i.e., $p_{\mathbf{S}}(\mathbf{s}) = \lambda \delta(\mathbf{s} - \mathbf{s}_1) + (1 - \lambda)\delta(\mathbf{s} - \mathbf{s}_2)$. Suppose also that a single MAP detector is employed at the receiver. Then, (7) reduces to the optimization problem in (16); hence, (7) covers (16) as a special case. Since both (16) and (18) are special cases of (7) for any choice of the channels $i \in \{1, 2, \ldots, K\}, l \in \{1, 2, \ldots, K\}$ and $i \neq l$, the maximum value of the objective function in (7) should be larger than or equal to the maximum given by (21). This, in turn, implies that $P_c^\dagger \geqslant P_c^\star$. On the other hand, the optimization problem in (7) has been replaced with the upper bound given in (8), the solution of which is shown to reduce to that given in (21); that is, $P_c^\dagger \leqslant P_c^\star$. Therefore, it is concluded that $P_c^\dagger = P_c^\star$. □

Proposition 1 implies that the solution of the original optimization problem stated in (7), which considers the joint optimization of switching factors among channels, channel specific signal PDFs employed at the transmitter and the corresponding detectors used at the receiver, can be obtained as the solution of the much simpler optimization problem specified in (21). Formally, when multiple channels are available for signal transmission (i.e., $K \geqslant 2$), it is sufficient to either employ switching between two channels with deterministic signaling over each channel (i.e., there is no need to employ stochastic signaling over a channel to achieve the optimal solution while switching channels); or randomize between at most two signal constellations over a single channel, whichever results in the highest average probability of correct decision.

The solution of the optimization problem in (21) can be obtained via global optimization techniques (since it is a nonlinear nonconvex optimization problem in general due to arbitrary noise PDFs), or a convex relaxation approach as in [5] can be employed to obtain approximate solutions in polynomial time.

3. Improvability and non-improvability conditions

Although the solution given in (19)–(21) has simplified the search over all possible channel switching factors, signal PDFs and decision rules (see (7)) to a search over a few variables (see (16) and (18)), it is still computationally intensive. Specifically, for the optimal stochastic signaling strategy given in (19), the maximum correct decision probabilities should be computed for all K channels. Similarly, for the optimal channel switching strategy given in (20), the maximum correct decision probabilities should be computed for all $K(K - 1)/2$ distinct pairs of channels. In total, it is required to solve $K(K + 1)/2$ optimization problems, and there are $2MN + 1$ optimization variables in each problem (i.e., $\mathbf{s}_1$ and $\mathbf{s}_2$ are two signal constellations employed for M-ary communications in an N-dimensional signal space and λ is a scalar parameter). Therefore, it is very important to know, before attempting to solve the overall optimization problem, whether channel switching in the presence of stochastic signaling can help improve the performance of the communication system under an average power constraint.

Remark. From this point on, the terms *channel switching* and *stochastic signaling* are used to refer to "switching between two channels with deterministic signaling over each channel" and "randomization between at most two signal constellations over a single channel", respectively.

In order to define *improvability* and *non-improvability*, we refer to a *conventional* communications scenario, in which the transmitter employs a fixed constellation with average signal power A

(e.g., antipodal signaling with $\{-\sqrt{A}, \sqrt{A}\}$ for binary communications) over the channel that results in the highest correct decision probability and the receiver uses the corresponding MAP detector. Then, the system is called *improvable* if either stochastic signaling or channel switching[3] can improve the average probability of correct decision over the conventional signaling method. Otherwise, the system is called *nonimprovable*.

Before writing down the expression for the average correct decision probability of the conventional system, we need to introduce more notation. Recall from (18) that $G_{i,\mathrm{MAP}}(\mathbf{s})$ represents the average probability of correct decision when the deterministic signal constellation $\mathbf{s}$ is used for the transmission of M symbols over the additive noise channel i and the corresponding MAP detector is employed at the receiver for the same channel. Next, $G_{\mathrm{MAP}}(\mathbf{s})$ is defined as the maximum of these correct decision probabilities for the given signal constellation $\mathbf{s}$ over all K additive noise channels. Namely,

$$G_{\mathrm{MAP}}(\mathbf{s}) \triangleq \max_{i \in \{1, 2, \ldots, K\}} G_{i,\mathrm{MAP}}(\mathbf{s}) \quad \text{for all } \mathbf{s}, \tag{22}$$

where $G_{i,\mathrm{MAP}}(\mathbf{s}) = \int_{\mathbb{R}^N} \max_{j \in \{0,1,\ldots,M-1\}}\{\pi_j \ p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_j)\} \, d\mathbf{y}$. It is also recalled that $H(\mathbf{s})$ denotes the average power of the signal constellation $\mathbf{s}$ over the prior probabilities (see its definition after (10)). With this notation, the probability of correct decision for the conventional system can be expressed as $P_c^{\mathrm{cv}} = G_{\mathrm{MAP}}(\mathbf{s}_{\mathrm{cv}})$, where $\mathbf{s}_{\mathrm{cv}}$ represents the *conventional* deterministic signal constellation employed for the transmission of all the M symbols, and $H(\mathbf{s}_{\mathrm{cv}}) = \mathrm{A}$ is satisfied. The *max* operator in (22) ensures that $\mathbf{s}_{\mathrm{cv}}$ is transmitted over the channel with the highest correct decision probability. It is also sensible to assume that the components of the constellation vector $\mathbf{s}_{\mathrm{cv}}$, i.e., the signal vectors employed for symbol transmission, are designed to maximize the correct decision probability under the average power constraint, but some popular choices can also be assumed such as M-PAM or M-QAM [36]. The aim is to improve upon P_c^{cv} under the average power constraint. Next, $G_{i,\phi}(\mathbf{s})$ is defined as the probability of correct decision when the signal constellation $\mathbf{s}$ is transmitted over channel i and decoded using a given fixed decision rule ϕ. Similar to the above discussion, $G_\phi(\mathbf{s})$ is defined as the maximum of these correct decision probabilities over all the additive noise channels.

$$G_\phi(\mathbf{s}) \triangleq \max_{i \in \{1, 2, \ldots, K\}} G_{i,\phi}(\mathbf{s}) \quad \text{for all } \mathbf{s}, \tag{23}$$

where $G_{i,\phi}(\mathbf{s}) = \sum_{j=0}^{M-1} \int_{\Gamma_j} \pi_j p_{\mathbf{N}^{(i)}}(\mathbf{y} - \mathbf{s}_j) \, d\mathbf{y}$. In (22), each channel is allowed to employ its own MAP detector that is tuned according to the channel noise and signal constellation, whereas in (23), the same decision rule is used for all the channels.

Suppose that the conventional system transmits over a specific channel $\hat{i}$ using the signal constellation $\mathbf{s}_{\mathrm{cv}}$ and decoding is performed using the corresponding MAP detector $\hat{\phi}$, thereby achieving the highest correct decision probability P_c^{cv} via deterministic signaling with $\mathbf{s}_{\mathrm{cv}}$ over a single channel. That is, $P_c^{\mathrm{cv}} = G_{\mathrm{MAP}}(\mathbf{s}_{\mathrm{cv}}) = G_{\hat{i},\mathrm{MAP}}(\mathbf{s}_{\mathrm{cv}}) = G_{\hat{i},\hat{\phi}}(\mathbf{s}_{\mathrm{cv}}) = G_{\hat{\phi}}(\mathbf{s}_{\mathrm{cv}})$. Let $\mathsf{S}_h = \{\mathbf{s}: H(\mathbf{s}) = h\}$. For a given value h of H, we have $\mathbf{s} = H^{-1}(h)$, where H^{-1} is the inverse mapping of H. Since H is not a one-to-one function, there exists a set of values $\mathbf{s}$ which satisfy $H(\mathbf{s}) = h$. A new function $\mathcal{J}_{\hat{\phi}}(h)$ is defined as

$$\mathcal{J}_{\hat{\phi}}(h) \triangleq \max_{\mathbf{s} \in \mathsf{S}_h} G_{\hat{\phi}}(\mathbf{s}), \tag{24}$$

which specifies the maximum probability of correct decision that can be attained for a given value of the average signal power h us-

---

3 Together, they constitute the solution for the optimal signaling and detector design problem in the presence of multiple additive noise channels.

ing the detector of the conventional system $\hat{\phi}$ [1,37].[4] More clearly, the maximum is computed over the performances of the individual channels for the set of signal vectors with the given average power value h when deterministic signaling is performed over each channel and decoding is accomplished at the receiver via the detector $\hat{\phi}$ of the conventional system. In other words, an equivalent representation for $\mathcal{J}_{\hat{\phi}}(h)$ can be given as $\mathcal{J}_{\hat{\phi}}(h) = \max_{\mathbf{s} \in S_h, i \in \{1,2,...,K\}} G_{i,\hat{\phi}}(\mathbf{s})$. Lastly, it should be noted that $\mathcal{J}_{\hat{\phi}}(A) \geqslant P_c^{cv}$ by definition. Based on these preliminaries, the following improvability condition is obtained first.

