Optimal Transmission-Sensing-Reception Strategies for Full-duplex Dynamic Spectrum Access

Wessam Afifi and Marwan Krunz
Department of Electrical and Computer Engineering, University of Arizona, USA
E-mail: {wessamafifi, krunz}@email.arizona.edu

Technical Report
TR-UA-ECE-2013-4
Last update: November 27, 2013

Abstract

In this paper, we exploit recent advances in full-duplex (FD) communications and self-interference suppression (SIS) to improve the performance of an opportunistic spectrum access (OSA) system. Specifically, we consider secondary users (SUs) that are equipped with SIS-capable radios. These radios can operate in a simultaneous transmission-and-sensing (TS) mode to improve the detection probability of primary users (PUs), or in a simultaneous transmission-and-reception (TR) mode to enhance the SU throughput. The radios can also revert to the standard sensing-only (SO) mode or perform channel switching (CS). The competing goals of the full-duplex TS and TR modes give rise to a spectrum-awareness/efficiency tradeoff, which can be optimized by allowing the SU link to adaptively switch between various modes, depending on the forecasted PU dynamics. In practice, SIS is imperfect, resulting in residual self-interference that degrades the sensing performance in the TS mode. Accordingly, we adopt a waveform-based sensing approach, which allows an SU to detect (with high accuracy) the PU signal in the presence of self-interference (and noise). In such a context, we analyze the sensing performance in the TS mode by deriving the false-alarm and detection probabilities. We also derive the throughput and the PU-SU collision probability for the TS and TR modes, which we then use to establish an optimal mode-selection strategy that maximizes an SU utility function subject to a constraint on the PU collision probability. This utility rewards the SU instantly for successful communication (throughput), but also includes a long-term component that depends on the outcomes of the action taken by the SU (the selected mode from the set {TR, TS, SO, CS}). Our results show that the proposed adaptive strategy results in about 50% reduction in the collision probability and twice the throughput of the half-duplex case. The results also indicate that the SU should operate in the TR mode if it has a high belief regarding the PU idleness over a given channel. As this belief decreases, the SU should switch to the TS mode to monitor any change in the PU activity while transmitting. At very low belief values, where the PU is highly likely to be active, the SU should switch to another channel.

Index Terms

Self-interference suppression, full-duplex communications, opportunistic spectrum access, spectrum awareness/efficiency trade-off, cognitive radios.

I. Introduction

Until recently, the idea of simultaneous transmission and reception over the same channel (STAR-S) was not deemed possible. The reason is that while a wireless device is receiving data, its own transmission produces strong self-interference, which makes the decoding process impossible. One way to solve this problem is to suppress the node's self-interference. However, traditional self-interference suppression (SIS) techniques (e.g., RF and digital interference cancellation) have not been sufficient. Even simultaneous transmission and reception over different frequencies (STAR-D) is not straightforward, particularly when the transmit and receive bands are not sufficiently separated (in practice, filters are not ideal, and sidelobes/spectral leakage is deemed to occur). In this paper, we focus on STAR-S (the more challenging case), which we simply refer to as FD communication.

By combining novel and traditional SIS techniques, the authors in [1–5] have demonstrated the feasibility of FD communication. In [1], the authors proposed an antenna-based SIS technique in which two properly placed transmit antennas and one receive antenna are used to nullify self-interference at the receiving antenna. This technique has two problems. First, it generates additional interference in the far field, i.e., it increases the interference at other receivers. Second, it has a bandwidth limitation, as antenna placement is determined by a single carrier frequency. However, wireless transmissions typically involve multiple carriers. These concerns were addressed in [2], where the authors used only 2 antennas and proposed an interference cancellation mechanism based on signal inversion. Very recently, the authors in [6] proposed an FD system for 802.11ac devices using only one antenna. The main objective of these works is to bring down a node's self-interference to the noise level. As an example, a WiFi device has to suppress around 110 dB of its own transmitted signal (assuming a transmit power of 20 dBm) to reduce it to the noise level [6].

This is a technical report for our paper titled "Adaptive Transmission-Reception-Sensing Strategy for Cognitive Radios with Full-duplex Capabilities", which was submitted to IEEE DySPAN 14 conference.

In parallel with the developments in SIS techniques, there have been a number of works that exploit SIS/FD capabilities in network-protocol design, in the contexts of MIMO communications [7–10] and dynamic spectrum access (DSA) [11, 12]. Assuming multiple antennas per node, the authors in [10] addressed the issue of choosing between MIMO and FD, as both need multiple antennas. They showed that the optimal strategy is a combination of both schemes. In [13] the authors studied power control in wireless FD devices with imperfect SIS. They developed an optimal dynamic power allocation scheme that maximizes the sum-rate of a number of users.

In this paper, we consider a DSA network, where secondary users (SUs) have imperfect SIS capabilities, allowing them to suppress a fraction of their self-interference. This partial SIS capability can be exploited to support simultaneous transmission-and-sensing (TS) by the SU so as to reduce the collision probability with primary users (PUs), or simultaneous transmission-and-reception (TR) to enhance the SU throughput. The ability to operate in either mode gives rise to a spectrum awareness/efficiency tradeoff. More specifically, the SU may improve the spectrum utilization by operating in the TR mode, which will dramatically increase the throughput of the SU link. On the other hand, the SU may exploit its SIS capabilities in the TS mode, enabling it to monitor the PU activity while transmitting and to quickly vacate the channel whenever such activity is detected. This motivates the need for the optimal transmission-sensing-reception strategy introduced in this paper.

Because the sensing efficiency in the TS mode decreases as the SIS capabilities decrease, in some cases the SU needs to operate in a sensing-only (SO) mode to achieve an acceptable sensing outcome. Also, having a relatively high belief that the PU is active may return a high collision probability in the TS/TR modes. In that case, the SU should stop transmission and just monitors the channel. Considering the availability of multiple idle channels, the SU may decide to perform channel switching (CS) if the PU is more likely to return to the current operational channel.

An important aspect of the system design is to determine how the SU adaptively switch between different modes (TR, TS, SO, and CS), considering the highly dynamic spectrum environment and the possibility of colliding with PUs. Our objective is to find the optimal strategy that maximizes the SU's utility (e.g., goodput) under a constraint on the PU collision probability. This strategy is found to be threshold-based, with thresholds that depend on the SU's belief about the PU state. Based on this belief, the SU will take an optimal action and then update this belief according to the outcome of the action taken. The outcome is ACK/NACK in case of a transmission decision, free/busy in case of a sensing decision, and decoded/undecoded in case of reception. The SU may also get a combination of these outcomes in the TR and TS modes.

The problem of finding the optimal access strategy at an SU device has been studied before [14–17], but for half-duplex (HD) devices. In [14], the authors considered the quickest detection problem of the PU idle period when multiple PUs are present. In their scheme, the SU chooses an action from the following: spectrum sensing, channel switching, or data transmission. The authors in [15] studied the sensing-throughput tradeoff and proposed a scheme in which the SU can have multiple consecutive sensing or transmission periods, determined according to the SU's belief about the state of the PU. In their scheme, the SU has only two options: spectrum sensing or data transmission. The objective was to maximize the SU's utility, which rewards the SU for successful transmission and penalizes him for collisions. Another adaptive scheme was proposed in [17], where the authors added another possible action to the SU, namely staying idle. The motivation behind this action is to save energy when the probability that the PU is idle is very low.

In [11], we proposed applying SIS/FD in DSA systems and introduced the TR and TS modes. However, our treatment was limited to energy-based spectrum sensing (for the TS mode). Energy detection cannot differentiate between a PU signal and a residual self-interference signal. Hence, it is inefficient under low SIS capabilities. This problem is not present in waveform-based sensing, whereby the sensed waveform is contrasted with well-known patterns (pilots, preambles, etc.) of the PU signal. In [11], we also studied the traditional sensing-throughput tradeoff for the TR and TS modes and determined the optimal sensing and transmission durations for the SU that maximizes its throughput subject to a constraint on the SU/PU collision probability.

The contributions of this paper are as follows. First, we consider a DSA system where SUs are partially capable of SIS. We analyze the waveform-based spectrum sensing technique for the TS mode, which is crucial especially at imperfect SIS, and derive the false-alarm and detection probabilities. Second, we derive the probability of successful transmission for the SU, its achievable throughput, and the PU collision probability in both TS and TR modes, taking into consideration that SIS may be imperfect and assuming different channel conditions at the communicating SUs. Third, we propose an optimal adaptive strategy at the SU for switching between the TR, TS, SO, and CS modes. The criteria for choosing the optimal action is to maximize the SU's utility subject to a constraint on the PU collision probability. To achieve this goal, we formulate the problem as a partially observable decision process and analyze the four actions by formulating the myopic and long-term rewards. To the best of our knowledge, this is the first paper to address the optimal transmission-reception-sensing strategy for SUs with imperfect FD/SIS capabilities.

The rest of the paper is organized as follows. We describe the system model in Section II. In Section III, we derive the false-alarm and detection probabilities under waveform-based sensing for the TS mode and compare them with the HD case. We formulate the SU decision process and obtain the optimal adaptive SU spectrum access strategy in Section IV. Finally, we present our numerical results and conclude the paper in Sections V and VI, respectively.

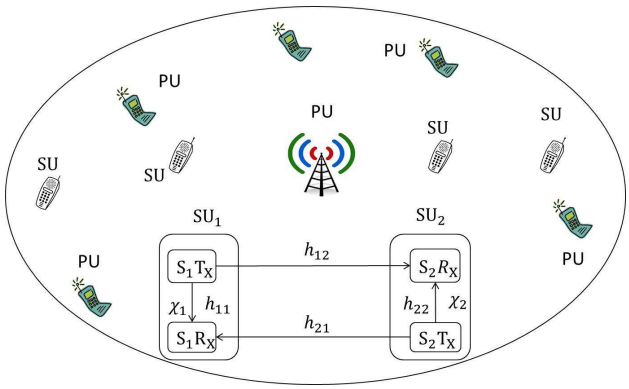


Fig. 1. System model of our DSA network, where SUs are equipped with SIS/FD capabilities and opportunistically access the spectrum of PUs. Each SU consists of a transceiver with a given SIS capability factor χ.

II. System Model

As shown in Figure 1, we consider a DSA network, where SUs are opportunistically operating on the licensed PUs channels. PUs can access the available channels at will, and are not aware of the SU's presence. The PU activity is modeled as an alternating ON/OFF random process. Let the OFF and ON durations be denoted by X and Y, with corresponding probability distributions f_X and f_Y, and means $\bar{X}$ and $\bar{Y}$, respectively. These distributions are assumed to be independent and can be constructed at the SU through measurements [18, 19].

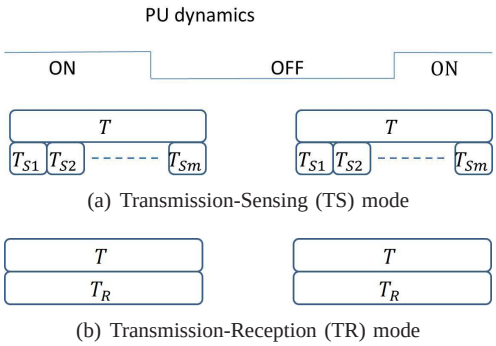
Each secondary device is capable of partial or complete SIS, enabling it to operate in the TS and TR modes, along with the SO and CS modes. We use χ_i to quantify the SIS capability of the ith SU, $\chi_i \in [0, 1]$. Specifically, χ_i is the ratio between the residual self-interference and the SIS reduction needed to reach the noise floor (in dB). If $\chi_i = 0$, the node can bring its self-interference down to the noise floor; otherwise, it can only suppress a fraction $1 - \chi_i$ of its self-interference (imperfect SIS). As an example, for a transmission power of 20 dBm and a noise floor of -90 dBm, an 90 dB SIS translates into $\chi_i = (110 - 90)/110 = 0.18$. χ_i may differ from one node to another, depending on the employed SIS technique.

We assume that at a given time instant, and a given frequency, only one SU link is active in a given geographical area. Hence, we focus on the case where different SU links cannot interfere with each other, for example, by implementing an appropriate channel access scheme. Existing techniques can be used to tackle the issue of secondary-secondary interference (see [20], for example), and will not be addressed here. Let P_i and σ_i^2 denote the transmission power and noise variance at node i, and let h_{ij} be the channel gain between transmitter i and receiver j. Although the opportunistic spectrum consists of multiple channels, the SU can only monitor/operate on one channel at a time. Sensing multiple channels has already been discussed in several papers, and can be easily incorporated [21].

