

# Average Capacity Maximization via Channel Switching in the Presence of Additive White Gaussian Noise Channels and Switching Delays

Ahmet Dundar Sezer, *Student Member, IEEE*, and Sinan Gezici, *Senior Member, IEEE*

**Abstract**—The optimal channel switching problem is studied for average capacity maximization in the presence of additive white Gaussian noise channels and channel switching delays. First, an optimization problem is formulated for the maximization of the average channel capacity considering channel switching delays and constraints on average and peak powers. Then, an equivalent optimization problem is obtained to facilitate theoretical investigations. The optimal strategy is derived and the corresponding average capacity is specified when channel switching is performed among a given number of channels. Based on this result, it is shown that channel switching among more than two different channels is not optimal. In addition, the maximum average capacity achieved by the optimal channel switching strategy is formulated as a function of the channel switching delay parameter and the average and peak power limits. Then, scenarios under which the optimal strategy corresponds to the exclusive use of a single channel or to channel switching between two channels are described. Furthermore, sufficient conditions are obtained to determine when the optimal single channel strategy outperforms the optimal channel switching strategy. Numerical results are presented to provide examples of the theoretical results and to illustrate effects of channel switching delays.

**Index Terms**—Channel switching, capacity, switching delay, time sharing.

## I. INTRODUCTION

Optimal resource allocation is an important approach for enhancing performance of communication systems. One common metric that is optimized in resource allocation problems is the channel capacity [1]. In [2], the optimal dynamic resource allocation in fading broadcast channels is investigated for code division, time division, and frequency division in the presence of perfect channel side information at the transmitter and the receivers, and ergodic capacity regions are obtained. In [3], an adaptive resource allocation technique is proposed for multiuser orthogonal frequency division multiplexing (OFDM) systems in the presence of proportional fairness constraints among users, and optimal and suboptimal algorithms are designed for sum capacity maximization under constraints on the minimum required data rate for each user. Optimal joint power and channel allocation strategies are studied in [4] for cognitive radio systems, and a near-optimal algorithm is

proposed for the total sum capacity maximization of power-limited secondary users in a centralized cognitive radio system. In [5], capacity maximizing antenna selection is investigated for a multiple-input multiple-output (MIMO) system and low-complexity antenna subset selection algorithms are provided. It is shown that near optimal capacity of a full-complexity system can be achieved by selecting the number of antennas at the receiver to be greater than or equal to the number of antennas at the transmitter. In [6], the optimal antenna selection is studied in correlated channels for both the transmitter and receiver to reduce the number of radio frequency (RF) chains. The proposed algorithm leads to a near-optimal capacity that is achieved without antenna selection. In addition to the capacity, other metrics such as probability of error, probability of detection, and outage probability are considered in various resource allocation problems; e.g., [7]–[15]. For example, in the detector randomization problem, the aim is to minimize the average probability of error of a communication system by optimizing time sharing factors and transmit power (signal) levels corresponding to different detectors at the receiver [7]–[9]. Also, a jammer can maximize the average probability of error or minimize the detection probability of a victim receiver by performing optimal time sharing among multiple power levels [12]–[14]. In [14], the optimal power allocation is performed for an average power constrained jammer to minimize the detection probability of an instantaneously and fully adaptive receiver employing the Neyman-Pearson criterion, and it is shown that the optimal jamming performance is achieved via time sharing between at most two different power levels. In [15], the optimal time sharing of power levels is implemented for minimizing the outage probability in a flat block-fading Gaussian channel under an average power constraint and in the presence of channel distribution information at the transmitter.

In the presence of multiple channels between a transmitter and a receiver, optimal time sharing and power allocation can be implemented for performance improvement [12], [16]–[20]. In other words, channel switching, which involves the use of each channel with a certain fraction of time and a certain power level, can be optimized for enhancing performance of a communication system. In [12], the channel switching problem is studied for the optimal detection of binary antipodal signals under an average power constraint, and it is shown that the optimal strategy is either to communicate over one channel exclusively, or to switch between two channels with a certain time sharing factor. In [18], the channel switching problem is analyzed for an  $M$ -ary communication system over an additive noise channel in the presence of time sharing among multiple

This research was supported in part by the Distinguished Young Scientist Award of Turkish Academy of Sciences (TUBA-GEBIP 2013). Part of this work was presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), April 2016.

A. D. Sezer and S. Gezici are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey, Tel: +90 (312) 290-3139, Fax: +90 (312) 266-4192, e-mails: {adsezer,gezici}@ee.bilkent.edu.tr.

signal constellations over each channel. It is proved that the optimal strategy that minimizes the average probability of error under an average power constraint corresponds to one of the following techniques: deterministic signaling (i.e., use of one signal constellation) over a single channel; time sharing between two different signal constellations over a single channel; or, switching (time sharing) between two channels with deterministic signaling over each channel [18]. The study in [19] investigates the optimal channel switching problem over Gaussian channels in the presence of average power and cost constraints. In particular, each channel is assigned a certain utilization cost, and the average probability of error is minimized in the presence of an average cost constraint. It is shown that the optimal strategy involves the use of at most three different channels [19]. In [20], the optimal channel switching strategy is developed for the maximization of average capacity, and it is stated that the optimal strategy can be realized by channel switching between at most two different channels. Also, a low-complexity optimization problem is presented to obtain the optimal channel switching strategy.

In most of the previous studies on optimal channel switching strategies, delays (costs) associated with the channel switching operation are not considered [12], [16]–[20]. However, due to hardware limitations, the channel switching operation takes a certain time in practice. In particular, when switching to a new channel, the parameters at the transmitter and the receiver are set according to the characteristics (i.e., frequency) of the new channel, which induces a channel switching delay and consequently reduces the available time for data transmission [21], [22]. Most of the studies in the literature omit the channel switching overhead (delay) by assuming that it is negligible due to improved hardware technologies. However, the study in [23] shows that the state-of-the-art algorithms related to scheduling in wireless mesh networks experience performance degradation in the presence of the channel switching latency. Similarly, in [24], the channel switching cost is considered in the design of the energy efficient centralized cognitive radio networks, and an energy efficient heuristic scheduler is proposed to allocate each idle frequency to the cognitive radio with the highest energy efficiency at that frequency. In [25], effects of channel switching time and energy on cooperative sensing scheduling are analyzed for cognitive radio networks. In [26], a spectrum aware routing algorithm for multi-hop cognitive radio networks is proposed with the consideration of the channel switching overhead.

Although the channel switching problem has been investigated from various perspectives, no studies in the literature have considered channel switching for average capacity maximization in the presence of channel switching delays. In this study, the optimal channel switching strategy is proposed for average capacity maximization under power constraints and considering a time delay for each channel switching operation during which data communication cannot be performed. After presenting an optimization theoretic formulation of the proposed problem, an equivalent optimization problem is obtained to facilitate theoretical investigations. It is observed that consideration of channel switching delays leads to significant differences in the formulation and analyses

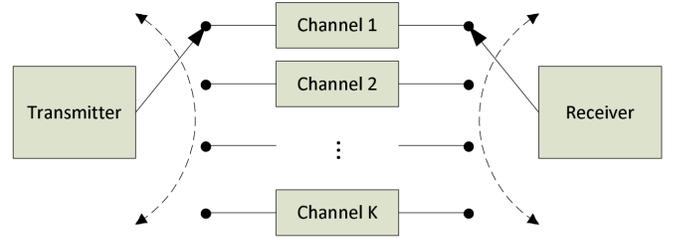


Fig. 1. Block diagram of a communication system in which transmitter and receiver can switch among  $K$  channels.

compared to those obtained by omitting the effects of channel switching delays [20]. First, the optimal strategy is obtained and the corresponding average capacity is specified when channel switching is performed among a given number of channels. Based on this result, it is then shown that channel switching among more than two different channels cannot be optimal. Also, the maximum average capacity achieved by the optimal channel switching strategy is formulated for various values of the channel switching delay parameter and the average and peak power limits. In addition, scenarios under which the optimal strategy corresponds to the utilization of a single channel or to channel switching between two channels are described. Furthermore, sufficient conditions are derived to determine when the optimal single channel strategy outperforms the optimal channel switching strategy. Numerical examples are presented for the theoretical results and effects of channel switching delays are investigated.

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

- The channel switching problem for average capacity maximization in the presence of channel switching delays is studied for the first time in the literature.
- An alternative optimization problem, which facilitates theoretical investigations, is formulated in terms of the number of channels employed in the channel switching process (Proposition 1 and Proposition 2).
- When the channel switching is to be performed among a certain number of channels, the optimal strategy and the corresponding average capacity are derived (Proposition 3).
- It is shown that channel switching among more than two different channels is not optimal, and an expression for the maximum average capacity of the optimal channel switching strategy is presented (Proposition 4).
- Conditions are specified for the cases in which the optimal strategy corresponds to the exclusive use of a single channel or to channel switching between two channels (Proposition 5 and Remark 4).

## II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a communication system in which  $K$  different channels are available in the communication link between a transmitter and a receiver. The channels are assumed to introduce independent additive Gaussian noise with constant

spectral density levels over the channel bandwidths.<sup>1</sup> It is assumed that the spectral density levels and the bandwidths of the channels can be different in general. The transmitter and the receiver can switch among these  $K$  channels in order to enhance the capacity of the communication system. At any given time, only one channel can be utilized for the transmission and the transmitter informs the receiver about which channel is occupied for the given time so that the transmitter and the receiver are synchronized [12], [20]. Fig. 1 illustrates the system with  $K$  different channels with possibly various bandwidths and noise levels. In practice, the transmitter can perform communication over one channel for a certain fraction of time; then, it switches to another channel and continues communication for another fraction of time, and so on. This scenario is applicable for cognitive radio systems in which a secondary user utilizes multiple available frequency bands that are not in use by primary users [30], [31]. Hence, secondary users can improve their average channel capacity by employing the channel switching strategy proposed in this study.

The main motivation behind the use of a single channel at a time is to realize a system with low cost/complexity. Since the channels considered in the system model in Fig. 1 have different center frequencies which can be dispersed over a wide range of frequencies in general (e.g., in cognitive radio systems [30], [31]), simultaneous utilization of multiple channels requires either multiple RF units (one for each channel) at the transmitter and the receiver, or single RF units that operate over the whole possible range of frequencies (i.e., over a very wide bandwidth)<sup>2</sup>. Therefore, simultaneous utilization of multiple channels leads to high complexity/cost compared to the use of one channel at a time. In the latter case, the single RF units at the transmitter and the receiver can be designed for a relatively narrowband scenario, and only one channel is used at a time by tuning the filters and amplifiers in the RF units and adjusting the upconversion/downconversion frequency according to the employed channel [32], [33].

In fact, if the frequency bands of two channels are adjacent to each other, they can be treated as a single channel with a larger bandwidth if the total bandwidth is within the operating range of the RF components. Hence, the theoretical analysis in the manuscript is also valid for scenarios in which two (multiple) such frequency bands (channels) are used simultaneously. In that case, all the theoretical results would hold by updating the definitions of the channels.

In the considered system model, before data communication commences, the transmitter determines a channel switching strategy that will be employed during a time duration of  $T_d$  seconds and informs the receiver about the channels to be utilized and the respective utilization times according to that strategy. It is assumed that the channel characteristics do not change during  $T_d$  seconds. To start data communication, the

transmitter and the receiver set their parameters for the first channel to be utilized (i.e., they switch to the same channel), and this process is assumed to take a time duration  $T_{cs}$  seconds, which is called the *channel switching delay (cost)*. During  $T_{cs}$  seconds, there is no data communication and consequently no power is transmitted. Then, data transmission starts and lasts for a certain time duration based on the employed strategy. Next, the transmitter and the receiver switch to the second channel to be utilized, which again takes  $T_{cs}$  seconds, and then data communication occurs over that channel for a specified time. The process continues in this manner according to the employed channel switching strategy, which may utilize a subset of all channels in general. For the next period of  $T_d$  seconds, the optimal channel switching strategy is calculated again according to the new channel characteristics, and communication continues in the same fashion as described above.

In Fig. 2, a sample time frame structure is presented for channel switching over 4 channels. In this case, the transmitter and the receiver communicate during  $3T_d$  seconds. In first  $T_d$  seconds, the channel switching strategy is to communicate over channel 1 and channel 3 for  $T_1^1$  and  $T_3^1$  seconds, respectively, where  $T_1^1 + T_3^1 = T_d$ . Before the data transmission over each channel, there exists a channel switching time (cost) of  $T_{cs}$  seconds, which is required for the transmitter and the receiver to set their parameters for communication over the desired channel. During the second  $T_d$  seconds, the communication is performed over only channel 2 for a time duration of  $T_2^2$  seconds, where  $T_2^2 = T_d$ , and there is no channel switching to another channel in this case. Finally, channels 1, 2 and 3 are utilized for the communication in the last  $T_d$  seconds. It is important to note that it is not necessary to utilize all the channels in a given channel switching strategy. For example, channel 4 is not utilized in any of the channel switching strategies in Fig. 2.

Let  $B_i$  and  $N_i/2$  denote, respectively, the bandwidth and the constant power spectral density level of the additive Gaussian noise for channel  $i$ , where  $i \in \{1, \dots, K\}$ . Then, the capacity of channel  $i$  is expressed as

$$C_i(P) = B_i \log_2 \left( 1 + \frac{P}{N_i B_i} \right) \text{ bits/sec} \quad (1)$$

where  $P$  represents the average transmit power [34].

The main aim of this study is to characterize the optimal channel switching strategy that maximizes the average capacity of the communication system in Fig. 1 under average and peak power constraints and in the presence of channel switching delays. To that aim, channel time-sharing (channel switching) factors are expressed as  $\lambda_1 \triangleq \frac{T_1}{T_d}, \dots, \lambda_K \triangleq \frac{T_K}{T_d}$ , where  $T_i$  denotes the amount of time allocated for channel  $i$  and  $T_d$  is the duration over which the channel switching strategy is employed. In addition,  $\varepsilon \triangleq \frac{T_{cs}}{T_d}$  is defined as the *channel switching delay factor*, and  $(\lambda_i - \varepsilon) \mathbb{I}_{\{\lambda_i > 0\}}$  represents the fraction of time when channel  $i$  is used for communication, where  $\mathbb{I}_{\{\lambda_i > 0\}}$  denotes the indicator function, which is equal to 1 if  $\lambda_i > 0$  and 0 otherwise. Then, the following optimal channel switching problem is proposed for capacity

<sup>1</sup>The additive Gaussian channel is an accurate model in the presence of thermal noise. In addition, it can also be employed in the presence of interference and jamming if they can be approximated by a Gaussian distribution; e.g., multiuser interference due to a large number of users with similar power levels and Gaussian jamming [27]-[29].