Proposition 2. *If $\mathcal{J}_{\hat{\phi}}(h)$ is second-order continuously differentiable around* A *and satisfies $\mathcal{J}_{\hat{\phi}}''(A) > 0$, then the communication system is improvable.*

Proof. When $\mathcal{J}_{\hat{\phi}}''(A) > 0$ and $\mathcal{J}_{\hat{\phi}}(h)$ in (24) is second-order continuously differentiable around $h = A$, there exists $\varepsilon > 0$ such that $\mathcal{J}_{\hat{\phi}}(h)$ is convex on the interval $(A - \varepsilon, A + \varepsilon)$. Consider a signaling scheme with PDF $p_{\mathbf{S}}(\mathbf{s}) = 0.5\delta(\mathbf{s} - \mathbf{s}_1) + 0.5\delta(\mathbf{s} - \mathbf{s}_2)$ where $H(\mathbf{s}_1) = A - \varepsilon$ and $H(\mathbf{s}_2) = A + \varepsilon$. Since $H(\mathbf{s})$ is a continuous mapping from $\mathbb{R}^{MN}$ to $[0, \infty)$, the existence of $\mathbf{s}_1$ and $\mathbf{s}_2$ is assured. First, it is observed that the average transmit power under the proposed signaling scheme does not violate the power constraint. Formally, $0.5H(\mathbf{s}_1) + 0.5H(\mathbf{s}_2) = 0.5(A - \varepsilon) + 0.5(A + \varepsilon) = A$. Next, due to the strict convexity of $\mathcal{J}_{\hat{\phi}}$ around $h = A$, we have $0.5\mathcal{J}_{\hat{\phi}}(A - \varepsilon) + 0.5\mathcal{J}_{\hat{\phi}}(A + \varepsilon) > \mathcal{J}_{\hat{\phi}}(A) \geqslant P_c^{cv}$. From the definition of $\mathcal{J}_{\hat{\phi}}(h)$, it is also observed that there exist channels m and n such that $0.5G_{m,\hat{\phi}}(\mathbf{s}_1) + 0.5G_{n,\hat{\phi}}(\mathbf{s}_2) > P_c^{cv}$. Depending on the channel PDFs, the performance scores $G_{m,\hat{\phi}}(\mathbf{s}_1)$ and $G_{n,\hat{\phi}}(\mathbf{s}_2)$ can be attained by either transmitting over the same channel ($m = n$) or on distinct channels ($m \neq n$), and employing the detector $\hat{\phi}$ of the conventional system. In the case of transmitting over the same channel, the performance can further be increased by designing the optimal MAP detector corresponding to the PDF $p_{\mathbf{S}}(\mathbf{s}) = 0.5\delta(\mathbf{s} - \mathbf{s}_1) + 0.5\delta(\mathbf{s} - \mathbf{s}_2)$ instead of using the detector $\hat{\phi}$. Similarly, in the case of different channels, each channel can employ its own optimal MAP detector resulting in a better performance score, that is $0.5G_{m,\text{MAP}}(\mathbf{s}_1) + 0.5G_{n,\text{MAP}}(\mathbf{s}_2) \geqslant 0.5G_{m,\hat{\phi}}(\mathbf{s}_1) + 0.5G_{n,\hat{\phi}}(\mathbf{s}_2) > P_c^{cv}$. Hence, it is concluded that under the assumptions in the proposition, the correct decision probability can be improved using either stochastic signaling or channel switching depending on the channel noise PDFs. $\square$

In order to evaluate the improvability condition in Proposition 2, explicit knowledge about the behavior of $\mathcal{J}_{\hat{\phi}}(h)$ around $h = A$ is required. This could be a difficult task since $H(\mathbf{s})$ is not an injective function and the exact form of $G_{\hat{\phi}}(\mathbf{s})$ can be hard to compute. In such cases, the relationship between $G_{\hat{\phi}}(\mathbf{s})$ and $H(\mathbf{s})$ can be learned by Monte Carlo simulation using importance sampling [1]. Once this is accomplished, the check for the improvability condition stated in Proposition 2 can be carried out in the single-dimensional domain of $\mathcal{J}_{\hat{\phi}}(h)$ instead of the multi-dimensional domain of $G_{\hat{\phi}}(\mathbf{s})$. In the following, we present improvability conditions that can be evaluated in a more direct manner without relying on auxiliary functions like $\mathcal{J}_{\hat{\phi}}(h)$.

Proposition 3. *The average probability of correct decision can be improved if there exists signal constellations $\mathbf{s}_1$ and $\mathbf{s}_2$ such that*

- $H(\mathbf{s}_1) > A > H(\mathbf{s}_2)$, *and*
- $(A - H(\mathbf{s}_2))(G_{\hat{\phi}}(\mathbf{s}_1) - G_{\hat{\phi}}(\mathbf{s}_2)) > (H(\mathbf{s}_1) - H(\mathbf{s}_2))(P_c^{cv} - G_{\hat{\phi}}(\mathbf{s}_2))$

are satisfied.

Proof. Consider a signaling scheme with PDF $p_{\mathbf{S}}(\mathbf{s}) = \lambda\delta(\mathbf{s} - \mathbf{s}_1) + (1 - \lambda)\delta(\mathbf{s} - \mathbf{s}_2)$, which utilizes all the average power, i.e. $\lambda H(\mathbf{s}_1) + (1 - \lambda)H(\mathbf{s}_2) = A$. From this equality, λ can be expressed as $\lambda = (A - H(\mathbf{s}_2))/(H(\mathbf{s}_1) - H(\mathbf{s}_2))$. Since $\lambda \in (0, 1)$ must be satisfied for physically realizable configurations, the first condition follows. Secondly, observe that the expression $\lambda G_{\hat{\phi}}(\mathbf{s}_1) + (1 - \lambda)G_{\hat{\phi}}(\mathbf{s}_2)$ provides a lower bound on the performance of optimal design given in (21). Then, $\lambda G_{\hat{\phi}}(\mathbf{s}_1) + (1 - \lambda)G_{\hat{\phi}}(\mathbf{s}_2) > P_c^{cv}$ is sufficient for improvability of the conventional system. The second condition in the proposition can be obtained by substituting the expression for λ into the preceding inequality. $\square$

The following corollary follows from Proposition 3 by focusing on the improvements due to channel switching only.

Corollary 1. *The average probability of correct decision can be improved if there exists different channels i_1 and i_2, and signal constellations $\mathbf{s}_1$ and $\mathbf{s}_2$ that satisfy*

- $H(\mathbf{s}_1) > A > H(\mathbf{s}_2)$, *and*
- $(A - H(\mathbf{s}_2))(G_{i_1,\text{MAP}}(\mathbf{s}_1) - G_{i_2,\text{MAP}}(\mathbf{s}_2)) > (H(\mathbf{s}_1) - H(\mathbf{s}_2))(P_c^{cv} - G_{i_2,\text{MAP}}(\mathbf{s}_2))$.

If a pair of signal constellations that satisfy the conditions in Proposition 3 and Corollary 1 is found, one can continue to solve for the optimal set of parameters using the approach given in (21). Next, some alternative improvability conditions are stated.

Proposition 4. *Suppose that $G_{\hat{\phi}}(\mathbf{s})$ is second-order continuously differentiable around $\mathbf{s} = \mathbf{s}_{cv}$. Define $g^{(1)}(\mathbf{s}, \mathbf{x}) \triangleq \mathbf{J}_G(\mathbf{s})\mathbf{x}$, $g^{(2)}(\mathbf{s}, \mathbf{x}) \triangleq \mathbf{x}^T\mathbf{H}_G(\mathbf{s})\mathbf{x}$, $h^{(1)}(\mathbf{s}, \mathbf{x}) \triangleq 2\mathbf{s}^T\boldsymbol{\Pi}\mathbf{x}$, and $h^{(2)}(\mathbf{x}) \triangleq 2\mathbf{x}^T\boldsymbol{\Pi}\mathbf{x}$, where $\mathbf{J}_G(\mathbf{s})$ denotes the Jacobian (gradient) of $G_{\hat{\phi}}$ evaluated at $\mathbf{s}$, $\mathbf{H}_G(\mathbf{s})$ denotes the Hessian of $G_{\hat{\phi}}$ evaluated at $\mathbf{s}$, and $\boldsymbol{\Pi}$ is a diagonal matrix of prior probabilities obtained by repeating each prior N times consecutively along the diagonal.[5] Then, the probability of correct decision can be improved if there exists a signal constellation $\mathbf{x}$ such that*

- $g^{(1)}(\mathbf{s}, \mathbf{x})h^{(1)}(\mathbf{s}, \mathbf{x}) < 0$ *is satisfied at $\mathbf{s} = \mathbf{s}_{cv}$, or*
- $g^{(1)}(\mathbf{s}, \mathbf{x}) > 0$, $h^{(1)}(\mathbf{s}, \mathbf{x}) > 0$, *and* $g^{(2)}(\mathbf{s}, \mathbf{x})h^{(1)}(\mathbf{s}, \mathbf{x}) > h^{(2)}(\mathbf{x})g^{(1)}(\mathbf{s}, \mathbf{x})$ *are satisfied at $\mathbf{s} = \mathbf{s}_{cv}$.*

Proof. Please see Appendix A. $\square$

Proposition 4 presents a sufficient condition for improvability that is based solely on the first and second derivatives of the functions $G_{\hat{\phi}}$ and H. In the following, sufficient conditions for non-improvability of the correct decision performance over the conventional system are derived. Similar to the previous discussion, we present two approaches, the first one is based on an auxiliary function and the latter facilitates direct evaluation.