To sum up how our system works, the SU starts its communication with an initial belief value. Depending on the adaptive access strategy, the SU chooses the optimal action that maximizes its utility while maintaining a certain QoS threshold for the PU communication. The action's outcome may be ACK/NACK in case of transmission, free/busy in case of sensing, and decoded/undecoded in case of reception. The SU may also get a combination of these outcomes in the TR and TS modes. Depending on these outcomes, the SU updates its belief about the PU state. Getting an ACK/free/decoded outcome will increase your belief that the PU is idle with a certain degree. However, getting a NACK/busy/undecoded outcome will increase your belief that the PU is busy. Based on this belief, and according to our spectrum access strategy, the SU will take the optimal action, and so on.

A. SU Operation Modes

1) TS mode: Using SIS techniques, the SU can carry out the spectrum sensing process while transmitting its data. This has two advantages over the Listen-Before-Talk (LBT) scheme. First, from the SU's perspective, transmitting while sensing increases the SU throughput, and reduces the frequency of interrupting its transmission (such interruptions are detrimental to any real-time communications). Second, the SU can monitor the PU activity while transmitting. Hence, a better PU detection performance is achieved. This parallel sensing process may be done over multiple (consecutive) short periods instead of one long sensing period. To do that, the SU performs m sensing actions T_{Si}, $i = 1, 2, \ldots, m$, while transmitting data for a period of T seconds (see Figure 2(a)). The motivation behind this approach is to account for the tradeoff between sensing efficiency and the timeliness in detecting PU activity. On the one hand, increasing the sensing duration improves the sensing efficiency. However, such an increase implies delaying the time to make a decision regarding the change of PU activity. Thus, in the TS mode, we have a total of m sensing durations. If at the end of any given sensing period, PU activity is detected, the SU aborts its current transmission and updates its belief to determine the next action. We use the term *FD sensing* to refer to the sensing process



(a) Transmission-Sensing (TS) mode

(b) Transmission-Reception (TR) mode

Fig. 2. FD operation modes for the SU.

in the TS mode. Note that under imperfect SIS, such sensing has worse performance than the traditional SO mode due to the residual self-interference signal.

2) TR mode: In the TR mode, the SU transmits and receives data simultaneously over the same channel, as shown in Figure 2(b). Denote the transmission and reception durations by T and T_R, respectively. For simplicity, we assume that $T_R = T$. Although operating in the TR mode enhances the SU's throughput, the SU will not be able to monitor the PU state. Hence, the probability of colliding with the PU will be higher than in the TS mode.

3) SO mode: In this mode, the SU senses the spectrum for a duration T_S, which we refer to as *HD sensing*. Under imperfect SIS, the TS mode is not always efficient. Hence, the SU may switch to the SO mode to get more accurate sensing results.

4) CS mode: The SU may switch to another channel and carry out spectrum sensing on this new channel if the SU believes that the PU is very likely to return to the currently used channel. Existing techniques can be used to select the channel sensing order (see [22], for example). However, any previous information about the new channel that was obtained from prior sensing attempts are discarded.

Although we will not consider the transmission-only (TO) mode as an option, we will use it for comparison purposes. The reason for not considering it is that the sensing cost is almost negligible. Hence, there is no advantage of the TO mode over the TS mode.

III. Waveform-based FD Sensing

A significant amount of DSA literature has focused on energy-based sensing. Despite its simplicity, this technique cannot differentiate between different types of users. In the TS mode, residual self-interference can cause energy detection to wrongly indicate PU activity. In this paper, we study the sensing performance of the TS mode, assuming waveform-based sensing.

Waveform-based sensing utilizes known patterns in the PU signal, such as preambles and pilot symbols. These patterns are typically used for channel estimation, synchronization, equalization, etc. To detect the presence of the PU signal, waveform-based sensing correlates a known pattern with the received signal [23, 24]. In this section, we analyze waveform-based sensing under FD operation and derive the false-alarm and detection probabilities for the SU, assuming a given SIS factor χ.

In the TS mode, the hypothesis test of whether the channel is occupied by a PU or not can be formulated as follows:

$$r(n) = \begin{cases} \chi\, s(n) + w(n) & H_0 \text{ (if PU is idle)} \quad (1a) \\ l(n) + \chi\, s(n) + w(n) & H_1 \text{ (if PU is busy)} \quad (1b) \end{cases}$$

where $r(n)$ is the nth sample of the discretized received signal, $s(n)$ is the self-interfering SU signal, $l(n)$ is the received PU signal, and $w(n)$ is the additive white Gaussian noise with variance σ_w^2. We assume that $s(n)$ is a zero-mean complex random signal with variance σ_s^2. We also assume that all signal samples are independent, hence $r(n)$'s are also independent.

In the case of HD sensing, where no self-interference is present, the hypothesis test can be written as:

$$\tilde{r}(n) = \begin{cases} w(n) & H_0 \text{ (if PU is idle)} \quad (2a) \\ l(n) + w(n) & H_1 \text{ (if PU is busy)} \quad (2b) \end{cases}$$

where $\tilde{r}(n)$ is the nth sample of the received signal in the HD case.

The performance of any sensing technique is measured by the false-alarm probability (P_f) and the detection probability (P_d). P_f and P_d are defined as the probabilities that the SU declares the sensed channel to be busy given hypothesis H_0 and H_1, respectively. A good system should have high P_d to reduce collisions between SUs and PUs. At the same time, a lower P_f values results in a higher SU throughput due to a reduction in the missed transmission opportunities.

The decision metric, denoted by M, in waveform-based sensing is based on correlating the received samples $(r(n)'s)$ with known pattern samples, and then comparing M against a given threshold γ to determine the state of the sensed channel. Formally,

M is given by:

$$M = \mathrm{Re}\left[\sum_{n=1}^{N} r(n)\, l^*(n)\right] \tag{3}$$

where $l^*(n)$ is the conjugate of l(n). Substituting (1a) and (1b) into (3), we obtain M under hypothesis H_0 and H_1, denoted by M_0 and M_1, respectively. respectively:

$$M_0 = \mathrm{Re}\left[\sum_{n=1}^{N} \left(\chi\, s(n)\, l^*(n) + w(n)\, l^*(n)\right)\right] \tag{4}$$

$$M_1 = \sum_{n=1}^{N} |l(n)|^2 + \mathrm{Re}\left[\sum_{n=1}^{N} \left(\chi\, s(n)\, l^*(n) + w(n)\, l^*(n)\right)\right]. \tag{5}$$

For FD sensing, P_f and P_d are given by:

$$P_f = \Pr\left[M_0 > \gamma\right] = 1 - F_{M_0}(\gamma) \tag{6}$$

$$P_d = \Pr\left[M_1 > \gamma\right] = 1 - F_{M_1}(\gamma) \tag{7}$$

where $F_{M_0}(\gamma)$ and $F_{M_1}(\gamma)$ are the CDFs of the random variables M_0 and M_1, respectively.

***Proposition** 1:* Using the central limit theorem (for a large N), the pdf of M_0 can be approximated by a Gaussian distribution with mean $\mu_{M_0} = 0$ and the following variance:

$$\sigma^2_{M_0} = \frac{N}{2}\left[\chi^2 \mathrm{E}\,|s(n)|^2\, \mathrm{E}\,|l(n)|^2 + \mathrm{E}\,|w(n)|^2\, \mathrm{E}\,|l(n)|^2\right]. \tag{8}$$

Hence, the false-alarm probability can be written as:

$$P_f = Q\left(\frac{\gamma - \mu_{M_0}}{\sigma_{M_0}}\right), \tag{9}$$

where Q is the complementary distribution function of a standard Gaussian random variable. Substituting for μ_{M_0} and $\sigma^2_{M_0}$ in (9), we get the false-alarm probability for FD sensing:

$$P_f = Q\left(\frac{\gamma}{\chi^2\sigma_s^2 + \sigma_w^2}\sqrt{\frac{2}{N\mathit{SNR}^{(FD)}}}\right) \tag{10}$$

where $\mathit{SNR}^{(FD)}$ is the SNR at the secondary receiver in the FD case and is given by:

$$\mathit{SNR}^{(FD)} = \frac{\mathrm{E}\,|l(n)|^2}{\chi^2 \mathrm{E}\,|s(n)|^2 + \mathrm{E}\,|w(n)|^2}. \tag{11}$$

Note that $\mathit{SNR}^{(FD)}$ contains the self-interference term, in addition to noise. Furthermore, the number of samples N can be described as a function of the sensing duration T_{Si}, $i = 1, 2, \ldots, m$ and the sampling rate f_S as follows: $N = T_{Si} f_S$.

***Proposition** 2:* For a large N, the pdf of M_1 can be approximated by a Gaussian distribution with mean $\mu_{M_1} = N\,\mathrm{E}\,|l(n)|^2$ and the following variance

$$\sigma^2_{M_1} = N\Bigg[\mathrm{E}\,|l(n)|^4 - \mathrm{E}^{\,2}\,|l(n)|^2 + \frac{1}{2}\left(\chi^2 \mathrm{E}\,|s(n)|^2\, \mathrm{E}\,|l(n)|^2 + \mathrm{E}\,|w(n)|^2\, \mathrm{E}\,|l(n)|^2\right)\Bigg].$$

See the Appendix for the proof of the previous two propositions.

The detection probability for the waveform-based FD sensing can be written as follows:

$$P_d = Q\left(\frac{\gamma - \mu_{M_1}}{\sigma_{M_1}}\right). \tag{12}$$

Substituting for μ_{M_1} and $\sigma^2_{M_1}$ in (12), we get:

$$P_d = Q\left(\frac{\gamma/\left(\chi^2\sigma_s^2 + \sigma_w^2\right) - N\mathit{SNR}^{(FD)}}{\sqrt{N\left[(\alpha-1)\left(\mathit{SNR}^{(FD)}\right)^2 + \mathit{SNR}^{(FD)}/2\right]}}\right) \tag{13}$$

where α is a parameter of the PU signal that relates to its randomness [23]. As an example $\alpha = 2$ for complex Gaussian signals and can range from 1 to 2 for other signal types. Formally, α is defined as follows:

$$\alpha \stackrel{\text{def}}{=} \mathrm{E}\,|l(n)|^4 / \mathrm{E}^{\,2}\,|l(n)|^2\,. \tag{14}$$

The false-alarm and detection probabilities in (10) and (13) for FD sensing converge to HD sensing at perfect SIS (i.e., $\chi = 0$), as shown in the following equations for a specific sensing duration T_{Si}, $i = 1, 2, \ldots, m$:

$$\tilde{P}_f = Q\left(\frac{\gamma}{\sigma_w^2}\sqrt{\frac{2}{N\mathit{SNR}^{(HD)}}}\right) \tag{15}$$

$$\tilde{P}_d = Q\left(\frac{\gamma/\left(\sigma_w^2\right) - N\mathit{SNR}^{(HD)}}{\sqrt{N\left[(\alpha-1)\left(\mathit{SNR}^{(HD)}\right)^2 + \mathit{SNR}^{(HD)}/2\right]}}\right) \tag{16}$$

where $\tilde{P}_f$ and $\tilde{P}_d$ are the false-alarm and detection probabilities for HD sensing, respectively, and $\mathit{SNR}^{(HD)}$ is the SNR at the secondary receiver in the HD case:

$$\mathit{SNR}^{(HD)} = \frac{\mathrm{E}\,|l(n)|^2}{\mathrm{E}\,|w(n)|^2}. \tag{17}$$

P_f and P_d derived in (10) and (13) for FD sensing are functions of the sensing threshold γ. The optimal sensing threshold γ^* can be determined according to the system requirements on P_f and $(1 - P_d)$. For a target P_f or P_d, γ^* can be calculated by finding the inverse of the Q-functions in (10) and (13), respectively. As an example, for a system with a requirement that P_f and $1 - P_d$ are equal. The optimal sensing threshold γ^* can be determined by equating P_f with $1 - P_d$ in (9) and (12), resulting in:

$$\gamma^* = \frac{\mu_{M_0}\sigma_{M_1} + \mu_{M_1}\sigma_{M_0}}{\sigma_{M_0} + \sigma_{M_1}}. \tag{18}$$

Substituting this γ^* in (9) and (12), and after some mathematical manipulations, we obtain the following for P_f and P_d:

$$P_f = Q\left(\frac{\sqrt{N\,\mathit{SNR}^{(FD)}}}{\sqrt{(\alpha-1)\mathit{SNR}^{(FD)} + 1/2} + \sqrt{1/2}}\right) \tag{19}$$

$$P_d = 1 - Q\left(\frac{\sqrt{N\,\mathit{SNR}^{(FD)}}}{\sqrt{(\alpha-1)\mathit{SNR}^{(FD)} + 1/2} + \sqrt{1/2}}\right). \tag{20}$$

IV. Optimal SU Strategy

In this section, we present an optimal strategy for operating an FD-capable SU link.