<sup>2</sup>In this case, very high rates would be required for analog-to-digital converters, which would lead to increased cost and high power consumption.

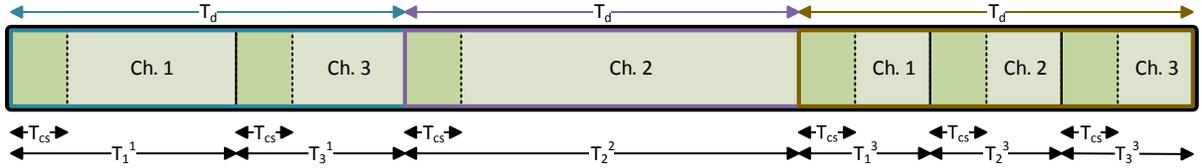


Fig. 2. A sample time frame structure of a communication system in which transmitter and receiver can switch among 4 channels.

TABLE I  
SYMBOLS AND THEIR DEFINITIONS

Symbol	Definition
$K$	Number of channels in the system
$B_i$	Bandwidth of channel $i$
$N_i$	Noise power spectral density level for channel $i$
$\varepsilon$	Channel switching delay factor
$P_{\text{pk}}$	Peak power limit
$P_{\text{av}}$	Average power limit
$P_i$	Average transmit power allocated to channel $i$
$C_i(P)$	Capacity of channel $i$ for average power $P$

maximization in the presence of channel switching delays:

$$\begin{aligned}
 & \max_{\{\lambda_i, P_i\}_{i=1}^K} \sum_{i=1}^K \mathbb{I}_{\{\lambda_i > 0\}} (\lambda_i - \varepsilon) C_i(P_i) \\
 & \text{subject to } \sum_{i=1}^K \mathbb{I}_{\{\lambda_i > 0\}} (\lambda_i - \varepsilon) P_i \leq P_{\text{av}}, \\
 & P_i \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, K\}, \\
 & \sum_{i=1}^K \lambda_i = 1, \lambda_i \in \{0\} \cup [\varepsilon, 1], \forall i \in \{1, \dots, K\}
 \end{aligned} \tag{2}$$

where  $C_i(P_i)$  is as in (1),  $P_i$  is the average transmit power allocated to channel  $i$ ,  $P_{\text{pk}}$  denotes the peak power limit, and  $P_{\text{av}}$  represents the average power limit for the transmitter. It is assumed that  $P_{\text{av}} < P_{\text{pk}}$  and  $0 < \varepsilon < 1$ . From (2), it is noted that due to the channel switching delay, a channel can be utilized only if its time-sharing factor is larger than or equal to the channel switching delay factor,  $\varepsilon$ . In addition,  $\varepsilon$  fractions are subtracted from both the average capacity and the average power terms since no data transmission occurs during channel switching. It should be emphasized that the objective function in (2) is referred to as the ‘‘average’’ capacity due to the averaging operation over time, considering the use of different channels and the channel switching delays.

For convenience, the symbols that are frequently used throughout the manuscript are summarized in Table I.

### III. OPTIMAL CHANNEL SWITCHING WITH SWITCHING DELAYS

In its current form, the optimization problem in (2) is difficult to solve in general since it is not a convex optimization problem and requires a search over a  $2K$  dimensional space. Therefore, our aim is to derive an equivalent formulation of the problem in (2), which leads to a low-complexity solution

for the optimal channel switching strategy. To achieve such a formulation, the optimization problem in (2) is first converted into another problem, the solution of which achieves the same maximum average capacity as (2) does. In the following proposition, this alternative optimization problem is presented.

**Proposition 1:** Define set  $A$  as  $A = \{1, \dots, K\}$  and let  $\mathcal{P}(A)$  denote the power set of set  $A$ . Then, the solution of the following optimization problem results in the same maximum value that is achieved by the problem in (2):

$$\begin{aligned}
 & \max_{\tilde{K} \in A} \max_{S \in B^{\tilde{K}}} \max_{\{\nu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} \sum_{i=1}^{\tilde{K}} (\nu_{s_i} - \varepsilon) C_{s_i}(P_{s_i}) \\
 & \text{subject to } \sum_{i=1}^{\tilde{K}} (\nu_{s_i} - \varepsilon) P_{s_i} \leq P_{\text{av}} \\
 & P_{s_i} \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, \tilde{K}\} \\
 & \sum_{i=1}^{\tilde{K}} \nu_{s_i} = 1, \nu_{s_i} \geq \varepsilon, \forall i \in \{1, \dots, \tilde{K}\}
 \end{aligned} \tag{3}$$

where  $s_i$  represents the  $i$ th element of set  $S$ , and  $B^{\tilde{K}}$  is defined as

$$B^{\tilde{K}} \triangleq \{\chi \in \mathcal{P}(A) \mid |\chi| = \tilde{K}\} \tag{4}$$

for  $\tilde{K} \in \{1, \dots, K\}$ , with  $|\chi|$  denoting the cardinality of set  $\chi$ .

*Proof:* Please see Appendix A.

In the optimization problem in (3), parameter  $\tilde{K}$  indicates the number of employed channels in a channel switching strategy; that is, the optimization is performed for all possible numbers of employed channels explicitly. In this way, the indicator functions in (2) are removed. Since there exist  $K$  available channels in the system, the optimization problem in (3) requires a search over all possible values of  $\tilde{K} \in A$ , where  $A = \{1, \dots, K\}$ . For each  $\tilde{K}$ , set  $B^{\tilde{K}}$  in (4) consists of the sets that are subsets of set  $A$  with  $\tilde{K}$  elements; that is,  $B^{\tilde{K}}$  corresponds to all possible  $\tilde{K}$  combinations of  $K$  different channels. Hence,  $B^{\tilde{K}}$  consists of  $\binom{K}{\tilde{K}}$  sets. For example, if  $K = 3$  and  $\tilde{K} = 2$ , then  $B^{\tilde{K}} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . For each element of  $B^{\tilde{K}}$ , which is denoted by  $S$  in (3), the optimization is performed over  $\{\nu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}$ , where  $s_i$  selects the  $i$ th channel in  $S$  and  $\nu_{s_i}$  and  $P_{s_i}$  denote, respectively, the time-sharing factor and the average transmit power allocated to channel  $s_i$ ; i.e., the  $i$ th employed (selected) channel.

The optimization problem in (3) is not only more convenient than the one in (2), which involves indicator functions, but also leads to simpler formulations of the optimal channel switching problem. To that end, the following proposition provides a

scaled and more compact version of the optimization problem in (3), the solution of which achieves the same maximum average capacity as (2) and (3) do.

**Proposition 2:** *The optimization problem in (3) can be expressed in the form of the following optimization problem:*

$$\begin{aligned} & \max_{\tilde{K} \in A} \max_{S \in B^{\tilde{K}}} \max_{\{\mu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} \left(1 - \tilde{K}\varepsilon\right) \sum_{i=1}^{\tilde{K}} \mu_{s_i} C_{s_i}(P_{s_i}) \\ & \text{subject to } \sum_{i=1}^{\tilde{K}} \mu_{s_i} P_{s_i} \leq \frac{P_{\text{av}}}{(1 - \tilde{K}\varepsilon)} \\ & P_{s_i} \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, \tilde{K}\} \\ & \sum_{i=1}^{\tilde{K}} \mu_{s_i} = 1, \mu_{s_i} \geq 0, \forall i \in \{1, \dots, \tilde{K}\} \\ & \tilde{K} < \frac{1}{\varepsilon} \end{aligned} \quad (5)$$

where  $A$ ,  $B^{\tilde{K}}$ , and  $s_i$  are as defined in Proposition 1.

*Proof:* Please see Appendix B.

The optimization problem in (5) can be separated into two optimization problems based on the value of  $\tilde{K}$  as follows:

- **Case-1 (Single Channel):** For the case in which a single channel is employed for communication, that is,  $\tilde{K} = 1$ , the optimization problem in (5) can be stated as follows:

$$\begin{aligned} & \max_{S \in B^1} \max_{\mu_{s_1}, P_{s_1}} (1 - \varepsilon) \mu_{s_1} C_{s_1}(P_{s_1}) \\ & \text{subject to } \mu_{s_1} P_{s_1} \leq \frac{P_{\text{av}}}{(1 - \varepsilon)} \\ & P_{s_1} \in [0, P_{\text{pk}}] \\ & \mu_{s_1} = 1, \mu_{s_1} \geq 0 \\ & \varepsilon < 1 \end{aligned} \quad (6)$$

where  $B^1 = \{\{1\}, \{2\}, \dots, \{K\}\}$  and  $s_1$  denotes the (first) element of  $S$ . The optimization problem in (6) achieves the maximum average capacity that can be obtained by employing a single channel during data communication. This approach corresponds to the case of no channel switching and is easily solvable by using simple algebra. Let  $C_{\text{scs}}$  denote the solution of (6). Then, the achieved maximum capacity via the optimal *single channel strategy* can be expressed as

$$C_{\text{scs}} = \max_{l \in \{1, \dots, K\}} (1 - \varepsilon) C_l \left( \min \left\{ \frac{P_{\text{av}}}{(1 - \varepsilon)}, P_{\text{pk}} \right\} \right) \quad (7)$$

and the channel index  $m$  employed in this strategy can be obtained as

$$m = \arg \max_{l \in \{1, \dots, K\}} C_l \left( \min \left\{ \frac{P_{\text{av}}}{(1 - \varepsilon)}, P_{\text{pk}} \right\} \right). \quad (8)$$

In the optimal single channel strategy, it is optimal to use all the available and attainable power,  $\min \left\{ \frac{P_{\text{av}}}{(1 - \varepsilon)}, P_{\text{pk}} \right\}$  over a single channel since  $C_i(P)$  in (1) is a monotone increasing and continuous function.

- **Case-2 (Channel Switching):** Consider the optimization problem in (5) in the presence of channel switching; that

is,  $\tilde{K} \geq 2$ . Then, the following optimization problem is obtained:

$$\begin{aligned} C_{\text{css}} &= \max_{\tilde{K} \in A \setminus \{1\}} \max_{S \in B^{\tilde{K}}} \max_{\{\mu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} (1 - \tilde{K}\varepsilon) \sum_{i=1}^{\tilde{K}} \mu_{s_i} C_{s_i}(P_{s_i}) \\ & \text{subject to } \sum_{i=1}^{\tilde{K}} \mu_{s_i} P_{s_i} \leq \frac{P_{\text{av}}}{(1 - \tilde{K}\varepsilon)} \\ & P_{s_i} \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, \tilde{K}\} \\ & \sum_{i=1}^{\tilde{K}} \mu_{s_i} = 1, \mu_{s_i} \geq 0, \forall i \in \{1, \dots, \tilde{K}\} \\ & \tilde{K} < \frac{1}{\varepsilon} \end{aligned} \quad (9)$$

The solution of the optimization problem in (9) results in the maximum average capacity that can be achieved by employing at least two different channels. In general, it is difficult to obtain the solution of (9). Therefore, further analysis is performed in the remainder of this study to obtain the optimal solution of (9) with low computational complexity.

Based on Case-1 and Case-2, the solution of (5) corresponds to either the single channel strategy or the channel switching strategy. Let  $C_{\text{scs}}$  and  $C_{\text{css}}$  denote the solutions of the optimization problems in (7) and (9), respectively. Then, the solution of (5) can be calculated as

$$\max \{C_{\text{scs}}, C_{\text{css}}\}. \quad (10)$$

As discussed in Case-1, the optimal single channel strategy has a simple closed-form solution. However, it is difficult to solve the channel switching problem in the form of (9). Therefore, the following proposition is presented to simplify the optimization problem in (9).

**Proposition 3:** *Assume that  $\tilde{K} \geq 2$  channels are employed in the channel switching strategy and  $\varepsilon < 1/\tilde{K}$  holds. Then, the maximum average capacity achieved via the optimal channel switching strategy over  $\tilde{K}$  channels can be expressed as*

$$\psi(\tilde{K}) = \begin{cases} \max_{\substack{\tilde{P}_1 \in [\frac{P_{\text{av}}}{1 - \tilde{K}\varepsilon}, P_{\text{pk}}] \\ \tilde{P}_2 \in [0, \frac{P_{\text{av}}}{1 - \tilde{K}\varepsilon}]} (1 - \tilde{K}\varepsilon) \left( \frac{\frac{P_{\text{av}}}{1 - \tilde{K}\varepsilon} - \tilde{P}_2}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_1) \right. \\ \quad \left. + \frac{\tilde{P}_1 - \frac{P_{\text{av}}}{1 - \tilde{K}\varepsilon}}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_2) \right), & \text{if } \frac{P_{\text{av}}}{1 - \tilde{K}\varepsilon} < P_{\text{pk}} \\ (1 - \tilde{K}\varepsilon) C_{\text{max}}(P_{\text{pk}}), & \text{otherwise} \end{cases} \quad (11)$$

where  $C_{\text{max}}(P)$  is defined as

$$C_{\text{max}}(P) \triangleq \max \{C_1(P), \dots, C_K(P)\}. \quad (12)$$

*Proof:* Please see Appendix C.

**Remark 1:** *For the case of  $P_{\text{av}}/(1 - \tilde{K}\varepsilon) \geq P_{\text{pk}}$  in (11), the average capacity of  $(1 - \tilde{K}\varepsilon) C_{\text{max}}(P_{\text{pk}})$  can be achieved by the following approach: First, switching to the best channel that achieves the maximum capacity for power level  $P_{\text{pk}}$  and*

transmitting at power level  $P_{\text{pk}}$  over that channel<sup>3</sup> for a time fraction of  $(1 - \bar{K}\varepsilon)$ ; then, switching among any  $(\bar{K} - 1)$  channels, except for the best channel, without transmitting any power (i.e., by only consuming a time fraction of  $\varepsilon$  for each channel). As will be proved towards the end of this section, it is always better to employ a single channel and not to perform channel switching in the case of  $P_{\text{av}}/(1 - \bar{K}\varepsilon) \geq P_{\text{pk}}$ . Hence, the solution of the optimal channel switching problem in (5) does not correspond to  $(1 - \bar{K}\varepsilon) C_{\text{max}}(P_{\text{pk}})$  for  $\bar{K} \geq 2$ . Therefore, the approach in this remark is optimal only under the condition that  $\bar{K} \geq 2$  channels are employed, but not optimal for the overall problem in (5).