In light of the definitions of G_{MAP} and H given at the beginning of Section 3, we note the following observations. We recall that S_h is defined as $S_h = \{\mathbf{s}: H(\mathbf{s}) = h\}$. A set of values g of G_{MAP} can be obtained correspondingly by $g = G_{\text{MAP}}(\mathbf{s}) = G_{\text{MAP}}(H^{-1}(h))$. By introducing the joint PDF $p_{\mathbf{S},h}(\cdot)$ for the signal distribution in the

---

[4] The average signal power is considered because it is recalled that the expectation is taken over the prior probabilities of all M symbols.

[5] $\boldsymbol{\Pi} = \text{diag}(\underbrace{\pi_1, \ldots, \pi_1}_{N \text{ times}}, \underbrace{\pi_2, \ldots, \pi_2}_{N \text{ times}}, \ldots, \underbrace{\pi_M, \ldots, \pi_M}_{N \text{ times}})$.

h domain, the upper bound for the original optimization problem given in (9) can be equivalently expressed as

$$\max_{p_{\mathbf{S},h}} \int_0^\infty g p_{\mathbf{S},h}(h)\, dh \qquad \text{subject to} \qquad \int_0^\infty h p_{\mathbf{S},h}(h)\, dh \leqslant \mathrm{A}. \qquad (25)$$

In fact, a finite upper limit can be used in the integrals instead of infinity in practical scenarios [1]. Next, an auxiliary function is defined as follows.

$$\mathcal{F}(h) = \max_{\mathbf{s} \in \mathsf{S}_h} G_{\mathrm{MAP}}(\mathbf{s}) \qquad (26)$$

where $\mathcal{F}(h)$ represents the maximum probability of correct decision over all the channels when a deterministic signal constellation with average power h is employed for symbol transmission and decoding is performed using the corresponding optimal MAP detector for each channel.

Proposition 5. *If there exists a non-decreasing concave function $\Upsilon(h)$ that satisfies $\Upsilon(h) \geqslant \mathcal{F}(h)$, $\forall h$ and $\Upsilon(\mathrm{A}) = \mathcal{F}(\mathrm{A}) = \mathrm{P}_c^{\mathrm{cv}}$, then neither channel switching nor stochastic signaling can improve the probability of correct decision.*

Proof. Consider the optimization problem given in (25), which is an upper bound on the original optimization problem as mentioned before. Suppose that a signaling scheme characterized with a PDF $p_{\mathbf{S},h}(h)$ is employed. Let the corresponding average probability of correct decision be denoted by $\mathrm{P}_c(p_{\mathbf{S},h})$. Similar to [1, Theorem 2], we have

$$\mathrm{P}_c(p_{\mathbf{S},h}) = \int_0^\infty g p_{\mathbf{S},h}(h)\, dh$$

$$\leqslant \int_0^\infty o^\infty \mathcal{F}(h) p_{\mathbf{S},h}(h)\, dh,$$

$$\text{from (26)}$$

$$\leqslant \int_0^\infty \Upsilon(h) p_{\mathbf{S},h}(h)\, dh,$$

from the assumption in the proposition

$$\leqslant \Upsilon\!\left(\int_0^\infty h p_{\mathbf{S},h}(h)\, dh \right),$$

due to the concavity of $\Upsilon(h)$

$$\leqslant \Upsilon(\mathrm{A}) = \mathrm{P}_c^{\mathrm{cv}},$$

$$\Upsilon(h) \text{ is non-decreasing.} \qquad (27)$$

Since the above inequality holds for all possible signal distributions satisfying the average power constraint, it is also valid for the optimal signal PDF that maximizes the optimization problem given in (25). Hence, (27) is an upper bound on the original optimization problem given in (6), and it is concluded that the communication system is nonimprovable. $\square$

Although the sufficient condition for non-improvability suggested in Proposition 5 relies on single-variable functions $\Upsilon(h)$ and $\mathcal{F}(h)$, they are not easy to obtain in general. The following condition depends directly on $G_{\mathrm{MAP}}(\mathbf{s})$ and $H(s)$.

Proposition 6. *Let $\mathcal{C}$ denote the convex hull of the set of all possible values for the signal constellation vector $\mathbf{s}$. Suppose also that $H(\mathbf{s}) \leqslant \mathrm{A}$ implies $G_{\mathrm{MAP}}(\mathbf{s}) \leqslant \mathrm{P}_c^{\mathrm{cv}}$ for all $\mathbf{s} \in \mathcal{C}$. If $G_{\mathrm{MAP}}(\mathbf{s})$ is a concave function over $\mathcal{C}$, then the communications system is nonimprovable.*

Proof. Please see Appendix B. $\square$

In this section, sufficient conditions are provided for the improvability and non-improvability of the correct decision performance via stochastic signaling or channel switching over a conventional communications scenario, in which the transmitter employs a fixed constellation at the average transmit signal power over the channel that results in the highest correct decision probability while the receiver employs the corresponding MAP detector. At this point, it should be recalled that a total of $K(K+1)/2$ optimization problems, each having $2MN+1$ optimization variables, are needed to be solved to obtain the optimal signaling strategy, where K is the number of available channels, M is the number of signals in the constellation, and N is the constellation dimensionality. Therefore, before attempting to solve these optimization problems, which are not necessarily convex, it would be helpful to know in advance whether an improvement via stochastic signaling or channel switching is possible. The sufficient conditions presented in Propositions 2–6 are derived mainly for this reason. If an improvability condition is satisfied, we can start searching for the optimal strategy by solving the proposed optimization problems. On the contrary, if a non-improvability condition is satisfied, it indicates that the correct decision performance cannot be improved via stochastic signaling or channel switching. In this case, the sufficient condition lets us know beforehand that the optimal strategy is to employ deterministic signaling over a single channel at the average transmit signal power. Therefore, no performance improvement is possible via stochastic signaling or channel switching, and there is no reason to solve the proposed optimization problems. In summary, by checking the sufficient conditions, it may be possible to identify whether an optimal solution that involves stochastic signaling or channel switching exists or not. However, if such a solution exists (i.e., one of the improvability conditions is satisfied), it is necessary to solve the proposed optimization problem to find that solution. Compared with the computational complexity of obtaining the optimal solution, the sufficient conditions can be much easier to check depending on the exact form of the channel noise PDFs. For example, in Propositions 3 and 4, it is sufficient to find a feasible point that satisfies the given conditions in order to conclude that the performance is improvable. On the contrary, Proposition 5 relies on the existence of a non-decreasing concave function that satisfies a certain condition, which may be relatively harder to find.

4. Numerical results

In this section, numerical examples are presented to illustrate the performance of the proposed signaling strategies in the presence of multiple channels. A scalar binary communication system with equiprobable information symbols is considered and the average power limit is set to $\mathrm{A} = 1$. It is assumed that $K \geqslant 2$ channels are available between the transmitter and the receiver, and only one of them can be used for transmission at any given time. The following four strategies are considered for performance comparison.