A. Problem Formulation

To optimize the selection of the operational mode at an SU, we formulate the problem as a partially observable decision process. Let $S = \{0, 1\}$ be the state space, which defines the actual state (idle or busy) of the channel currently being observed by the SU. The action space at an SU is given by $A = \{\mathit{TR}, \mathit{TS}, \mathit{SO}, \mathit{CS}\}$. While observing the PU channel, the SU has to choose an action from the set A. The outcome/observation space for the SU depends on the action taken. If the SU takes the TR action, it will later observe the outcome $\{D\}$, which means that the SU was able to decode the received message, or the outcome $\{U\}$, which stands for undecoded message. For the transmission part (in the TR and TS modes), the SU may get ACK or a NACK from the peer SU, which are denoted by $\{A\}$ and $\{N\}$, respectively. For a TS action, the SU will also observe two additional outcomes: ($\{F\}$ for free or $\{B\}$ for busy). Finally, the observed outcomes for the SO/CS actions are $\{F\}$ or $\{B\}$. Altogether, these various actions result in an observation space $O = \{D, U, A, N, F, B\}$. Later on, we present a reward function, which maps the state and action space to a reward value.

Our objective is to let the SU choose actions sequentially in time so as to maximize the expected reward over some random finite horizon. It is known that the sufficient statistics for choosing the optimal action at each time t is the belief [25], which is defined as the a posteriori probability $p_t \in [0, 1]$ that the PU is idle at time t given the whole observation history. We consider a similar setup as in [15] for the partially observable decision process part, where the time index t is defined as the time elapsed since the PU has switched from ON to OFF. Hence, $t = 0$ is the start of the PU idle period, which is assumed to be known to the SU, and therefore $p_0 = \tilde{P}_d$. Starting from $t = 0$, the SU keeps tracking of time, and applying the optimal mode selection policy until switching to a new channel (CS action). At this time, the SU resets the algorithm and keeps sensing/switching between

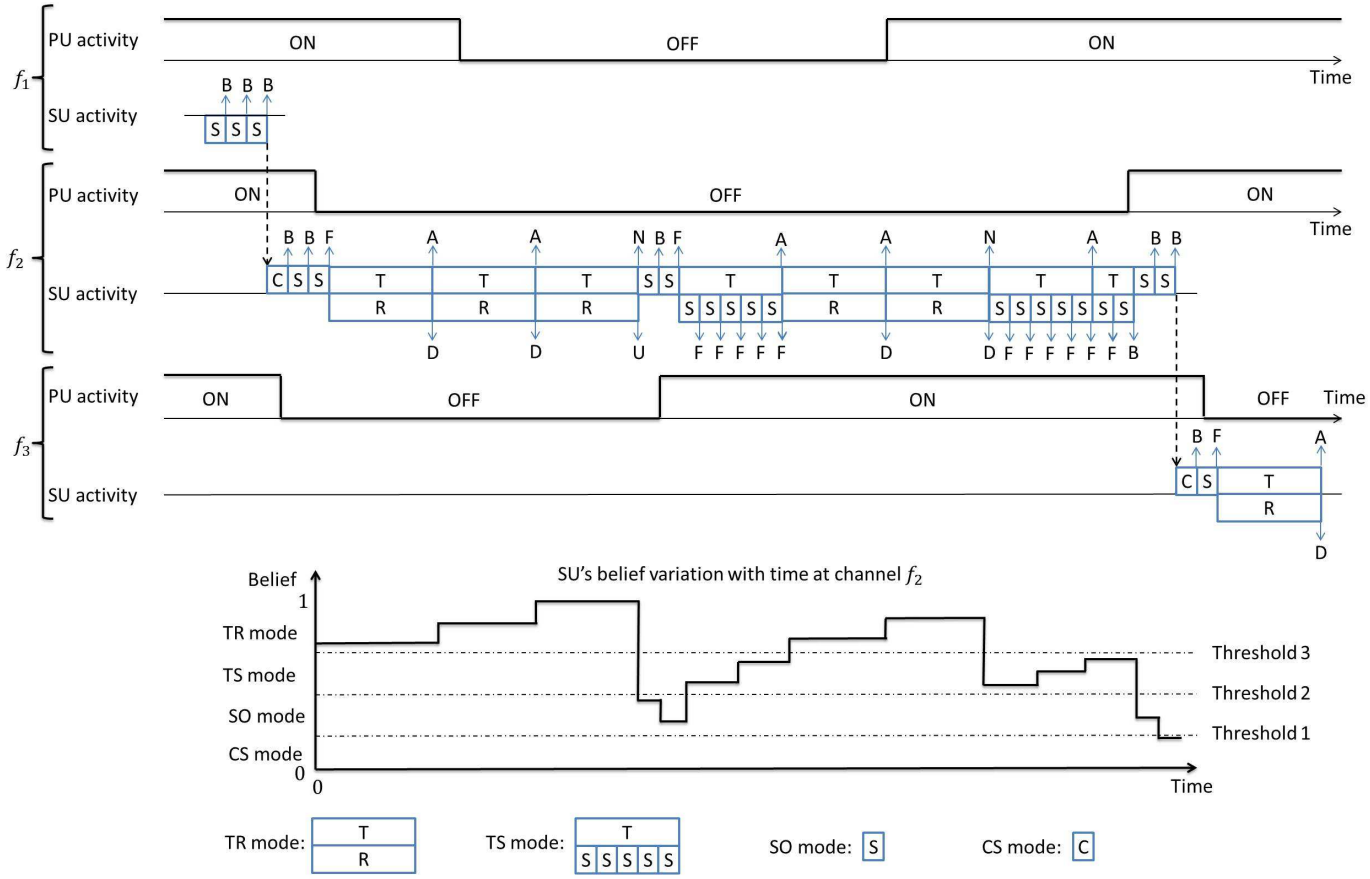


Fig. 3. A time diagram for PUs activities at different frequencies (f_1, f_2, and f_3) and the SU interaction with them using our optimal strategy (top). The SU's belief variation with time is then shown below the time diagram (middle). A description of the shapes used to describe the TR, TS, SO, and CS modes is shown at the bottom. The outcomes notations are: $\{A\}$ for ACK, $\{N\}$ for NACK, $\{D\}$ for decoded, $\{U\}$ for undecoded, $\{F\}$ for free, and $\{B\}$ for busy. Note that this figure aims to deliver the main general idea of our adaptive policy (some simplifications are done to not complicate the figure).

different channels until catching the start of the PU idle period. Once the beginning of an idle period is detected, the SU sets its timer to $t = 0$, and then starts applying the optimal policy. While deriving our optimal policy, we assume that both SUs always have data to transmit. However, at the end of this section, we will discuss the more general case that accounts for the traffic flow between different SUs.

Figure 3 shows a simplified example to illustrate how the SU will adaptively choose its optimal actions and update its belief according to the resulting outcome. In this example, we have three channels (f_1, f_2, and f_3) occupied with PUs and an SU link that is trying to opportunistically communicate on one of these channels. Assume that the SU starts monitoring channel f_1, which is happened to be busy. After getting multiple busy outcomes, it decides to switch to another channel f_2, which is sensed free after multiple busy outcomes. At this point the SU starts its timer (i.e., $t = 0$). In that case, the belief is higher than threshold 3, then the SU starts with the TR mode. The outcomes of the first TR action are ACK and decoded. Hence, as shown in the figure, the SU's belief increases and recommends the TR mode again. However, it happened that the SU receives a NACK and undecoded for the third TR action (may be for a deep fading at this time instance). The SU then updates its belief, which decreases below threshold 2, which implies that the SU should sense the spectrum (i.e., SO mode). The SU keeps sensing until the belief goes over threshold 2. In that case, the SU switches to the TS mode and updates its belief according to the outcomes that it gets. The SU continues switching between different operation modes and updates its belief accordingly as shown in figure 3 until its belief goes below threshold 1. In that case the SU switches to another channel, and so on. Note that the PU has returned to channel f_2, while the SU is operating on the same channel. However, it happened that the SU is in the TS mode, which makes the SU detect the PU activity while transmitting. In that case the SU stops its transmission quickly to prevent collision with the PU (see the figure 3).

Following any given action $a \in A$ and depending on the observation $o \in O$, the SU updates its belief p_t and will also gain a certain reward. Let π be the policy that maps the SU's belief p_t to the action space $a \in A$ at each time t. Define the value function $U(p_t, t)$ as the maximum expected total reward at time t when the current belief is p_t. This function specifies the

performance of the optimal policy π^*, starting from belief p_t. Based on Bellman equation, we have the following:

$$U(p_t,t)=\max\left\{U_{TR}(p_t,t),U_{TS}(p_t,t),U_{SO}(p_t,t),U_{CS}(p_t,t)\right\} \tag{21}$$

where $U_{TR}(p_t,t),U_{TS}(p_t,t),U_{SO}(p_t,t)$, and $U_{CS}(p_t,t)$ are the expected total rewards if the SU decides to operate in the TR, TS, SO, and CS modes, respectively, at time t and then follows the optimal policy π^*

***Lemma** 1:* $U(p_t,t)$ is a convex function of p for a given t.

Proof: We use a similar argument as in [15, 26] to prove this lemma. Let $0\leq\lambda\leq 1$ and $0\leq p_1,p_2\leq 1$. Assume that the initial state p is determined according to the outcome of flipping a biased-coin with a probability λ that a head appears. We set $p=p_1$ if a head appears, and $p=p_2$ if a tail appears. The best reward that we can get if we know the outcome of the coin flipping is $\lambda U(p_1,t)+(1-\lambda)U(p_2,t)$. However, if we do not know the outcome of the coin flipping, the best achieved reward is $U(\lambda p_1+(1-\lambda)p_2,t)$. Since, the best reward with no information will not be higher than that achieved with information available. Therefore, $U(\lambda p_1+(1-\lambda)p_2,t)\leq\lambda U(p_1,t)+(1-\lambda)U(p_2,t)$. ■

B. Reward Function

In this section, we formulate the SU utility for various actions. Define the immediate and expected future reward that the SU gains from taking action i as $R_i^{(M)}$ (M for myopic) and $R_i^{(L)}$ (L for long-term), respectively. The probability that the ith SU observes outcome o is denoted by $w_o^{(i)}$. The updated belief probability for outcome o at node i is denoted by $\mathcal{E}_o^{(i)}$. Define $q_t^{(T)}$ as the probability that the PU will remain idle during the transmission period T, given that the PU is idle at time t. Similarly, define $q_t^{(S)}(i)$ as the probability that the PU will remain idle during $T_{Si},i=1,2,\ldots,m$, given that the PU is idle at the start of this sensing duration. These two quantities can be expressed as follows:

$$q_t^{(T)}=\frac{1-F_X(t+T)}{1-F_X(t)} \tag{22}$$

$$q_t^{(S)}(i)=\frac{1-F_X\left(t+\sum_{j=1}^{i}T_{Sj}\right)}{1-F_X\left(t+\sum_{j=1}^{i-1}T_{Sj}\right)}. \tag{23}$$

where $F_X(t)$ is the CDF of random variable X evaluated at point t.

Next, we derive the reward function for various SU modes.

1) TR mode: The myopic reward for the SU link, consisting of nodes a and b, under the TR mode can be formulated as follows:

$$R_{TR}^{(M)}=\sum_{i\in\{a,b\}}w_D^{(i)}T\log\left(1+SNR_{TR}^{(i)}\right). \tag{24}$$

where $w_D^{(i)}$ is the probability that the ith SU has successfully decoded the received message, and $SNR_{TR}^{(i)}$ is the SNR in the TR mode at node i, which is given by:

$$SNR_{TR}^{(i)}=\frac{P_j\left|h_{ji}\right|^2}{\sigma_i^2+\chi_i^2P_i\left|h_{ii}\right|^2}. \tag{25}$$

In (25), h_{ii} is the gain of the self-interfering channel at node i.

