Title
Spectrum Sharing by Cognitive Radios: Opportunities and Challenges

Author
Tandra, Rahul

Publication Date
2009

Peer reviewed|Thesis/dissertation

Spectrum Sharing by Cognitive Radios: Opportunities and Challenges

by

Rahul Tandra

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in

Engineering—Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Anant Sahai, Chair
Professor David Tse
Professor David Aldous

Fall 2009

Spectrum Sharing by Cognitive Radios: Opportunities and Challenges

by

Rahul Tandra

Abstract

Spectrum Sharing by Cognitive Radios: Opportunities and Challenges
by
Rahul Tandra
Doctor of Philosophy in Engineering—Electrical Engineering and Computer Sciences
University of California, Berkeley
Professor Anant Sahai, Chair

Under the current regulatory model of static frequency assignment, most of the spectrum is allocated while the actual usage is sparse. This seeming waste is commonly referred to as the problem of "regulatory overhead". The advent of frequency-agile cognitive radios represents a potential opportunity to improve performance and reduce this regulatory overhead. The white space ruling on Nov 4th 2008, legalizing the reuse of television band whitespaces, is a first step taken by the FCC to address the issue of regulatory overhead. However, the problem is that we do not yet know what the right regulatory changes are that will reduce the regulatory overhead.

In this thesis we focus on one fundamental aspect of this problem, namely sensing spectrum for empty bands that are not being used at their current time and location for their primary purpose. It turns out that obtaining a suitable technical formulation for this seemingly simple problem is a highly non-trivial task. Traditionally, sensing problems are formulated mathematically using a binary hypotheses test between the two hypotheses — "signal present" and "signal absent". In the latter part of this thesis we show that the traditional framework does not completely capture all the interesting dimensions of the sensing problem. However, there are several 1-bit decision problems in spectrum sharing that can be directly modeled using the hypothesis testing framework, and our technical results are directly applicable for those problems.

In the traditional binary hypothesis testing framework, we show that designing robust algorithms that can distinguish between the two hypotheses at low signal to noise ratios is a very hard problem. Real world uncertainties in the noise plus interference, and the fading process make spectrum sensing very challenging. In particular, we prove that there exist fundamental limits called SNR walls, below which robust detection is impossible, regardless of how many samples we take.

We show that the presence of signal features in the primary signal can significantly improve robustness of detection for the secondary user. Coherent detection algorithms that look for commonly existing signal features like pilot tones and cyclostationary features are also limited by SNR walls, although they are more robust than signals without any features. We also explicitly construct signals with macroscale features that can be robustly detected at arbitrarily low SNRs. These results suggest that in order to allow cognitive radio operation,

the primary user must pay in terms of restrictions on its freedom to choose any possible signaling scheme.

The SNR wall result leads us to the natural question: why does a cognitive radio need to sense at extremely low SNRs? To answer this question, we are forced to look at the spatial dimension of the sensing problem, which introduces new tradeoffs that cannot be fully understood using the traditional hypothesis testing formulation. In this thesis, we propose a new space-time sensing framework that helps us better understand this problem and brings out new tradeoffs that are otherwise not apparent. We give two new metrics, *Fear of harmful interference* (F_{HI}) and the *Weighted probability of space-time recovered* ($WPSTR$) which characterize the safety for the primary user and the performance of the secondary user respectively. These metrics show that single-radio based sensing algorithms can only recover a small fraction of the available opportunities even if they take an infinite number of samples. The key reason for this is that single-radio sensors are forced to be conservative to protect the primary user from atypical fading events. This helps us make a concrete claim for looking at other approaches like collaborative sensing algorithms, multiband algorithms etc.

Acknowledgements

The six years I spent at Berkeley have been a great learning experience. Interacting with some of the brightest minds in the world on a daily basis, and learning from them has helped me grow immensely, both as a person as well as a researcher. In many ways the time I spent in Berkeley has played a significant role in defining the person I am today.

Among all the people I interacted with during my time at Berkeley, my advisor Prof. Anant Sahai has had a tremendous impact on me. I am thankful to him for his great mentorship, for the patience he showed in me, and above all for his passion towards learning. His never ending urge to learn new things, not only in engineering, but in all other aspects of life rubs onto his students, including me. When I first started at Berkeley, I was under the false impression that research is mostly about problem solving. Working with Anant, I quickly realized how important and hard it is to formulate the right questions.

I thank Professors David Tse, Michael Gastpar, Jan Rabaey and David Aldous for serving on my dissertation/qualifying-examination committees. Professor David Tse has been a role model for all the students in the Wireless Foundations. His course of fundamentals of wireless communications taught me the essential ingredients required for research. I am also grateful to him for providing me with invaluable feedback during my job search and in preparation of several important presentations. Interactions with Steve Shellhammer (Qualcomm) were also very helpful in shaping the course of my research in cognitive radios. I am also thankful for all the professors at Berkeley whose courses taught me the fundamentals required to excel as a researcher.

My fellow students in Wireless Foundations became an integral part of my life at Berkeley. It has been an honor to have known and interacted with students in Professor Sahai's research group — Cheng Chang, Shridhar Mubaraq Mishra, Hari Palaiyanur, Pulkit Grover, Kristen Woyach, Niels Hoven, and Se Yong Park. Shridhar Mubaraq Mishra has been my closest collaborator, and it was lot of fun working together for endless hours on various research problems. I am thankful to Hari, Pulkit and Kristen for their feedback in the preparation of this dissertation, and also several of my other research papers. Interactions with my group mates extended more than just about research. I thoroughly enjoyed the time spent with them during several hiking trips, workout sessions, basketball games (thanks Hari for being such a great coach!), dinners, and potluck parties. They became great friends of mine and I will cherish their friendship for the rest of my life.

I am thankful to Vinod Prabhakaran for being very helpful to me, especially when I had a knee surgery. He was very selfless and offered to drive me to wherever I needed to go. He was like my elder brother, always ready to help me when I am in need! My first technical meeting in Berkeley was with Sheila Ross to discuss the semantics of the course for which I was her teaching assistant. It was in the meeting when I met Bobak Nazer, John Secord and Joe Makin, who we also teaching assistants for the same course. Bobak and I

have become great friends ever since, and I enjoyed having several stimulating conversations with him throughout my time at Berkeley. John Secord and Joe Makin were captains of the softball team named Headgear, and it was fun playing alongside them for several years. I am also happy to have know and interacted with students in Wireless Foundations: Krishnan Eswaran, Pablo Minero, Alex Dimakis, Salman Avestimehr, June Wang, Anand Sarwate, Animesh Kumar, Mark Johnson, Dapo Omidiran, Galen Reeves, Guy Bresler, Jiening Zhan, Nebojsa Milosavljevic, Naveen Goela, Sahand Negahban, Barlas Oguz, Changho Suh, I-Hsiang Wang, Hao Zhang, Amin Gohari, Sudeep Kamath, Raul Etkin, Lara Dolecek, Prasad Santhanam, Aaron Wagner, and Parvathinathan Venkita Subramaniam.

I am also thankful to my roommates and friends with whom I spent six memorable years in Berkeley: Pankaj Kalra, Kaushik Ravindran, Krishnendu Chatterjee, Mohan Vamsi Dunga, Arkadeb Ghosal, Satrajit Chatterjee, Rohit Ambekar, Sujit Kirpekar, Rohit Karnik, Vineet Gupta, Puneet Gupta, Gautham Gupta, Ankit Jain, Shariq Rizvi, and Nadathur Satish.

Last but not least, I am thankful to my family for supporting me through all my endeavors and for giving me the freedom to chase my dreams.

To my family,
for their unconditional love and affection.

Contents

Chapter 1

Introduction

Significant advances in wireless technologies over the last decade have left us on the verge of another IT revolution. Billions of users will carry portable communication/computation devices with wireless as their primary mode of connectivity. To support this grand vision of an Internet-like revolution in wireless it is important to scale the performance of systems appropriately. This leads us to the big question: how can we get the next $10 - 100$ factor improvement in the performance of wireless systems?

To address this question we take a look at the different layers (physical layer (PHY), MAC layer, network layer etc.) in a communication system architecture. Each of these layers are primarily designed by engineers and there are reasonably well defined metrics that can measure performance of these layers. For example, data rate and bit-error probability in the PHY layer, average medium occupancy and packet collision probability in the MAC layer, and latency and average queue length in the network layer. With the aid of these metrics it is easy to get an idea of how much room for improvement is possible in each of these layers.

As engineers, we often ignore the existence of another layer: the "regulatory layer" that provides access to spectrum for wireless systems. This layer provides a system with details of the band of operating frequencies, power constraints, interference margins etc. The design of this layer is primarily done by non-engineers like policy makers, economists and lawyers. This leads to an important question: is there a lot of room for improvement in the design of this layer? Furthermore, can we use engineering principles to derive reasonably-approximate metrics for the regulatory layer?

To answer these questions we first look at the current regulatory layer. In the United States, the *Federal Communications Commission* (FCC) regulates interstate and international communications by radio, television, wire, satellite and cable under a *command-and-control* model [1]. The FCC allocates frequency bands to be exclusively used for a particular service, within a given spatial region, and for a specified time duration. Figure 1.1 shows the National Telecommunications and Information Administration's (NTIA) chart of spectrum

allocation in the United States [2]. From the spectrum allocation chart it is evident that most of the usable frequencies are already allocated and that there is very little room for future innovative services. On the other hand, Figure 1.2 shows a plot of the real-time usage of spectrum in the 0 to 2.5 GHz band [3]. The usage picture shows that only a small fraction (about 5%) of the spectrum is actually used[1]. The inefficient use of spectrum due to the static and exclusive-use allocation model is sometimes referred to as a problem of *regulatory overhead* [6].

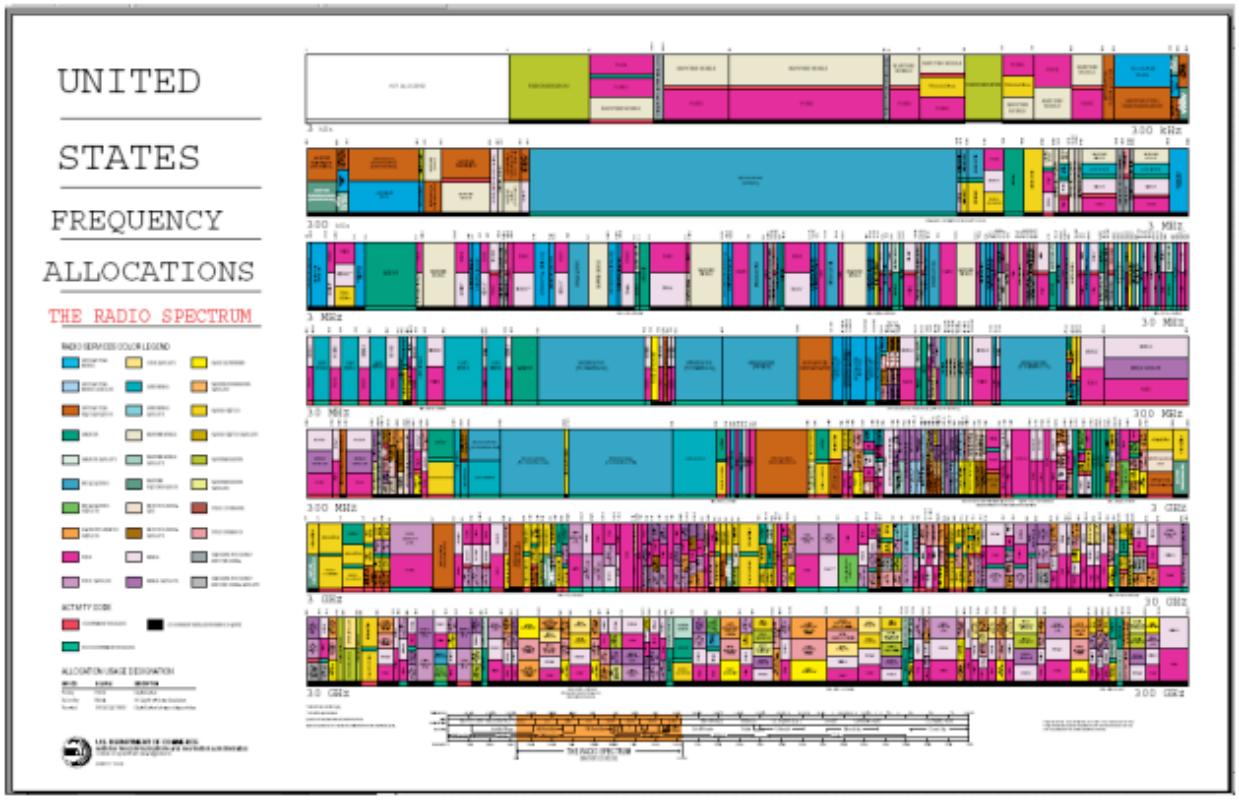


Figure 1.1: Spectrum allocation chart for the United States. It is evident from the chart that most of the frequencies are already allocated, and there is very little room for new and innovative services in the future.

[1]There were several other spectrum-measurement studies done [4; 5] in the United States, each of these confirmed the fact that most bands in most places are underused most of the time.

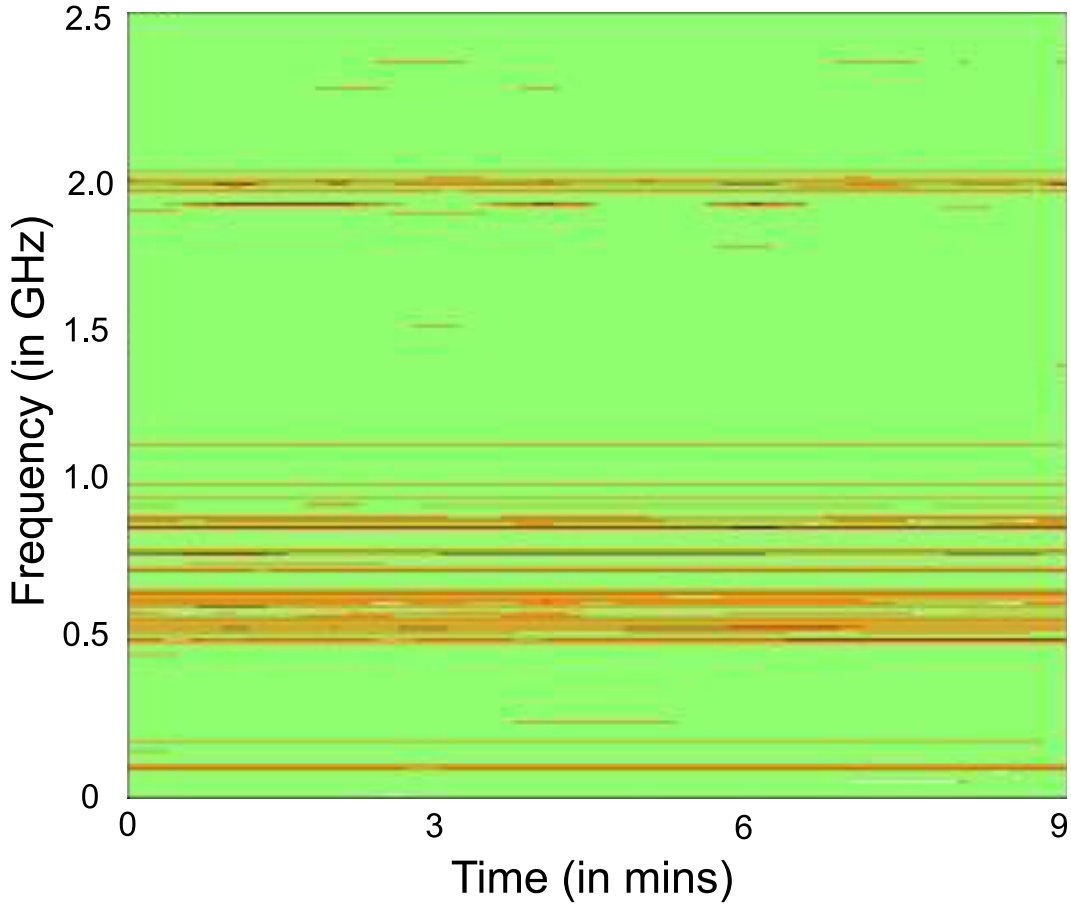


Figure 1.2: A ten minute snap shot of the spectral activity in the 0-2.5 GHz spectrum. The figure shows real-time measurements taken at the Berkeley Wireless Research Center (BWRC) in downtown Berkeley. The 'brown' regions in the plot shows activity in time and frequency. As it is evident from the figure, most of the frequencies are unused.

1.1 Dynamic spectrum access

The unused spectrum opportunities, referred to as spectrum holes are a natural consequence of the gap between the distinct scales at which regulation and use occur. This is because spectrum allocations and planning is done over a time-span of several years/decades, whereas the use of spectrum occurs at time scales of the order of seconds/minutes. This suggests that in order to solve this problem of regulatory overhead a dynamic approach to spectrum access is required. However, the exact form of these dynamic allocation strategies is debated. There have been two extreme proposals to solve the problem of regulatory overhead — spectrum privatization and spectrum commons model. Spectrum privatization advocates allocation of spectrum through real-time markets [1; 7], whereas the spectrum commons proponents argued that the task of sharing spectrum must be left to the devices themselves and that advances in technology will make the devices capable of dynamically sharing spectrum efficiently [8; 9]. The reader is advised to refer to [10] for a detailed survey on the history of the commons versus privatization debate, and also other regulatory issues in spectrum sharing.

A hybrid model for spectrum access known as *opportunistic spectrum sharing* has been proposed to reduce the regulatory overhead. In this model, radios (secondary users) can opportunistically refill the spectrum holes in licensed bands (primary users) as long as they do not cause significant harmful interference to the licensed users. This model has the advantage of being implementable with very little changes to the currently existing system, and hence is a more politically feasible solution than both the privatization or commons models. The opportunistic spectrum sharing approach had gained significant traction in the last few years due to technological advances in radio technology. In particular, the advent of frequency agile and software defined radios, also known as cognitive radios has made the vision of opportunistic spectrum access technically feasible.

The term "cognitive radio" was initially coined by Mitola in the late 1990's [11; 12]. In a broad sense, *cognitive radios* are devices that can sense their environments and autonomously adapt and optimize their system parameters based on the changing operating conditions [13; 14]. Existing wireless systems like cellular networks and wireless local-area networks (WLANs) already have cognitive capabilities like adaptive power control, dynamic channel selection etc. However, the grand vision of the cognitive radio paradigm is for situation-awareness and system-level adaptation [15]. There are many projects that are currently working to incorporate this vision into the next-generation wireless networks. For example, the Defense Advanced Research Projects Agency (DARPA) XG program [16], the IEEE 802.22 wireless regional area network (WRAN) standard [17] (secondary use of television bands), and the European end-to-end reconfigurability (E^2R) research program [18].

In this thesis we mainly focus on opportunistic spectrum sharing using cognitive radios. We consider an opportunistic model because it is an easy extension to the current regulatory system with potential to improve spectral efficiency. Also, the key questions we answer under this model show up both in the privatization and commons models. In an opportunistic

model the cognitive radios must make sure that the band is unused by any primary user in its vicinity before transmitting in the band. So reliably identifying spectrum holes is one of the core technical challenges for cognitive radios. Also, the rules of operation for an opportunistically shared band of spectrum is an important policy question. Presumably, a policymaker would want the rules to be such that they preserve as much flexibility as possible for both the primary and secondary users. Flexibility is an important factor because it gives systems the freedom to innovate and deploy new services in the future. This is an important consideration from a policymaker point of view because the time scale of technological advances is faster than that of rulemaking.

In this thesis, we formulate the following key questions: What are the core technical challenges that need to be overcome in order to make opportunistic spectrum sensing work? How does the solution to these technical challenges influence the regulations? In particular, what are the implication of these results from a 'freedom of use' point of view?

Even though our results hold for any general opportunistic spectrum sharing system, we use the example of cognitive use of television bands ($50 - 800$ MHz) as a case study throughout the thesis. This example is also practically significant because the television (TV) bands were the first licensed bands (licensed to the TV broadcasters) which were identified by the FCC as candidate bands for cognitive radio use (see the Spectrum Policy Task Force report [19]).

The licensed users of TV bands are television stations broadcasting information from high power transmitters mounted on towers (about 500 m high), and wireless microphones[2] (part 74 devices). At any given time and spatial location a subset of the TV channels are unused, and hence these unused opportunities (also called as 'white space') can be used for cognitive transmissions. Since the FCC task force report in 2002, there has been lot of work both in academic circles as well as in industry to identify the key problems that needed to be solved in order to deploy cognitive radios in the TV band[3]. In particular, the IEEE 802.22 task group was formed to come up with a standard for Wireless Rural Area Networks that can operate opportunistically within the TV bands [17]. Finally, on November 4, 2008 the FCC passed a ruling that legalized the use of 'white space' devices in the television bands [26]. The white space ruling allows devices with a combination of geo-location capability (GPS measurement of its position as well as the TV tower locations) and sensing technology to operate in locations and bands that are not currently registered as occupied. The FCC ruling

[2]Technically, a wireless microphone needs to be registered with the TV broadcasters before it can transmit in the TV bands. However, there are a significant number of wireless microphones that are currently operating (illegally) in the TV bands, which are not registered with the broadcasters. However, any cognitive radio operating in the TV bands is supposed to protect wireless microphones that are both legally and illegally operating.

[3]A comprehensive review of all the literature on dynamic spectrum sharing can be found in the proceedings of the IEEE symposia on new frontiers in dynamic spectrum access networks (DySpAN) [20; 21; 22]. There are several other special issues of journals which are exclusively devoted to papers on cognitive radios and dynamic spectrum access [23; 24; 25].

was made after considering all the work done by researchers on spectrum sensing, including some of our key results discussed in this thesis.

1.2 Opportunistic spectrum sharing

We now formulate the problem of opportunistic reuse of television bands by cognitive radios. For simplicity, we focus on a single television band (6 MHz wide) with primary transmitters[4] (TV towers) distributed over a large geographic area with non-overlapping service regions[5]. A television station's transmitter is mounted on a high tower and serves a large area (radius of $\approx$ 135 km). As we move away from the TV tower, the primary signal is very weak and hence the band is clean enough for the cognitive radio to use it for its transmissions. This is a pollution-oriented viewpoint [31]. Moreover a secondary user at such far away locations can transmit without causing too much interference to the primary receivers. This is a protection-oriented viewpoint for the secondary user. Our attention will mostly be focused on a single one of those towers and the area around it.

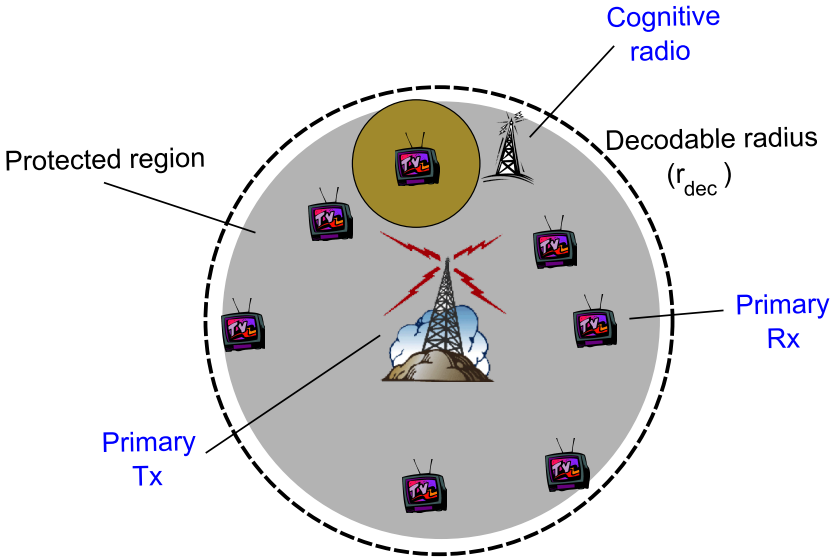
Figure 1.3(a) shows a primary transmitter broadcasting information to its receivers. In the absence of interference from cognitive radios, a receiver within the dashed circle (Figure 1.3(a)) with radius r_{dec} would be able to decode a signal from the transmitter, while a receiver outside the circle would not. To tolerate any secondary users, the primary receiver needs to accept some additional interference. The grey circle represents the protected radius (denoted r_p) where decodability is guaranteed to primary receivers. Primary receivers between the two circles may not be able to get service once secondary systems come on, but this is considered to be an acceptable loss of primary user's quality of service (QoS)[6]. We call the annulus between r_{dec} and r_p as a "sacrificial zone". The time-dimension equivalent of this annulus of width $r_{dec} - r_p$ is the short sacrificial time-segment at the beginning of a primary transmission during which secondary users are permitted to cause interference (see Chapter 5). Like its spatial equivalent, this can be viewed as either a loss of QoS for the primary user in the sense of a dropped frame or as requiring the primary user to lengthen its synchronization preamble before commencing data transmission. Without this provision, a secondary user could never transmit due to the fear of primary user reappearance during the secondary transmission.

Given the primary transmit power (p_t), the protection radius r_p, the minimum decodable

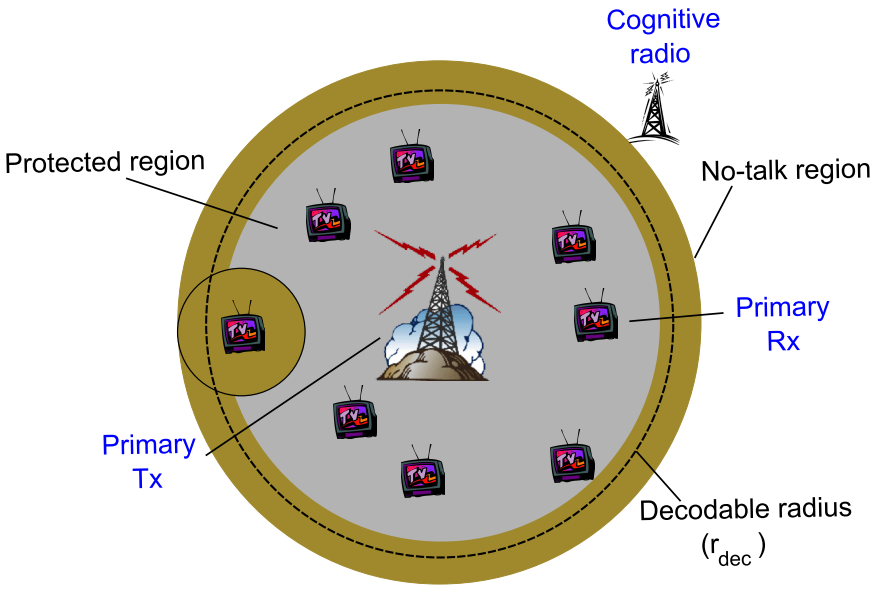
[4]For simplicity, we ignore the issue of peaceful coexistence with wireless microphones operating in the television band. Such smaller scale primary users introduce additional challenges [27; 28].

[5]We can also leverage gains by considering multiple primary bands simultaneously for sensing. The gains in this case arise from the fact that shadowing is highly correlated across frequency and hence measurements in one band can provide robust information about the shadowing environment in other bands [29; 30].

[6]This can be viewed as either the loss of service to certain customers of the primary system or an additional cost of transmit power that must be spent by the primary user to maintain service to all the same customers.



(a)



(b)

Figure 1.3: Opportunistic spectrum sensing

signal to noise plus interference ratio[7] (η), and the pathloss function $l(r)$, we can compute the maximum allowable secondary interference at the primary receivers

$$\begin{aligned} \frac{p_t \cdot l(r_p)}{\sigma_n^2 + I_s} &\geq \eta \\ \Rightarrow I_s &\leq \frac{p_t \cdot l(r_p)}{\eta} - \sigma_n^2, \end{aligned}$$

where σ_n^2 is the background noise power in the frequency band, and I_s is the interference from the secondary users at the location of the primary receivers.

The maximum allowable secondary interference (I_s) at the primary receiver translates into a no-talk region around each primary receiver where a secondary user cannot safely transmit. The radius of this no-talk region depends on the nature of the secondary transmission. If it has low transmit power, the no-talk zone around each receiver can be small. If it has high transmit power, the radius of the no-talk zone becomes much larger[8].

The overall no-talk area is thus the union of the no-talk regions of all primary receivers. The spectrum hole is the complement of this union. To recover this area, the secondary system must know the locations of all primary receivers (see Figure 1.3(a)). Since a primary user may know this information, such complete area recovery might be possible with explicit primary participation. In addition, secondary users themselves may be able to determine the locations of receivers for particular TV channels by sensing the TV receivers themselves [34].

However, if we want the primary user to have the flexibility of deploying receivers anywhere within its protected region, or to have the flexibility to deploy passive receivers (which cannot be sensed), then the secondary user cannot hope to recover spectrum holes within the protected region. Moreover, even if a secondary transmitter can safely transmit in a particular location on a particular band it does not imply that it should want to do so. After all, close to a functioning primary receiver there will usually be a lot of interference from the primary signal itself. This is a pollution-oriented view for the secondary users [31]. Consequently, in this thesis we focus on recovering the region outside the global no-talk zone (r_n), as shown in Figure 1.3(b). The global no-talk radius is the union of the no-talk radius around primary receivers corresponding to all possible locations for protected primary receivers[9]. For simplicity, we refer to the global no-talk radius as the no-talk radius in the

[7]The FCC ruling specifies this protection margin to the primary receivers as a lower bound on the desired to undesired signal power (signal to interference) ratio at the edge of the protection radius. The value of this protection ratio in the FCC ruling is 23 dB [26].

[8]For simplicity, this discussion assumes a single simultaneous secondary transmission. In practice, the secondary system is likely to contain many transmitters operating simultaneously over a distributed area. Such systems can have their user footprints considered in terms of their power density as shown in [32] and [33]. However, the analysis in [27] shows that this interpretation becomes problematic when we really try to scale to secondary users with very different footprints.

[9]The global no-talk radius can be computed from the knowledge of the protection radius, the transmit

rest of the thesis. The cognitive radio can use the band only if it lies outside the no-talk[10] radius of all the towers transmitting on this frequency band.

1.2.1 The spectrum sensing approach

If a cognitive radio wants to opportunistically reuse primary spectrum then it must be able to reliably decide whether it is inside or outside the primary transmitter's no-talk radius r_n. Ideally, a cognitive radio should make this decision without imposing too many restrictions on the flexibility of the primary system. Note that we already assumed that the primary transmitter's power p_t and its service region r_p is known to the cognitive radio, and that they are fixed. This is a restriction on the primary's flexibility to use any transmit power it wants. However, this restriction is tolerable as the primary has to obey the power spectral density mask in the band, and it has no incentive to transmit at any power lesser than the maximum allowable spectral mask.

One possible way for the cognitive radio to know whether it is inside or outside the no-talk radii of all primary transmitters is to have access to a GPS receiver that tells its own location and also a database of all primary transmitter locations. The problem with this approach is that additional infrastructure in the form of a GPS receiver is needed at the cognitive radio. Moreover, this approach imposes a restriction on the freedom of mobility for the primary transmitters. For instance, maintaining a database of all the TV transmitter locations does not impose any restriction on the TV broadcasters as the TV towers are stationary. However, maintaining a database of all the wireless microphones can impose severe restrictions on their freedom of mobile deployment.

So, in this thesis we assume that the cognitive radio does not have access to its relative location with respect to the primary transmitter. Instead, the cognitive radio can sense the primary transmitter's signal and use it as a proxy for the distance between the primary transmitter and itself. This information can be implicitly used by the cognitive radio to decide whether it is inside or outside the primary's no-talk radius. We call this a 'sense before talk' approach. In this approach the primary transmitter's signal at the edge of its no-talk radius can be weak enough that it can no longer be decoded, especially if the cognitive radio transmits at moderately high powers [35]. This means that the cognitive radio must be

power of the cognitive radio, and the propagation model from the cognitive radio to the primary receiver[35; 33; 31].

[10] An empirical computation of the no-talk radius around every TV tower, and across all the TV channels in the United States is done in [36; 31; 37]. Using this computation the authors show that on an average ≈ 31 TV channels per person are available for white space use if an uniform population density is assumed, and an average of ≈ 19 TV channels per person are available when using the actual population density. The authors in [31] used the US Census data of 2000 which lists the population density per zip code [38]. The amount of white space in the TV bands when viewed from a capacity point of view is estimated in [39]. These results show that the opportunity provided by TV whitespaces is shown to be potentially of the same order as the recent release of 700MHz spectrum for wireless data service.

able to detect undecodable signals. Designing sensors that can detect undecodable signals is uncommon as in most traditional communication systems there is no point in detecting signals if they are too weak to decode.

There have been several other formulations for spectrum sharing that are different than the 'sense before talk' approach. Information-theoretic models for cognitive radios have also been proposed [40] in which the spectrum sharing problem is viewed as a joint primary-secondary communication design problem with constraints on the interference at the primary receiver. This model assumes that the primary explicitly participates in spectrum sharing with the secondary system. In these results, the secondary transmitter decodes the primary signal using dirty paper coding (DPC) techniques and simultaneously boost the primary signal in the direction of interference [41; 42]. The key assumption in these results is that the secondary transmitter assumes knowledge of the codeword of the primary transmitter, which might not always be practical. Furthermore, it has also been shown that DPC based approaches are not robust since simple phase uncertainty can significantly lower the performance of such schemes [43].

Other forms of partial information, like knowledge of the primary user's codebook, are also not useful unless the secondary receiver can actually decode the primary signal and use multiuser detection. Otherwise, it has been shown that the secondary system is forced to treat the primary transmission as noise [49]. Implicit feedback from the primary system has also been shown to be useful in the spectrum sharing context. For example, [44; 45] proposes a spectrum-sharing architecture in which the secondary user eavesdrops on a packetized primary user's automatic repeat request (ARQ) messages to stay within the interference budget of the primary users.

1.2.2 Hypothesis-testing based formulation for spectrum sensing

There are different ways in which one can mathematically formulate the sensing problem described in the previous section. We now describe the most popular formulation, in which the sensing problem is cast as a binary hypothesis test between the following two hypotheses — "primary ON" and "primary OFF". This formulation has been used in many research papers on cognitive radios, including our initial paper [35]. We use the same binary hypothesis testing formulation in Chapters 2, 3, and, 4.

Let $X(t)$ denote the band-limited primary signal we are trying to sense, and let the additive noise process be $W(t)$. The discrete-time version is obtained by sampling the received signal at the appropriate rate. The two hypotheses are:

$$\begin{aligned} \text{Signal absent } \mathcal{H}_0 : Y[n] &= W[n] \\ \text{Signal present } \mathcal{H}_1 : Y[n] &= \sqrt{P}H(X[n]) + W[n], \end{aligned} \tag{1.1}$$

for $n = 1, 2, \cdots, N$. Here P is the received signal power, $X[n]$ are the unattenuated samples

(normalized to have unit power) of the primary signal, $H(\cdot)$ is the fading operator (also normalized to have unit power), $W[n]$ are noise samples and $Y[n]$ are the received signal samples. We assume that the *signal is independent of both the noise and the fading process.*

The goal is to design detectors that will distinguish between the two hypothesis (with low probability of error) for low values of P (weak received signal). For instance, the FCC white space ruling says that sensing based cognitive radios must be able to detect the primary signal at -114 dBm [26], which corresponds to an SNR of about -18 dB. Even the IEEE 802.22 standard [46] requires the WRAN devices to sense primary signals at around -20 dB.

For the sake of completeness, we review the traditional performance metrics for binary hypothesis testing [47]. To be concrete, consider test-statistic/threshold based detection algorithms. Let the detector be given by $T(\mathbf{Y}) := \frac{1}{N}\sum_{n=1}^{N}\phi(Y[n]) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda$, where $\phi(\cdot)$ is a known deterministic function and λ is the detector threshold. For a fixed detector threshold λ, and the sensing time N, the error probabilities are defined as

$$\begin{aligned} P_{FA}(N,\gamma) &:= \mathcal{P}_{\mathcal{W}}\left(T(\mathbf{Y}) > \lambda | \mathcal{H}_0\right), \\ P_{MD}(N,\gamma) &:= \mathcal{P}_{\mathcal{W}}\left(T(\mathbf{Y}) < \lambda | \mathcal{H}_1\right). \end{aligned} \tag{1.2}$$

The lowest signal to noise ratio, $SNR := \frac{P}{\sigma_n^2}$ (σ_n^2 is the nominal noise power) for which the constraints in (1.2) are met is called the *sensitivity*[11] of the detector. Furthermore, eliminating γ from (1.2) we can solve for N as a function of the SNR (sensitivity), P_{FA}, and P_{MD}. Hence, we can write

$$N = \xi(SNR, P_{FA}, P_{MD}). \tag{1.3}$$

This is called the *sample complexity* of the detector. Expressions for the asymptotic sample-complexity of commonly-used detectors like the radiometer ($N = O(SNR^{-2})$ [48]), and the sample complexity of a matched filter ($N = O(SNR^{-1})$ [49]) are known.

The traditional metrics triad of sensitivity, P_{FA}, and P_{MD}, are used along with the sample complexity to evaluate the performance of detection algorithms. For reasonable detectors, $\xi(SNR, P_{FA}, P_{MD})$ is a monotonically decreasing function of SNR, P_{FA} and P_{MD}. In particular, when the noise and fading distributions are ergodic, we can choose a threshold λ such that arbitrarily high sensitivities (low SNRs) can be achieved by increasing the number of samples [50; 51].

[11] In signal processing jargon, a highly sensitive detector implies that it can detect extremely weak signals. This usage is somewhat counter intuitive because 'high' sensitivity is related to 'low' SNRs.

1.2.3 Is the sensing problem trivial?

Detection theory tells us that one can always design highly sensitive cognitive radios at the cost of increased sensing time. However, the key assumption that was made in the hypothesis testing formulation in Section 1.2.2 was that the cognitive radio sensor had precise knowledge about the distribution of the noise and fading processes. In reality, the distributions of the noise and fading processes are never known to infinite precision. Thermal noise depends on the circuit elements in the cognitive radio sensor, and it is impossible to fit a precise model for these random fluctuations. It is commonly modeled as 'white' Gaussian, whereas in reality this is only an approximation. Similarly, Rayleigh and Rician fading models are also approximations.

So, an important design question is: how robust are sensing algorithms to modeling uncertainties? This is a very practical question faced by any communications engineer deploying systems in the field. Naturally, there has been lot of work on developing detection algorithms that are robust to distributional uncertainties in the noise and fading processes [52; 53; 54; 55; 56; 57]. However, most of these results on robust detection algorithms assume that the underlying SNR is moderate to high (larger than 0 dB). Whereas, for cognitive radios the functional requirement is to reliably detect signals as weak as -20 dB.

There has been much work on designing sensing algorithms for cognitive radio systems. Table 1.1 gives a brief sampling of some representative single-user sensing techniques. The techniques given there are by no means exhaustive. The reader is encouraged to look into the references within these references for more.

1.3 Thesis outline and main contributions

The discussion in Section 1.2.3 leads to one of the core questions addressed in this thesis: *Can we design robust detection algorithms that work at extremely low signal to noise ratios?* We formulate this question as a robust hypothesis testing problem and analyze the performance of non-coherent detection algorithms in Chapter 2. For simplicity, in this chapter we model the uncertainty in the noise process using a bounded-moment distributional uncertainty model (see Section 2.4.1). Also, we assume that the primary user is flexible to use any signaling scheme of his choice, and we only assume knowledge of the average primary received signal power. Freedom of use for the primary user translates into uncertainty for the cognitive radio sensor. Under these modeling assumptions, we show that there exist fundamental limits to robust detection. These limits are in the form of SNR thresholds, called SNR *walls*, below which robust detection is impossible, regardless of how many samples we take. We first show the SNR wall result for a radiometer. Uncertainty in the noise power is sufficient to induce a wall for the radiometer, and this wall is very severe for moderately low values of the uncertainty.

Furthermore, if the primary use has the flexibility of transmitting *iid* signal samples

$X[n]$ drawn out of a zero-mean symmetric constellation, then even knowledge of the signal constellation at the detector is not helpful. That is, even with knowledge of the signal constellation, *all* possible detection algorithms suffer from an SNR wall limitation and their performance is at most 3 dB better than that of a radiometer. These results have significant impact for policymakers writing regulations for whitespace use. In particular, the SNR wall result says that if the primary user is given the freedom to transmit *iid* symbols from a symmetric signal constellation, then the cognitive radios cannot guarantee robust detection, and hence they will not be able to recover the whitespaces. This means that opportunistic spectrum sharing with complete freedom to the primary user will not solve the regulatory overhead problem.

This leads to the natural question of: what kind of restrictions on the primary user will help the secondary sensor robustly detect the primary signal? Many communication signals use pilot/training symbols to help their receivers achieve phase and symbol synchronization. These known 'signal features' are also useful in estimating the channel for coherent communication. In chapter 3 we analyze how the presence of *known signal features* in the primary signal help in robust detection for the secondary user. In particular, we focus on two commonly used signal features — narrowband pilots, and cyclostationary signal features. The key idea is that known signal features allow the secondary detector to coherently process the primary signal and average out the noise. So, averaging preserves the strength of the signal feature but reduces the uncertainty in the noise process. In the limit of infinite coherent processing (coherent averaging for large sensing time N), any signal with arbitrarily low SNR can be robustly detected in the presence of uncertain noise. This suggests that the presence of known signal features will avoid the SNR wall limitation.

However, we have not yet modeled the uncertainty in the wireless channel between the primary transmitter and the secondary sensor. Fading can be modeled as a linear time-varying filter, with the rate of variation characterized by the coherence time of the channel [58]. As the received primary signal strength can be very low, the secondary sensor cannot track the fading channel coefficients reliably. We model this by assuming an *iid* block fading model for the wireless channel between the primary transmitter and secondary sensor with the latter knowing only a lower bound on the channel coherence time. Under this model, we show that the finite coherence time imposes a limit on the amount of coherence processing gains feasible for coherent detection of the known signal feature. In particular, for the case of narrowband pilots, we show that the uncertainty in the phase of the fading process (phase changes randomly after every coherence time and hence the pilot tone can be coherently matched only for the length of a coherence block.) imposes an SNR wall limitation for pilot detection. However, the SNR wall for a pilot detector is better than that of a radiometer due to the robustness gains achieved by coherent processing for the duration of the coherence time.

In the case of narrowband pilots we show that robustness can be further improved by *run-time noise calibration*, which is an accurate calibration of uncertain statistical quantities like

noise at run-time. If a signal contains a narrowband pilot, then we can measure the average noise plus interference power at frequency locations away from the pilot tone and use it to calibrate the power of the noise plus interference at the location of the pilot tone. The idea is that, whatever the power spectral density of the noise plus interference is, as long as it is approximately 'white' around the pilot tone frequency then calibration can significantly reduce uncertainty in the noise plus interference power during run-time!

In Chapter 3 we also consider signals with cyclostationary features. Most communication signals have built-in periodicities in them, like the symbol rate, carrier frequency, cyclic prefix in OFDM packets, etc. Hence, signals can be modeled as cyclostationary, whereas noise is typically modeled as stationary. Many researchers have worked on designing detection algorithms to exploit the in-built periodicities in signals, of which the class of cyclostationary feature detectors proposed by William Gardner [59] are the most popular ones. These detectors have been widely studied in the spectrum sharing community and are believed to be good for both signal detection as well as signal identification/classification[60; 61].

A serious drawback of Gardners form of cyclostationary feature detectors is that they need the signal to have spectral redundancy in order for the detector to work. That is, these detectors work only when the symbol rate of the signal is slower than the Nyquist rate of the signal. Hence, spectral redundancy was believed to be a fundamental cost one needs to pay for the cyclostationary signal to be detectable. In Chapter 3, we disprove this fact by explicitly constructing examples of signals that do not have spectral redundancy, and yet there exist *histogram-based* sensing algorithms that can detect the underlying cyclostationarity. The key idea behind the counter-example shown in this chapter is the fact that Gardners cyclostationary feature detectors depend on the periodicity of second-order statistics (autocorrelation function) of the signal. Whereas in our example we look for distributional periodicity, which exists in every cyclostationary signal even when the signal is wide-sense stationary (second-order statistics are time-invariant).

From a robustness point of view, we show that cyclostationary signals are non-robust to time-varying frequency selective fading processes. Frequency selective fading mixes the signal samples and as the fading coefficients vary with time it destroys the periodicity in the signal statistics. The signal is locally cyclostationary within the channel coherence time, but it is stationary across multiple coherence blocks. So, even in this case the finite coherence time limits the gains from cyclostationary feature detection algorithms, leading to an SNR wall. Both Gardner's form of cyclostationary feature detectors and our histogram-based detectors suffer from these SNR wall limitations.

Both narrowband pilots as well as cyclostationary signals suffer from SNR wall limitations, and the location of the SNR wall in each of the case depends on the channel coherence time. This leads to the question of: what are the relevant channel coherence times in both cases? In the narrowband pilot case, the change of phase of the fading coefficients causes an SNR wall. So, the relevant coherence time is the time after which the phase of the signal changes significantly. We call this the *phase coherence time*. Note that the phase coherence

time could be finite even when the physical wireless channel is time-invariant (the transmitter, receiver and all reflectors are stationary) due to the frequency jitter of the oscillator at the secondary sensor relative to the oscillator at the primary transmitter. We show that the location of the SNR wall varies as $\log_{10} N_c$ (in dB scale), where N_c is a lower bound on the length of the phase coherence time of the channel.

In the case of cyclostationary signals, the channel forgets the timing of the signal samples relative to the period of the cyclostationary process, causing an SNR wall. So, the relevant coherence time in this case is the time after which the shape of the impulse response of the channel changes significantly. We call this the *delay coherence time* of the channel. The delay coherence time of the channel could be finite even when the fading process is stationary due to the clock drift and jitter at the secondary sensor. Also, the location of the SNR wall for cyclostationary signals varies as $\frac{1}{2}\log_{10} D_c$ (in dB scale), where D_c is a lower bound on the length of the delay coherence time of the channel. More importantly, there can be physical situations in which the delay coherence time D_c is significantly larger than the phase coherence time N_c. So, it is possible that $\frac{1}{2}\log_{10} D_C > \log_{10} N_c$.

If the primary signal has known features, then the results from Chapter 3 tell us that the secondary sensor can robustly detect weak primary signals (SNR walls lower than -20 dB are achievable). However, the cost of this improved robustness for the secondary sensor is the loss in freedom for the primary user, as it is forced to transmit strong features in its signals. So, in Chapter 4 we consider the tradeoff between the loss in freedom for the primary user and the gain in robustness for the secondary sensor. We capture the loss in freedom for the primary user by the hit in data rate (to the primary receivers) it suffers in order to provide robustness gains for the secondary sensor. We call this the *capacity-robustness tradeoff*. We derive the tradeoff curves for several combinations of primary signaling schemes and secondary sensing algorithms. These tradeoff curves help the regulators choose a suitable point in the capacity-robustness tradeoff space.

This formulation leads us to the question of what the optimal capacity-robustness tradeoff curve is. We give an answer to this question in Chapter 4 by constructing an example of a primary signaling scheme that can be robustly detected by the secondary sensor at arbitrarily low SNRs. Furthermore, the loss to capacity for the primary user under this signaling scheme can also be made arbitrarily close to zero. This shows that the optimal capacity-robustness tradeoff is a trivial rectangle!

The signaling scheme is motivated from the results on cyclostationary feature detection in Chapter 3. This signal is constructed by block-interleaving two *iid* white Gaussian signals with distinct average power levels. This gives rise to a cyclostationary signal. The length of each Gaussian block is chosen such that the scale of the feature (period of the cyclostationary process) is larger than the delay spread of the fading channel between the primary user and secondary sensor. This ensures that the fading process cannot completely mix the signal samples and destroy the underlying cyclostationarity. For this reason we say that the signal contains *macroscale* features. This signal is shown to be completely robust to arbitrarily

varying noise and fading processes. On the other hand, the power levels of the two Gaussian processes can be chosen close to make the data rate arbitrarily close to the primary channel capacity.

Having proved that one can design robust sensing algorithms that can detect weak (of the order of -20 dB) signals, it is clear that cognitive radios can opportunistically share spectrum while guaranteeing protection to the primary users. The question is: what fraction of the spectrum holes does the cognitive radio recover? More concretely, in the TV band case can we design sensing algorithms that do as well as a geo-location based spectrum sharing solution? The traditional sensing metrics of P_{FA}, P_{MD} and the sensing time N are insufficient to give answers to these questions. The reason is that the hypothesis testing formulation given in Section 1.2.2 does not capture the spatial perspective for spectrum sensing, which is essential for cases with primary transmitters distributed over space.

We formulate the joint time-space spectrum sensing problem in Chapter 5. We present two new metrics, the *fear of harmful interference* (F_{HI}) and the weighted probability of space-time recovered ($WPSTR$) metric that capture protection to the primary user and fraction of whitespace opportunities recovered by the secondary user respectively. The main idea in this formulation is the asymmetric nature of the two performance metrics. The F_{HI} metric is computed as a worst case over uncertainty models for the various statistical quantities like noise and multipath and shadow fading. On the other hand, the $WPSTR$ metric is computed as an average using probabilistic models for the statistical quantities involved in problem. The reason for this is that the primary user does not trust the models used by the secondary and hence we need a worst case metric for protection, whereas the secondary is free to use any probabilistic model for measuring its own performance[12].

The key advantages of using the new space-time sensing framework are:

- This framework gives a uniform way to compare the performance of different sensing algorithms. For instance, it can be used to compare single-user sensing algorithms, collaborative sensing algorithms, and multi-band sensing algorithms [62; 63; 64]. Furthermore, this framework quantifies the intuition that single-radio sensing algorithms have poor performance and one needs collaborative sensing algorithms to improve the whitespace recovery performance.