Gaussian solution over the best channel: In this approach, antipodal signals $\{-\sqrt{\mathrm{A}}, \sqrt{\mathrm{A}}\}$ are transmitted for binary information symbols over the most favorable channel, i.e., the one that yields the highest probability of correct decision, and the corresponding MAP detector is employed at the receiver. Since deterministic antipodal signaling is optimal in the presence of Gaussian noise (not
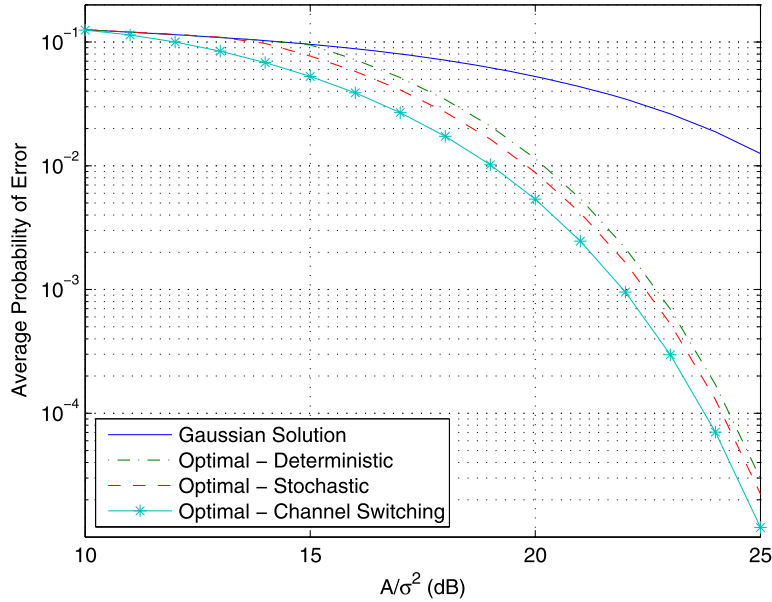
Fig. 3. Average probability of error versus A/σ^2 for various strategies, where $L = 3$ and $\boldsymbol{\mu} = [-0.9\ 0\ 0.9]$ for the Gaussian mixture noise.

necessarily optimal for other types of noise), this approach is called Gaussian solution over the best channel.

Optimal deterministic solution over the best channel: In this scheme, the optimal deterministic signal constellation and the corresponding MAP decision rule are obtained to maximize the probability of correct decision in the absence of stochastic signaling and channel switching. K optimization problems in the form of (14) are solved and the most favorable channel is employed for symbol transmission.

Optimal stochastic solution over the best channel: This scheme employs a single MAP detector at the receiver and randomizes between at most two signal constellations. The optimization problem in (16) is solved for all K channels and the most favorable channel is selected for symbol transmission as shown in (19).

Optimal channel switching with deterministic signaling: In this scheme, switching is performed between at most two channels with deterministic signaling over each channel. $K(K - 1)/2$ optimization problems in the form of (18) are solved and the most favorable channel pair is selected as shown in (20).

It should be noted that the maximum of the last two strategies constitute the solution to the optimal signaling and detector design problem in the presence of multiple channels, as stated in (21).

In the following numerical examples, it is assumed that the channel noise is modeled by a Gaussian mixture distribution [1, 5,32,34], which is represented by

$$p_{N^{(i)}}(n) = \frac{1}{\sqrt{2\pi}\,\sigma_i L_i} \sum_{l=1}^{L_i} \exp\left\{-\frac{(n - \mu_l^{(i)})^2}{2\sigma_i^2}\right\} \quad (28)$$

for $i \in \{1, \ldots, K\}$, where L_i is the number of components in the mixture for channel i. As noted from (28), the components of the Gaussian mixture noise have the same weight $1/L_i$ and the same variance σ_i^2. For notational simplicity, the component means of the Gaussian mixture for channel i are collected in the vector $\boldsymbol{\mu}^{(i)} = [\mu_1^{(i)} \ \ldots \ \mu_{L_i}^{(i)}]$. Based on (28), the average noise power of the ith channel can be calculated as $\mathbb{E}\{|N^{(i)}|^2\} = \sigma_i^2 + \frac{1}{L_i}\|\boldsymbol{\mu}^{(i)}\|_2^2$, where $\|\boldsymbol{\mu}^{(i)}\|_2$ denotes the L_2 norm of vector $\boldsymbol{\mu}^{(i)}$.

First, we consider a scenario in which $K \geqslant 2$ identical channels (i.e., channels with the same noise PDF) are available; i.e., $\sigma_i = \sigma$, $L_i = L$, and $\boldsymbol{\mu}^{(i)} = \boldsymbol{\mu}$, $\forall i \in \{1, \ldots, K\}$, where $\boldsymbol{\mu} = [\mu_1 \ \ldots \ \mu_L]$. Since

identical channels are considered and at most two channels are required for the optimum solution as discussed in Section 2, K can be any number that is larger than or equal to 2. Hence, the results in this part are valid for all $K \geqslant 2$. In Fig. 3, the average probabilities of error corresponding to the four strategies discussed above are plotted versus A/σ^2 for $L = 3$ and $\boldsymbol{\mu} = [-0.9\ 0\ 0.9]$. From the figure, it is observed that the Gaussian solution has the worst performance among all the approaches as expected since it is optimized for Gaussian noise and is not expected to achieve good performance in the presence of multimodal channel noise. When optimal deterministic signaling is employed, significant gains can be achieved over the Gaussian solution in this example. In addition, further improvements are possible when stochastic signaling is used instead of deterministic signaling. As A/σ^2 increases, the overlap between the class conditional PDFs corresponding to binary symbols decreases and there is more room in the signal space for performance improvement via randomized approaches. Overall, the best performance is achieved when switching is performed between two MAP detectors corresponding to two signal constellations. Since identical channels are considered in this example, channel switching can also be regarded as detector randomization via time-sharing for this scenario [13]. Furthermore, the performance of detector randomization is guaranteed to exceed that of stochastic signaling in the case of identical channels, which is also evident from Fig. 3.[6]

In order to further investigate the results in Fig. 3, the parameters for the proposed strategies are presented in Table 1 for some values of A/σ^2. Due to the symmetry of the Gaussian mixture noise, antipodal signaling is employed for binary communications. More explicitly, for optimal deterministic signaling, $\mathbf{s}_0$ and $\mathbf{s}_1$ denote the signals transmitted for information symbols 0 and 1, respectively, and we have $\mathbf{s}_0 = -\mathbf{s}_1$. For optimal stochastic signaling, the optimal signal for information symbol $i \in \{0, 1\}$ is expressed in the form of $p_{\mathbf{S}_i}(\mathbf{s}) = \lambda\delta(\mathbf{s} - \mathbf{s}_{1,i}) + (1 - \lambda)\delta(\mathbf{s} - \mathbf{s}_{2,i})$ with $\mathbf{s}_{1,0} = -\mathbf{s}_{1,1}$ and $\mathbf{s}_{2,0} = -\mathbf{s}_{2,1}$. Finally, the optimal channel switching solution employs the signal pair $\{-\mathbf{s}_1^{(1)}, \mathbf{s}_1^{(1)}\}$ and the corresponding MAP

---

[6] Additional results were obtained for $\boldsymbol{\mu} = [-0.9\ -0.2\ 0.2\ 0.9]$, $\boldsymbol{\mu} = [-0.9\ -0.2\ 0\ 0.2\ 0.9]$, and $\boldsymbol{\mu} = [-1.2\ -0.6\ -0.1\ 0.1\ 0.6\ 1.2]$ as well, and similar observations to those for Fig. 3 were made. The resulting figures are not presented since they are quite similar to Fig. 3.

Table 1
Optimal signal parameters for the scenario in Fig. 3.

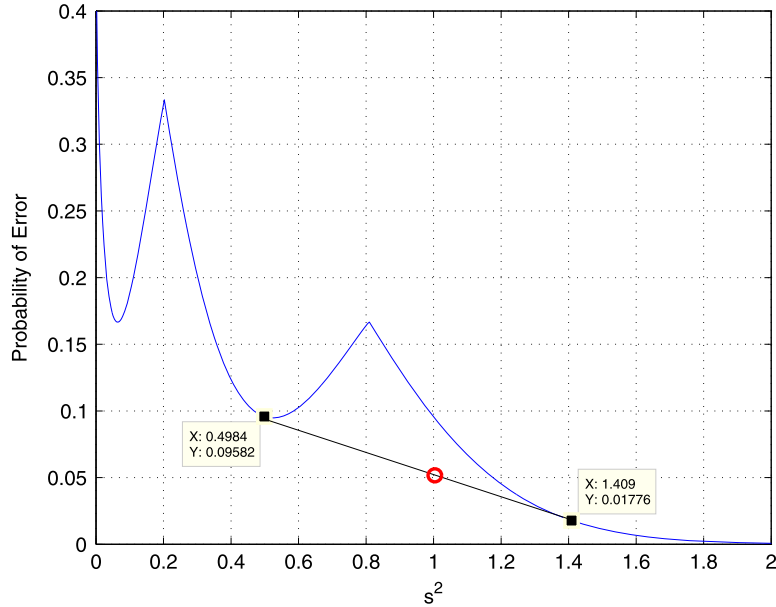| A/σ^2 (dB)	Deterministic signaling	Stochastic signaling			Channel switching		
	$\mathbf{s}_1$	λ	$\mathbf{s}_{1,1}$	$\mathbf{s}_{2,1}$	λ	$\mathbf{s}_1^{(1)}$	$\mathbf{s}_1^{(2)}$
10	1	N/A	1	1	0.1533	0.7271	1.0418
15	0.7239	0.7885	0.7160	1.6783	0.4492	0.7060	1.1870
20	0.6904	0.7650	0.6894	1.6456	0.4254	0.6880	1.1790
25	0.6799	0.7482	0.6798	1.6120	0.3843	0.6796	1.1558