Since the two communicating SUs may experience different channel conditions, the ability to receive data differ from one node to another. Although, we assume that the PU signal affects both SUs equally, the interference level may differ from one node to another because of other interference sources. Hence, a successful decoding process at one node does not imply that the other node will be able to decode its packet. Also, the SU might get an ACK although the PU is ON, due to deep channel fading between the primary transmitter and the secondary receiver. All of these features are captured in the following two probabilities, which may differ from one SU to another [15]:

$\delta_0^{(i)}$: probability that the ith secondary transmitter receives a NACK although the PU is OFF.

$\delta_1^{(i)}$: probability that the ith secondary transmitter receives a NACK given that the PU is ON.

When the ACK/NACK reflects only whether a collision occurs with the PU or not, we have $\delta_0^{(i)}=0$ and $\delta_1^{(i)}=1$.

For the reception part of the TR mode, the probability that the ith SU, $i=a,b$ successfully decode the received message is as follows:

$$w_D^{(i)}=p_tq_t^{(T)}\left(1-\delta_0^{(i)}\right)+\left(1-p_tq_t^{(T)}\right)\left(1-\delta_1^{(i)}\right). \tag{26}$$

Using Bayes' rule, the probability that the PU is idle after T given that the ith secondary transmitter successfully decode the received message (i.e., the belief update) is:

$$\mathcal{E}_D^{(i)}=\left[p_tq_t^{(T)}\left(1-\delta_0^{(i)}\right)\right]/w_D^{(i)}. \tag{27}$$

Similarly, the probability that the ith SU failed to decode the received message and the belief update in that case can be written, respectively, as follows:

$$w_U^{(i)} = p_t q_t^{(T)} \delta_0^{(i)} + \left(1 - p_t q_t^{(T)}\right) \delta_1^{(i)} \tag{28}$$

$$\mathcal{E}_U^{(i)} = \left[p_t q_t^{(T)} \delta_0^{(i)}\right] / w_U^{(i)}. \tag{29}$$

For the transmission part, the probability of receiving an ACK/NACK at node i is the same as the probability that the other node succeed/fail in decoding the message as we assume that transmission errors in ACK/NACK are negligible. This is also applied to the belief update for the corresponding cases. Hence, $w_A^{(i)}, \mathcal{E}_A^{(i)}, w_N^{(i)}$, and $\mathcal{E}_N^{(i)}$ will be formulated similarly as $w_D^{(i)}, \mathcal{E}_D^{(i)}, w_U^{(i)}$, and $\mathcal{E}_U^{(i)}$, respectively.

There are four possible outcomes for the TR mode. An SU may receive an ACK for correct transmission and be able to successfully decode the received message, or the SU might get an ACK and an undecoded message. The other two outcomes of the TR mode is to either get a NACK and a decoded message, or a NACK and an undecoded message. Hence, the expected future reward for an SU link obtained at the ith SU can be formulated as follows:

$$R_{TR}^{(L)} = \sum_{\substack{k=\{A,N\} \\ l=\{D,U\}}} w_k^{(i)} w_l^{(i)} \, U\left(\mathcal{E}_k^{(i)} \mathcal{E}_l^{(i)}, t+T\right) \tag{30}$$

where the summation over the two indices (k, l) can be calculated by considering the four possible combinations $(A, D), (A, U)$, $(N, D), (N, U)$.
Finally,

$$U_{TR}\left(p_t, t\right) = R_{TR}^{(M)} + \eta \, R_{TR}^{(L)} \tag{31}$$

where $\eta \in [0, 1]$ is the discount factor, which determines how far you take the future reward into consideration while formulating the secondary utilities. At $\eta = 0$, the SU only cares for the immediate reward. The final belief p_{t+T} will be the multiplication of the two updates $\mathcal{E}_{O_1}^{(i)} \mathcal{E}_{O_2}^{(i)}$, where $O_1 \in \{A, N\}$, and $O_2 \in \{D, U\}$.

2) TS mode: The myopic reward of the TS mode is different from that of the TR mode because the SU is monitoring the spectrum while transmission. Hence, the SU could abort transmission if a busy outcome is observed after any $T_{Sj}, j = 1, 2, \ldots, m$. Therefore, the myopic reward in the TS mode, assuming that SU i is transmitting to SU j will be formulated as follows:

$$R_{TS}^{(M)} = \prod_{l=1}^{m} w_F^{(i)}(l) \, w_A^{(i)} T \log\left(1 + SNR_{TS}^{(j)}\right) \tag{32}$$

where the SNR in the TS mode at node j is given by:

$$SNR_{TS}^{(j)} = \frac{P_i \left|h_{ij}\right|^2}{\sigma_j^2}. \tag{33}$$

$\prod_{l=1}^{m} w_F^{(i)}(l) \, w_A^{(i)}$ is the probability of successful transmission in the TS mode, which has two conditions. First, SU i gets a free outcome after each and every sensing period of the m sensing durations. This probability is denoted by $\prod_{l=1}^{m} w_F^{(i)}(l)$. Second, SU i receives an ACK from SU j at the end of T, which is denoted by probability $w_A^{(i)}$. Define $\mathbb{P}_f = [P_{f,1} \; P_{f,2} \; \ldots \; P_{f,m}]$ and $\mathbb{P}_d = [P_{d,1} \; P_{d,2} \; \ldots \; P_{d,m}]$ as m-dimensional vectors that represent the false-alarm and detection probabilities, respectively, for the m FD sensing periods in the TS mode.

The probability of getting ACK/NACK from the transmission process and the belief update in the corresponding cases are the same as that of the TR mode. The sensing process has also two outcomes, either free or busy. We assume that if the PU is sensed free/busy at time t at one end of the SU link, then the other SU will experience the same situation. Also, if the PU is sensed busy at any sensing period, this yields a failure communication and the SU should abort the TS mode.

Hence, the probability that the ith SU, $i = a, b$ gets a free outcome after T_{S1} can be expressed as follows.

$$w_F^{(i)}(1) = p_t q_t^{(S)}(1) \left(1 - P_{f,1}\right) + \left(1 - p_t q_t^{(S)}(1)\right) \left(1 - P_{d,1}\right). \tag{34}$$

Similarly, the probability that the ith SU gets a free outcome after $T_{Sj}, j = 2, 3, \ldots, m$ given that it got a free outcome at $T_{S(j-1)}$ is as follows:

$$w_F^{(i)}(j) = q_t^{(S)}(j) \left(1 - P_{f,j}\right) + \left(1 - q_t^{(S)}(j)\right) \left(1 - P_{d,j}\right). \tag{35}$$

The belief update after T_{S1} and $T_{Sj}, j = 2, 3, \ldots, m$ in the case of a free outcome can be written, respectively, as follows:

$$\mathcal{E}_F^{(i)}\left(1\right) = \left[p_t q_t^{(S)}(1) \left(1 - P_{f,1}\right)\right] / w_F^{(i)}(1) \tag{36}$$

$$\mathcal{E}_F^{(i)}(j) = \left[q_t^{(S)}(j)\left(1 - P_{f,j}\right)\right] / w_F^{(i)}(j). \tag{37}$$

Similarly, the probability that the ith SU, $i = a, b$ gets a busy outcome after T_{S1} is:

$$w_B^{(i)}(1) = p_t q_t^{(S)}(1) P_{f,1} + \left(1 - p_t q_t^{(S)}(1)\right) P_{d,1}. \tag{38}$$

Generally, the probability of getting a busy outcome after $T_{Sj}, j = 2, 3, \ldots, m$ given that it got a free outcome at $T_{S(j-1)}$ is as follows:

$$w_B^{(i)}(j) = q_t^{(S)}(j) P_{f,j} + \left(1 - q_t^{(S)}(j)\right) P_{d,j}. \tag{39}$$

The belief update after T_{S1} and $T_{Sj}, j = 2, 3, \ldots, m$ in the case of a busy outcome can be written, respectively, as follows:

$$\mathcal{E}_B^{(i)}(1) = \left[p_t q_t^{(S)}(1) P_{f,1}\right] / w_B^{(i)}(1) \tag{40}$$

$$\mathcal{E}_B^{(i)}(j) = \left[q_t^{(S)}(j) P_{f,j}\right] / w_B^{(i)}(j). \tag{41}$$

TS mode is different from other modes as the SU may not continue till the end of the transmission duration. This happens if the SU gets a busy outcome at the end of any sensing duration $T_{Si}, i = 1, 2 \ldots, m$. However, if the SU gets a free outcome after every sensing period $T_{Si}, i = 1, 2 \ldots, m$, there are two possible outcomes. The SU might get a Free and ACK for correct reception or Free and NACK for incorrect reception. Putting all together, the expected future reward in the TS mode can be formulated as follows:

$$\begin{aligned} R_{TS}^{(L)} = & \sum_{j=1}^{m} w_B^{(i)}(j) \prod_{l=1}^{j-1} w_F^{(i)}(l)\, U\left(\mathcal{E}_B^{(i)}(j) \prod_{l=1}^{j-1} \mathcal{E}_F^{(i)}(l), t + \sum_{l=1}^{j} T_{Sl}\right) \\ & + \sum_{k=\{A,N\}} w_k^{(i)} \prod_{l=1}^{m} w_F^{(i)}(l)\, U\left(\mathcal{E}_k^{(i)} \prod_{l=1}^{m} \mathcal{E}_F^{(i)}(l), t + T\right) \end{aligned} \tag{42}$$

Finally,

$$U_{TS}(p_t, t) = R_{TS}^{(M)} + \eta\, R_{TS}^{(L)}. \tag{43}$$

3) SO mode: The immediate reward $R_{SO}^{(M)}$ in the SO mode will be zero as no transmission takes place. The outcome of the sensing process is either free or busy. The probability of getting a free/busy outcome and the belief update in each case can be formulated similarly as the sensing part of the TS mode taking into consideration that the SO mode consists of a single sensing period T_S. Hence, the expected future reward in the SO mode can be expressed as follows:

$$R_{SO}^{(L)} = \sum_{k=\{F,B\}} w_k^{(i)} U\left(\mathcal{E}_k^{(i)}, t + T_S\right) \tag{44}$$

where $w_F^{(i)}$ and $w_B^{(i)}$ are the probabilities of getting a free and busy outcomes, respectively, after T_S. $\mathcal{E}_F^{(i)}$ and $\mathcal{E}_B^{(i)}$ are the corresponding belief updates. Hence we can write the maximum expected utility that the SU gains from sensing the spectrum as:

$$U_{SO}(p_t, t) = \eta\, R_{SO}^{(L)}. \tag{45}$$

4) CS mode: The SU might choose to switch to another frequency channel (where no information about the PU state is available) and carry out spectrum sensing, if the probability that the PU returns is very high. The analysis for this operation mode is the same as that of SO mode, except for the belief p_t because the belief in the new channel $\hat{p}_t$ will be the probability that the PU is idle at time t given that no previous information is available, which can be written generally as follows:

$$\hat{p}_t = \bar{X} / \left(\bar{X} + \bar{Y}\right). \tag{46}$$

The maximum expected utility for the CS mode is as follows:

$$U_{CS}(\hat{p}_t, t) = \eta\, R_{CS}^{(L)} \tag{47}$$

where

$$R_{CS}^{(L)} = \sum_{k=\{F,B\}} \hat{w}_k^{(i)} U\left(\hat{\mathcal{E}}_k^{(i)}, t + T_S\right) \tag{48}$$

where $\hat{w}_F^{(i)}, \hat{w}_B^{(i)}, \hat{\mathcal{E}}_F^{(i)}$ and $\hat{\mathcal{E}}_B^{(i)}$ are formulated similarly as $w_F^{(i)}, w_B^{(i)}, \mathcal{E}_F^{(i)}$ and $\mathcal{E}_B^{(i)}$, respectively, after replacing p_t by $\hat{p}_t$.

In the Appendix, we discuss the convexity and other properties of the SU's utilities $U_{TR}(p_t, t), U_{TS}(p_t, t), U_{SO}(p_t, t)$, and $U_{CS}(\hat{p}_t, t)$ with respect to the belief. We also discuss how $U(p_t, t)$ varies with p for a given t.