- This framework brings out the tradeoff between the performance in space versus the performance in time, which is otherwise not at all obvious to see. In Chapter 5, we show that in some cases the $WPSTR$ performance is *not* monotonic with the sensing time N. Increasing the sensing time improves the detector performance, but it also means that the secondary user has a smaller fraction of time in which it can use the spectrum hole.

[12] After all, an engineer designing a secondary system has no incentive to lie to himself, unlike his incentive to lie to his investors!

- Finally, this framework also brings out the analogous SNR wall story in the spatial domain. We show that in the presence of uncertainties, there is a finite F_{HI} (protection) threshold below which the performance of the cognitive radio goes to zero. That is to protect the primary below this threshold the secondary has to give up all the spectral opportunities.

Detection algorithm	*Description of algorithm*	*What is modeled?*	*What to gain?*
Energy detection [65] [66]	Get empirical estimate of energy in a frequency band and compare against a detection threshold.	Average power	Baseline detector for comparison.
FFT for DTV pilot signal [67; 68; 69]	Partial coherent detection using Filter around pilot to reduce noise power. Use FFT as partial coherent detector for sinusoids.	DTV pilot. Signal contains narrowband pilot tone	Sensing time and robustness
Run-time noise calibrated detection [70]	Noise is calibrated during run-time leading to robustness gains.	Asymmetric use of degrees of freedom	Robustness
Cyclostationary detection [59] [71] [72]	Spectral correlation function reveals peaks at multiples of the modulation rate/pilot frequency.	Signal is modeled as wide-sense cyclostationary	Robustness
Dual FPLL pilot sensing [73]	Use two Digital PLLs which are preset to $\pm$30kHz around the pilot. Use time to converge as test statistic.	Signal contains narrowband pilot tone	Simplicity of implementation
Eigenvalue based detection [74; 75]	Utilizes the fact that white noise is uncorrelated across samples/antennas while a band-limited external signal is correlated	Band-limited primary signal and secondary radio has multiple receive antennas	Sensing time
Event-based detection [76; 77]	The detector tries to detect arrival/departure of signals. This technique can be used for identifying time-domain holes.	Primary user ON/OFF durations are much shorter than the time between secondary user movement.	Robustness

Table 1.1: Comparison of representative single-user sensing algorithms for DTV detection. These algorithms use various facets of the transmitted signal to obtain a better detection sensitivity over simple energy detection.

Chapter 2

SNR walls for detection

2.1 Introduction

As discussed in Chapter 1, the goal is to design a secondary sensor that meets a given target probability of false-alarm (pfa) and probability of missed-detection (pmd) constraints at low SNRs (high sensitivity). Classical detection theory suggests that degradation in the pfa and pmd due to reduced SNR can be countered by increasing the sensing time [47], [50]. Hence, the achievable sensitivity is considered to be limited by the maximum possible sensing time, which depends on higher layer design considerations. For instance, the QoS requirements (delay constraints) of the application drive the protocol layer design, which in turn dictates the time available for sensing by the physical layer. This traditional perspective implies that a cognitive radio system can always be engineered at the cost of low enough QoS.

However, it is impossible to model parameters in real-world physical systems with infinite precision. For example, real-world background "noise" is neither perfectly Gaussian, perfectly white, nor perfectly stationary. Fluctuations in the thermal noise floor were observed even in a calibrated[1] experimental setup [78]. The channel fading is neither flat nor is it constant over time. Real-world filters are not ideal, A/D converters have finite precision, I and Q signal pathways in a receiver are never perfectly matched and local oscillators are never perfect sine-waves. In this chapter[2], we argue that *these model uncertainties impose fundamental limitations on detection performance. The limitations cannot be countered by increasing the sensing duration.* At very low SNRs, the ergodic view of the world is no longer valid, i.e., one cannot count on infinite averaging to combat the relevant uncertainties.

To illustrate the impact of model uncertainties, consider a simple thought experiment. Imagine a detector that computes a test-statistic and compares it to a threshold to decide if the primary is present/absent. The threshold is set so the target false alarm probability is met

[1] Variation in the noise power due to temperature changes was accounted for in the experiment. Furthermore, the experiment was setup such that there was no in-band interference.

[2] The discussion in this chapter is based on out results in [79; 80; 66].

using the nominal model. Now, this detector is deployed in physically distinct scenarios with their own slight variations in local noise and fading characteristics. The actual performance can deviate significantly from the prediction. An example using the radiometer is illustrated in Figure 2.1. In fact, below a certain SNR threshold, at least one of the error probabilities can become worse than $\frac{1}{2}$. We call this sort of failure a lack of *robustness* in the detector. The nominal SNR threshold below which this phenomenon manifests is called the SNR *wall* for the detector.

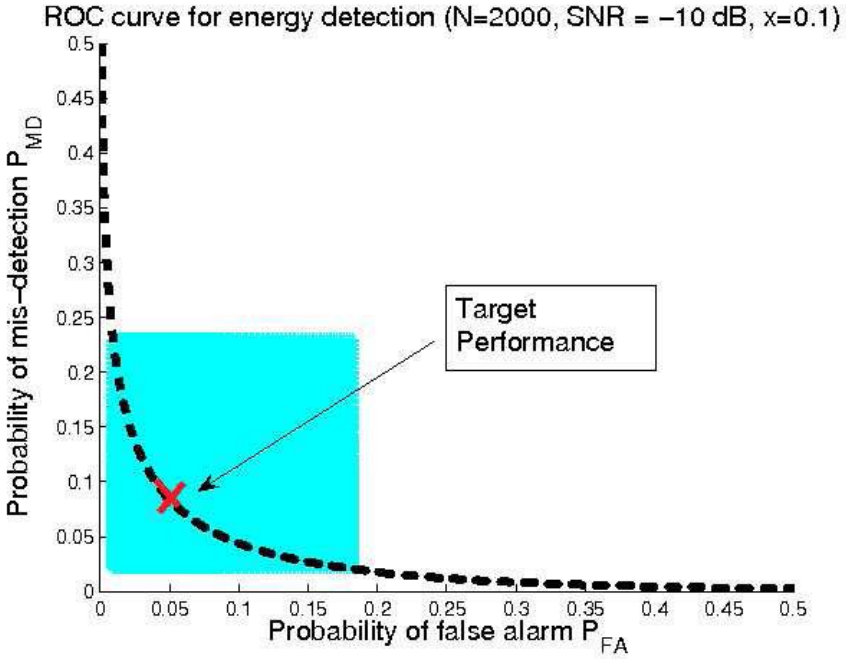


Figure 2.1: Error probabilities of radiometric detection under noise uncertainty. The point marked by an 'X' is the target theoretical performance under a nominal SNR of -10 dB. The shaded area illustrates the range of performance achieved with actual SNR's varying between -9.9 dB and -10.1 dB ($x = .1$ in Definition 2.1).

Such robustness limits were first shown in the context of radiometric (energy) detection of spread spectrum signals [81]. Robustness limits to radiometric detection in the context of cognitive radios were first shown in [82]. Subsequently, these limitations were also verified by controlled experiments that were carefully calibrated to limit uncertainties [83].

The outline of the chapter is as follows. The spectrum sensing problem in presence of modeling uncertainties is formulated as a *robust hypothesis test* in Section 2.2. We review the robustness performance of a radiometer in Section 2.3. We show that uncertainty in the noise power is sufficient to impose severe limitations on the performance of a radiometer. In

particular, we show that there exists an SNR threshold (called an SNR wall) below which robust detection is impossible for the radiometer. To make progress for general classes of signals and detection algorithms, we distill the relevant uncertainties into tractable mathematical models in Section 2.4.1. These models enable us to focus on those uncertainties that are most critical. Using these models we show in Section 2.4 that even the optimal detector suffers from SNR wall limitations when the 'signal of interest' does not contain any known feature. In particular, we show that the robustness performance of an optimal non-coherent detector is as bad as the radiometer. The existence of *absolute SNR walls* for non-coherent detection was first proved in my Masters thesis [80]. However, in my Masters thesis we only conjectured that the lower bound on the location of the absolute SNR wall was as shown in (2.10). The proof for this conjecture is given in this chapter in Theorem 2.2. We conclude this chapter with the policy implications of the SNR wall result in Section 2.5. Our results imply that if the primary user has complete flexibility of use, then the secondary user cannot reliably discover whitespaces. Hence, the regulators must impose some restrictions on the primary user in order to allow the secondary user to opportunistically recover spectrum holes.

2.2 Robust hypothesis test formulation

Let $X(t)$ denote the band-limited signal we are trying to sense and let the additive noise process be $W(t)$. The discrete-time version is obtained by sampling the received signal at the appropriate rate. In this chapter we ignore the fading in the wireless channel between the primary transmitter and the secondary detector. The idea is to isolate the impact of uncertainty in the noise process from the uncertainty in the fading process. Our goal is to incorporate real-life uncertainties in noise into our hypothesis testing framework, and quantify the impact of these uncertainties on the performance of sensing algorithms. This is done using the robust hypothesis testing framework proposed by Huber [84]. The two hypotheses are:

$$\begin{aligned} \mathcal{H}_0 : Y[n] &= W[n]; & W[n] \sim W \in \mathcal{W} \\ \mathcal{H}_1 : Y[n] &= \sqrt{P} \cdot X[n] + W[n]; & X[n] \sim X \in \mathcal{X}, W[n] \sim W \in \mathcal{W} \end{aligned} \tag{2.1}$$

Here P is the received signal power, $X[n]$ are the unattenuated samples (normalized to have unit power) of the primary signal, $W[n]$ are noise samples and $Y[n]$ are the received signal samples. We assume that $X[n]$ is an *independent and identically* (*iid*) distributed random process, $W[n]$ is an *iid* random process and that the signal ($X[n]$) is independent of the noise process ($W[n]$). We model uncertainty in the noise process by assuming that the underlying noise distribution lies in some neighborhood of an idealized noise distribution. That is, assume that the underlying noise distribution $W \in \mathcal{W}$, where $\mathcal{W}$ is a class of

distributions containing the nominal noise distribution W_n. For simplicity, we assume that the nominal noise distribution is a Gaussian, $W_n \sim \mathcal{N}(0, \sigma_n^2)$. Define the signal to noise ratio to be $SNR := \frac{P}{\sigma_n^2}$. As the primary user can have freedom to change its signaling scheme, and the secondary sensor might not have precise knowledge of the primary signaling scheme, we assume that the signal distribution can lie anywhere in a class of distributions, i.e., $X \in \mathcal{X}$.

In this thesis we focus on threshold based detection rules. For a given test-statistic based detection rule $T(\mathbf{Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda$, where $\mathbf{Y} := (Y[1], Y[2], \cdots, Y[N])$ and λ is the detection threshold, the probability of false alarm and the probability of missed-detection are defined as (this is the *min-max* formulation introduced in [85])

$$\begin{aligned} P_{FA}(N, \lambda) &:= \sup_{W \in \mathcal{W}} \mathcal{P}_W \left(T(\mathbf{Y}) > \lambda | \mathcal{H}_0 \right), \\ P_{MD}(N, \lambda) &:= \sup_{W \in \mathcal{W}} \mathcal{P}_{W,X} \left(T(\mathbf{Y}) < \lambda | \mathcal{H}_1 \right). \end{aligned} \tag{2.2}$$

Furthermore, eliminating λ from (2.2) we can solve for N as a function of the SNR (sensitivity), P_{FA}, and P_{MD}, and the uncertainty set $\mathcal{W}$. Hence, we can write

$$N = \xi(SNR, P_{FA}, P_{MD}, \mathcal{W}). \tag{2.3}$$

This is called the *sample complexity* of the detector. The traditional metrics triad of sensitivity, P_{FA}, and P_{MD}, are used along with the sample complexity to evaluate the performance of detection algorithms. For reasonable detectors, $\xi(SNR, P_{FA}, P_{MD}, \mathcal{W})$ is a monotonically decreasing function of SNR, P_{FA} and P_{MD}. For a given SNR, the aim is to design detectors that minimize the sample complexity, while achieving the target P_{FA} and P_{MD}.

2.3 Impact of uncertainty: radiometer example

In this Section we review the performance of a simple as well as very commonly used detector, namely the radiometer [48] under uncertain noise distributions. The advantage of the radiometer is that it does not need any knowledge of the signal distribution, and all it needs to know is the average power of the signal. As the radiometer only sees the power, the distributional uncertainty of noise can be summarized by uncertainty in its variance, i.e., we assume that the noise variance $\sigma^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]$ where σ_n^2 is the nominal noise power and $\rho > 1$ is a parameter that quantifies the size of the uncertainty. The impact of noise-power uncertainty on a radiometer was first quantified in [81].

2.3.1 Radiometer robustness

The test-statistic for the radiometer is given by $T(\mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N}|Y[n]|^2$. For a given noise variance σ^2, the central limit theorem (see [86]) gives the following approximations [47]:

$$\begin{aligned} T(\mathbf{Y})|\mathcal{H}_0 &\sim \mathcal{N}(\sigma^2, \frac{1}{N}2\sigma^4), \\ T(\mathbf{Y})|\mathcal{H}_1 &\sim \mathcal{N}(P+\sigma^2, \frac{1}{N}2(P+\sigma^2)^2), \end{aligned} \tag{2.4}$$

Using these approximations

$$\begin{aligned} P_{FA}(N,\lambda) &= \sup_{\sigma^2\in[\frac{1}{\rho}\sigma_n^2,\rho\sigma_n^2]} Prob\left(T(\mathbf{Y})>\lambda|\mathcal{H}_0,\sigma^2\right) \\ &= \sup_{\sigma^2\in[\frac{1}{\rho}\sigma_n^2,\rho\sigma_n^2]} \mathcal{Q}\left(\frac{\lambda-\sigma^2}{\sqrt{\frac{2}{N}}\sigma^2}\right) \\ &= \sup_{\sigma^2\in[\frac{1}{\rho}\sigma_n^2,\rho\sigma_n^2]} \mathcal{Q}\left(\frac{\frac{\lambda}{\sigma^2}-1}{\sqrt{\frac{2}{N}}}\right) \\ &\overset{(a)}{=} \mathcal{Q}\left(\frac{\lambda-\rho\sigma_n^2}{\sqrt{\frac{2}{N}}\rho\sigma_n^2}\right), \end{aligned} \tag{2.5}$$

where λ is the detector threshold and $\mathcal{Q}(\cdot)$ is the standard Gaussian complementary CDF. The equality in (a) follows from the observation that $\mathcal{Q}(\cdot)$ is a monotonically decreasing function and $\frac{\lambda}{\sigma^2}$ is minimized at $\sigma^2=\rho\sigma_n^2$.

Similarly,

$$\begin{aligned} P_{MD}(N,\lambda) &= 1-\inf_{\sigma^2\in[\frac{1}{\rho}\sigma_n^2,\rho\sigma_n^2]} \mathcal{Q}\left(\frac{\lambda-(P+\sigma^2)}{\sqrt{\frac{2}{N}}(P+\sigma^2)}\right) \\ &= 1-\mathcal{Q}\left(\frac{\lambda-(P+\frac{1}{\rho}\sigma_n^2)}{\sqrt{\frac{2}{N}}(P+\frac{1}{\rho}\sigma_n^2)}\right). \end{aligned} \tag{2.6}$$

The probability of false-alarm and probability of missed-detection targets ($P_{FA}(N,\lambda)\leq$ pfa and $P_{MD}(N,\lambda)\leq$ pmd for some $0\leq \text{pfa},\text{pmd}\leq 1$) can be achieved by choosing the

sensing time

$$N \approx \frac{2[\mathcal{Q}^{-1}(\text{pfa}) - \mathcal{Q}^{-1}(1-\text{pmd})]^2}{\left[SNR - \left(\rho - \frac{1}{\rho}\right)\right]^2}. \tag{2.7}$$

This is the sample complexity of the radiometer under uncertain noise-power[3]. The above sample-complexity expression was obtained by eliminating λ from (2.5) and (2.6), and by approximating $1 + SNR \approx 1$.

From the sample-complexity expression in (2.7) it is clear that $N \to \infty$ as $SNR \downarrow \left(\rho - \frac{1}{\rho}\right)$. This suggests that robust detection is impossible for $SNR \leq \left(\rho - \frac{1}{\rho}\right)$. We now formalize this observation.

Definition 2.1. Consider the robust hypothesis testing problem in (2.1). Fix an $SNR > 0$, and a detector characterized by the test-statistic $T(\cdot)$. The detector is said to be *non-robust* at the given SNR if for any $0 \leq \text{pfa} < 0.5$, $0 \leq \text{pmd} < 0.5$, we cannot choose an $N > 0$ and a detector threshold λ such that both the false-alarm and missed-detection constraints are met, i.e., $P_{FA}(N, \lambda) \leq \text{pfa}$ and $P_{MD}(N, \lambda) \leq \text{pmd}$ cannot simultaneously be satisfied for any choice of (N, λ).

Definition 2.2. Consider the robust hypothesis testing problem in (2.1). For a given detector we define an "SNR wall" to be the maximum signal to noise ratio (SNR) threshold below which the detector is non-robust.

Theorem 2.1. *Consider the robust hypothesis testing problem in (2.1) with uncertain noise variance, i.e., $\sigma^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]$. Then, there exists an SNR threshold below which the radiometer is non-robust to uncertainty in the noise power. Furthermore, the "SNR wall" for the radiometer is given by*

$$SNR_{wall}^{energy} = \frac{\rho^2 - 1}{\rho}. \tag{2.8}$$

Proof. Fix an $0 \leq SNR \leq \frac{\rho^2-1}{\rho}$. Suppose, if possible assume that the radiometer is robust at the given SNR. Then, from Definition 2.1, there exists a $0 \leq pfa < 0.5$, and $0 \leq pmd < 0.5$, such that we can choose a $N > 0$ and λ such that $P_{FA} \leq pfa$ and $P_{MD} \leq pmd$.

From (2.5), it is clear that $P_{FA} \leq pfa$ implies $\lambda > \rho\sigma_n^2$. This follows from the choice of $pfa \leq 0.5$. Similarly, $P_{MD} \leq pmd$ implies $\lambda < P + \frac{1}{\rho}\sigma_n^2$ (see (2.6)). The above inequalities

[3]Substituting $\rho = 1$ in (2.7), which corresponds to the no uncertainty case, gives us the expression for the sample complexity of a radiometer with perfectly known noise variance [48], [87], [80]

give

$$\begin{aligned} \rho\sigma_n^2 < \lambda &< P + \frac{1}{\rho}\sigma_n^2 \\ \Rightarrow \rho\sigma_n^2 &< P + \frac{1}{\rho}\sigma_n^2 \\ \Rightarrow \frac{P}{\sigma_n^2} &> \left(\rho - \frac{1}{\rho}\right) \\ \Rightarrow SNR &> \frac{\rho^2 - 1}{\rho}, \end{aligned} \tag{2.9}$$

which is a contradiction. So, the radiometer is non-robust for all $0 \leq SNR \leq \frac{\rho^2-1}{\rho}$. □

Remarks:

- The above result shows that the radiometer is highly non-robust to uncertainty in the noise power. This is true even when the signal distribution X is known, i.e., where $\mathcal{X}$ is a singleton.

- The key intuition in traditional binary hypothesis testing is that as we increase the number of samples (N) to infinity, the distribution of the empirical test-statistic concentrates around the corresponding mean for each hypothesis [50]. So, if the test-statistic means under both hypotheses are distinct, the two hypotheses can be distinguished with arbitrarily low error probabilities.

  However, the situation is different under noise uncertainty. For instance, the test-statistic means for the radiometer with noise uncertainty are given by

$$\begin{aligned} \mathbb{E}T(\mathbf{Y}|\mathcal{H}_0) &\in \left[\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2\right], \\ \mathbb{E}T(\mathbf{Y}|\mathcal{H}_1) &\in \left[P + \frac{1}{\rho}\sigma_n^2, P + \rho\sigma_n^2\right]. \end{aligned}$$

  The test-statistic means under each hypothesis can take any arbitrary value in the corresponding interval. Using the same intuition as before, the two hypotheses can be robustly distinguished if the intervals for the test-statistic means are disjoint. For the radiometer this happen if $\rho\sigma_n^2 < P + \frac{1}{\rho}\sigma_n^2$, which is same as the condition in (2.9).

Figure 2.2 illustrates the overlap of the test-statistic means for $SNR \leq \frac{\rho^2-1}{\rho}$. Figure 2.3(a) plots the location of the SNR wall for a radiometer (see (2.8)) as a function of the noise uncertainty parameter $x = 10\log_{10}\rho$, as expressed in dB terms. Extensive simulations showing the limits on radiometer performance under noise uncertainty have been reported

in [88]. Our theoretical SNR wall limits were also experimentally validated in [78]. The sample-complexity of a radiometer under noise uncertainty is illustrated in Figure 2.3(b). From the figure, we can see that the sample complexity approaches infinity as the SNR approaches the radiometer SNR wall. From the figure it is clear why we call it an "SNR Wall".

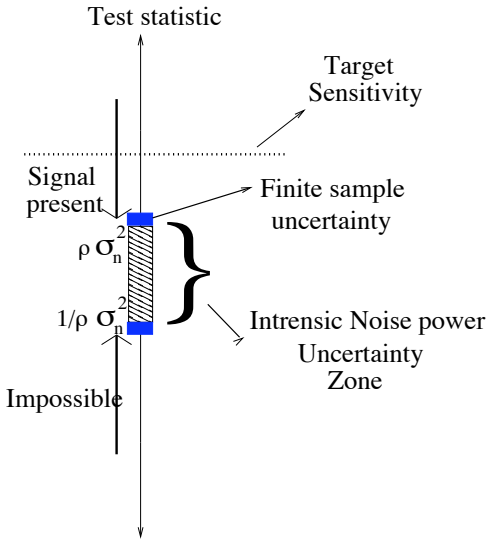
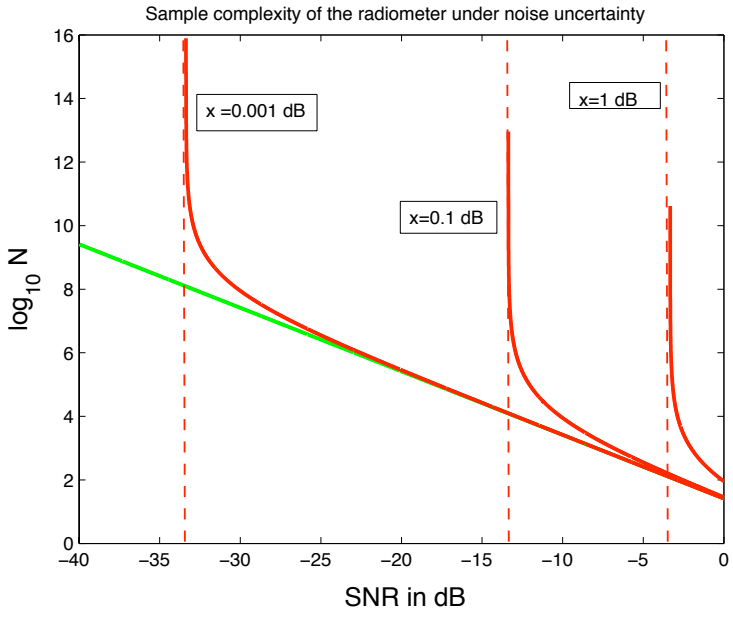


Figure 2.2: Understanding noise uncertainty for a radiometer. The shaded area in the figure represents the uncertainty in the noise power. It is clear that if the test statistic falls within the shaded region, there is no way to distinguish between the two hypotheses.

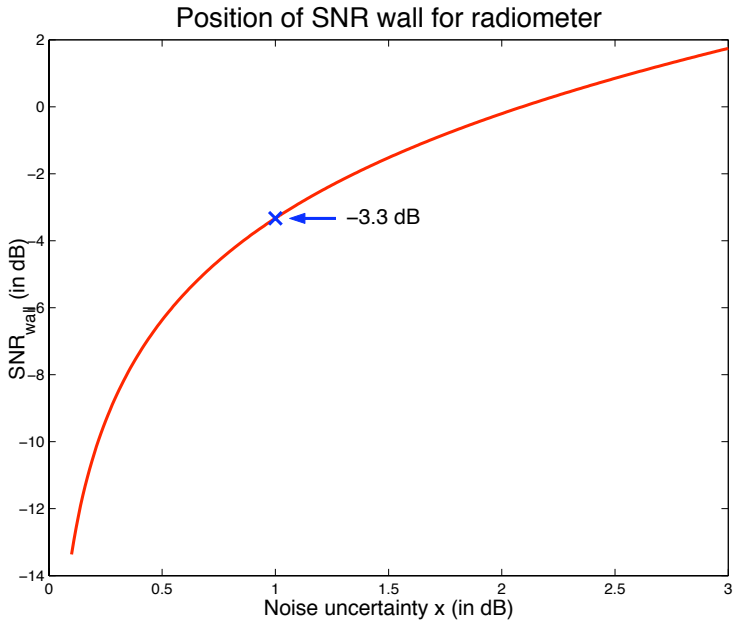
2.3.2 ROC-based Intuition of the SNR wall result

So what happens on the other side of the wall? An understanding can be obtained by looking at a detector's *Receiver Operating Characteristic* (ROC) curves. The ROC curve for a detector is the tradeoff between the probability of false-alarm and the probability of missed-detection for a fixed SNR and sensing time N [47]. Figure 2.4 plots the ROC curves with/without (solid/dashed) noise uncertainty for a radiometer.

Consider the case without noise uncertainty. Every point on the ROC curve corresponds to a particular choice of the detector threshold λ. From the dashed curves in Figure 2.4 we can clearly observe the performance of the detector as a function of the SNR and the sensing time N. Firstly, as we increase the SNR (top left plot to the top right plot in Figure 2.4) the ROC curves shift towards the $(0, 0)$ corner. This shows that the detector performance can be improved by increasing the SNR for a fixed sensing time N. The reason for this is



(a)



(b)

Figure 2.3: Figure (a) shows how the sample complexity N varies for the radiometer as the SNR approaches the SNR wall (See Equation (2.7) and use $x = 10\log_{10}\rho$). Figure (b) plots the SNR_{wall}^{energy} in (2.8) as a function of noise level uncertainty $x = 10\log_{10}\rho$. The point marked by a "x" on the plot corresponds to x=1 dB of noise power uncertainty.

that as the SNR increases the separation between the test-statistic means under both the hypotheses increases[4]. Secondly, for a fixed SNR the radiometer performance improves as N increases (the ROC curve shifts towards the $(0, 0)$ corner). The reason for this is that as N increases the distribution of the test-statistic concentrates around its mean due to ergodic averaging [89] [51].

Now, consider the performance of a radiometer with noise uncertainty. As before, an ROC curve is defined as the tradeoff between the probability of false-alarm and the probability of missed-detection obtained by varying the detector threshold λ, while keeping the SNR and sensing time N fixed. However, with noise uncertainty there is a subtle difference in the interpretation of an ROC curve. Recall the definition of the probability of false-alarm and the probability of missed-detection under noise uncertainty given in (2.2). From (2.2) it is clear that under noise uncertainty the error probabilities P_{FA} and P_{MD} are computed using the worst-case distribution from the uncertainty set $\mathcal{W}$. Furthermore, the worst-case distribution from a false-alarm point of view can be different than the worst-case distribution from a missed-detection point of view. In particular, for a radiometer with uncertain noise power, every point on the ROC curve corresponds to two different choices of the underlying noise power[5]. This means that for every point (pfa, pmd) on the ROC curve, the actual probability of false-alarm is upper bounded by pfa and the actual probability of missed-detection is upper bounded by pmd. As there is no way of knowing the actual noise power we use the upper bound (pfa, pmd) as a measure of detector performance.

The solid curves in the top right plot in Figure 2.4 show the ROC curves for a radiometer with $x = 1$ dB of uncertainty, and a nominal SNR of -2 dB (above the radiometer SNR wall). The ROC curves for each N are worse (shift away from the $(0, 0)$ corner) than the corresponding curves without uncertainty (dashed curves in the plot). The intuition for this is given in the bottom right plot in Figure 2.4 which shows the set of possible test-statistics means under both hypotheses. The plot also shows the two least-favorable distributions. The difference in means of the two least-favorable distributions is smaller than the difference in means of the nominal distributions. So, the performance with noise uncertainty is degraded as compared to the performance without noise uncertainty. In this case, the performance degradation due to noise uncertainty can be compensated for by increasing the number of samples. The sets of test-statistic means under both hypotheses do not overlap and hence ergodic averaging is helpful. However, if the SNR is below the SNR wall, then the sets

[4]In binary hypothesis testing, the performance of a detector is a function of the Kullback-Leibler (KL) divergence between the two distributions corresponding to the two hypotheses. In particular, the KL divergence between two Gaussian distributions is a monotonically increasing function of the difference of means of the Gaussian distributions.

[5]These two distributions are called the *least-favorable* distributions in robust hypothesis testing literature [85]. The standard likelihood-ratio test between these two least-favorable distributions gives us an optimal min-max robust hypothesis test for the original problem [85]. The bottom plots in Figure 2.4 shows the set of possible test-statistic distributions under both hypotheses for a radiometer. The curves shown in bold correspond to the probability density function of the least-favorable distributions under each hypothesis.

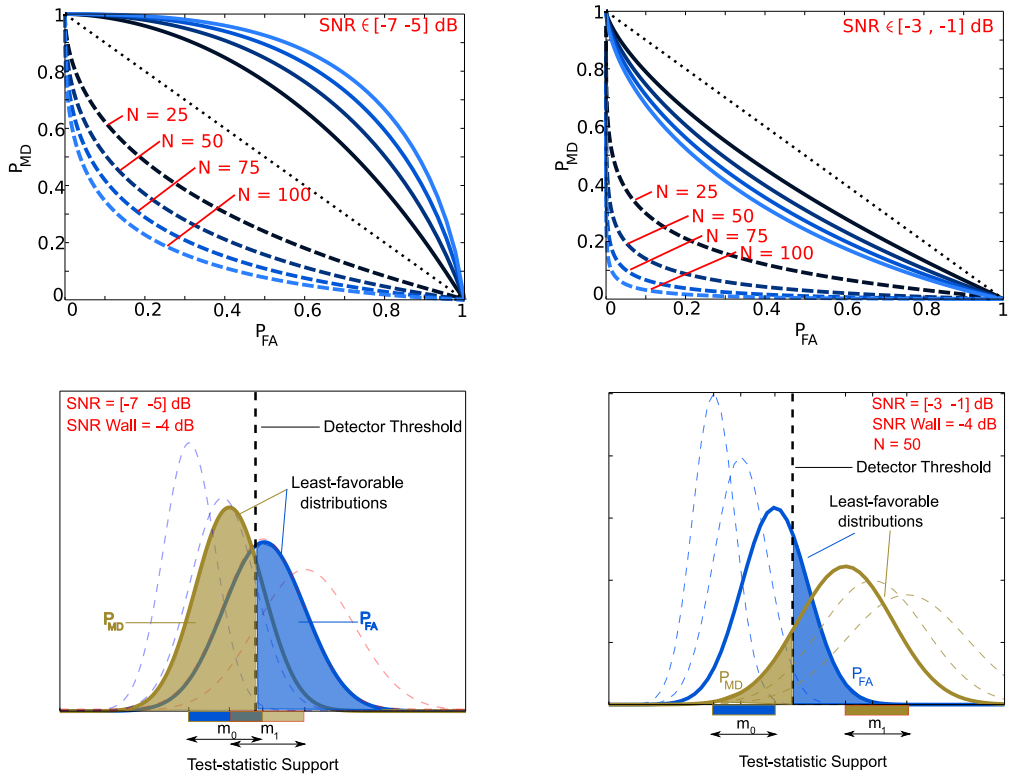


Figure 2.4: The solid ROC curves correspond to the case with noise uncertainty and the dotted ROC curves correspond to the case without noise uncertainty. The plots on the right correspond to the case where the operating SNR is above the SNR wall. The plots on the left correspond to the case when the operating SNR is below the SNR wall.

of test-statistic means under both hypotheses overlap as shown in the bottom left plot in Figure 2.4. The ROC curves in this case are given by the solid curves in the top left plot in Figure 2.4. The ROC curves can be worse than those of the random coin-tossing detector! (dotted straight line in Figure 2.4).

2.4 Optimal non-coherent detection

The results in Section 2.3.1 tell us that the radiometer is non-robust to uncertainty in the variance of the noise process. Moreover, the SNR wall for the radiometer is close to 0 dB for moderate values ($x \approx 1$ dB) of the noise uncertainty. This suggests that secondary users cannot use a radiometer to sense weak primary signals. The obvious question now is whether there are other detectors whose robustness performance is significantly better than that of

the radiometer. In order to answer this question, we need to work with a generalized noise uncertainty model.

2.4.1 White-noise uncertainty model

For the radiometer it was sufficient to model the uncertainty in the noise power. This model was backed up by the experimental observation that the thermal noise power fluctuates in an non-ergodic fashion [78]. Since the noise power is uncertain it is reasonable to assume that the other moments of the noise process will also be uncertain. So, we propose a bounded moment noise uncertainty model that captures the idea of approximate Gaussianity. See Figure 2.5 for a pictorial description of the noise uncertainty set.

Definition 2.3. Define $\mathcal{W}_x$, the class of uncertain noise distributions to be the set of all distributions W that satisfy:

- The noise process is symmetric $\mathbb{E}W^{2k-1} = 0, \forall k = 1, 2, \ldots$.
- Even moments of the noise must be close to the nominal noise moments in that $\mathbb{E}W^{2k} \in [\frac{1}{\rho^k}\mathbb{E}W_n^{2k}, \rho^k\, \mathbb{E}W_n^{2k}]$, where $W_n \sim \mathcal{N}(0, \sigma_n^2)$ is a nominal Gaussian noise random variable and $\rho = 10^{x/10} > 1$.

The parameter x in Definition 2.3 is used to quantify the amount of uncertainty in the noise power, i.e., we allow for $\pm x$ dB of uncertainty in the noise variance. The above model is reasonable for the following reasons:

- If the nominal noise distribution is believed to be $W_n \sim \mathcal{N}(0, \sigma_n^2)$. Then, any reasonable uncertainty model must include Gaussian distributions with variance slightly different than σ_n^2, i.e., any Gaussian with variance $\sigma^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]$, for some $\rho > 1$.
- Background noise also includes weak signals from interferers, quantization noise, and other random disturbances. This means that assuming perfect Gaussianity[6] is unrealistic and the noise uncertainty set should include other distributions that are close to Gaussian. For mathematical tractability, closeness to Gaussian distributions is expressed in terms of the moments. It is clear that similar results hold if the noise uncertainty set is modeled as a small ball (in the variational distance) around the nominal Gaussian CDF for W_n.

[6] In information theory, it turns out that the Gaussian distribution is a saddle point for many interesting optimization problems. Therefore, we are sometimes safe in ignoring the distributional uncertainty and just overbounding with a worst case Gaussian. However, this turns out not to be true in our problem of spectrum sensing.

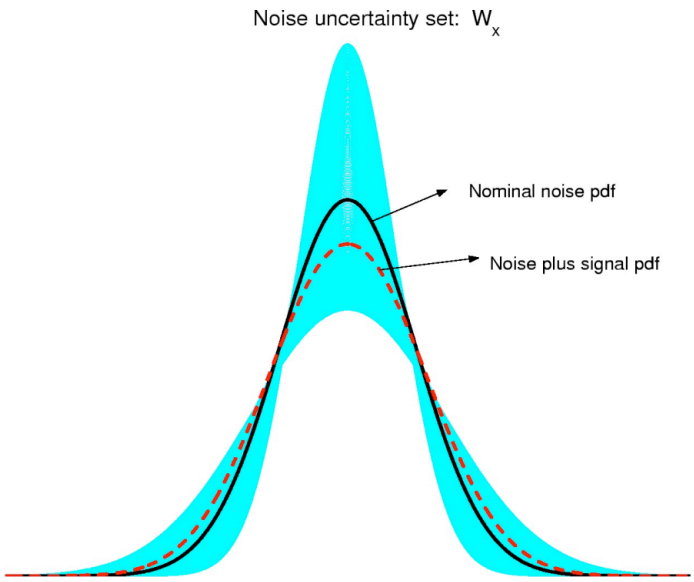


Figure 2.5: An example of a noise uncertainty set $\mathcal{W}_x$, for $x = 3dB$. The shaded region is a set of possible pdfs for the noise random variable. The bold curve inside the region is the pdf of a nominal Gaussian, W_n. The dashed curve is the pdf of $X + W_n$ for BPSK X at $SNR = -6dB$. This illustrates that at low signal powers, it is impossible to robustly distinguish the signal from noise.

2.4.2 Absolute SNR walls

Given the mathematical model for distributional uncertainty in the noise process, the key questions we ask is: do all detectors suffer from SNR wall limits? If so, how robust is the optimal detector as compared to the radiometer? One approach to answer these questions is to find the optimal detection rule and then compute the SNR wall for that detector. The problem with this approach is that finding optimal detection rules for *min-max* robust hypothesis testing problems is hard[7]. So, we take an alternative approach —we show the existence of SNR thresholds below which *all possible* detection algorithms will be non-robust. We call such thresholds *absolute SNR walls*. Having shown the existence of absolute SNR walls, we try to give upper and lower bounds on the locations of these walls.

[7]The only known approach, proposed by Huber [85], is to identify two distributions, one from each hypothesis such that the likelihood ratio test between these two distributions is the optimal robust test for the original min-max robust hypothesis test. These distributions are called a *least-favorable* pair of distributions. Huber gives sufficient conditions for a pair of distributions to be a least-favorable pair of distributions [85]. However, there is no technique to construct a least-favorable pair of distributions. The only option is to guess a pair of distributions and test whether they satisfy Huber's conditions.

We now give a sufficient condition for the existence of absolute SNR walls. Consider the robust hypothesis testing problem in (2.1). As the received signal samples $\mathbf{Y}$ are *iid* under both hypotheses, the distribution of $\mathbf{Y}$ is characterized by its marginal distribution. Let $Y[n] \sim Y^{(i)} \in \mathcal{Y}^{(i)}$ for $i \in \{0,1\}$, where i denotes the underlying hypothesis. Here $\mathcal{Y}^{(i)}$ denotes the set of possible distributions for the received signal samples under hypothesis $\mathcal{H}_i$, $i = 0, 1$. Suppose $\mathcal{Y}^{(0)} \bigcap \mathcal{Y}^{(1)} \neq \Phi$, then it is clear that no possible detection algorithm can robustly distinguish between the two hypotheses.

Assume that the distributional uncertainty in noise is as given in Definition 2.3. That is, $\mathcal{Y}^{(0)} = \mathcal{W}_x$ for some $x > 0$. Let the distribution of the signal samples $X[n] \sim X$, with X being known. Then, $\mathcal{Y}^{(1)} = \{X + W | W \in \mathcal{W}_x\}$. In particular, $X + W_n \in \mathcal{Y}^{(1)}$, where $W_n \sim \mathcal{N}(0, \sigma_n^2)$ is the nominal noise distribution. Intuitively, it is clear that if the signal power P is sufficiently low, then by continuity the distribution of $X + W_n$ is close to the distribution of W_n and hence can lie in $\mathcal{Y}^{(0)}$, in which case robust detection is impossible for any detection algorithm.

To get a better understanding of absolute SNR walls and the conditions under which these walls exist, we look at the quantized version of the sensing problem. Assume that the secondary sensor has access only to quantized versions of the received signal samples. This is a reasonable assumption because in reality every receiver chain has an analog-to-digital converter that samples and quantizes the received signals. We assume that the received signal vector $\mathbf{Y}$ is quantized by passing each sample $Y[n]$ through a symmetric scalar $2K$-bin quantizer. Let the quantization points be denoted by $a_{-(K-1)}, a_{-(K-2)}, \cdots a_0, a_1, a_2, \cdots, a_{K-1}$ with $a_0 = 0$. This quantizer is symmetric if the quantization points are symmetric around the origin, i.e., $a_{-i} = a_i$, for all $i = 1, 2, \cdots, K-1$. Figure 2.6 is an example of a symmetric scalar 6-bin quantizer. Let $\mathbf{Y}_q$ denote the quantized received signal samples. As $\mathbf{Y}$ is *iid*, $\mathbf{Y}_q$ will also be *iid*. So, the distribution of $\mathbf{Y}_q$ is uniquely characterized by its marginal distribution. Let $Y_q[n] \sim Y_q^{(i)} \in \mathcal{Y}_q^{(i)}$ for $i \in \{0,1\}$, where i denotes the underlying hypothesis. Here $\mathcal{Y}_q^{(i)}$ is used to denote the set of possible distributions of the quantized signal under hypothesis $\mathcal{H}_i$. Note that $Y_q^{(i)}$ is a discrete random variable with a support set of size $2K$. As the quantizer is symmetric and since the received signal samples $Y[n]$ are symmetric around 0, the probability mass function of $Y_q^{(i)}$ is completely characterized by the probabilities of the K non-negative bins. Denote these probabilities by $p_0^{(i)}, p_1^{(i)}, \cdots, p_K^{(i)}$ (see Figure 2.6). Note that these probabilities must satisfy $\sum_{j=0}^{K} p_j^{(i)} = \frac{1}{2}$ for $i \in \{0,1\}$.

Figure 2.7 illustrates the set of distributions in $\mathcal{Y}_q^{(0)}$ and $\mathcal{Y}_q^{(1)}$ for the case when the received signal samples $Y[n]$ are passed through a 6-bin symmetric quantizer shown in Figure 2.6. Each distribution in the set $\mathcal{Y}_q^{(i)}$ for $i \in \{0,1\}$ is characterized by a probability vector $(p_0^{(i)}, p_1^{(i)}, p_2^{(i)})$ ($K = 3$ in this example) with $p_0^{(i)} + p_1^{(i)} + p_2^{(i)} = 0.5$. So, every probability distribution in $\mathcal{Y}_q^{(i)}$ can be uniquely represented by a point in the 2-dimensional space spanned by the two probabilities $p_0^{(i)}$ and $p_1^{(i)}$. In the Figure 2.7, as the signal power P is decreased

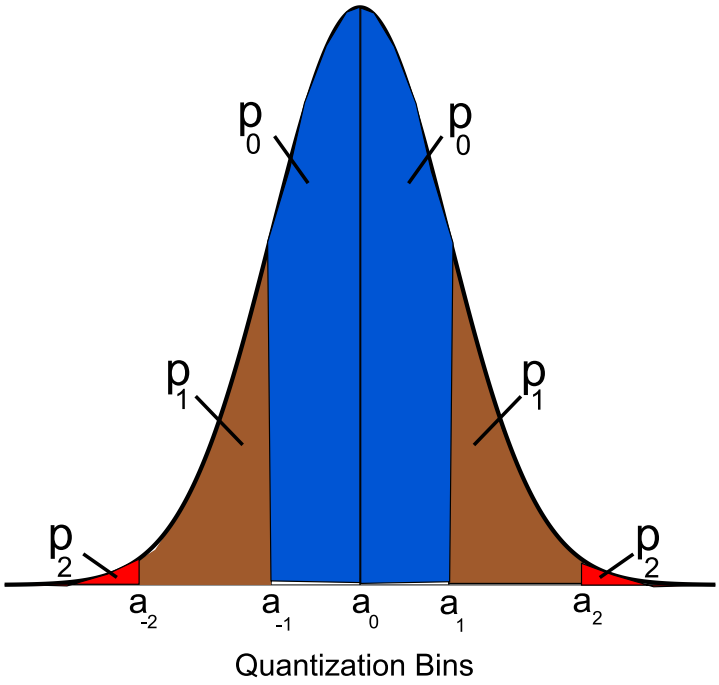


Figure 2.6: A six-bin symmetric scalar quantizer is shown in this figure. The quantizer is symmetric around the origin if $a_0 = 0$, and $a_{-2} = a_2$ and $a_{-1} = a_1$. Each bin is colored differently, while the bins symmetric around the origin are shown in the same color. In this example the probability mass function is uniquely characterized by the probability vector (p_0, p_1, p_2), which satisfy $p_0 + p_1 + p_2 = \frac{1}{2}$.

the distributional uncertainty sets $\mathcal{Y}_q^{(i)}$ get close to each other and eventually overlap when P is below a certain threshold. Robust detection is impossible for any detector (working with a 6-bin symmetric quantized signal) below this threshold, and hence this threshold is an absolute SNR wall.

The existence of absolute SNR walls can be shown for quantized signal with any number of quantization bins. Intuitively, the location of this absolute SNR wall is a function of the number of quantization bins, and it improves as the number of quantization bins increase. In the next section we show that there exists an absolute SNR wall even in the limiting case of infinite quantization bins, i.e., even when the detector has access to continuous valued signal samples $Y[n]$.

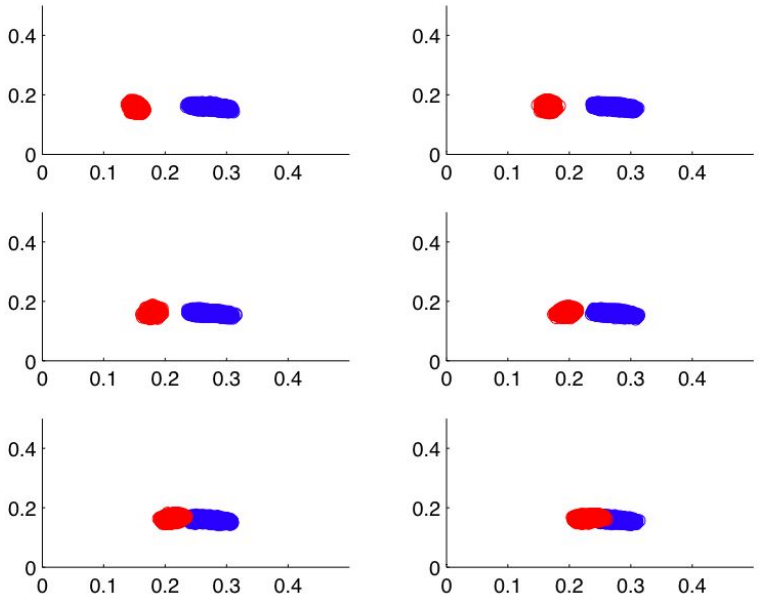


Figure 2.7: The distributional uncertainty sets $\mathcal{Y}_q^{(i)}$ for $i = 0, 1$ are shown for different values of the signal power P. The blue colored set corresponds to distributions under hypothesis $\mathcal{H}_0$ and the red colored set corresponds to distributions under hypothesis $\mathcal{H}_1$. It is clear from the figure that the distributional uncertainty sets overlap when the power P is sufficiently small, and hence this leads to an absolute SNR wall.

2.4.2.1 Main theorem

Theorem 2.2. *Consider the robust hypothesis testing problem defined in (2.1) with the noise uncertainty set ($\mathcal{W} = \mathcal{W}_x$, $x > 0$) described in Definition 2.3. Let $\rho = 10^{\frac{x}{10}}$. Also, assume that the primary signal model $\mathcal{X}$ includes signals $X[n]$ satisfying the following properties:*

1. *$X[n]$ are independent of the noise samples $W[\hat{n}]$, for all $n, \hat{n} > 0$.*
2. *$X[n]$ are independent and identically distributed as a random variable X.*
3. *X is zero-mean and all its odd moments are zero, i.e., $\mathbb{E}[X^{2k-1}] = 0$ for all $k = 1, 2, \cdots$*
4. *X has bounded support so $|X| \leq \bar{X}$.*

Define $SNR_{peak} = \frac{\bar{X}^2}{\sigma_n^2}$. Under these assumptions, there exists a threshold SNR_{wall}^* such that robust detection is impossible if $SNR_{peak} \leq SNR_{wall}^*$. Furthermore, there are easy to

compute bounds on this SNR wall:

$$\rho - 1 \leq SNR^{*}_{wall} \leq \frac{\rho^2 - 1}{\rho}. \tag{2.10}$$

Proof. See Appendix A.1 for the proof. □

The upper bound in (2.10) is the SNR wall for the radiometer from (2.8). Furthermore, the gap between the upper and lower bounds is at most a ratio of 2 when the uncertainty parameter ρ is very close to 1, and the ratio approaches 1 if the uncertainty factor ρ gets large. Thus, Theorem 2.2 asserts the existence of an absolute SNR wall that is essentially like that of the radiometer. This wall holds for white primary signals with bounded support in cases without fading.

2.5 Implication of SNR walls

The results in this chapter suggest that if the primary user has complete freedom to choose its signaling scheme (within a given power constraint), then it is impossible for the secondary sensor to guarantee robust detection at low SNRs. In other words, if the regulations for opportunistic spectrum access are written to provide complete flexibility to the primary system then the secondary system will fail to recover any spectral opportunities. So, even opportunistic spectrum sharing cannot reduce the regulatory overhead.

This leads to the natural policy question: how much loss in freedom should the primary incur in order to provide sufficient robustness to the secondary sensor? Restrictions on the primary user's freedom in the form of mandating 'known features' in its signal is explored in Chapter 3. We show that the presence of known signal features in the primary signal can significantly improve robustness for the secondary sensor. The tradeoff between loss in freedom for the primary user and gain in robustness for the secondary sensor is technically quantified in Chapter 4. The results in Chapters 3 and 4 tell us that achieving robustness to modeling uncertainties at low SNRs is extremely hard.

In fact, our results on robust detection motivated us to ask the question: do we really want to sense signals at such low SNRs? We answer this question in Chapter 5 by showing that it is not desirable to design detectors that sense at low SNRs as this forces the secondary users to give up a significant fraction of the spectral opportunities in space. The results in Chapter 5 make the discussion of robust detection of signal at low SNRs in Chapters 3 and 4 potentially redundant.