Fig. 4. Error probability versus signal power $\mathbf{s}^2$ for the channel characterized by the parameters $L = 3$ and $\boldsymbol{\mu} = [-0.9\ 0\ 0.9]$ and $A/\sigma^2 = 15$ dB (cf. Fig. 3 and Table 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

detector with probability λ, and the signal pair $\{-\mathbf{s}_1^{(2)}, \mathbf{s}_1^{(2)}\}$ and the corresponding MAP detector with probability $1 - \lambda$. From Table 1, it is observed that all the solutions converge to the Gaussian solution as the noise variance increases. This is due to the fact that the Gaussian mixture noise approximates a unimodal PDF at high values of the variance for which the Gaussian solution is optimal. However, as the noise variance decreases (i.e., A/σ^2 increases), the multimodal nature of the noise PDF prevails and the best performance is achieved by the optimal channel switching solution.

The results depicted in Fig. 3 and Table 1 can also be verified by plotting the error probability of the optimal MAP detector as a function of the signal power in the presence of deterministic antipodal signaling, i.e., $\mathbf{s}_1 = -\mathbf{s}_0 = \mathbf{s}$. This is shown in Fig. 4 for the channel characterized by the parameters $L = 3$, $\boldsymbol{\mu} = [-0.9\ 0\ 0.9]$ and $A/\sigma^2 = 15$ dB, where $A = 1$ as specified before. Due to multimodal noise, the error probability is a nonmonotonic and nonconvex function of the signal power [3,9]. From Fig. 4, it is seen that the optimal deterministic solution is obtained as $\mathbf{s}_1 = -\mathbf{s}_0 = \sqrt{0.524} = 0.7239$, which corresponds to the minimum value (0.0948) of the error probability curve for $\mathbf{s}^2 \leqslant 1$. The best performance is achieved by switching between two power levels 0.4984 and 1.409 using the corresponding antipodal signal pairs $\{-0.7060, 0.7060\}$ and $\{-1.1870, 1.1870\}$, which are in compliance with Table 1. Also, the switching factor λ can be calculated based on the average power limit, $A = 1$, as follows: $0.4984\lambda + 1.409(1 - \lambda) = 1$, which yields $\lambda = 0.4492$ as in Table 1. It is observed from Fig. 4 that switching between two MAP detectors can reduce the average probability of error down to nearly 0.05, which is indicated by the red circle in the figure.

Next, we consider a scenario in which all the channels have distinct noise PDFs. In this case, the best performance can be achieved

by either the optimal channel switching with deterministic signaling approach or the optimal stochastic solution over the best channel approach. For the Gaussian mixture noise model in (28), it is assumed that $\sigma_i = \sigma$ and $L_i = L$, $\forall i \in \{1, \ldots, K\}$, and that the component means of the Gaussian mixture are chosen as

$$\boldsymbol{\mu}^{(i)} = \sqrt{E}\,\frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2} \qquad (29)$$

for $i = 1, \ldots, K$, where E is a constant and $\mathbf{v}_i$'s are L-dimensional distinct vectors. It is noted that $\|\boldsymbol{\mu}^{(i)}\|_2^2 = E$. Hence, the average noise power is the same for all the channels. Namely, $\mathbb{E}\{|N^{(i)}|^2\} = \sigma^2 + \frac{E}{L}$, $\forall i \in \{1, \ldots, K\}$. In Fig. 5, the average probabilities of error for the four strategies are plotted versus A/σ^2 for $K = 3$, $\mathbf{v}_1 = [-3\ -2\ 0\ 2\ 3]$, $\mathbf{v}_2 = [-4\ -3\ 0\ 3\ 4]$, $\mathbf{v}_3 = [-5\ -3\ 0\ 3\ 5]$, and $E = 3$. From Fig. 5, it is concluded that the optimal channel switching strategy achieves the lowest average probability of error and the Gaussian solution has the worst performance over the whole range of A/σ^2 values.

The optimal parameters of the strategies in Fig. 5 are shown for some values of A/σ^2 in Table 2. For the Gaussian solution and the optimal deterministic solution, the channel that results in the lowest probability of error is indicated in the first column of the respective area in the table and the second column specifies the scalar signal value employed for the transmission of information symbol 1. Again, antipodal signals are considered for symbol 0 and symbol 1. It is observed that either channel 2 or channel 3 is employed for these solutions depending on the noise level. For the optimal stochastic solution, the same notation is employed as in Table 1 together with the channel index employed for communications. In the case of optimal channel switching, Table 2 shows the two channels between which switching is performed
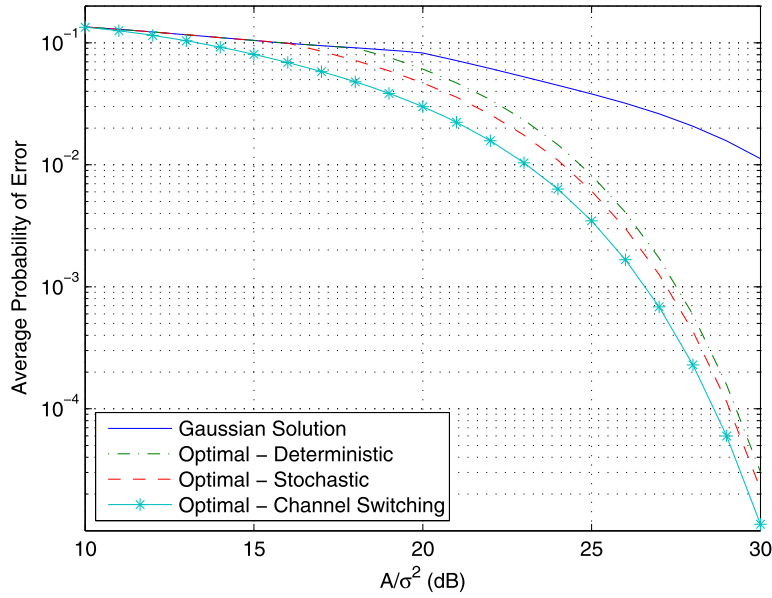
Fig. 5. Average probability of error versus A/σ^2 for various approaches, where $K = 3$, $\mathbf{v}_1 = [-3\ -2\ 0\ 2\ 3]$, $\mathbf{v}_2 = [-4\ -3\ 0\ 3\ 4]$, $\mathbf{v}_3 = [-5\ -3\ 0\ 3\ 5]$, and $E = 3$ (see (29)).

Table 2
Optimal signal parameters for the scenario in Fig. 5.

A/σ^2 (dB)	Gaussian solution		Deterministic signaling		Stochastic signaling				Channel switching			
	Channel	$\mathbf{s}_1$	Channel	$\mathbf{s}_1$	Channel	λ	$\mathbf{s}_{1,1}$	$\mathbf{s}_{2,1}$	λ	$\mathbf{s}_1^{(1)}$	$\mathbf{s}_1^{(2)}$	$\mathbf{s}_1^{(3)}$
10	2	1	2	1	2	N/A	1	1	0.8450	1.0601	0.5697	X
15	2	1	2	1	2	0.0502	1.0078	0.9996	0.5642	1.202	0.6509	X
20	3	1	2	0.6405	2	0.7547	0.6381	1.6805	0.5614	1.2108	0.6353	X
25	3	1	2	0.6213	2	0.7348	0.6210	1.6439	0.6023	1.1848	0.6206	X
30	3	1	2	0.6152	2	0.7222	0.6152	1.6174	0.6369	1.1638	0.6151	X

(the "X" mark indicates that the corresponding channel is not utilized). As an example, for $A/\sigma^2 = 20$ dB in Fig. 5, the optimal channel switching strategy transmits over channel 1 using the constellation $\{-1.2108, 1.2108\}$ with probability 0.5614 (i.e., 56.14% of the time), and transmits over channel 2 using the constellation $\{-0.6353, 0.6353\}$ with probability 0.4386. Since the average noise power is the same for all channels, the optimal parameters for each strategy are determined by the variance and the means of the Gaussian mixture components. In order to determine the improvability in this scenario, the conditions in Proposition 3 can be evaluated. For example, at $A/\sigma^2 = 25$ dB, the calculations show that the improvability conditions in Proposition 3 are satisfied for $\mathbf{s}_1 = 1.45$ and $\mathbf{s}_2 = 0.95$.