C. Optimal Policy

After formulating the SU utilities in the four possible actions and adding a constraint on the collision probability with the PU, our problem can be formulated as follows:

$$\begin{aligned} \underset{\pi}{\text{maximize}} \quad & U(p_t, t) \\ \text{subject to} \quad & P_i \leq P_i^* \qquad i \in \{TR, TS\} \end{aligned} \tag{49}$$

where $P_i, i \in \{TR, TS\}$ is the PU collision probability in the TR and TS modes, respectively and P_i^* is the threshold PU collision probability. P_{TR} and P_{TS} can formulated as follows:

$$P_{TR} = (1 - p_t) + p_t \left(1 - q_t^{(T)}\right) = 1 - p_t q_t^{(T)} \tag{50}$$

$$P_{TS} = p_t \sum_{i=1}^{m} \left\{ \prod_{j=1}^{i-1} (1 - P_{f,j}) \left(1 - q_t^{(S)}(i)\right) \right\} + (1 - p_t). \tag{51}$$

For certain probability distributions for the PU idle period as Gaussian distribution, uniform distribution, and Rayleigh distribution (as an example), $q_t^{(T)}$ will approach zero at large values of t. This means that the probability that the PU returns to utilize the channel increases with t, which is very intuitive for this type of distributions. To be able to derive our optimal threshold-based policy, we define a technical condition similar to the approach followed in [15]. This technical condition states that for all $t > t^*$, the SU should not transmit any data (i.e. should not operate in either TR, or TS modes) as the collision probability constraint will not be satisfied and hence, zero reward will be gained even at $p_t = 1$. This threshold time t^* is defined as the minimum time where the PU collision constraint is not satisfied. Hence, $U(1, t) = 0, \forall t > t^*$.

$$t^* = \min \left\{ t : q_t^{(T)} < 1 - P_{TR}^* \right\}. \tag{52}$$

***Theorem** 1:* The optimal policy for the SU can be written as follows:

$$\pi^*(p_t) = \begin{cases} CS, & p_t < \beta_c \\ SO, & \beta_c \leq p_t < \beta_s \\ TS, & \beta_s < p_t < \beta_t \\ TR, & p_t \geq \beta_t \end{cases} \tag{53}$$

Proof: Since $U_{CS}(\hat{p}, t)$ is constant with p_t, and $U_{SO}(p_t, t)$ is a convex increasing function of p_t (lemma 6 in the Appendix). Therefore, There exist at most one intersection between the two functions because $U_{CS}(\hat{p}, t) > U_{SO}(0, t)$. This intersection occurs when $p_t = \hat{p}$, which is denoted by β_c. This intersection exists when $\hat{p} < \beta_s$ (i.e., under two conditions: strict PU collision constraint and highly loaded PU networks). Although the first condition is guaranteed, the second one is not. In that case the SO region might disappear, and hence $\beta_c = \beta_s$. Since $U_{CS}(\hat{p}, t) > U_{SO}(p_t, t)$ for $p_t < \beta_c$ and $U_{CS}(\hat{p}, t) < U_{SO}(p_t, t)$ for $p_t > \beta_c$, the first two lines of the policy shown in (53) are optimal given that the aforementioned condition is satisfied.

$U_{TR}(p_t, t)$ and $U_{TS}(p_t, t)$ are proved to be convex increasing functions of p_t at a given t (Lemmas (2), (3), (4), and (5) in the Appendix). Since $U_{TR}(1, t) > U_{TS}(1, t)$ and since $\beta_t \geq \beta_s$ (Generally, the amount of interference induced by the SU on the PU in the TR mode is higher than that of the TS mode), therefore both functions intersects together at β_t. Since $U_{TS}(p_t, t) > U_{TR}(p_t, t)$ for $p_t < \beta_t$ (as $U_{TR}(p_t, t)$ goes to zero for violating the PU collision constraint) and $U_{TS}(p_t, t) < U_{TR}(p_t, t)$ for $p_t > \beta_t$, the last two lines of the policy shown in (53) are optimal. Note that $U_{TS}(p_t, t)$ goes to zero $\forall p_t < \beta_s$ due to the violation of the PU collision constraint. ■

The above theorem states that the SU should utilize the opportunity of having a high belief that the PU is idle and operate in the TR mode if $p_t \geq \beta_t$, where β_t is the transmission-reception threshold. In that case the SU will dramatically increase the throughput by transmitting and receiving data simultaneously over the same channel. If the belief decreases and falls in the following range $\beta_s < p_t < \beta_t$, the SU should monitor the spectrum while transmitting (i.e., operate in the TS mode) as the probability that the SU returns is now relatively high. β_s is called the sensing threshold. In that case, the SU still getting some throughput (lower than TR mode), however a lower collision probability is achieved. The SU should stop transmitting, and carry out HD sensing (i.e., SO mode) if p_t is relatively low $\beta_c \leq p_t < \beta_s$ because in that case the probability that the PU returns to the channel is too high and the PU collision constraint will not be satisfied. Hence, a better sensing quality and a temporarily channel vacation is required. β_c is called the channel switching threshold. At very low belief values $p_t < \beta_c$, where the PU is most likely to return to use the channel, the SU should take the CS action. This happens when the probability that the PU is idle in a new channel (where no information is available) is higher than the current belief.

To solve our problem, we have to find the threshold time t^*, where our condition is satisfied and then apply backward induction to find the thresholds $\beta_c, \beta_s, \beta_t$ and the maximum utility for the SU $U(p, t)$ for different values of p and t.

D. Discussion of the optimal policy

In this section, we will highlight some important features of our optimal policy that should be considered in our FD DSA network. Until this point of discussion, we assumed that SUs always have data to transmit. That is, if the optimal policy recommends the TR mode, then SUs will carry out simultaneous transmission and reception over the same frequency. But, what if one of the nodes does not have data to its peer? This motivates the refinement of our adaptive strategy to account for SU's traffic flow. Before doing that, let us first discuss how we define master/slave nodes, and data/control phases in the following two items:

- Master and slave nodes: A master node is a designation given to any SU device that executes the optimal adaptive decision strategy. The master node is the one that takes a final decision about the operation mode, whether receiving data while transmitting, sensing the spectrum while transmitting, switching to another channel, or solely sensing the spectrum. The slave node is the SU device that is receiving orders from the master node with regard to starting, continuing, or stopping transmission. This is done through control packets. In the present design, the node that initiates the communication is the master node, and the other one is the slave node. However, these roles may change over time, depending on which node has traffic to send (i.e., the traffic directionality).
- Data and Control Phases: This concept is related to the organizing protocol used between SUs, equipped with SIS/FD capabilities, to communicate with each other. Once a channel is thought to be idle by the initiating SU, the time frame will be divided into two alternating phases: data phase and control phase. The SUs will start communicating together in the data phase where data packets only are exchanged. Also, spectrum sensing is executed in the data phases as well. After that, a control phase is established between the two nodes, which has two goals. First, this control phase is used to confirm the correct reception of packets transmitted in the previous data phase in both directions, if applicable. Second, the control phase is used by the master node to trigger the slave node to start, continue, or stop transmission while receiving. If the two nodes switch to another channel, the data and control phases will be terminated (until finding a free channel) and the sensing durations can be optimized separately, without the data and control phases restrictions.

The four possible modes (TR, TS, SO, and CS) represent a single SUs perspective. However, the links operation is determined by the mode of operation at the two communicating nodes. Hence, these modes of operation should be written as follows: TR-TR, TS-R (or R-TS), SO, and CS. For instance, if an SU is transmitting and receiving data simultaneously, the other secondary node will also be transmitting and receiving data, which is indicated by the TR-TR mode. However, if an SU is transmitting and sensing, its peer will be only in the reception mode, which defines the TS-R mode ($\{\mathrm{R}\}$ stands for reception). The threshold-based structure discussed so far has to be adjusted with the SUs traffic load. For instance, assume that the optimal strategy recommends the SU to operate in the transmission-reception mode, while the other communicating node does not have data to transmit. In this case, the SU node can operate in the transmit-sense mode while the other SU can only receive data. To implement this, the two communicating SUs can utilize the more packets (MP) bit in the header of each packet to find the final decision. This MP bit determines whether a certain node has more packets in its queue or not.

Another crucial point, in FD DSA networks, to be considered in our design is the FCC requirements. Consider the scenario where SUs are operating in the idle PU period using the TR mode. In the case of good channel conditions, both nodes will keep ACKing their packets, updating their beliefs, and will not switch to any other mode unless they collide with the PU (in that case they will get NACKs), or when $t > t^*$. To avoid this blind communication without monitoring the PU channel, SUs should periodically switch to any of the sensing modes (TS or SO), if they violate the FCC requirements discussed next. The FCC imposes rules for operating opportunistic wireless networks. One of these rules is the periodic sensing interval, which means that any channel used by an SU has to be sensed every T_{req} seconds to check for the PU activity. The SU has to vacate the channel quickly if a PU activity is detected. Hence, this part of the decision strategy states that each SU link has to maintain a maximum duration of T_{req} seconds between sensing periods, whether it was operating in the SO or TS modes.

V. Numerical Results

We use the following parameters unless otherwise is mentioned. The sampling frequency and the SU signal power are $f_S = 6MHz$ and $\sigma_s^2 = 5$, respectively and $SNR^{(HD)} = -20dB$. For evaluating P_f and P_d, we consider a complex Gaussian primary signal with $\alpha = 2$. The PU idle period is uniformly distributed in the range $[0, 1000]$. We also set $T_S = 1$, $m = 30$, $T = 30$, $\delta_0 = 0.01$, and $\delta_1 = 0.99$.

A. Performance Metrics

1) False-Alarm And Detection Probabilities: The impact of the residual self-interference signal on P_f and P_d for waveform-based sensing is shown in Figures 4 and 5, respectively. As χ increases the performance of the waveform-based sensing get worse (i.e., P_f increases and P_d decreases) due to the increment in the residual self-interference. We also notice that P_f and P_d converges to HD sensing at perfect SIS. In imperfect sensing schemes, increasing the sensing duration improves the performance of the sensing technique. At low SNR regions, the SU needs around 20% increment in the sensing duration to achieve the same P_f and P_d (achieved for HD sensing) for 20% residual self-interference from the original SU signal and needs about 80% increment in T_S for 40% increase in the residual self-interference.

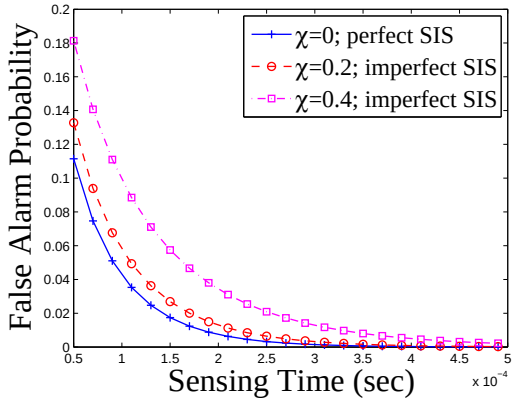


Fig. 4. False-alarm probability vs. sensing time for FD sensing at different values of χ.

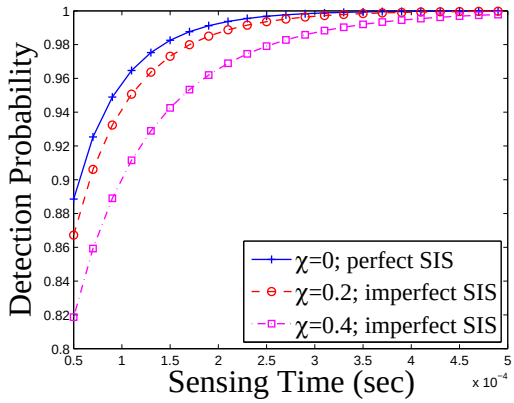


Fig. 5. Detection probability vs. sensing time for FD sensing at different values of χ.

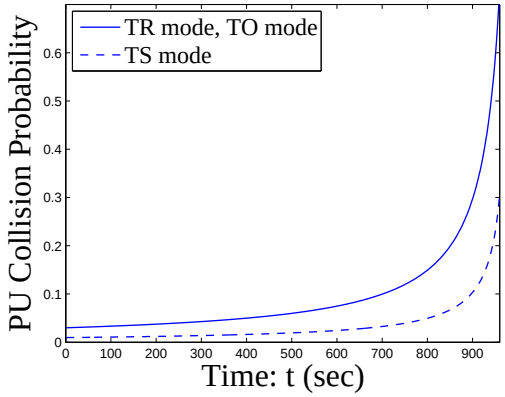


Fig. 6. PU collision probability vs. t for TR, TS, and TO modes at $p = 1$ and $P_f = 0.1$.