However, there are several 1-bit decision problems that arise in opportunistic spectrum sharing, which are naturally modeled as binary hypothesis tests. For instance, providing robust identity to secondary users so that the primary can decide whether a secondary user is interfering with it or not [90]. The results on designing robust detection algorithms at low

SNR are directly applicable to solve such problems. Moreover, the discussion in Chapters 3 and 4 is given to maintain a chronology of results in opportunistic spectrum sensing.

Chapter 3

Signals with known features

3.1 Introduction

In Chapter 2 we saw that robust detection of signals without any special structure (or knowledge of one) is impossible below certain thresholds, called SNR walls. In this chapter, we explore how additional structure in the primary signal can improve robustness. We consider primary signals that contain signal features that are known to the secondary sensor. In particular, we consider two kinds of signal features: pilot tones/training sequences and cyclostationary signal features.

If the secondary sensor knows the primary signal feature, then it can project the received signal in the direction of the known primary signal feature. For example, consider a primary signal with a known pilot tone. In this case a secondary user can run a matched filter (projection in the direction of the pilot tone) and compare the averaged filter output to a threshold to detect the presence/absence of the pilot tone. If the received signal contains a pilot tone (primary is present), then averaging the matched filter output reduces the noise power, while preserving the pilot tone. As the noise averages out, the uncertainty in the noise statistics also shrinks. This improves the effective SNR of the pilot tone. Hence, coherently combining the received signal samples with the known pilot tone provides robustness gains for the secondary user by increasing the effective SNR. We call this improvement in robustness as gains from coherent processing.

In the above discussion we ignored the fading in the wireless channel from the primary transmitter to the secondary sensor. Fading changes the direction of the pilot tone, and hence perfect matched filtering is only possible if the channel coefficients are known at the secondary sensor. However, as the primary signal can be very weak, it is impossible to accurately track the channel fading coefficients. The lack of knowledge of the fading coefficients at the secondary sensor limits the amount of coherent processing gains feasible. We can expect to coherently combine the signal feature only as long as the channel remains relatively static. So, there is only a finite amount of robustness gain possible, which is a

function of the channel coherence time. This shows that even when the primary signal has a known feature, it suffers from SNR wall limitations. However, the SNR walls in this case are significantly better than those for the radiometer.

The improvement in the SNR wall also depends on the kind of signal feature present. For pilot tones, the improvement in the SNR wall is proportional to $\log_{10} N_c$ (in dB scale), where N_c is the length of the coherence time of the channel. For cyclostationary signals, the improvement in the SNR wall is proportional to $\frac{1}{2}\log_{10} N_c$ (in dB scale). More subtly, the relevant coherence times can be qualitatively different for these two cases.

Also, if the signal feature occupies a fraction of the total available degrees of freedom, then we can further improve robustness by calibrating noise during run-time. The idea is that by taking measurements along a subset of the signal degrees of freedom, we can calibrate the noise in the degrees of freedom where the signal feature is present. This reduces the uncertainty in the noise during run-time!

Table 3.1 summarizes the robustness results proved in this chapter for signals with known features. The scaling of the SNR wall in each case is listed along with the key uncertainty to which the feature is non-robust to.

Signal model	Strategy	Key reason for non-robustness	SNR wall scaling
Unknown or 'white' signal with a power constraint	Energy detection	Uncertain noise power	constant
	Optimal detection	Uncertain noise moments	
'White' signal with narrow-band pilot	Matched filter	Finite phase-coherence time and uncertain noise power	$\log_{10} N_c$
	Matched filter with run-time noise calibration	Finite phase-coherence time and uncertain noise color	
Cyclostationary signals	Histogram-based detector defined in (3.16)	Finite delay-coherence time and uncertain noise power	$\frac{1}{2}\log_{10} D_c$
	Histogram-based detector defined in (3.20) with run-time noise calibration	Finite delay-coherence time and uncertain noise color	

Table 3.1: Comparison of robustness results for different classes of signal models.

3.2 Signals with deterministic pilot tones

Assume that the primary signal allocates a fraction of its total power to transmit a known and deterministic pilot tone and the secondary sensor just focuses on that part for detection. This model covers many practical communication schemes that use pilot tones/training sequences for data frame synchronization and timing acquisition. For example, a digital television (ATSC) signal is mandated to have a pilot tone that is -11 dB weaker than the average signal power [91].

3.2.1 Pilot detection

We now consider detection algorithms for signals containing pilot tones. For concreteness, we state the hypothesis test for this problem, which is similar to the one in (2.1)

$$\begin{aligned} \mathcal{H}_0 : Y[n] &= W[n]; & W[n] \sim W \in \mathcal{W} \\ \mathcal{H}_1 : Y[n] &= \sqrt{P} \cdot H(X[n]) + W[n]; & X[n] \in \mathcal{X}_p, W[n] \in \mathcal{W}, \end{aligned} \tag{3.1}$$

where we assume that all potential primary signals $X[n] \in \mathcal{X}_P$ are of the form $\sqrt{\theta}X_p[n] + \sqrt{(1-\theta)}X_d[n]$. Here $X_p[n]$ is a deterministic pilot tone (normalized to have unit power) with θ being the fraction of the total power guaranteed to be allocated to the pilot tone. Let $X_d[n]$ denote the data carrying part of the primary signal, modeled as a zero-mean *iid* random process. Again, without loss of generality we assume that $X_d[n]$ has unit power. The noise samples $W[n]$ are independent of the signal, and can take any distribution in the noise uncertainty set $\mathcal{W}$. The fading process is denoted by the operator $H(\cdot)$, and can also be uncertain, i.e., $H \in \mathcal{H}_f$. We also assume that the fading operator is normalized to have unit norm. This is justified by capturing the fluctuations in the signal power (due to path-loss and fading) using the received signal power parameter P. This assumption simplifies the detector specification, and the impact of fading is dealt by mandating detectors to have high sensitivity (can detect signal with low values of P). In Chapter 5 we show that this modeling assumption might not always reveal the true underlying tradeoffs, and present a framework in which we can explicitly incorporate models for path-loss and fading into the hypothesis testing formulation.

The matched filter[1] is the most commonly used detector to sense signals with pilot tone. The test-statistic of a matched filter is given by $T(\mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N} Y[n]X_p[n]$.

[1] In the case of completely known noise statistics, the matched filter is asymptotically optimal. It achieves a dwell time of $N \approx [\mathcal{Q}^{-1}(P_D) - \mathcal{Q}^{-1}(P_{FA})]^2 \ \theta^{-1} SNR^{-1}$ [80]

3.2.1.1 Robustness analysis of the matched filter

We consider two different cases, with and without fading.

No fading: It is easy to see that when there is no fading ($H(\cdot)$ is an identity operator), and the noise uncertainty set is $\mathcal{W}_x$ (see Definition 2.3), the matched filter is completely robust to uncertainties in the noise distribution. Notice that $\mathbb{E}[T(\mathbf{Y})|\mathcal{H}_0] = 0$ and $\mathbb{E}[T(\mathbf{Y})|\mathcal{H}_1] = \sqrt{\theta P}\frac{1}{N}\sum_{n=1}^{N}|X_p[n]|^2 \neq 0$. This shows that the set of means of $T(\mathbf{Y})$ under both hypotheses do not overlap and hence the detector can robustly achieve any (P_{FA}, P_{MD}) pair by choosing a sufficiently large N.

Faded signal: In this case the matched filter is non-robust to time-selectivity of the fading process. Let us consider the simplest possible time-selective fading process: $H(\cdot)$ is modeled as acting on the signal by simple scalar multiplication $H(X[n]) = h[n]X[n]$ where $h[n]$ is the unknown time-varying flat fade. Even with this simple fading model, it is clear that we cannot reap the gains of coherent signal processing forever. As soon as the channel assumes an independent realization, we can no longer gain from coherent signal processing.

To simplify analysis we assume a block fading model for the fades $h[n]$. That is, $h[n]$ is assumed to be a random variable (constant within each coherence block) taking an independent realization every coherence time. Furthermore, assume that the length of the coherence time is N_c. It is clear that a good modification to the matched filter gives the following test statistic

$$T(\mathbf{Y}) = \frac{1}{M}\sum_{n=0}^{M-1}\left[\frac{1}{\sqrt{N_c}}\sum_{k=1}^{N_c} Y[nN_c+k]X_p[nN_c+k]\right]^2, \tag{3.2}$$

where M is the number of coherent blocks over which we can listen for the primary.

This detector can be visualized as a combination of two detectors. First, the signal is coherently combined within each coherence block. Coherent processing gain boosts the signal power by N_c while the noise uncertainty is unchanged. Second, the new boosted signal is detected by passing it through a radiometer. The radiometer aspect remains non-robust to noise uncertainties, in spite of the boost in the signal strength. The effective SNR of the coherently combined signal is given by $SNR_{eff} = SNR \cdot \theta \cdot N_c$.

Hence, the modified matched filter will be non-robust if

$$\begin{aligned} SNR \cdot \theta \cdot N_c &\leq \left(\frac{\rho^2-1}{\rho}\right) \\ \Rightarrow SNR^{mf}_{wall} &= \frac{1}{N_c \cdot \theta}\left(\frac{\rho^2-1}{\rho}\right). \end{aligned} \tag{3.3}$$

Using computations similar to the ones in Section 2.3.1, we can also show that the sample complexity of the modified matched filter given in (3.2) is

$$N^{mf} \approx \frac{2N_c[\mathcal{Q}^{-1}(P_{FA}) - \mathcal{Q}^{-1}(1-P_{MD})]^2}{\left[\theta \cdot N_c \cdot SNR - \left(\frac{\rho^2-1}{\rho}\right)\right]^2} \tag{3.4}$$

Figure 3.1 plots the sample complexity of the radiometer and the matched filter with and without (solid and dashed curves) noise uncertainty. Notice that at high SNR, the radiometer has better sample-complexity performance than the matched filter. This is because the radiometer uses the total power in the signal for detection, but the matched filter uses only a small fraction of the total signal power. The most prominent feature of the figure is that under noise uncertainty, the sample complexity curves blow up to infinity as the SNR approaches the corresponding SNR walls.

Notice that the sample-complexity curves for the radiometer and the matched filter (dashed lines) without uncertainty have slopes of $-\frac{2}{10}$ and $-\frac{1}{10}$ respectively. With noise uncertainty, the sample-complexity curves deviate from their nominal (no uncertainty) behavior as the SNR approaches the wall. The matched-filter's curve (blue) has an interesting intermediate phase. For moderately low SNRs, the number of samples required for the matched filter is on the order of multiple coherence times and here the slope transitions to $-\frac{2}{10}$ before ramping to $-\infty$ near the SNR wall.

Coherent processing gains could also be limited due to implementation issues. For example, the clock-instability of both the sensing radio and the primary transmitter imposes a limit on the coherent processing time. The primary signal itself might also contain pilots that are only coherent for a certain length of time. For example, the primary "user" might actually be a collection of wireless nodes doing TDMA among each other and sending packets within their assigned time-slots. Each packet has its own pilot, but the different primary nodes are not all phase-synchronized with each other so no coherent processing gain is available across packets.

Extensive simulation results evaluating the performance of coherent detectors for captured DTV signals can be found in [69; 92].

3.2.2 Noise calibration

So far we have shown the existence of SNR walls for detection algorithms that do not attempt to explicitly do anything about the noise uncertainty. The key question is whether it is possible to obtain accurate calibration of uncertain statistical quantities like noise or fading at *run-time*. The focus on run-time is important because both the fading process and the noise/interference are likely to be at least mildly non-stationary in practice. This means that calibration will get stale (leaving substantial residual uncertainty) if it is done far in advance.

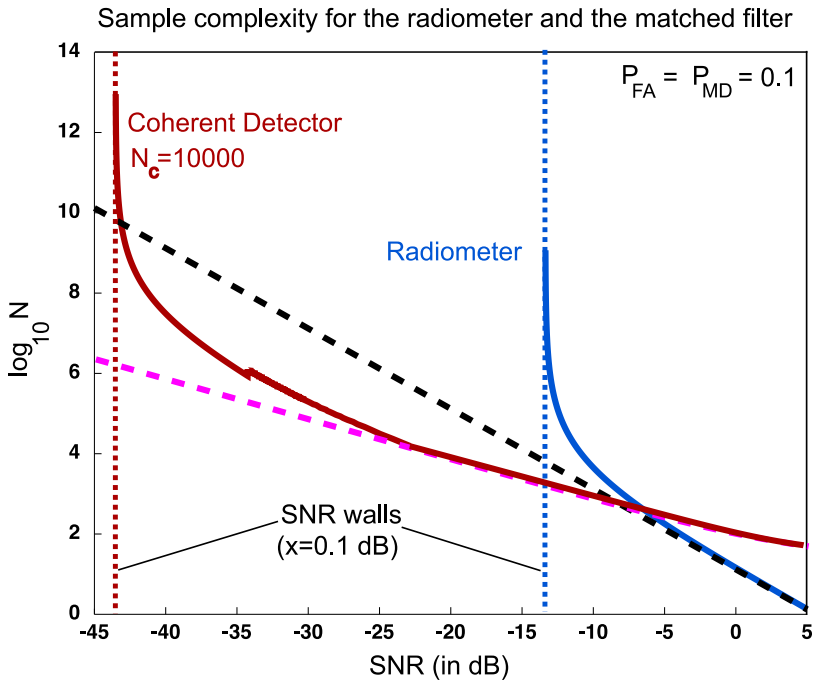


Figure 3.1: Under noise uncertainty the sample complexities for both the radiometer (blue curve) and the matched filter (red curve) blow up to infinity as the SNR approaches the corresponding wall. The dashed lines show the sample complexity without noise uncertainty. In the plot, the pilot tone contains a fraction $\theta = 0.1$ of the total signal power.

This leads to a tension. Ideally, we would like access to $\mathcal{H}_0$ in a parallel universe that is identical to our own except that the primary user is guaranteed to be absent. However, our detector is necessarily confined to our own universe where the presence or absence of the primary user is unknown. Any run-time calibration that aims to reduce the noise uncertainty must use data that might be corrupted by the presence of the primary signal. When the primary user is allowed to be white, then Theorem 2.2 can be interpreted as telling us that noise-calibration is impossible. Intuitively, there is no degree of freedom that is guaranteed to be uncorrupted by the primary signal.

Therefore, a prerequisite for getting any noise-calibration gains is that the primary signal must be constrained to discriminate among the available degrees of freedom. *In a run-time setting, noise-calibration is only possible if the signal features allow it.* Pilot-detection provides a natural case to explore since the pilot signal is confined to a specific frequency, leaving measurements at other frequencies available to do noise calibration.

One can think of noise calibration as a method for adjusting the detector threshold using the additional run-time measurements. For instance, in Section 2.3.1 the radiometer sets its detector threshold corresponding to the worst case noise variance in $[\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]$. If the radiometer had access to additional measurements of noise at run-time, then it could adjust its threshold accordingly and hence improve performance.

3.2.2.1 Pilots with white noise

Consider the basic model of Section 3.2.1, except with the additional assumption that the pilot $X_p[n]$ is a narrowband pilot-tone given by $X_p[n] = \sqrt{2}\sin 2\pi\frac{f_p}{B}n$, where P is the average signal power and f_p is the pilot tone frequency. Rather than modeling the fading process using a block-fading model with a coherence time of N_c samples, it is convenient to think about the fading model $\mathcal{H}_f$ as being arbitrary but bandlimited to a Doppler bandwidth of B_d. This means that the coherence time $T_c \approx \frac{1}{B_d}$. Since $N_c = T_c B$ where B is the original width of the channel, the Doppler bandwidth can be considered $B_d = \frac{B}{N_c}$. Thus, the coherent processing using the matched filter of the previous section can be reinterpreted as an ideal bandpass filter that selects a bandwidth that is a factor $\frac{1}{N_c}$ as wide as the total primary band and centered on f_p. The bandpass filter gives a reduction in noise power, and thus boosts the effective SNR by a factor of N_c.

The approach of the previous section throws away all the information in the received signal that lies outside that bandpass filter. Noise calibration is about using that information to improve detection robustness.

Figure 3.2 pictorially describes the noise calibration technique in the case of signals with narrowband pilots. The figure plots the power spectral density (PSD) of the primary signal plus noise. The shaded region is the noise PSD and the solid region corresponds to the signal PSD. The sinusoidal pilot tone is shown as a Dirac delta in the frequency domain. Let f_p be the frequency of the pilot tone and $f_m \neq f_p$ be any other frequency within the channel.

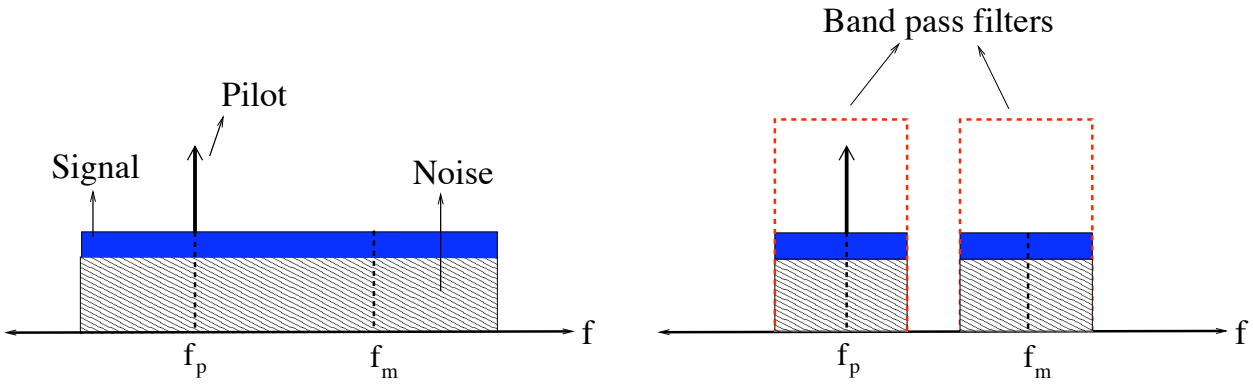


Figure 3.2: Noise calibration in the case of stationary white noise and white signals with narrow-band pilot tones.

The noise-calibration detection algorithm can be formally described as follows:

- Pass the received signal through two identical bandpass filters with disjoint passbands centered at f_p and f_m respectively. For mathematical convenience assume that these bandpass filters are ideal "brick-wall" filters and denote their bandwidth by B_{BPF}. Let the output of these filters be denoted by $Y_p[n]$ and $Y_m[n]$ respectively.

- Measure the empirical power of both outputs. Let them be denoted by $\widehat{P}(f_p, N)$ and $\widehat{P}(f_m, N)$ respectively. That is $\widehat{P}(f_p, N) = \frac{1}{N}\sum_{n=1}^{N} |Y_p[n]|^2$ and $\widehat{P}(f_m, N) = \frac{1}{N}\sum_{n=1}^{N} |Y_m[n]|^2$.

- Compare $\widehat{P}(f_p, N)$ to a detector threshold λ to distinguish between the two hypotheses. The threshold λ is set so as to robustly achieve the target probability of false alarm, P_{FA}. If we ignore the additional noise measurements $\widehat{P}(f_m, N)$, then this detector would get coherent processing gains due to reduction in noise power, but would not get any gains from noise calibration. However, $\widehat{P}(f_m, N)$ gives a reliable estimate of the actual noise level and hence can be used to adjust the detector threshold, i.e., set $\lambda = \lambda' + \widehat{P}(f_m, N)$.

- Adjusting the detector threshold is the same as considering the following new test statistic: $T(\mathbf{Y}, N) := [\widehat{P}(f_p, N) - \widehat{P}(f_m, N)]$. Compare this test statistic to a static threshold λ' and declare the signal is present if it exceeds the threshold.

We call this detector the *pilot power detector* [69]. $\widehat{P}(f_p, N)$ estimates the total power in a narrow band around the pilot tone and this includes the pilot power itself, signal power (the solid region in Figure 3.2) and the noise power. Since we are concerned with the low

SNR regime, the signal power in the band is much lower than the noise power and can be safely neglected. Similarly, $\widehat{P}(f_m, N)$ estimates the signal power plus noise power in a narrow band around f_m. If we assume that the noise is guaranteed to be white, the noise power in the output of both the band pass filters is the same. Hence, $\mathbb{E}[T(\mathbf{Y}, N)|\mathcal{H}_0] = 0$ and $\mathbb{E}[T(\mathbf{Y}, N)|\mathcal{H}_1] = \theta P \neq 0$. This clearly implies that this detector is robust to distributional uncertainties in the marginals of a white noise process.

3.2.2.2 Colored noise uncertainty model

The robustness of the pilot power detector described in the previous section critically depends on the absolute whiteness of noise. It is unreasonable to assume that noise is completely white. Since "noise" includes thermal noise and interference from other transmitters sharing the same band of spectrum, leakage from adjacent bands, etc, it has some *a priori* unknown color to its spectrum. Real-world noise also contains seemingly arbitrarily placed spurs and other terms that can change quite rapidly across the channel [93].

Our noise model is motivated by the idea that the noise uncertainty is dominated by the unknown interference that is coming from an aggregation of many sources. The uncertainty in the level of the overall noise corresponds to the uncertainty in the activity level of the interference. This can vary wildly over possible sites because it is a function of how far away the sources of interference are. To understand the color, suppose that each interferer is roughly white to begin with, but hits our spectrum sensor after passing through its own frequency-selective channel representing multipath fading. The channels between our spectrum sensor and the different interferers are independent. Thus the autocorrelation function of the sum has a large term at a lag of zero, but the non-zero lag terms coming from different interferers tend to cancel out by the independence of the fades. However, there are a finite number of dominant interferers and so the autocorrelation function of the aggregate is not a perfect delta-function. This type of scenario naturally leads to a model of unknown, but slight, color for the noise.

Definition 3.1. The colored noise uncertainty set $\widehat{\mathcal{W}}_{\rho,\nu}$ is defined as the set of all noise distributions satisfying the following properties:

- The noise process $W[n] \in \widehat{\mathcal{W}}_{\rho,\nu}$ is assumed to be stationary.
- Let $W[n] \sim W$ denote the marginal distribution of the stationary process $W[n]$. Then, $W \in \mathcal{W}_x$ (see Section 2.4.1), where $x = 10 \log_{10} \rho$.
- The power spectral density of the noise process $S_W(f)$ satisfies $\frac{1}{\nu} N_a \leq S_W(f) \leq \nu N_a$ for some $\rho > \nu > 1$. Here N_a satisfies $\frac{\nu}{\rho} N_0 \leq N_a \leq \frac{\rho}{\nu} N_0$ and represents the unknown overall PSD level actually experienced by the cognitive radio in this particular site at the time of sensing. N_0 is the nominal PSD level at design time and this is related to the discrete-time σ_n^2 using $N_0 B = \sigma_n^2$.

Figure 3.3(a) illustrates the set of possible noise power spectral densities under the earlier model $\mathcal{W}_x$ of Section 2.4.1. The PSD is flat, but its variance can vary within the bounds as shown in the figure. Figure 3.3(b) plots the set of possible power spectral densities under the new colored noise uncertainty model $\widehat{\mathcal{W}}_{\rho,\nu}$. The two solid (red) curves correspond to two different nominal levels around which the noise has some ripples.

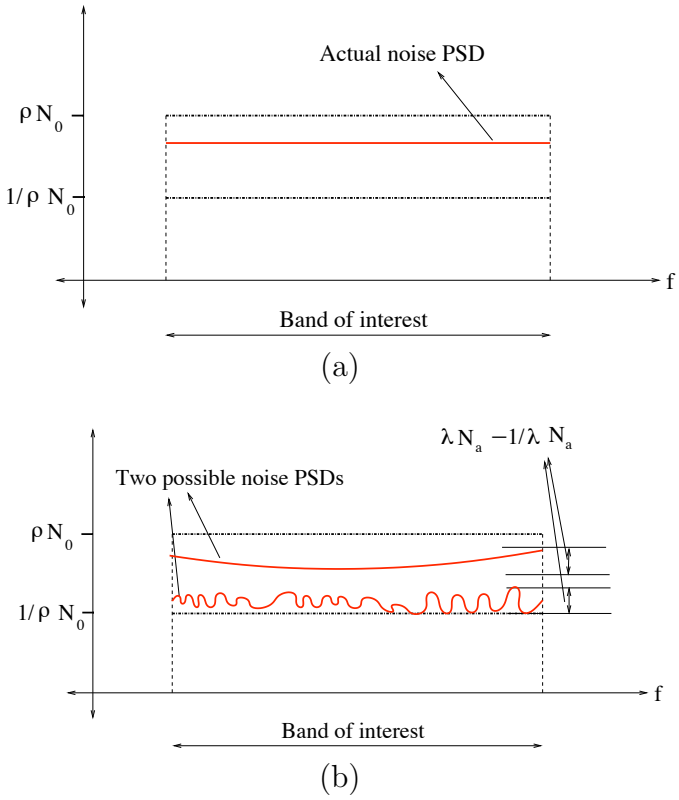


Figure 3.3: Figure (a) describes the power spectral density of noise under the white noise uncertainty model, and Figure (b) describes the power spectral density of noise under the colored noise uncertainty model.

3.2.2.3 Pilots with colored noise

Consider the detector in Section 3.2.2.1 under our new colored noise uncertainty model. Recall that the test statistic for the detector is $T(\mathbf{Y}, N) = \widehat{P}(f_p, N) - \widehat{P}(f_m, N)$, i.e., the test statistic measures the difference in powers at the pilot frequency f_p and the noise calibration frequency f_m. Since the noise can be colored, it is clear that the difference between the noise powers in the two passbands (centered at f_p and f_m) need not be zero. The gains from noise

calibration are limited and the detector faces SNR wall limitations.

Theorem 3.1. *Consider the detection problem in (3.1). Assume that* $X[n] = \sqrt{\theta}X_p[n] + \sqrt{(1-\theta)}X_d[n]$*, the fading uncertainty set* $\mathcal{H}_f$ *contains arbitrary bandlimited processes* H *with Doppler bandwidth* $\frac{B}{N_c}$ *and unit gain, and* $W[n]$ *are noise samples whose distribution lies in the colored noise uncertainty set* $\widehat{\mathcal{W}}_{\rho,\lambda}$*. Assume* $X_p[n] = \sqrt{2}\sin 2\pi\frac{f_p}{B}n$*.*

In this case the pilot power detector *defined in Section 3.2.2.1 is non-robust if*

$$SNR \leq \frac{\rho}{\nu\theta} \cdot \left(\frac{\nu^2-1}{\nu}\right) \cdot \left(\frac{2B_{BPF}}{B}\right), \tag{3.5}$$

where B *is the bandwidth of the primary signal and* $B_{BPF} \geq \frac{B}{N_c}$ *is the bandwidth of the bandpass filter in the detection algorithm. If the filter's width* $B_{BPF} = \frac{B}{N_c}$ *is set to be as narrow as possible, then (3.5) simplifies to:*

$$SNR \leq \frac{\rho}{\nu\theta} \cdot \left(\frac{\nu^2-1}{\nu}\right) \cdot \left(\frac{2}{N_c}\right). \tag{3.6}$$

Proof. The test statistic in Section 3.2.2.1 is $T(\mathbf{Y}, N) = \widehat{P}(f_p, N) - \widehat{P}(f_m, N)$. This detector is non-robust if

$$\begin{aligned}
\inf_{W\in\widehat{\mathcal{W}}_{\rho,\nu}, H\in\mathcal{H}_f} \mathbb{E}_{W,H}[T(\mathbf{Y},N)|\mathcal{H}_1] &\leq \sup_{W\in\widehat{\mathcal{W}}_{\rho,\nu}} \mathbb{E}_W[T(\mathbf{Y},N)|\mathcal{H}_0] \\
\overset{(a)}{\Rightarrow} \theta P + \inf_{W\in\widehat{\mathcal{W}}_{\rho,\nu}} \mathbb{E}_W[T(\mathbf{W},N)] &\leq \sup_{W\in\widehat{\mathcal{W}}_{\rho,\nu}} \mathbb{E}_W[T(\mathbf{W},N)] \\
\overset{(b)}{\Rightarrow} \theta P + \left(\frac{1}{\nu}N_aB_{BPF} - \nu N_aB_{BPF}\right) &\leq \left(\nu N_aB_{BPF} - \frac{1}{\nu}N_aB_{BPF}\right) \\
\overset{(c)}{\Rightarrow} SNR := \frac{P}{\sigma_n^2} &\leq \frac{1}{\theta}\cdot\left(\frac{\nu^2-1}{\nu}\right)\cdot\left(\frac{2B_{BPF}}{B}\right)\cdot\left(\frac{N_a}{N_0}\right) \\
&\leq \frac{\rho}{\nu\theta}\cdot\left(\frac{\nu^2-1}{\nu}\right)\cdot\left(\frac{2B_{BPF}}{B}\right).
\end{aligned}$$

- In step (a) above, $\mathbf{W}$ is an N dimensional vector of noise samples drawn from a distribution $W \in \widehat{\mathcal{W}}_{\rho,\nu}$. The result follows from the fact that the test statistic under hypothesis $\mathcal{H}_1$ is the sum of the faded pilot power and test statistic on the noise alone. This is true since the bandpass filter is wide enough to capture all the pilot power even after it is spread in frequency by the channel's Doppler spread. So $\mathbb{E}\left[|H(X_p)|^2\right] = \mathbb{E}[|H|^2]\mathbb{E}[|X_p|^2] = \theta P$ since the unknown bandlimited fading process H is restricted to have unit gain on average.

In addition, we are neglecting the power of the data part of the signal ($X_d[n]$) in the passband. This assumption is reasonable because we are interested in the regime where $P \ll \sigma_n^2$.

- Step (b) is obtained by using the fact that the minimum difference in noise powers in the two bands is $\left(\frac{1}{\nu}N_aB_{BPF} - \nu N_aB_{BPF}\right)$. Similarly, the maximum possible difference is $\left(\nu N_aB_{BPF} - \frac{1}{\nu}N_aB_{BPF}\right)$.

- (c) follows from the fact that $N_0B = \sigma_n^2$. The final factor of $\frac{\rho}{\nu}$ is just present to account for the fact that $\frac{N_a}{N_0}$ could be as big as $\frac{\rho}{\nu}$. This is an artifact of the fact that we are expressing the wall in terms of the nominal SNR and the uncertainty is modeled multiplicatively.

□

From Theorem 3.1, the only way to have the pilot power detector's SNR wall go to zero is to let $\left(\frac{B_{BPF}}{B}\right) \to 0$. This in turn requires the coherence-time $N_c \to \infty$. This is because the pilot tone is a Dirac delta in frequency domain and is smeared out due to the fading process. The pilot power is spread over the bandwidth of the fading process in some unknown way. Under this uncertainty model $\mathcal{H}_f$, tightening the bandpass filter any further could result in the loss of all the pilot power and hence missed detection.

It is easy to see that because the time allowed for sensing is unbounded, the above noise calibration strategy does not really depend on the width of the bandpass filter used for calibration. Even if a detector used all the information outside of the tight window of width B_d containing the pilot signal, the detector cannot completely calibrate the noise process within that tight window due to the model $\widehat{\mathcal{W}}_{\rho,\nu}$ of unknown color. Within the tight window of width B_d, all detectors essentially face the same problem as Theorem 2.2 — an arbitrary (and hence possibly white) signal in the presence of noise with a residual unknown level coming from the color ripple.

3.2.2.4 Primary signals with perfect guard-bands

The case of pilot tones is just one way that the primary signal can leave room for noise calibration. Suppose instead that the primary signal leaves empty a contiguous $(1-\beta)B$ of its total licensed bandwidth B for a guard-band, but is free to use the remaining βB bandwidth as it sees fit. It is immediately obvious that the pilot power detector can be generalized to this case:

- Pass the received signal through two different ideal bandpass filters. Let the first encompass the tightest band guaranteed to contain the entire primary signal. By the assumptions above, this is of width $B_{BPF} = (\beta + \frac{1}{N_c})B$ since the βB must be expanded by B_d to account for the fading Doppler. Let the second bandpass filter have width

$0 < B_c \leq (1 - \beta - \frac{1}{N_c})B$ and be placed entirely in the guard-band. Let the output of these filters be denoted by $Y_p[n]$ and $Y_m[n]$ respectively.

- Measure the empirical power $\widehat{P}(f_p, N)$ and $\widehat{P}(f_m, N)$ of both filter outputs. Since there is no limit on the sample complexity, assume that we listen for long enough for both to have converged to their limits.
- The test statistic is given by $T(\mathbf{Y}, N) := [\widehat{P}(f_p, N) - \frac{B_{BPF}}{B_c}\widehat{P}(f_m, N)]$. Compare this test statistic to a threshold and declare the signal is present if it exceeds the threshold. The normalization term $\frac{B_{BPF}}{B_c}$ would make the test statistic perfectly zero if there were no signal and the noise were perfectly white.

It is easy to see that Theorem 3.1 generalizes to this scenario and detector. The SNR wall from (3.5) holds with $\theta = 1$ and B_{BPF} as above. This shows that the noise calibration gains add to the bandwidth reduction gains in a straightforward way by reducing the noise uncertainty to a factor that depends on the coloring uncertainty ν.

However, it is important to note that run-time noise calibration does not improve the sample complexity of detection. Noise calibration reduces the uncertainty in the noise variance, but does not reduce the value of the actual noise variance. Hence, the sample-complexity of detection is constant even after noise calibration.

3.3 Signals with cyclostationary features

Most communication signals have in-built periodicities in them, like the symbol rate, carrier frequency, cyclic prefix in OFDM packets, etc. Hence, signals can be modeled as *cyclostationary*, whereas noise is typically modeled as stationary. Many researchers have worked on designing detection algorithms to exploit the in-built periodicities in signals, of which the class of *cyclostationary feature detectors* proposed by William Gardner [59] are the most popular ones. These detectors have been widely studied in the spectrum sharing community and are believed to be good for both signal detection as well as signal identification/classification [61; 60]. We now discuss robustness properties of cyclostationary signals and see what kind of sensing algorithms can exploit the presence of cyclostationary features.

3.3.1 Gardner's Cyclostationary feature detectors

3.3.1.1 Cyclostationary signal model

We first review the basic properties of cyclostationary random processes. The notation used here is same as in [59]. We work with continuous time signals in this section as it is intuitively easier to think of cyclostationary signals in continuous time. The signal of interest, $X(t)$ is called *wide-sense cyclostationary* if its mean and autocorrelation function

are periodic in time. That is, there exists a $T_0 > 0$ such that $m_X(t) = m_X(t + T_0)$ and $R_X(t + \frac{\tau}{2}, t - \frac{\tau}{2}) = R_X(t + T_0 + \frac{\tau}{2}, t + T_0 - \frac{\tau}{2})$, where

$$m_X(t) \;:=\; \mathbb{E}[X(t)], \quad R_X(t_1, t_2) := \mathbb{E}\left[X(t_1)X^*(t_2)\right].$$

This immediately implies that $R_X(t + \frac{\tau}{2}, t - \frac{\tau}{2})$ admits a Fourier series representation given by

$$R_X(t + \frac{\tau}{2}, t - \frac{\tau}{2}) = \sum_{\alpha} R_X^{\alpha}(\tau) e^{i2\pi\alpha t},$$

where $\alpha = \frac{k}{T_0}$, $k = 0, 1, 2, \ldots$. Assume that the signal $X(t)$ is zero-mean, i.e., $m_X(t) = 0$. The Fourier coefficient $R_X^{\alpha}(\tau)$, which depends on the lag parameter τ, is called the *cyclic autocorrelation function*, and α is called the cycle frequency. For $\alpha = 0$, it reduces to the conventional autocorrelation function $R_X^0(\tau)$. One can visualize $R_X^0(\tau)$ as the DC component of the lag-product waveform $X(t+\frac{\tau}{2})X^*(t-\frac{\tau}{2})$ for each value of τ, and $R_X^{\alpha}(\tau)$ can be thought of as the AC component corresponding to the frequency α. The Fourier transform of $R_X^{\alpha}(\tau)$ is given by

$$S_X^{\alpha}(f) := \int_{-\infty}^{\infty} R_X^{\alpha}(\tau) e^{-i2\pi f\tau}\, d\tau$$

is called the *spectral correlation function*. For $\alpha = 0$ this reduces to the conventional power spectral density. For $\alpha \neq 0$, $S_X^{\alpha}(f)$ can be thought of as the density of the correlation between spectral components at frequencies $f + \frac{\alpha}{2}$ and $f - \frac{\alpha}{2}$. Explicit formulas of $S_X^{\alpha}(f)$ for many classes of modulated signals can be found in [94; 95; 96]. It is easy to show that if $W(t)$ is wide-sense stationary, then its spectral correlation function $S_W^{\alpha}(f) = 0$ for $\alpha \neq 0$ [97].

3.3.1.2 Single cycle feature detectors

It is well known that we need to over-sample the received signal in order to reconstruct the features from the signal samples since the features are capturing the *shape* of the pulses onto which the symbols have been modulated. We assume that the sampling rate $f_s > 2\max\{\alpha, B\}$, where α is the cycle frequency of the feature we are trying to detect, and B is the bandwidth of $X(t)$.

The test statistic for a single cycle feature detector is defined as

$$Z_{sc}(N) := \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \left| \widetilde{S}_Y^{\alpha}(f, N) \widehat{S}_X^{\alpha}(f)^* \right| df, \tag{3.7}$$

where $\widehat{S}_X^{\alpha}(f) := \frac{S_X^{\alpha}(f)}{\sqrt{\mathcal{E}}}$, with $\mathcal{E} = \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} |S_X^{\alpha}(f)|^2\, df$, and N is the number of signal samples used for sensing. Note that $\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} |\widehat{S}_X^{\alpha}(f)|^2\, df = 1$. Intuitively, the detector (3.7) is doing

matched filtering of the estimated spectral correlation function $\widetilde{S}_Y^{\alpha}(f)$ with the normalized spectral correlation function $\widehat{S}_X^{\alpha}(f)$ of the signal we are trying to detect. Figure 3.4 shows the block diagram for a single cycle feature detector.

The empirical estimate of the spectral correlation function, $\widetilde{S}_Y^{\alpha}(f, N)$, can be computed from the received signal samples $Y[n]$ by [94; 98]

$$\begin{aligned} \widetilde{S}_Y^{\alpha}(f, N) &= \frac{1}{N}\frac{1}{T}\sum_{n=0}^{N} Y_T\left((2n+1)\frac{T}{2}, f+\frac{\alpha}{2}\right) \\ &\quad Y_T^*\left((2n+1)\frac{T}{2}, f-\frac{\alpha}{2}\right), \end{aligned} \tag{3.8}$$

where $Y_T(n, f)$ is a T point DTFT of the signal $Y[n]$ centered at time n, i.e.,

$$Y_T(n, f) = \sum_{u=n-\frac{T}{2}}^{n+\frac{T}{2}} Y[u]e^{-i2\pi\left(\frac{f}{f_s}\right)u}. \tag{3.9}$$

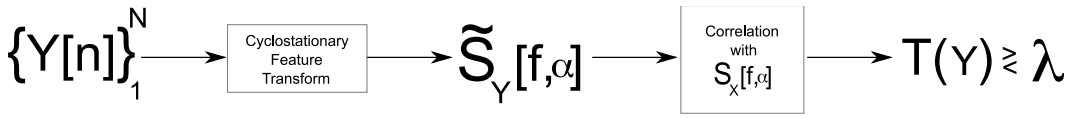


Figure 3.4: The block diagram showing various stages in the single-cycle cyclostationary feature detection algorithm.

The key idea behind the single cycle feature detector is that for $\alpha \neq 0$,

$$\lim_{N\to\infty} \widetilde{S}_Y^{\alpha}(f, N) \begin{cases} = 0 & Y[n] \text{ is stationary } (\mathcal{H}_0) \\ \neq 0 & Y[n] \text{ is cyclostationary } (\mathcal{H}_1) \end{cases}$$

Hence, by increasing the sensing time N, one can extract the signal feature from noise and thereby detect the presence/absence of the signal. Figure 3.5 shows the empirical estimate of the spectral correlation function for $Y[n]$ under hypotheses $\mathcal{H}_1$ for different values of the sensing time N. In this figure, the signal $X(t)$ is a raised-cosine pulse modulated with *iid* BPSK samples. It is clear from the figure that as the sensing time increases, averaging forces the noise to zero, but the signal feature stands out.

3.3.2 Spectral redundancy

A key requirement for the single-cycle cyclostationary feature detector (see Eqn. (3.7)) to work is that the signal of interest should have spectral redundancy. We now define what

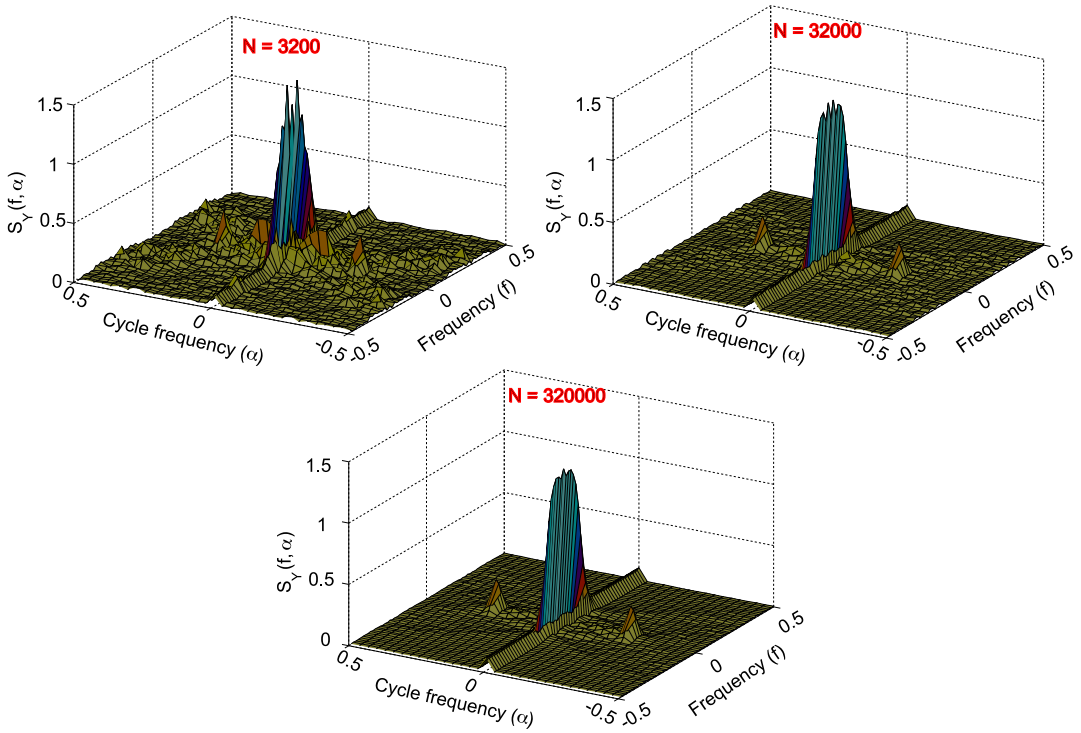


Figure 3.5: The empirical spectral correlation function $\widetilde{S}_Y^{\alpha}(f, N)$ for a single-cycle cyclostationary feature detector is illustrated in this figure. In this figure, the signal $X(t)$ is a raised-cosine function modulated with *iid* $BPSK$ symbols, and $W(t)$ is a white Gaussian noise process. As the raised-cosine pulse shape has spectral redundancy, the empirical spectral correlation function shows cyclostationary features. In fact, as N increases the noise averages to zero, but the cyclostationary signal feature stands out.

we mean by spectral redundancy and then give an example of a class of signals with and without spectral redundancy.

Definition 3.2. Consider a digital modulated signal $X(t)$ being transmitted over a given frequency band. The signal $X(t)$ is said to have *spectral redundancy* if the symbol rate is strictly less than the Nyquist rate.

For example, consider the class of pulse-amplitude modulated (PAM) signals.

$$\begin{aligned} X(t) &= \left(\sum_{m=-\infty}^{\infty} A_m \delta(t - mT_0) \right) * p(t) \\ &= \sum_{m=-\infty}^{\infty} A_m p(t - mT_0), \end{aligned}$$

where $p(t)$ is the pulse shape onto which the data symbols A_m are modulated.

- $p(t) = \text{sinc}\left(\frac{t}{T_0}\right)$. In this example, the Nyquist rate is $\frac{1}{T_0}$ (twice the bandwidth of the signal), and the symbol rate is also $\frac{1}{T_0}$ (one symbol is transmitted even t_0 seconds). Hence, a PAM signal with Sinc modulation has no spectral redundancy.

- $p(t) = \text{sinc}\left(\frac{t}{T_0}\right) \frac{\cos\left(\frac{\pi\beta t}{T_0}\right)}{1-\frac{4\beta^2 t^2}{T^2}}$ is a raised-cosine pulse function with roll-off factor $\beta \neq 0$. In this case the Nyquist rate is $(1+\beta)\cdot\frac{1}{T_0}$, and the symbol rate is $\frac{1}{T_0}$. This is an example where the symbol rate is strictly less than the Nyquist rate, and hence this signal is said to have spectral redundancy.

Given the formal definition of spectral redundancy it is easy to show that the signal cycle cyclostationary feature detector described in Section 3.3.1.2 will not work if the signal has no spectral redundancy. The main idea is that the spectral correlation function $S_X^\alpha(f)$ vanishes when the spectral redundancy is zero [99].

We now give a counter-example showing that spectral redundancy is not fundamental to cyclostationary feature detection.

3.3.3 Counter-example: PAM signals with Sinc modulation

Consider a pulse-amplitude modulated signal using a Sinc pulse function. That is

$$X_s(t) = \sum_{m=-\infty}^{\infty} A_m \text{sinc}\left(\frac{t-mT_0}{T_0}\right), \tag{3.10}$$

for some $T_0 > 0$, and let A_m be *independent and identically distributed* data symbols. For concreteness, assume that A_m are drawn out of a symmetric BPSK constellation, i.e., $A_m \sim \text{Bernoulli}(\frac{1}{2}) \in \{-\sqrt{P}, \sqrt{P}\}$. $X(t)$ is cyclostationary with no spectral redundancy. This implies that the spectral correlation function $S_{X_s}^\alpha(f) = 0$. Hence, we cannot use the single cycle detection described in Section 3.3.1.2 to distinguish the presence/absence of $X_s(t)$ in noise.

We now show the existence of *histogram* based detection algorithms that can distinguish the presence/absence of the cyclostationary signal $X(t)$ in noise. Detection algorithms similar to the histogram-based detectors proposed in this section have also been proposed before in the cognitive radio literature. For instance, the idea of using higher-order statistics for signal detection is given in [100]. Entropy-based detection algorithms for signal detection have been proposed in [101]. Most of these algorithms rely on the fact that noise is Gaussian (in some of the papers, the noise variance is unknown). However, our algorithm works for any stationary noise distribution.

We now state the hypothesis testing problem.

$$\begin{aligned} \text{Signal absent } \mathcal{H}_0 : Y(t) &= W(t) \\ \text{Signal present } \mathcal{H}_1 : Y(t) &= \sqrt{P}X_s(t) + W(t), \end{aligned} \tag{3.11}$$

where $W(t)$ is assumed to be a stationary noise process, $X_s(t)$ is the unattenuated signal defined in (3.10) and $Y(t)$ is the received signal.

The key steps in the *histogram* based cyclostationary signal detection algorithm are described below:

1. First oversample the received signal $Y(t)$ in (3.10) to give a discrete signal $Y[n]$. For concreteness, assume that the signal is sampled at twice the Nyquist rate, i.e., $Y[n] = Y\left(n \cdot \frac{T_0}{2}\right)$.

2. Collect the *odd* and *even* numbered samples separately and label these processes are $Y_o[n] := Y[2n-1]$ and $Y_e[n] := Y[2n]$. Figure 3.6 shows the oversampling process pictorially. The odd and even numbered samples are colored in 'black' and 'blue' respectively.

3. The test-statistic for the detector is given by

$$T(\mathbf{Y}) = \frac{2}{N}\sum_{n=1}^{\frac{N}{2}} \phi(Y_o[n]) - \frac{2}{N}\sum_{n=1}^{\frac{N}{2}} \phi(Y_e[n]), \tag{3.12}$$

   where N is the total number of signal samples (assumed to be even) collected, and the function $\phi(\cdot)$ is chosen such that $\mathbb{E}\phi(X_e[n]) \neq \mathbb{E}\phi(X_o[n])$. Here $X_o[n] := X_s[2n-1]$ and $X_e[n] := X_s[2n]$, with $X_s[n] := X_s\left(n \cdot \frac{T_0}{2}\right)$. The existence of such a function $\phi(\cdot)$, and the proof of why this detection algorithm works in given below.

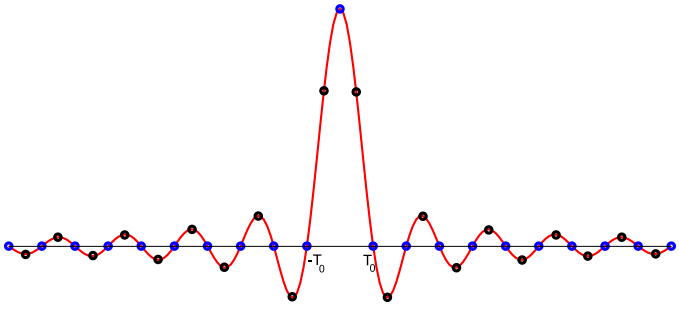


Figure 3.6: Oversampled Sinc function, resulting in a cyclostationary signal

Theorem 3.2. *Consider the hypothesis testing problem in (3.11), with the signal of interest $X_s(t)$ as given in (3.10). Let the detector test-statistic be given by the expression in (3.12). Then, this detector can distinguish between the two hypotheses with arbitrarily low probabilities of error.*

Proof. We begin by proving the existence of the function $\phi(\cdot)$. It is easy to verify that the random processes $\{X_e[n]\}_{n=1}^{\infty}$ and $\{X_o[n]\}_{n=1}^{\infty}$ are stationary processes. This is because each of them are Nyquist-rate sampled versions of the original PAM process $X_s(t)$. This implies that $\{X_e[n]\}_{n=1}^{\infty}$ and $\{X_o[n]\}_{n=1}^{\infty}$ are identically distributed (marginal distributions are same for all n) random processes. However, the marginal distributions of each of these processes are distinct. This is true because $X_e[n]$ is a BPSK random variable, whereas $X_o[n]$ is a particular linear combination of an infinite number of BPSK random variables. Figure 3.7 shows the histograms of the even and odd samples. Clearly, the histogram of the even samples has a Bernoulli distribution, whereas the histogram of the odd samples is something different. This implies that at least one of the moments[2] of $X_e[n]$ will be distinct from the moments of $X_o[n]$. Hence, there exists some function $\phi(\cdot)$ such that $\mathbb{E}\phi(X_e[n]) \neq \mathbb{E}\phi(X_o[n])$.