In order to illustrate the improvements via channel switching, Fig. 6 presents the error probabilities of the three channels considered in Fig. 5 and Table 2 as a function of the signal power in the presence of antipodal signaling when $A/\sigma^2 = 15$ dB. As shown in the figure, the optimal channel switching strategy performs time sharing between Channel 1 and Channel 2 with power levels 1.445 and 0.4238 (i.e., signal constellations $\{-1.202, 1.202\}$ and $\{-0.6509, 0.6509\}$), respectively. The results are in compliance with Table 2, as expected. It should also be noted that a lower average probability of error can be achieved for the scenario in Fig. 6 if detector randomization is allowed for each channel; that is, if multiple detectors can be implemented and time shared for the detection of symbols acquired over each channel. In that case, a randomization between two constellations and the corresponding MAP detectors over Channel 2 can result in a lower average probability of error. Fortunately, as previously stated in footnote 2, such scenarios can be covered using the proposed framework in

this study by considering multiple channels with identical distributions.

Finally, a scenario with just two channels is considered. The parameters of the first channel are given by $\mathbf{v}_1 = [-6\ -3\ -2\ 2\ 3\ 6]$, $L_1 = 6$, and $E = 4$ (see (29)). The second channel is modeled to have zero-mean Gaussian noise with the same average power as the first one; i.e., $L_2 = 1$, $\mu^{(2)} = 0$, and $\sigma_2^2 = \sigma_1^2 + \frac{E}{L_1}$ in (28). The average probabilities of error for the proposed strategies are plotted versus A/σ_1^2 in Fig. 7. Unlike the cases in Fig. 3 and Fig. 5, the best performance is achieved by stochastic signaling over the best channel in this scenario. It should be emphasized that the possibility of an optimal solution in the form of stochastic signaling is stated in Section 2 (see (19)–(21)). It is also observed that the optimal channel switching strategy performs very closely to the optimal deterministic signaling approach over the best channel. In other words, channel switching does not provide significant performance improvements in this scenario. The optimal parameters of the strategies depicted in Fig. 7 are presented for some values of A/σ_1^2 in Table 3.

5. Concluding remarks

Optimal signaling and detector design have been studied under an average transmit power constraint for generic noise distributions in the presence of multiple channels and stochastic signaling. It has been shown that the optimal solution to the joint channel switching, stochastic signaling, and detector design problem corresponds to one of the following strategies: (i) deterministic signaling over a single channel, (ii) randomizing (time sharing) between at most two signal constellations over a single
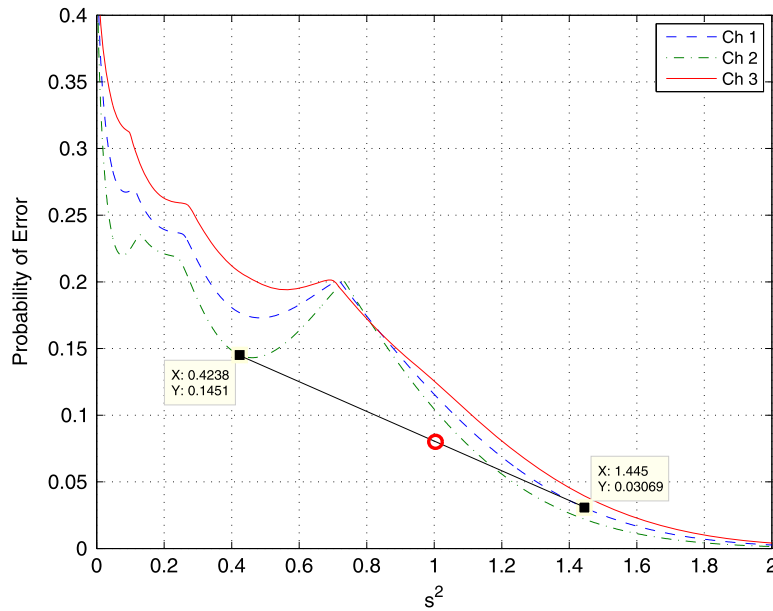
Fig. 6. Error probability versus signal power $\mathbf{s}^2$ for the three channels when $A/\sigma^2 = 15$ dB (cf. Fig. 5 and Table 2).
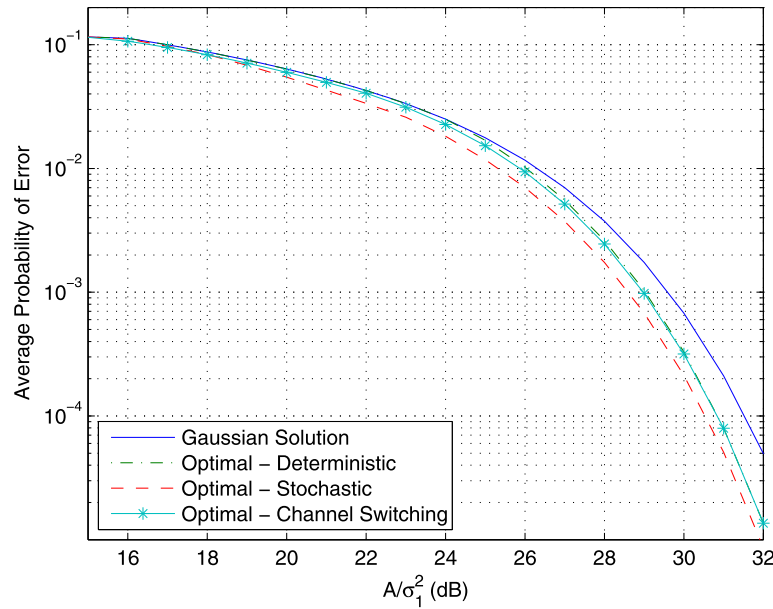

Fig. 7. Average probability of error versus A/σ_1^2 for various approaches, where the first channel is characterized by the parameters $K = 2$, $\mathbf{v}_1 = [-6\ -3\ -2\ 2\ 3\ 6]$, $E = 4$ (see (29)), and the second channel has zero-mean Gaussian noise with the same average power as the first channel.

Table 3
Optimal signal parameters for the scenario in Fig. 7.

A/σ_1^2 (dB)	Gaussian solution		Deterministic signaling		Stochastic signaling				Channel switching		
	Channel	$\mathbf{s}_1$	Channel	$\mathbf{s}_1$	Channel	λ	$\mathbf{s}_{1,1}$	$\mathbf{s}_{2,1}$	λ	$\mathbf{s}_1^{(1)}$	$\mathbf{s}_1^{(2)}$
15	2	1	2	1	2	N/A	1	1	0.1823	0.6683	1.0599
20	1	1	1	1	1	0.0857	0.2068	1.0439	0.9134	1.0266	0.6576
25	1	1	1	0.6963	1	0.6725	0.6964	1.4344	0.8810	0.6961	2.1951
30	1	1	1	0.7037	1	0.6378	0.7037	1.3743	0.9495	0.7037	3.2388

channel, or (iii) switching (time sharing) between at most two channels with deterministic signaling over each channel. For all cases, the optimal strategies employ the corresponding MAP detectors at the receiver. Optimization problems have been formulated to obtain the parameters of the proposed strategies. In addition, sufficient conditions have been provided to specify whether or not the proposed strategy can improve the error performance over the conventional approach, in which a single channel is employed with deterministic signaling at the average power limit. Various numerical examples have been presented to illustrate the theoretical results. It has been observed that significant performance improvements can be achieved in some cases via the proposed optimal approach in the presence of multimodal noise.

Acknowledgments

The authors would like to thank Prof. Orhan Arikan from Bilkent University for his insightful comments.