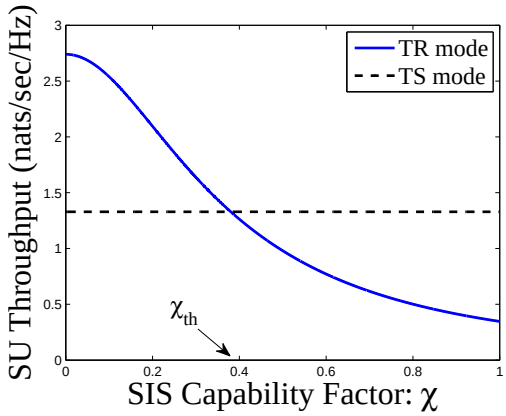


Fig. 7. SU's throughput vs. χ for TR and TS modes at $p = 1$, $P_f = 0$, $P_d = 1$, $\sigma_s^2 = 15$ and $SNR^{(HD)} = 5dB$.

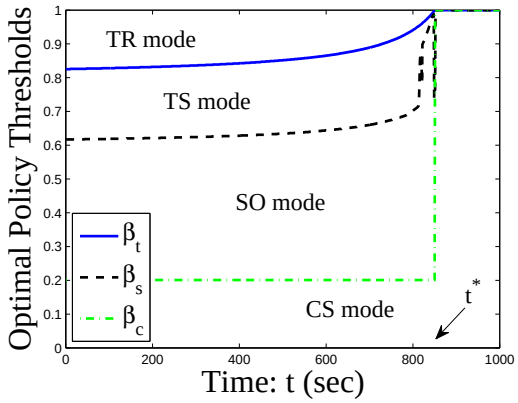


Fig. 8. Optimal policy thresholds vs. t. The decision region is divided into four parts: TR, TS, SO , and CS.

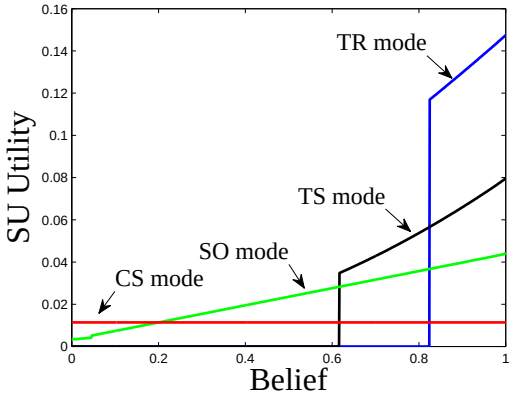


Fig. 9. SU utilities in the TR, TS, SO, and CS modes vs. the belief at $t = 0$. The final SU utility is the maximum of these four utilities.

2) PU Collision Probability: The advantage of the TS mode over the TR mode is the lower collision probability. Figure 6 shows the variation of the PU collision probability for the TR, TS, and TO mode with t. As t increases, the collision probability increase as the PU is more likely to return to utilize the channel. As shown in Figure 6, the SU can achieve a lower collision probability in the TS mode than that of the TR mode. This improvement may reach in certain scenarios up to 50% reduction in the collision probability.

3) SU Throughput: Figure 7 shows a comparison between the achievable SU's throughput in the TS mode (assuming long enough T_S) and the TR mode at different values of χ. The SU's throughput in the TR mode decreases as χ increases due to residual self-interference. Hence, working in a FD fashion is not always optimal especially at high χ values. According to our simulation setup, the threshold SIS factor where the SU should switch to the TS mode is $\chi_{th} = 0.38$. Also, as t increases the SU's throughput decreases due to the increment in the probability that PU returns.

B. Transmission-Sensing-Reception Strategy

We use backward induction to find the optimal thresholds and the maximum SU utility. We set $\eta = 0.3, P^*_{TR} = 0.2, P^*_{TS} = 0.4, P_f = 0.01,\ P_d = 0.99, SNR^{(HD)} = 20dB,$ and $\bar{Y} = 2000$.

The variation of β_t, β_s and β_c with t is shown in Figure 8, which shows the mechanism of our optimal policy. The SU should operate in TR mode as long as $p_t \geq \beta_t$, switch to TS mode when $\beta_s < p_t < \beta_t$, switch to SO mode when $\beta_c \leq p_t < \beta_s$, and finally switch to a new channel when $p_t < \beta_c$. The SU should also switch to a new channel if $t > t^*$ as the PU is more likely to return to utilize the channel, which justifies the convergence of β_t, β_s and β_c to 1 for $t > t^*$. Note that β_c is constant (for $t < t^*$) because it depends on the channel availability, when no information is known. Hence, $\beta_c = 500/2500 = 0.2$ according to our setup.

Figure 9 shows the variation of the maximum SU utilities for the TR, TS, SO, and CS modes with p_t. The final SU utility is the maximum of these four utilities. Note that the utility in the CS mode is constant with p because it is independent of the SU belief in the currently used channel. The abrupt reduction for the SU utilities in the TR and TS modes is due to the violation of the PU collision probability constraints.

VI. Conclusions

We consider a novel application of FD communications in DSA networks, where we analyzed the performance metrics of the TR and TS modes, namely the throughput and collision probability. We determined the optimal switching policy of SUs, equipped with SIS/FD capabilities, that maximizes the SU's utility subject to a constraint on the PU collision probability. To enable the TS mode, we analyzed the waveform-based sensing in the case of imperfect SIS, and derived the false-alarm and detection probabilities. Using our adaptive strategy, the SU can achieve about 50% reduction in the collision probability and double the throughput comparing to the HD case. Finally, an optimal threshold-based strategy is obtained, which depends on the SU's belief regarding the idleness of the PU. Our results indicate that SU should operate in the TR mode if it has a high belief that the PU is idle. As this belief decreases, the SU should adaptively switch to the TS mode to monitor any change in the PU activity while transmitting. At very low belief values, where the PU is more likely to be active, the SU should switch to another channel. One possible direction of future work is to address how SUs will negotiate together in the control phase to determine the final action given that both nodes may have different traffic flows.

References

[1] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. of the ACM Mobicom'10 Conf.*, Sep. 2010, pp. 1–12.

[2] M. Jain, J. I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. of the ACM Mobicom'11 Conf.*, Sep. 2011, pp. 301–312.

[3] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. of the ASILOMAR'10 Conf.*, Nov. 2010, pp. 1558–1562.

[4] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in *Proc. of the ASILOMAR'11 Conf.*, 2011, pp. 2002–2006.

[5] B. Radunovic, D. Gunawardena, P. Key, A. Proutiere, N. Singh, V. Balan, and G. Dejean, "Rethinking indoor wireless mesh design: Low power, low frequency, full-duplex," in *Proc. of the Fifth IEEE Workshop on Wireless Mesh Networks*, 2010.

[6] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proc. of the ACM SIGCOMM'13 Conf.*, Aug. 2013.

[7] S. Barghi, A. Khojastepour, K. Sundaresan, and S. Rangarajan, "Characterizing the throughput gain of single cell MIMO wireless systems with full duplex radios," in *Proc. of the WiOpt'12 Conf.*, 2012, pp. 68–74.

[8] D. Ng, E. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1291–1304, 2012.

[9] B. Day, A. Margetts, D. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3702–3713, 2012.

[10] E. Aryafar, M. A. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, "MIDU: Enabling MIMO full duplex," in *Proc. of the ACM Mobicom'12 Conf.*, 2012, pp. 257–268.

[11] W. Afifi and M. Krunz, "Exploiting self-interference suppression for improved spectrum awareness/efficiency in cognitive radio systems." in *Proc. of the IEEE INFOCOM'13 Conf.*, 2013, pp. 1258–1266.

[12] W. Cheng, X. Zhang, and H. Zhang, "Full duplex spectrum sensing in non-time-slotted cognitive radio networks," in *Proc. of the MILCOM'11 Conf.*, Nov. 2011.

[13] ——, "Optimal dynamic power control for full-duplex bidirectional-channel based wireless networks," in *Proc. of the IEEE INFOCOM'13 Conf.*, Turin, Italy, Apr. 2013.
[14] Q. Zhao and J. Ye, "Quickest detection in multiple on-off processes," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 5994–6006, Dec. 2010.
[15] S. Huang, X. Liu, and Z. Ding, "Optimal sensing-transmission structure for dynamic spectrum access," in *Proc. of the IEEE INFOCOM'09 Conf.*, April 2009, pp. 2295–2303.
[16] ——, "Short paper: On optimal sensing and transmission strategies for dynamic spectrum access," in *Proc. of the IEEE DySPAN'08 Conf.*, Oct. 2008, pp. 1–5.
[17] W. Afifi, A. Sultan, and M. Nafie, "Adaptive sensing and transmission durations for cognitive radios," in *Proc. of the IEEE DySPAN'11 Conf.*, 2011, pp. 380–388.
[18] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary users in cellular networks: A large-scale measurement study," in *Proc. of the IEEE DySPAN'08 Conf.*, Oct. 2008, pp. 1 –11.
[19] H. Kim and K. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 533–545, May 2008.
[20] Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu, "Allocating dynamic time-spectrum blocks in cognitive radio networks," in *Proc. of the ACM MobiHoc'07 Conf.*, Sep. 2007, pp. 130–139.
[21] T. Shu and M. Krunz, "Throughput-efficient sequential channel sensing and probing in cognitive radio networks under sensing errors," in *Proc. of the ACM Mobicom'09 Conf.*, Sep. 2009, pp. 37–48.
[22] H. Jiang, L. Lai, R. Fan, and H. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 297–307, 2009.
[23] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in *Proc. of the IEEE DySPAN'05 Conf.*, 2005, pp. 151–159.
[24] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.
[25] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable markov processes over a finite horizon," *Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973.
[26] S. Ross, *Introduction to stochastic dynamic programming.* Orlando, FL, USA: Academic Press, 1983, ch.3, pp. 58-59.

APPENDIX

A. Proofs for Waveform-based Sensing

Proof of Proposition 1: The mean of M_0 can be expressed as follows:

$$\mu_{M_0} = \mathrm{Re}\left[\sum_{n=1}^{N} \mathrm{E}\left(\chi\, s(n)\, l^*(n) + w(n)\, l^*(n)\right)\right] = 0. \tag{54}$$

Since $s(n), l(n)$, and $w(n)$ are independent, the above result holds. The SU signal (and similarly for other signals) can be written as a function of the real and imaginary components as follows: $s(n) = s_r(n) + j s_i(n)$. Hence, the variance of M_0 is:

$$\begin{aligned}
\sigma^2_{M_0} &= \sum_{n=1}^{N} \mathrm{Var}\left(\mathrm{Re}\left[\left(\chi\, s(n)\, l^*(n) + w(n)\, l^*(n)\right)\right]\right) \\
&= N\left[\chi^2\left\{\mathrm{Var}\left(s_r(n) l_r(n)\right) + \mathrm{Var}\left(s_i(n) l_i(n)\right)\right\} + \mathrm{Var}\left(w_r(n) l_r(n)\right) + \mathrm{Var}\left(w_i(n) l_i(n)\right)\right] \\
&= N\left[\chi^2\left\{\mathrm{E}\left(s_r^2(n)\right)\mathrm{E}\left(l_r^2(n)\right) + \mathrm{E}\left(s_i^2(n)\right)\mathrm{E}\left(l_i^2(n)\right)\right\} + \mathrm{E}\left(w_r^2(n)\right)\mathrm{E}\left(l_r^2(n)\right) + \mathrm{E}\left(w_i^2(n)\right)\mathrm{E}\left(l_i^2(n)\right)\right] \\
&= \frac{N}{2}\left[\chi^2 \mathrm{E}\left|s(n)\right|^2 \mathrm{E}\left|l(n)\right|^2 + \mathrm{E}\left|w(n)\right|^2 \mathrm{E}\left|l(n)\right|^2\right].
\end{aligned}$$

■

Proof of Proposition 2: Due to independence, the mean of M_1 is expressed as follows:

$$\mu_{M_1} = \sum_{n=1}^{N} \mathrm{E}\left|l(n)\right|^2 + \mathrm{Re}\left[\sum_{n=1}^{N} \mathrm{E}\left(\chi\, s(n)\, l^*(n) + w(n)\, l^*(n)\right)\right] = N\,\mathrm{E}\left|l(n)\right|^2 \tag{55}$$

The variance of M_1 can be shown to be:

$$\begin{aligned}
\sigma^2_{M_1} &= \sum_{n=1}^{N}\left[\mathrm{Var}\left(\left|l(n)\right|^2\right) + \mathrm{Var}\left(\mathrm{Re}\left[\left(\chi\, s(n)\, l^*(n) + w(n)\, l^*(n)\right)\right]\right)\right] \\
&= N\left[\mathrm{Var}\left(\left|l(n)\right|^2\right) + \chi^2\left\{\mathrm{Var}\left(s_r(n) l_r(n)\right) + \mathrm{Var}\left(s_i(n) l_i(n)\right)\right\} + \mathrm{Var}\left(w_r(n) l_r(n)\right) + \mathrm{Var}\left(w_i(n) l_i(n)\right)\right] \\
&= N\Big[\mathrm{E}\left|l(n)\right|^4 - \mathrm{E}^2\left|l(n)\right|^2 \\
&\quad + \chi^2\left\{\mathrm{E}\left(s_r^2(n)\right)\mathrm{E}\left(l_r^2(n)\right) + \mathrm{E}\left(s_i^2(n)\right)\mathrm{E}\left(l_i^2(n)\right)\right\} + \mathrm{E}\left(w_r^2(n)\right)\mathrm{E}\left(l_r^2(n)\right) + \mathrm{E}\left(w_i^2(n)\right)\mathrm{E}\left(l_i^2(n)\right)\Big] \\
&= N\left[\mathrm{E}\left|l(n)\right|^4 - \mathrm{E}^2\left|l(n)\right|^2 + \frac{1}{2}\left(\chi^2 \mathrm{E}\left|s(n)\right|^2 \mathrm{E}\left|l(n)\right|^2 + \mathrm{E}\left|w(n)\right|^2 \mathrm{E}\left|l(n)\right|^2\right)\right].
\end{aligned}$$

■

B. Proofs for SU utilities

Lemma 2: $U_{TR}(p_t, t)$ increases in p for a given t.