Since $\{X_e[n]\}_{n=1}^{\infty}$ and $\{X_o[n]\}_{n=1}^{\infty}$ are identically distributed random processes with distinct marginal distributions, we can show that $\{Y_e[n]\}_{n=1}^{\infty}$ and $\{Y_o[n]\}_{n=1}^{\infty}$ also are also identically distributed random processes with distinct marginal distributions under hypothesis $\mathcal{H}_1$. On the other hand, under hypothesis $\mathcal{H}_0$, $\{Y_e[n]\}_{n=1}^{\infty}$ and $\{Y_o[n]\}_{n=1}^{\infty}$ are samples from a stationary noise process $W(t)$. So under hypothesis $\mathcal{H}_0$, $\{Y_e[n]\}_{n=1}^{\infty}$ and $\{Y_o[n]\}_{n=1}^{\infty}$ are identically distributed random processes with the same marginal distributions.

Using the above observations we have

$$\begin{aligned}
\lim_{N\to\infty} T(\mathbf{Y}) &= \lim_{N\to\infty} \frac{2}{N}\sum_{n=1}^{\frac{N}{2}} \phi(Y_o[n]) - \frac{2}{N}\sum_{n=1}^{\frac{N}{2}} \phi(Y_e[n]) \\
&= \mathbb{E}\phi(Y_o[n]) - \mathbb{E}\phi(Y_e[n]) \\
&= \begin{cases} \mathbb{E}\phi(W_o[n]) - \mathbb{E}\phi(W_e[n]) & \text{under } \mathcal{H}_0 \\ \mathbb{E}\phi(X_o[n]+W_o[n]) - \mathbb{E}\phi(X_e[n]+W_e[n]) & \text{under } \mathcal{H}_1 \end{cases} \\
&\quad \begin{cases} = 0 & \text{under } \mathcal{H}_0 \\ \neq 0 & \text{under } \mathcal{H}_1 \end{cases}
\end{aligned}$$

This shows that the test-statistic converges to zero under hypothesis $\mathcal{H}_0$, whereas it converges to a non-zero value under hypothesis $\mathcal{H}_1$ as the sensing time N approaches infinity. Hence, this detection algorithm can distinguish between the 'noise only' hypothesis and the 'signal plus noise' hypothesis. □

[2]For the case of a BPSK modulation all the even moments of $X_e[n]$ excluding the 2nd moment are distinct from the corresponding moments of $X_o[n]$.

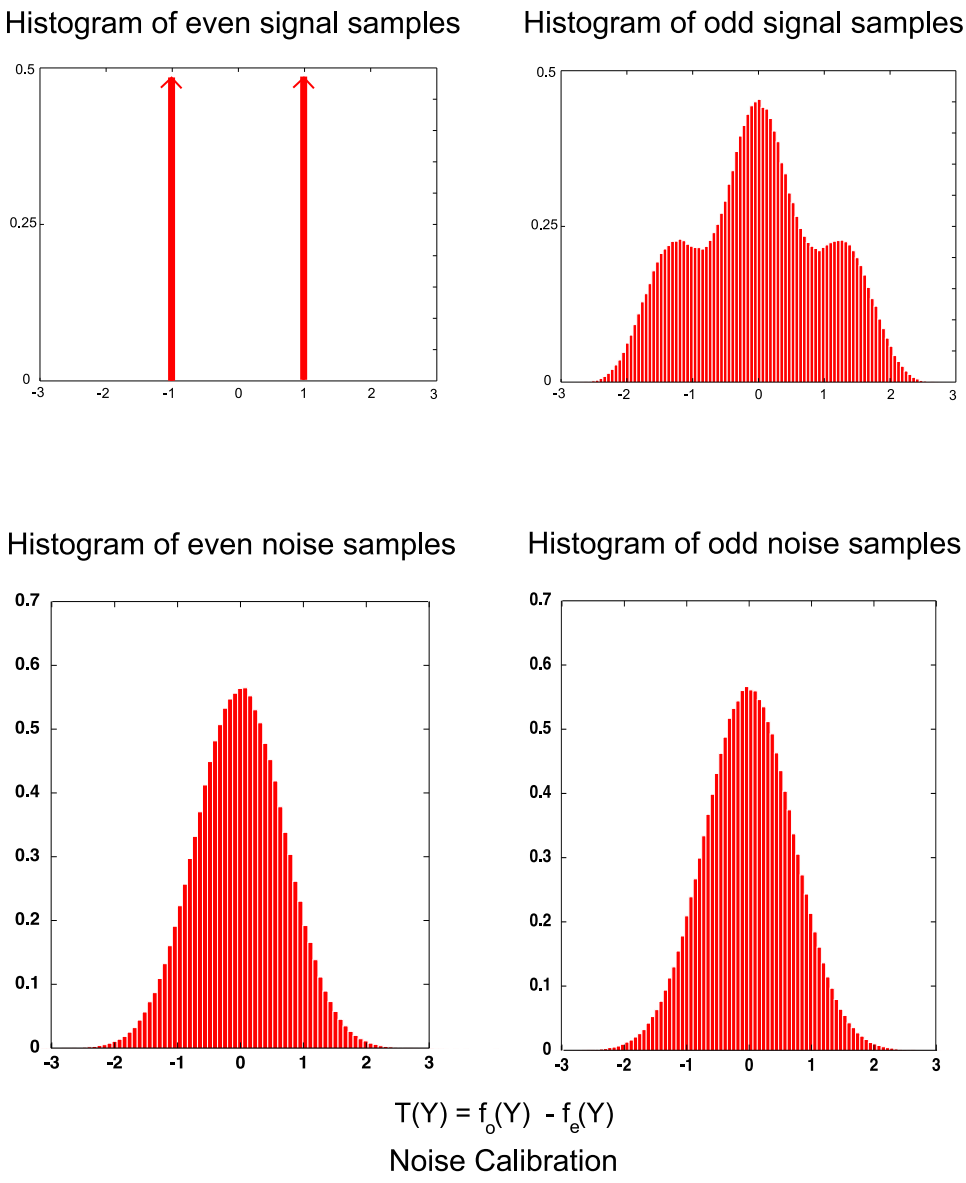


Figure 3.7: The histograms of the even and odd signal samples are distinct, whereas the histograms of the even and odd noise samples are identical. Hence, any test-statistic that calibrates out the noise can be used to robustly distinguish the signal from noise. This is an example of a cyclostationary signal without spectral redundancy, yet it can be robustly detected in the presence of unknown stationary noise.

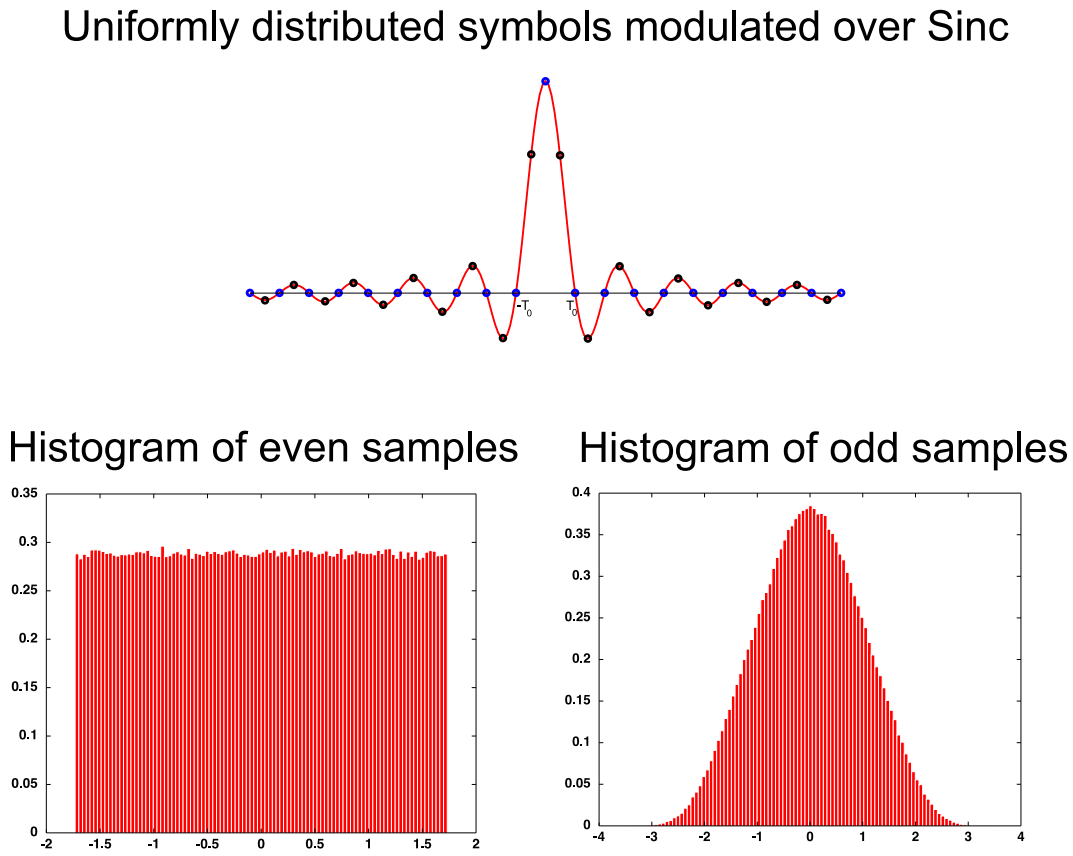


Figure 3.8: The histogram of the even and odd signal samples are shown for the case when the data symbols A_m are Uniform $\sim [-\sqrt{3}, \sqrt{3}]$. Even in this case the histogram are distinct and hence run-time calibration can distinguish the presence/absence of the signal in noise with any unknown stationary distribution.

Remarks:

- The main idea behind the counter-example is that the oversampled cyclostationary process $X_s[n]$ is wide-sense stationary (both the mean and auto-correlation function are time-invariant). This can be shown by using Parseval's identity and the fact that there is no aliasing when the pulse function $p(t)$ is a Sinc.
- On the other hand, if one looks at the empirical distribution of $X_s[n]$ we can clearly see two distinct distributions — the distribution of the even samples and the distributions of the odd samples. This is shown by the histogram of the even and odd samples in Figure 3.7. So, looking at the distributions of the even and odd samples helps us identify the presence of the signal versus noise alone (the distribution of the even and odd samples of the noise process are identical). As a contrast Figure 3.7 also shows the histogram of the even and odd samples of a white Gaussian noise process, and they look identical. This result also holds true for any stationary noise distribution.
- The counter-example style above is not restricted to BPSK modulated data. Figure 3.8 shows the histogram of the even and odd samples of the signal for the case when the data symbols A_m are uniformly distributed random variables. Even in this case the two histograms are distinct from each other.

3.3.4 Robustness of cyclostationary detection

So far we have seen two types of detectors, Gardner's cyclostationary feature detectors and our histogram-based cyclostationary signal detectors, both of which can detect wide-sense cyclostationary signals. However, the histogram-based detectors can detect any cyclostationary signal, even if it is wide-sense stationary. In this section we analyze the robustness performance of these detectors. For simplicity, we only present the robustness analysis of the histogram-based detectors here. The readers are encouraged to refer to our paper [72] for the robustness analysis of Gardner's cyclostationary feature detectors. It turns out that the robustness performance of both these detectors is qualitatively the same. The only difference is that the histogram-based detectors are more general than Gardner's cyclostationary feature detectors.

Again, for concreteness we rewrite the two hypotheses in the detection problem.

$$\begin{aligned} \mathcal{H}_0 : Y[n] &= W[n]; \qquad\qquad W[n] \in \widehat{\mathcal{W}}_{\rho,\nu} \\ \mathcal{H}_1 : Y[n] &= \sqrt{P} \cdot H(X[n]) + W[n]; \qquad W[n] \in \widehat{\mathcal{W}}_{\rho,\nu}. \end{aligned} \tag{3.13}$$

We assume that that the signal $X[n]$ is a zero-mean, symmetric (all odd moments are zero at each time n) cyclostationary process. We normalize the signal such that the average signal

power is 1. For simplicity, we consider cyclostationary signals with period 2, and assume that the odd and even samples of $X[n]$ are stationary processes, and have distinct marginals[3]. That is, if we define $X_e[n] := X[2n]$, and $X_o[n] := X[2n-1]$, then we assume that $X_e[n]$ and $X_o[n]$ are stationary processes respectively, and the marginal distribution of $X_e[n]$ is distinct from the marginal distribution of $X_o[n]$.

For this sensing problem we consider histogram-based detection algorithms. As the marginal distribution of $X_e[n]$ and $X_o[n]$ are distinct there exist some integer $k > 0$ such that the the $2k$th moment of $X_e[n]$ is distinct from the $2k$th moment of $X_o[n]$, i.e., $\mathbb{E}X_e[n]^{2k} \neq \mathbb{E}X_o[n]^{2k}$. Then, we define the test-statistic of the sensing algorithm to be

$$T(\mathbf{Y}) = \frac{2}{N}\sum_{n=1}^{\frac{N}{2}} |Y_o[n]|^{2k} - \frac{2}{N}\sum_{n=1}^{\frac{N}{2}} |Y_e[n]|^{2k}, \tag{3.14}$$

We analyse the robustness of this detector for three different models for the fading operator $H(\cdot)$.

- *No channel fading:* Suppose that there was no fading ($H(\cdot)$ is an identity operator in (3.13)). In this case it is easy to verify that $\mathbb{E}[T(\mathbf{Y})|\mathcal{H}_0] = 0$. This follows from the assumption that noise is stationary, and $\mathbb{E}[T(\mathbf{Y})|\mathcal{H}_1] \neq 0$. This easily follows from the fact that $\mathbb{E}X_e[n]^{2k} \neq \mathbb{E}X_o[n]^{2k}$. Therefore, it is clear that we robustly detect the signal $X[n]$ for arbitrarily low values of P by increasing the sensing time N. This argument is identical to the proof of Theorem 3.2.

- *Flat fading*: Assume that the channel fading is flat, i.e., $H(X[n]) = h[n]$, where $h[n]$ are unknown flat fading coefficients. Assume that $h[n]$ is a stationary random process, and that the average power of $h[n]$ is normalized to be one. Recall that for pilot-detection, this is enough to cause an SNR wall (see Section 3.2.1.1). In this case however, even after the scalar multiplication by $h[n]$ the average $2k$-th moments of the even and odd samples are distinct (stationarity of fading). That is, $\mathbb{E}[T(\mathbf{Y})|\mathcal{H}_1] \neq 0$. As before the stationarity of the noise guarantees that the test-statistic mean under hypothesis $\mathcal{H}_0$, $\mathbb{E}[T(\mathbf{Y})|\mathcal{H}_0] = 0$. So, the signal is robust to flat-fading uncertainty. Figure 3.10 shows the histogram of even and odd samples of a flat-faded signal $h[n] \cdot X_s[n]$ (*iid* BPSK data modulated over a Sinc pulse as given in (3.10)). It is clear from the figure that flat fading affects the marginal distributions of the even and odd samples, but the marginal distributions are still distinct. Hence, the signal can be robustly

[3]Any pulse-amplitude-modulated signal with a raised cosine pulse function (or a Sinc pulse), *iid* data symbols drawn from a signal constellation (except a Gaussian), and an oversampling rate of 2 satisfy the above criterion. The results can easily be extended to any general cyclostationary process with a period $N_0 > 1$. The key property that is required for our results to hold is that the signal $X[n]$ must have at least two sub processes that have distinct marginal distributions.

detectable. Robustness of cyclostationary signals to flat-fading can also be observed by using Gardner's cyclostationary feature detectors, as shown in Figure 3.9.

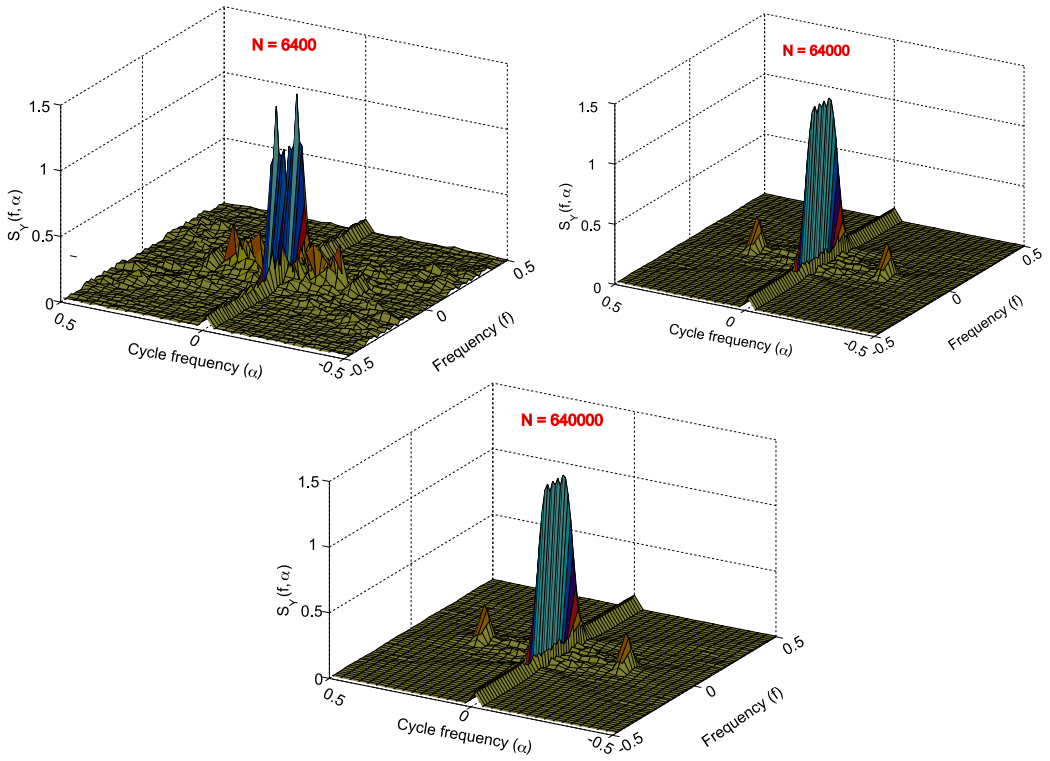


Figure 3.9: The empirical spectral correlation function $\widetilde{S}_Y^{\alpha}(f, N)$ for a single-cycle cyclostationary feature detector is illustrated in this figure. In this figure, the received signal $Y[n] = \sqrt{P}h[n]X[n] + W[n]$, where the signal $X[n]$ is a raised-cosine function modulated with *iid* $BPSK$ symbols, the fading process $h[n]$ is chosen to be a Gauss-Markov process, and $W[n]$ are samples of a white Gaussian noise process. This figures clearly shows that cyclostationary features are robust to flat-fading processes.

- *Frequency-selective fading*: The previous result is somewhat surprising as the matched filter for pilot tones is not-robust to this kind of flat-fading uncertainty, even with noise calibration when the noise is colored. However, frequency-selective fading does introduce an SNR wall for the detector in (3.14). Figure 3.12 illustrates this idea for the signal $X_s[n]$ (*iid* BPSK data modulated over a Sinc pulse as given in (3.10)) when passed through a 2-tap Rayleigh fading process. As shown in the figure, the empirical distributions of the even and odd samples are identical. Frequency-selective fading mixes the even and odd samples and hence destroys the cyclostationarity. Figure 3.11

illustrates the same qualitative behavior using Gardner's cyclostationary feature transform.

To mathematically quantify the non-robustness to frequency selective fading, we consider the simplest possible frequency-selective fading: a pure delay. Let the received signal under hypothesis $\mathcal{H}_1$ be $Y[n] = X[n - l[n]] + W[n]$. In this case, we show that uncertainty in the delay $l[n]$ will lead to an SNR wall. To be concrete, consider a block-fading model in which the delay $l[n]$ is piecewise constant for D_c (assumed to be even) time steps before taking on an independent realization. Since D_c models the minimum time over which the delay of the fading process changes significantly, we call it the *delay-coherence time*. More generally, the delay-coherence time refers to the minimum time over which the shape of the impulse response of the fading process changes significantly.

3.3.4.1 Interpretation of delay-coherence time

The reason for considering the example of a pure delay is two fold: firstly, from a robustness point of view $H(X[n]) = h[n]X[n - l[n]]$ would behave the same as just $X[n - l[n]]$; secondly, a pure-delay is the simplest example for which we can illustrate the non-robustness of cyclostationary signals. We now give an intuitive interpretation for the delay-coherence time and compare it with the phase-coherence time and amplitude-coherence time.

The amplitude-coherence time is the minimum time over which the magnitude response of the fading process changes significantly. The change in the magnitude response is caused due to motion and the amplitude-response time is inversely proportional to the Doppler spread of the channel. In a traditional communication system, a significant change in the magnitude-response leads to a change in the power allocation across frequency. The phase-response time is the minimum time over which the phase of an individual channel tap changes significantly. This change occurs at a much faster scale than the change in the amplitude-response time. On the other hand the delay-coherence time is the minimum time over which the shape of the impulse response changes significantly. Alternatively, we can also think of it as the minimum time over which the relative phase response changes significantly. Typically, the delay-coherence time is larger than the phase-coherence time, but is smaller than the amplitude-coherence time.

In most traditional communication systems, receivers have a phase lock loop (PLL). However, for receivers trying to detect the presence/absence of primary signals it is not possible to have an accurate PLL because the receiver is not sure whether the primary signal is present or absent. This leads to a frequency offset error, which impacts the delay-coherence time. Further, clock offset and clock jitter can leads to missed samples, which impacts the delay-coherence time of the channel.

To further understand the nature of the delay-coherence time D_c, it is useful to consider two cases:

- Assume that the delay spread of the channel is significantly larger than 1 sample and all of the channel taps arise from the sum of many paths. (There is no dominant line-of-sight path) As these path-lengths change over time, all of the taps will get different values. In such a case, the delay-coherence time D_c and the traditional phase-coherence time N_c (or equivalently, the reciprocal of the bandwidth of the fading process) are the same since the relative phase response can change significantly in this time. This is the model that is likely to hold for wideband primary users being detected at significant distances from their transmitters. This basically corresponds to when ISI is significant.

- Now suppose there is only one dominant line-of-sight path that is arriving at the spectrum sensor, but is subject to local reflections in the near neighborhood of the sensor. (e.g. A user in an airplane) The delay spread of the channel can be smaller than 1 sample. In this case, there will just be a single tap in the filter. The amplitude might change rapidly since the local paths can go in and out of phase with each other, but the overall delay is not going to be changing as fast. In such cases, the delay-coherence time $D_c > N_c$ and the dominant effect may very well be the clock skew of the sensor's local oscillator relative to the primary oscillator. In such cases, the factor difference between D_c and N_c is comparable to the factor difference between the signal bandwidth and the carrier frequency.

3.3.4.2 Robustness analysis for the pure-delay case

For this simplest example of a fading process with finite D_c, the strategy of (3.14) is doomed to failure if the sums are taken beyond one coherence time. The signal will corrupt both the even and odd time samples eventually due to the random delay. This makes the empirical marginal distribution, when computed across several coherence blocks, of both the even and odd samples identical. By analogy with the case of matched filtering, a natural detection strategy is to compute the square of the test statistic within each coherence time and average that over multiple coherence times.

We now derive SNR wall expressions for a simple example signal $X[n]$ in which the odd samples $X_o[n]$ are *iid* BPSK data with average power 2, and the even samples $X_e[n]$ are identically zero. The behavior of the SNR wall for the general case is qualitatively similar. However, evaluating closed form expressions for the general case (with arbitrary distributions for $X_e[n]$ and $X_o[n]$) and for higher moments ($k \geq 2$ in (3.14)) is very tedious. The signal $X[n]$ can be described as

$$X[n] = \begin{cases} \pm\sqrt{2} & \text{with probability } \frac{1}{2}, \quad \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even.} \end{cases} \tag{3.15}$$

Theorem 3.3. *Let the received signal under hypothesis $\mathcal{H}_1$ in (3.13) be given by $Y[n] = X[n - l[n]] + W[n]$, with $X[n]$ as in (3.15). Let $W[n]$ be stationary uncertain noise indepen-*

dent of the signal. Assume that the distribution of the noise arises from the colored noise uncertainty set $\widehat{\mathbb{W}}_{\rho,\nu}$. *Suppose the delay fading satisfies:*

- $l[n]$ *is constant within a delay-coherence block (of length* D_c*) and* iid *across different delay-coherence blocks.*
- $l[n]$ *is uniformly distributed on* $\{0, 1\}$.

Define, $SNR = \frac{P}{\sigma_n^2}$. *Then, the detector with test-statistic given by*

$$T(\mathbf{Y}, D_c) = \frac{1}{M}\sum_{k=0}^{M-1}\left[\Big|\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\{|Y[kM+2n-1]|^2 - \sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}|Y[kM+2n]|^2\}\Big|^2\right], \quad (3.16)$$

where $N = MD_c$ *is non-robust if*

$$SNR \leq \frac{-2\left(\frac{\nu}{\rho}\right) + \sqrt{4\left(\frac{\nu^2}{\rho^2}\right) + 2D_c\left(\frac{\rho^2}{\nu^2} - \frac{\nu^2}{\rho^2}\right)}}{D_c}. \quad (3.17)$$

Proof. From (3.16) it is clear that

$$\mathbb{E}[T(\mathbf{Y}, D_c)|\mathcal{H}_i] = \mathbb{E}\left[\left|\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\left\{|Y[2n-1]|^2 - |Y[2n]|^2\right\}\right|^2 \Big|\mathcal{H}_i\right], \quad (3.18)$$

for $i = 0, 1$. Since we are computing the square of the test statistic within each coherence block, assume without loss of generality that the delay in the first coherence block is zero, i.e., $l[n] = 0$ for $n = 1, 2, \cdots, D_c$. The detector in (3.16) is non-robust if there exists a class of noise distributions such that the set of means under both hypotheses overlap. We now exhibit a specific class of noise distributions that satisfy this condition.

Assume $W[n]$ are *iid* random variables with marginals given by $W[n] \sim \mathcal{N}(0, \sigma^2)$ for some $\frac{\nu}{\rho} \leq \sigma^2 \leq \frac{\rho}{\nu}$. The following lemma gives the desired quantities.

Lemma 3.4.

$$\begin{aligned}\mathbb{E}[T(\mathbf{Y}, D_c)|\mathcal{H}_0] &= 4\sigma^4, \\ \mathbb{E}[T(\mathbf{Y}, N)|\mathcal{H}_1] &= 2P^2D_c + 8P\sigma^2 + 4\sigma^4, \end{aligned} \quad (3.19)$$

where $\sigma^2 = \mathbb{E}[|W[n]|^2]$.

Proof. See Appendix B.1. □

Using Lemma 3.4, it is clear that the detector in (3.16) is non-robust if the set of means under both hypotheses overlap. This condition is equivalent to

$$\begin{aligned} & \min_{\frac{\nu}{\rho}\sigma_n^2 \le \sigma^2 \le \frac{\rho}{\nu}\sigma_n^2} 2P^2 D_c + 8P\sigma^2 + 4\sigma^4 \\ \le \quad & \max_{\frac{\nu}{\rho}\sigma_n^2 \le \sigma^2 \le \frac{\rho}{\nu}\sigma_n^2} 4\sigma^4 \\ \Rightarrow \quad & P^2 D_c + 4\left(\frac{\nu}{\rho}\right) P\sigma_n^2 - 2\left(\frac{\rho^2}{\nu^2} - \frac{\nu^2}{\rho^2}\right)\sigma_n^4 \le 0 \end{aligned}$$

The last inequality in $SNR := \frac{P}{\sigma_n^2}$ can be solved to give the required result by the quadratic formula. □

This theorem suggests that it is the uncertainty corresponding to the frequency selectivity of the fading that induces limits on the detection of very weak signals that do not fully utilize all the degrees of freedom in time-domain. There are two interesting aspects of this result:

- The dependence of the SNR wall for this detector on the delay coherence time is $O(\frac{1}{\sqrt{D_c}})$. This gain from increasing delay-coherence time is slower than the $O(\frac{1}{N_c})$ gain for the matched filter, but it *is* better than the lack of any gain with coherence time for a primary signal that only uses half the degrees of freedom in the frequency domain [66].

- The SNR wall result so far does not depend on the color of the noise! The proposed detector in (3.16), while being intuitively very natural, is non-robust in the presence of frequency-selective fading, even when the noise is truly white (use $\nu = 1$ in (3.17)). As the next section shows, this is an artifact of the detection strategy used.

3.3.5 Noise calibration for cyclostationary signals

In matched filtering, we obtained the approximate level of noise from noise power measurements in an adjacent band. In the current example, the test statistic is a difference and so intuitively, the sum can provide access to the approximate level of noise. Explicitly, compute the empirical power in the received signal $\widehat{P} = \frac{1}{N}\sum_{n=1}^{N} |Y[n]|^2$. This converges to $P + \sigma^2$ under $\mathcal{H}_1$ and to σ^2 under $\mathcal{H}_0$ as $N \to \infty$, where $\sigma^2 = \mathbb{E}[|W[n]|^2]$. We normalize the received signal with this empirical power estimate and compute the new test statistic given in (3.20). Intuitively, one can think of this as passing the received signal through an automatic gain controller (AGC), which is tuned such that its output is normalized to unit power.

Theorem 3.5. *Let the received signal under hypothesis $\mathcal{H}_1$ in (3.13) be given by $Y[n] = X[n - l[n]] + W[n]$, with $X[n]$ given in (3.15), and $W[n]$ being independent of the signal*

with an unknown distribution from the colored noise uncertainty set $\widehat{\mathcal{W}}_{\rho,\nu}$. *Suppose the delay fading* $l[n]$ *satisfies:*

- $l[n]$ *is constant within a delay-coherence block and* iid *across different delay-coherence blocks.*
- *The length of the delay-coherence block is* $D_c > 2$.
- $l[n]$ *is uniformly distributed on* $\{0, 1\}$.

Then, the detector with test-statistic given by

$$\widetilde{T}(\mathbf{Y}, D_c) = \frac{1}{M}\sum_{k=0}^{M-1}\left[\Big|\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\{|\widehat{Y}[kM+2n-1]|^2 - \sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}|\widehat{Y}[kM+2n]|^2\}\Big|^2\right], \tag{3.20}$$

where $N = MD_c$ *and* $\widehat{Y}[n] = \frac{Y[n]}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}|Y[n]|^2}}$ *is non-robust if*

$$SNR \leq \left(\frac{\rho}{\nu}\right)\cdot\frac{2\left(1-\frac{1}{\sqrt{\nu}}\right)}{\sqrt{D_c}}. \tag{3.21}$$

Proof. Assume that a genie guarantees that the times when the delay $l[n]$ assumes an independent realization are integer multiples of D_c itself. From (3.20) it is clear that

$$\mathbb{E}[\widetilde{T}(\mathbf{Y}, D_c)|\mathcal{H}_i] = \mathbb{E}\left[\left|\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\left\{|\widehat{Y}[2n-1]|^2 - |\widehat{Y}[2n]|^2\right\}\right|^2\Big|\mathcal{H}_i\right], \tag{3.22}$$

for $i = 0, 1$. Since we are computing the square of the test statistic within each coherence block, assume without loss of generality that the delay in the first coherence block is zero, i.e., $l[n] = 0$ for $n = 1, 2, \cdots, D_c$. The detector in (3.20) is non-robust if there exist two distinct noise distributions $\mathcal{W}_1, \mathcal{W}_2 \in \widehat{\mathbb{W}}_{\rho,\nu}$ such that $\mathbb{E}_{\mathcal{W}_1}[\widetilde{T}(\mathbf{Y}, D_c)|\mathcal{H}_1] = \mathbb{E}_{\mathcal{W}_2}[\widetilde{T}(\mathbf{Y}, D_c)|\mathcal{H}_0]$. We now exhibit a specific class of noise distributions that satisfy this condition.

Let $M[n]$ be an *iid* white Gaussian noise process, with the variance $\sigma_m^2 \in [\frac{\nu}{\rho}\sigma_n^2, \frac{\rho}{\nu}\sigma_n^2]$. Let $W[n] = M[n] - \alpha M[n-1]$ for some α within some small range $0 \leq |\alpha| \leq \alpha_{max}$.

To apply the central limit theorem, define

$$(G_1, G_2) := \left(\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}|\widehat{Y}[2n-1]|^2, \sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}|\widehat{Y}[2n]|^2\right). \tag{3.23}$$

The parameters of the jointly Gaussian approximation for (G_1, G_2) are computed in the following lemma.

Lemma 3.6. *Assume $N \gg 0$. Then, the statistics of (G_1, G_2) under hypothesis $\mathcal{H}_1$ are given by*

$$\begin{aligned} \bar{m} &= \frac{1}{\kappa}\begin{pmatrix} \sqrt{\frac{D_c}{2}}(2P+\sigma^2) \\ \sqrt{\frac{D_c}{2}}\sigma^2 \end{pmatrix}, \\ \nu &= \frac{1}{\kappa^2}\begin{pmatrix} 8P\sigma^2+2\sigma^4 & 4\alpha^2\sigma_m^4\left[\frac{D_c-2}{D_c}\right] \\ 4\alpha^2\sigma_m^4\left[\frac{D_c-2}{D_c}\right] & 2\sigma^4 \end{pmatrix}, \end{aligned} \tag{3.24}$$

where $\kappa := \frac{1}{P+\sigma^2}$, $\sigma^2 = (1+\alpha^2)\sigma_m^2$, and

$$\begin{aligned} \bar{m} &:= \begin{pmatrix} \mathbb{E}G_1 \\ \mathbb{E}G_2 \end{pmatrix}, \\ \nu &:= \begin{pmatrix} Var(G_1) & Cov(G_1,G_2) \\ Cov(G_1,G_2) & Var(G_1) \end{pmatrix}. \end{aligned}$$

The corresponding statistics under the hypothesis $\mathcal{H}_0$ are computed by substituting $P = 0$ in (3.24).

Proof. See Appendix B.2. □

From (3.22) we have

$$\begin{aligned} \mathbb{E}[\widetilde{T}(\mathbf{Y}, D_c)|\mathcal{H}_1] &= \mathbb{E}[|G_1 - G_2|^2] \\ &= \mathrm{Var}(G_1) + \mathrm{Var}(G_2) - 2\mathrm{Cov}(G_1, G_2) + (\mathbb{E}G_1 - \mathbb{E}G_2)^2. \end{aligned} \tag{3.25}$$

Substituting (3.24) in (3.25) gives

$$\begin{aligned} \mathbb{E}[\widetilde{T}(\mathbf{Y}, D_c)|\mathcal{H}_1] &= \frac{4[(1+\alpha^2)2P + (1+\alpha^4)\sigma_m^2]\sigma_m^2}{(P+(1+\alpha^2)\sigma_m^2)^2} \\ &\quad + \frac{8\alpha^2\sigma_m^4\left(\frac{2}{D_c}\right) + \left(\frac{D_c}{2}\right)4P^2}{(P+(1+\alpha^2)\sigma_m^2)^2}. \end{aligned} \tag{3.26}$$

Setting $P = 0$ in (3.26) gives

$$\begin{aligned}\mathbb{E}[\widetilde{T}(\mathbf{Y}, D_c)|\mathcal{H}_0] &= \frac{4(1+\alpha^4) + 8\left(\frac{2}{D_c}\right)\alpha^2}{(1+\alpha^2)^2} \\ &= 4\left[1 - \frac{2\alpha^2\left(1-\frac{2}{D_c}\right)}{(1+\alpha^2)^2}\right] \\ &\leq 4, \end{aligned} \tag{3.27}$$

$\forall D_c > 2$. From (3.26) and (3.27), it is clear that the means of the test statistic under both hypotheses overlap if

$$\begin{aligned} 4 &\geq \frac{4[(1+\alpha_{max}^2)2P + (1+\alpha_{max}^4)\sigma_m^2]\sigma_m^2}{(P+(1+\alpha_{max}^2)\sigma_m^2)^2} \\ &\quad + \frac{8\alpha_{max}^2\sigma_m^4\left(\frac{2}{D_c}\right) + \left(\frac{D_c}{2}\right)4P^2}{(P+(1+\alpha_{max}^2)\sigma_m^2)^2} \\ \Rightarrow 0 &\overset{(a)}{\geq} -(1-\frac{2}{D_c})8\alpha_{max}^2\sigma_m^4 + \left(\frac{D_c}{2}-1\right)4P^2 \\ \Rightarrow SNR &\leq \frac{2\alpha_{max}}{\sqrt{D_c}}\left(\frac{\sigma_m^2}{\sigma_n^2}\right) \leq \left(\frac{\rho}{\nu}\right)\frac{2\alpha_{max}}{\sqrt{D_c}}. \end{aligned} \tag{3.28}$$

In the above chain of inequalities, (a) follows by completing squares on the right hand side.

To prove the result of the theorem we need to give conditions on α_{max} such that the distribution of the noise process $W[n]$ lies in $\widehat{\mathcal{W}}_{\rho,\nu}$.

Lemma 3.7. *Let $\alpha_{max} = 1 - \frac{1}{\sqrt{\nu}}$, where $\nu > 1$ is the parameter in the definition of the colored noise uncertainty set $\widehat{\mathcal{W}}_{\rho,\nu}$. Then the class of noise processes defined by $W[n] = M[n] - \alpha M[n-1]$ (for white $M[n]$ with variance $\sigma_m^2 \in [\frac{\nu}{\rho}\sigma_n^2, \frac{\rho}{\nu}\sigma_n^2]$) have distributions $W \in \widehat{\mathcal{W}}_{\rho,\nu}$ for all $0 \leq |\alpha| \leq \alpha_{max}$.*

Proof. See Appendix B.3. □

By Lemma 3.7, setting $\alpha_{max} = 1 - \frac{1}{\sqrt{\nu}}$ in (3.28) gives us the required result. □

3.4 Discussion and conclusions

3.4.1 Comparing different SNR walls

In this chapter we have shown the existence and computed the location of SNR walls for pilot detectors and histogram-based cyclostationary feature detectors (for the 50%-duty-cycle example), both with and without noise calibration. Figure 3.13 plots the SNR wall locations for the radiometer, the matched filter, and the 50% duty-cycle example with and without noise calibration as a function of the channel-coherence time N_c. In this example, the delay-coherence time is assumed to be the same as the phase-coherence time, i.e., $N_c = D_c$. In this plot we have assumed $\theta = 0.5$ (See (3.3))) for the matched filter. The noise uncertainty parameters for this Figure are $\rho = 10$ and $\nu = 10^{\frac{1}{10}}$.

In Figure 3.13, the curve with 'triangle' markers corresponds to the energy detection SNR wall, and this is independent of the channel-coherence time. The dashed curve corresponds to the matched filter. The dashed-dotted curved corresponds to the 50%-duty-cycle example without noise calibration (See (3.17)), and the solid curve represents the SNR wall for the 50%-duty-cycle example with noise calibration (See (3.21)). Notice that in Fig 3.13, the matched filter SNR wall is better than the SNR wall for the 50%-duty-cycle example with noise calibration for $N_c \geq 300$.

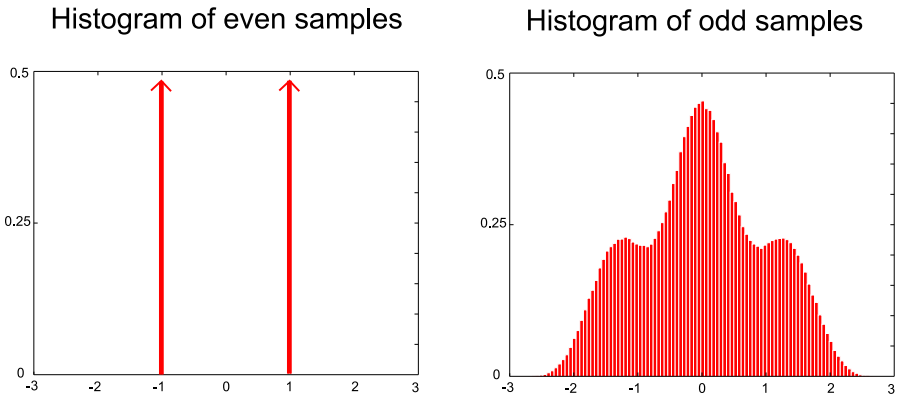
3.4.2 Concluding remarks

In this chapter we considered the problem of detecting signals containing known features. We considered signal features like pilot tones and cyclostationary features that exist in most communication signals. The presence of signal features allow for *coherent processing* as well as *run-time noise calibration*, which improves the robustness performance for the secondary user. However, the robustness gains are limited due to the uncertainty in the fading process. The presence of a known pilot tone improves the SNR wall by $\log_{10} N_c$ dB, where N_c is the minimum length of the phase coherence time. The presence of a cyclostationary signal feature improves the SNR wall by $\frac{1}{2} \log_{10} D_c$ dB, where D_c is the delay coherence time of the channel.

More subtly, pilot tones are non-robust even to uncertain flat fading processes, whereas cyclostationary signals are robust to such uncertainties. However, cyclostationary signals are non-robust to frequency-selective fading processes. Furthermore, we have shown that there exist scenarios in which the delay coherence time D_c can be much longer than the phase coherence time N_c.

In this chapter we considered commonly existing signal features like pilot tones and cyclostationary features, and focused on designing sensing algorithms that can detect these features. However, one can ask the following question: can we explicitly design robustly detectable signal features? This question is very useful and arises naturally in many spectrum

sharing scenarios. For instance, the primary user can put out a robustly detectable beacon to indicate that it is back in the band and that all secondary systems should vacate the band. On the flip side, the secondary users might be mandated to include a robustly detectable feature in their signals to help in spectrum enforcement. In Chapter 4, the ideas developed in this chapter are used to construct macroscale signal features which can be robustly detected at arbitrarily low SNRs.



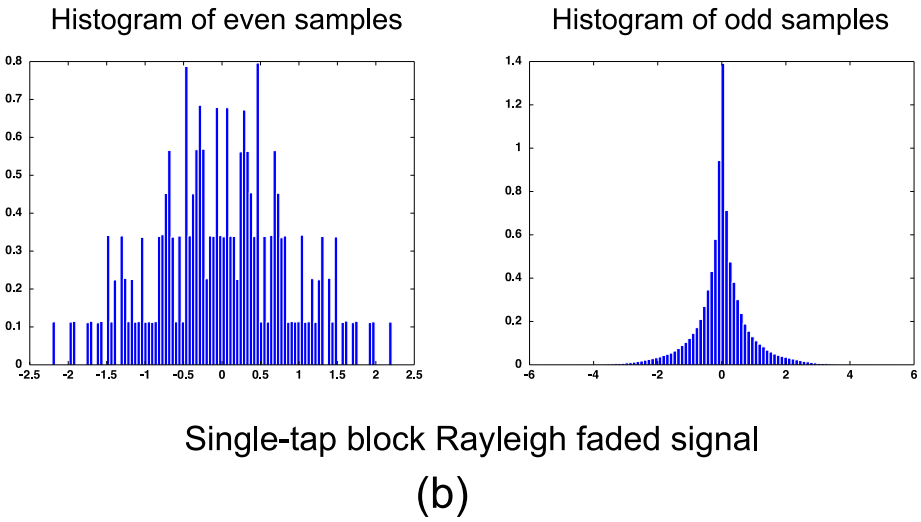


Figure 3.10: This figure illustrates the change in the empirical distribution of the even and odd samples of the signal $X_s[n]$ (see Eqn. (3.10)) when it is passed through a single-tap block Rayleigh fading process. The histograms in (a) show the empirical distributions when there is no fading. As expected, the even samples are *iid* Bernoulli random variables (BPSK data) and odd samples have a distribution different than a Bernoulli process. Figure (b) shows the histogram of the even and odd samples of the faded signal $h[n]X_s[n]$. Even after fading the two distributions are distinguishable. This shows that the signal X_s can be robustly detected under uncertain flat fading.

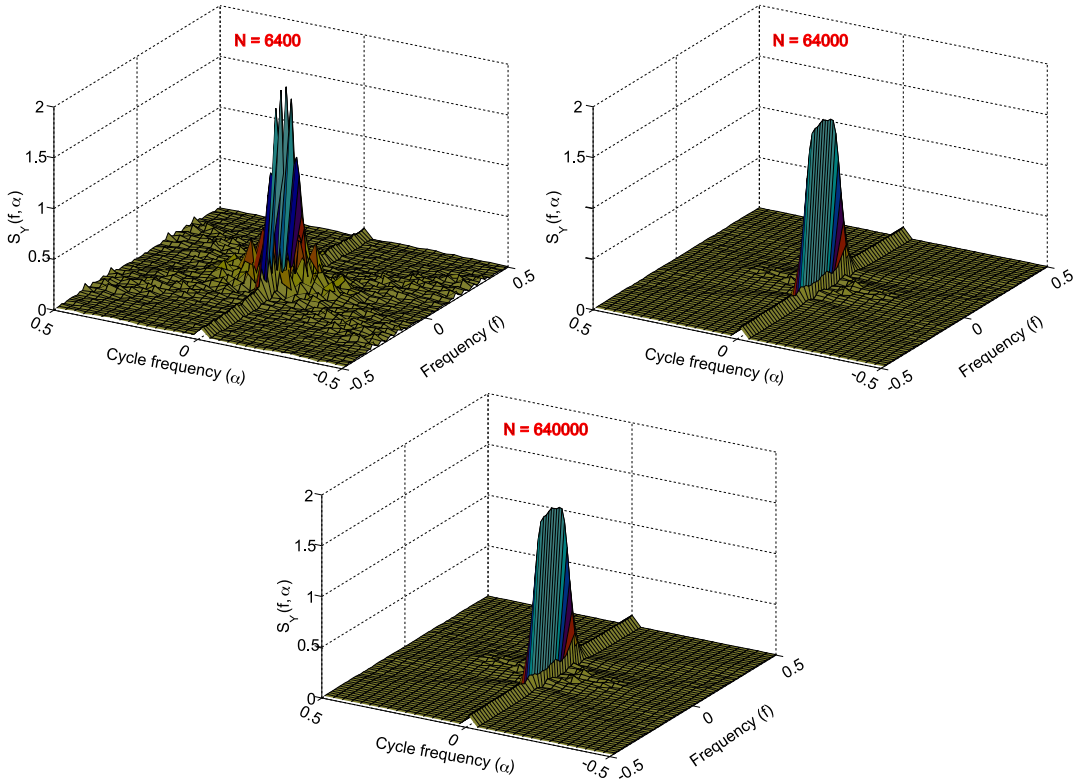
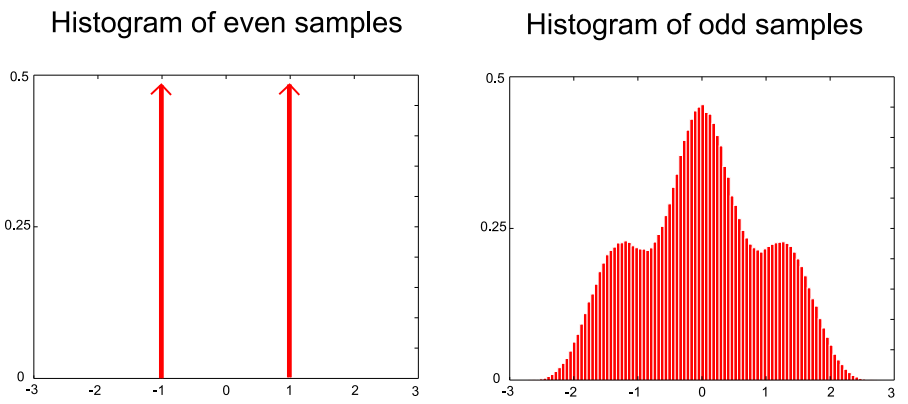
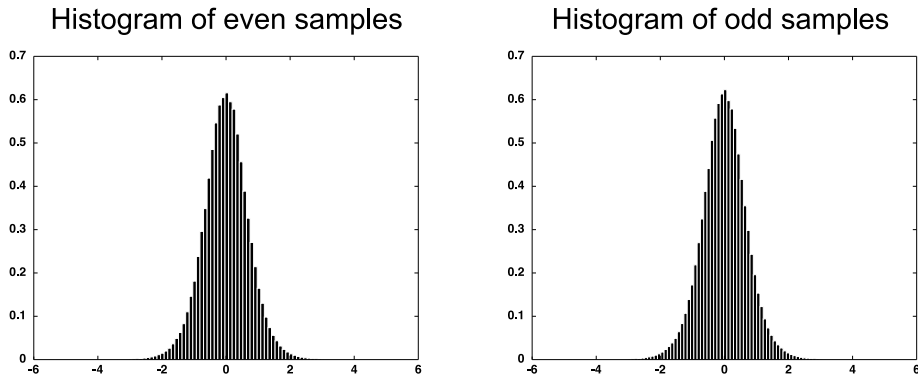


Figure 3.11: The empirical spectral correlation function $\widetilde{S}_Y^{\alpha}(f, N)$ for a single-cycle cyclostationary feature detector is illustrated in this figure. In this figure, the received signal $Y[n] = \sqrt{P}H(X[n]) + W[n]$, where the signal $X[n]$ is a raised-cosine function modulated with *iid* $BPSK$ symbols, the fading process $H(\cdot)$ is a 3-tap Gauss-Markov process, with the filter taps independent of each other, and $W[n]$ are samples of a white Gaussian noise process. This figures clearly shows that cyclostationary features are non-robust to frequency-selective fading processes. As you can see increasing the number of samples N destroys the signal feature.