Proposition 3 provides a significant simplification for the solution of the optimization problem in (9) and leads to the following formulation for the optimal channel switching strategy (Case-2):

$$\begin{aligned} & \max_{\tilde{K} \in A \setminus \{1\}} \psi(\tilde{K}) \\ & \text{subject to } \tilde{K} < \frac{1}{\varepsilon} \end{aligned} \quad (13)$$

where  $\psi(\tilde{K})$  is as in (11). Compared to (9), the problem in (13) has significantly lower computational complexity since its search space is only two-dimensional for each feasible  $\tilde{K}$  (see (11)) whereas a search over a  $2\tilde{K}$  dimensional space is required in (9) for each  $(\tilde{K}, S)$  pair.

Towards the aim of specifying the solution of (13), the following lemma is presented first, which states a useful inequality for  $C_{\text{max}}(\cdot)$  in (12).

**Lemma 1:** Let  $C_{\text{max}}(P/\alpha)$  and  $C_{\text{max}}(P/\beta)$  denote the capacities of the best channels for power levels  $P/\alpha$  and  $P/\beta$ , respectively, where  $C_{\text{max}}$  is as in (12),  $\alpha, \beta \in (0, 1)$  and  $P > 0$ . Then, the following inequality holds for  $\alpha > \beta$ :

$$\alpha C_{\text{max}}\left(\frac{P}{\alpha}\right) > \beta C_{\text{max}}\left(\frac{P}{\beta}\right) \quad (14)$$

*Proof:* Please see Appendix D.

It is noted that although  $C_{\text{max}}$  in (12) is not a concave function in general (cf. Fig. 3), the inequality in (14) always holds due to the fact that the capacity curve for each channel is nonnegative, concave, monotone increasing, and continuous.

In the following proposition, a general solution for (13) is provided, and it is shown that the optimal channel switching strategy (Case-2) corresponds to switching between two of the channels.

**Proposition 4:** The optimal channel switching strategy (Case-2) is to switch between two channels; that is, switching among more than two channels is not optimal. In addition, the maximum average capacity  $C_{\text{css}}$  achieved by the optimal channel switching strategy, which is obtained as the solution

of (13), can be expressed as follows:

$$C_{\text{css}} = \begin{cases} 0, & \text{if } \varepsilon \geq \frac{1}{2} \\ (1 - 2\varepsilon)C_{\text{max}}(P_{\text{pk}}), & \text{if } \varepsilon < \frac{1}{2} \text{ and } \frac{P_{\text{av}}}{1 - 2\varepsilon} \geq P_{\text{pk}} \\ \max_{\substack{\tilde{P}_1 \in [\frac{P_{\text{av}}}{1 - 2\varepsilon}, P_{\text{pk}}] \\ \tilde{P}_2 \in [0, \frac{P_{\text{av}}}{1 - 2\varepsilon})}} (1 - 2\varepsilon) \left( \frac{\frac{P_{\text{av}}}{1 - 2\varepsilon} - \tilde{P}_2}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_1) \right. \\ \left. + \frac{\tilde{P}_1 - \frac{P_{\text{av}}}{1 - 2\varepsilon}}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_2) \right), & \text{otherwise} \end{cases} \quad (15)$$

*Proof:* Please see Appendix E.

Based on Proposition 4, the optimal channel switching strategy can be specified in various scenarios. For the first scenario in (15), i.e., for  $\varepsilon \geq 1/2$ ,  $C_{\text{css}} = 0$  since channel switching is not feasible, as noted from the constraint in (13). For  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1 - 2\varepsilon) \geq P_{\text{pk}}$ , the solution of the optimal channel switching problem is to transmit at power level  $P_{\text{pk}}$  over the best channel (that achieves the maximum capacity for power level  $P_{\text{pk}}$ ) for a time fraction of  $(1 - 2\varepsilon)$ , then switching to another channel and not transmitting any power (i.e., by consuming a time fraction of  $\varepsilon$ ), which results in  $C_{\text{css}} = (1 - 2\varepsilon)C_{\text{max}}(P_{\text{pk}})$  (see Remark 1). Finally, for  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1 - 2\varepsilon) < P_{\text{pk}}$ , the achieved maximum average capacity can be calculated based on (65) in Appendix C as  $C_{\text{css}} = (1 - 2\varepsilon)(\mu^* C_{\text{max}}(\tilde{P}_1^*) + (1 - \mu^*) C_{\text{max}}(\tilde{P}_2^*))$ , where  $\tilde{P}_1^*$  and  $\tilde{P}_2^*$  are the optimizers of the maximization problem in (15),

$$\mu^* = \left( \frac{P_{\text{av}}}{1 - 2\varepsilon} - \tilde{P}_2^* \right) / \left( \tilde{P}_1^* - \tilde{P}_2^* \right), \quad (16)$$

and the optimal channel switching strategy is to switch between channel  $i$  and channel  $j$  with power levels  $\tilde{P}_1^*$  and  $\tilde{P}_2^*$ , respectively, where  $i$  and  $j$  are given by<sup>4</sup>

$$i = \arg \max_{l \in \{1, \dots, K\}} C_l(\tilde{P}_1^*) \quad (17)$$

$$j = \arg \max_{l \in \{1, \dots, K\}} C_l(\tilde{P}_2^*). \quad (18)$$

**Remark 2:** It is important to note that  $\mu^*$  in (16) and  $1 - \mu^*$  do not directly correspond to the time-sharing factors defined in the optimization problem in (2). In terms of the notation of the optimization problem in (2), the optimal time-sharing factors, denoted by  $\lambda_i^*$  and  $\lambda_j^*$ , for the optimal channel switching strategy between channel  $i$  and channel  $j$  can be obtained based on the transformations in Proposition 1 and Proposition 2 as

$$\lambda_i^* = (1 - 2\varepsilon)\mu^* + \varepsilon \quad (19)$$

$$\lambda_j^* = (1 - 2\varepsilon)(1 - \mu^*) + \varepsilon \quad (20)$$

where  $\mu^*$  is as defined in (16). Since the optimal channel switching strategy is to switch between two channels as stated in Proposition 4,  $\lambda_k^* = 0$  for  $k \in \{1, \dots, K\} \setminus \{i, j\}$ .

Next the solutions of the optimal single channel strategy in (7) and the optimal channel switching strategy in (15) are

<sup>3</sup>In the case of multiple channels that achieve the maximum capacity for power level  $P_{\text{pk}}$ , any of them can be chosen as the best channel.

<sup>4</sup>In the case of multiple maximizers in (17) or (18), any of them can be chosen for the optimal strategy.

considered together. Overall, the optimal strategy corresponds to one of them, which achieves the higher average capacity, as expressed in (10).

- If  $\varepsilon \geq 1/2$ , then the optimal single channel strategy outperforms the optimal channel switching strategy since  $C_{\text{scs}}$  in (7) always satisfies  $C_{\text{scs}} > 0$  whereas  $C_{\text{css}} = 0$  in this case.
- If  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) \geq P_{\text{pk}}$ , then the following expressions can be obtained for  $C_{\text{scs}}$ :

$$C_{\text{scs}} = (1-\varepsilon) \left( C_{\text{max}} \left( \frac{P_{\text{av}}}{1-\varepsilon} \right) \mathbb{I}_{\left\{ \frac{P_{\text{av}}}{1-\varepsilon} < P_{\text{pk}} \right\}} + C_{\text{max}}(P_{\text{pk}}) \mathbb{I}_{\left\{ \frac{P_{\text{av}}}{1-\varepsilon} \geq P_{\text{pk}} \right\}} \right) \quad (21)$$

$$> (1-2\varepsilon) \left( C_{\text{max}} \left( \frac{P_{\text{av}}}{1-2\varepsilon} \right) \mathbb{I}_{\left\{ \frac{P_{\text{av}}}{1-2\varepsilon} < P_{\text{pk}} \right\}} + C_{\text{max}}(P_{\text{pk}}) \mathbb{I}_{\left\{ \frac{P_{\text{av}}}{1-2\varepsilon} \geq P_{\text{pk}} \right\}} \right) \quad (22)$$

$$\geq (1-2\varepsilon) \left( C_{\text{max}}(P_{\text{pk}}) \mathbb{I}_{\left\{ \frac{P_{\text{av}}}{1-2\varepsilon} < P_{\text{pk}} \right\}} + C_{\text{max}}(P_{\text{pk}}) \mathbb{I}_{\left\{ \frac{P_{\text{av}}}{1-2\varepsilon} \geq P_{\text{pk}} \right\}} \right) \quad (23)$$

$$= (1-2\varepsilon) C_{\text{max}}(P_{\text{pk}}) \quad (24)$$

where the equality in (21) is obtained from (7), the inequality in (22) follows from (14) in Lemma 1, the relation in (23) is due to the condition  $P_{\text{av}}/(1-2\varepsilon) \geq P_{\text{pk}}$  and the monotone increasing property of  $C_{\text{max}}$  in (12), and the final expression in (24) follows from the definition of the indicator function. From (21)-(24), is obtained that  $C_{\text{scs}} > (1-2\varepsilon)C_{\text{max}}(P_{\text{pk}}) = C_{\text{css}}$ ; that is, the optimal single channel strategy achieves a higher average capacity than the optimal channel switching strategy for  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) \geq P_{\text{pk}}$ .

- Finally, for the case of  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) < P_{\text{pk}}$ , the optimal strategy is either the single channel strategy or the channel switching strategy, and the achieved maximum average capacity is expressed as

$$C_{\text{av}}^{\text{max}} = \max \{ C_{\text{scs}}, C_{\text{css}} \} \quad (25)$$

where  $C_{\text{scs}}$  is as in (7) and  $C_{\text{css}}$  can be calculated as specified in (15), namely,

$$\max_{\substack{\tilde{P}_1 \in \left[ \frac{P_{\text{av}}}{1-2\varepsilon}, P_{\text{pk}} \right] \\ \tilde{P}_2 \in \left[ 0, \frac{P_{\text{av}}}{1-2\varepsilon} \right)}} (1-2\varepsilon) \left( \frac{\frac{P_{\text{av}}}{1-2\varepsilon} - \tilde{P}_2}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_1) + \frac{\tilde{P}_1 - \frac{P_{\text{av}}}{1-2\varepsilon}}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_2) \right). \quad (26)$$

**Remark 3:** The fact that the optimal single channel strategy outperforms the optimal channel switching strategy for  $\varepsilon \geq 1/2$  is valid not only for the capacity metric in (1) but also for any performance metric that is a nonnegative function of the average transmit power. Similarly, the result in Proposition 3 can be extended for any performance metric that is a continuous and bounded function of the transmit power  $P$  for  $P \in [0, P_{\text{pk}}]$ . On the other hand, in the proof of Proposition 4, additional properties of nonnegativity,

monotonicity, and concavity are also employed since Lemma 1 is utilized (see Appendix D and Appendix E). For example, the capacity of a discrete memoryless channel (not necessarily Gaussian) with average transmit power constraint  $P$  is a nondecreasing, concave, and continuous function of  $P$  [35].

For the case of  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) < P_{\text{pk}}$ , the following result can be obtained in a similar fashion to Proposition 2 of [20], which presents a sufficient condition for the optimal single channel strategy to achieve a higher average capacity than the optimal channel switching strategy.

**Proposition 5:** Suppose that  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) < P_{\text{pk}}$  hold, and  $C_{\text{max}}(P)$  in (12) is first-order continuously differentiable in an interval around  $P_{\text{av}}/(1-2\varepsilon)$ . Then, the optimal single channel strategy outperforms the optimal channel switching strategy in terms of the maximum average capacity if

$$(P - P_{\text{av}}) \frac{B_{i^*} \log_2 e}{(1-2\varepsilon)N_{i^*}B_{i^*} + P_{\text{av}}} \geq C_{\text{max}} \left( \frac{P}{1-2\varepsilon} \right) - C_{\text{max}} \left( \frac{P_{\text{av}}}{1-2\varepsilon} \right) \quad (27)$$

for all  $P \in [0, (1-2\varepsilon)P_{\text{pk}}]$ , where  $i^* = \arg \max_{i \in \{1, \dots, K\}} C_i \left( \frac{P_{\text{av}}}{1-2\varepsilon} \right)$ .

*Proof:* Please see Appendix F.

Based on Proposition 5, if the condition in (27) is satisfied for the case of  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) < P_{\text{pk}}$ , and  $C_{\text{max}}(P)$  in (12) is first-order continuously differentiable in an interval around  $\frac{P_{\text{av}}}{1-2\varepsilon}$ , then the optimal strategy corresponds to the optimal single channel strategy and there is no need for channel switching. Otherwise, the optimal strategy cannot be directly determined and it requires the comparison of the average capacities obtained by the optimal single channel and the optimal channel switching strategies.