Proof: To prove this lemma we have to show that the first order derivative of $U_{TR}\left(p_t,t\right)$ with regard to p is non-negative. We will start first with the myopic reward and then the long term reward of (31).

$$R_{TR}^{M'} = \sum_{i\in\{a,b\}} q_t^T \left(\delta_1^{(i)} - \delta_0^{(i)}\right) T \log\left(1+SNR_{TR}^{(i)}\right) \tag{56}$$

Which is non-negative as $\delta_1^{(i)} > \delta_0^{(i)}$. The long term reward consists of four terms as shown in (30). For the ACK/Decoded term, we have the following:

$$\begin{aligned} R_{TR}^{L'} = &\left(w_A^{(i)}w_D^{(i)}\right)' U\left(\mathcal{E}_A^{(i)}\mathcal{E}_D^{(i)},t+T\right) + w_A^{(i)}w_D^{(i)}U'\left(\mathcal{E}_A^{(i)}\mathcal{E}_D^{(i)},t+T\right)\left(\mathcal{E}_A^{(i)}\mathcal{E}_D^{(i)}\right)' \\ &+ \left(w_A^{(i)}w_U^{(i)}\right)' U\left(\mathcal{E}_A^{(i)}\mathcal{E}_U^{(i)},t+T\right) + w_A^{(i)}w_U^{(i)}U'\left(\mathcal{E}_A^{(i)}\mathcal{E}_U^{(i)},t+T\right)\left(\mathcal{E}_A^{(i)}\mathcal{E}_U^{(i)}\right)' \\ &+ \left(w_N^{(i)}w_D^{(i)}\right)' U\left(\mathcal{E}_N^{(i)}\mathcal{E}_D^{(i)},t+T\right) + w_N^{(i)}w_D^{(i)}U'\left(\mathcal{E}_N^{(i)}\mathcal{E}_D^{(i)},t+T\right)\left(\mathcal{E}_N^{(i)}\mathcal{E}_D^{(i)}\right)' \\ &+ \left(w_N^{(i)}w_U^{(i)}\right)' U\left(\mathcal{E}_N^{(i)}\mathcal{E}_U^{(i)},t+T\right) + w_N^{(i)}w_U^{(i)}U'\left(\mathcal{E}_N^{(i)}\mathcal{E}_U^{(i)},t+T\right)\left(\mathcal{E}_N^{(i)}\mathcal{E}_U^{(i)}\right)' \end{aligned} \tag{57}$$

$$\begin{aligned} R_{TR}^{L'} = &\left[q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)w_D^{(a)} + q_t^T\left(\delta_1^{(a)}-\delta_0^{(a)}\right)w_D^{(b)}\right] U\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right) + w_D^{(b)}w_D^{(a)}U'\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right)\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)}\right)' \\ &+ \left[q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)w_U^{(a)} - q_t^T\left(\delta_1^{(a)}-\delta_0^{(a)}\right)w_D^{(b)}\right] U\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right) + w_D^{(b)}w_U^{(a)}U'\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right)\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)}\right)' \\ &+ \left[-q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)w_D^{(a)} + q_t^T\left(\delta_1^{(a)}-\delta_0^{(a)}\right)w_U^{(b)}\right] U\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right) + w_U^{(b)}w_D^{(a)}U'\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right)\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)}\right)' \\ &+ \left[-q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)w_U^{(a)} - q_t^T\left(\delta_1^{(a)}-\delta_0^{(a)}\right)w_U^{(b)}\right] U\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right) + w_U^{(b)}w_U^{(a)}U'\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right)\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)}\right)' \end{aligned} \tag{58}$$

$$\begin{aligned} R_{TR}^{L'} = &q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)\left[w_D^{(a)}\left(U\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right) - U\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right)\right) + w_U^{(a)}\left(U\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right) - U\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right)\right)\right] \\ &+ q_t^T\left(\delta_1^{(a)}-\delta_0^{(a)}\right)\left[w_D^{(b)}\left(U\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right) - U\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right)\right) + w_U^{(b)}\left(U\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right) - U\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right)\right)\right] \\ &+ w_D^{(b)}w_D^{(a)}U'\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right)\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)}\right)' + w_D^{(b)}w_U^{(a)}U'\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right)\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)}\right)' \\ &+ w_U^{(b)}w_D^{(a)}U'\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right)\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)}\right)' + w_U^{(b)}w_U^{(a)}U'\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right)\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)}\right)' \end{aligned} \tag{59}$$

Assume, for the time being, that $U\left(p_t,t\right)$ is an increasing function of p (we will justify this assumption while proving lemma (7) by backward induction). Since $\mathcal{E}_D^{(i)} \geq \mathcal{E}_U^{(i)}$ for $i=a,b$. Therefore, the following inequalities hold:

$$\begin{aligned} U\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right) &> U\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right) \\ U\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right) &> U\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right) \\ U\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)},t+T\right) &> U\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)},t+T\right) \\ U\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)},t+T\right) &> U\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)},t+T\right) \end{aligned}$$

Therefore, the first four terms of (59) are non-negative. It can also be proved that $\left(\mathcal{E}_D^{(b)}\mathcal{E}_D^{(a)}\right)'$, $\left(\mathcal{E}_D^{(b)}\mathcal{E}_U^{(a)}\right)'$, $\left(\mathcal{E}_U^{(b)}\mathcal{E}_D^{(a)}\right)'$, and $\left(\mathcal{E}_U^{(b)}\mathcal{E}_U^{(a)}\right)'$ are non-negative. Hence, the last four terms of (59) are also non-negative, which completes the proof. ■

***Lemma** 3: $U_{TR}\left(p_t,t\right)$ is a convex function of p for a given t.*

Proof: To simplify the analysis of this lemma, we consider the case where $\delta_0^{(a)} = \delta_0^{(b)}$ and $\delta_1^{(a)} = \delta_1^{(b)}$. From (56), the second order derivative of the myopic reward $R_{TR}^{M''} = 0$. Next, we find the second order derivative of the long term reward.

$$R_{TR}^{L'} = q_t^T\left(\delta_1^{(a)} - \delta_0^{(a)}\right)\left[U\left(\mathcal{E}_D^{(a)}, t+T\right) - U\left(\mathcal{E}_U^{(a)}, t+T\right)\right] + \frac{q_t^T\left(1-\delta_0^{(a)}\right)\left(1-\delta_1^{(a)}\right)U'\left(\mathcal{E}_D^{(a)}, t+T\right)}{w_D^{(a)}} + \frac{q_t^T\delta_0^{(a)}\delta_1^{(a)}U'\left(\mathcal{E}_U^{(a)}, t+T\right)}{w_U^{(a)}} \tag{60}$$

$R_{TR}^{L'}$ is non-negative as $\mathcal{E}_D^{(a)} \geq \mathcal{E}_U^{(a)}$ and $U\left(p_t, t\right)$ is an increasing function of p. The second order derivative of R_{TR}^{L} can be expressed as follows.

$$\begin{aligned} R_{TR}^{L''} = {} & q_t^T\left(\delta_1^{(a)} - \delta_0^{(a)}\right)\left[\frac{q_t^T\left(1-\delta_0^{(a)}\right)\left(1-\delta_1^{(a)}\right)U'\left(\mathcal{E}_D^{(a)}, t+T\right)}{\left[w_D^{(a)}\right]^2} - \frac{q_t^T\delta_0^{(a)}\delta_1^{(a)}U'\left(\mathcal{E}_U^{(a)}, t+T\right)}{\left[w_U^{(a)}\right]^2}\right] \\ & + \frac{q_t^T\left(1-\delta_0^{(a)}\right)\left(1-\delta_1^{(a)}\right)\mathcal{E}_D^{'(a)}U''\left(\mathcal{E}_D^{(a)}, t+T\right)w_D^{(a)} - q_t^T\left(1-\delta_0^{(a)}\right)\left(1-\delta_1^{(a)}\right)U'\left(\mathcal{E}_D^{(a)}, t+T\right)q_t^T\left(\delta_1^{(a)} - \delta_0^{(a)}\right)}{\left[w_D^{(a)}\right]^2} \\ & + \frac{q_t^T\delta_0^{(a)}\delta_1^{(a)}\mathcal{E}_U^{'(a)}U''\left(\mathcal{E}_U^{(a)}, t+T\right)w_U^{(a)} + q_t^T\delta_0^{(a)}\delta_1^{(a)}U'\left(\mathcal{E}_U^{(a)}, t+T\right)q_t^T\left(\delta_1^{(a)} - \delta_0^{(a)}\right)}{\left[w_U^{(a)}\right]^2} \end{aligned} \tag{61}$$

$$R_{TR}^{L''} = \frac{\left[q_t^T\left(1-\delta_0^{(a)}\right)\left(1-\delta_1^{(a)}\right)\right]^2 U''\left(\mathcal{E}_D^{(a)}, t+T\right)}{\left[w_D^{(a)}\right]^3} + \frac{\left[q_t^T\delta_0^{(a)}\delta_1^{(a)}\right]^2 U''\left(\mathcal{E}_U^{(a)}, t+T\right)}{\left[w_U^{(a)}\right]^3} \tag{62}$$

Since $U\left(p_t, t\right)$ is a convex function of p. Therefore $R_{TR}^{L''} \geq 0$ and hence $U_{TR}\left(p_t, t\right)$ is convex. ■

***Lemma** 4:* $U_{TS}\left(p_t, t\right)$ increases in p for a given t.

Proof:

$$R_{TS}^{M'} = \left(w_F^{(a)}(1)w_D^{(b)}\right)' \prod_{l=2}^{m} w_F^{(a)}(l) T \log\left(1 + SNR_{TS}^{(b)}\right) \tag{63}$$

$$R_{TS}^{M'} = \left(q_t^S(1)\left(P_{d,1} - P_{f,1}\right)w_D^{(b)} + w_F^{(a)}(1)q_t^T\left(\delta_1^{(b)} - \delta_0^{(b)}\right)\right)\prod_{l=2}^{m} w_F^{(a)}(l) T \log\left(1 + SNR_{TS}^{(b)}\right) \tag{64}$$

Generally $P_d > P_f$ and $\delta_1 > \delta_0$. Hence $R_{TS}^{M'} \geq 0$. Next we will find the first order derivative for the long term reward in the TS mode. To do that, we will split the long term reward shown in (42) into two parts to be $R_{TS}^{L} = Z_1 + Z_2$. Let $C_1 = \prod_{l=2}^{m} w_F^{(a)}(l)$.