BPSK modulated symbols over a Sinc pulse (before fading)

(a)



Two-tap block Rayleigh faded signal

(b)

Figure 3.12: This figure illustrates the change in the empirical distribution of the even and odd samples of the signal $X_s[n]$ (see Eqn. (3.10)) when it is passed through a two-tap block Rayleigh fading process. The histograms in (a) show the empirical distributions when there is no fading. As expected, the even samples are *iid* Bernoulli random variables (BPSK data) and odd samples have a distribution different than a Bernoulli process. Figure (b) shows the histogram of the even and odd samples of the faded signal $H(X_s[n])$, when $H(\cdot)$ is a two-tap block Rayleigh fading process. The distribution of the even and odd samples of the faded process are identical! This shows that the signal X_s is non-robust to uncertain frequency selective fading.

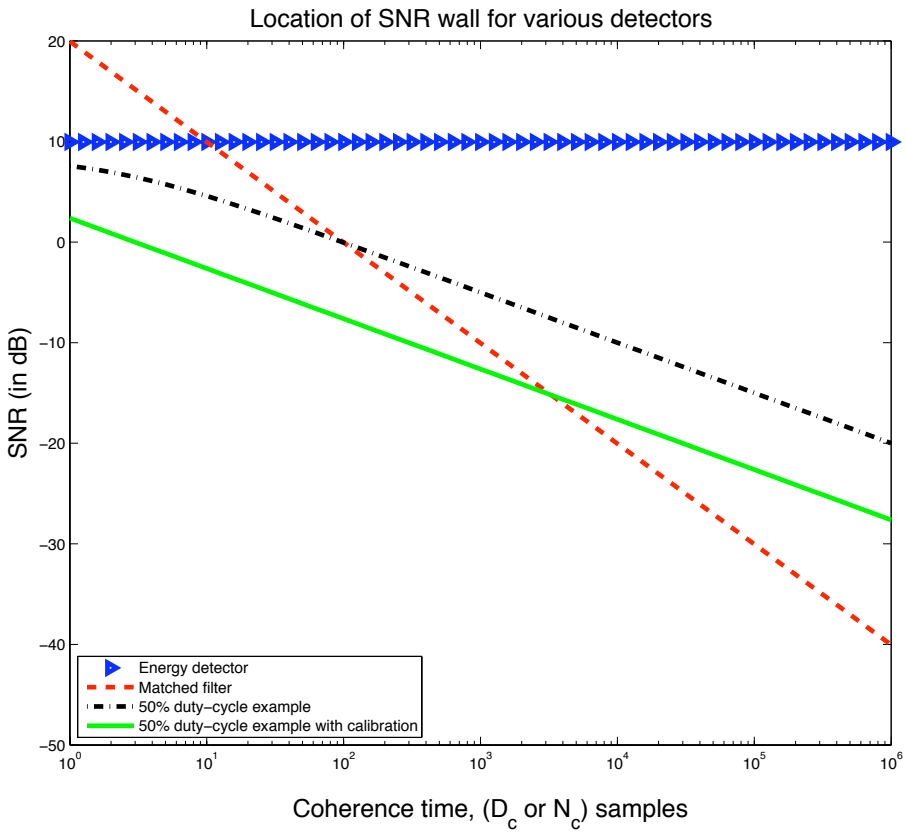


Figure 3.13: The location of SNR walls for various detection algorithms is plotted as a function of the delay-coherence time. The curve with 'triangle' markers (blue) is the location of the SNR wall for the energy detector. The dashed curve (red) is the location of the SNR wall for the matched filter. The dashed-dotted curve (black) is the location of the SNR wall for the 50%-duty-cycle example without noise calibration. Finally, the solid curve (green) is the location of the SNR wall for the 50%- duty-cycle example with noise calibration. These curves were obtained from Equations (2.8), (3.3), (3.17), and (3.21) respectively

Chapter 4

Overcoming SNR walls

4.1 Introduction

In Chapters 2 and 3 we derived robustness results for several sensing algorithms and showed that their robustness can be characterized by the location of their SNR walls. The robustness performance of a secondary sensor depends on two main factors — the amount of modeling uncertainty in the system, and the presence/absence of *known* signal features in the primary signal. The robustness of a secondary sensor can be improved by reducing the amount of modeling uncertainty in the system. For instance, the secondary user can build a receiver with expensive and finely calibrated circuit components[1]. On the other hand, the secondary user's robustness can also be improved by mandating the primary user to transmit known features in its signal. This loss of freedom for the primary user leads to improved robustness for the secondary user. Our aim in this chapter is to quantify the tradeoff between loss of freedom for the primary user and the gain in robustness for the secondary user.

The goal of a primary user is to communicate information to the primary receivers (for example, the TV transmitter broadcasts information to TV receivers) and it optimizes its system accordingly. Also, the primary transmitter wants the flexibility to change its signaling scheme over time. The goal of a secondary sensor is to robustly sense the primary transmitter's signal at low SNRs. The results in Chapter 2 show that complete flexibility of signaling for the primary transmitter implies that the secondary sensor cannot robustly sense the primary signal. Hence, the key question that the policymakers must address is: how much freedom should the primary sacrifice so that the secondary user can robustly sense it at low SNRs?

To technically study this problem we must model the cost of "loss in freedom" for the primary user. This leads to the following important philosophical question: "what forms of

[1]For example a radio astronomy receiver is designed such that the fluctuations in the noise power are significantly less than those in a cellular phone receiver.

loss in freedom are acceptable, and what are not acceptable?". For example, mandating the primary user to transmit a pilot tone is one way of restricting the primary's freedom (this can potentially leads to loss in primary data rate). However, this form of loss in freedom may be tolerable as the primary might need to transmit pilot signals for its own needs (to assist its receivers in channel equalization and timing/phase synchronization). On the other hand, restricting the primary user's signal to a narrow class of signaling schemes might be unacceptable, even if these schemes achieve high data rates.

In this chapter we quantify the pain due to loss in freedom for the primary system by the data rate it must sacrifice in order to provide robustness to the secondary system. This gives us a tradeoff between the primary's data rate and the secondary's robustness. We call this the *capacity-robustness* tradeoff[2]. The detailed technical assumptions in our formulation and some tradeoff curves are presented in Section 4.2. This formulation immediately prompts the following question: what is the optimal capacity-robustness tradeoff? We give an answer to this question in Section[3] 4.3. We show the existence of signaling schemes that can be robustly detected by the secondary sensor at arbitrarily low SNRs (SNR wall at negative infinity). Moreover, the loss in data rate for these schemes can be made arbitrarily small, which shows that the optimal capacity-robustness tradeoff is the trivial rectangle corresponding to *no* loss in primary data rate and complete robustness to the secondary user. However, for these signaling schemes, as the data rates approach the channel capacity for the primary system, the number of samples required for the secondary user to detect these signals approaches infinity, even at a finite SNR. This observations leads to the conjecture that the tradeoff between the primary capacity and the delay of the secondary user is a fundamental and non-trivial tradeoff. We comment on this in Section 4.4.

4.2 Capacity-robustness tradeoffs

In this section we formulate the capacity-robustness tradeoff problem. The primary performance is quantified using its achievable data rate and the secondary's robustness is quantified by the location of its SNR wall (see the results in Chapter 2 and Chapter 3).

In reality the primary transmitter also cares for robust detection for its own receivers, but at high SNR. The primary receivers must robustly detect the start of a primary packet, they must lock onto the phase of the primary signal and must also be able to time synchronize to the primary signal. Furthermore, they must also detect the pilot/training sequences in order to track the channel fading coefficients. There have been numerous papers [103; 104; 105; 106] and books [107; 108; 109; 110; 111; 112] written over the last four decades on robust techniques to achieve carrier and timing synchronization. The key difference between

[2]The capacity-robustness tradeoff problem was first proposed by us in [66]. The results in the first half of this chapter are mainly drawn from this paper.

[3]The discussion in this section is based on the results in [102].

robustness requirement for primary receivers and secondary sensors is the operating SNR. Primary receivers care about robust carrier and symbol synchronization at moderate to high SNRs (greater than 0 dB, whereas the secondary user must robustly sense the primary signal at low SNRs (of the order of -20 dB). The reason for this is that the primary transmitter cares about communicating information to its receivers, and communication over wireless fading channels at low SNRs is very inefficient. For instance, it is shown in [113] that the information-theoretic capacity of a very wideband CDMA system is inversely proportional to the number of resolvable multipaths, and in fact goes to zero if the number of paths is too large. The limitation comes from the fact that when there are many multipaths, it is very difficult to track the channel. Hence, the primary system will not want to operate at low SNRs, and so it does not care about robustness at low SNRs either. On the other hand the secondary sensor has to robustly detect weak undecodable primary signals.

4.2.1 Problem formulation

Figure 4.1 shows the block diagram for the capacity-robustness problem. The primary transmitter is communicating information to its receiver via the 'primary-primary' channel which we assume to be a bandlimited AWGN channel[4]. The capacity of this channel is $C = B \log(1 + SNR_P)$ assuming complex signaling. Here, B is the channel bandwidth and the primary signal to noise ratio $SNR_P = \frac{P_P}{N_0 B}$ where P_P is the received primary signal power at the primary receiver.

A secondary user measures the signal in the band and tries to sense for the presence/absence of the primary transmitter. We call the sensing channel between the primary transmitter and the secondary user the 'primary-secondary' channel. When the primary transmitter is ON, we assume that the 'primary-secondary' channel is a Rayleigh fading channel with additive noise. As the received primary signal strength at the secondary user, P_S can be very low, we assume that the secondary user cannot track the Rayleigh fading coefficients, and that it only knows a lower bound on the channel coherence time. Furthermore, we assume that the noise distribution at the secondary receiver is uncertain and lies in the noise uncertainty set described in Definition 3.1. Finally, we assume that the nominal noise power spectral density is given by N_0.

Given the above assumptions about the 'primary-primary' and 'primary-secondary' channel, our goal is to determine the tradeoff between the primary's data rate and the secondary's robustness as we vary the signaling scheme and the detection algorithm chosen by the primary transmitter and the secondary user respectively.

[4]The AWGN channel assumption is made to make the mathematical analysis simple. We can also consider fading in the primary-primary channel. In this case the capacity-achieving distribution might not be a Gaussian [114; 115; 116]. This can potentially give interesting results from the point of view of the capacity-robustness tradeoff. However, we do not explore this dimension in the thesis.

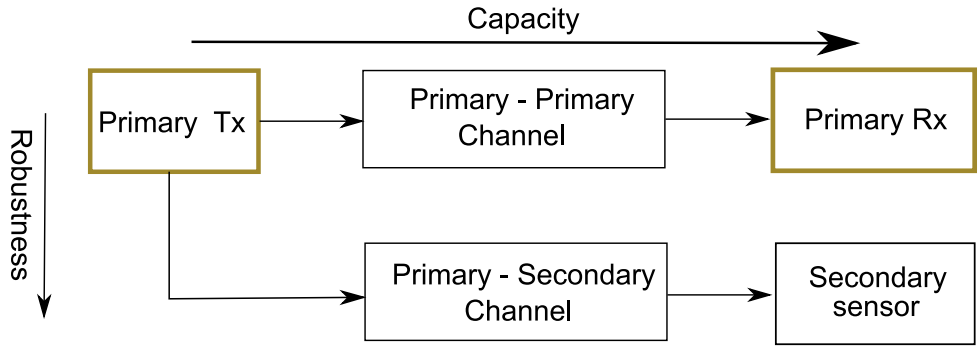


Figure 4.1: Block diagram showing the primary and secondary systems in the capacity-robustness problem formulation.

4.2.2 Capacity-robustness tradeoff: radiometric detection

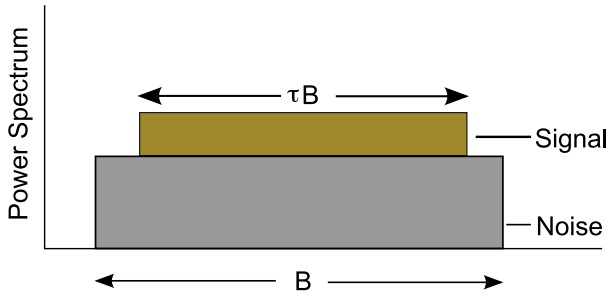


Figure 4.2: Primary transmitter signaling over a fraction of the total available degrees of freedom.

Suppose the spectrum sensor uses a radiometer to detect the primary signal. As seen in Section 2.3, the relevant SNR wall is given by $SNR_{wall}^{energy} = \frac{\rho^2-1}{\rho}$, with the nominal SNR at the secondary sensor given by $SNR_S = \frac{P_S}{N_0 B}$. It is clear that the common parameter that affects both the primary's data rate and the sensing SNR is the bandwidth B. The primary user can tradeoff its rate in exchange for robustness at the spectrum sensor by reducing the fraction of degrees of freedom used (see Figure 4.2). Let $B_{actual} = \beta B$, where $\beta \in [0, 1]$. We also assume that the fading at the spectrum sensor is bandlimited and has a Doppler bandwidth of $B_d = \frac{B}{N_c}$. In this case, the total bandwidth occupied by the primary signal at the spectrum sensor is $\beta B + \frac{B}{N_c}$. The effective SNR wall is given by $(\beta + \frac{1}{N_c})\left(\frac{\rho^2-1}{\rho}\right)$. Hence the capacity-robustness tradeoff for the simple radiometer is given by the parametric curve

$$\left[(\beta + \frac{1}{N_c})\left(\frac{\rho^2-1}{\rho}\right), \beta B \log\left(1 + \frac{P_P}{\beta N_0 B}\right)\right], \tag{4.1}$$

where $\beta \in [0,1]$. The first coordinate in (4.1) is the robustness of secondary detection and the second coordinate is an upper bound on the data rate achievable for the primary system. The factor $\beta + \frac{1}{N_c}$ in the robustness is the processing gain obtained because the primary does not use all the degrees of freedom.

One can also obtain noise calibration gains by taking measurements in the frequencies over which the primary does not spread its signal. Using the model in Section 3.2.2.4, the uncertainty in the noise is reduced from ρ to some smaller number λ. Therefore, the tradeoff with noise calibration is given by:

$$\left[(\beta + \frac{1}{N_c})\frac{2\rho}{\lambda}\left(\frac{\lambda^2 - 1}{\lambda}\right), \beta B \log\left(1 + \frac{P_P}{\beta N_0 B}\right)\right]. \tag{4.2}$$

4.2.3 Capacity-robustness tradeoff: coherent detection

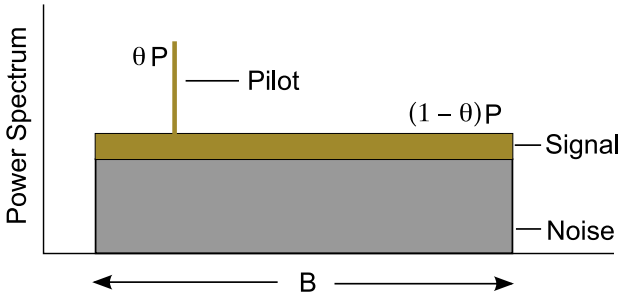


Figure 4.3: Primary transmitter signaling using a narrowband pilot tone with a fraction θ of the total power.

Assume that the spectrum sensor uses a coherent detector and the primary signal has a deterministic pilot tone as in Section 3.2.1. Any power spent on the pilot improves robustness but reduces the power available for communicating messages. Let θ be the fraction of total power spent in the pilot tone (see Figure 4.3). The primary data rate is bounded by $R_P = B \log\left(1 + \frac{(1-\theta)P_P}{N_0 B}\right)$. From (3.3), the SNR wall is given by $SNR_{wall}^{mf} = \frac{1}{N_c \cdot \theta}\left(\frac{\rho^2 - 1}{\rho}\right)$, where $\theta \in [0,1]$ and N_c is the channel coherence time. Thus, the capacity-robustness tradeoff for the matched filter is given by the parametric curve

$$\left[\frac{1}{N_c \cdot \theta}\left(\frac{\rho^2 - 1}{\rho}\right), B \log\left(1 + \frac{(1-\theta)P_P}{N_0 B}\right)\right], \tag{4.3}$$

where $\theta \in [0,1]$. The gains in robustness given in (4.3) are coherent processing gains. As discussed in Section 3.2.2.3, noise calibration can also be used to reduce the uncertainty down to the λ level. Hence, by (3.6) the capacity-robustness tradeoff with noise calibration

is given by

$$\left[\frac{2\rho}{\lambda \cdot N_c \cdot \theta}\left(\frac{\lambda^2-1}{\lambda}\right), B\log\left(1+\frac{(1-\theta)P_P}{N_0 B}\right)\right]. \tag{4.4}$$

4.2.4 Tradeoff curves

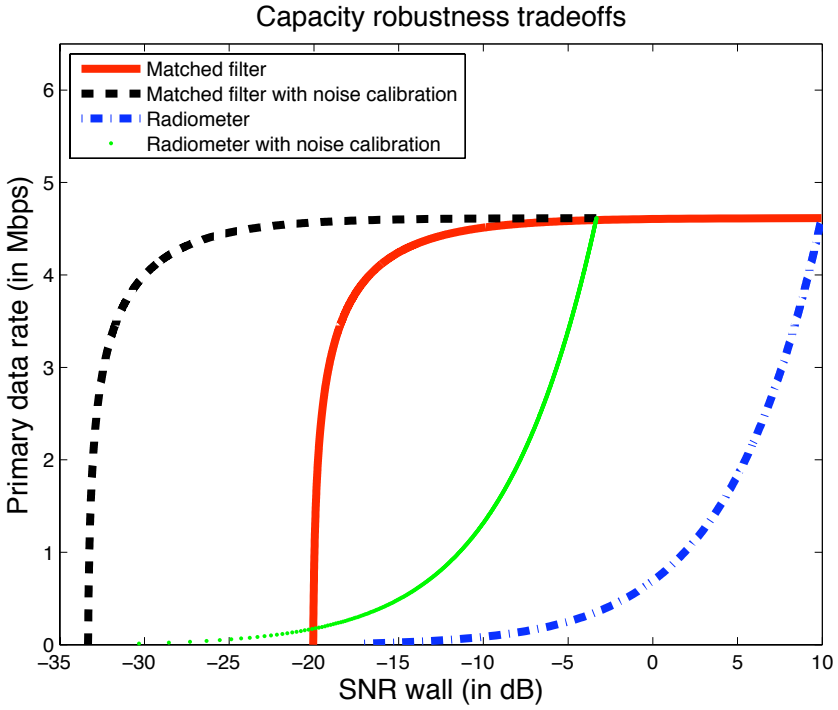


Figure 4.4: The capacity-robustness tradeoff curves for the radiometer and coherent detectors where the primary is operating at high spectral efficiency and large noise uncertainty ($\rho = 10, \lambda = 10^{0.1}$).

Figure 4.4 compares the capacity-robustness tradeoff curves for the radiometer and coherent detection. These curves were plotted using Equations (4.1), (4.2), (4.3) and (4.4). These plots correspond to a bandwidth B of 1 MHz, the signal to noise ratio of the primary channel $SNR_P = 20$ dB, the channel coherence time $N_c = 1000$, the noise uncertainty parameters $\rho = 10$ (to account for potentially significant interference) and $\lambda = 10^{0.1}$. Qualitatively, one can see that coherent detection is much better than the radiometer, i.e., for a given loss in primary data rate, it gives better robustness gains for the secondary system. Essentially, the primary system is not power limited at an $SNR_P = 20$ dB. So, sacrificing some power to the

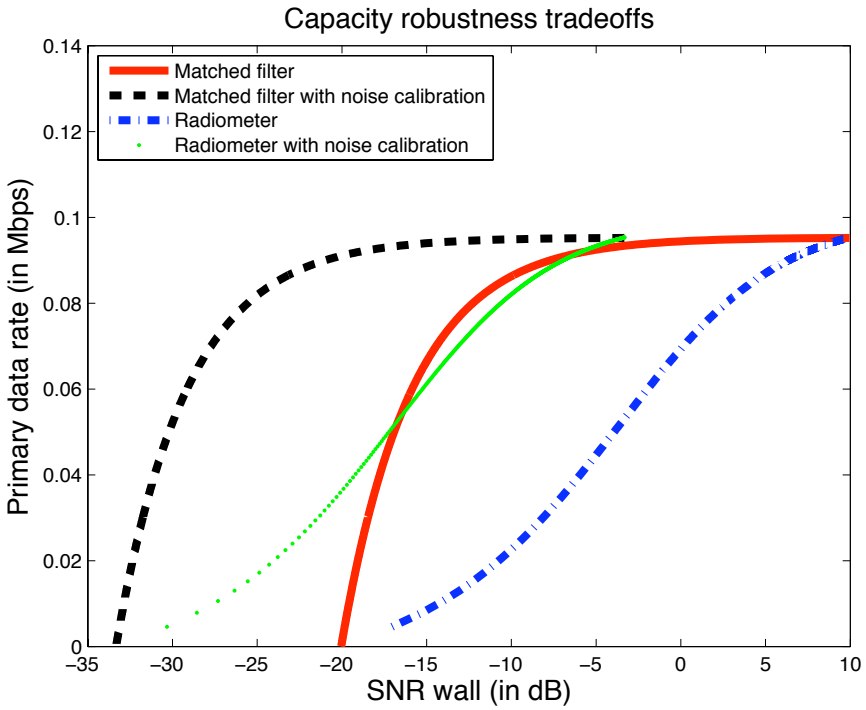


Figure 4.5: The capacity-robustness tradeoff curves for the radiometer and coherent detectors where the primary is operating at low spectral efficiency and large noise uncertainty ($\rho = 10, \lambda = 10^{0.1}$)

pilot tone is not a significant cost, but it gives significant robustness gains. However, shrinking the bandwidth and using fewer degrees of freedom to improve radiometer robustness is painful since it is degrees of freedom that are the critical resource for high SNR primary users.

Figure 4.5 compares the capacity-robustness tradeoffs for the radiometer and coherent detection when the primary SNR_P is low. These tradeoffs were plotted for $SNR_P = -10$ dB. The remaining parameters are the same as in Figure (4.4). Since the primary is power limited, the loss of power to the pilot has a non-trivial rate penalty. Meanwhile, giving up degrees of freedom incurs less of a penalty to help the radiometer. The robustness curves for the two strategies come closer to each other, but deterministic pilot tones and coherent detection still win out. Also, both Figure 4.4 and 4.5 are plotted for the case when the ratio of the noise uncertainty parameters $\rho = 10$ and $\lambda = 10^{0.1}$ is large. In this case it is clear that run-time noise calibration gives significant gains to both the radiometer and coherent detector.

Figure 4.6 and 4.7 plot the capacity robustness tradeoff curves for the radiometer and coherent detection when the ratio of the noise uncertainty parameters $\rho = 10^{0.2}$ and $\lambda = 10^{0.1}$ is small. As before the two figures correspond to high and low spectral efficiency for the primary system.

Finally, in order to make a fair comparison, the nominal SNRs in all these plots correspond to the worst possible noise power within the noise uncertainty set as opposed to the middle noise power. This way, it is clear that after noise calibration, it makes no difference what the original ρ was. But noise-calibration makes less of a difference when λ is very close to 1.

4.3 Overcoming SNR walls

In Section 4.2 we have derived some achievable capacity-robustness tradeoff curves. A natural question one can ask is: what is the optimal capacity-robustness tradeoff curve? In this section we construct primary signaling schemes with rates arbitrarily close to the AWGN channel capacity, while they are also robust to uncertainties in the noise and fading processes at arbitrarily low signal to noise ratios. That is, we construct signals with no SNR wall limitations and with primary data rates close to the channel capacity. This result shows that the optimal capacity-robustness tradeoff is a trivial tradeoff!

In reality, we must also consider delay into the equation. How long does the secondary sensor take to detect the presence/absence of the primary signal? We know that the sensing time is a function of the received SNR at the secondary sensor. However, is it also dependent on the particular signaling scheme chosen by the primary transmitter. In fact, for the example given in this section the sensing time for the secondary approaches infinity as the primary data rate approaches its capacity. This prompts us to consider the tradeoff between the

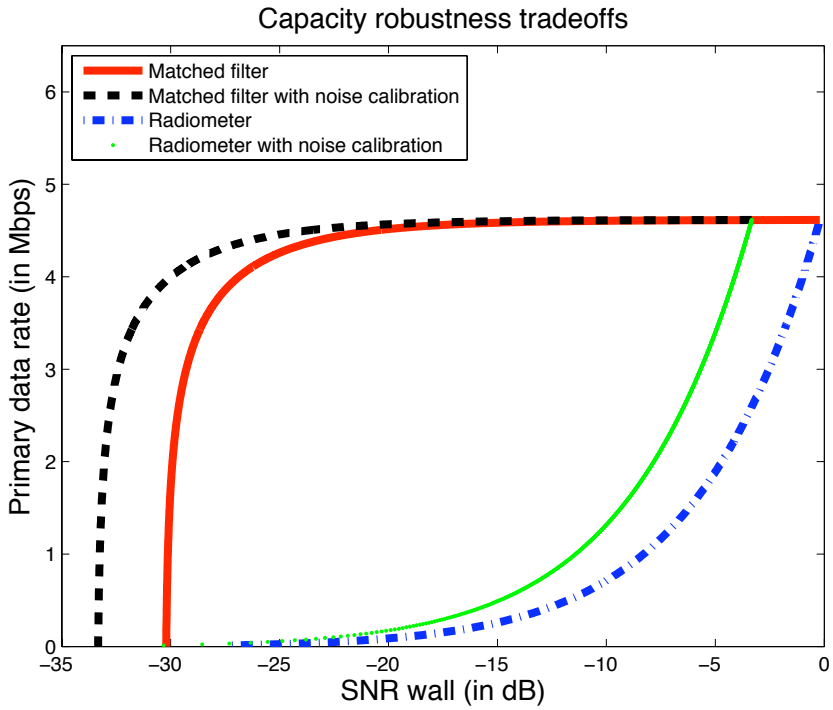


Figure 4.6: The capacity-robustness tradeoff curves for the radiometer and coherent detectors where the primary is operating at high spectral efficiency and moderate noise uncertainty ($\rho = 10^{0.2}, \lambda = 10^{0.1}$)

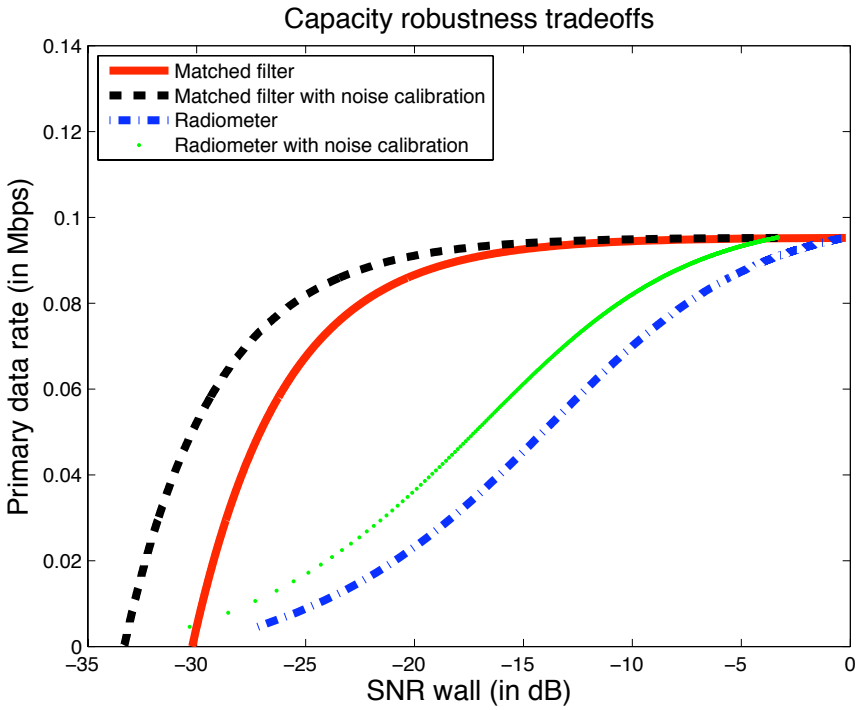


Figure 4.7: The capacity-robustness tradeoff curves for the radiometer and coherent detectors where the primary is operating at low spectral efficiency and moderate noise uncertainty ($\rho = 10^{0.2}, \lambda = 10^{0.1}$)

capacity of the primary user and the sensing delay for the secondary user. We comment on this tradeoff in Section 4.4.

4.3.1 Modeling philosophy

To show that a signal can be robustly detected at arbitrarily low SNRs, it is no longer sufficient to work with the bounded moment noise uncertainty model given in Definition 3.1. The reason for this is that the uncertainties in real life are worse than the ones captured in the bounded-moment model. Hence, we propose an *arbitrarily varying* uncertainty model, and use this to prove our robustness results. The idea is that if a signal can be robustly detected under an adversarial uncertainty (arbitrarily varying uncertainty model), then it must be robust to the uncertainties in practice.

Figure 4.8 pictorially illustrates our modeling philosophy. Let the 'green' shaded region in the figure denote the set of all uncertainty model in reality, the innermost 'tan' region in the figure correspond to the bounded-moment uncertainty model, and the outermost region denote the arbitrarily varying uncertainty model. As shown in the Figure, the bounded-moment model is a very conservative uncertainty model, the arbitrarily varying model is a very pessimistic model for noise uncertainty, and the actual uncertainties encountered in practice are somewhere in between these two models. The main idea is that if detectors have SNR wall limitations under the conservative uncertainty model ('bounded moment'), then their performance will only be worse in real life. Hence, the SNR wall results are of significant practical importance. Similarly, if a signal is robust to arbitrary variations in the noise process, then it will definitely be robust to uncertainties in real life.

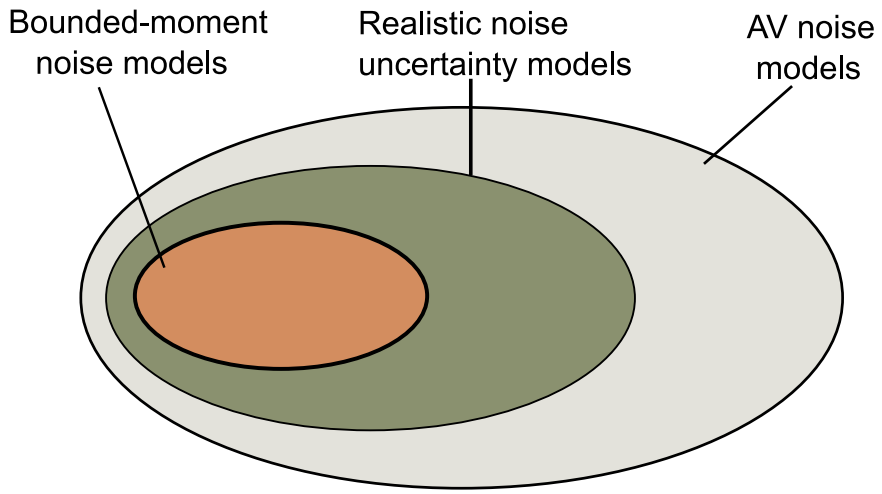


Figure 4.8: An illustrative picture to capture the modeling philosophy behind the uncertainty model used in this Chapter.

4.3.1.1 Arbitrarily varying uncertainty model

We now define the set of possible noise and fading processes within the arbitrarily fading uncertainty model:

- We model noise as an arbitrarily varying sequence $\{W[n]\}_{n=1}^{\infty}$ with a finite empirical second moment. That is, $\frac{1}{N}\sum_{n=1}^{N}|W[n]|^2 =: \sigma^2$, for some unknown $\sigma > 0$.
- We model the fading process as an L-tap linear time-varying filter with coefficients $h_l[n]$, where the fading coefficients are arbitrarily varying but are normalized to satisfy $\sum_{l=0}^{L}|h_l[n]|^2 = 1$ for all $n = 1, 2, 3, \cdots$. In other words we are assuming that most of the energy in the received signal is coming from paths than have a delay falling within the first L filter taps. This modeling assumption on the fading process is not unrealistic if we choose L sufficiently large.
- We also make some boundedness assumptions on the fourth moments of the noise and fading processes. The exact form of these assumptions are given in Theorem 4.1.
- Finally, we assume that the primary transmitter and the secondary sensor have access to common randomness.

The arbitrarily varying uncertainty model described above is similar to the ones used in the *arbitrarily varying channel* (AVC) literature [117; 118; 119]. The AVC is an information-theoretic model for communication in the presence of a malicious interferer or a jammer, whose objective is to minimize the reliable rate of communication of the link it is trying to jam. For our problem, a malicious-jammer model for the uncertainty in noise and fading is probably pessimistic. Realistically, noise and fading processes are approximately stationary and they might have time-varying fluctuations. Hence, they cannot be modeled by any stationary, ergodic class of uncertain distributions. We tackle this problem by treating the noise and fading process as potentially malicious, but to counter them we give the primary transmitter and secondary sensor access to common randomness. The idea of sharing a source of common randomness between the encoder and decoder allowing them to randomize their coding strategy is well known in the AVC literature [117; 119]. Recent results have also shown the capacity of AVCs under limited common randomness constraints [120].

4.3.2 Signals with Macroscale features

We start off by constructing a signal which can be robustly detected at arbitrarily low signal to noise ratios. However, the infinite robustness gain is achieved at the cost of reducing the primary data rate by a constant factor. The key idea is to introduce *macroscale* features in signals, which in turn enable robust detection even in the presence of uncertainties. The idea of introducing macroscale features for robust detection was first proposed by us in [102].

Subsequently, this result lead to the idea of using macroscale features to construct codes that provide robust identity in wireless systems [121; 90; 6].

Let $X[n]$ be a discrete time signal given by

$$X[n] = \begin{cases} \pm\sqrt{2P} & \text{w.p. } \frac{1}{2} \quad \text{if} \quad \lceil \frac{n}{K} \rceil \quad \text{is odd,} \\ 0 & \quad\quad\quad\quad \text{if} \quad \lceil \frac{n}{K} \rceil \quad \text{is even.} \end{cases}, \tag{4.5}$$

for some $K \gg L$. Time is divided into blocks of size K, the odd numbered blocks contain *iid* Bernoulli $\pm\sqrt{2P}$ random variables, and the even numbered blocks have zeros. We call the former 'data' blocks and the latter 'silence' blocks. Figure 4.9 pictorially describes the construction of the signal $X[n]$.

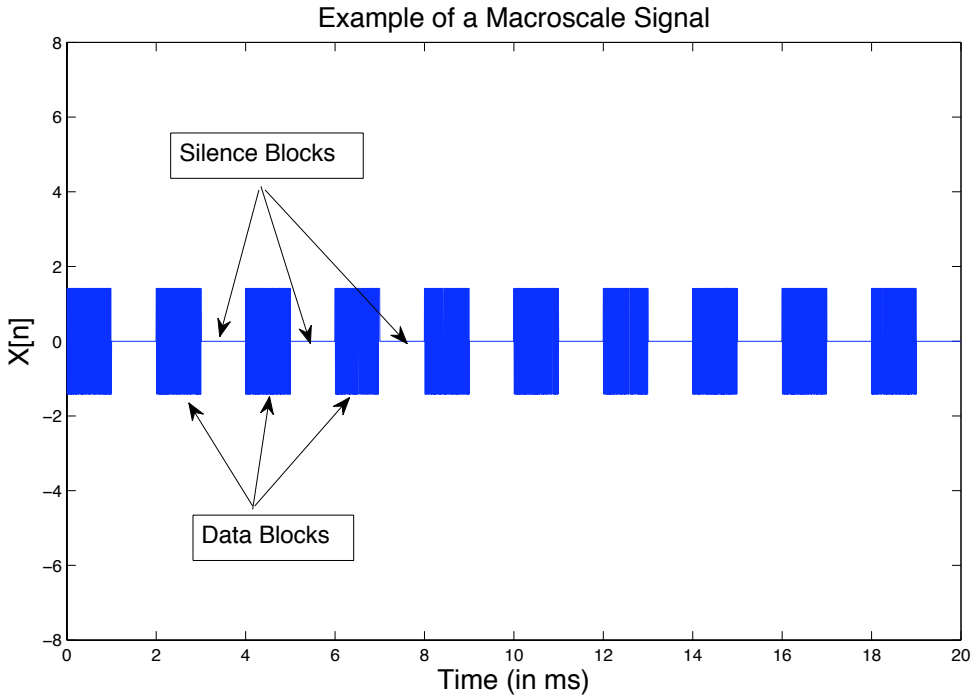


Figure 4.9: This figure pictorially shows the signal described in Equation (4.5). The solid regions correspond to the "data" blocks, which contain *iid* BPSK($\pm\sqrt{2P}$) signal samples, and the empty regions correspond to the "silence" blocks, and they contain zeros.

It can easily be verified that $X[n]$ is a cyclostationary signal and it exhibits cyclostationary features described in Chapter 3. The key difference between this signal and the ones discussed in Chapter 3 is that the scale of the feature in $X[n]$ is much larger than the delay spread of the channel. This is true because the width of the ON/OFF periods in $X[n]$ is K,

which is chosen to be larger than L (the first L taps contain more than half of the signal energy). For this reason, we say that $X[n]$ has a macroscale feature (relative to the delay spread of the fading channel).

When $K > L$, even an arbitrarily varying fading process cannot completely mix the 'data' blocks and the 'silence' blocks. So, the faded signal will have some parts that are guaranteed to have lower average power than the other parts of the signal. Moreover, these lower power blocks occur periodically in time. This means that we can run a detector that searches for these periodic fluctuations in the signal power to detect the signal in noise. Note that the number of alternatives in the search does not grow exponentially with the sensing time, and hence our algorithm will successfully detect the signal given a sufficiently long sensing duration[5]. Also, observe that the important thing is the width of the fluctuation in the signal power rather than the absolute value itself. Hence, this signal can be robustly distinguished from noise at arbitrarily low signal to noise ratios.

Notice that for the above result to be true we must make the assumption that the noise process is stationary and hence its power does not vary with time. In Section 4.3.3 we relax the stationarity assumption in the noise process.

The signal $X[n]$ has no SNR wall limitation, but the signal uses only half the degrees of freedom. So, even though this achieves complete robustness it achieves only half the channel capacity to the primary user. Hence, this example does not achieve the best possible tradeoff point in the capacity-robustness space.

4.3.3 Macroscale feature example: mixture of Gaussian signals

We now give another example of a signal that contains macroscale features, and which also makes better use of the degrees of freedom. Define $\widetilde{X}[n]$ to be

$$\widetilde{X}[n] = \begin{cases} \mathcal{N}(0, P+\epsilon) & \text{if } \lceil \frac{n}{K} \rceil \text{ is odd,} \\ \mathcal{N}(0, P-\epsilon) & \text{if } \lceil \frac{n}{K} \rceil \text{ is even.} \end{cases}, \tag{4.6}$$

for some $0 < \epsilon < P$. This signal is obtained by interleaving two *iid* Gaussian signals with power $P+\epsilon$ and $P-\epsilon$ respectively. A candidate detection algorithm for this example is to compute the difference in the average power between the odd numbered blocks and the even numbered blocks and compare it to a threshold. If the received signal is stationary noise, then this difference converges to zero as N increases to infinity. If the received signal

[5]The faded signal $H(X[n])$ has silence blocks of length at least $K-L$ occurring periodically every K samples. As there can be a time-offset between the secondary sensor and the primary transmitter, the secondary sensor has to run K parallel detectors each with a different time-offset. This will ensure that at least one of these K detectors will align with the periodic silences perfectly. The P_{FA} and P_{MD} of each of the parallel detectors can be made arbitrarily small by increasing the sensing time N. As the number of parallel detectors is fixed, the overall P_{FA} and P_{MD} will also decrease as N increases.

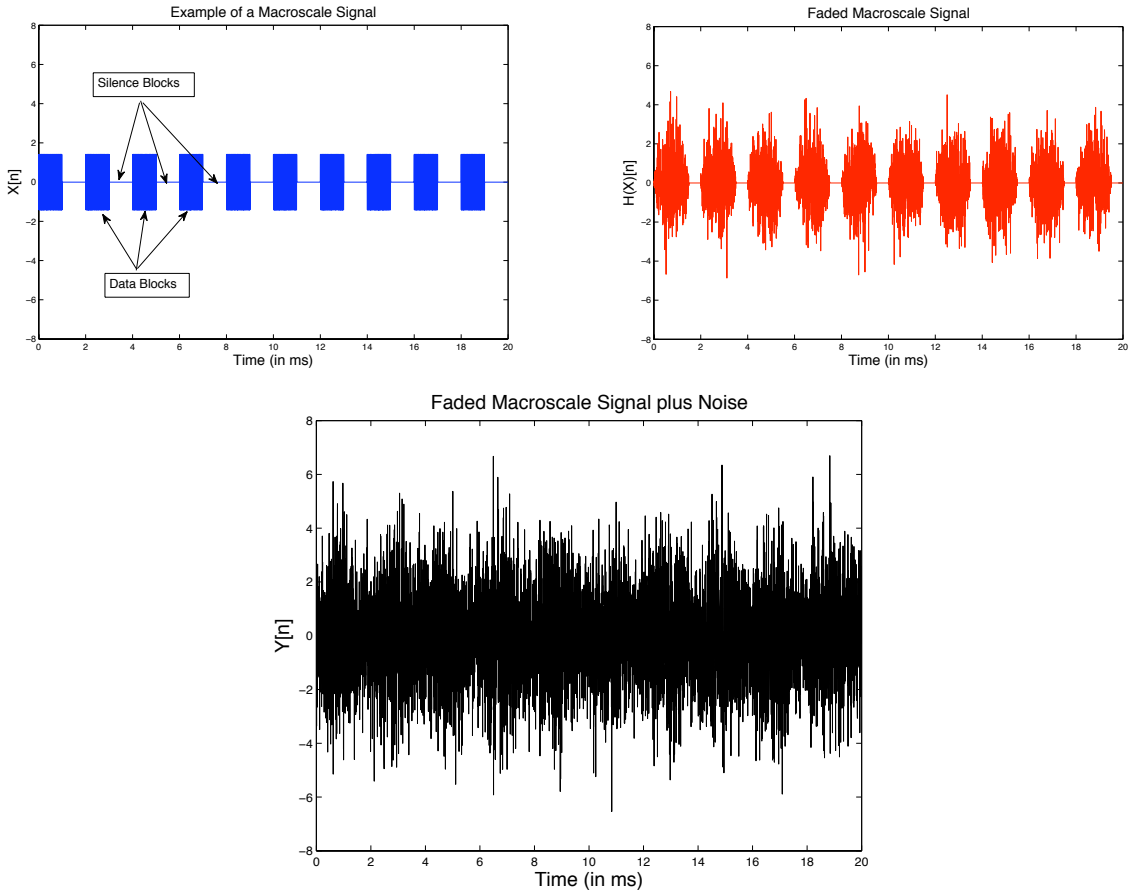


Figure 4.10: This figure shows the transmitted signal $X[n]$, faded version of the signal $H(X)[n]$, and the received signal $Y[n] = \sqrt{P}H(X[n]) + W[n]$. The fading process $H(\cdot)$ is assumed to be a linear time-varying filter with L taps. From the figure it is clear that at certain time instances, the fading mixes with all zeros in the signal and so the output is also zero. Hence, in these time slots we can have access to noise even under hypothesis $\mathcal{H}_1$.

contains $\widetilde{X}[n]$, then the difference in powers converges to a non-zero constant proportional to ϵ. Hence, this signal can be robustly distinguished from stationary noise.

However, if the noise itself has a time-varying variance then it is possible that the difference in the average power between the odd numbered blocks and the even numbered blocks is non-zero even under hypothesis $\mathcal{H}_0$. So, robust detection is not guaranteed at low SNRs. The possibility of noise having time-varying variance is not unrealistic because noise also contains interference, which is due to man-made signals transmitted by other users over the wireless medium.

One way to overcome this problem is to choose a complex signature that does not occur either in nature or in other licensed systems. This is analogous to the idea proposed by Carl Sagan that space-aliens would signal to us by modulating the signal with prime numbers that could not occur by any natural process unlike pure pilot tones [122; 123]. The complex signature is used to choose the blocks to transmit a Gaussian signal with power $P+\epsilon$, while the remaining blocks are used to transmit a Gaussian signal with power $P-\epsilon$. To model this different level of complexity, we make a technical assumption of common randomness to capture codes that are effectively random when viewed from the perspective of natural effects. Specifically, we assume common randomness between the secondary detector and the primary transmitter. This allows the secondary detector to know the random partitioning of blocks created by the primary transmitter. The detailed description of the construction of a common-randomness based signal with macroscale features is given below.

Let $X[n]$ denote the signal we are trying to construct. Divide time into blocks of length $K = M + L - 1$ for some $M > 0$. Rewrite the sequence $\{X[n]\}_{n=1}^{\infty}$ as $\{\mathbf{X}_i\}_{i=1}^{\infty}$, where $\mathbf{X}_i \in \mathbb{R}^{1\times K}$ is a K dimensional row vector given by $\mathbf{X}_i = (X[(i-1)K+1], X[(i-1)K+2], \cdots, X[(i-1)K+K])$. Let $\{B_i\}_{i=1}^{\infty}$ be a sequence of *iid* Bernoulli-$\frac{1}{2}$ random variables taking values in the set $\{0,1\}$. Think of this Bernoulli sequence as labeling the sequence $\{\mathbf{X}_i\}_{i=1}^{\infty}$ by '1's and '0's.

Given the Bernoulli sequence $\{B_i\}_{i=1}^{\infty}$, define

$$\mathbf{X}_i = \begin{cases} \mathcal{N}(\mathbf{0}, (P+\epsilon)\mathbf{I}_{K\times K}) & \text{if} \quad B_i = 1 \\ \mathcal{N}(\mathbf{0}, (P-\epsilon)\mathbf{I}_{K\times K}) & \text{if} \quad B_i = 0 \end{cases}, \tag{4.7}$$

where $\mathcal{N}(\mathbf{m}, \Lambda_{K\times K})$ denotes a K-dimensional Gaussian random vector with mean $\mathbf{m}$ and covariance matrix $\Lambda_{K\times K}$. Here ϵ is chosen such that $0 < \epsilon < P$, where P is the average power of the signal. Intuitively, the signal $\{\mathbf{X}_i\}_{i=1}^{\infty}$ is a mixture of two *iid* random Gaussian processes, one with power $(P+\epsilon)$ and another with power $(P-\epsilon)$. The choice of the block length K ensures that fading cannot completely mix the two Gaussian processes.

Assume that the same Bernoulli sequence $\{B_i\}_{i=1}^{\infty}$ is also available at the detector. This can be thought of as common randomness between the primary transmitter and the secondary detector. Given this common randomness, the detection algorithm can be described as follows:

- For any K-dimensional vector $\mathbf{X} := (X[1], X[2], \cdots, X[K]) \in \mathbb{R}^{1\times K}$, $K > L$, define the function $\kappa(\mathbf{X}) := \frac{1}{K-L+1}\sum_{k=L}^{K} |X[k]|^2$. This function computes the empirical power (excluding the first $L-1$ samples) of the vector $\mathbf{X}$. excluding the first $L-1$ samples is to prevent fading from mixing samples within two adjacent blocks. This property will be used in the proof of Theorem 4.1.

- Assume that N is an integer multiple of K, i.e., $N = K \cdot G$ for some integer $G > 0$. Rewrite the received sequences $\{Y[n]\}_{n=1}^{N}$ as a sequence of K-dimensional vectors $\{\mathbf{Y}_i\}_{i=1}^{G}$, where $\mathbf{Y}_i = (Y[(i-1)K+1], Y[(i-1)K+2], \cdots, Y[(i-1)K+K])$.

- Compute

$$\begin{aligned} T_1 &:= \frac{1}{G}\sum_{i=1}^{G} B_i \cdot \kappa(\mathbf{Y}_i); \\ T_2 &:= \frac{1}{G}\sum_{i=1}^{G} (1-B_i) \cdot \kappa(\mathbf{Y}_i) \end{aligned} \tag{4.8}$$

  T_1 and T_2 are the empirical average of the received signal powers in the blocks labeled '1' and '0' respectively.

- The detector is given by

$$T\left(\{\mathbf{Y}_i\}_{i=1}^{G}\right) := T_1 - T_2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \tag{4.9}$$

  where $\lambda > 0$ is the detection threshold.

The intuitive justification for choosing the test-statistic in (4.9) is as follows: if the signal is present, the difference in powers in the blocks labeled '1' and the blocks labeled '0' converges in probability to a non-zero number as $G \to \infty$. On the other hand, if the signal is absent, this difference converges to zero as $G \to \infty$.

Figure 4.11 pictorially describes the signal transmitted $X[n]$, the received signal $Y[n] = H(X[n]) + W[n]$, and the test-statistic of the detection algorithm given in (4.9).