**Remark 4:** Overall, the solution of the optimal channel switching problem in the presence of switching delays can be specified as follows:

- If  $\varepsilon \geq 1/2$  or if  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) \geq P_{\text{pk}}$ , then the optimal strategy is to transmit over a single channel, which has the maximum capacity for power level  $\min \left\{ \frac{P_{\text{av}}}{(1-\varepsilon)}, P_{\text{pk}} \right\}$  (see (7) and (8)).
- If  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1-2\varepsilon) < P_{\text{pk}}$ ,
  - if  $C_{\text{max}}(P)$  in (12) is first-order continuously differentiable in an interval around  $P_{\text{av}}/(1-2\varepsilon)$  and the condition in (27) holds, then the optimal strategy is to transmit over a single channel, which has the maximum capacity for power level  $P_{\text{av}}/(1-\varepsilon)$ .
  - otherwise, depending on which one achieves a higher average capacity, the optimal solution is either transmission over a single channel that has the maximum capacity for power level  $\min \left\{ \frac{P_{\text{av}}}{(1-\varepsilon)}, P_{\text{pk}} \right\}$  or channel switching between channel  $i$  and channel  $j$  with time-sharing factors  $\lambda_i^* = (1-2\varepsilon) \left( \frac{P_{\text{av}}}{1-2\varepsilon} - \tilde{P}_2^* \right) / (\tilde{P}_1^* - \tilde{P}_2^*) + \varepsilon$  and  $\lambda_j^* = 1 - \lambda_i^* = (1-2\varepsilon) \left( \tilde{P}_1^* - \frac{P_{\text{av}}}{1-2\varepsilon} \right) / (\tilde{P}_1^* - \tilde{P}_2^*) + \varepsilon$  (see Remark 2) and power levels  $P_i^* = \tilde{P}_1^*$  and  $P_j^* = \tilde{P}_2^*$ , respectively, where  $i$  and  $j$  are given by (17) and

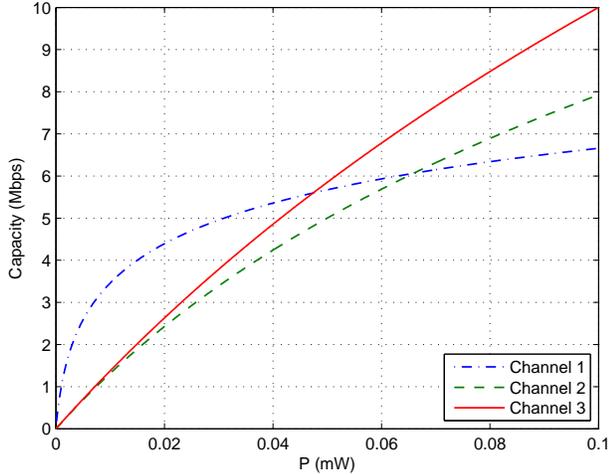


Fig. 3. Capacity of each channel versus power, where  $B_1 = 1$  MHz,  $B_2 = 5$  MHz,  $B_3 = 10$  MHz,  $N_1 = 10^{-12}$  W/Hz,  $N_2 = 10^{-11}$  W/Hz, and  $N_3 = 10^{-11}$  W/Hz.

(18), and  $\tilde{P}_1^*$  and  $\tilde{P}_2^*$  are the optimizers of (26).

#### IV. NUMERICAL RESULTS

In this section, numerical examples are presented to investigate the effects of the channel switching delay on the proposed optimal channel switching strategy, and to compare performance of the optimal channel switching and optimal single channel strategies in terms of average capacity maximization. Consider a scenario with  $K = 3$  channels where the bandwidths and the noise levels (cf. (1)) are given by  $B_1 = 1$  MHz,  $B_2 = 5$  MHz,  $B_3 = 10$  MHz,  $N_1 = 10^{-12}$  W/Hz,  $N_2 = 10^{-11}$  W/Hz, and  $N_3 = 10^{-11}$  W/Hz. Suppose that the peak power constraint and the channel switching delay factor in (2) are set to  $P_{\text{pk}} = 0.1$  mW and  $\varepsilon = 0.1$ , respectively. In Fig. 3, the capacity of each channel is plotted versus power based on the capacity formula in (1). For the scenario in Fig. 3, the proposed optimal channel switching strategies and the optimal single channel strategy are calculated for various average power limits ( $P_{\text{av}}$ ), and the achieved maximum average capacities are plotted versus  $P_{\text{av}}$  in Fig. 4. As discussed in the previous section, the optimal single channel strategy achieves a capacity of  $(1 - \varepsilon) C_{\text{max}}(\phi)$ , where  $\phi \triangleq \min\left\{\frac{P_{\text{av}}}{(1 - \varepsilon)}, P_{\text{pk}}\right\}$  and  $C_{\text{max}}(\phi) = \max\{C_1(\phi), C_2(\phi), C_3(\phi)\}$  in the considered scenario. It is observed from Fig. 3 and Fig. 4 that  $C_{\text{max}}(\phi) = C_1(\phi)$  for  $P_{\text{av}} \in (0, 0.0426)$  mW and  $C_{\text{max}}(\phi) = C_3(\phi)$  for  $P_{\text{av}} \in [0.0426, 0.1]$  mW; that is, channel 1 is the best channel up to  $P_{\text{av}} = 0.0426$  mW, and channel 3 is the best after that power level. Among the optimal channel switching strategies discussed in the previous section, it can be observed from Fig. 4 that the optimal channel switching strategy with two channels outperforms the optimal channel switching strategy with three channels for all  $P_{\text{av}} \in [0, 0.1]$  mW in accordance with Proposition 4. Overall, the optimal strategy is to employ the optimal channel switching strategy with two channels for  $P_{\text{av}} \in (0.0332, 0.0582)$  mW and the optimal single channel

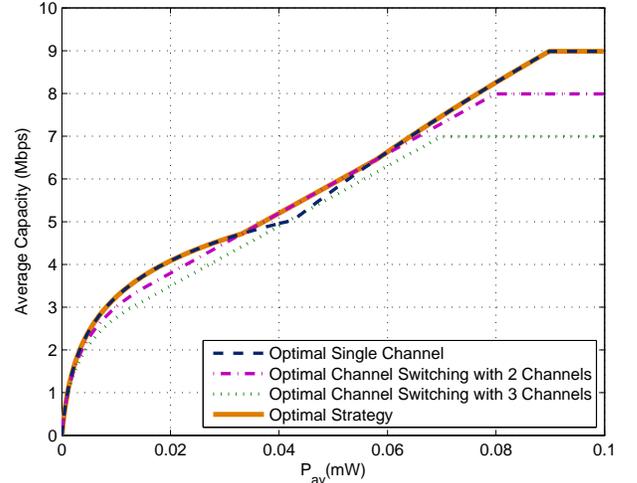


Fig. 4. Average capacity versus average power limit for the optimal channel switching and the optimal single channel strategies for the scenario in Fig. 3, where  $P_{\text{pk}} = 0.1$  mW.

strategy for  $P_{\text{av}} \in [0, 0.0332] \cup [0.0582, 0.1]$  mW. From (15) in Proposition 4, the behaviour of the optimal channel switching strategy with two channels in Fig. 4 can be explained as follows: For  $P_{\text{av}}/(1 - 2\varepsilon) \geq P_{\text{pk}}$ ; that is, for  $P_{\text{av}} \geq 0.08$  mW,  $C_{\text{css}}$  in (15) is given by  $(1 - 2\varepsilon)C_{\text{max}}(P_{\text{pk}}) = 0.8C_{\text{max}}(0.1)$ . On the other hand, for  $P_{\text{av}} < 0.08$  mW,  $C_{\text{css}}$  is calculated from the third expression in (15). In a similar fashion, based on (11) in Proposition 3, the optimal channel switching strategy with three channels achieves an average capacity of  $(1 - 3\varepsilon)C_{\text{max}}(P_{\text{pk}}) = 0.7C_{\text{max}}(0.1)$  for  $P_{\text{av}} \geq 0.07$  mW and yields the average capacity obtained from the first expression in (11) for  $P_{\text{av}} < 0.07$  mW. In addition, in accordance with Proposition 5, the optimal strategy is the optimal single channel strategy for  $P_{\text{av}} \in [0, 0.0176]$  mW since the condition in (27) holds for  $P_{\text{av}} \in [0, 0.0176]$  mW.

In order to investigate the optimal strategy in Fig. 4 in more detail, Table II presents the solutions of the optimal strategy for various values of the average power limit,  $P_{\text{av}}$ . In the table, the optimal solution is represented by parameters  $\lambda^*$ ,  $P_1^*$ ,  $P_2^*$ ,  $i$ , and  $j$ , meaning that channel  $i$  is used with time-sharing factor  $\lambda^*$  and power  $P_1^*$ , and channel  $j$  is employed with time-sharing factor  $1 - \lambda^*$  and power  $P_2^*$ . From Table II, it is observed that the optimal channel switching strategy with two channels is the optimal strategy for  $P_{\text{av}} = 0.04$  mW and  $P_{\text{av}} = 0.05$  mW, where switching between channel 1 and channel 3 is performed. For the other  $P_{\text{av}}$  values in Table II, it is optimal to employ the optimal single channel strategy which achieves higher average capacities than the optimal channel switching strategy.

To provide benchmarks on the performance of the proposed optimal channel switching strategy, two scenarios are considered: In the first one, the optimal channel switching strategy is performed in the absence of channel switching delays (i.e.,  $\varepsilon = 0$ ), which leads to an upper performance limit. In the second one, a lower performance limit is obtained by

TABLE II  
OPTIMAL STRATEGY FOR THE SCENARIO IN FIG. 3, WHICH EMPLOYS CHANNEL  $i$  AND CHANNEL  $j$  WITH TIME-SHARING FACTORS  $\lambda^*$  AND  $(1 - \lambda^*)$  AND POWER LEVELS  $P_1^*$  AND  $P_2^*$ , RESPECTIVELY.

$P_{av}$ (mW)	$\lambda^*$	$P_1^*$	$i$	$(1 - \lambda^*)$	$P_2^*$	$j$
0.01	—	—	—	1	0.0111	1
0.02	—	—	—	1	0.0222	1
0.03	—	—	—	1	0.0333	1
0.04	0.4026	0.1	3	0.5974	0.0196	1
0.05	0.527	0.1	3	0.473	0.0196	1
0.06	—	—	—	1	0.0667	3
0.07	—	—	—	1	0.0778	3
0.08	—	—	—	1	0.0889	3
0.09	—	—	—	1	0.1	3
0.099	—	—	—	1	0.1	3

designing the “optimal” channel switching strategy without the consideration of channel switching delays (i.e., assuming that  $\varepsilon$  is zero even though it is not). This scenario corresponds to the use of the approach in [20] (which is optimal for  $\varepsilon = 0$ ) in the presence of channel switching delays. Fig. 5 presents the average capacities achieved in these two scenarios, together with that achieved by the proposed optimal strategy obtained from (2) for the system in Fig. 3, where  $P_{pk} = 0.1$  mW and  $\varepsilon = 0.1$ . For the calculation of the average capacities achieved by the “optimal” strategy without the consideration of channel switching delays, the problem in [20] is solved first, and then the obtained solution is substituted into the objective function in (2). Namely, if  $\lambda^*$ ,  $P_1^*$ , and  $P_2^*$  denote the solution of [20], the maximum average capacity obtained via the strategy in which the delays are neglected is given by  $\max\{\lambda^* - \varepsilon, 0\}C_{\max}(P_1^*) + \max\{1 - \lambda^* - \varepsilon, 0\}C_{\max}(P_2^*)$ . On the other hand, the maximum average capacity achieved by the optimal channel switching strategy in the absence of channel switching delays (i.e., for  $\varepsilon = 0$ ) can be expressed as  $\lambda^*C_{\max}(P_1^*) + (1 - \lambda^*)C_{\max}(P_2^*)$ . Based on these strategies, it is observed from Fig. 5 that the optimal strategy in the absence of channel switching delays outperforms the other strategies; hence, presents an upper limit, as expected. In addition, the delay-ignorant strategy (i.e., assuming no delays) cannot achieve a higher average capacity than that achieved by (2) (i.e., the proposed approach) due to the inefficient use of the average power and the optimization of the channel switching factors and power levels based on an unrealistic setting. On the other hand, the proposed optimal strategy obtained from (2) takes into account the fact that no data transmission occurs during channel switching and consequently no power is transmitted. Therefore, it optimizes the channel switching factors and power levels by using the average power efficiently. It is also noted that the abrupt behavioral changes in the average capacity curve of the delay-ignorant strategy occurs due to the change in the number of channels employed in the strategy and the decrease in the efficiency of average power usage.

Based on the scenario in Fig. 3, the maximum average capacities for the strategies specified in Fig. 4 are plotted

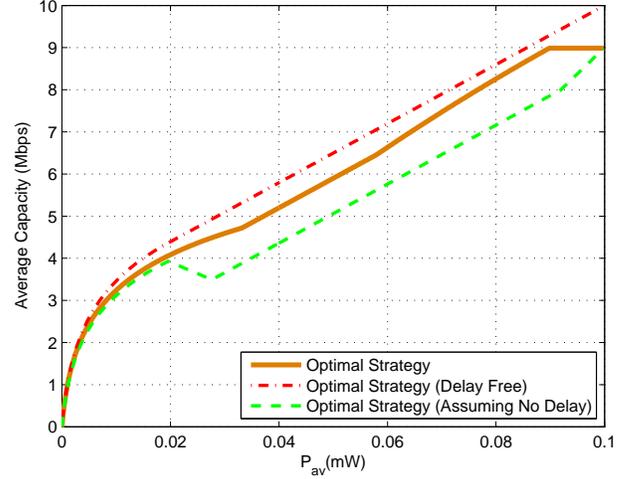


Fig. 5. Average capacity versus average power limit for the optimal strategy in the absence of channel switching delays ( $\varepsilon = 0$ ) and the optimal strategy without considering channel switching delays ( $\varepsilon = 0.1$ ), together with the proposed optimal strategy for the scenario in Fig. 3, where  $P_{pk} = 0.1$  mW and  $\varepsilon = 0.1$ .

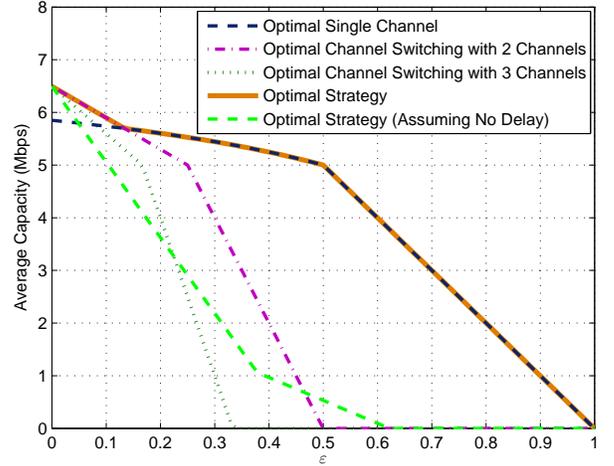


Fig. 6. Average capacity versus channel switching delay factor for various optimal strategies for the scenario in Fig. 3, where  $P_{av} = 0.05$  mW and  $P_{pk} = 0.1$  mW.

versus the channel switching delay factor ( $\varepsilon$ ) in Fig. 6 to investigate the effects of the channel switching delay factor on the average capacity. The average power limit and the peak power constraint in (2) are set to  $P_{av} = 0.05$  mW and  $P_{pk} = 0.1$  mW, respectively. From Fig. 6, it is noted that, in accordance with Proposition 4, the optimal channel switching strategy with two channels achieves a higher average capacity than the optimal channel switching strategy with three channels whenever channel switching is feasible; i.e.,  $\varepsilon < 0.5$ . For  $\varepsilon \in (0, 0.134)$ , the optimal strategy is the optimal channel switching strategy between two channels, whereas the optimal single channel strategy is the overall optimal for  $\varepsilon \in [0.134, 1)$ . It is important to note that the behavioral

TABLE III  
OPTIMAL STRATEGY FOR THE SCENARIO IN FIG. 3, WHICH EMPLOYS CHANNEL  $i$  AND CHANNEL  $j$  WITH TIME-SHARING FACTORS  $\lambda^*$  AND  $(1 - \lambda^*)$  AND POWER LEVELS  $P_1^*$  AND  $P_2^*$ , RESPECTIVELY.