Appendix A. Proof of Proposition 4

Consider a signaling scheme with infinitesimally small perturbations around the conventional signal constellation, $p_\mathbf{S}(\mathbf{s}) = \lambda\delta(\mathbf{s} - (\mathbf{s}_{cv} + \mathbf{\Delta}_1)) + (1 - \lambda)\delta(\mathbf{s} - (\mathbf{s}_{cv} + \mathbf{\Delta}_2))$. A sufficient set of conditions for improvability can then be expressed as

$$\lambda G_{\hat{\phi}}(\mathbf{s}_{cv} + \mathbf{\Delta}_1) + (1 - \lambda)G_{\hat{\phi}}(\mathbf{s}_{cv} + \mathbf{\Delta}_2) > P_c^{cv},$$
$$\lambda H(\mathbf{s}_{cv} + \mathbf{\Delta}_1) + (1 - \lambda)H(\mathbf{s}_{cv} + \mathbf{\Delta}_2) \leqslant A. \tag{A.1}$$

A second-order approximation for $G_{\hat{\phi}}(\mathbf{s}_{cv} + \mathbf{\Delta}_k)$ and $H(\mathbf{s}_{cv} + \mathbf{\Delta}_k)$ can be obtained using the Taylor series expansion as $G_{\hat{\phi}}(\mathbf{s}_{cv} + \mathbf{\Delta}_k) \approx G_{\hat{\phi}}(\mathbf{s}_{cv}) + \mathbf{J}_G(\mathbf{s}_{cv})\mathbf{\Delta}_k + 0.5\mathbf{\Delta}_k^T\mathbf{H}_G(\mathbf{s}_{cv})\mathbf{\Delta}_k$ and $H(\mathbf{s}_{cv} + \mathbf{\Delta}_k) \approx H(\mathbf{s}_{cv}) + \mathbf{J}_H(\mathbf{s}_{cv})\mathbf{\Delta}_k + 0.5\mathbf{\Delta}_k^T\mathbf{H}_H(\mathbf{s}_{cv})\mathbf{\Delta}_k$ respectively, where $\mathbf{J}_G(\mathbf{s}_{cv})$ and $\mathbf{H}_G(\mathbf{s}_{cv})$ (similarly $\mathbf{J}_H(\mathbf{s}_{cv})$ and $\mathbf{H}_H(\mathbf{s}_{cv})$) are the gradient and Hessian of $G_{\hat{\phi}}(\mathbf{s})$ ($H(\mathbf{s})$) evaluated at $\mathbf{s} = \mathbf{s}_{cv}$, respectively. From (A.1),

$$\lambda\mathbf{\Delta}_1^T\mathbf{H}_G(\mathbf{s}_{cv})\mathbf{\Delta}_1 + (1 - \lambda)\mathbf{\Delta}_2^T\mathbf{H}_G(\mathbf{s}_{cv})\mathbf{\Delta}_2$$
$$+ 2\mathbf{J}_G(\mathbf{s}_{cv})[\lambda\mathbf{\Delta}_1 + (1 - \lambda)\mathbf{\Delta}_2] > 0, \tag{A.2}$$
$$\lambda\mathbf{\Delta}_1^T\mathbf{H}_H(\mathbf{s}_{cv})\mathbf{\Delta}_1 + (1 - \lambda)\mathbf{\Delta}_2^T\mathbf{H}_H(\mathbf{s}_{cv})\mathbf{\Delta}_2$$
$$+ 2\mathbf{J}_H(\mathbf{s}_{cv})[\lambda\mathbf{\Delta}_1 + (1 - \lambda)\mathbf{\Delta}_2] \leqslant 0. \tag{A.3}$$

Let $\mathbf{\Delta}_1 = \nu\mathbf{x}$ and $\mathbf{\Delta}_2 = \omega\mathbf{x}$, where ν and ω are infinitesimal real numbers, and $\mathbf{x}$ is an MN-dimensional real vector. Using the definitions from the statement of the proposition, the conditions in (A.2) can be expressed, after some manipulation[7] as

$$\left(g^{(2)}(\mathbf{s}, \mathbf{x}) + \ell \cdot g^{(1)}(\mathbf{s}, \mathbf{x})\right)\big|_{\mathbf{s} = \mathbf{s}_{cv}} > 0, \tag{A.4}$$
$$\left(h^{(2)}(\mathbf{x}) + \ell \cdot h^{(1)}(\mathbf{s}, \mathbf{x})\right)\big|_{\mathbf{s} = \mathbf{s}_{cv}} \leqslant 0, \tag{A.5}$$

where $\ell \triangleq 2(\lambda(\nu - \omega) + \omega)/(\lambda(\nu^2 - \omega^2) + \omega^2)$. By varying λ in the interval $(0, 1)$ and choosing appropriate values for infinitesimal quantities ν and ω, any real value can be assigned to ℓ. The first part of the proposition states $g^{(1)}(\mathbf{s}, \mathbf{x})h^{(1)}(\mathbf{s}, \mathbf{x}) < 0$ at $\mathbf{s} = \mathbf{s}_{cv}$, meaning that $g^{(1)}(\mathbf{s}_{cv}, \mathbf{x})$ and $h^{(1)}(\mathbf{s}_{cv}, \mathbf{x})$ must have different signs. Under this condition, it is easy to see that a suitable choice of ℓ satisfies the requirements given in (A.4) and (A.5). Namely, any choice of ℓ with a sufficiently high absolute value and the correct sign is adequate. In the second part of the proposition, it is assumed that $g^{(1)}(\mathbf{s}, \mathbf{x})$ and $h^{(1)}(\mathbf{s}, \mathbf{x})$ are positive at $\mathbf{s} = \mathbf{s}_{cv}$. Multiplying both terms in (A.4) with $h^{(1)}(\mathbf{s}, \mathbf{x})$, and similarly multiplying both terms in (A.5) with $g^{(1)}(\mathbf{s}, \mathbf{x})$, an equivalent condition for the improvability can be written as

$$\left(g^{(2)}(\mathbf{s}, \mathbf{x})h^{(1)}(\mathbf{s}, \mathbf{x}) + \ell \cdot h^{(1)}(\mathbf{s}, \mathbf{x})g^{(1)}(\mathbf{s}, \mathbf{x})\right)\big|_{\mathbf{s} = \mathbf{s}_{cv}} > 0, \tag{A.6}$$
$$\left(h^{(2)}(\mathbf{x})g^{(1)}(\mathbf{s}, \mathbf{x}) + \ell \cdot h^{(1)}(\mathbf{s}, \mathbf{x})g^{(1)}(\mathbf{s}, \mathbf{x})\right)\big|_{\mathbf{s} = \mathbf{s}_{cv}} \leqslant 0. \tag{A.7}$$

Notice that the second terms in (A.6) and (A.7) are the same. Under the condition of $g^{(2)}(\mathbf{s}, \mathbf{x})h^{(1)}(\mathbf{s}, \mathbf{x}) > h^{(2)}(\mathbf{x})g^{(1)}(\mathbf{s}, \mathbf{x})$ at $\mathbf{s} = \mathbf{s}_{cv}$, meaning that the first term in (A.6) is greater than the first term in (A.7), an appropriate value of ℓ can always be found such that the improvability conditions in (A.6) and (A.7) are satisfied. $\square$

---

[7] Recall that $H(\mathbf{s}) = \sum_{j=0}^{M-1} \pi_j \|\mathbf{s}_j\|_2^2$ where $\mathbf{s}_j$ represents the signal vector transmitted for the jth symbol. Then, the gradient and the Hessian are given as $\mathbf{J}_H(\mathbf{s}) = 2\mathbf{s}^T\mathbf{\Pi}$ and $\mathbf{H}_H(\mathbf{s}) = 2\mathbf{\Pi}$, respectively. Since the only nonzero entries of $\mathbf{\Pi}$ are the prior probabilities on the diagonal, the Hessian of $H(\mathbf{s})$ is a constant and positive definite matrix that is independent of $\mathbf{s}$.

Appendix B. Proof of Proposition 6

Since $H(\mathbf{s})$ is a convex function for all values of $\mathbf{s}$, Jensen's inequality implies that for any given signal distribution $p_\mathbf{S}$, $\mathbb{E}\{H(\mathbf{S})\} \geqslant H(\mathbb{E}\{\mathbf{S}\})$, which in turn implies that $H(\mathbb{E}\{\mathbf{S}\}) \leqslant A$ due to the average power constraint. Since $\mathcal{C}$ is convex, $\mathbb{E}\{\mathbf{S}\} \in \mathcal{C}$. Then, from the assumption in the proposition, $H(\mathbb{E}\{\mathbf{S}\}) \leqslant A$ implies that $G_{MAP}(\mathbb{E}\{\mathbf{S}\}) \leqslant P_c^{cv}$. Since $G_{MAP}(\mathbf{s})$ is a concave function over $\mathcal{C}$, $\mathbb{E}\{G_{MAP}(\mathbf{S})\} \leqslant G_{MAP}(\mathbb{E}\{\mathbf{S}\}) \leqslant P_c^{cv}$ for any given distribution $p_\mathbf{S}$. Consequently, the previous inequality also holds for the optimal signal distribution obtained as the solution of the optimization problem given in (9), which is itself an upper bound on the solution of the original optimization problem given in (6). Hence, it is concluded that under the conditions in the proposition, the communication system is nonimprovable.

References

[1] H. Chen, P.K. Varshney, S.M. Kay, J.H. Michels, Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors, IEEE Trans. Signal Process. 55 (7) (July 2007) 3172–3184.

[2] H. Chen, P.K. Varshney, Theory of the stochastic resonance effect in signal detection: Part II—Variable detectors, IEEE Trans. Signal Process. 56 (10) (Oct. 2007) 5031–5041.

[3] A. Patel, B. Kosko, Optimal noise benefits in Neyman–Pearson and inequality-constrained signal detection, IEEE Trans. Signal Process. 57 (5) (May 2009) 1655–1669.