4.3.3.1 Robustness analysis

Theorem 4.1. *Consider the robust hypothesis testing problem in (2.1). Assume that the sensing time can be written as $N = K \cdot G$, for some positive integers K and G. Let $H(X)[n] = \sum_{l=1}^{L-1} h_l[n]X[n-l]$, for some integer $K > L > 0$. For any K-dimensional vector $\mathbf{W} = (W[1], W[2], \cdots, W[K])$, and a $L \times K$ matrix $\mathbb{H} = \{h_l[k]\}$, $l = 0, 1, \cdots L-1$ and $k =$*

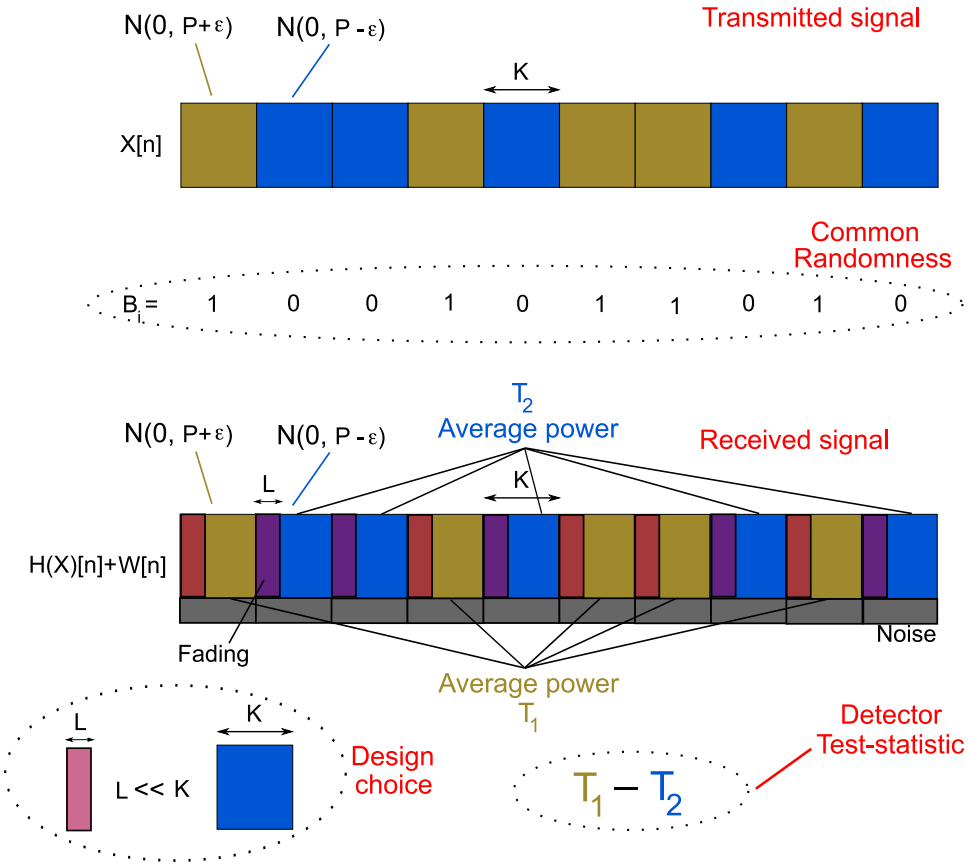


Figure 4.11: This figure illustrates the construction of the signal $X[n]$, and also shows how the signal can be distinguished in the presence of arbitrarily varying noise and fading processes.

$1, 2, \cdots, K$, *define*

$$\kappa(\mathbf{W}) := \frac{1}{K-L+1}\sum_{k=L}^{K}|W[k]|^2$$

$$\kappa(\mathbb{H}) := \frac{1}{K-L+1}\sum_{k=L}^{K}\sum_{l=0}^{L-1}|h_l[k]|^2.$$

Assume that the fading and noise processes satisfy the following constraints.

- $\frac{1}{G}\sum_{i=1}^{G}\kappa(\mathbf{W}_i) = \sigma^2$, *for some unknown constant* $\sigma > 0$.
- $\frac{1}{G}\sum_{i=1}^{G}\kappa(\mathbb{H}_i) = 1$.
- *There exist known constants* $G_0 > 0$, *and* $C_i > 0$, $i = 1, 2, 3$, *such that for all* $G > G_0$,

we have

$$\begin{aligned} \frac{1}{G}\sum_{i=1}^{G}\kappa^2(\mathbf{W}_i) &\leq C_1 \\ \frac{1}{G}\sum_{i=1}^{G}\kappa^2(\mathbb{H}_i) &\leq C_2 \\ \frac{1}{G}\sum_{i=1}^{G}\kappa(\mathbf{W}_i)\cdot\kappa(\mathbb{H}_i) &\leq C_3, \end{aligned} \tag{4.10}$$

where $\mathbf{W}_i := (W[(i-1)K+1], W[(i-1)K+2], \cdots, W[(i-1)K+K])$*, and* $\mathbb{H}_i$ *is a* $L \times K$ *matrix with* $\mathbb{H}_i := \{h_l[(i-1)K+k]\}$*,* $l = 0, \cdots, L-1$*, and* $k = 1, \cdots, K$*.*

Let $\{B_i\}_{i=1}^{G}$ *be a sequence of Bernoulli-*$\frac{1}{2}$ iid *random variables taking values in the set* $\{0, 1\}$*. Rewrite the signal samples* $\{X[n]\}_{n=1}^{N}$ *as* $\{\mathbf{X}_i\}_{i=1}^{G}$*, where* $\mathbf{X}_i := (X[(i-1)K+1], \cdots, X[(i-1)K+K])$*, and*

$$\mathbf{X}_i = \begin{cases} \mathcal{N}(\mathbf{0}, (P+\epsilon)\mathbf{I}_{K\times K}) & \text{if} \quad B_i = 1 \\ \mathcal{N}(\mathbf{0}, (P-\epsilon)\mathbf{I}_{K\times K}) & \text{if} \quad B_i = 0 \end{cases}, \tag{4.11}$$

where $0 < \epsilon < P$*.*

Consider the detection algorithm in (4.9). Define,

$$\begin{aligned} P_{FA}(G) &:= \mathcal{P}\left(T\left(\{\mathbf{Y}_i\}_{i=1}^{G}\right) > \lambda | \mathcal{H}_0\right) \\ P_{MD}(G) &:= \mathcal{P}\left(T\left(\{\mathbf{Y}_i\}_{i=1}^{G}\right) < \lambda | \mathcal{H}_1\right). \end{aligned} \tag{4.12}$$

Then, for any given choice of $1 \geq P_{FA} > 0$*,* $1 \geq P_{MD} > 0$*, and* $P > 0$*, there exists a* G *such that* $P_{FA}(G) \leq P_{FA}$*,* $P_{MD}(G) \leq P_{MD}$*. Hence, the detection algorithm in (4.9) is robust to uncertainties for all* $P > 0$*.*

Proof. See Appendix C.1 for the proof. □

In Theorem 4.1 we have shown that the signal $X[n]$ is completely robust to uncertainty noise and fading processes and arbitrarily low SNRs. The question we ask is whether the primary user has to pay a penalty for the improved robustness of the secondary user? Recall that the channel from the primary transmitter to the primary receiver was assumed to be an AWGN channel with noise power $\sigma_p^2 := N_0 B$. Hence, the rate achieved at the primary receiver is the average of the rates achieved by the two Gaussian streams in the signal $X[n]$,

i.e.,

$$R_p := \frac{1}{2}\left[\frac{1}{2}\log\left(1+\frac{P+\epsilon}{\sigma_p^2}\right)+\frac{1}{2}\log\left(1+\frac{P-\epsilon}{\sigma_p^2}\right)\right] \leq C_p \tag{4.13}$$

The concavity-$\cap$ of the capacity function C_p shows that the primary takes a hit in the data rate to its receivers in order to provide robustness guarantee to the secondary users. This rate penalty depends on the parameter ϵ, and as $\epsilon \downarrow 0$, the rate penalty converges to zero. This suggests that the cost to the primary (in terms of loss in data rate) for providing robustness to the secondary can be made arbitrarily close to zero.

On the other hand, the parameter ϵ affects the sensing time (number of samples) required by the secondary user to robustly attain a given target P_{FA} and P_{MD}. Eliminating λ in (C.15), and using the fact that $\Theta \approx 1$ for large G, we get

$$N = G \cdot K = K\left[\sqrt{\frac{\lambda_1}{P_{FA}}}+\sqrt{\frac{\lambda_2}{P_{MD}}}\right]\epsilon^{-2} \tag{4.14}$$

Equation (4.14) tells us that the sensing delay at the secondary user monotonically increases as $\epsilon \downarrow 0$. Therefore, in the Gaussian mixture example there is tension between the primary's data rate and the sensing delay at the secondary detector.

4.4 Capacity-delay tradeoff

Comparing Equations (4.13) and (4.14) we can see that there is a tradeoff between the data rate (R_p) achieved by the primary user and the sensing delay (N) for the secondary user. We call this the *capacity-delay tradeoff*. An important question arising from this discussion is — whether there is a fundamental capacity-delay tradeoff? That is, does the primary user always have to take a hit in its data rate to provide sensing delay guarantees to the secondary users. The intuitive reasoning for the existence of a fundamental capacity-delay tradeoff is given below.

We know that the capacity-achieving output distribution is unique for most channels [124; 125]. Appealing to the continuity of mutual information this means that in most generic cases the capacity achieving input distribution is also unique. The results in this chapter show that the secondary user can robustly detect the primary user if the primary signal is a mixture of two distinct distributions. So, if the primary user is operating close to capacity using an input signal that is a mixture of two distributions, then both these distinct distributions must be close to the unique capacity-achieving input distribution. By continuity, the two output distributions at the secondary receiver must also be close to each other. Hence, to distinguish between these two output distributions, the sensing delay at the secondary must

be large. So, as the primary rate approaches its channel capacity, the sensing delay at the secondary must approach infinity.

4.5 Policy implications

In this chapter we considered the tradeoff between the loss in freedom for the primary user and robustness for the secondary user. We modeled the loss in freedom for the primary user by the hit in data rate it suffers in order to provide robustness (measured by the SNR wall) gains for the secondary user. We called this the *capacity-robustness tradeoff*.

The technical results in this chapter have non-trivial policy implications. For instance, regulators can mandate the primary users to transmit commonly used signal features like pilot tones and cyclostationary signals. This leads to a loss in data rate for the primary user, but provides sufficient robustness for the secondary user. On the other hand, the regulators can also mandate the primary users to transmit macroscale features which are complex and unnatural. The loss in data rate in this case can be made arbitrarily small, while also providing sufficient robustness to the secondary user.

From a purely data rate point of view, mandating macroscale features seems to be the right thing to do. However, this can be severely restrictive from a freedom point of view. The primary is forced to choose from the class of signals containing macroscale features, which can be unnatural and complex signaling schemes (like the mixture Gaussian example in Section 4.3.3). Unlike macroscale features, most communication signals already have pilot tones and cyclostationary features. So, mandating the primary user to transmit these features (pilots/cyclostationary features) might not be severely restrictive, even though it causes a loss in data rate. Also, the primary might anyway want to transmit these features for timing/synchronization to its own receivers.

These results suggest that there are different kinds of loss in freedom for the primary user. Some of them are good forms of loss in freedom even if they lead to loss in data rate, while others are not so good from a freedom point of view even though they do not incur any loss in data rate. A regulator needs to take these into consideration while writing rules for whitespace use.

The results so far indicate that with some restrictions on the flexibility of primary users, and by using carefully designed sensing algorithms, the secondary user can robustly detect very weak primary signals (SNRs of the order of -20 dB). The important question we need to ask is: What fraction of the whitespaces do the secondary users recover if they use these sensing algorithms? This is important because the whole point of opportunistic spectrum sharing is to reuse whitespaces and reduce the regulatory overhead of static spectrum allocation. We answer this question in Chapter 5.

Chapter 5

Space-time sensing metrics

5.1 Introduction

Philosophically, sensing is a decision problem: is it safe to use the spectrum where we are or is it unsafe? This is a question with a binary answer and so it seems natural to encapsulate the entire problem of spectrum sensing by mathematically casting it as a binary hypothesis testing problem. So far, in Chapters 2, 3, 4 we modeled the binary hypotheses as *primary present*, and *primary absent*. This suggested that the key metrics are the *probability of missed-detection* and the *probability of false-alarm*. This chapter shows that while such formulations are seemingly natural, if we are not careful, they can blind us to some of the true fundamental limits involved in spectrum sensing. In particular, the core spatial dimension to the problem and its interaction with fading in wireless channels introduces tradeoffs that need a better formulation to understand.

The SNR wall limits shown in Chapters 2, 3, 4 suggest that it is very hard to design highly sensitive detectors. However, why does one need sensitive detectors? The main reason is because the cognitive radio needs to be sure that it is far away (outside the no-talk radius r_n) from the primary transmitter. The strength of the primary transmitter's signal received at the cognitive radio is just a proxy to ensure that we are 'far enough'. This prompts the question: what is the right level of sensitivity at which the cognitive radio is sufficiently 'far away' from the primary transmitter? If there were no wireless fading, there would be a single 'right level' of sensitivity. It is the reality of fading that makes us demand additional sensitivity.

Because fading can effect different detectors differently, this reveals that a head-to-head comparison of the sensitivity of two detectors can be misleading. Instead, the possibility of fading has to be incorporated into the "signal present" hypothesis itself. The logical tradeoff is then between the effective probability of missed detection (it matters not whether the miss is due to an unfortunate fade or a quirk in the noise realization) and the sample complexity [63].

The bigger conceptual challenge comes in trying to understand "false alarms." The traditional hypothesis-testing formulation would say that a false alarm consists of when the detector says that we should not use the channel when the primary is truly absent ('noise only' hypotheses). But this is not always the problem facing a cognitive radio. There are two possible situations when the cognitive radio can use the band — the primary is truly absent, or the primary is present but is far away from the cognitive radio. The cognitive radio must guard against saying that it is close to the primary when it is indeed far enough away. This suggest that the "signal absent" hypothesis needs to be modified in some reasonable way.

At this point, a joint space-time perspective is essential. For ease of exposition we first review the 'space only' situation. We assume that the primary is always ON and the only spectrum opportunities are the spatial locations farther than r_n from the primary transmitter. We show that this problem can also be cast into a binary hypothesis-testing framework. However, the there are two fundamental differences in this formulation from the binary hypothesis-testing formulation in Section 1.2.2. Firstly, both the hypotheses are composite-hypotheses, parametrized by the spatial location of the cognitive radio. Secondly, we assume a worst-case spatial distribution under hypothesis $\mathcal{H}_1$ ('band used'), and a Bayesian approach for hypothesis $\mathcal{H}_0$ ('band unused'). The asymmetry in the formulation is due to the fact that the primary need not trust the deployment model for the secondary user, whereas the secondary might have a good model of its own spatial deployment.

This hypothesis-testing framework leads to two metrics (similar to P_{MD} and P_{FA}) first introduced by us[1] in [62]. The first metric, namely the *Fear of Harmful interference* F_{HI}, captures the safety to the primary users. The second metric, namely the *Weighted Probability of Area Recovered* $WPAR$, captures the performance of spectrum sensing by appropriately weighting the probability of false alarm (P_{FA}) across different spatial locations. These metrics give a unifying framework in which to compare different spectrum-sensing algorithms. Reasonable metric parameters from real-world data are given in [63; 64]. The tradeoff between $WPAR$ and F_{HI} is thus the correct ROC curve for spectrum sensing. However, the probabilistic uncertainty underlying the hypotheses is non-ergodic and so the tradeoffs are interesting even if we allow an infinite number of samples.

The new metrics show that fading uncertainty forces the $WPAR$ performance of single-radio sensing algorithms to be very low for desirably small values of F_{HI}, even with infinite samples. The physical reason for such poor performance is that a single radio cannot distinguish whether it is close to the primary user and severely (and atypically) shadowed, or if it is far away and (as typical) not shadowed. Furthermore, these metrics shed a new perspective on the impact of noise uncertainty on the sensing performance of spatial spectrum holes as well as on the comparison of different detectors. We show that under noise uncertainty, there exits an F_{HI} threshold beyond which the $WPAR$ vanishes to zero, i.e., if we need to

[1]The metrics defined in Section 5.3.1 and the spatial domain formulation of the spectrum sensing problem is joint work with Shridhar Mubaraq Mishra. The usefulness of the spatial metrics to analyze multi-user sensing algorithms has been shown in Shridhar Mubaraq Mishra's PhD dissertation [64]

guarantee protection to the primary below this threshold, then one cannot robustly recover any spectrum holes in space. In addition, a head-to-head comparison is made between the energy detector and the coherent pilot detector. This reveals that even with an infinite number of samples, the energy detector will do better for high values of F_{HI} and it is only noise-uncertainty that allows the coherent pilot detector to do better at low F_{HI}. This fact would be invisible without using the right metrics. These new spatial metrics also bring new insights into the fundamental tradeoffs involved in cooperative spectrum sensing [63; 62; 126; 64].

After reviewing the framework for spatial sensing metrics, we focus on the joint space-time spectrum sensing problem. There has been lot of work on the design of spectrum sensing algorithms to dynamically reuse the OFF times in a primary ON/OFF process [127; 128; 129; 130]. There has also been work on using change-detection techniques to detect a change in the state of the primary signal (ON to OFF transition), and use it to identify the unused primary channels [131; 77]. All these papers focus on the time-domain aspect of the problem, i.e, they assume that the secondary is located within the service area of the primary user and hence it can only transmit when the primary user is OFF. In this chapter, we consider the general scenario where the location of the secondary user can also vary in space. We propose a framework to analyze this problem and we give metrics that characterize the protection to the primary user, and also quantify the space-time opportunities recovered.

The key aspects of our space-time spectrum sensing framework are

- We argue the need for a sacrificial buffer zone (δ samples) at the start of each primary ON period within which secondaries are allowed to interfere with the primary. This is the time-domain equivalent to the spatial area sacrificed (between the decodable radius and the protected radius as shown in Figure 1.3(a)) by the primary user to allow the operation of secondary radios. This buffer zone is required to give the secondary time to evacuate the band when the primary comes back ON.

- We propose the *Weighted Probability of Space-Time Recovered* (WPSTR) metric that quantifies the opportunities in space-time recovered by the secondary user.

- We give an example in which there is a tension between performance in time versus performance in space. In this example, there exists an optimal choice of the sensing time for which the WPSTR metric is maximized. This is contrary to traditional wisdom which says that the performance improves monotonically with the sensing time. This non-trivial observation is hard to see without these space-time metrics.

5.2 Problems with the traditional formulation

The key idea behind the formulation in Section 1.2.2 is that the received signal power, P, averaged over fading is a monotonically decreasing function of distance between the cognitive

radio and the primary transmitter. So, a detector with high sensitivity (that can sense weak signals) ensures an appropriately low probability of mis-declaring that we are outside the no-talk radius whenever we are actually inside. However, this formulation has some fundamental flaws, which we discuss below.

5.2.1 How much sensitivity do we really need?

One is forced to ask the question – *At what level should we set the detector sensitivity?* The right level of sensitivity should correspond to the signal power at the no-talk radius. If there were no fading, the required sensitivity would immediately follow from the path-loss model. However, multipath fading and shadowing exist. We may hope to average over the multipath fading since it changes every coherence time. However, the coherence time is itself uncertain since it depends on physical motion — there is a real possibility of an infinite coherence time since the transmitter and the cognitive radio may both be stationary. This is thus potentially a nonergodic uncertainty, even though it presumably has a probabilistic model. One approach to deal with this nonergodic fading is to incorporate a margin for fading into the sensitivity, i.e., set the sensitivity low enough to account for the 1% worst-case fades. The problem with this approach is that firstly, different detectors may be effected differently by the details of the fading distribution. Hence, a single level of sensitivity for all detectors is a flawed. Secondly, setting extremely low values of sensitivity leads to performance loss as shown below.

5.2.2 How to measure performance of cognitive radios?

Traditionally, the "band unused" hypothesis ($\mathcal{H}_0$) has been modeled as receiving noise alone, and the performance of the cognitive radio is measured by the probability of false-alarm (P_{FA}). However, that does not reflect what we actually care about for cognitive radio systems. It is perfectly fine for the primary transmitter to be ON, as long as the cognitive radio can verify that it is outside the primary's no-talk radius. So, comparing against the "noise only" hypothesis is incorrect. In fact, as the received signal depends on the distance of the cognitive radio from the primary transmitter, hypothesis $\mathcal{H}_0$ is different for different locations of the cognitive radio. In particular, the performance of a cognitive radio is a function of its position, and how we set our detector's threshold impacts how much area we can recover for cognitive radio operation [31].

5.3 Spectrum Sensing: spatial-domain perspective

The discussion in the previous section forces us to rethink the hypothesis-testing formulation. The goal is to decide whether the cognitive radio is inside or outside r_n based on the measured

received signal samples $Y[n]$. The received signal can be modeled as $Y[n] = \sqrt{P(R)}X[n] + W[n]$, where R is the distance of the cognitive radio from the primary transmitter[2]. For a fixed distance R, the received signal power $P(R)$ is a random variable depending on the path-loss and fading distributions. Hence, we come up with the following hypothesis testing model:

$$\begin{aligned} \mathcal{H}_0 &: Y[n] = \sqrt{P(R)}X[n] + W[n], \quad R > r_n \\ \mathcal{H}_1 &: Y[n] = \sqrt{P(R)}X[n] + W[n], \quad 0 \leq R \leq r_n \end{aligned}$$

The only problem with the above is that both the hypotheses are parametrized by the distance R, and we need to come up with model for R for the formulation to make sense. The true position of the cognitive radio is unknown. For $\mathcal{H}_1$, it is natural to take the worst-case position of being just within r_n. This is because the primary might not trust the deployment model for the cognitive radios. Suppose we took the same approach to $\mathcal{H}_0$. Then the worst-case location under $\mathcal{H}_0$ would be just outside r_n. After all, if we can recover this location then we can recover all the area greater than r_n. This approach is fatally flawed since the distribution of the signal strength just within r_n and outside of r_n are essentially the same. No interesting tradeoff is possible. We miss the fundamental fact that we must give up some area immediately outside of r_n to avoid having a cognitive radio use the channel within r_n.

Simply averaging over R (distance from the primary transmitter) also poses a challenge. The interval (r_n, ∞) is infinite and hence there is no uniform distribution over it. This mathematical challenge corresponds to the physical fact that if we take a single primary-user model, the area outside r_n that can be potentially recovered is infinite. With an infinite opportunity, it does not matter how much of it that we give up! We need to come up with probability distribution over R or in other words, we need to weight/discount area outside r_n appropriately. Weighting area by "utility" is a possibility, but as discussed in [62], this would tightly couple the evaluation of sensing with details of the business model and system architecture. It is useful to find an approximate utility function that decouples the evaluation of the sensing approach from all of these other concerns.

With the above discussion in mind, we come up with the following hypothesis testing model:

$$\begin{aligned} \mathcal{H}_0 &: Y[n] = \sqrt{P(R)}X[n] + W[n], \quad R \sim w(r)r \\ \mathcal{H}_1 &: Y[n] = \sqrt{P(R)}X[n] + W[n], \quad R \in [0, r_n] \end{aligned} \tag{5.1}$$

where $w(r)$ must satisfy $\int_{r_n}^{\infty} w(r)\ r\ dr = 1$ and $w(r) = 0$ if $r < r_n$. The numerical results in this paper have been computed using an exponential weighting function, $w(r) = A\exp\left(-\kappa(r - r_n)\right)$, for $r > r_n$ and $w(r) = 0$ for $0 \leq r \leq r_n$. One possible justification for

[2]Throughout this chapter we assume radial symmetry and hence it is sufficient to consider the radial distance of the cognitive radio from the primary transmitter to characterize its location.

the choice of the weighting function comes from the fact that an exponential distribution has the *maximum entropy* among the set of all continuous valued distributions on $[0, \infty)$ with a given mean. In our case the mean corresponds to the average distance between two primary towers transmitting on the same frequency band.

From the above, the asymmetry between the two hypotheses is clear. $\mathcal{H}_0$ is a well-defined probability and so P_{FA} can be calculated for a detector. Meanwhile, $\mathcal{H}_1$ has R as a non-probabilistic uncertainty and so we would have to require something like

$$\sup_{r \in [0, r_n]} \mathcal{P}_{R=r} \left(T(\mathbf{Y}) < \lambda | \mathcal{H}_1 \right) \leq P_{MD}.$$

The resulting mixed ROC curve for a detector reveals the fundamental tradeoffs.

However, such a formulation mixing worst-case and Bayesian uncertainties in different ways across the two hypotheses is novel. Using the traditional names P_{FA} and P_{MD} in this context is likely to lead to confusion. Therefore, as in [62] we gave them new more descriptive names that better reflected their roles.

5.3.1 Spatial sensing metrics

5.3.1.1 Safety: Controlling the Fear of Harmful Interference

The idea behind the safety metric is to measure the worst-case safety that the cognitive radio can assure the primary (the worst case is calculated over the fading distribution negotiated between the cognitive radio and the primary, and over the location of the cognitive radio within the no-talk radius). To capture the strength of the safety guarantee under uncertainty we define a new metric called the *Fear of Harmful Interference*. This is the same as the average P_{MD} but takes into account all uncertainties in location and fading.

Definition 5.1. *Fear of Harmful Interference* (F_{HI}) metric is

$$F_{HI} = \sup_{0 \leq r \leq r_n} \sup_{F_r \in \mathbb{F}_r} \mathcal{P}_{F_r}(D = 0 | R = r). \tag{5.2}$$

where $D = 0$ is detector's decision declaring that the cognitive radio is outside the no-talk radius and $\mathbb{F}_r$ is the set of possible distributions for $P(r), W$ at a distance of r from the primary transmitter. The outer supremum is needed to issue a uniform guarantee to all protected primary receivers and also reflects the uncertainty in cognitive radio deployments. The inner supremum reflects the non-probabilistic uncertainty in the distributions of the path-loss, fading, noise, or anything else.

5.3.1.2 Performance

Next we consider a metric to deal with the cognitive radio's performance — its ability to identify spectrum opportunities. From a traditional perspective, this is basically a weighted P_{FA}. Every point at a radial distance $r > r_n$ is a spectrum opportunity. For any detection algorithm, there is a probability associated with identifying an opportunity there, called the probability of finding the hole $P_{FH}(r)$:

$$P_{FH}(r) = \mathcal{P}_{F_r}(D = 0|R = r), \quad r > r_n. \tag{5.3}$$

where F_r represents the propagation, fading, and noise models believed by the cognitive radio designer.

Definition 5.2. The *Weighted Probability of Area Recovered* (WPAR) metric is

$$WPAR = \int_{r_n}^{\infty} P_{FH}(r)w(r)\ r dr, \tag{5.4}$$

where $w(r)$ is a weighting function that satisfies $\int_{r_n}^{\infty} w(r)\ r\ dr = 1$.

The name $WPAR$ reminds users of the weighting of performance over spatial locations that is fundamental to the cognitive radio context. $1 - WPAR$ is the appropriate analog of P_{FA} and measures the sensing overhead [37]. Note that $WPAR = 1$ corresponds to the cognitive radio recovering all the area outside r_n.

5.3.2 Models for fading uncertainty

The received primary signal strength P (in dBm) can be modeled as $P(r) = p_t - (l(r) + S + M)$, where p_t is the power of the transmitted signal, $l(r)$ is the loss in power due to attenuation at a distance r from the primary transmitter, S is the loss due to shadowing and M is the loss due to multipath fading. Unless specifically mentioned, in this chapter we assume that all powers are measured in the Decibel (dB) scale. We assume that $l(r) = 10\log_{10}(r^{\alpha})$, and α is the true attenuation exponent.[3]

5.3.2.1 Nominal model

For convenience, S and M are assumed to be independent of r and to follow a nominal model for $S + M$ that is Gaussian ($S + M \sim \mathcal{N}(\mu_S, \sigma^2)$) on a dB scale. This implies that $P \sim \mathcal{N}(\mu(r), \sigma^2)$, where $\mu(r) = p_t - (l(r) + \mu_S)$. This is the distribution used to compute the WPAR as in (5.4). For the plots in this paper, $\mu_S = 0\ dB$ and the standard deviation $\sigma = 5.5$ dB were chosen to match standard assumptions in the IEEE 802.22 literature [132].

[3] We could include an uncertainty model for the attenuation exponent and include this in the computation of F_{HI}. However, for simplicity we assume complete knowledge of the attenuation exponent in this chapter.

5.3.2.2 Quantile models

To compute F_{HI}, we cannot always use the nominal model for shadowing and multipath as it is important to model the fact that the primary user does not trust this model completely. Instead, it is possible that the primary user trusts only a quantized version (or a coarse histogram) of the fading distribution. Mathematically, we model this as a class of distributions $(\mathbb{F}_r)$ that satisfy given quantile constraints.

Definition 5.3. A *single-quantile model* $\mathbb{F}_r$ is a set of distributions for the received signal power defined by a single number $0 \leq \beta \leq 1$ and a function of r denoted $\gamma(r, \beta)$. A distribution $F_r \in \mathbb{F}_r$ iff

$$\mathcal{P}_{F_r}(P < \gamma(r, \beta)) = \beta. \tag{5.5}$$

A *k-quantile model* is a set of distributions $\mathbb{F}_r$ for the received signal power defined by a list of numbers $(\beta_1 < \ldots < \beta_k)$ and a corresponding list of functions $(\gamma_1(r, \beta_1), \ldots, \gamma_k(r, \beta_k))$. A distribution $F_r \in \mathbb{F}_r$ iff $\forall i \leq k$

$$\mathcal{P}_{F_r}(P < \gamma_i(r, \beta_i)) = \beta_i. \tag{5.6}$$

For consistency, the quantiles are chosen so that the nominal Gaussian $\mathcal{N}(\mu(r), \sigma^2)$ is always one of the possible distributions for P.

$$\gamma(r, \beta) = \mathcal{Q}^{-1}(1 - \beta)\sigma + \mu(r), \tag{5.7}$$

where $\mathcal{Q}^{-1}(\cdot)$ is the inverse of the standard Gaussian tail probability function.

Figure 5.1 shows a picture of the distributions allowed under the quantile model (5 learned quantiles) defined in this section. The set of allowed Cumulative Distribution Functions (CDF's) for P under our quantile model is precisely the set of all possible curves sandwiched between the upper and lower bounds shown in Figure 5.1. The dashed (black) curve in the figure shows that nominal Gaussian CDF for P, and the quantile constraints can be thought of as samples of the nominal CDF (the triangle points (in red) in the figure).

5.3.3 Single radio sensing performance

The tradeoff between F_{HI} and $WPAR$ depends on the detector used by the secondary user. We start with a hypothetical single-radio detector (with infinite sensing time) that meets the current FCC sensing requirement [26]. The issue of finite sensing time is illustrated next through the example of a radiometer.

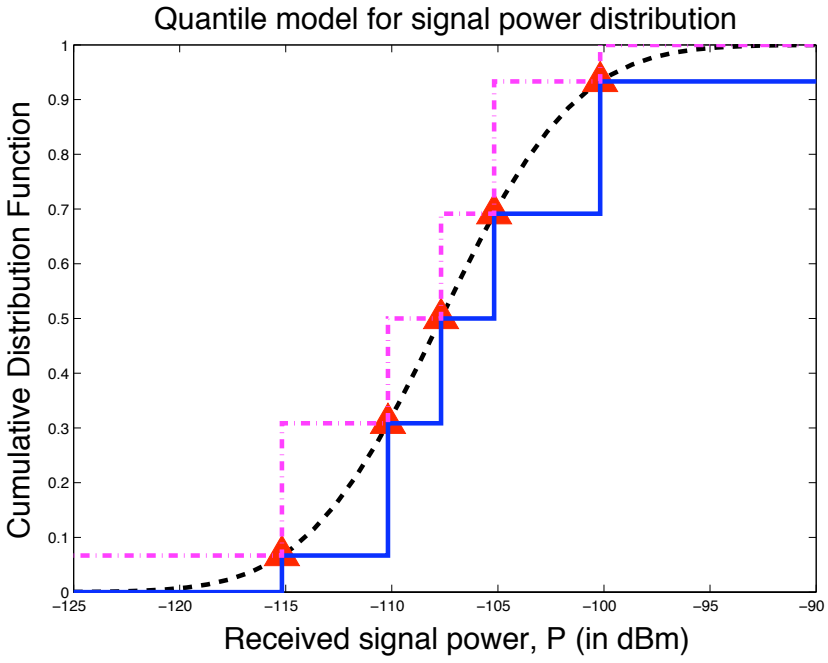


Figure 5.1: The quantile model for the received signal power (P) distribution. The dashed (black) curve is the nominal Gaussian CDF for P, and the triangle points (in red) show the quantile constraints on the CDF. The dashed-dotted (magenta) curve is the upper bound and the solid (blue) curve is the lower bound on the allowable CDF for P. The actual CDF can lie anywhere in between these two bounding functions.

5.3.3.1 Evaluating an ideal -114 dBm detector

The detector specifications (for sensing-based cognitive radio systems) in the FCC whitespace ruling require any proposed sensing algorithm to be able to detect digital television signals at -114 dBm to a probability of mis-detection $P_{MD} \leq 0.1$ and probability of false alarm $P_{FA} \leq 0.1$ [26].[4] We now show that detectors based on such specifications lead to very poor area recovery.

We assume a generic detection algorithm that computes a test statistic $T(\mathbf{Y})$ as a function of the received signal samples $\mathbf{Y} = (Y[1], Y[2], \cdots, Y[N])$ and compares it to a threshold to decide whether the band is free to use or not, i.e,

$$T(\mathbf{Y}) \underset{D=0}{\overset{D=1}{\gtrless}} \lambda.$$

Let P (in dBm) denote the average power of the received signal samples. P is a random variable whose distribution depends on the shadowing and fading distributions. The probability of mis-detection and probability of false-alarm to be

$$\begin{aligned} P_{MD}(p) &:= \mathcal{P}(T(\mathbf{Y}) < \lambda | P = p), \\ P_{FA} &:= \mathcal{P}(T(\mathbf{Y}) \geq \lambda | P = -\infty). \end{aligned} \tag{5.8}$$

Note that the probability of mis-detection is a function of the received signal power, where as the probability of false-alarm is a single number. For most reasonable detection algorithms $P_{MD}(p)$ is a monotonically increasing function, i.e., as the received signal power increases there is a smaller chance of missing the signal.

Theorem 5.1. *Consider a generic detection algorithm that computes a test-statistic $T(\mathbf{Y})$ and compares it to a detector threshold λ to decide whether the band is free to use. Assume that the detection algorithm satisfies the following specifications:*

- *The mis-detection probability function (defined in (5.8)) satisfies, $P_{MD}(-114) \leq 0.1$.*
- *$P_{MD}(p)$ is a monotonically decreasing function of p.*
- *The probability of false-alarm (defined in (5.8)) satisfies $P_{FA} \leq 0.1$.*

Assume that the nominal distribution (to compute WPAR) for the received single power P is $\mathcal{N}(\mu(r), \sigma^2)$, where $\mu(r)$ is a deterministic monotonically decreasing function of r. Furthermore, assume that the set of distributions $\mathbb{F}_r$, that the primary trusts (to compute F_{HI})

[4]The FCC report does not explicitly state the probability of false alarm and the probability of missed-detection to be 0.1. However, we choose these values based on the sensing requirements in the IEEE 802.22 standard.

for the received signal power P satisfies the single-quantile model in Definition 5.3. That is, any distribution $F_r \in \mathbb{F}_r$ satisfies[5]

$$\mathcal{P}_{F_r}(P < -114) = \beta(r), \tag{5.9}$$

where $\beta(r) = 1 - \mathcal{Q}\left(\frac{\mu(r)-(-114)}{\sigma}\right)$. *Then, we have*

$$\begin{aligned} F_{HI} &= 0.9\beta(r_n) + 0.1, \\ P_{FH}(r) &\leq \beta(r) + 0.1, \end{aligned} \tag{5.10}$$

Proof. The secondary user can use any detection algorithms as long as it meets the specifications given in the statement of the Theorem. To compute the fear of harmful interference to the primary user, one must consider the worst-possible detection algorithm satisfying the given specifications.

Let $\mathcal{D}$ denote the set of all detection algorithms satisfying the specifications given in the statement of the theorem. Then, the fear of harmful interference is

$$\begin{aligned} F_{HI} &= \sup_{0\leq r\leq r_n} \sup_{F_r\in\mathbb{F}_r} \sup_{D\in\mathcal{D}} \mathbb{E}_{F_r}\left[P_{MD}^D(P)\right] \\ &\overset{(a)}{=} \sup_{0\leq r\leq r_n} \sup_{D\in\mathcal{D}} P_{MD}^D(-\infty)\mathcal{P}(P < -114) + P_{MD}^D(-114)\mathcal{P}(P \geq -114) \\ &\overset{(b)}{=} \sup_{0\leq r\leq r_n} \sup_{D\in\mathcal{D}} [(1 - P_{FA}^D)\mathcal{P}(P < -114) + P_{MD}^D(-114)\mathcal{P}(P \geq -114)] \\ &\overset{(c)}{=} \sup_{0\leq r\leq r_n} [\beta(r) + 0.1(1 - \beta(r))] \\ &= \sup_{0\leq r\leq r_n} (0.9\beta(r) + 0.1) \\ &\overset{(d)}{=} 0.9\beta(r_n) + 0.1. \end{aligned}$$

In the above chain of equalities the superscript D is used to denote a detection algorithm from the class of allowed detection algorithms $\mathcal{D}$. Equality (a) follows from the fact that the maximizing distribution $F_r^* \in \mathbb{F}_r$ corresponds to placing a mass of $\beta(r)$ at $-\infty$ and $(1-\beta(r))$ at -114 dBm. This is the maximizing distribution irrespective of the actual detection algorithm $D \in \mathcal{D}$. This is because $P_{MD}^D(p)$ is a monotonically decreasing function of p, for all $D \in \mathcal{D}$. Equality (b) follows from the fact that $P_{MD}^D(-\infty)$ is the mis-detection probability when the signal is absent ($p = -\infty$). This corresponds to the event when noise-only received signal samples do not cause a false-alarm. Hence, $P_{MD}^D(-\infty) = 1 - P_{FA}^D$. Equality (c) follows

[5]Since only the -114 dBm level is specified, it is natural to assume that the primary only has confidence in a single-quantile that corresponds to that level.

from the fact that $\sup_{D\in\mathcal{D}}(1-P_{FA}^D)=1$, and $\sup_{D\in\mathcal{D}}(1-P_{MD}^D(-114))=0.1$. Finally, equality (d) follows from the fact that $\beta(r)$ is a monotonically increasing function.

Now, for any $D\in\mathcal{D}$, the probability of finding a hole is given by

$$\begin{aligned} P_{FH}(r) &= \mathbb{E}_P\left[\mathcal{P}(T^D(\mathbf{Y})<\lambda|P)\right] \\ &= \mathbb{E}_P[P_{MD}^D(P)] \\ &\overset{(e)}{\leq} P_{MD}^D(-\infty)\mathcal{P}(P<-114)+P_{MD}^D(-114)\mathcal{P}(P\geq -114) \\ &\overset{(f)}{\leq} \beta(r)+0.1, \quad \text{for} \quad r>r_n. \end{aligned} \tag{5.11}$$

The bound in (e) follows from the fact that the function $P_{MD}^D(p)\leq P_{MD}^D(-\infty)$, for $-\infty<p<-114$, and $P_{MD}^D(p)\leq P_{MD}^D(-114)$, for $-114\leq p<\infty$. These inequalities follows from the fact that $P_{MD}^D(p)$ is a monotonically decreasing function of p. The bound in (f) follows from observing that $P_{MD}^D(-\infty)\leq 1$, $\mathcal{P}(P<-114):=\beta(r)$, $P_{MD}^D(-114)\leq 0.1 \ \ \forall D\in\mathcal{D}$, and $\mathcal{P}(P\geq -114)\leq 1$. □

Shockingly, there is a benefit from missed detections above in step (f) of (5.11)! This suggests that a clever detector designer would do well to introduce intentional missed detections to improve performance while still meeting the official specification. This calls into question the unspoken assumption that deployed detector implementations would have better probabilities of missed detection when the primary signal is stronger than −114 dBm.

Using (5.11) in the definition of WPAR (See (5.4)) and applying our nominal model gives a $WPAR\leq 0.16$. This clearly shows that while the −114 dBm requirement seems very conservative, the detector specification is actually not very safe and simultaneously has a poor area recovery irrespective of the actual detector used. In the worst-case, the signal can indeed fall as low as -114 dBm at the no-talk radius. However in the average case the signal is a lot stronger and this leads to a lot of valuable area going unrecovered.

5.3.3.2 The radiometer

Recall from Chapter 2 that the test-statistic for the radiometer can be written as

$$T(\mathbf{Y})=\frac{1}{N}\sum_{n=1}^{N}|Y[n]|^2 \underset{D=0}{\overset{D=1}{\gtrless}} \lambda, \tag{5.12}$$

where λ is the design parameter and is called the detector threshold. Here

$$Y[n]=\sqrt{P(r)}X[n]+W[n],$$

where $X[n]$ is the faded primary signal at time n, and $W[n]$ is the background noise at time n. For convenience assume that all $W[n]$ are independent and identically distributed as $\mathcal{N}(0, \sigma_w^2)$.

Theorem 5.2. *Consider the radiometer whose test-statistic is defined in (5.12). Let P denote the average received primary signal power in the Decibel scale. Let $\mathbb{F}_r$ be the set of possible distributions for P at a distance of r from the primary transmitter.*

Assume that the received power distribution is completely known and is given by $P \sim \mathcal{N}(\mu(r), \sigma^2)$, where $\mu(\cdot)$ is a known monotonically decreasing function. This corresponds to the case of total consensus between primary and secondary and $\mathbb{F}_r$ contains just the nominal model.

In this case, the safety/performance of the radiometer given in (5.12) is

$$\begin{aligned} F_{HI}^{ck} &= \int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right)\right] \cdot \xi \exp^{\frac{-(p-\mu(r_n))^2}{2\sigma^2}} \; dp \; + O\left(\frac{1}{\sqrt{N}}\right), \\ P_{FH}(r) &= \int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right)\right] \cdot \xi \exp^{\frac{-(p-\mu(r))^2}{2\sigma^2}} \; dp \; + O\left(\frac{1}{\sqrt{N}}\right), \quad (5.13) \end{aligned}$$

where $\xi := \frac{1}{\sqrt{2\pi\sigma^2}}$, and the WPAR is given by

$$WPAR^{ck} = \int_{r_n}^{\infty} w(r) P_{FH}(r) \; r dr$$

Proof. As the received power distribution is completely known, the fear of harmful interference is given by

$$F_{HI}^{ck} = \mathbb{E}_P \left[\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P\right)\right], \quad (5.14)$$

where $P \sim \mathcal{N}(\mu(r_n), \sigma^2)$. Let $P = p$ be a given received power realization, then using the Central Limit Theorem we can approximate $T(\mathbf{Y}) \sim \mathcal{N}\left(10^{\frac{p}{10}} + \sigma_w^2, \frac{2}{N}[10^{\frac{p}{10}} + \sigma_w^2]^2\right)$ [66]. Furthermore, if $F_T^N(x)$ is the CDF of $\frac{T(\mathbf{Y}) - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}[10^{\frac{p}{10}} + \sigma_w^2]}$, then the Berry-Esseen Theorem (see Theorem 2.4.9 in [86]) about the convergence of the CLT gives us

$$\left|F_T^N(x) - G(x)\right| \leq \frac{c}{\sqrt{N}},$$

where $G(x)$ is the CDF of a $\mathcal{N}(0,1)$ distribution, and $c > 0$ is a known constant.

Therefore,

$$\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = p\right) = 1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right) + O\left(\frac{1}{\sqrt{N}}\right). \tag{5.15}$$

Using (5.15) in (5.14) we get

$$\begin{aligned} F_{HI}^{ck} &= \int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right) + O\left(\frac{1}{\sqrt{N}}\right)\right] \cdot \xi \exp^{\frac{-(p-\mu(r_n))^2}{2\sigma^2}} \; dp \\ &= \int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right)\right] \cdot \xi \exp^{\frac{-(p-\mu(r_n))^2}{2\sigma^2}} \; dp \quad + O\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{5.16}$$

The probability of finding a hole at a distance r from the transmitter is given by

$$P_{FH}(r) = \mathbb{E}_P\left[\mathcal{P}(T(\mathbf{Y}) < \lambda | P)\right], \tag{5.17}$$

where $P \sim \mathcal{N}(\mu(r), \sigma^2)$. Therefore,

$$P_{FH}(r) = \int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right)\right] \cdot \xi \exp^{\frac{-(p-\mu(r))^2}{2\sigma^2}} \; dp \quad + O\left(\frac{1}{\sqrt{N}}\right). \tag{5.18}$$

Substituting (5.18) in (5.4) gives us the expression for WPAR

$$\begin{aligned} WPAR^{ck} = & \int_{r_n}^{\infty} w(r) \left(\int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right)\right] \cdot \xi \exp^{\frac{-(p-\mu(r))^2}{2\sigma^2}} \; dp\right) \; r dr \\ & + \; O\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{5.19}$$

□

Note that the expressions for the performance/safety of the radiometer in Theorem 5.2 contain error terms from the CLT approximation. If one wishes to compute exact expressions then they must use the Chi-squared distribution for the test-statistic $T(\mathbf{Y})$. In the rest of the chapter, we omit the error terms arising due to the CLT approximation. This is done to keep the mathematical equations simple.

Theorem 5.2 gives us the safety/performance of a radiometer with complete distribu-

tional knowledge. We now give the safety/performance under the single-quantile model of Section 5.3.2 for the received signal distribution.

Theorem 5.3. *Consider the radiometer whose test-statistic is defined in (5.12). Let P denote the average received primary signal power in the Decibel scale. Let $\mathbb{F}_r$ be the set of possible distributions for P at a distance of r from the primary transmitter. Assume that the uncertain set $\mathbb{F}_r$ satisfies the single-quantile model of Section 5.3.2. That is, given a $0 \leq \beta \leq 1$, the primary agrees on a threshold $\gamma(r,\beta)$ such that for every distribution $F_r \in \mathbb{F}_r$, $\mathcal{P}_{F_r}(P < \gamma(r,\beta)) = \beta$, where $\gamma(r,\beta) = \sigma\mathcal{Q}^{-1}(1-\beta) + \mu(r)$.*

In this case, the fear of harmful interference is given by

$$F_{HI}^{sq} = \beta \left[1 - \mathcal{Q}\left(\frac{\lambda - \sigma_w^2)}{\sqrt{\frac{2}{N}}\sigma_w^2}\right)\right] + (1-\beta)\left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{\gamma(r_n,\beta)}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{\gamma(r_n,\beta)}{10}} + \sigma_w^2)}\right)\right], \tag{5.20}$$

where $\gamma(r_n,\beta) = \sigma\mathcal{Q}^{-1}(1-\beta) + \mu(r_n)$. The WPAR performance is same as Eqn. (5.19) in Theorem 5.2.

Proof. See Appendix D.1 for the proof. □

Note that the secondary user has two parameters to adjust. It can adjust the threshold λ on its own and it can negotiate with the regulator/primary regarding the appropriate value for β. We assume that it does both and satisfies (D.4). The optimal β and λ are chosen to maximize $WPAR^{ck}$ (See Eqn. (5.19)). This can be done numerically. Figure 5.2 shows the resulting safety/performance tradeoff for a single radio with both a finite and infinite number of samples. From Figure 5.2 we can see that the impact of the uncertainty is substantial when the sensing time is finite.

A special limiting case of interest is when the number of samples available for detection is infinite. This allows us to isolate the effect of fading/shadowing from the effect of finite samples. We call this a 'perfect' radiometer (in general a single-user perfect detector). The test-statistic for a perfect radiometer is $T^{per}(\mathbf{Y}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |Y[n]|^2$. As $Y[n]$ are independent and identically (*iid*) distributed random variables, the strong law of large numbers implies $\frac{1}{N}\sum_{n=1}^{N} |Y[n]|^2 \xrightarrow{a.s.} 10^{\frac{p}{10}} + \sigma_w^2$, where p (in dBm) is the average received signal power. Therefore, the perfect radiometer decides whether the band is used/unused according to the following rule

$$\begin{aligned} T^{per}(\mathbf{Y}) = 10^{\frac{P}{10}} + \sigma_w^2 &\underset{D=0}{\overset{D=1}{\gtrless}} \lambda \\ \Leftrightarrow P &\underset{D=0}{\overset{D=1}{\gtrless}} 10\log_{10}(\lambda - \sigma_w^2), \end{aligned} \tag{5.21}$$

We now derive the safety/performance of the perfect radiometer.

Theorem 5.4. *Consider a perfect radiometer, whose test-statistic is defined in (5.21), where P is the received signal power, $\widetilde{\lambda} := 10\log_{10}(\lambda - \sigma_w^2)$ is the detection threshold and $\mathbb{F}_r$ is the set of possible distributions for P at a distance of r from the primary transmitter.*

- ***A**: Assume that the received power distribution is completely known ($\mathbb{F}_r$ is a singleton) and is given by $P \sim \mathcal{N}(\mu(r), \sigma^2)$, where $\mu(\cdot)$ is a known monotonically decreasing function. Then, the safety/performance is given by*

$$\begin{aligned} F_{HI}^{per,ck} &= 1 - \mathcal{Q}\left(\frac{\widetilde{\lambda} - \mu(r_n)}{\sigma}\right), \\ WPAR^{per,ck} &= \int_{r_n}^{\infty} w(r)\left[1 - \mathcal{Q}\left(\frac{\widetilde{\lambda} - \mu(r)}{\sigma}\right)\right] r\ dr. \end{aligned} \tag{5.22}$$

- ***B**: Assume that the uncertain set $\mathbb{F}_r$ satisfies the single-quantile model of Section 5.3.2. That is, given a $0 \le \beta \le 1$, the primary agrees on a threshold $\gamma(r,\beta)$ such that for every distribution $F_r \in \mathbb{F}_r$, $\mathcal{P}_{F_r}(P < \gamma(r,\beta)) = \beta$, where $\gamma(r,\beta) = \sigma\mathcal{Q}^{-1}(1-\beta) + \mu(r)$. Then, the fear of harmful interference is given by*

$$F_{HI}^{per,sq} = \begin{cases} \beta & \text{if} \quad \widetilde{\lambda} \le \gamma(r_n,\beta) \\ 1 & \text{otherwise.} \end{cases} \tag{5.23}$$

The WPAR in this case is same as in the complete knowledge case (See Eqn. (5.22)).