$\varepsilon$	$\lambda^*$	$P_1^*$	$i$	$(1 - \lambda^*)$	$P_2^*$	$j$
0.05	0.4526	0.1	3	0.5474	0.0196	1
0.1	0.527	0.1	3	0.473	0.0196	1
0.2	—	—	—	1	0.0625	3
0.3	—	—	—	1	0.0714	3
0.4	—	—	—	1	0.0833	3
0.5	—	—	—	1	0.1	3
0.6	—	—	—	1	0.1	3
0.7	—	—	—	1	0.1	3
0.8	—	—	—	1	0.1	3
0.9	—	—	—	1	0.1	3

change in the average capacity curve of the optimal strategy at  $\varepsilon = 0.5$  is observed due to the peak power constraint in (2). Since  $P_{\text{av}}/(1 - \varepsilon) \geq P_{\text{pk}}$  for  $\varepsilon \geq 0.5$ , the optimal strategy achieves an average capacity of  $(1 - \varepsilon)C_{\text{max}}(P_{\text{pk}})$  by allocating all the available and attainable power to a single channel and transmitting over this single channel. For comparison purposes, Fig. 6 also presents the average capacity achieved by the “optimal” strategy which assumes no channel switching delays and optimizes the parameters accordingly [20]. It is noted that this strategy is outperformed by the proposed optimal strategy, which takes into account the channel switching delays. Fig. 6 clearly points out that the consideration of channel switching delays in the strategy design becomes more crucial for improved average capacity as the channel switching delay factor increases.

Similar to Table II, Table III presents the solutions corresponding to the optimal strategy for various values of the channel switching delay factor. For  $\varepsilon = 0.05$  and  $\varepsilon = 0.1$ , it is observed that the optimal strategy is to switch between channel 1 and channel 3. For the other  $\varepsilon$  values satisfying  $\varepsilon \geq 0.134$  in Table III, the optimal strategy is to transmit over channel 3 exclusively with power level  $P_2^* = P_{\text{av}}/(1 - \varepsilon)$  for  $\varepsilon < 0.5$  and  $P_2^* = P_{\text{pk}}$  otherwise.

In order to investigate whether channel switching can provide any benefits for practical modulation schemes, consider the achievable capacity of the discrete-input continuous-output memoryless channel (DCMC) with 64-QAM signaling in the presence of additive white Gaussian noise [36, eqn. (23.23)] for the scenario in Fig. 3. As an example, for  $P_{\text{av}} = 0.04$  mW and  $\varepsilon = 0.1$ , the calculations show that when the optimal strategy for  $P_{\text{av}} = 0.04$  mW in Table II (that is,  $\lambda^* = 0.4026$ ,  $P_1^* = 0.1$  mW,  $P_2^* = 0.0196$  mW,  $i = 3$ , and  $j = 1$ ) is employed for the achievable capacity of 64-QAM [36], an average achievable capacity of 5.033 Mbps is obtained whereas the optimal single channel approach yields an achievable capacity of 4.5819 Mbps. Hence, it is observed that it is possible to achieve performance improvements via channel

switching also for practical modulation schemes.<sup>5</sup>

## V. EXTENSIONS

In this study, the optimal channel switching problem is investigated for a single user. In the presence of multiple users, the results in this study can be extended in various directions. First, if orthogonal resource allocation is employed such that each user utilizes a different channel at a given time, then the results in this manuscript would still hold. In such a scenario, a central unit can provide coordination by informing each user about the available channels for that user in each time frame. Secondly, if users are allowed to employ the same channels and possible interference to a user is modeled by a Gaussian noise process, then the channel switching problem in (2) can be extended for nonorthogonal resource allocation, as well. In this case, when a user wishes to commence communications over the available channels, it first performs spectrum sensing and determines the interference level in each channel. Then, the capacity of each channel is given by

$$C_i(P) = B_i \log_2 \left( 1 + \frac{P}{2B_i(N_i/2 + I_i)} \right) \text{ bits/sec} \quad (28)$$

where  $I_i$  is the spectral density level of the interference (due to the other users) in channel  $i$  and the other parameters are as defined for (2). When the channel switching problem in (2) is solved based on the capacity expression in (28), the solution corresponds to the optimal channel switching strategy in the presence of multiuser interference. Since the structure of this new problem is the same as that of the original problem (cf. (1)), all the theoretical results apply to this scenario, as well. An example application for this scenario is a cognitive radio system with the underlay approach, where a secondary user utilizes the channels of primary users as long as it does not cause significant performance degradation for primary users [37, Ch. 2]. In that case, the secondary user performs channel (spectrum) sensing and determines the presence of primary users and the corresponding interference levels. Then, the proposed optimal channel switching strategy can be obtained as described above.

In case of non-orthogonal multiple access, fairness should be considered to satisfy certain average capacity requirements for all users. One way of achieving fairness is related to the limitation of power levels over different channels so that interference to users is limited; hence, no significant capacity degradations are observed. In other words, for each user, the maximum amount of power that can be transmitted over each channel can be determined according to a fairness criterion, which is set by a central unit. To provide a generic analysis that covers various fairness strategies, let  $\bar{P}_i$  represent the maximum power that can be transmitted over channel  $i$ . When a user wants to start communications over the available

<sup>5</sup>It is noted that this performance improvement is achieved without performing specific optimization for the achievable capacity function corresponding to a practical modulation scheme, which can be implemented to obtain further improvements.

channels, it designs the optimal channel switching strategy as follows (cf. (2)):

$$\begin{aligned} & \max_{\{\lambda_i, P_i\}_{i=1}^K} \sum_{i=1}^K \mathbb{I}_{\{\lambda_i > 0\}} (\lambda_i - \varepsilon) C_i(P_i) \\ & \text{subject to } \sum_{i=1}^K \mathbb{I}_{\{\lambda_i > 0\}} (\lambda_i - \varepsilon) P_i \leq P_{av}, \\ & P_i \in [0, \min\{\hat{P}_i, P_{pk}\}], \forall i \in \{1, \dots, K\}, \\ & \sum_{i=1}^K \lambda_i = 1, \lambda_i \in \{0\} \cup [\varepsilon, 1], \forall i \in \{1, \dots, K\} \end{aligned} \quad (29)$$

where  $C_i(P_i)$  is as in (28),  $\hat{P}_i$  is the power limit for channel  $i$ , and the other parameters are as in (2). In this way, fairness among various users can be achieved by adjusting the power limits of each user over different channels.

The results in the manuscript can be extended for the problem in (29) as follows: Similar to Proposition 1, an alternative optimization problem to (29) can be obtained as in (3) by updating the definition of  $C_{s_i}(\cdot)$  and replacing the peak power constraints with  $P_{s_i} \in [0, \min\{\hat{P}_{s_i}, P_{pk}\}]$ ,  $\forall i \in \{1, \dots, \tilde{K}\}$ . It can be shown based on similar arguments to those in the proof of Proposition 1 that the alternative problem achieves the same maximum average capacity as (29). Next, define the following function:

$$\hat{C}_i(P) \triangleq \begin{cases} C_i(P), & \text{if } P \leq \min\{\hat{P}_i, P_{pk}\} \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

for  $i \in \{1, \dots, K\}$ , where  $C_i(P)$  is as in (28). Based on a similar approach to that in Proposition 2, the alternative optimization problem can be expressed as in (5) by replacing  $C_{s_i}(P_{s_i})$  and  $P_{s_i} \in [0, P_{pk}]$  in (5) with  $\hat{C}_{s_i}(P_{s_i})$  in (30) and  $P_{s_i} \in [0, \min\{\hat{P}_{s_i}, P_{pk}\}]$ , respectively. Then, the resulting optimization problem can be separated into two optimization problems in a similar fashion:

**Case-1 (Single Channel):** In this case, the following optimization problem can be obtained:

$$\begin{aligned} & \max_{S \in B^1} \max_{\mu_{s_1}, P_{s_1}} (1 - \varepsilon) \mu_{s_1} \hat{C}_{s_1}(P_{s_1}) \\ & \text{subject to } \mu_{s_1} P_{s_1} \leq \frac{P_{av}}{(1 - \varepsilon)} \\ & P_{s_1} \in [0, \min\{\hat{P}_{s_1}, P_{pk}\}] \\ & \mu_{s_1} = 1, \mu_{s_1} \geq 0 \\ & \varepsilon < 1 \end{aligned} \quad (31)$$

where the parameters are as defined in (6). Let  $\hat{C}_{scs}$  denote the solution of (31). Then,  $\hat{C}_{scs}$  can be expressed as

$$\hat{C}_{scs} = \max_{l \in \{1, \dots, K\}} (1 - \varepsilon) C_l \left( \min \left\{ \frac{P_{av}}{(1 - \varepsilon)}, \min\{\hat{P}_l, P_{pk}\} \right\} \right) \quad (32)$$

and the channel index  $m$  employed in this strategy can be

obtained as

$$\hat{m} = \arg \max_{l \in \{1, \dots, K\}} C_l \left( \min \left\{ \frac{P_{av}}{(1 - \varepsilon)}, \min\{\hat{P}_l, P_{pk}\} \right\} \right). \quad (33)$$

**Case-2 (Channel Switching):** In this case, the following optimization problem can be obtained:

$$\begin{aligned} & \hat{C}_{css} = \max_{\tilde{K} \in A \setminus \{1\}} \max_{S \in B^{\tilde{K}}} \max_{\{\mu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} (1 - \tilde{K}\varepsilon) \sum_{i=1}^{\tilde{K}} \mu_{s_i} \hat{C}_{s_i}(P_{s_i}) \\ & \text{subject to } \sum_{i=1}^{\tilde{K}} \mu_{s_i} P_{s_i} \leq \frac{P_{av}}{(1 - \tilde{K}\varepsilon)} \\ & P_{s_i} \in [0, \min\{\hat{P}_{s_i}, P_{pk}\}], \forall i \in \{1, \dots, \tilde{K}\} \\ & \sum_{i=1}^{\tilde{K}} \mu_{s_i} = 1, \mu_{s_i} \geq 0, \forall i \in \{1, \dots, \tilde{K}\} \\ & \tilde{K} < \frac{1}{\varepsilon} \end{aligned} \quad (34)$$

where the parameters are as in (9). Based on Case-1 and Case-2, the solution can be calculated as  $\max\{\hat{C}_{scs}, \hat{C}_{css}\}$ .

For the optimization problem in (34), the statement in Proposition 3 can be extended as follows: Assume that  $\tilde{K} \geq 2$  channels are employed in the channel switching strategy and  $\varepsilon < 1/\tilde{K}$  holds. Also,  $P_{max}$  is defined as  $P_{max} = \max_{i \in \{1, \dots, K\}} \min\{\hat{P}_i, P_{pk}\}$ . Then, the maximum average capacity achieved via the optimal channel switching strategy over  $\tilde{K}$  channels can be expressed as

$$\psi(\tilde{K}) = \begin{cases} \max_{\substack{\tilde{P}_1 \in [\frac{P_{av}}{1 - \tilde{K}\varepsilon}, P_{max}] \\ \tilde{P}_2 \in [0, \frac{P_{av}}{1 - \tilde{K}\varepsilon})}} (1 - \tilde{K}\varepsilon) \left( \frac{\frac{P_{av}}{1 - \tilde{K}\varepsilon} - \tilde{P}_2}{\tilde{P}_1 - \tilde{P}_2} \hat{C}_{max}(\tilde{P}_1) \right. \\ \quad \left. + \frac{\tilde{P}_1 - \frac{P_{av}}{1 - \tilde{K}\varepsilon}}{\tilde{P}_1 - \tilde{P}_2} \hat{C}_{max}(\tilde{P}_2) \right), & \text{if } \frac{P_{av}}{1 - \tilde{K}\varepsilon} < \hat{P} \\ (1 - \tilde{K}\varepsilon) \hat{C}_{max}(\hat{P}), & \text{otherwise} \end{cases} \quad (35)$$

where  $\hat{C}_{max}(P)$  is defined as

$$\hat{C}_{max}(P) \triangleq \max\{\hat{C}_1(P), \dots, \hat{C}_K(P)\} \quad (36)$$

and  $\hat{P}$  is given by

$$\hat{P} \triangleq \arg \max_{P \in [0, P_{max}]} \hat{C}_{max}(P). \quad (37)$$

The solution of the optimization problem in (34) can be obtained from (13) where  $\psi(\tilde{K})$  is as in (35). In addition, the statement in Lemma 1 also holds for positive  $\hat{C}_{max}(\cdot)$ ; i.e., it holds if  $P/\alpha$  and  $P/\beta$  satisfy  $P/\alpha, P/\beta \in [0, P_{max}]$ . Then, the optimal channel switching strategy is to switch between two channels and the maximum average capacity  $\hat{C}_{css}$  achieved by the optimal channel switching strategy can be

expressed, similar to Proposition 4, as follows:

$$\hat{C}_{\text{css}} = \begin{cases} 0, & \text{if } \varepsilon \geq \frac{1}{2} \\ (1 - 2\varepsilon)C_{\text{max}}(\hat{P}), & \text{if } \varepsilon < \frac{1}{2} \text{ and } \frac{P_{\text{av}}}{1-2\varepsilon} \geq \hat{P} \\ \max_{\substack{\tilde{P}_1 \in [\frac{P_{\text{av}}}{1-2\varepsilon}, P_{\text{max}}] \\ \tilde{P}_2 \in [0, \frac{P_{\text{av}}}{1-2\varepsilon})}} (1 - 2\varepsilon) \left( \frac{P_{\text{av}} - \tilde{P}_2}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_1) \right. \\ \left. + \frac{\tilde{P}_1 - P_{\text{av}}}{\tilde{P}_1 - \tilde{P}_2} C_{\text{max}}(\tilde{P}_2) \right), & \text{otherwise} \end{cases} \quad (38)$$

Based on (32) and (38), it can be obtained that the optimal strategy corresponds to the optimal single channel strategy if  $\varepsilon \geq 1/2$  or if  $\varepsilon < 1/2$  and  $P_{\text{av}}/(1 - 2\varepsilon) \geq \hat{P}$ . Otherwise, the optimal strategy is either the single channel strategy or the channel switching strategy based on the comparison of the average capacities obtained from (32) and (38). Overall, it is concluded that in the presence of generic power limits for different channels for each user (due to a fairness criterion), the results in this manuscript are still valid with slight modifications in the optimization problems and the statements in the propositions.