[4] S. Bayram, S. Gezici, Noise-enhanced M-ary hypothesis-testing in the minimax framework, in: Proc. Int. Conf. Signal Process. Commun. Systems, Omaha, Nebraska, September 2009, pp. 31–36.

[5] S. Bayram, S. Gezici, H.V. Poor, Noise enhanced hypothesis-testing in the restricted Bayesian framework, IEEE Trans. Signal Process. 58 (8) (August 2010) 3972–3989.

[6] M. Azizoglu, Convexity properties in binary detection problems, IEEE Trans. Inform. Theory 42 (4) (July 1996) 1316–1321.

[7] S. Loyka, V. Kostina, F. Gagnon, Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications, IEEE Trans. Inform. Theory 56 (4) (April 2010) 1948–1960.

[8] C. Goken, S. Gezici, O. Arikan, Optimal stochastic signaling for power-constrained binary communications systems, IEEE Trans. Wirel. Commun. 9 (12) (December 2010) 3650–3661.

[9] C. Goken, S. Gezici, O. Arikan, Optimal signaling and detector design for power-constrained binary communications systems over non-Gaussian channels, IEEE Commun. Lett. 14 (2) (Feb. 2010) 100–102.

[10] M.D. McDonnell, Is electrical noise useful?, Proc. IEEE 99 (2) (Feb. 2011) 242–246.

[11] B. Dulek, S. Gezici, Optimal signaling and detector design for power constrained on-off keying systems in Neyman–Pearson framework, in: 16th IEEE Wksp. Stat. Signal Process. (SSP), Nice, France, 28–30 June 2011, pp. 93–96.

[12] B. Dulek, S. Gezici, Optimal stochastic signal design and detector randomization in the Neyman–Pearson framework, in: 37th IEEE Int. Conf. Acous., Speech, Signal Process. (ICASSP), Kyoto, Japan, 25–30 Mar. 2012, pp. 3025–3028.

[13] B. Dulek, S. Gezici, Detector randomization and stochastic signaling for minimum probability of error receivers, IEEE Trans. Commun. 60 (4) (April 2012) 923–928.

[14] B. Dulek, S. Gezici, O. Arikan, Convexity properties of detection probability under additive Gaussian noise: Optimal signaling and jamming strategies, IEEE Trans. Signal Process. 61 (13) (July 2013) 3303–3310.

[15] M.E. Tutay, S. Gezici, O. Arikan, Optimal detector randomization for multiuser communications systems, IEEE Trans. Commun. 61 (7) (July 2013) 2876–2889.

[16] B. Dulek, N.D. Vanli, S. Gezici, P.K. Varshney, Optimum power randomization for the minimization of outage probability, IEEE Trans. Wirel. Commun. 12 (9) (Sep. 2013) 4627–4637.

[17] H.V. Poor, An Introduction to Signal Detection and Estimation, Springer-Verlag, New York, 1994.

[18] C. Goken, S. Gezici, O. Arikan, Stochastic signaling in the presence of channel state information uncertainty, Digital Signal Process. 23 (2) (2013) 635–645, http://dx.doi.org/10.1016/j.dsp.2012.10.004.

[19] E.G. Larsson, Improving the frame-error-rate of spatial multiplexing in block fading by randomly rotating the signal constellation, IEEE Commun. Lett. 8 (8) (August 2004) 514–516.

[20] E.G. Larsson, Constellation randomization (CoRa) for outage performance improvement on MIMO channels, in: IEEE Global Commun. Conf., vol. 1, Nov.–Dec. 2004, pp. 386–390.

[21] Y. Li, C.N. Georghiades, G. Huang, Transmit diversity over quasi-static fading channels using multiple antennas and random signal mapping, IEEE Trans. Commun. 51 (11) (Nov. 2003) 1918–1926.

[22] C. Lamy, J. Boutros, On random rotations diversity and minimum MSE decoding of lattices, IEEE Trans. Inform. Theory 46 (July 2000) 1584–1589.

[23] A. Hiroike, F. Adachi, N. Nakajima, Combined effects of phase sweeping transmitter diversity and channel coding, IEEE Trans. Veh. Tech. 41 (May 1992) 170–176.

[24] X. Ma, G.B. Giannakis, Space–time-multipath coding using digital phase sweeping, in: IEEE Global Commun. Conf., vol. 1, Nov. 2002, pp. 384–388.

[25] E.L. Lehmann, Testing Statistical Hypotheses, 2nd ed., Chapman & Hall, New York, 1986.

[26] Y. Ma, C.C. Chai, Unified error probability analysis for generalized selection combining in Nakagami fading channels, IEEE J. Select. Areas Commun. 18 (11) (Nov. 2000) 2198–2210.

[27] J.A. Ritcey, M. Azizoglu, Performance analysis of generalized selection combining with switching constraints, IEEE Commun. Lett. 4 (5) (May 2000) 152–154.

[28] B. Dulek, M.E. Tutay, S. Gezici, P.K. Varshney, Optimal channel switching in the presence of stochastic signaling, in: IEEE Int. Symp. Inform. Theory (ISIT), Istanbul, Turkey, 7–12 July 2013, pp. 2189–2193.

[29] J. Mitola, G.Q. Maguire, Cognitive radio: Making software radios more personal, IEEE Personal Commun. Mag. 6 (4) (August 1999) 13–18.

[30] S. Haykin, Cognitive radio: Brain-empowered wireless communications, IEEE J. Select. Areas Commun. 23 (2) (Feb. 2005) 201–220.

[31] S. Gezici, H. Celebi, H.V. Poor, H. Arslan, Fundamental limits on time delay estimation in dispersed spectrum cognitive radio systems, IEEE Trans. Wirel. Commun. 8 (1) (Jan. 2009) 78–83.

[32] V. Bhatia, B. Mulgrew, Non-parametric likelihood based channel estimator for Gaussian mixture noise, Signal Process. 87 (Nov. 2007) 2569–2586.

[33] S. Verdu, Multiuser Detection, 1st ed., Cambridge University Press, Cambridge, UK, 1998.

[34] T. Erseghe, V. Cellini, G. Dona, On UWB impulse radio receivers derived by modeling MAI as a Gaussian mixture process, IEEE Trans. Wirel. Commun. 7 (6) (June 2008) 2388–2396.

[35] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1968.

[36] J.G. Proakis, Digital Communications, 4th ed., McGraw–Hill, New York, 2001.

[37] S. Bayram, S. Gezici, On the improvability and nonimprovability of detection via additional independent noise, IEEE Signal Process. Lett. 16 (11) (Feb. 2009) 1001–1004.

Berkan Dulek received the B.S., M.S. and Ph.D. degrees in electrical and electronics engineering from Bilkent University in 2003, 2006 and 2012, respectively. From 2007 to 2010, he worked at TÜBİTAK UEKAE – İLTAREN Research and Development Group. From 2012 to 2013, he was a postdoctoral research associate at the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY. His research interests are in statistical signal processing, detection and estimation theory, and wireless communications.

Mehmet Emin Tutay received the B.S. degree in 2008 and the M.S. degree in 2010, both from the Department of Electrical and Electronics Engineering, Bilkent University, Turkey. He is currently working towards the Ph.D. degree in the same department. His main research interests are in the fields of statistical signal processing and wireless communications.

Sinan Gezici received the B.S. degree from Bilkent University, Turkey in 2001, and the Ph.D. degree in electrical engineering from Princeton University in 2006. From 2006 to 2007, he worked at Mitsubishi Electric Research Laboratories, Cambridge, MA. Since 2007, he has been with the Department of Electrical and Electronics Engineering at Bilkent University, where he is currently an Associate Professor. Dr. Gezici's research interests are in the areas of detection and estimation theory, wireless communications, and localization systems. Among his publications in these areas is the book Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols (Cambridge University Press, 2008). Dr. Gezici is an associate editor for IEEE Transactions on Communications, IEEE Wireless Communications Letters, and the Journal of Communications and Networks.

Pramod K. Varshney was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana–Champaign in 1972, 1974, and 1976, respectively.

During 1972–1976, he held teaching and research assistantships at the University of Illinois. Since 1976 he has been with Syracuse University, Syracuse, NY where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He served as the Associate Chair of the department during 1993–1996. He is also an Adjunct Professor of Radiology at Upstate Medical University in Syracuse, NY. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing and remote sensing. He has published extensively. He is the author of Distributed Detection and Data Fusion, published by Springer-Verlag in 1997. He has served as a consultant to several major companies.

While at the University of Illinois, Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the guest editor of the special issue on data fusion of the Proceedings of the IEEE, January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award. He is on the editorial board of Journal on Advances in Information Fusion. He was the President of International Society of Information Fusion during 2001.