Proof. See Appendix D.2 for the proof. □

Corollary 5.5. *Consider the perfect radiometer under the single-quantile uncertainty model for $\mathbb{F}_r$. For a given choice of target fear of harmful interference F_{HI}^t, the optimal β^* and $\widetilde{\lambda}^*$ that maximizes $WPAR^{per,sq}$ subject to $F_{HI}^{per,sq} \le F_{HI}^t$ is given by*

$$\beta^* = F_{HI}^t, \qquad \widetilde{\lambda}^* = \sigma\mathcal{Q}^{-1}(1 - F_{HI}^t) + \mu(r_n). \tag{5.24}$$

Hence, $F_{HI}^t = \mathcal{Q}\left(\frac{\lambda^* - \mu(r_n)}{\sigma}\right)$, which is the same as the safety performance given in (5.22) for the perfect radiometer under complete distributional knowledge.

Proof. From (5.23) it is clear that, to meet the constraint $F_{HI}^{per,sq} \le F_{HI}^t$ we must have $\beta \le F_{HI}^t$ and $\gamma(r_n,\beta) \ge \widetilde{\lambda}$. Therefore, the optimization problem is reduced to

$$\sup_{\beta,\widetilde{\lambda}} \left[\int_{r_n}^{\infty} w(r)\left[1 - \mathcal{Q}\left(\frac{\widetilde{\lambda} - \mu(r)}{\sigma}\right)\right] r\ dr\right] : \qquad \text{s.t. } \beta \le F_{HI}^t,\ \ \widetilde{\lambda} \le \sigma\mathcal{Q}^{-1}(1-\beta) + \mu(r_n) \tag{5.25}$$

Note that the objective function is monotonically increasing function of $\widetilde{\lambda}$. This is because $\left[1-\mathcal{Q}\left(\frac{\widetilde{\lambda}-\mu(r)}{\sigma}\right)\right]$ is monotonically increasing with $\widetilde{\lambda}$ and $w(r) \geq 0$. Also, note that the constraint on $\widetilde{\lambda}$, $\sigma\mathcal{Q}^{-1}(1-\beta)+\mu(r_n)$ is monotonically increasing with β. These two observations imply that the optimizing β^* and $\widetilde{\lambda}^*$ are the extreme values of the constraint functions, i.e., $\beta^* = F_{HI}^t$ and $\widetilde{\lambda}^* = \sigma\mathcal{Q}^{-1}(1-F_{HI}^t)+\mu(r_n)$. □

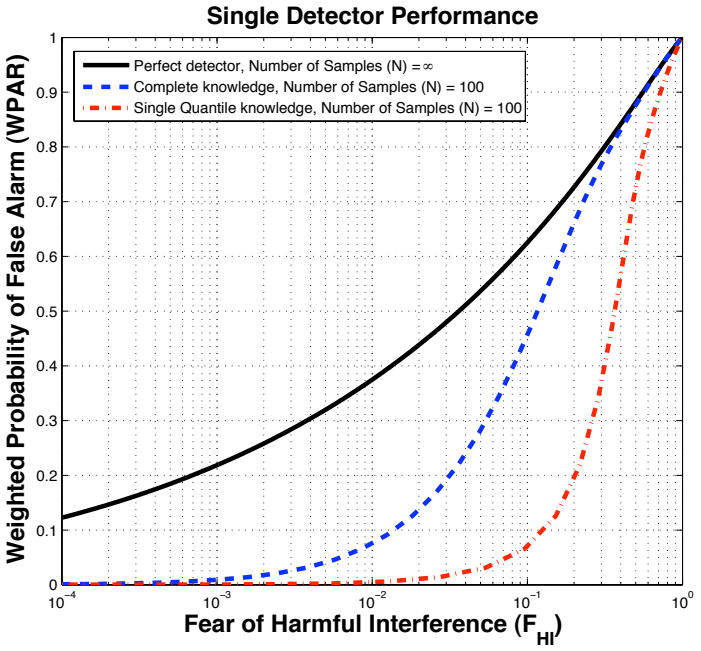


Figure 5.2: Performance of a perfect detector (infinite samples) as compared with a radiometer using a finite number of samples.

The black curve in Figure 5.3 shows the F_{HI} vs $WPAR$ tradeoff for a single user running a radiometer. The $WPAR$ performance at low F_{HI} is bad even for the perfect radiometer. This captures the physical intuition that guaranteeing strong protection to the primary user forces the detector to budget for deep fading events. Unlike in traditional communication problems where there is no harm if the fading is not bad, here there is substantial harm since a spectrum opportunity is left unexploited. The reader is encouraged to read [62], [126] to understand the $(F_{HI}, WPAR)$ tradeoffs for detectors limited to a finite number of samples as well as those for uncertain fading models.

5.3.3.3 Impact of noise uncertainty: SNR walls in space

In the analysis of the perfect radiometer (see Eqn (5.21)) we assumed that the noise power σ_w^2 is completely known. We now see the impact of uncertainty in the noise power on the F_{HI} vs $WPAR$ tradeoff for the perfect radiometer.

Theorem 5.6. *Consider a perfect radiometer, whose test-statistic is defined in (5.21), where $P(r)$ is the received signal power, $\widetilde{\lambda} := 10\log_{10}(\lambda - \sigma_w^2)$ is the detection threshold, σ_w^2 is the nominal noise power, and $\mathbb{F}_r$ is the set of possible distributions for $P(r)$ at a distance of r from the primary transmitter. Assume,*

- *The received power distribution is completely known ($\mathbb{F}_r$ is a singleton) and is given by $P(r) \sim \mathcal{N}(\mu(r), \sigma^2)$, where $\mu(\cdot)$ is a known monotonically decreasing function.*
- *The noise power is uncertain, and is known only within a certain range given by $\sigma_w^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]$, where σ_n^2 is the nominal noise power, and ρ is a parameter that captures the uncertainty in the noise power.*

Then, there exists an F_{HI} threshold $F_{HI}^t := 1 - \mathcal{Q}\left(\frac{10\log_{10}([1-\frac{1}{\rho}]\sigma_n^2) - \mu(r_n)}{\sigma}\right)$, below which the area recovered is zero, i.e., $WPAR = 0$.

Proof. From the definition of the *fear of harmful interference*, we have

$$\begin{aligned} F_{HI} &= \sup_{\sigma_w^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]} \mathcal{P}\left(P(r) < \widetilde{\lambda} | r = r_n\right) \\ &= \sup_{\sigma_w^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]} \left[1 - \mathcal{Q}\left(\widetilde{\lambda} - \mu(r_n)\sigma\right)\right] \\ &= \left[1 - \mathcal{Q}\left(\frac{10\log_{10}(\lambda - \frac{1}{\rho}\sigma_n^2) - \mu(r_n)}{\sigma}\right)\right] \end{aligned}$$

From the above expression the threshold λ can be computed to be

$$\lambda = \frac{1}{\rho}\sigma_n^2 + 10^{\left(\frac{\mu(r_n) + \sigma\mathcal{Q}^{-1}(1-F_{HI})}{10}\right)} \tag{5.26}$$

Now, the probability of finding a hole is given by

$$\begin{aligned} P_{FH}(r) &= \mathcal{P}_{\mathbb{F}_r}\left(P(r) < 10\log_{10}(\lambda - \sigma_n^2)\right) \\ &= \mathcal{P}_{\mathbb{F}_r}\left(10^{\frac{P(r)}{10}} < \lambda - \sigma_n^2\right) \end{aligned} \tag{5.27}$$

Substituting the expression for λ from (5.26) in (5.27), we get

$$P_{FH}(r) = \mathcal{P}_{\mathbb{F}_r}\left(10^{\frac{P(r)H^2}{10}} < 10^{\left(\frac{\mu(r_n)+\sigma\mathcal{Q}^{-1}(1-F_{HI})}{10}\right)} - (\sigma_n^2 - \frac{1}{\rho}\sigma_n^2)\right) \tag{5.28}$$

Since $10^{\frac{P(r)H^2}{10}} > 0$, $P_{FH}(r) = 0\ \forall r \geq r_n$

$$\begin{aligned}
&\Leftrightarrow\ 10^{\left(\frac{\mu(r_n)+\sigma\mathcal{Q}^{-1}(1-F_{HI})}{10}\right)} - (\sigma_n^2 - \frac{1}{\rho}\sigma_n^2) \leq 0 \\
&\Leftrightarrow\ 10^{\left(\frac{\mu(r_n)+\sigma\mathcal{Q}^{-1}(1-F_{HI})}{10}\right)} \leq (\sigma_n^2 - \frac{1}{\rho}\sigma_n^2) \\
&\Leftrightarrow\ F_{HI} \leq 1 - \mathcal{Q}\left(\frac{10\log_{10}([1-\frac{1}{\rho}]\sigma_n^2) - \mu(r_n)}{\sigma}\right),
\end{aligned} \tag{5.29}$$

This implies, $WPAR = 0$, for all $F_{HI} \leq F_{HI}^t$. □

Theorem 5.6 gives an F_{HI} threshold such that a safety guarantee to the primary beyond this threshold will force the cognitive radio to lose all the recoverable area ($WPAR = 0$). In order to guarantee very low F_{HI} for the primary, the threshold must be set such that the primary is protected against extremely deep fading events. In traditional terms, the resulting sensitivity requirement is beyond the SNR wall. Recall that the traditional $\mathcal{H}_0$ corresponds to $R = \infty$ since an infinitely far away primary transmitter might as well not exist at all.

Figure 5.3 shows the F_{HI} vs $WPAR$ performance for the radiometer and the matched filter under fading and noise uncertainty. The noise uncertainty model used is described in the statement of Theorem 5.6 and the fading is assumed to be uncertain, and only knowledge of the minimum length of the uncertain channel coherence time, N_c is assumed. The black curve in Figure 5.3 is the tradeoff for the radiometer with fading uncertainty and no noise uncertainty, whereas the red curve is the tradeoff for the radiometer with both fading and noise uncertainty. Notice how the noise uncertainty introduces a F_{HI} threshold below which the $WPAR$ is zero.

5.3.3.4 Dual detection: how to exploit time-diversity

A diversity perspective is interesting to consider. Since the number of samples N is infinite, one could exploit time diversity for multipath if we believed that the actual coherence time is finite $N_c < \infty$. However, for the radiometer all the thresholds must be set based on the primary user's fear of an infinite coherence time. So, the radiometer cannot do anything even if the cognitive radio designer actually believed that the coherence time is most likely finite.

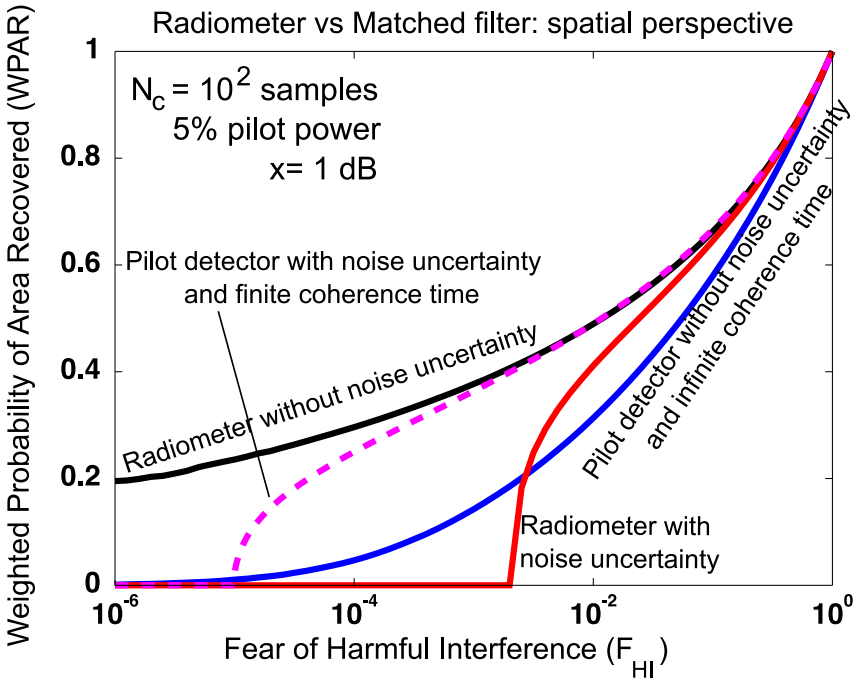


Figure 5.3: The impact of noise uncertainty from a spatial perspective is illustrated in this figure. Under noise uncertainty, there is a finite F_{HI} threshold such that if the cognitive radio needs to guarantee protection below this threshold, the area recovered by a radiometer is zero ($WPAR = 0$). The coherent detector (modified matched filter) has a more interesting set of plots discussed in the text.

Now consider the matched filter trying to sense for a sinusoidal pilot tone. As the sinusoidal pilot is narrowband the matched filter suffers from the lack of frequency diversity as compared to the radiometer: effectively, fading is more variable. This means that we have to be more conservative, which can cost us in area. However, the matched filter gives us an opportunity to deal with the situation where the coherence-time is uncertain. We can run two parallel matched filters, one assumes that the coherence time is infinity and the other assumes that the coherence time is N_c, with the thresholds set according to their respective assumptions on the coherence time. If either of these detectors declares that the band is used then the decision is that the band is used (this is an "OR" rule among the two sub-detectors). This ensures that the F_{HI} constraint is met irrespective of the actual coherence time.

From a $WPAR$ point of view, the matched filter assuming an infinite coherence time has no SNR wall (as infinite coherent processing kills the uncertainty in the noise), but is susceptible to multipath (no time-diversity to exploit) and shadowing. So, as Figure 5.3 shows, this detector loses a lot of area. The other matched filter runs using a coherence time of N_c. This enjoys time-diversity that completely wipes out multipath and so has better performance. However, this detector still has an SNR wall due to noise uncertainty. This SNR wall shows up as the $WPAR$ crashing to zero at an appropriately low F_{HI}.

The dual-detector approach leads to two different F_{HI} vs $WPAR$ curves depending on what the mix of underlying coherence times is (stationary devices or moving devices). The good thing about the dual-detector approach is that the F_{HI} is met even when the primary is uncertain, simultaneously guaranteeing the best possible $WPAR$ based on the true channel coherence time. Figure 5.3 shows this for a very short coherence time $N_c = 100$. For any realistic coherence time, the SNR wall effect would becomes negligible at all but extremely paranoid values for F_{HI}.

So Figure 5.3 shows an interesting effect. In the case when the actual coherence time is infinite, the radiometer (red curve) has a better $WPAR$ performance than the matched filter for $F_{HI} \geq 2 \cdot 10^{-3}$, even under noise uncertainty! This suggests that diversity is very important, and the lack of it can lead to poor performance. This effect of the radiometer outperforming a matched filter at high F_{HI} is analogous to the time-domain effect of the radiometer sometimes having a better sample complexity than the matched filter (see Figure 3.1).

5.4 Spectrum sensing: space-time perspective

In Section 5.3 we considered the situation when the primary is always "ON", and proposed metrics (F_{HI} and $WPAR$) to evaluate the performance of sensing algorithms. In this section, we consider the general sensing problem where the primary signal has both "ON" and "OFF" periods. Unlike the previous case, where space was very critical (the spectrum hole was

infinite in the time-dimension), in this case both space and time must be considered together.

As in Section 1.2, we assume that a primary transmitter is providing service to its receivers within a service area. As before, there is a no-talk region around the primary transmitter. The only change now is that the primary transmitter turns ON/OFF according to some random process (whose knowledge might not be known apriori to the secondary user). We assume that the length of the ON and OFF durations of the primary transmitter are much longer than the typical sensing time for the secondary user. In this situation, protection to the primary corresponds to the secondary remaining quiet within the no-talk region when the primary is ON. On the other hand, the secondary can transmit under two conditions: if it is outside the primary's no-talk radius, or if it is inside the primary's no-talk radius and the primary transmitter if OFF.

5.4.1 Space-time sensing formulation

We now formulate the most general version of the space-time sensing problem. Let $U[n]$ denote the state of the primary transmitter at each time instance n, $U[n] = 1$ if the primary transmitter is ON, and $U[n] = 0$ if the primary transmitter is OFF. Assume that the secondary decides to sense at time instances n_i, $i = 1, 2, 3, \cdots$, and let the length of each sensing interval be denoted by δ_i, where $\delta_i < n_{i+1} - n_i$. At the end of each sensing interval, the secondary makes a decision of whether the band is safe to use ($D_i = 0$) or not safe to use ($D_i = 1$). If the sensing algorithm declares the band to be unused, then we assume that the secondary transmits its signal until the start of the next sensing epoch n_{i+1}, otherwise the secondary user does not transmit and waits for the outcome of the next sensing decision. Based on the sensing decisions, we can construct a random process $Z^r[n] \in \{0, 1\}$ which denotes the state of the secondary user located at a distance r from the primary transmitter. $Z^r[n] = 1$ implies that the secondary user is transmitting (ON), and $Z^r[n] = 0$ implies it is not transmitting (OFF). An example scenario illustrating $U[n]$, the sensing epochs and $Z^r[n]$ are shown in Figure 5.4. Given the problem setup we now define the two main metrics of interest:

5.4.1.1 Protection

Intuitively, interference to the primary can be quantified by measuring the fraction of the time (ON time of the primary) the secondary is transmitting within the no-talk region when the primary is also ON. However, there is a problem with this definition as illustrated by the scenario below.

Assume that the primary transmitter if OFF, and that a secondary user senses the primary for N samples and declares that the primary is OFF and hence starts transmitting its packets. Now, there is a finite probability (strictly non-zero and does not scale with the sensing time N) that the primary comes back ON while the secondary is transmitting,

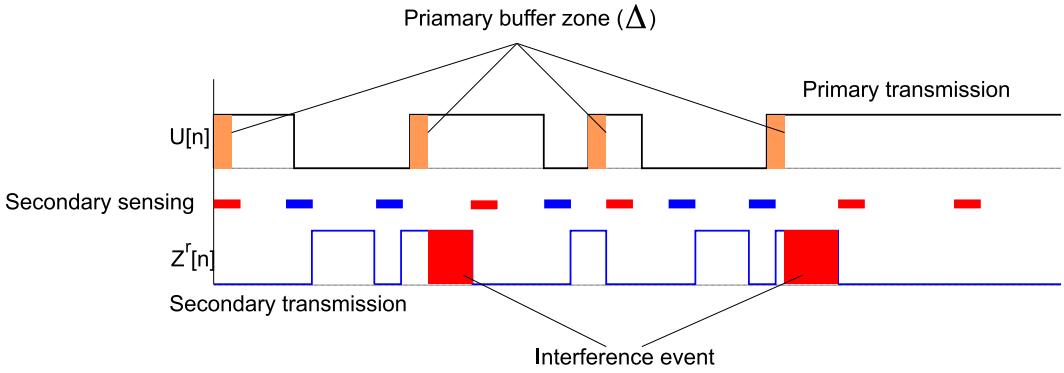


Figure 5.4: The figure illustrates the problem set up for space-time spectrum sensing. The state of the primary $U[n]$, the sensing epochs, as well as the secondary ON/OFF process $Z^r[n]$ are shown in the figure. The 'red' colored sensing windows indicate the events where the detector declares the band to be used, and the 'green' colored sensing windows indicate the events where the detector declared the band to be unused. The primary buffer zones are shown by the 'tan' shaded regions on the function $U[n]$, and the actual interference events are shown by shaded 'red' regions on $Z^r[n]$.

leading to an interference event. According to our definition, the secondary has caused interference to the primary even when its detector was correct. So, with the current definition of interference, the only way to drive the probability of interference to zero is to scale the secondary transmission time to zero. This gives an uninteresting tradeoff between the protection to the primary system and the performance of the secondary user[6].

To avoid running into the scenario described above, we assume that there is a sacrificial buffer zone (Δ samples) at the beginning of every OFF to ON transition of the primary user within which interference from the secondary user is tolerated. This is shown by the shaded 'tan' region in Figure 5.4. This is equivalent to the primary user potentially losing some marginal users (operating close to the noise-limited contour) due to the potential operation of the secondary user [31].

In this formulation we assume that the primary packets are of moderate length and if the secondary interferes with the primary's packet (outside the buffer zone), then we fear that the whole packet may be lost. So, a reasonable metric to protect the primary from losing packets is to constraint the total fraction of mis-detection by the secondary. So, the fear of

[6]A similar conclusion was presented in [133]. In that paper the authors model the primary ON/OFF distributions as exponentially distributed, and they show that the fraction of time the secondary can use the band (performance) is proportional to the fraction of time the primary is interfered with (interference).

harmful interference is same as the F_{HI} definition in Section 5.3.1.1.

$$F_{HI} = \sup_{0 \le r \le r_n} \sup_{F_r \in \mathbb{F}_r} \mathcal{P}_{F_r}(D = 0 | R = r). \tag{5.30}$$

As before, we assume a worst-case location of the secondary user to compute the interference caused to the primary, i.e., we measure the interference caused by the secondary located at the edge of the primary no-talk radius.

5.4.1.2 Performance

The metric that measures the performance of the secondary user can be defined similar to the definition in Section 5.3.1.2. For each location r, we can compute the total fraction of space-time opportunities recovered by the sensing algorithm

$$P_{STR}(r) = \begin{cases} \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathbf{I}_{(Z^r[n]=1)} & \text{if} \quad r > r_n \\ \lim_{M \to \infty} \frac{\frac{1}{M} \sum_{n=1}^{M} \mathbf{I}_{(Z^r[n]=1, U[n]=0)}}{\frac{1}{M} \sum_{n=1}^{M} \mathbf{I}_{(U[n]=0)}} & \text{if} \quad r \le r_n \end{cases}$$

Finally, we assume a distribution (weighting function) $w(r)$ on the location of the secondary users, and use this to compute the *weighted probability of space-time recovered* ($WPSTR$) by

$$WPSTR = \int_0^\infty P_{STR}(r) w(r) r dr, \tag{5.31}$$

where $\int_0^\infty w(r) r dr = 1$. Notice that the integral spans locations inside and outside the no-talk radius (0 to ∞). Also, the weighting function $w(r)$ is non-zero for $r \le r_n$. This is different than the definition of $WPAR$ in (5.4) where the integral is only from the edge of the no-talk radius to infinity, and $w(r) = 0$ for $r \le r_n$. In fact the only information about the value of the weighting function inside the no-talk radius $w(r), 0 \le r \le r_n$ that is relevant to compute $WPSTR$ is the integral $\int_{r=0}^{r_n} w(r) r dr$. This is because the probability of finding a hole inside the no-talk radius is the probability that the noise does not trigger the detection threshold $(1 - pfa)$, which is independent of the spatial location r.

5.4.2 Memoryless sensing algorithms

Having formulated the metrics for space-time spectrum sensing in Section 5.4.1, we now consider a sub-optimal class of algorithms and evaluate their F_{HI} vs $WPSTR$ performance. Even these restrictive class of algorithms bring out some interesting tradeoffs which were absent in the space only scenario. The assumptions we make are:

- We consider memoryless detection algorithms with fixed sensing time. That is, we assume that the detector senses for N samples and makes a decision based on the measured N samples, and it does not consider any past information about the channel. This is clearly sub-optimal because memory could help significantly [77]. However, memoryless algorithms are easy to certify as compared to the ones with memory.

- We assume that the transmit time for the secondary is fixed. That is, whenever the sensing algorithm declares the band to be unused, the secondary transmits for N_{TX} number of samples before sensing the channel again. Also, if the detector declares the channel to be used, then it senses the channel again after N_{TX} samples. We make this assumption to make the analysis simple.

- We assume that the mean length of the ON/OFF time of the primary are much larger than the sensing time N.

- As the sensing time N is much smaller than the mean ON/OFF duration, the chance of a sensing window of N samples containing a primary state transition (ON $\longrightarrow$ OFF, or OFF $\longrightarrow$ ON) is very low. In this section, we ignore such transitions and assume that every sensing window (N samples) has either the primary ON or the primary OFF. In particular, to compute F_{HI} we assume that the primary is ON for the whole of the sensing window (N samples). This approximation can lead to higher rate of missed-detections because if the primary turns ON somewhere within the sensing window, there is a chance that the energy detector will not trigger the threshold due to less than typical energy in the sensing window. To counter this we enforce $N + N_{TX} \leq \Delta$.

- Finally, for simplicity we consider a radiometer (energy detector). The analysis in this section can easily be extended to other single user as well as cooperative sensing algorithms.

5.4.2.1 F_{HI} for a radiometer

Given a sensing window of N samples, the secondary interferes with the primary user only if the secondary is within the no-talk radius, the primary is ON, and the secondary user

mis-detects the primary. The fear of harmful interference is given by

$$\begin{aligned}
&F_{HI} \\
&= \mathcal{P}\left(T(\mathbf{Y}) < \lambda | r = r_n, ON\right) \\
&= \mathcal{P}\left(\frac{1}{N}\sum_{n=1}^{N} |\sqrt{P(r_n)}X[n] + W[n]|^2 < \lambda\right) \\
&= \int_{-\infty}^{\infty} \mathcal{P}\left(\frac{1}{N}\sum_{n=1}^{N} |\sqrt{10^{p/10}}X[n] + W[n]|^2 < \lambda \right. \\
&\qquad \left. \Big| P(r_n) = 10^{\frac{p}{10}}\right) \mathcal{P}(P(r_n) = 10^{\frac{p}{10}}) dp \\
&\approx \int_{-\infty}^{\infty} \left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{2/N}(10^{\frac{p}{10}} + \sigma_w^2)}\right)\right] f_{r_n}(p) dp. \qquad (5.32)
\end{aligned}$$

In the above derivation we assumed that the received signal power (in dBm) is log-normal distributed. That is, $10\log_{10} P(r) \sim \mathcal{N}(\mu(r), \sigma^2)$, where $\mu(\cdot)$ is a deterministic function of r. Also, σ_w^2 is used to denote the noise variance, $\mathcal{Q}(\cdot)$ is the Gaussian tail-probability function and $f_r(\cdot)$ is the Gaussian probability density function given by $f_r(p) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{p-\mu(r)}{2\sigma^2}\right)$, where p is the signal power in dBm.

5.4.2.2 $WPSTR$ for a radiometer

The performance of a radiometer can be characterized along two dimensions: time-recovery and spatial-recovery. As the radiometer uses N samples out of a total of Δ samples for sensing, the fraction of time available for secondary transmissions is given by $1 - \frac{N}{\Delta}$. This can be thought of as the time recovery metric.

The space-recovered metric is similar to the WPAR computed in Section 5.3.1.2. The only difference now is that the secondary can also reuse the spectrum inside the no-talk radius when the primary is OFF. As before, to compute the WPAR we first compute the probability of finding a hole as a function of the spatial location

$$\begin{aligned}
P_{FH}(r) &= \mathcal{P}\left(T(\mathbf{Y}) < \lambda\right) \\
&= \begin{cases} p_{off}\left(T(\mathbf{Y}) < \lambda \big| OFF\right) & \text{if } r \le r_n \\ \\ p_{off}\left(T(\mathbf{Y}) < \lambda \big| OFF\right) & \\ +p_{on}\left(T(\mathbf{Y}) < \lambda \big| OFF\right) & \text{if } r > r_n \end{cases} \qquad (5.33)
\end{aligned}$$

where p_{on} is the probability that the primary transmitter is ON, and p_{off} is the probability

that the primary transmitter is OFF. Furthermore,

$$\begin{aligned}\mathcal{P}\left(T(\mathbf{Y})<\lambda\middle|OFF\right) &= \mathcal{P}\left(\frac{1}{N}\sum_{n=1}^{N}|W[n]|^2<\lambda\right)\\ &= 1-\mathcal{Q}\left(\frac{\lambda-\sigma_w^2}{\sqrt{\frac{2}{N}}\sigma_w^2}\right)=:1-pfa,\end{aligned}$$

where pfa is used to denote the false-alarm probability of the radiometer. Similarly,

$$\begin{aligned}&\mathcal{P}\left(T(\mathbf{Y})<\lambda\middle|ON\right)\\ &=\mathcal{P}\left(\frac{1}{N}\sum_{n=1}^{N}|\sqrt{P(r)}X[n]+W[n]|^2<\lambda\right)\\ &\approx\int_{-\infty}^{\infty}\left[1-\mathcal{Q}\left(\frac{\lambda-(10^{\frac{p}{10}}+\sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}}+\sigma_w^2)}\right)\right]f_r(p)dp\end{aligned}$$

Weighting the probability of finding a hole $P_{FH}(r)$ using the weighting function $w(r)r$ gives us the weighted probability of space recovered metric:

$$\begin{aligned}&WPSR\\ &=\int_0^{\infty}P_{FH}(r)w(r)rdr\\ &=p_{off}(1-pfa)+p_{on}\int_{r_n}^{\infty}P_{FH}(r)w(r)rdr\\ &=p_{off}(1-pfa)\\ &+p_{on}\left(\int_{r_n}^{\infty}w(r)rdr\right)\int_{r_n}^{\infty}P_{FH}(r)\left[\frac{w(r)}{\int_{r_n}^{\infty}w(r)rdr}\right]rdr\\ &=p_{off}(1-pfa)+p_{on}\cdot\beta\cdot\int_{r_n}^{\infty}P_{FH}(r)\widetilde{w}(r)rdr,\end{aligned}\tag{5.34}$$

where $\beta:=\int_{r_n}^{\infty}w(r)rdr$ denotes the weighted fraction of the area outside the no-talk radius, with the convention that $\int_0^{\infty}w(r)rdr=1$. Also, $\widetilde{w}(r):=\frac{w(r)}{\int_{r_n}^{\infty}w(r)rdr}$ can be thought of as a new weighting function for area outside the no-talk radius. Note, by definition we have $\int_{r_n}^{\infty}\widetilde{w}(r)rdr=1$. This gives a new interpretation to Equation (5.34): WPAR is a linear weighting of the area recovered when the primary is OFF $(1-pfa)$ and the area recovered when the primary is ON $(\beta\cdot\int_{r_n}^{\infty}P_{FH}(r)\widetilde{w}(r)rdr\ )$.

Finally, the space-time recovery metric is given by

$$\begin{aligned} WPSTR &= \text{Time metric } \cdot \text{Space metric} \\ &= \left[1 - \frac{N}{\Delta}\right] \cdot WPSR \end{aligned} \tag{5.35}$$

Remark: The $WPSTR$ expression derived in (5.35) is same as the one obtained by evaluating the expression in (5.31) for the special case of memoryless sensing algorithms.

5.4.2.3 Numerical simulations

In this Section we give some numerical simulations for a single-radio energy detector trying to detect an ON/OFF primary transmitter with a given duty cycle. Figure 5.5 shows that F_{HI} vs $WPSTR$ performance of a radiometer for different values of the sensing time N. Each curve in the figure shows this tradeoff for a fixed sensing time N. It is clear from the figure that the optimal sensing time is a function of the protection metric F_{HI}. In fact, the best possible tradeoff is given by the envelope of these performance curves.

In Figure 5.6 we consider the $WPSTR$ performance as a function of the sensing time N for a fixed value of protection, $F_{HI} = 0.01$. The figure shows that there is an optimal value of N for which the space-time performance is maximized. This is contrary to the traditional wisdom which tells us that the sensing performance is a monotonic function of the sensing time N. However, our space-time framework shows that there is a tension between space and time. This tradeoff between performance in time and performance in space is shown in Figure 5.7.

Finally, Figure 5.8 shows the effect of increasing sensing time on $WPSTR$. However, the sensing time is constrained by the sacrificial buffer-zone Δ. So, to see the effect of very large sensing times N we scale Δ proportional to N while keep the ratio $\frac{N}{\Delta}$ to be constant at 0.5. The figure shows that the performance of a cognitive radio saturates to a finite value (strictly less than 0.5) even as N is increased to infinity. This shows that increasing the sensing time does not necessarily guarantee improved performance.

5.5 Concluding remarks

In this chapter we showed that the traditional binary hypothesis testing formulation with the two hypotheses being "primary ON" and "primary OFF" does not completely reveal the true fundamental tradeoffs involved in spectrum sensing. This formulation is particularly bad for spatially distributed primary users. The metrics of sensitivity, probability of false-alarm and probability of missed detection arising out of the traditional formulation fail to capture the tradeoffs in the spatial dimension of the spectrum sensing problem. In fact, using sensitivity

as a metric the secondary sensor is forced to sense at very low SNRs (high sensitivity) to guarantee protection to the primary user (as in the -114 dBm rule). Unfortunately, sensing at such low SNRs forces the secondary user to lose a significant portion of the spatial spectrum holes as shown in Section 5.3.3.1.

We proposed a new joint space-time formulation for the spectrum sensing problem that addresses the drawbacks in the traditional formulation. Using our new formulation, we derived two new metrics, F_{HI} and $WPSTR$ that quantify safety to the primary user as well as performance for the secondary users. The tradeoff between F_{HI} and $WPSTR$ is the new (and fundamental) ROC curve for this problem.

With our new space-time metrics framework we show that the F_{HI} vs $WPSTR$ performance for any single-radio detector is poor. The reason for this is that a single-radio detector cannot robustly quantify its shadowing environment and hence is forced to budget for deep (and atypical) shadow fades. Our result suggests that one needs to look at other sensing strategies like collaborative sensing, multiband sensing approaches, etc to improve performance. The key is to have a robust way for the secondary user to conclude that it is not deeply shadowed (which is the typical case) and hence can aggressively set its detector threshold.

One possible approach is to use memory to know your shadowing environment robustly. That is, if a secondary sees a strong primary signal in the near past, then it knows that it is not deeply shadowed. If the shadowing environment does not change rapidly (slowly moving secondary and primary users), the secondary user can use this to set an aggressive threshold at the current time. This suggests that change-detection based algorithms can improve performance.

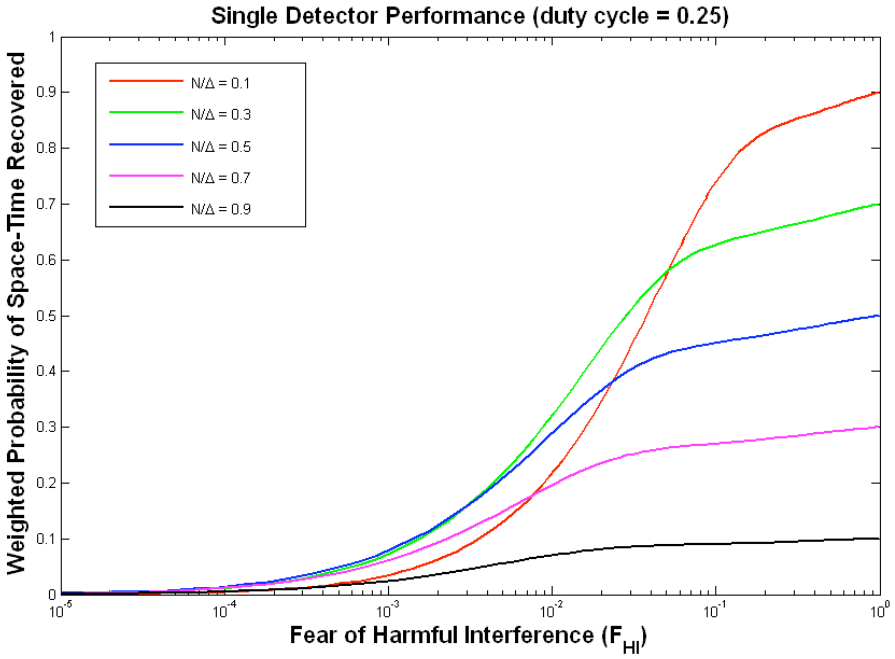


Figure 5.5: F_{HI} vs $WPSTR$ for different values of N. In this figure $p_{on} = 0.25$, and $p_{off} = 0.75$.

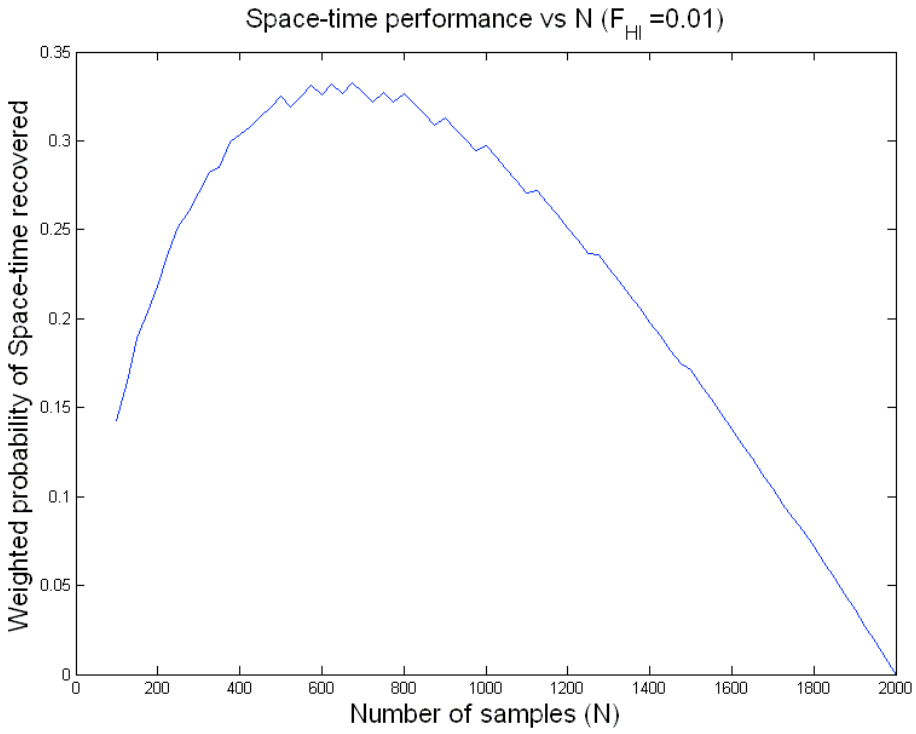


Figure 5.6: WPSTR as a function of sensing time N for fixed $F_{HI} = 0.01$. In this figure $p_{on} = 0.25$, and $p_{off} = 0.75$

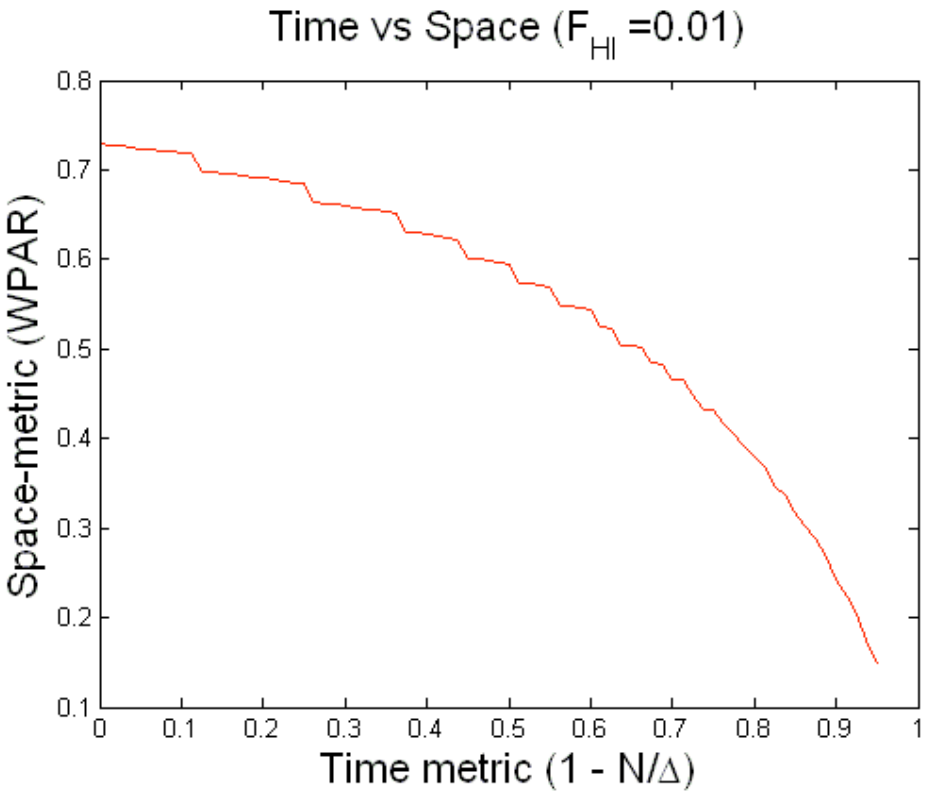


Figure 5.7: Time recovery metrics vs Space recovery metric for a fixed value of $F_{HI} = 0.01$. In this figure $p_{on} = 0.25$, and $p_{off} = 0.75$

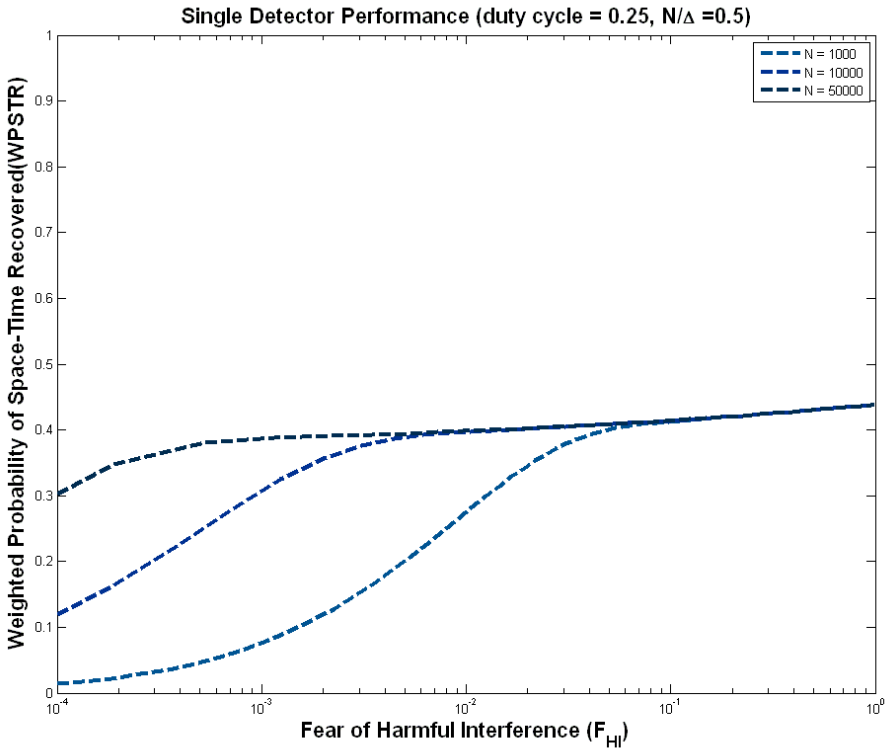


Figure 5.8: The $WPSTR$ performance as a function of the sensing time N as we increase N to infinity keeping $\frac{N}{\Delta} = 0.5$.

Chapter 6

Conclusions

This thesis gives a unified space-time formulation for the spectrum sensing problem. The two main objectives for cognitive radios are: to provide adequate safety guarantees to the primary user; and recover as many spectral opportunities as possible. Our formulation naturally results in two new performance metrics, namely the *Fear of Harmful Interference* (F_{HI}), and the Weighted Probability of Space-Time Recovered ($WPSTR$) that quantify the two objectives mentioned above.

The joint space-time sensing framework lays a common foundation to compare various detection algorithms. We show that the performance of single-radio sensing algorithms is poor, even in the limit of infinite sensing time. The key idea behind this result is that the performance of single-radio detectors is limited due to the fact that they cannot differentiate between atypical deep shadowing at locations close to the primary user and typical shadowing when it is far away. It is the uncertainty in the fading environment that is the bottleneck, which cannot be resolved by designing highly sensitive detection algorithms. This suggests that we should focus on other sensing strategies like cooperative sensing, multi-band sensing, and geo-location based sensing algorithms to recover most of the unused spectral opportunities.

The new framework also allows for a unified treatment of uncertainties in the system. Both device-level uncertainties like distributional uncertainties in the noise and multipath fading, and system-level uncertainties like uncertainty in propagation models, and shadow fading can be incorporated in our framework. This allows us to quantify the impact of uncertainties on the performance of sensing algorithms (single-user as well as multi-user based strategies). This framework also brings out the tradeoff between the space-domain performance and the time-domain performance of cognitive radios.

It is the quest for designing highly sensitive robust detectors that motivated us to come up with this new space-time perspective. Our results showed that designing robust algorithms that sense at low SNRs is very challenging. There exist fundamental limits to robust sensing in the presence of uncertainties. These limits are in the form of finite SNR thresholds, called

SNR walls, below which robust detection is impossible, irrespective of the number of samples we take. We showed the existence of SNR wall limits for several commonly used sensing algorithms like a radiometer, pilot detector, and cyclostationary feature detectors, although some are more robust than others. It is the presence of known features in signals (like pilot tones, cyclic prefixes in OFDM symbols etc) that allow for coherent processing and run-time noise calibration, which provides robustness gains. In fact, we also construct signals containing macroscale features that can be robustly detected at arbitrarily low SNRs.

In retrospect, we know that a cognitive radio sensor would not want to set its threshold to catch extremely weak signals (even if we can design such a robust detector). So, our results on designing robustly detectable signals at low SNRs are not directly useful in the problem of spectrum sensing by cognitive radios. However, these results are very useful in other problems in opportunistic spectrum sharing. For instance, in opportunistic spectrum sharing systems, primary users will want to catch secondary users interfering with its transmissions. In this problem, the macroscale feature idea can be used to design codes that provide robust identity to secondary users [121]. In some spectrum sharing systems, the primary can explicitly send beacon signals to inform the secondary user whether the primary band can be used or not. Even in this case, our results on robust signal features are very useful. Our results are also applicable for primary users like wireless microphones, in which case the spectral holes are mainly in the time-domain.

Appendices

Appendix A

Proofs for Chapter 2

A.1 Proof of Theorem 2.2

The goal is to show the existence of an absolute SNR wall, below which *every detector* is non-robust. This requires all possible test statistics to have overlapping means under the two hypotheses. The easiest way for this to happen is if the set of joint distributions of the received signal vector $\mathbf{Y}$ under both hypotheses overlap. As the noise and signal are both modeled as *iid*, it is sufficient to show that the set of marginal distributions of the scalar received signal samples $Y[n]$ overlap under both hypotheses. Mathematically, we just need to show the existence of two noise distributions $W_1, W_2 \in \mathcal{W}_x$, such that $W_1 + X$ has the same distribution as W_2.

Let $W_n \sim \mathcal{N}(0, \sigma_n^2)$, $W_1 \sim \frac{1}{\sqrt{\rho}} W_n$. Let $W_2 := X + W_1$. If $W_1 \in \mathcal{W}_x$, and $W_2 \in \mathcal{W}_x$, then it is clear that robust detection is impossible. We now prove $W_2 \in \mathcal{W}_x$ by showing that $\mathbb{E}W_2^{2k-1} = 0$ and $\frac{1}{\rho^k}\mathbb{E}W_n^{2k} \leq \mathbb{E}W_2^{2k} \leq \rho^k \mathbb{E}W_n^{2k} \ \ \forall\ k$.

Since $\mathbb{E}[X^{2k-1}] = 0$ and $W_2 =: X + W_1$, we have $\mathbb{E}W_2^{2k-1} = 0$ and $\mathbb{E}W_2^{2k} \geq \frac{1}{\rho^k}\mathbb{E}W_n^{2k}$.

Therefore, $W_2 \in \mathcal{W}_x$

$$\begin{aligned}
&\Leftrightarrow \quad \mathbb{E}W_2^{2k} \leq \rho^k \, \mathbb{E}W_n^{2k} \\
&\Leftrightarrow \quad \frac{\mathbb{E}(X+W_1)^{2k}}{\mathbb{E}W_n^{2k}} \leq \rho^k \\
&\Leftrightarrow \quad \sum_{i=0}^{k} \binom{2k}{2i} \frac{\mathbb{E}W_1^{2k-2i}}{\mathbb{E}W_n^{2k}} \mathbb{E}X^{2i} \leq \rho^k \\
&\Leftrightarrow \quad \sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1}{\rho^{k-i}}\right) \left(\frac{1 \cdot 3 \cdots (2k-2i-1)\ \sigma_n^{2k-2i}}{1 \cdot 3 \cdots (2k-1)\ \sigma_n^{2k}}\right) \mathbb{E}X^{2i} \leq \rho^k \\
&\Leftrightarrow \quad \sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1}{\rho^{k-i}}\right) \left(\frac{1 \cdot 3 \cdots (2k-2i-1)}{1 \cdot 3 \cdots (2k-1)}\right) \frac{\mathbb{E}X^{2i}}{\sigma_n^{2i}} \leq \rho^k, \quad \forall\ k = 1, 2, \cdots . \quad \text{(A.1)}
\end{aligned}$$

So, we just need to show that there exits an absolute threshold SNR_{wall}^* such that that the last inequality in (A.1) is true if $SNR_{peak} \leq SNR_{wall}^*$.