Another way of providing fairness can be realized via the joint optimization of the multiuser system. In that case, the aim is to maximize the sum of the average capacities of the users under constraints on the average capacity of each user (to guarantee a certain average capacity for all users), the average power, and the peak powers. In general, it is quite difficult to obtain the solution of this joint optimization problem. Theoretical and numerical investigations of this problem are considered as an important direction for future work.

## VI. CONCLUDING REMARKS

In this study, the optimal channel switching problem has been investigated for average capacity maximization in the presence of channel switching delays. First, an equivalent formulation of the optimal channel switching problem has been obtained to facilitate theoretical investigations. Then, the optimal strategy has been obtained and the corresponding average capacity has been specified when channel switching is performed among a given number of channels. Based on this result (and Lemma 1), it has been shown that optimal channel switching does not involve more than two different channels, and the resulting maximum average capacity has been formulated for various values of the channel switching delay parameter and the average and peak power limits. Then, the scenarios under which the optimal strategy corresponds to the exclusive use of a single channel or to channel switching between two channels have been specified. Furthermore, sufficient conditions have been obtained to determine when the optimal single channel strategy outperforms optimal channel switching. Via numerical examples, the theoretical results and the effects of channel switching delays have been illustrated.

The capacity metric in (1) specifies the maximum data rates, which can be achieved in practice via turbo coding or low density parity check codes [38]. The results in this study can also be extended for any other performance metric that is

a nonnegative, concave, monotone increasing, bounded, and continuous function of the transmit power. For example, considering a certain modulation/demodulation scheme, the average number of correctly received symbols can be defined as an alternative performance metric. Since, in Gaussian channels, the probability of correct decision is a concave function of the transmit power for many modulation types (for all modulation types at high signal-to-noise ratios) [39], it can be shown that the average number of correctly received symbols becomes a nonnegative, concave, monotone increasing, bounded, and continuous function of the transmit power. Therefore, it can be shown that the results in Propositions 1–4 and Lemma 1 hold for such a scenario, as well, and Proposition 5 can also be extended.

## APPENDIX

### A. Proof of Proposition 1

Let  $\{\lambda_i^*, P_i^*\}_{i=1}^K$  represent the solution of (2) and define  $C^*$  as the maximum average capacity achieved by the optimization problem in (2); that is,

$$C^* = \sum_{i=1}^K \mathbb{I}_{\{\lambda_i^* > 0\}} (\lambda_i^* - \varepsilon) C_i(P_i^*). \quad (39)$$

Also, define a set as

$$M \triangleq \{l \in \{1, \dots, K\} \mid \lambda_l^* > 0\} \quad (40)$$

which consists of the channel indices with nonzero (positive) time-sharing factors. Next, consider the following transformation:

$$\nu_{m_i}^* = \lambda_{m_i}^*, \bar{P}_{m_i}^* = P_{m_i}^*, \forall i \in \{1, \dots, |M|\} \quad (41)$$

where  $m_i$  represents the  $i$ th element of  $M$ , and  $|M|$  is the cardinality of set  $M$ . Then, the following relations can be obtained for  $C^*$ :

$$\begin{aligned} C^* &= \sum_{i=1}^K \mathbb{I}_{\{\lambda_i^* > 0\}} (\lambda_i^* - \varepsilon) C_i(P_i^*) \\ &= \sum_{m \in M} (\lambda_m^* - \varepsilon) C_m(P_m^*) \end{aligned} \quad (42)$$

$$= \sum_{i=1}^{|M|} (\nu_{m_i}^* - \varepsilon) C_{m_i}(\bar{P}_{m_i}^*) \quad (43)$$

where the equalities in (42) and (43) are obtained from the definitions in (40) and (41), respectively. Next, define  $\tilde{K}^*$  as  $\tilde{K}^* \triangleq |M|$  and  $S^*$  as  $S^* \triangleq M$ . Then, the relation in (43) implies that the optimization problem in (3) achieves  $C^*$  for  $\tilde{K}^*$ ,  $S^*$ , and  $\{\nu_{s_i^*}^*, \bar{P}_{s_i^*}^*\}_{i=1}^{\tilde{K}^*}$  (see (41)), where  $s_i^*$  denotes the  $i$ th element of  $S^*$ .<sup>6</sup> Hence, (3) is guaranteed to yield the maximum average capacity achieved by the optimization problem in (2), that is,  $C^* \leq C^\diamond$ , where  $C^\diamond$  represents the maximum average capacity achieved by (3).

Next, suppose that  $\tilde{K}^\diamond$ ,  $S^\diamond$ , and  $\{\nu_{s_i^\diamond}^\diamond, \bar{P}_{s_i^\diamond}^\diamond\}_{i=1}^{\tilde{K}^\diamond}$  denote the solution of the optimization problem in (3), where  $s_i^\diamond$  denotes

<sup>6</sup>Note that the constraints in (3) are satisfied for  $\tilde{K}^*$ ,  $S^*$ , and  $\{\nu_{s_i^*}^*, \bar{P}_{s_i^*}^*\}_{i=1}^{\tilde{K}^*}$ .

the  $i$ th element of  $S^\diamond$ . Consider the following functions that map the solution set of the problem in (3) to the possible solution set of the problem in (2):

$$\lambda_i^\diamond = \begin{cases} \nu_i^\diamond, & \text{if } i \in S^\diamond \\ 0, & \text{otherwise} \end{cases}, \forall i \in \{1, \dots, K\} \quad (44)$$

$$P_i^\diamond = \begin{cases} \bar{P}_i^\diamond, & \text{if } i \in S^\diamond \\ 0, & \text{otherwise} \end{cases}, \forall i \in \{1, \dots, K\} \quad (45)$$

Then, the following relations can be written for  $C^\diamond$ :

$$\begin{aligned} C^\diamond &= \sum_{i=1}^{\tilde{K}^\diamond} (\nu_{s_i^\diamond}^\diamond - \varepsilon) C_{s_i^\diamond}(\bar{P}_{s_i^\diamond}^\diamond) \\ &= \sum_{m \in S^\diamond} (\nu_m^\diamond - \varepsilon) C_m(\bar{P}_m^\diamond) \end{aligned} \quad (46)$$

$$= \sum_{m \in S^\diamond} (\lambda_m^\diamond - \varepsilon) C_m(P_m^\diamond) \quad (47)$$

$$= \sum_{i=1}^K \mathbb{I}_{\{i \in S^\diamond\}} (\lambda_i^\diamond - \varepsilon) C_i(P_i^\diamond) \quad (48)$$

$$= \sum_{i=1}^K \mathbb{I}_{\{\lambda_i^\diamond > 0\}} (\lambda_i^\diamond - \varepsilon) C_i(P_i^\diamond) \quad (49)$$

where the equality in (46) is due to the definition of set  $S^*$  (see (3)), the equalities in (47) and (48) follow from the mapping functions in (44) and (45), and (49) is obtained from the fact that  $\lambda_i^\diamond > 0$  only for  $i \in S^\diamond$ . Based on the transformations defined in (44) and (45),  $\{\lambda_i^\diamond, P_i^\diamond\}_{i=1}^K$  satisfies the constraints in (2) and the relation in (46)-(49) implies that (2) yields the average capacity of  $C^\diamond$  for  $\{\lambda_i^\diamond, P_i^\diamond\}_{i=1}^K$ ; hence, it is concluded that  $C^\diamond \leq C^*$ . Overall, it is concluded that  $C^\diamond = C^*$  must hold in order to satisfy both  $C^* \leq C^\diamond$  and  $C^\diamond \leq C^*$ . ■

### B. Proof of Proposition 2

Consider the optimization problem in (3) and define new variables  $\gamma_{s_i}$  as  $\gamma_{s_i} \triangleq \nu_{s_i} - \varepsilon$ ,  $\forall i \in \{1, \dots, \tilde{K}\}$ . Then, the problem in (3) can be written as follows:

$$\max_{\tilde{K} \in A} \max_{S \in B^{\tilde{K}}} \max_{\{\gamma_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} \sum_{i=1}^{\tilde{K}} \gamma_{s_i} C_{s_i}(P_{s_i}) \quad (50)$$

$$\text{subject to } \sum_{i=1}^{\tilde{K}} \gamma_{s_i} P_{s_i} \leq P_{\text{av}} \quad (51)$$

$$P_{s_i} \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, \tilde{K}\} \quad (52)$$

$$\sum_{i=1}^{\tilde{K}} \gamma_{s_i} = 1 - \tilde{K}\varepsilon, \gamma_{s_i} \geq 0, \forall i \in \{1, \dots, \tilde{K}\} \quad (53)$$

It is noted from (53) that  $1 - \tilde{K}\varepsilon \geq 0$  should be satisfied since  $\sum_{i=1}^{\tilde{K}} \gamma_{s_i} = 1 - \tilde{K}\varepsilon$  and  $\gamma_{s_i} \geq 0$ ,  $\forall i \in \{1, \dots, \tilde{K}\}$ . Suppose that  $\tilde{K}^\diamond$ ,  $S^\diamond$ , and  $\{\gamma_{s_i^\diamond}^\diamond, P_{s_i^\diamond}^\diamond\}_{i=1}^{\tilde{K}^\diamond}$  denote the solution of (50)-(53) such that  $1 - \tilde{K}^\diamond\varepsilon = 0$ . Then, based on the constraint in (53),  $\gamma_{s_i^\diamond}^\diamond = 0$ ,  $\forall i \in \{1, \dots, \tilde{K}^\diamond\}$ , and consequently  $\sum_{i=1}^{\tilde{K}^\diamond} \gamma_{s_i^\diamond}^\diamond C_{s_i^\diamond}(P_{s_i^\diamond}^\diamond) = 0$ . Also,  $\tilde{K}^\diamond$  satisfies  $\tilde{K}^\diamond > 1$  since  $\tilde{K}^\diamond = 1/\varepsilon$  and  $0 < \varepsilon < 1$  by assumption. Hence, more

than one channel is available for channel switching. Now, consider an alternative solution, denoted by  $\tilde{K}^*$ ,  $S^*$ , and  $\{\gamma_{s_i^*}^*, P_{s_i^*}^*\}_{i=1}^{\tilde{K}^*}$ , where  $\tilde{K}^* = 1$ ,  $S^* = \{1\}$ ,  $\gamma_{s_1}^* = 1 - \varepsilon$ , and  $P_{s_1}^* = \min\{\frac{P_{\text{av}}}{(1-\varepsilon)}, P_{\text{pk}}\}$ . Then, the alternative solution achieves an average capacity of  $\sum_{i=1}^{\tilde{K}^*} \gamma_{s_i^*}^* C_{s_i^*}(P_{s_i^*}^*) = (1 - \varepsilon) C_1(\min\{\frac{P_{\text{av}}}{(1-\varepsilon)}, P_{\text{pk}}\})$ , which is positive; hence, larger than the one achieved by  $\tilde{K}^\diamond$ ,  $S^\diamond$ , and  $\{\gamma_{s_i^\diamond}^\diamond, P_{s_i^\diamond}^\diamond\}_{i=1}^{\tilde{K}^\diamond}$ . Therefore,  $\tilde{K}^\diamond$ ,  $S^\diamond$ , and  $\{\gamma_{s_i^\diamond}^\diamond, P_{s_i^\diamond}^\diamond\}_{i=1}^{\tilde{K}^\diamond}$  with  $1 - \tilde{K}^\diamond\varepsilon = 0$  cannot be optimal, which contradicts with the initial assumption. Hence, the solution of (50) must satisfy  $1 - \tilde{K}\varepsilon > 0$ . Based on this inequality,  $\mu_{s_i}$  is defined as follows:

$$\mu_{s_i} \triangleq \gamma_{s_i} / (1 - \tilde{K}\varepsilon) \quad (54)$$

for  $i \in \{1, \dots, \tilde{K}\}$ . Thus, the optimization problem in (5) can be obtained by substituting the new variables defined in (54) into the optimization problem in (50)-(53). ■

### C. Proof of Proposition 3

Under the assumption in the proposition, the optimization problem in (9) can be expressed for  $\tilde{K}$  channels as follows:

$$\begin{aligned} &\max_{S \in B^{\tilde{K}}} \max_{\{\mu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} (1 - \tilde{K}\varepsilon) \sum_{i=1}^{\tilde{K}} \mu_{s_i} C_{s_i}(P_{s_i}) \\ &\text{subject to } \sum_{i=1}^{\tilde{K}} \mu_{s_i} P_{s_i} \leq \frac{P_{\text{av}}}{(1 - \tilde{K}\varepsilon)} \\ &P_{s_i} \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, \tilde{K}\} \\ &\sum_{i=1}^{\tilde{K}} \mu_{s_i} = 1, \mu_{s_i} \geq 0, \forall i \in \{1, \dots, \tilde{K}\} \end{aligned} \quad (55)$$

Then, based on a similar approach to that in Proposition 1 of [20], the problem in (55) can be stated as

$$\begin{aligned} &\max_{S \in B^{\tilde{K}}} \max_{\{\mu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}} (1 - \tilde{K}\varepsilon) \sum_{i=1}^{\tilde{K}} \mu_{s_i} C_{\text{max}}^S(P_{s_i}) \\ &\text{subject to } \sum_{i=1}^{\tilde{K}} \mu_{s_i} P_{s_i} \leq \frac{P_{\text{av}}}{(1 - \tilde{K}\varepsilon)} \\ &P_{s_i} \in [0, P_{\text{pk}}], \forall i \in \{1, \dots, \tilde{K}\} \\ &\sum_{i=1}^{\tilde{K}} \mu_{s_i} = 1, \mu_{s_i} \geq 0, \forall i \in \{1, \dots, \tilde{K}\} \end{aligned} \quad (56)$$

where  $C_{\text{max}}^S(P)$  is defined as

$$C_{\text{max}}^S(P) \triangleq \max_{m \in S} C_m(P). \quad (57)$$

That is, since the optimal solution involves the use of the best channel (among the given set of channels) for each power level (cf. (57)), the problem in (55) can be solved based on (56).