$$\begin{aligned} Z_2' = {} & \left(w_D^{(b)}w_F^{(a)}(1)\right)' C_1 U\left(\mathcal{E}_D^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}\left(l\right), t+T\right) + w_D^{(b)}w_F^{(a)}(1)C_1U'\left(\mathcal{E}_D^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}\left(l\right), t+T\right)\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)\right)'\prod_{l=2}^{m}\mathcal{E}_F^{(a)}(l) \\ & + \left(w_U^{(b)}w_F^{(a)}(1)\right)' C_1 U\left(\mathcal{E}_U^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}(l), t+T\right) + w_U^{(b)}w_F^{(a)}(1)C_1U'\left(\mathcal{E}_U^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}(l), t+T\right)\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)\right)'\prod_{l=2}^{m}\mathcal{E}_F^{(a)}(l) \end{aligned} \tag{65}$$

$$\begin{aligned} Z_2' = {} & \left[q_t^T\left(\delta_1^{(b)} - \delta_0^{(b)}\right)w_F^{(a)}(1) + w_D^{(b)}q_t^S(1)\left(P_{d,1} - P_{f,1}\right)\right]C_1U\left(\mathcal{E}_D^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}\left(l\right), t+T\right) \\ & + C_1\prod_{l=2}^{m}\mathcal{E}_F^{(a)}(l)U'\left(\mathcal{E}_D^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}\left(l\right), t+T\right)\left[q_t^S(1)\mathcal{E}_D^{(b)}\left(1-\delta_1^{(b)}\right)\left(1-P_{f,1}\right) + q_t^T\mathcal{E}_F^{(a)}(1)\left(1-\delta_0^{(b)}\right)\left(1-P_{d,1}\right)\right] \\ & + \left[-q_t^T\left(\delta_1^{(b)} - \delta_0^{(b)}\right)w_F^{(a)}(1) + w_U^{(b)}q_t^S(1)\left(P_{d,1} - P_{f,1}\right)\right]C_1U\left(\mathcal{E}_U^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}(l), t+T\right) \\ & + C_1\prod_{l=2}^{m}\mathcal{E}_F^{(a)}(l)U'\left(\mathcal{E}_U^{(b)}\prod_{l=1}^{m}\mathcal{E}_F^{(a)}(l), t+T\right)\left[q_t^S(1)\mathcal{E}_U^{(b)}\delta_1^{(b)}\left(1-P_{f,1}\right) + q_t^T\mathcal{E}_F^{(a)}(1)\delta_0^{(b)}\left(1-P_{d,1}\right)\right] \end{aligned} \tag{66}$$

Since $U\left(p_t,t\right)$ is an increasing function of p (To be justified in lemma (7)). Since $\mathcal{E}_D^{(b)} \geq \mathcal{E}_U^{(b)}$. Therefore, the first and third terms of $Z_2^{'}$ are non-negative. Hence, $Z_2^{'} \geq 0$

$$\begin{aligned} Z_1 = w_B^{(a)}(1)U\left(\mathcal{E}_B^{(a)}(1), t+T_{S1}\right) + w_F^{(a)}(1)w_B^{(a)}(2)U\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_B^{(a)}(2), t+T_{S1}+T_{S2}\right) \\ + w_F^{(a)}(1)w_F^{(a)}(2)w_B^{(a)}(3)U\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_F^{(a)}(2)\mathcal{E}_B^{(a)}(3), t+T_{S1}+T_{S2}+T_{S3}\right)+\ldots \end{aligned} \tag{67}$$

To simplify the proof, the first term of (67) can be safely neglected. The reason is that the probability of getting a busy outcome after T_{S1} (very small period) is almost negligible, given that p_t is sufficiently large (to satisfy the PU collision probability constraint) and m is large enough. Note also that the probability that the PU will switch its state during T_{S1} is very small given that the PU's ON and OFF periods are much longer than T_{S1}.

$$\begin{aligned} Z_1^{'} &= q_t^S(1)\left(P_{d,1}-P_{f,1}\right)w_B^{(a)}(2)U\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_B^{(a)}(2), t+\sum_{l=1}^{2}T_{Sl}\right) + w_F^{(a)}(1)w_B^{(a)}(2)U^{'}\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_B^{(a)}(2), t+\sum_{l=1}^{2}T_{Sl}\right)\mathcal{E}_F^{'(a)}(1)\mathcal{E}_B^{(a)}(2) \\ &+ q_t^S(1)\left(P_{d,1}-P_{f,1}\right)w_F^{(a)}(2)w_B^{(a)}(3)U\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_F^{(a)}(2)\mathcal{E}_B^{(a)}(3), t+\sum_{l=1}^{3}T_{Sl}\right) \\ &+ w_F^{(a)}(1)w_F^{(a)}(2)w_B^{(a)}(3)U^{'}\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_F^{(a)}(2)\mathcal{E}_B^{(a)}(3), t+\sum_{l=1}^{3}T_{Sl}\right)\mathcal{E}_F^{'(a)}(1)\mathcal{E}_F^{(a)}(2)\mathcal{E}_B^{(a)}(3)+\ldots \\ &\geq 0 \end{aligned}$$

■

***Lemma** 5: $U_{TS}\left(p_t,t\right)$ is a convex function of p for a given t.*

Proof: First for the Myopic reward:

$$R_{TS}^{M^{''}} = 2C_1T\,q_t^S(1)q_t^T\left(P_{d,1}-P_{f,1}\right)\left(\delta_1^{(b)}-\delta_0^{(b)}\right)\log\left(1+SNR_{TS}^{(b)}\right) \geq 0 \tag{68}$$

Next, we will find the second order derivative of the long term reward. Let $C_2 = \prod_{l=2}^{m}\mathcal{E}_F^{(a)}(l)$. We first start with finding $Z_2^{''}$ and then proceed to $Z_1^{''}$.

$$\begin{aligned} Z_2^{''} &= 2C_1q_t^Tq_t^S(1)\left(\delta_1^{(b)}-\delta_0^{(b)}\right)\left(P_{d,1}-P_{f,1}\right)\left[U\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)-U\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\right] \\ &+ C_1C_2\left[q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)w_F^{(a)}(1)+w_D^{(b)}q_t^S(1)\left(P_{d,1}-P_{f,1}\right)\right]\left[U\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\right]^{'} \\ &+ C_1C_2^2U^{''}\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)\right)^{'}\left[q_t^S(1)\mathcal{E}_D^{(b)}\left(1-\delta_1^{(b)}\right)\left(1-P_{f,1}\right)+q_t^T\mathcal{E}_F^{(a)}(1)\left(1-\delta_0^{(b)}\right)\left(1-P_{d,1}\right)\right] \\ &+ C_1C_2U^{'}\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\left[q_t^S(1)\mathcal{E}_D^{'(b)}\left(1-\delta_1^{(b)}\right)\left(1-P_{f,1}\right)+q_t^T\mathcal{E}_F^{'(a)}(1)\left(1-\delta_0^{(b)}\right)\left(1-P_{d,1}\right)\right] \\ &+ C_1C_2\left[-q_t^T\left(\delta_1^{(b)}-\delta_0^{(b)}\right)w_F^{(a)}(1)+w_U^{(b)}q_t^S(1)\left(P_{d,1}-P_{f,1}\right)\right]\left[U\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\right]^{'} \\ &+ C_1C_2^2U^{''}\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)\right)^{'}\left[q_t^S(1)\mathcal{E}_U^{(b)}\delta_1^{(b)}\left(1-P_{f,1}\right)+q_t^T\mathcal{E}_F^{(a)}(1)\delta_0^{(b)}\left(1-P_{d,1}\right)\right] \\ &+ C_1C_2U^{'}\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)\left[q_t^S(1)\mathcal{E}_U^{'(b)}\delta_1^{(b)}\left(1-P_{f,1}\right)+q_t^T\mathcal{E}_F^{'(a)}(1)\delta_0^{(b)}\left(1-P_{d,1}\right)\right] \end{aligned} \tag{69}$$

Since $U\left(p_t,t\right)$ is an increasing function of p (to be justified) and since $\mathcal{E}_D^{(b)} \geq \mathcal{E}_U^{(b)}$. Therefore $U\left(\mathcal{E}_D^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right) \geq U\left(\mathcal{E}_U^{(b)}\mathcal{E}_F^{(a)}(1)C_2, t+T\right)$ and hence the first term of $Z_2^{''}$ is non-negative. Since convex functions have always an increasing slope and since $U\left(p_t,t\right)$ is proved to be a convex function of p. Therefore the second and fifth terms of (69) are non-negative. Hence $Z_2^{''}$ is non-negative.

$$\begin{aligned} Z_1^{''} &= w_B^{(a)}(2)\mathcal{E}_B^{2(a)}(2)q_t^S(1)\left(1-P_{f,1}\right)\left(1-P_{d,1}\right)\mathcal{E}_F^{'(a)}(1)\left(1/w_F^{'(a)}(1)\right)U^{''}\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_B^{(a)}(2), t+\sum_{l=1}^{2}T_{Sl}\right) \\ &+ w_F^{(a)}(2)\mathcal{E}_F^{2(a)}(2)w_B^{(a)}(3)\mathcal{E}_B^{2(a)}(3)q_t^S(1)\left(1-P_{f,1}\right)\left(1-P_{d,1}\right)\mathcal{E}_F^{'(a)}(1)\left(1/w_F^{'(a)}(1)\right)U^{''}\left(\mathcal{E}_F^{(a)}(1)\mathcal{E}_F^{(a)}(2)\mathcal{E}_B^{(a)}(3), t+\sum_{l=1}^{3}T_{Sl}\right)+\ldots \end{aligned}$$

Since $U\left(p_t,t\right)$ is a convex function of p. Therefore $Z_1^{''}$ is non-negative, which completes the proof. ■

***Lemma** 6: $U_{SO}\left(p_t,t\right)$ is an increasing and convex function of p for a given t.*

Proof:

$$R_{SO}^{L'}=q_t^S\left(\tilde{P}_d-\tilde{P}_f\right)\left[U\left(\mathcal{E}_F^{(a)},t+T_S\right)-U\left(\mathcal{E}_B^{(a)},t+T_S\right)\right] + \frac{q_t^S\left(1-\tilde{P}_f\right)\left(1-\tilde{P}_d\right)U'\left(\mathcal{E}_F^{(a)},t+T_S\right)}{w_F^{(a)}}+\frac{q_t^S\tilde{P}_f\tilde{P}_dU'\left(\mathcal{E}_B^{(a)},t+T_S\right)}{w_B^{(a)}} \tag{70}$$

Where

$$q_t^S=\frac{1-F_X\left(t+T_S\right)}{1-F_X\left(t\right)}. \tag{71}$$

$R_{SO}^{L'}$ is non-negative as $\mathcal{E}_F^{(a)}\geq\mathcal{E}_B^{(a)}$ and $U\left(p_t,t\right)$ is an increasing function of p. The second order derivative of R_{SO}^{L} can be expressed as follows.

$$R_{SO}^{L''}=\frac{\left[q_t^S\left(1-\tilde{P}_f\right)\left(1-\tilde{P}_d\right)\right]^2U''\left(\mathcal{E}_F^{(a)},t+T_S\right)}{\left[w_F^{(a)}\right]^3}+\frac{\left[q_t^S\tilde{P}_f\tilde{P}_d\right]^2U''\left(\mathcal{E}_B^{(a)},t+T_S\right)}{\left[w_B^{(a)}\right]^3} \tag{72}$$

Since $U\left(p_t,t\right)$ is a convex function of p. Therefore $R_{SO}^{L''}\geq 0$ and hence $U_{SO}\left(p_t,t\right)$ is convex. ■

***Lemma** 7: $U(p_t,t)$ increases in p for a given t.*

Proof: This lemma can be proved using backward induction on t [15]. Since $U(p_t,t)=0\,\forall t>t^*$ (as $U(p_t,t)$ is convex in p and $U(1,t)=0,\forall t>t^*$). Therefore, $U(p_t,t)$ is an increasing function of $p,\forall t>t^*$. Assume that for $t\geq t^*-k$, $U(p_t,t)$ increases in p. Let us now check whether $U(p_t,t)$ increases in p or not for time instant $t=t^*-k-1$. Using (21), and since we proved that $U_{TR}\left(p_t,t\right),U_{TS}\left(p_t,t\right),U_{SO}\left(p_t,t\right)$, and $U_{CS}\left(\hat{p}_t,t\right)$ are increasing functions of p (Lemmas ((2), (4), (6)). Note that $U(p_t,t^*-k)$ increases in p by the induction hypothesis. Therefore, $U(p_t,t^*-k-1)$ also increases in p. ■