We now construct the threshold SNR_{wall}^*. For $k \geq 1$, define $f(k, SNR)$ to be the following uni-variate degree k polynomial

$$f(k, SNR) \quad = \quad \left[\sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1}{\rho^{k-i}}\right) \left(\frac{1 \cdot 3 \cdots (2k-2i-1)}{1 \cdot 3 \cdots (2k-1)}\right) SNR^i\right]. \qquad \text{(A.2)}$$

Note that all the coefficients of $f(k, SNR)$ are non-negative and hence $f(k, SNR)$ is monotonically increasing in SNR. This means that the equation $f(k, SNR) = \rho^k$ has a unique real valued root $SNR_{wall}^{(2k)}$. Because $f(k, 0) = \frac{1}{\rho^k} < \rho^k$ and $\rho > 1$, the root $SNR_{wall}^{(2k)} > 0$ for all k. Define SNR_{wall}^* to be

$$SNR_{wall}^* := \inf_{k=1,2,3,\cdots\infty} SNR_{wall}^{(2k)}. \qquad \text{(A.3)}$$

By definition of SNR_{peak}, we have $\frac{\mathbb{E}X^{2i}}{\sigma_n^{2i}} \leq (SNR_{peak})^i$. Now, if $SNR_{peak} \leq SNR_{wall}^*$, by

the monotonicity of $f(k, SNR)$, we have

$$\begin{aligned}
& \sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1}{\rho^{k-i}}\right) \left(\frac{1 \cdot 3 \cdots (2k-2i-1)}{1 \cdot 3 \cdots (2k-1)}\right) \frac{\mathbb{E}X^{2i}}{\sigma_n^{2i}} \\
\leq & \sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1}{\rho^{k-i}}\right) \left(\frac{1 \cdot 3 \cdots (2k-2i-1)}{1 \cdot 3 \cdots (2k-1)}\right) SNR_{peak}^{i} \\
= & f(k, SNR_{peak}) \\
\leq & f(k, SNR_{wall}^{*}) = \rho^k,
\end{aligned}$$

which is the last inequality in (A.1).

In terms of the upper-bound in (2.10), from (A.3) it is clear that $SNR_{wall}^{*} \leq SNR_{wall}^{(2)} = \frac{\rho^2-1}{\rho}$. However, the entire result is vacuous if the infimum in (A.3) is zero. Getting a nontrivial lower bound is more interesting.

Recall that $SNR_{wall}^{(2k)}$ is the root of the polynomial $f(k, SNR) - \rho^k$. The coefficient of SNR^i in $f(k, SNR) - \rho^k$ is $\binom{2k}{2i}\left(\frac{1}{\rho^{k-i}}\right)\left(\frac{1 \cdot 3 \cdot 5 \cdots (2k-2i-1)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}\right)$. Replace $\left(\frac{1}{\rho^{k-i}}\right)$ by 1 to get a new polynomial, $\widetilde{f}(k, SNR) - \rho^k$. Since $\rho > 1$, we have $\left(\frac{1}{\rho^{k-i}}\right) \leq 1$ for all $i = 1, 2, \cdots, k$, and hence the unique positive root of $\widetilde{f}(k, SNR) - \rho^k = 0$ must be smaller than $SNR_{wall}^{(2k)}$. Call this root $\widetilde{SNR}_{wall}^{(2k)}$. Clearly,

$$SNR_{wall}^{*} \geq \inf_{k=1,2,\cdots,\infty} \widetilde{SNR}_{wall}^{(2k)} \tag{A.4}$$

where $\widetilde{SNR}_{wall}^{(2k)}$ is the unique positive root of

$$\widetilde{f}(k, SNR) \;:=\; \left[\sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1 \cdot 3 \cdots (2k-2i-1)}{1 \cdot 3 \cdots (2k-1)}\right) SNR^i\right] = \rho^k. \tag{A.5}$$

Fix $k \geq 1$ and $SNR \geq 0$. From the definition of $\widetilde{f}(k, SNR)$ in (A.5), we get

$$\begin{aligned}
\widetilde{f}(k, SNR) &= \left[\sum_{i=0}^{k} \binom{2k}{2i} \left(\frac{1 \cdot 3 \cdot 5 \cdots (2k-2i-1)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}\right) SNR^i\right] \\
&= \left[\sum_{i=0}^{k} \frac{(2k) \cdot (2k-1) \cdots (2k-2i+1)}{(2i) \cdot (2i-1) \cdots 1} \left(\frac{1 \cdot 3 \cdot 5 \cdots (2k-2i-1)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}\right) SNR^i\right] \\
&\overset{(a)}{=} \left[\sum_{i=0}^{k} \binom{k}{i} \frac{(2k-1) \cdot (2k-3) \cdots (2k-2i+1)}{(2i-1) \cdot (2i-3) \cdots 1} \cdot \frac{1 \cdot 3 \cdot 5 \cdots (2k-2i-1)}{1 \cdot 3 \cdot 5 \cdots (2k-1)} SNR^i\right] \\
&\overset{(b)}{=} \left[\sum_{i=0}^{k} \binom{k}{i} \frac{1}{(2i-1) \cdot (2i-3) \cdots 1} SNR^i\right] \\
&\leq \left[\sum_{i=0}^{k} \binom{k}{i} SNR^i\right] = (1+SNR)^k. \qquad \text{(A.6)}
\end{aligned}$$

- (a) follows from the following calculation.

$$\begin{aligned}
&\frac{(2k) \cdot (2k-1) \cdots (2k-2i+1)}{(2i) \cdot (2i-1) \cdots 1} \\
= &\left(\frac{(2k) \cdot (2k-2) \cdots (2k-2i+2)}{(2i) \cdot (2i-2) \cdots 2}\right) \cdot \left(\frac{(2k-1) \cdot (2k-3) \cdots (2k-2i+1)}{(2i-1) \cdot (2i-3) \cdots 1}\right) \\
= &\left(\frac{2^i \cdot (k) \cdot (k-1) \cdots (k-i+1)}{2^i \cdot (i) \cdot (i-1) \cdots 1}\right) \cdot \left(\frac{(2k-1) \cdot (2k-3) \cdots (2k-2i+1)}{(2i-1) \cdot (2i-3) \cdots 1}\right) \\
= &\binom{k}{i} \left(\frac{(2k-1) \cdot (2k-3) \cdots (2k-2i+1)}{(2i-1) \cdot (2i-3) \cdots 1}\right).
\end{aligned}$$

- (b) follows from the following calculation.

$$\begin{aligned}
&\left[\frac{(2k-1) \cdot (2k-3) \cdots (2k-2i+1)}{(2i-1) \cdot (2i-3) \cdots 1} \cdot \frac{1 \cdot 3 \cdot 5 \cdots (2k-2i-1)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}\right] \\
= &\left[\frac{1}{(2i-1) \cdot (2i-3) \cdots 1} \cdot \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}\right] = \left[\frac{1}{(2i-1) \cdot (2i-3) \cdots 1}\right].
\end{aligned}$$

Using (A.6) in (A.5) gives

$$\rho^k = \widetilde{f}(k, \widetilde{SNR}_{wall}^{(2k)}) \leq (1 + \widetilde{SNR}_{wall}^{(2k)})^k.$$

This implies that $\widetilde{SNR}_{wall}^{(2k)} \geq \rho - 1$, for all $k \geq 1$. Substituting this in (A.4) gives the desired lower bound.

Appendix B

Proofs for Chapter 3

B.1 Proof of Lemma 3.4

Under hypothesis $\mathcal{H}_0$: $Y[n] = W[n] \sim \mathcal{N}(0, \sigma^2)$ for all $n > 0$. Using this in (3.18), we get

$$\mathbb{E}[T(\mathbf{Y}, D_c)|\mathcal{H}_0] = \mathbb{E}\left[\left|\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\left\{|W[2n-1]|^2 - |W[2n]|^2\right\}\right|^2\right]. \tag{B.1}$$

$|W[2n-1]|^2 - |W[2n]|^2$ are *iid* random variables with zero mean and variance $4\sigma^4$. Using the Central Limit Theorem[86], we get

$$\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\left\{|W[2n-1]|^2 - |W[2n]|^2\right\} \sim \mathcal{N}(0, 4\sigma^4).$$

Using this approximation in (B.1), we get $\mathbb{E}[T(\mathbf{Y}, D_c)|\mathcal{H}_0] = 4\sigma^4$.

Under hypothesis $\mathcal{H}_1$: Without loss of generality we can assume $l[n] = 0$. So, we have $Y[2n-1] = X[2n-1] + W[2n-1]$ and $Y[2n] = X[2n] + W[2n] = W[2n]$. This implies that $|Y[2n-1]|^2 - |Y[2n]|^2$ are *iid* random variables with mean $2P$ and variance $8P\sigma^2 + 4P^2$. Again using the Central Limit Theorem[86], we get

$$\sqrt{\frac{2}{D_c}}\sum_{n=1}^{D_c/2}\left\{|Y[2n-1]|^2 - |Y[2n]|^2\right\} \sim \mathcal{N}(\sqrt{\frac{D_c}{2}}(2P),\ 8P\sigma^2 + 4\sigma^4).$$

Using this approximation we get

$$\begin{aligned}\mathbb{E}[T(\mathbf{Y}, D_c)|\mathcal{H}_1] &= 8P\sigma^2 + 4\sigma^4 + \left(\sqrt{\frac{D_c}{2}}(2P)\right)^2 \\ &= 8P\sigma^2 + 4\sigma^4 + 2P^2 D_c\end{aligned}$$

B.2 Proof of Lemma 3.6

Recall $\widehat{Y}[n] = \frac{Y[n]}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}|Y[n]|^2}}$. Since we are interested in robustness results, we assume that N is very large. Specifically, we assume that $\frac{1}{N}\sum_{n=1}^{N} Y^2[n] = P + \sigma^2$ under hypothesis $\mathcal{H}_1$ and $\frac{1}{N}\sum_{n=1}^{N} Y^2[n] = \sigma^2$ under hypothesis $\mathcal{H}_0$, where $\sigma^2 = (1+\alpha^2)\sigma_m^2$.

Under hypothesis $\mathcal{H}_1$: From (3.23), we can write $G_1 = K_1 S_1$, and $G_2 = K_1 S_2$, where

$$S_1 = \sum_{n=1}^{\frac{D_c}{2}} Y^2[2n-1], \quad S_2 = \sum_{n=1}^{\frac{D_c}{2}} Y^2[2n],$$

and $K_1 = \sqrt{\frac{2}{D_c}} \cdot \frac{1}{P+\sigma^2}$. Now, $\mathbb{E}S_1 = \sum_{n=1}^{\frac{D_c}{2}} \mathbb{E}Y^2[2n-1] = \sum_{n=1}^{\frac{D_c}{2}} \mathbb{E}(X[2n-1] + W[2n-1])^2 = \frac{D_c}{2}[2P+\sigma^2]$. Similarly, $\mathbb{E}S_2 = \sum_{n=1}^{\frac{D_c}{2}} \mathbb{E}Y^2[2n] = \sum_{n=1}^{\frac{D_c}{2}} \mathbb{E}(W[2n])^2 = \frac{D_c}{2}[\sigma^2]$. Here we have used the fact that the random delay in the current coherent block is $l[n] = 0$. Therefore, $X[2n-1]$ is $\pm\sqrt{2P}$ with probability $\frac{1}{2}$ for each sign and $X[2n] = 0$. Hence, we get $\mathbb{E}[G_1] = K_1\mathbb{E}[S_1] = \frac{1}{P+\sigma^2}\left(\frac{D_c}{2}\right)[2P+\sigma^2]$ and $\mathbb{E}[G_2] = K_1\mathbb{E}[S_2] = \frac{1}{P+\sigma^2}\frac{D_c}{2}[\sigma^2]$.

We now compute the second order statistics of (G_1, G_2). By definition, we have $\mathbb{E}(G_1 - \mathbb{E}G_1)^2 = K_1^2\mathbb{E}(S_1 - \mathbb{E}S_1)^2$ and $\mathbb{E}(G_2 - \mathbb{E}G_2)^2 = K_1^2\mathbb{E}(S_2 - \mathbb{E}S_2)^2$. Now,

$$\begin{aligned}\mathbb{E}S_1^2 &= \mathbb{E}\left[\sum_{n=1}^{\frac{D_c}{2}} Y^2[2n-1]\right]^2 \\ &= \sum_{n=1}^{\frac{D_c}{2}} \mathbb{E}(Y^4[2n-1]) + \sum_{m\neq n} \mathbb{E}(Y^2[2n-1]Y^2[2m-1]). \qquad \text{(B.2)}\end{aligned}$$

Using the fact that $Y[2n-1] = X[2n-1] + W[2n-1]$, we can easily show that $\mathbb{E}(Y^4[2n-1]) = 4P^2 + 12P\sigma^2 + 3\sigma^4$ and $\mathbb{E}(Y^2[2n-1]Y^2[2m-1]) = 4P^2 + 4P\sigma^2 + \sigma^4$. Substituting this

in (B.2) we get

$$\mathbb{E}S_1^2 = \frac{D_c}{2}[4P^2 + 12P\sigma^2 + 3\sigma^4] + \frac{D_c}{2}\left(\frac{D_c}{2} - 1\right)[4P^2 + 4P\sigma^2 + \sigma^4].$$

Therefore,

$$\begin{aligned} \mathrm{Var}(S_1) = \mathbb{E}S_1^2 - (\mathbb{E}S_1)^2 &= \frac{D_c}{2}[8P\sigma^2 + 2\sigma^4] \\ \Rightarrow \mathrm{Var}(G_1) = K_1^2\mathrm{Var}(S_1) &= \frac{[8P\sigma^2 + 2\sigma^4]}{(P+\sigma^2)^2}. \end{aligned}$$

Identical computations show that $\mathrm{Var}(G_2) = \frac{[2\sigma^4]}{(P+\sigma^2)^2}$.

$$\begin{aligned} \mathbb{E}[S_1S_2] = \quad & \mathbb{E}\left[\left(\sum_{n=1}^{\frac{D_c}{2}}(X[2n-1] + M[2n-1] - \alpha M[2n-2])^2\right)\right. \\ & \left.\left(\sum_{m=1}^{\frac{D_c}{2}}(M[2m] - M[2m-1])^2\right)\right]. \end{aligned} \tag{B.3}$$

Expanding this product and taking expectations we get a total of $\left(\frac{D_c}{2}\right)^2$ terms. The terms can be grouped into two groups of the form $\mathbb{E}\left[(X[3] + M[3] - \alpha M[2])^2(M[2] - M[1])^2\right]$, and $\mathbb{E}\left[(X[3] + M[3] - \alpha M[2])^2(M[6] - M[5])^2\right]$. The first group has exactly one $M[\cdot]$ term in common, while the second group has no $M[\cdot]$ terms in common. By standard computations, we can see that the expectation of each term in group one is $(2P)\sigma^2 + \sigma^4 + 2\alpha^2\sigma_m^4$ and the expectation of each term in group two is $(2P)\sigma^2 + \sigma^4$. The number of terms in group one is $(D_c - 2)$ and the number of terms in group two is $\left(\frac{D_c}{2}\right)^2 - (D_c - 2)$. Using this in (B.3), we get

$$\begin{aligned} \mathbb{E}[S_1S_2] &= (D_c - 2)[(2P)\sigma^2 + \sigma^4 + 2\alpha^2\sigma_m^4] \\ &\quad + \left[\left(\frac{D_c}{2}\right)^2 - (D_c - 2)\right][(2P)\sigma^2 + \sigma^4] \\ \Rightarrow \mathrm{Cov}(G_1, G_2) &= K_1^2\mathrm{Cov}(S_1, S_2) = \left[\frac{4\alpha^2\sigma_m^4}{(P+\sigma^2)^2}\right]\left[\frac{D_c - 2}{D_c}\right]. \end{aligned}$$

B.3 Proof of Lemma 3.7

Let $W[n] = M[n] - \alpha M[n-1]$, where $M[n]$ is an *iid* Gaussian noise process with variance $\sigma_m^2 \in (\frac{\nu}{\rho}\sigma_n^2, \frac{\rho}{\nu}\sigma_n^2)$. The noise process $\{W[n]\}$ certainly has a distribution in $\widehat{\mathcal{W}}_{\rho,\nu}$ if

$$\frac{1}{\nu} \leq |H(f)|^2 \leq \nu, \quad \forall f \in [-\frac{1}{2}, \frac{1}{2}]. \tag{B.4}$$

Here $H(f) = 1 - \alpha e^{j2\pi f}$, is the DTFT of the discrete time filter with impulse response $h[0] = 1, h[1] = -\alpha$. Thus, we have $|H(f)|^2 = 1 - 2\alpha\cos(2\pi f) + \alpha^2$. Clearly, we have $(1-|\alpha|)^2 \leq |H(f)|^2 \leq (1+|\alpha|)^2$. Using this in (B.4), we conclude that $\{W[n]\}$ has a distribution in $\widehat{\mathcal{W}}_{\rho,\nu}$ if

$$\begin{aligned} \frac{1}{\nu} \leq (1-|\alpha|)^2 \leq (1+|\alpha|)^2 \leq \nu \\ \Leftrightarrow |\alpha| \leq 1 - \frac{1}{\sqrt{\nu}} =: \alpha_{max}. \end{aligned}$$

Appendix C

Proofs for Chapter 4

C.1 Proof of Theorem 4.1

From Equations (4.9) and (4.12),

$$P_{FA}(G) = \mathcal{P}\left(T_1 - T_2 > \lambda | \mathcal{H}_0\right) \tag{C.1}$$

By the definition of T_1 in (4.8),

$$\begin{aligned}
\mathbb{E}[T_1|\mathcal{H}_0] &= \frac{1}{G}\sum_{i=1}^{G}\mathbb{E}[B_i \cdot \kappa(\mathbf{Y}_i)|\mathcal{H}_0] \\
&= \frac{1}{G}\sum_{i=1}^{G}\mathbb{E}[B_i \cdot \kappa(\mathbf{W}_i)] \\
&= \frac{1}{2}\left[\frac{1}{G}\sum_{i=1}^{G}\kappa(\mathbf{W}_i)\right] = \frac{\sigma^2}{2},
\end{aligned} \tag{C.2}$$

where $\mathbf{W}_i = (W[(i-1)K+1], \cdots, W[(i-1)K+K])$. Similarly, we can show that

$$\mathbb{E}[T_2|\mathcal{H}_0] = \frac{1}{2}\left[\frac{1}{G}\sum_{i=1}^{G}\kappa(\mathbf{W}_i)\right] = \frac{\sigma^2}{2}. \tag{C.3}$$

From (C.2) and (C.3), we have $\mathbb{E}[T_1 - T_2|\mathcal{H}_0] = 0$. So, from Chebyshev's inequality [86] we get

$$\mathcal{P}\left(|T_1 - T_2| > \lambda|\mathcal{H}_0\right) \leq \frac{\mathrm{Var}[T_1 - T_2|\mathcal{H}_0]}{\lambda^2}. \tag{C.4}$$

Using (C.4) in (C.1),

$$\begin{aligned} P_{FA}(G) &\leq \mathcal{P}\left(|T_1 - T_2| > \lambda | \mathcal{H}_0\right) \\ &\leq \frac{\mathrm{Var}[T_1 - T_2 | \mathcal{H}_0]}{\lambda^2}. \end{aligned} \tag{C.5}$$

We now compute the expressions under hypothesis $\mathcal{H}_1$.

$$\begin{aligned} \mathbb{E}[T_1|\mathcal{H}_1] &= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G} B_i \cdot \kappa(\mathbf{Y}_i)|\mathcal{H}_1\right] \\ &= \left[\frac{1}{G}\sum_{i=1}^{G} \mathbb{E}[B_i \cdot \kappa(\mathbf{Y}_i)|\mathcal{H}_1]\right]. \end{aligned} \tag{C.6}$$

The expectation on the RHS of (C.6) can be simplified as

$$\begin{aligned} & \mathbb{E}[B_i \cdot \kappa(\mathbf{Y}_i)|\mathcal{H}_1] \\ = \; & \mathbb{E}\left[B_i \cdot \frac{1}{K-L+1}\sum_{k=L}^{K} |Y[(i-1)K+k]|^2 \big| \mathcal{H}_1\right] \\ = \; & \frac{1}{K-L+1}\sum_{k=L}^{K} \mathbb{E}\left[B_i |Y[(i-1)K+k]|^2 \big| \mathcal{H}_1\right]. \end{aligned} \tag{C.7}$$

Again, the expectation within the summation in the above equation can be simplified using $\tilde{k} = (i-1)K + k$ as

$$\begin{aligned} & \mathbb{E}\left[B_i |Y[\tilde{k}]|^2 \big| \mathcal{H}_1\right] \\ = \; & \mathbb{E}\left[B_i \left|\sum_{l=0}^{L-1} h_l[\tilde{k}]X[\tilde{k}-l] + W[\tilde{k}]\right|^2\right] \\ = \; & \frac{1}{2}\mathbb{E}\left[\left|\sum_{l=0}^{L-1} h_l[\tilde{k}]X[\tilde{k}-l] + W[\tilde{k}]\right|^2 \big| B_i = 1\right] \\ \overset{(a)}{=} \; & \frac{1}{2}\left[\left((P+\epsilon)\sum_{l=0}^{L-1} |h_l[\tilde{k}]|^2\right) + |W[\tilde{k}]|^2\right]. \end{aligned} \tag{C.8}$$

In the above chain of equalities, (a) follows from the fact that $X[n] \sim \mathcal{N}(0, P+\epsilon)$ when $B_i = 1$. Substituting (C.8) in (C.7), and substituting the result in (C.6), we get

$$\begin{aligned}\mathbb{E}[T_1|\mathcal{H}_1] &\overset{(b)}{=} \left[\frac{1}{G}\sum_{i=1}^{G}\left(\frac{1}{K-L+1}\sum_{\widetilde{k}}\frac{1}{2}\left\{\sum_{l=0}^{L-1}(P+\epsilon)|h_l[\tilde{k}]|^2 + |W[\tilde{k}]|^2\right\}\right)\right] \\ &= \frac{1}{G}\sum_{i=1}^{G}\left(\frac{P+\epsilon}{2}\kappa(\mathbb{H}_i) + \kappa(\mathbf{W}_i)\right), \end{aligned} \tag{C.9}$$

where $\widetilde{k}$ in the second summation in (b) ranges from $(i-1)K+L$ to $i \cdot K$. Similarly, we can also show that

$$\mathbb{E}[T_2|\mathcal{H}_1] = \frac{1}{G}\sum_{i=1}^{G}\left(\frac{P-\epsilon}{2}\kappa(\mathbb{H}_i) + \kappa(\mathbf{W}_i)\right) \tag{C.10}$$

From (C.9) and (C.10), we get

$$\begin{aligned}\mathbb{E}[T_1 - T_2|\mathcal{H}_1] &= \epsilon \cdot \left(\frac{1}{G}\kappa(\mathbb{H}_i)\right) \\ &= \epsilon. \end{aligned} \tag{C.11}$$

Again, using Chebyshev's inequality, we get

$$\mathcal{P}\left(|(T_1 - T_2) - \epsilon| > \tilde{\lambda}\big|\mathcal{H}_1\right) \leq \frac{\mathrm{Var}[T_1 - T_2|\mathcal{H}_1]}{\tilde{\lambda}^2}, \tag{C.12}$$

where $\tilde{\lambda} = \epsilon - \lambda$. Now, the probability of mis-detection is given by

$$\begin{aligned}P_{MD}(G) &= \mathcal{P}\left(T_1 - T_2 < \lambda|\mathcal{H}_1\right) \\ &= \mathcal{P}\left((T_1 - T_2) - \epsilon < \lambda - \epsilon|\mathcal{H}_1\right) \\ &= \mathcal{P}\left((T_1 - T_2) - \epsilon < -\tilde{\lambda}|\mathcal{H}_1\right) \\ &\leq \mathcal{P}\left(|(T_1 - T_2) - \epsilon| > \tilde{\lambda}|\mathcal{H}_1\right) \\ &\leq \frac{\mathrm{Var}[T_1 - T_2|\mathcal{H}_1]}{\tilde{\lambda}^2} \end{aligned} \tag{C.13}$$

From (C.4) and (C.12), it follow that $P_{FA}(G)$ and $P_{MD}(G)$ can be made arbitrarily small if $\mathrm{Var}[T_1 - T_2|\mathcal{H}_0]$ and $\mathrm{Var}[T_1 - T_2|\mathcal{H}_1]$ decrease to zero as G increases to ∞. This is proved in the following Lemma.

Lemma C.1. *For all $G > G_0$,*

$$\begin{aligned} Var[T_1 - T_2|\mathcal{H}_0] &\leq \lambda_1 \cdot G^{-1} \\ Var[T_1 - T_2|\mathcal{H}_1] &\leq \lambda_2 \cdot G^{-1}, \end{aligned} \tag{C.14}$$

where λ_1 and λ_2 are constants dependent only on C_i, $i = 1, 2, 3$.

Proof. The proof of this lemma is straightforward and is omitted here. □

Using the bounds in Lemma C.1, we have

$$\begin{aligned} P_{FA}(G) &\leq \frac{\lambda_1}{G \cdot \lambda^2}, \\ P_{MD}(G) &\leq \frac{\lambda_2}{G \cdot (\epsilon - \lambda)^2}, \qquad \forall \ G > G_0. \end{aligned} \tag{C.15}$$

Appendix D

Proofs for Chapter 5

D.1 Proof of Theorem 5.3

In the single-quantile model, the fear of harmful interference (F_{HI}) is then given by

$$F_{HI}^{sq} = \sup_{0\leq r\leq r_n} \sup_{F_r\in\mathbb{F}_r} \mathbb{E}_{F_r}\left[\mathcal{P}(T(\mathbf{Y}) < \lambda | P = p)\right]. \tag{D.1}$$

For a given received power realization $P = p$, the CLT gives us [66]

$$T(\mathbf{Y}) \sim \mathcal{N}\left(10^{\frac{p}{10}} + \sigma_w^2, \frac{2}{N}[10^{\frac{p}{10}} + \sigma_w^2]^2\right).$$

Therefore,

$$\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = p\right) = 1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{p}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{p}{10}} + \sigma_w^2)}\right). \tag{D.2}$$

So,

$$\begin{aligned} F_{HI}^{sq} &= \sup_{0\leq r\leq r_n} \sup_{F_r\in\mathbb{F}_r} \mathbb{E}_{F_r}\left[\mathcal{P}(T(\mathbf{Y}) < \lambda | P = p)\right] \\ &\overset{(a)}{=} \sup_{0\leq r\leq r_n} \left[\beta\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = -\infty\right) + (1-\beta)\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = \gamma(r,\beta)\right)\right] \\ &\overset{(b)}{=} \left[\beta\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = -\infty\right) + (1-\beta)\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = \gamma(r_n,\beta)\right)\right]. \end{aligned} \tag{D.3}$$

The equality in (a) follows from the fact that the optimizing distribution $F_r^* \in \mathbb{F}_r$ has a mass

of β at $-\infty$ and has a mass of $(1-\beta)$ at $\gamma(r,\beta)$. This is because, from (D.2) we can see that $\mathcal{P}\left(T(\mathbf{Y}) < \lambda | P = p\right)$ is a monotonically decreasing function of p. The Equality in (b) reveals that the worst-case location corresponds to $r = r_n$. This follows from these observations. Firstly, $\mathcal{P}(T(\mathbf{Y}) < \lambda | P = -\infty)$ is independent of r (use $p = -\infty$ in (D.2)). Secondly, $\mathcal{P}(T(\mathbf{Y}) < \lambda | P = \gamma(r,\beta))$ monotonically increases with decreasing $\gamma(r,\beta)$, whereas $\gamma(r,\beta)$ monotonically decreases with increasing r (recall, $\gamma(r,\beta) = \sigma\mathcal{Q}^{-1}(1-\beta) + \mu(r)$, and $\mu(r)$ decreases with increasing r). So,

$$F_{HI}^{sq} = \beta \left[1 - \mathcal{Q}\left(\frac{\lambda - \sigma_w^2)}{\sqrt{\frac{2}{N}}\sigma_w^2}\right)\right] + (1-\beta)\left[1 - \mathcal{Q}\left(\frac{\lambda - (10^{\frac{\gamma(r_n,\beta)}{10}} + \sigma_w^2)}{\sqrt{\frac{2}{N}}(10^{\frac{\gamma(r_n,\beta)}{10}} + \sigma_w^2)}\right)\right], \tag{D.4}$$

where $\gamma(r_n,\beta) = \sigma\mathcal{Q}^{-1}(1-\beta) + \mu(r_n)$.

D.2 Proof of Theorem 5.4

Proof of (A): The fear of harmful interference is given by

$$\begin{aligned} F_{HI}^{per,ck} &:= \sup_{0\le r\le r_n} \mathcal{P}(P < \widetilde{\lambda}) \\ &\overset{(a)}{=} \sup_{0\le r\le r_n}\left[1 - \mathcal{Q}\left(\frac{\widetilde{\lambda} - \mu(r)}{\sigma}\right)\right] \\ &\overset{(b)}{=} 1 - \mathcal{Q}\left(\frac{\widetilde{\lambda} - \mu(r_n)}{\sigma}\right), \end{aligned} \tag{D.5}$$

where (a) follows from the fact that $P \sim \mathcal{N}(\mu(r), \sigma^2)$ and (b) follows from the monotonicity of the $\mathcal{Q}(\cdot)$ function. Similarly, the probability of finding a hole is given by

$$\begin{aligned} P_{FH}(r) &= \mathcal{P}(P < \widetilde{\lambda}) \\ &= 1 - \mathcal{Q}\left(\frac{\widetilde{\lambda} - \mu(r)}{\sigma}\right), \end{aligned}$$

Substituting the above expression for $P_{FH}(r)$ in (5.4) gives us the desired result.

Proof of (B): Under the single-quantile model for $\mathbb{F}_r$, the fear of harmful interference is given by

$$F_{HI}^{per,sq} := \sup_{0\le r\le r_n}\sup_{F_r\in\mathbb{F}_r} \mathcal{P}(P < \widetilde{\lambda}). \tag{D.6}$$

We know that $\mathcal{P}(P < \gamma(r,\beta)) = \beta$. If $\gamma(r,\beta) \geq \widetilde{\lambda}$ then clearly the distribution $F_r^* \in \mathbb{F}_r$, maximizing $\sup_{F_r \in \mathbb{F}_r} \mathcal{P}(P < \widetilde{\lambda})$ has mass β at $-\infty$ and hence $\sup_{F_r \in \mathbb{F}_r} \mathcal{P}(P < \widetilde{\lambda}) = \beta$. On the other hand if $\gamma(r,\beta) < \widetilde{\lambda}$ then the distribution having mass of β at $-\infty$ and mass of $(1-\beta)$ at $\gamma(r,\beta)$ belongs to the set $\mathbb{F}_r$ and for this distribution $\mathcal{P}(P < \widetilde{\lambda}) = 1$. Therefore,

$$\sup_{F_r \in \mathbb{F}_r} \mathcal{P}(P < \widetilde{\lambda}) = \begin{cases} \beta & \text{if} \quad \widetilde{\lambda} \leq \gamma(r,\beta) \\ 1 & \text{otherwise.} \end{cases} \tag{D.7}$$

Note that $\gamma(r,\beta) := \sigma \mathcal{Q}^{-1}(1-\beta) + \mu(r)$ is monotonically decreasing in r. Using this fact in (D.6), we get

$$F_{HI}^{per,sq} := \sup_{0 \leq r \leq r_n} \sup_{F_r \in \mathbb{F}_r} \mathcal{P}(P < \widetilde{\lambda}) = \begin{cases} \beta & \text{if} \quad \widetilde{\lambda} \leq \gamma(r_n,\beta) \\ 1 & \text{otherwise.} \end{cases}$$

Bibliography

[1] R. H. Coase, "The Federal Communications Commission," *The Journal of Law and Economics*, vol. 2, pp. 1–40, Oct. 1959.

[2] NTIA, "US frequency allocation chart," 2003. [Online]. Available: http://www.ntia.doc.gov/osmhome/allochrt.html

[3] R. W. Broderson, A. Wolisz, D. Cabric, S. M. Mishra, and D. Willkomm, "White paper: CORVUS: A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum," University of California, Berkeley, Tech. Rep., 2004. [Online]. Available: http://bwrc.eecs.berkeley.edu/Research/MCMA/CR_White_paper_final1.pdf

[4] M. McHenry, "NSF spectrum occupancy measurements project summary," Shared Spectrum Company, Tech. Rep., Aug. 2005. [Online]. Available: www.sharedspectrum.com/inc/content/measurements/nsf/NSF_Project_Summary.pdf

[5] S. Shellhammer, A. Sadek, and W. Zhang, "Technical challenges for Cognitive Radio in the TV white space spectrum," in *Information Theory and Applications Workshop*, Feb. 2009.

[6] A. Sahai, K. Woyach, G. Atia, and V. Saligrama, "A technical framework for light-handed regulation of cognitive radios - [topics in radio communications]," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. 96–102, Mar. 2009.

[7] A. S. D. Vany, R. D. Eckert, C. J. Meyers, D. J. O'Hara, and R. C. Scott, "A property system for market allocation of the electromagnetic spectrum: A legal-economic-engineering study," *Stanford Law Review*, vol. 21, pp. 1499–1561, June 1969.

[8] E. Noam, "Taking the next step beyond spectrum auctions — open spectrum access," *IEEE Commun. Mag.*, vol. 33, pp. 66–73, Dec. 1995.

[9] Y. Benkler, "Overcoming Agoraphobia: Building the commons of the digitally networked environment," *Harvard Journal of Law and Technology*, vol. 11, pp. 287–400, Winter 1998.

[10] K. Woyach, "Crime and punishment for cognitive radios," Master's thesis, University of California, Berkeley, 2008.

[11] J. Mitola and G. Q. J. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[12] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," PhD Dissertation, Royal Institute of Technology, Stockholm, Sweden, May 2000.

[13] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 201–220, Feb. 2005.

[14] ——, "Cognitive dynamic systems," *Proc. IEEE*, vol. 94, pp. 1910–1911, Nov. 2006.

[15] Y. Zhao, S. Mao, J. O. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodology," *Proc. IEEE*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[16] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, "XG dynamic spectrum access field test results [topics in radio communications]," in *Communications Magazine, IEEE*, vol. 45, no. 6, Toronto, Ont., Canada,, June 2007, pp. 51–57.

[17] "Draft standrad for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Proceedings for Operation in the TV Bands," IEEE P802.22 Draft Standard, Tech. Rep. D0.3, May 2007.

[18] D. Grandblaise, C. Kloeck, K. Moessner, V. Rodriguez, E. Mohyeldin, M. K. Pereirasamy, J. Luo, and I. Martoyo, "Techno - economic of collaborative based secondary spectrum usage - e2r research project outcomes overview," in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, Baltimore, MD, USA,, Nov. 8–11, 2005, pp. 318–327.

[19] "Spectrum policy task force report," Federal Communications Commision, Tech. Rep. 02-135, Nov 2002. [Online]. Available: http://hraunfoss.fcc.gov/edocs$_$public/attachmatch/DOC-228542A1.pdf

[20] *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005.

[21] *Second IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Apr. 2007.

[22] *Third IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Oct. 2008.

[23] *Cognitive Radio Part 1: Practical Perspectives.* Proceedings of the IEEE, April 2009, vol. 97, no. 4.

[24] *Cognitive Radio Part 2: Fundamental Issues.* Proceedings of the IEEE, May 2009, vol. 97, no. 5.

[25] *Signal Processing and Networking for Dynamic Spectrum Access.* IEEE Journal of Selected Topics in Signal Processing, February 2008, vol. 2, no. 1.

[26] "In the Matter of Unlicensed Operation in the TV Broadcast Bands: Second Report and Order and Memorandum Opinion and Order," Federal Communications Commision, Tech. Rep. 08-260, Nov. 2008. [Online]. Available: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-260A1.pdf

[27] S. M. Mishra, R. Tandra, and A. Sahai, "Coexistence with primary users of different scales," in *Proc. of 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Dublin, Ireland, Apr. 2007, pp. 158–167.

[28] R. S. Dhillon and T. X. Brown, "Models for analyzing cognitive radio interference to wireless microphones in TV bands," in *3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Chicago, IL, Oct. 14–17, 2008, pp. 1–10.

[29] S. M. Mishra, R. Tandra, and A. Sahai, "The case for multiband sensing," in *Forty-fifth Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2007.

[30] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Wideband Spectrum Sensing in Cognitive Radio Networks," in *Proc. of the IEEE International Conference on Communications (ICC)*, May 2008, pp. 19–23.

[31] S. Mishra and A. Sahai, "How much white space is there?" Department of Electrical Engineering, University of California, Tech. Rep. UCB/EECS-2009-3, Jan. 2009. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-3.html

[32] N. Hoven and A. Sahai, "Power scaling for cognitive radio," in *Proc. of the WirelessCom 05 Symposium on Signal Processing*, Maui, HI, June 13–16 2005.

[33] N. Hoven, "On the feasibility of cognitive radio," Master's thesis, University of California, Berkeley, 2005.

[34] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *Proc. of 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Baltimore, USA, Nov. 2005, pp. 124–130.

[35] A. Sahai, N. Hoven, and R. Tandra, "Some fundamental limits on cognitive radio," in *Forty-second Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2004, pp. 1662–1671.

[36] M. Mishra and A. Sahai, "How much white space has the fcc opened up," *IEEE Commun. Lett.*, Jan. 2010.

[37] A. Sahai, S. M. Mishra, R. Tandra, and K. Woyach, "Cognitive radios for spectrum sharing," *IEEE Signal Processing Mag.*, vol. 26, no. 1, pp. 140–145, Jan. 2009.

[38] US Census Bureau, "US census 2000 Gazetteer files," Tech. Rep. [Online]. Available: http://www.census.gov/geo/www/gazetteer/places2k.html

[39] K. Harrison, M. Mishra, and A. Sahai, "How much white-space capacity is there?" in *Submitted to the IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2010.

[40] S. A. Jafar, S. Srinivasa, I. Maric, and A. Goldsmith, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[41] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1813–1827, May 2006.

[42] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," in *IEEE International Symposium on Information Theory*, Seattle, USA, July 2006, pp. 2413–2417.

[43] P. Grover and A. Sahai, "On the need for knowledge of the phase in exploiting known primary transmissions," in *Proc. of 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Dublin, Ireland, Apr. 2007, pp. 462–471.

[44] K. Eswaran, M. Gastpar, and K. Ramchandran, "Bits through ARQs: Spectrum sharing with a primary packet system," in *IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 2171–2175.

[45] K. Eswaran, "Communication and third parties: Costs, cues, and confidentiality," PhD Dissertation, University of California, Berkeley, California, USA, Oct. 2009.

[46] C. R. Stevenson, C. Cordeiro, E. Sofer, and G. Chouinard, "Functional requirements for IEEE 802.22 WRAN standard," Tech. Rep., Sept. 2005. [Online]. Available: https://mentor.ieee.org/802.22/file/05/22-05-0007-48-0000-draft-wran-rqmts-doc.doc

[47] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory.* Prentice-Hall, 1998, vol. 2.

[48] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, pp. 523–531, Apr. 1967.

[49] R. Tandra, "Fundamental limits on detection in low SNR," Master's thesis, University of California, Berkeley, 2005.

[50] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer, March 1998.

[51] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer, 1998.

[52] F. R. Hampel, "A general qualitative definition of robustness," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, 1971.

[53] P. J. Huber and V. Strassen, "Minimax tests and the neyman-pearson lemma for capacities," *The Annals of Statistics*, vol. 1, no. 2, pp. 251–263, 1973.

[54] A. El-Sawy and V. VandeLinde, "Robust detection of known signals," *IEEE Trans. Inform. Theory*, vol. 23, no. 6, pp. 722–727, Nov. 1977.

[55] H. Poor and J. Thomas, "Asymptotically robust quantization for detection," *IEEE Trans. Inform. Theory*, vol. 24, no. 2, pp. 222–229, Mar. 1978.

[56] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, Mar. 1985.

[57] V. Veeravalli, T. Basar, and H. V. Poor, "Minimax robust decentralized detection," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 35–40, Jan. 1994.

[58] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge, United Kingdom: Cambridge University Press, 2005.

[59] W. A. Gardner, "Signal interception: A unifying theoretical framework for feature detection," *IEEE Trans. Commun.*, vol. 36, pp. 897–906, Aug. 1988.

[60] K. Kim, I. Akbar, K. Bae, J. S. Um, C. Spooner, and J. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Apr. 2007, pp. 212–215.

[61] P. Sutton and K. N. L. Doyle, "Cyclostationary signatures for rendezvous in OFDM-based dynamic spectrum access networks," in *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Apr. 2007, pp. 220– 231.

[62] R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one?" *Proc. IEEE*, vol. 97, pp. 824–848, May 2009.

[63] A. Sahai, R. Tandra, and M. Mishra, *Cognitive Radios: System Design Perspective*, 2009, ch. Spectrum Sensing: Fundamental Limits. [Online]. Available: http://www.eecs.berkeley.edu/~sahai/Papers/SensingChapter.pdf

[64] S. M. Mishra, "Maximizing available spectrum for cognitive radios," PhD Dissertation, University of California, Berkeley, California, USA, Dec. 2009.

[65] Y. Li, "Signal processing issues in cognitive radio," *Proc. IEEE*, 2008.

[66] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, pp. 4 – 17, Feb. 2008.

[67] M. Ghosh, "Text on FFT-based pilot sensing - For Informative Annex on Sensing Techniques," IEEE 802.22 Meeting Documents, July 2007.

[68] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy and collaborative detection," in *Military Communications Conference*, Washington, DC, Oct. 2006, pp. 1–7.

[69] S. Shellhammer and R. Tandra, "An evaluation of DTV pilot power detection," in *IEEE 802.22-06/0188r0*, Sept. 2006.

[70] R. Tandra and A. Sahai, "Noise calibration, delay coherence and SNR walls for signal detection," in *Prof. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Chicago, IL, Oct. 2008.

[71] S. Haykin, J. Reed, and D. Thomson, "Spectrum sensing in Cognitive Radio," *Proc. of the IEEE*, 2008.

[72] R. Tandra and A. Sahai, "SNR walls for feature detectors," in *Proc. of 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Dublin, Ireland, Apr. 2007, pp. 559–570.

[73] G. Trukenich, V. Gaddam, and M. Ghosh, "Text on Dual FPLL pilot sensing - For Informative Annex on Sensing Techniques," IEEE 802.22 Meeting Documents, July 2007.

[74] Y. Zeng and Y.-C. Liang, "Text on eigenvalue based sensing - For Informative Annex on Sensing Techniques," IEEE 802.22 Meeting Documents, July 2007.

[75] ——, "Maximum-Minimum Eigenvalue Detection for Cognitive Radio," in *Proc. of the IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC)*, Sept. 2007, pp. 1–15.

[76] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[77] A. Parsa, A. A. Gohari, and A. Sahai, "Exploiting interference diversity for event-based spectrum sensing," in *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Chicago, IL, 2008.

[78] D. Cabric, "Cognitive radios: System design perspective," PhD Dissertation, University of California Berkeley, California, USA, 2006.

[79] R. Tandra and A. Sahai, "Fundamental limits on detection in low SNR under noise uncertainty," in *International Conference on Wireless Networks, Communications and Mobile Computing*, June 2005, pp. 464–469.

[80] R. Tandra, "Fundamental limits on detection in low SNR," Master's thesis, University of California, Berkeley, 2005. [Online]. Available: http://www.eecs.berkeley.edu/~tandra/pub1.htm

[81] A. Sonnenschein and P. M. Fishman, "Radiometric detection of spread-spectrum signals in noise of uncertain power," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 654–660, July 1992.

[82] R. Tandra and A. Sahai, "Fundamental limits on detection in low SNR under noise uncertainty," in *Proc. of the WirelessCom 05 Symposium on Signal Processing*, Maui, HI, June 13–16 2005.

[83] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation," in the proceedings of IEEE MILCOM 2006.

[84] P. J. Huber, *Robust Statistics*. Wiley-Interscience, February 1981.

[85] ——, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, pp. 1753–1758, 1965.

[86] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Belmont: Duxbury Press, 2004.

[87] D. Slepian, "Some comments on the detection of Gaussian signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 4, pp. 65–68, June 1958.

[88] S. Shellhammer and R. Tandra, "Performance of the power detector with noise uncertainty," in *IEEE 802.22-06/0134r0*, July 2006.

[89] M. Talagrand, "A new look at independence," *The Annals of Probability*, vol. 24, pp. 1–34, 1996.

[90] K. A. Woyach, A. Sahai, G. Atia, and V. Saligrama, "Crime and punishment for cognitive radios," in *Forty-sixth Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2008.

[91] *ATSC Digital Television Standard*, ATSC Std. A/53, 2007. [Online]. Available: "http://www.atsc.org/standards.html"

[92] S. Mathur and S. Shellhammer, "An evaluation of the PN sequence based detection of DTV signals in the draft," in *IEEE 802.22-06/0189r0*, September 2006.

[93] N. Agarwal, A. Sahai, and J. Tsitsiklis, "Narrowband noise mitigation in location-determining signal processing," U.S. Patent 7 177 614, February, 2007.

[94] W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[95] ——, "Spectral correlation of modulated signals, part i-analog modulation," *IEEE Trans. Commun.*, vol. COM-35, pp. 584–594, June 1987.

[96] W. A. Gardner, W. A. Brown, and C. K. Chen, "Spectral correlation of modulated signals, part ii-digital modulation," *IEEE Trans. Commun.*, vol. COM-35, pp. 595–601, June 1987.

[97] W. A. Gardner, *Introduction to Random Processes with Applications to Signals and Systems.* New York: Macmillan, 1985.

[98] ——, "The spectral correlation theory of cyclostationary time-series," *Signal Processing*, vol. 11, pp. 13–36, July 1986.

[99] A. Sahai and D. Cabric, "Cyclostationary feature detection," Tutorial presented at DySpAN, Baltimore, MD, oct 2005. [Online]. Available: http://www.eecs.berkeley.edu/~sahai/Presentations/DySPAN05_part2.ppt

[100] Y. Sun, Y. Liu, and X. Tan, "Spectrum sensing for cognitive radio based on higher-order statistics," in *4th International Conference on Wireless Communications, Networking and Mobile Computing*, Oct. 2008, pp. 1–4.

[101] X. Chen and S. Nagaraj, "Entropy-based spectrum sensing in cognitive radio," in *Wireless Telecommunications Symposium*, Apr. 2008, pp. 57–61.

[102] R. Tandra and A. Sahai, "Overcoming SNR walls through macroscale features," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, Sept. 23–26, 2008, pp. 583–590.

[103] L. Franks, "Carrier and bit synchronization in data communication–a tutorial review," *IEEE Trans. Commun.*, vol. 28, pp. 1107–1121, Aug. 1980.

[104] S. Gupta, "Phase-locked loops," *Proc. IEEE*, vol. 63, pp. 291–306, Feb. 1975.

[105] W. Lindsey and C. M. Chie, "A survey of digital phase-locked loops," *Proc. IEEE*, vol. 69, pp. 410–431, Apr. 1981.

[106] J. W. M. Bergmans, "Efficiency of data-aided timing recovery techniques," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1397–1408, Sept. 1995.

[107] A. Vitterbi, *Principles of Coherent Communication*. McGraw-Hill, 1966.

[108] F. M. Gardner, *Phaselock Techniques*. Wiley, 1979.

[109] W. Lindsey, *Synchronization Systems in Communications*. Prentice-Hall, 1972.

[110] W. Lindsey and M. Simon, *Telecommunication Systems Engineering*. Prentice-Hall, 1973.

[111] H. Meyr and G. Ascheid, *Synchronization in Digital Communications*. Wiley Interscience, 1990.

[112] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.

[113] E. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fadingchannels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1384–1400, July 2000.

[114] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low-SNR capacity of noncoherent fading channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 1555–1574, Apr. 2009.

[115] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3102–3120, Sept. 2005.

[116] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high-SNR rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3390–3396, Dec. 2004.

[117] L. B. David Blackwell and A. J. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, pp. 558–567, Sept. 1960.

[118] A. N. P. Lapidoth, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.

[119] I. Csiszar and P. Narayan, "Capacity of the gaussian arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 37, pp. 18–26, 1991.

[120] A. D. Sarwate, "Robust and adaptive communication under uncertain interference," PhD Dissertation, University of California Berkeley, California, USA, 2008.

[121] G. Atia, A. Sahai, and V. Saligrama, "Spectrum enforcement and liability assignment in cognitive radio systems," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, Chicago, IL, Oct. 2008, pp. 1–12.

[122] C. Sagan, *Cosmos*. Random House, 1980.

[123] ——, *Communication with Extraterrestrial Intelligence*. MIT Press, 1972.

[124] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[125] R. Gallager, *Information Theory and Reliable Communication*. Wiley, January 1968.

[126] R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one: extended version," University of California, Berkeley, Berkeley, CA, Tech. Rep., Aug. 2008.

[127] Q. Zhao and A. Swami, "A decision-theoretic framework for opportunistic spectrum access," *IEEE Wireless Communications Magazine: Special Issue on Cognitive Wireless Networks*, pp. 14–20, Aug. 2007.

[128] Q. Z. L. T. A. Swami and Y. Chen, "Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: A pomdp framework," *IEEE Journal on Selected Areas in Communications (JSAC): Special Issue on Adaptive, Spectrum Agile and Cognitive Wireles Networks*, pp. 589–600, Apr. 2007.

[129] J. Unnikrishnan and V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *Submitted to IEEE Trans. Signal Processing*, 2008.

[130] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access," *IEEE Trans. Inform. Theory*, vol. 55, pp. 4040–4050, Sept. 2009.

[131] J. Unnikrishnan and V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, pp. 18 – 27, Feb. 2008.

[132] S. J. Shellhammer, S. N. Shankar, R. Tandra, and J. Tomcik, "Performance of power detector sensors of DTV signals in IEEE 802.22 WRANs," in *First International Workshop on Technology and Policy for Accessing Spectrum*, June 2006.

[133] S. Huang, X. Liu, and Z. Ding, "On optimal sensing and transmission strategies for dynamic spectrum access," in *3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Chicago, IL, Oct. 2008, pp. 1–5.