It is noted from (56) that, for each  $S$ , the aim is to find the optimal  $\{\mu_{s_i}, P_{s_i}\}_{i=1}^{\tilde{K}}$  for maximizing the convex combination of the  $C_{\text{max}}^S(P_{s_i})$  terms subject to the constraints on the average and peak powers. This formulation for each  $S$  has the same form as the problem formulation in eqn. (3) of [20];

hence, similar to Proposition 4 in [20], it can be shown that the optimal  $\{\mu_{s_i}, P_{s_i}\}_{i=1}^K$  has at most two nonzero  $\mu_{s_i}$  for each  $S$  (i.e., channel switching between at most two different channels is optimal for each  $S$ ). Therefore, the problem in (56) can be expressed as follows:

$$\max_{S \in B^K} \max_{\mu, \tilde{P}_1, \tilde{P}_2} (1 - \bar{K}\varepsilon) \left( \mu C_{\max}^S(\tilde{P}_1) + (1 - \mu) C_{\max}^S(\tilde{P}_2) \right) \quad (58)$$

$$\text{subject to } \mu \tilde{P}_1 + (1 - \mu) \tilde{P}_2 \leq \frac{P_{\text{av}}}{1 - \bar{K}\varepsilon} \quad (59)$$

$$\tilde{P}_1 \in [0, P_{\text{pk}}], \tilde{P}_2 \in [0, P_{\text{pk}}] \quad (60)$$

$$\mu \in [0, 1] \quad (61)$$

where  $\tilde{P}_1$  and  $\tilde{P}_2$  denote the average transmit powers allocated to channel  $i$  and channel  $j$ , respectively, with  $i = \arg \max_{l \in S} C_l(\tilde{P}_1)$  and  $j = \arg \max_{l \in S} C_l(\tilde{P}_2)$ .

It is noted that  $C_{\max}^S$  in (58) is maximized with respect to set  $S$ , and  $S$  does not depend on the other parameters,  $\mu$ ,  $\tilde{P}_1$ , and  $\tilde{P}_2$ . Therefore, the maximization with respect to  $S$  can be considered first for simplifying the problem in (58)-(61). For that purpose, the following expressions are obtained for  $C_{\max}^S$ :

$$\max_{S \in B^K} C_{\max}^S(P) = \max_{S \in B^K} \max_{m \in S} C_m(P) \quad (62)$$

$$= \max_{l \in \{1, \dots, K\}} C_l(P) \quad (63)$$

$$= C_{\max}(P) \quad (64)$$

where (62) follows from the definition of  $C_{\max}^S$  in (57), (63) is obtained based on the definition of  $B^K$  in (4), and finally (64) is due to (12). Based on (62)-(64), the problem in (58)-(61) can be stated as follows:

$$\max_{\mu, \tilde{P}_1, \tilde{P}_2} (1 - \bar{K}\varepsilon) \left( \mu C_{\max}(\tilde{P}_1) + (1 - \mu) C_{\max}(\tilde{P}_2) \right) \quad (65)$$

$$\text{subject to } \mu \tilde{P}_1 + (1 - \mu) \tilde{P}_2 \leq \frac{P_{\text{av}}}{1 - \bar{K}\varepsilon} \quad (66)$$

$$\tilde{P}_1 \in [0, P_{\text{pk}}], \tilde{P}_2 \in [0, P_{\text{pk}}] \quad (67)$$

$$\mu \in [0, 1] \quad (68)$$

where  $\tilde{P}_1$  and  $\tilde{P}_2$  denote the average transmit powers allocated to channel  $i$  and channel  $j$ , respectively, with  $i = \arg \max_{l \in \{1, \dots, K\}} C_l(\tilde{P}_1)$  and  $j = \arg \max_{l \in \{1, \dots, K\}} C_l(\tilde{P}_2)$ .

Next, consider the optimization problem in (65)-(68) for  $\frac{P_{\text{av}}}{(1 - \bar{K}\varepsilon)} < P_{\text{pk}}$ . Similarly to Lemma 1 in [20], it is obtained that the optimal  $\mu$ ,  $\tilde{P}_1$ , and  $\tilde{P}_2$  satisfy the average power constraint with equality; that is,  $\mu \tilde{P}_1 + (1 - \mu) \tilde{P}_2 = \frac{P_{\text{av}}}{1 - \bar{K}\varepsilon}$ . Then, by considering (66) as an equality constraint and substituting the constraints in (66)-(68) into the objective function and specifying the search space, it is obtained that the achieved capacity for  $\frac{P_{\text{av}}}{1 - \bar{K}\varepsilon} < P_{\text{pk}}$  can be calculated by solving the optimization problem in (11). Otherwise, i.e., if  $\frac{P_{\text{av}}}{1 - \bar{K}\varepsilon} \geq P_{\text{pk}}$ , then the solution of the optimization problem in (65)-(68) can easily be obtained as  $(1 - \bar{K}\varepsilon) C_{\max}(P_{\text{pk}})$ . ■

#### D. Proof of Lemma 1

Let channel  $i$  and channel  $j$  denote the channels corresponding to the maximum capacities for power levels  $P/\alpha$  and  $P/\beta$ , respectively; that is,  $C_{\max}(P/\alpha) = C_i(P/\alpha)$  and  $C_{\max}(P/\beta) = C_j(P/\beta)$  where  $i = \arg \max_{l \in \{1, \dots, K\}} C_l(P/\alpha)$  and  $j = \arg \max_{l \in \{1, \dots, K\}} C_l(P/\beta)$ .

First, consider the case of  $i = j$ . Then,  $C_{\max}(P/\alpha) = C_i(P/\alpha)$  and  $C_{\max}(P/\beta) = C_i(P/\beta)$ . Since the capacity curves are strictly concave and  $C_i(P) = 0$  for  $P = 0$ ,  $\forall i \in \{1, \dots, K\}$  (cf. (1)), the following relation can be obtained based on the definition of concavity:

$$\frac{\beta}{\alpha} C_i\left(\frac{P}{\beta}\right) + \left(1 - \frac{\beta}{\alpha}\right) C_i(0) < C_i\left(\frac{P}{\alpha}\right) \quad (69)$$

where  $\beta/\alpha < 1$  as the statement in the lemma is for  $\alpha > \beta$  and  $\alpha, \beta \in (0, 1)$ . Thus, it is obtained from (69) that  $\beta C_{\max}(P/\beta) < \alpha C_{\max}(P/\alpha)$  as claimed in the lemma.

Next, consider the case of  $i \neq j$ . Since  $C_{\max}(P/\alpha) = C_i(P/\alpha)$ ,  $C_{\max}(P/\beta) = C_j(P/\beta)$ , and  $C_i$  and  $C_j$  are monotone increasing and continuous functions, then there exists a single point  $P/\gamma \in (P/\alpha, P/\beta)$  for  $\beta < \gamma < \alpha$  at which the capacity curves of channel  $i$  and channel  $j$  intersect; that is,  $C_i(P/\gamma) = C_j(P/\gamma)$ . Now considering the capacity of channel  $j$  for power levels  $P/\gamma$  and  $P/\beta$ , it can be shown that  $\beta C_j(P/\beta) < \gamma C_j(P/\gamma)$  based on a similar approach to that in (69). Similarly, for channel  $i$ , the following relation is obtained:  $\gamma C_i(P/\gamma) < \alpha C_i(P/\alpha)$ . Since  $C_i(P/\gamma) = C_j(P/\gamma)$ , these two inequalities imply that  $\beta C_j(P/\beta) < \alpha C_i(P/\alpha)$ , which is equivalent to  $\beta C_{\max}(P/\beta) < \alpha C_{\max}(P/\alpha)$  as claimed in the lemma. ■

#### E. Proof of Proposition 4

The aim is to prove that the statement in the proposition holds for all the cases specified in (15). Firstly, for  $\varepsilon \geq \frac{1}{2}$ , the constraint in (13) cannot be satisfied for any  $\bar{K}$ , and consequently, channel switching is not feasible in this case. Therefore, if  $\varepsilon \geq \frac{1}{2}$ , the maximum average capacity via channel switching can be specified as  $C_{\text{css}} = 0$ .<sup>7</sup> Secondly, if  $\varepsilon < \frac{1}{2}$  and  $P_{\text{av}}/(1 - 2\varepsilon) \geq P_{\text{pk}}$ , then the maximum average capacity achieved by performing optimal channel switching between two channels is obtained based on (11) as  $\psi(2) = (1 - 2\varepsilon) C_{\max}(P_{\text{pk}})$ . On the other hand, for optimal channel switching among  $M > 2$  channels, the following arguments can be provided. Since  $P_{\text{av}}/(1 - 2\varepsilon) \geq P_{\text{pk}}$  in this case, it is obtained that  $P_{\text{av}}/(1 - M\varepsilon) > P_{\text{pk}}$  for  $M > 2$  and  $M < 1/\varepsilon$ , which is the constraint in (13). Then, it follows from (11) that  $\psi(M) = (1 - M\varepsilon) C_{\max}(P_{\text{pk}})$  for  $M > 2$ . Since  $\psi(2) > \psi(M)$  for  $M > 2$ , it is concluded for  $\varepsilon < \frac{1}{2}$  and  $P_{\text{av}}/(1 - 2\varepsilon) \geq P_{\text{pk}}$  that the optimal channel switching strategy with two channels achieves a higher average capacity than the optimal channel switching strategies with more than two channels, and that the maximum average capacity achieved by the optimal channel switching strategy with two channels is equal to  $C_{\text{css}} = (1 - 2\varepsilon) C_{\max}(P_{\text{pk}})$ , as

<sup>7</sup>In this case, the solution of the optimization problem in (5) corresponds to the optimal single channel strategy (Case-1).

specified in (15). Finally, if  $\varepsilon < \frac{1}{2}$  and  $P_{av}/(1-2\varepsilon) < P_{pk}$ , the maximum average capacity for the channel switching strategy with  $\bar{K}$  channels can be obtained based on (11) as follows:<sup>8</sup>

$$\psi(\bar{K}) = \max_{\substack{\tilde{P}_1 \in [\frac{P_{av}}{1-\bar{K}\varepsilon}, P_{pk}] \\ \tilde{P}_2 \in [0, \frac{P_{av}}{1-\bar{K}\varepsilon}]}} (1 - \bar{K}\varepsilon) \left( \frac{\frac{P_{av}}{1-\bar{K}\varepsilon} - \tilde{P}_2}{\tilde{P}_1 - \tilde{P}_2} C_{\max}(\tilde{P}_1) + \frac{\tilde{P}_1 - \frac{P_{av}}{1-\bar{K}\varepsilon}}{\tilde{P}_1 - \tilde{P}_2} C_{\max}(\tilde{P}_2) \right) \quad (70)$$

Define  $\bar{P}_1$  and  $\bar{P}_2$  as  $\bar{P}_1 \triangleq (1 - \bar{K}\varepsilon) \tilde{P}_1$  and  $\bar{P}_2 \triangleq (1 - \bar{K}\varepsilon) \tilde{P}_2$ . Then,  $\psi(\bar{K})$  in (70) can be expressed as follows:

$$\psi(\bar{K}) = \max_{\substack{\bar{P}_1 \in [P_{av}, (1-\bar{K}\varepsilon)P_{pk}] \\ \bar{P}_2 \in [0, P_{av}]}} (1 - \bar{K}\varepsilon) \left( \frac{P_{av} - \bar{P}_2}{\bar{P}_1 - \bar{P}_2} \times C_{\max} \left( \frac{\bar{P}_1}{1 - \bar{K}\varepsilon} \right) + \frac{\bar{P}_1 - P_{av}}{\bar{P}_1 - \bar{P}_2} C_{\max} \left( \frac{\bar{P}_2}{1 - \bar{K}\varepsilon} \right) \right). \quad (71)$$

For the optimal channel switching strategy with two channels, the maximum average capacity is given by

$$\psi(2) = \max_{\substack{\bar{P}_1 \in [P_{av}, (1-2\varepsilon)P_{pk}] \\ \bar{P}_2 \in [0, P_{av}]}} (1 - 2\varepsilon) \left( \frac{P_{av} - \bar{P}_2}{\bar{P}_1 - \bar{P}_2} C_{\max} \left( \frac{\bar{P}_1}{1 - 2\varepsilon} \right) + \frac{\bar{P}_1 - P_{av}}{\bar{P}_1 - \bar{P}_2} C_{\max} \left( \frac{\bar{P}_2}{1 - 2\varepsilon} \right) \right). \quad (72)$$

The aim is to prove that  $\psi(2) > \psi(M)$  for  $M > 2$ , where  $\psi(M)$  denotes the maximum average capacity achieved by optimal channel switching among  $M > 2$  channels. To that aim, define a new optimization problem identical to (72) except that the search space for  $\bar{P}_1$  is  $[P_{av}, (1 - M\varepsilon)P_{pk}]$  instead of  $[P_{av}, (1 - 2\varepsilon)P_{pk}]$ . Let  $\xi$  denote the solution of this problem, which can be stated as follows:

$$\xi = \max_{\substack{\bar{P}_1 \in [P_{av}, (1-M\varepsilon)P_{pk}] \\ \bar{P}_2 \in [0, P_{av}]}} (1 - 2\varepsilon) \left( \frac{P_{av} - \bar{P}_2}{\bar{P}_1 - \bar{P}_2} C_{\max} \left( \frac{\bar{P}_1}{1 - 2\varepsilon} \right) + \frac{\bar{P}_1 - P_{av}}{\bar{P}_1 - \bar{P}_2} C_{\max} \left( \frac{\bar{P}_2}{1 - 2\varepsilon} \right) \right). \quad (73)$$

The optimization problem in (73) is the same as the problem in (72) except that the search space for  $\bar{P}_1$  in (73) is a subset of that in (72). Therefore, it is obtained that  $\psi(2) \geq \xi$ . Also, the following relations can be derived for  $M > 2$  based on

(14) in Lemma 1:

$$(1 - 2\varepsilon) C_{\max} \left( \frac{\bar{P}_1}{1 - 2\varepsilon} \right) > (1 - M\varepsilon) C_{\max} \left( \frac{\bar{P}_1}{1 - M\varepsilon} \right), \quad \forall \bar{P}_1 \in [P_{av}, (1 - M\varepsilon)P_{pk}] \quad (74)$$

$$(1 - 2\varepsilon) C_{\max} \left( \frac{\bar{P}_2}{1 - 2\varepsilon} \right) \geq (1 - M\varepsilon) C_{\max} \left( \frac{\bar{P}_2}{1 - M\varepsilon} \right), \quad \forall \bar{P}_2 \in [0, P_{av}] \quad (75)$$

where the equality sign in (75) is included to cover the case of  $\bar{P}_2 = 0$ . Based on (74), (75), and the fact that the search spaces of the optimization problems in (71) and (73) are the same for  $\bar{K} = M$ , it is obtained that  $\xi > \psi(M)$  for  $M > 2$ . (Note that  $\frac{(P_{av} - \bar{P}_2)}{(\bar{P}_1 - \bar{P}_2)} > 0$  and  $\frac{(\bar{P}_1 - P_{av})}{(\bar{P}_1 - \bar{P}_2)} \geq 0$ .) Therefore, it is concluded that  $\psi(2) > \psi(M)$  for  $M > 2$  since  $\psi(2) \geq \xi$  as shown previously. Hence, in accordance with (15), it is shown that  $C_{css} = \psi(2)$  for  $\varepsilon < \frac{1}{2}$  and  $P_{av}/(1 - 2\varepsilon) < P_{pk}$ , where  $\psi(\cdot)$  is as defined in (11) (cf. (72)). To sum up, the optimal channel switching strategy is to switch between two channels and the achieved maximum average capacity can be obtained as in (15). ■

#### F. Proof of Proposition 5

For  $\varepsilon < 1/2$  and  $P_{av}/(1 - 2\varepsilon) < P_{pk}$ , the optimal single channel strategy achieves an average capacity of  $C_{scs} = (1 - \varepsilon)C_{\max}(\frac{P_{av}}{1 - \varepsilon})$ , which is obtained from (7) since  $\frac{P_{av}}{1 - \varepsilon} < \frac{P_{av}}{1 - 2\varepsilon}$  and  $\frac{P_{av}}{1 - 2\varepsilon} < P_{pk}$ . Also, the maximum average capacity  $C_{css}$  obtained by the optimal channel switching strategy can be calculated from (26) in this case. The aim is to prove that under the assumptions in the proposition, if the condition in (27) holds, then the optimal single channel strategy achieves a higher average capacity than the optimal channel switching strategy; that is,  $C_{scs} > C_{css}$ . The assumption in the proposition states that the first-order derivative of  $C_{\max}(P)$  in (12) exists in an interval around  $\frac{P_{av}}{1 - 2\varepsilon}$ . Then its derivative at  $\frac{P_{av}}{1 - 2\varepsilon}$  can be obtained from (1) as follows:

$$C'_{\max} \left( \frac{P_{av}}{1 - 2\varepsilon} \right) = \frac{(1 - 2\varepsilon)B_{i^*} \log_2 e}{(1 - 2\varepsilon)N_{i^*} B_{i^*} + P_{av}} \quad (76)$$

where  $i^* = \arg \max_{i \in \{1, \dots, K\}} C_i(\frac{P_{av}}{1 - 2\varepsilon})$ . From (76) and the definition of  $\bar{P} \triangleq \frac{P}{1 - 2\varepsilon}$ , the condition in (27) can be expressed in the following form:

$$C_{\max}(\bar{P}) \leq C_{\max} \left( \frac{P_{av}}{1 - 2\varepsilon} \right) + \left( \bar{P} - \frac{P_{av}}{1 - 2\varepsilon} \right) C'_{\max} \left( \frac{P_{av}}{1 - 2\varepsilon} \right), \quad \forall \bar{P} \in [0, P_{pk}]. \quad (77)$$

It is noted that the problem for  $C_{css}$  in (26) can be expressed similarly to (65)-(68) as follows:

$$\max_{\mu, \tilde{P}_1, \tilde{P}_2} (1 - 2\varepsilon)(\mu C_{\max}(\tilde{P}_1) + (1 - \mu)C_{\max}(\tilde{P}_2)) \quad (78)$$

$$\text{subject to } \mu \tilde{P}_1 + (1 - \mu)\tilde{P}_2 = \frac{P_{av}}{1 - 2\varepsilon} \quad (79)$$

$$\tilde{P}_1 \in [0, P_{pk}], \tilde{P}_2 \in [0, P_{pk}] \quad (80)$$

$$\mu \in [0, 1] \quad (81)$$

<sup>8</sup>The equation in (70) is valid if  $P_{av}/(1 - \bar{K}\varepsilon) < P_{pk}$ . Otherwise, it is easy to prove that  $\psi(2) > \psi(M)$  for  $M > 2$  based on Lemma 1 since  $\psi(2) \geq (1 - 2\varepsilon)C_{\max}(P_{av}/(1 - 2\varepsilon)) > (1 - M\varepsilon)C_{\max}(P_{av}/(1 - M\varepsilon)) \geq (1 - M\varepsilon)C_{\max}(P_{pk}) = \psi(M)$ .

Then, for the solution of the channel switching strategy in (78)-(81) denoted as  $\mu^*$ ,  $\tilde{P}_1^*$ , and  $\tilde{P}_2^*$ , the following expressions can be obtained:

$$C_{\text{css}} = (1 - 2\varepsilon)(\mu^* C_{\text{max}}(\tilde{P}_1^*) + (1 - \mu^*)C_{\text{max}}(\tilde{P}_2^*)) \quad (82)$$

$$\leq (1 - 2\varepsilon) \left( \left( \mu^* \tilde{P}_1^* + (1 - \mu^*) \tilde{P}_2^* - \frac{P_{\text{av}}}{1 - 2\varepsilon} \right) \right. \\ \left. \times C'_{\text{max}} \left( \frac{P_{\text{av}}}{1 - 2\varepsilon} \right) + C_{\text{max}} \left( \frac{P_{\text{av}}}{1 - 2\varepsilon} \right) \right) \quad (83)$$

$$= (1 - 2\varepsilon) C_{\text{max}} \left( \frac{P_{\text{av}}}{1 - 2\varepsilon} \right) \quad (84)$$

$$< (1 - \varepsilon) C_{\text{max}} \left( \frac{P_{\text{av}}}{1 - \varepsilon} \right) \quad (85)$$

$$= C_{\text{scs}} \quad (86)$$

where  $\tilde{P}_1^*, \tilde{P}_2^* \in [0, P_{\text{pk}}]$  and  $\mu^* \geq 0$ . The equality in (82) follows from (78)-(81), and the inequality in (83) is obtained based on (77). The equality in (84) holds since  $\mu^*$ ,  $\tilde{P}_1^*$ , and  $\tilde{P}_2^*$  satisfy the average power constraint in (79); that is,  $\mu^* \tilde{P}_1^* + (1 - \mu^*) \tilde{P}_2^* = \frac{P_{\text{av}}}{1 - 2\varepsilon}$ , and since  $C'_{\text{max}}(\frac{P_{\text{av}}}{1 - 2\varepsilon})$  is finite. Finally, (85) is obtained due to (14) in Lemma 1, which results in the maximum average capacity achieved via the optimal single channel strategy as noted in (86). From (82)-(86), it is concluded that the optimal single channel strategy outperforms the optimal channel switching strategy in terms of the maximum average capacity if the assumptions and the condition in the proposition hold. ■

## REFERENCES

- [1] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.
- [2] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels— Part I: Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.
- [3] Z. Shen, J. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.
- [4] F. F. Digham, "Joint power and channel allocation for cognitive radios," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Mar. 2008, pp. 882–887.
- [5] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, July 2005.
- [6] L. Dai, S. Sfar, and K. B. Letaief, "Optimal antenna selection based on capacity maximization for MIMO systems in correlated channels," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 563–573, Mar. 2006.
- [7] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Sig. Processing*, vol. 57, no. 5, pp. 1655–1669, May 2009.
- [8] B. Dulek and S. Gezici, "Detector randomization and stochastic signaling for minimum probability of error receivers," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 923–928, Apr. 2012.
- [9] M. E. Tutay, S. Gezici, and O. Arikan, "Optimal detector randomization for multiuser communications systems," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2876–2889, July 2013.
- [10] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1948–1960, April 2010.
- [11] C. Goken, S. Gezici, and O. Arikan, "Optimal stochastic signaling for power-constrained binary communications systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3650–3661, Dec. 2010.
- [12] M. Azizoglu, "Convexity properties in binary detection problems," *IEEE Trans. Inform. Theory*, vol. 42, no. 4, pp. 1316–1321, July 1996.
- [13] B. Dulek, S. Gezici, and O. Arikan, "Convexity properties of detection probability under additive Gaussian noise: Optimal signaling and jamming strategies," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3303–3310, July 2013.
- [14] S. Bayram, N. D. Vanli, B. Dulek, I. Sezer, and S. Gezici, "Optimum power allocation for average power constrained jammers in the presence of non-Gaussian noise," *IEEE Commun. Letters*, vol. 16, no. 8, pp. 1153–1156, Aug. 2012.
- [15] B. Dulek, N. D. Vanli, S. Gezici, and P. K. Varshney, "Optimum power randomization for the minimization of outage probability," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4627–4637, Sep. 2013.
- [16] Y. Ma and C. C. Chai, "Unified error probability analysis for generalized selection combining in Nakagami fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 11, pp. 2198–2210, Nov. 2000.
- [17] J. A. Ritcey and M. Azizoglu, "Performance analysis of generalized selection combining with switching constraints," *IEEE Commun. Letters*, vol. 4, no. 5, pp. 152–154, May 2000.
- [18] B. Dulek, M. E. Tutay, S. Gezici, and P. K. Varshney, "Optimal signaling and detector design for M-ary communication systems in the presence of multiple additive noise channels," *Digital Signal Processing*, vol. 26, pp. 153–168, Mar. 2014.
- [19] M. E. Tutay, S. Gezici, H. Soganci, and O. Arikan, "Optimal channel switching over Gaussian channels under average power and cost constraints," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1907–1922, May 2015.
- [20] A. D. Sezer, S. Gezici, and H. Inaltekin, "Optimal channel switching strategy for average capacity maximization," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2143–2157, June 2015.
- [21] P. Bahl, R. Chandra, and J. Dunagan, "SSCH: Slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks," in *Proc. Annual Int. Conf. on Mobile Computing and Networking (Mobicom)*, 2004, pp. 216–230.
- [22] S. Krishnamurthy, M. Thoppian, S. Venkatesan, and R. Prakash, "Control channel based MAC-layer configuration, routing and situation awareness for cognitive radio networks," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, vol. 1, Oct. 2005, pp. 455–460.
- [23] M. Yun, Y. Zhou, A. Arora, and H.-A. Choi, "Channel-assignment and scheduling in wireless mesh networks considering switching overhead," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, June 2009, pp. 1–6.
- [24] S. Bayhan and F. Alagoz, "Scheduling in centralized cognitive radio networks for energy efficiency," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 582–595, Feb. 2013.
- [25] S. Eryigit, S. Bayhan, and T. Tugcu, "Channel switching cost aware and energy-efficient cooperative sensing scheduling for cognitive radio networks," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, June 2013, pp. 2633–2638.
- [26] H. Ma, L. Zheng, X. Ma, and Y. Luo, "Spectrum aware routing for multi-hop cognitive radio networks with a single transceiver," in *Proc. Int. Conf. on Cognitive Radio Oriented Wireless Netw. and Commun. (CrownCom)*, May 2008, pp. 1–6.
- [27] S. Verdú, *Multiuser Communications*. New York: Cambridge University Press, 1998.
- [28] G. Amariuca, S. Wei, and R. Kannan, "Gaussian jamming in block-fading channels under long term power constraints," in *IEEE International Symposium on Info. Theory (ISIT)*, June 2007, pp. 1001–1005.
- [29] J. Gao, S. Vorobyov, H. Jiang, and H. Poor, "Worst-case jamming on MIMO Gaussian channels," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5821–5836, Nov. 2015.
- [30] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.
- [31] S. Gezici, H. Celebi, H. V. Poor, and H. Arslan, "Fundamental limits on time delay estimation in dispersed spectrum cognitive radio systems," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 1, pp. 78–83, Jan. 2009.
- [32] B. Sadhu and R. Harjani, *Cognitive Radio Receiver Front-Ends RF/Analog Circuit Techniques*. Springer, 2014.
- [33] A. Ramadan, J. Costantine, K. Kaban, Y. Tawk, C. Christodoulou, and M. Al-Husseini, "A tunable RF front-end receiver for cognitive radio applications," in *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, July 2014, pp. 1212–1213.
- [34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.
- [35] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York: Cambridge University Press, 2012.
- [36] L. L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised*

and *Space-Time Coded OFDM, CDMA and MC-CDMA Systems*, 2nd ed. Wiley-IEEE Press, 2004.

- [37] E. Biglieri, A. J. Goldsmith, L. J. Greenstein, N. Mandayam, and H. V. Poor, *Principles of Cognitive Radio*. New York: Cambridge University Press, 2012.
- [38] T. J. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.
- [39] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.



**Ahmet Dundar Sezer** was born in 1989 in Emet, Kutahya, Turkey. He received both his B.S. and M.S. degrees in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey, in 2011 and 2013, respectively. He is currently working towards the Ph.D. degree at Bilkent University. His current research interests include signal processing, wireless communications, and optimization.



**Sinan Gezici** (S'03, M'06, SM'11) 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 *Journal of Communications and Networks*.